Enchained growth and cluster dislocation : a possible mechanism for microbiota homeostasis

Florence Bansept[1,2], Kathrin Schumann-Moor[3,4], Médéric Diard[3,5], Wolf-Dietrich Hardt[3], Emma Slack[3,6], and Claude Loverdo[1,*]

[1]Laboratoire Jean Perrin, Sorbonne Université / CNRS, Paris, France.
[2]Present address: Department for Evolutionary Theory, Max Planck Institute for Evolutionary Biology, Plön, Germany
[3]Institute of Microbiology, ETH Zürich, Zürich, Switzerland.
[4]Present address: Center of Dental Medicine / Oral and Maxillofacial Surgery, University of Zürich, Switzerland
[5]Present address: Biozentrum, University of Basel, Switzerland
[5]Present address: Institute of Food, Nutrition and Health, ETH Zürich, Switzerland.
[*]Corresponding author

Abstract

Immunoglobulin A is a class of antibodies produced by the adaptive immune system and secreted into the gut lumen to fight pathogenic bacteria. We recently demonstrated that the main physical effect of these antibodies is to enchain daughter bacteria, i.e. to cross-link bacteria into clusters as they divide, preventing them from interacting with epithelial cells, thus protecting the host. These links between bacteria may break over time. We study several models using analytical and numerical calculations. We obtain the resulting distribution of chain sizes, that we compare with experimental data. We study the rate of increase in the number of free bacteria as a function of the replication rate of bacteria. Our models show robustly that at higher replication rates, bacteria replicate before the link between daughter bacteria breaks, leading to growing cluster sizes. On the contrary at low growth rates two daughter bacteria have a high probability to break apart. Thus the gut could produce IgA against all the bacteria it has encountered, but the most affected bacteria would be the fast replicating ones, that are more likely to destabilize the microbiota. Linking the effect of the immune effectors (here the clustering) with a property directly relevant to the potential bacterial pathogeneicity (here the replication rate) could avoid to make complex decisions about which bacteria to produce effectors against.

Author Summary

Inside the organism, the immune system can fight generically against any bacteria. However, the lumen of the gut is home to a very important microbiota, so the host has to find alternative ways to fight dangerous bacteria while sparing beneficial ones. While many studies have focused on the complex molecular and cellular pathways that trigger an immune response, little is known about how the produced antibodies act once secreted into the intestinal lumen. We recently demonstrated that the main physical effect of these antibodies is to cross-link bacteria into clusters as they divide, preventing them from interacting with epithelial cells, thus protecting the host. These links between bacteria may break over time. Using analytical and numerical calculations, and comparing with experimental data, we studied the dynamics of these clusters. At higher replication rates, bacteria replicate before the link between daughter bacteria breaks, leading to growing cluster sizes, and conversely. Thus the gut could produce IgA against all the bacteria it has encountered, but the most affected bacteria would be the fast replicating ones, that are more likely destabilize the microbiota. Studying the mechanisms of the immune response may uncover more such processes that enable to target properties hard to escape through evolution.

Introduction

The digestive system has a large surface area[1][2], covered by a single layer of epithelial cells, essential for nutrient absorption, but also a gateway for many pathogens. Contrary to the inside of the body, where the presence of any bacteria is abnormal, the lumen of the digestive system is home to a very important microbiota. These microbiota bacteria are present in extremely high densities[3]. Bacteria are necessary to break down and absorb certain nutrients, and can compete against potentially pathogenic intruders[4]. Inside the organism, the immune system can fight generically against any bacteria. However, in the digestive system, the host has to find alternative ways to fight dangerous bacteria while sparing beneficial ones. As closely related bacteria (e.g. *Salmonella* spp. and commensal *E. coli*) can show highly variable behaviors in the intestine, identifying which bacteria are good or bad is challenging. Besides, the overgrowth of any type of bacteria, even those that do not cause acute pathology, can impair the functionality of the microbiota. Thus the host needs mechanisms to maintain the gut microbiota homeostasis.

The adaptive response is the only strong handle that the host has on directly controlling microbiota composition at the species level[5][6]. The main effector of the adaptive immune response in the digestive system is secretory IgA, an antibody. sIgA specifically bind to targets that the organism has already encountered and can be elicited by vaccination. It was observed more than 40 years ago that this prevents infection by pathogenic bacteria such as salmonella[7]. Many studies have focused on the complex molecular and cellular pathways that trigger an immune response on the host side of the digestive surface[8]. However, we are only just beginning to understand by which physical mechanisms the immune effectors act once secreted into the intestinal lumen, which are crucial for the control of both commensal and pathogenic bacteria.

The influence on bacteria dynamics of abiotic factors such as the flow in the gut has recently started being quantitatively studied [9, 10].

We have shown that mice vaccinated with inactivated *Salmonella* Typhimurium do produce specific sIgA which bind to *S.*Typhimurium, but this neither kills them nor prevents them from reproducing[11][12]. The initial colonization of the intestinal lumen by *S.*Typhimurium is in fact unchanged in either kinetics or magnitude in vaccinated animals. These mice are nevertheless protected against pathogen spread from the gut lumen to systemic sites like lymph nodes, liver or spleen. A classic idea in immunology is that, as one antibody has several binding sites, antibodies aggregate bacteria when they collide into each other. But this effect would be negligible at realistic densities of a given bacterium in the digestive system, simply due to very long typical encounter times between bacteria recognized by the same sIgA (see section 1 in appendix). We have shown that actually, the main effect is that upon replication, daughter bacteria remain attached to one another by sIgA, driving the formation of clusters derived from a single infecting bacterium[12]. This "enchained growth" is effective at any bacterial density. Clustering has physical consequences: the produced clusters do not come physically close to the epithelial cells. And as interaction with the epithelial cells is essential for *S.*Typhimurium virulence, this is sufficient to explain the observed protective effect.

If sIgA was perfectly sticky, we would expect all bacteria to be in clusters of ever increasing size. In these experiments, despite observing *S.*Typhimurium clusters in the presence of sIgA, there are still free *S.*Typhimurium, and small clusters. One possibility would be that not all bacteria are coated with sIgA. But in these experiments, it has been demonstrated that they are (extended figure 2c of[12]). Indeed, a gram of digestive content contains at most 10^{11} bacteria, and typically 50 micrograms or more of sIgA[13], of molecular mass of about 385kD. This leads to about 800 sIgA per bacteria. sIgA may not be all bound to bacteria, and sIgA for different specific antigens may be produced in proportions not matching the proportions of antigens present in the digestive system, so that not all bacteria are coated with 800 sIgA. Nevertheless, most bacteria already encountered by the organism will be coated with many sIgA, and thus the cluster size is not limited by the number of available sIgA. Another possibility is that the sIgA-mediated links break. Such breaking has been demonstrated to be dependent on the applied forces in related systems[14][15]. As there is shear in the digestive system, because mixing is needed for efficient nutrients absorption, it is plausible that links break over time.

Small clusters are linear chains of bacteria, bound by sIgA, with these links being broken over time by the forces induced by the flow. As bacteria are similar to each other, it is, at another scale, analogous to other physical systems[16], such as polymers breaking under flow[17]. The main difference is that these chains grow by bacterial replication. Growth and fragmentation are competing effects, and the modelling of these chains can be viewed as statistical physics, to predict their length distribution, whether there is a typical chain length, or if large chains of ever-increasing length dominate the distribution, and how the growth in number of free bacteria depends on the bacterial replication rate.

This could have very important biological consequences. To illustrate this point, let us consider a simplified model: bacteria remain enchained by sIgA when they grow (replication time τ_{div}), and this link between 2 bacteria breaks at a specific time τ_{break} (although this latter hypothesis is not realistic, we make

it for now for the sake of simplicity). If $\tau_{div} > \tau_{break}$, then when a bacterium divides, it forms a 2-bacteria cluster, which dislocates into 2 free bacteria before the next replication steps, so that the bacteria remain in the state of free or 2-bacteria clusters and there are no larger clusters. If $\tau_{div} < \tau_{break}$, when a bacterium divides, it forms a 2-bacteria cluster, which becomes a 4 bacteria cluster before the first link breaks, so there cannot be free bacteria. In this model, the fast-growing bacteria are selectively targeted by the action of the immune system. The immune system does not need to sense which bacteria are growing faster, it only has to produce sIgA targeted to all the bacteria it has encountered, and bacteria with $\tau_{div} > \tau_{break}$ are unaffected, whereas bacteria with $\tau_{div} < \tau_{break}$ are trapped in clusters. That could be a simple physical mechanism to target the action of the immune system to the fast-growing bacteria which are destabilizing the microbiota, and thus to preserve microbiota homeostasis.

In the following, we present different plausible models of bacteria clusters dynamics, and the methods to study them. Then we give, for each model, the resulting dynamics and chain length distribution, before putting these results in perspective with experimental data. Eventually, we discuss the results. As some biological details are unknown, studying different models enables to show which key results are robust; and differences confronted to experimental data give some indications about which are the most likely.

Models and methods

Ethics statement

All animal experiments were approved by the legal authorities (licenses 223/2010, 222/2013 and 193/2016; Kantonales Veterinäramt Zürich, Switzerland) and performed according to the legal and ethical requirements.

Experimental methods

We perform a new analysis on microscopy images that were produced for [12]. We analyzed images of cecal content in vaccinated mice for the early data points (4 and 5 hours) of experiments starting from a low inoculum (10^5), to minimize the clustering from random encounters. Further details on our analysis can be found in appendix 7, as well as a brief description of the experiments from which the images were produced.

Models and general methods

We consider low bacterial densities, so encounters between unrelated bacteria are negligible. Thus, we consider each free bacteria and each cluster of bacteria independently of the others. *Salmonella* are rod shaped bacteria, which replicate by dividing in two daughter bacteria at the middle of the longitudinal axis. Thus if the daughter bacteria remain enchained, they are linked to each other by their poles. With further bacterial replications, the cluster will then be a linear chain. This is consistent with experimental observations, in which clusters are either linear chains, with bacteria attached to one or two neighbors by their poles,

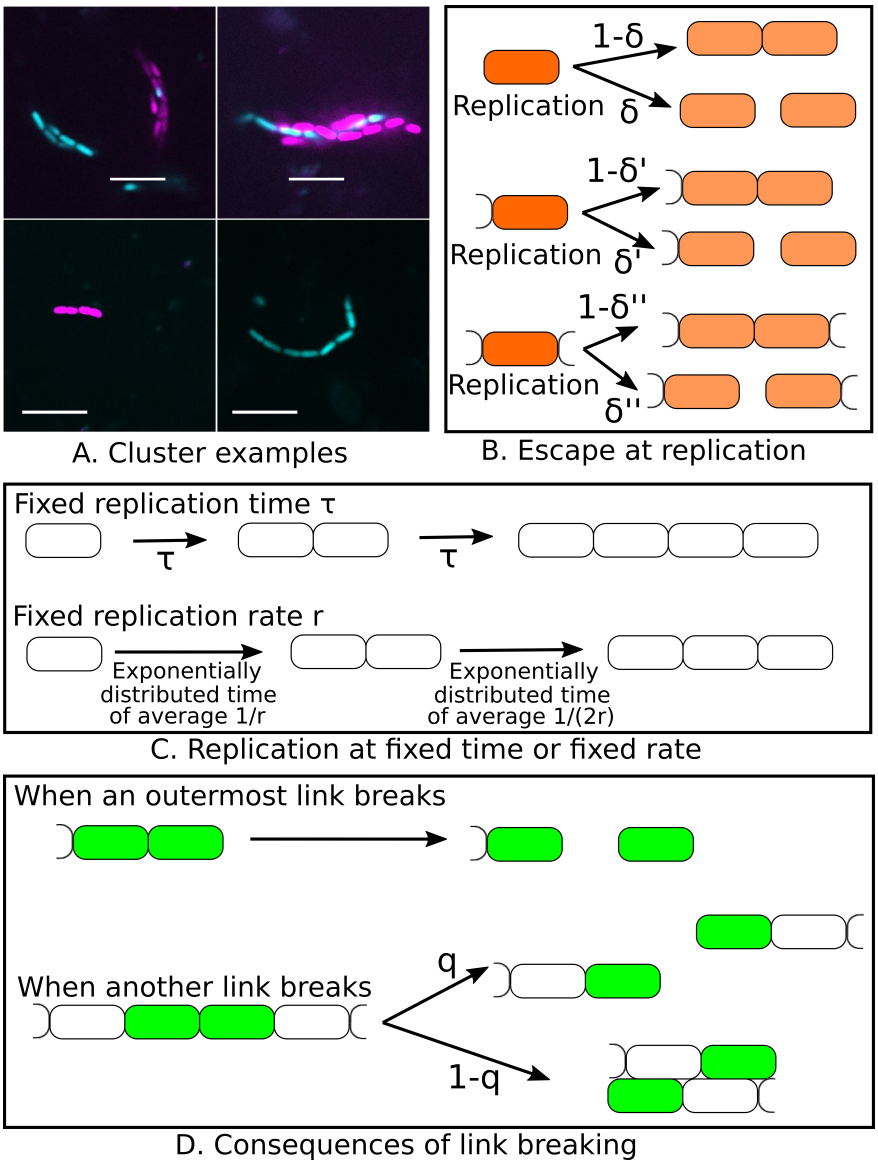


Figure 1: Bacterial cluster modeling. A. Representative experimental images of bacterial clusters in cecal content of vaccinated mouse at 5h post infection with isogenic GFP and mCherry expressing S.typhimurium (Experiments performed for [12]). The scale bar is $10\mu m$. Top images: complex clusters made from bundles of linear clusters, which could be re-linked single chains (left) or formed from at least two independent clones (indicated by fluorescence, right). Bottom images: linear clusters which dynamics we aim to model. B. Potential bacterial escape at replication (in the base model, $\delta = \delta' = \delta''$). C. Fixed replication time or fixed replication rate (the latter is chosen for the base model). D. Consequences of link breaking. In the base model, $q = 0$.

or larger clusters which seem to be formed as bundles of such linear clusters (pannel A figure 1). Our aim is to model the dynamics of these chains.

A first element is the bacterial replication (see figure 1 C). One way to model it is to assume that bacteria replicate every τ_{div}. Another way, that we will generally use, less realistic but easier for calculations, is to assume that there is a fixed replication rate r.

A second element is that when bacteria replicate, they may be able to escape enchainment (see figure 1 B), but likely with low probability (see discussion in section 2 in appendix). In general, we will take the limit with perfect enchainment upon replication ($\delta = \delta' = \delta'' = 0$).

A crucial element is the possibility for the links between bacteria to break. We usually assume that the breaking rate α is the same for all links and over time. We will also explore the case when the link breaking rate is force-dependent, in which case not all the links have the same breaking rate.

Another crucial element, is to model what happens when the chain breaks (see figure 1 D). If the subparts come in contact again at the same poles and get linked again, then this could simply be modeled by an effectively lower breaking rate. More likely, if the subparts come in contact again, they do so laterally, forming larger clusters of more complex shapes. Because in these clusters, most bacteria have more than two neighbors, and more contact surface, they are much less likely to escape. To simplify, we will consider that these clusters do not contribute anymore to releasing either free bacteria or linear chains. Thus when a link breaks, either the two subparts move sufficiently away and become two independent chains (probability q); or collide and become a more complex cluster which does not contribute anymore to both free bacteria and linear chains (probability $1 - q$). For simplicity, we consider that when an outermost link breaks, the single bacterium, more mobile, always escapes ($q_{outermost} = 1$), but that else q is length independent. The simplest values to study are either $q = 0$ or $q = 1$. As we will see, when we study the case in which q can take any value between 0 and 1, we find that the case $q = 1$ is qualitatively different from other values of q. Consequently, we will take $q = 0$ for the base model.

As digestive content leaves the digestive system, or the part of the digestive system under consideration, due to flow, we define c the loss rate of free bacteria, and c' the loss rate of chains. We assume no deat. Bacterial death would break chains. It would thus have a similar effect to a larger breaking rate α. As free bacteria have more autonomous motility, enabling them to swim towards the epithelial cells, it is likely that $c' \geq c$. We will usually take $c = c'$. Crucially, in this latter case, free bacteria, and all chains are lost at the same rate. The c value has a complex effect on stochastic quantities, such as the probability to have at least one chain of a given length. However, here we study the mean numbers of free bacteria and chains of different lengths, then the case with $c = c'$ is equivalent to $c = c' = 0$, with all numbers of bacteria and chains multiplied by $\exp(-ct)$.

We start with the most basic model, with a replication rate r, bacteria perfectly bound upon replication ($\delta = \delta' = \delta'' = 0$), a fixed breaking rate per link α, and bacterial chains always binding into a more complex cluster when a link breaks (except for the outermost links) ($q = 0$). We then study variations of the model to test the robustness of the results: with an non-zero escape probability upon replication and $c \neq c'$; with a replication time τ instead of a replication rate r; with the possibility for chains to escape when an inner link

breaks ($q > 0$); with a force-dependent breaking rate (see table below for a list of symbols).

We consider the beginning of the process, early enough so that the carrying capacity is far from reached, and thus the replication rate is constant. We do not consider generation of escape mutants which are not bound by IgA. We consider only the average numbers of free bacteria and linear chains of different lengths, and we do not count more complex clusters, as they do not contribute to free bacteria dynamics in our model. For each model, we write the equations for the derivative of these numbers with respect to time. With N the vector of the mean number of free bacteria, linear chains of length 2, 3, etc., these equations give the coefficients of the matrix M, such that $dN/dt = MN$. The results are obtained in part via analytical derivations and in part via numerical studies. The latter are obtained in Mathematica by numerically solving the eigensystem written for chains up to length n_{max}, chosen large enough not to impact the results. In the long term limit, $N(t) \to Ce^{\lambda t}P$, with C a constant, λ the largest eigenvalue of M, and P the corresponding eigenvector, normalized such that the sum of its components is equal to 1. λ is thus the long term growth rate of the free bacteria and the linear chains. For each model, we study how the growth of free bacteria – the ones which are capable of causing systemic infection[12] – which is λ in the steady state, depends on the bacterial replication rate. Besides, we obtain chain length distributions (the components p_i of P), which could be compared to experimentally observed distributions.

Recapitulation table of the symbols used

Base model	
r	Bacterial replication rate
α	Breaking rate of the link between two bacteria
$n_i(t)$	Number of linear chains of length i at t (n_1: free bacteria)
λ	Largest eigenvalue of the matrix M, which is the growth rate of the free bacteria in the steady state
P	Eigenvector of components p_i associated to λ, normalized such as $\sum_{i=0}^{\infty} p_i = 1$.
Model with bacterial escape and bacterial loss (all these parameters are taken as 0 in the base model)	
δ	When a free bacterium replicates, the probability that this will lead to 2 free bacteria
δ'	When a bacterium at the tip of a chain replicates, the probability that the daughter bacterium at the exterior side escapes
δ''	When a bacterium replicates within a chain, the probability that the daughter bacterium will not be bond to each other, resulting to the chain breaking in two
c	Loss rate for the free bacteria
c'	Loss rate for the chains
Model with fixed replication time	
τ	Time between one bacterial division and the next (the bacterial growth rate is $r_{eff} = \log(2)/\tau$)
$\mathcal{N}$	Largest eigenvalue of the matrix in this model. $\mathcal{N} = \exp(\lambda\tau)$
Model with linear chains independent after breaking	
q	Probability that when a inner link of a chain breaks, the two subparts become independent linear chains. In the base model, $q = 0$.
Model with force-dependent breaking rate	
β	A constant expressing the strength of the coupling between hydrodynamic forces and link breaking. In the base model, $\beta = 0$.

Results

Base model : replication rate, no bacteria escape upon replication, fixed breaking rate, $q = 0$

Equations

In the base model, bacteria have a replication rate r, daughters are perfectly bound upon replication, each link has a breaking rate α, and when a link which is not at a tip breaks, the resulting two chains of bacteria always bind into more complex clusters and thus do not contribute to free bacteria dynamics anymore ($q = 0$). With $n_i(t)$ the number of linear chains of length i as a function of time, (n_1 is the number of free bacteria),

$$\frac{dn_1}{dt} = -rn_1 + \sum_{i=2}^{\infty} 2\alpha n_i \tag{1}$$

and for $i \geq 2$,

$$\frac{dn_i}{dt} = rn_{i-1}(i-1) - irn_i - (i-1)n_i\alpha + 2\alpha n_{i+1} \tag{2}$$

In a chain of length i, i bacteria may replicate, thus the total rate for one replication occurring in any bacterium of the chain is ir. This explains the terms $rn_{i-1}(i-1) - irn_i$. Such a chain is made of $i-1$ links, so the total breaking rate is $(i-1)\alpha$, explaining the term $-(i-1)n_i\alpha$. When the link within a chain of length two breaks (rate α per chain), two free bacteria are released. For longer chains, there are two outermost links (on each side of the chain) which breaking releases one free bacterium and one one-bacterium shorter chain. This explains the term $\sum_{i=2}^{\infty} 2\alpha n_i$ in equation (1) and the term $2\alpha n_{i+1}$ in equation (2).

Free bacteria growth rate as a function of the bacterial replication rate

Even for this simple version, the system of equations is hard to solve in the general case. We start by studying numerically the growth rate in the long term (the maximum eigenvalue λ of the matrix of coefficients $m_{i,j} = r(i-1)\delta_{i-1,j} - ir\delta_{i,j} - (i-1)\alpha\delta_{i,j} + 2\alpha(\delta_{i+1,j} + \delta_{i,1}(1 - \delta_{j,1} - \delta_{j,2}))$, with $\delta_{i,j}$ the Kronecker symbol, which takes the value 1 when $i = j$, and 0 otherwise), as a function of the replication rate (see figure 2A). The growth rate has a maximum for a finite replication rate, of the order of α (the link breaking rate): the higher the replication rate, the higher the potential for growth in the number of free bacteria, but when the replication rate becomes too large compared to the breaking rate, the bacteria get trapped in clusters, which break and re-attach in more complex clusters from which independent bacteria cannot escape.

Chain length distribution

In the long time limit, the number of chains of length i is of the order of $Cp_i \exp(\lambda t)$, with λ the largest eigenvalue. Equation (2) simplifies to:

$$\lambda p_i = -irp_i + rp_{i-1}(i-1) - (i-1)p_i\alpha + 2\alpha p_{i+1}$$

Assuming that i is large,

$$p_i \simeq \frac{r}{r+\alpha} p_{i-1} \tag{3}$$

is required. Using this approximation for all i, the proportion of chains of length k among linear chains and free bacteria is:

$$p_k = \left(1 - \frac{r}{r+\alpha}\right)\left(\frac{r}{r+\alpha}\right)^{k-1} \tag{4}$$

This approximation works relatively well, especially for smaller r/α values (see figure 2B). Part of the discrepancy is that equation (3) is an approximation for large i, and thus does not hold at small chain length.

Model with bacteria escape

Equations

This is similar to the base model presented before, except that we take into account that upon replication, bacteria may not be perfectly bound, and may escape (pannel B of figure 1). We denote δ the probability for the two daughter bacteria to become free bacteria upon replication of a free bacterium. We denote δ' the probability that when a bacteria at the tip of a chain replicates, the daughter bacterium on the outside of the chain escapes the enchainment. We denote δ'' the probability that when a bacterium at the interior of the chain divides, the daughter bacteria will not be enchained, effectively clipping the chain in two. As free bacteria are more motile than clusters, then $\delta \geq \delta' \geq \delta''$. We also add here the possibility that the loss rate c for free bacteria and c' for chains are different. Then the base equations are:

$$\frac{dn_1}{dt} = r(-1+2\delta)n_1 + \sum_{i=2}^{\infty} 2r\delta' n_i + \sum_{i=2}^{\infty} 2\alpha n_i - cn_1$$

$$\frac{dn_2}{dt} = r(1-\delta)n_1 - 2r(1-\delta')n_2 - \alpha n_2 + 2n_3\alpha - c'n_2$$

and for $i \geq 3$,

$$\frac{dn_i}{dt} = r(2\delta' - i)n_i + rn_{i-1}(i-1-2\delta'+3\delta''-i\delta'') - (i-1)n_i\alpha + 2\alpha n_{i+1} - c'n_i.$$

Free bacteria growth rate as a function of the bacterial replication rate

Similarly to the base model, we study numerically the growth rate as a function of the replication rate (see figure 2C). The larger the replication rate, the more the deviation between the growth rate and the replication rate, which would be its value in the absence of clusters. If δ, δ', δ'' are small enough, the qualitative behavior is similar to the base model. But for larger δ, δ' and δ'', the growth rate continues to increase monotonically with the replication rate. The same is true when δ, δ' and δ'' are different (see supplementary figure 1). If $c = c'$, the growth rate is simply offset by minus the loss rate (see supplementary figure 1), and if $c \neq c'$, the effect is more complex, but for small r/α values it corresponds to an offset of $-c$.

Chain length distribution

We can reason similarly to the base model (more details in section 3 in appendix), and find:

$$p_k = \left(1 - (1-\delta'')\frac{r}{r+\alpha}\right)\left((1-\delta'')\frac{r}{r+\alpha}\right)^{k-1} \tag{5}$$

This approximation works relatively well (figure 2D, and supplementary figure 2). The approximation (5) depends on δ'', but neither on δ nor δ', but δ and δ' could actually matter when i is small, and indeed we observe (see supplementary figures 3 and 4) that the approximation (5) works slightly less well when δ'' is

different from δ or δ'. If $c = c'$, the distribution is exactly the same as for $c = c' = 0$, and if $c \neq c'$, the distribution changes very little (see supplementary figure 5).

Model with fixed replication time

In this variant of the base model, bacteria divide every τ. The effective growth rate is r_{eff} such that $\exp(r_{eff}t) = 2^{t/\tau}$, thus $r_{eff} = \log(2)/\tau$.

Equations

Let us start by considering a chain of n bacteria at $t = 0$, right after a replication event. Let us denote $l(n, i, t)$ the probability that at t, this chain has lost i bacteria in total on the extremities, and consequently is of length $n - i$ at t ($n \geq 2$ and $0 \leq i \leq n - 2$). Before the next replication event, since we assume $q = 0$ as in the base model (meaning that if the chain breaks somewhere else, the subparts form a more complex cluster and thus are "lost" for the system), we have:

$$\frac{dl(n, i, t)}{dt} = -\alpha(n - 1 - i)l(n, i, t) + 2\alpha l(n, i - 1, t). \tag{6}$$

At $t = 0$, $l(n, 0, 0) = 1$ and for $0 < i < n - 1$, $l(n, i, 0) = 0$. It can be checked easily that the following expression is the solution of (6), with $l(n, 0, 0) = 1$ and $l(n, i > 0, 0) = 0$, for any $0 \leq i \leq n - 2$:

$$l(n, i, t) = \frac{2^i}{i!} \exp(-\alpha t(n - 1))(\exp(\alpha t) - 1)^i.$$

Any chain of length > 2, has two outermost links, each breaking at rate α, liberating one free bacterium. A chain of length 2 breaks at rate α, but liberates two free bacteria. Starting from a unique chain of length n, the probability that a linear chain of length $n - i$ is present at time t is $l(n, i, t)$. Consequently, the total rate of production of free bacteria at time t is $2\alpha \sum_{i=0}^{n-2} l(n, i, t)$. Thus, the average number of free bacteria generated during τ by this chain of n bacteria is:

$$\aleph_{free}(n, \tau) = 2\alpha \sum_{i=0}^{n-2} \int_0^\tau l(n, i, t)dt = 2\alpha \sum_{i=0}^{n-2} \int_0^\tau \frac{2^i}{i!} \exp(-\alpha t(n-1))(\exp(\alpha t)-1)^i dt.$$

A chain of length n right before replication becomes a chain of length $2n$ upon it, and will have contributed to chains of length k by $l(2n, 2n - k, \tau)$, and to free bacteria by $\aleph_{free}(2n, \tau)$ right before the next replication event. With $n_i(t)$ the number of chains of length i at time t (right before a replication event), we can write the matricial relation between the $n_i(t)$ and $n_i(t+\tau)$ (right before the

next replication event) as follows:

$$\begin{pmatrix} n_1(t+\tau) \\ n_2(t+\tau) \\ n_3(t+\tau) \\ n_4(t+\tau) \\ \vdots \end{pmatrix} = \begin{pmatrix} \aleph_{free}(2,\tau) & \aleph_{free}(4,\tau) & \aleph_{free}(6,\tau) & ... & ... \\ l(2,0,\tau) & l(4,2,\tau) & l(6,4,\tau) & ... & ... \\ 0 & l(4,1,\tau) & l(6,3,\tau) & ... & ... \\ 0 & l(4,0,\tau) & l(6,2,\tau) & ... & ... \\ 0 & 0 & l(6,1,\tau) & ... & ... \\ 0 & 0 & \vdots & \ddots & \ddots \end{pmatrix} \begin{pmatrix} n_1(t) \\ n_2(t) \\ n_3(t) \\ \vdots \\ \vdots \end{pmatrix} \tag{7}$$

This matrix is then cut to size $n_{max} \times n_{max}$, and the corresponding eigensystem is solved numerically.

Free bacteria growth rate as a function of the bacterial replication rate

The shape of the relation between free bacteria growth rate and (effective) replication rate (figure 2E) is very similar in the fixed replication time vs. fixed replication rate models, with a maximum of the growth rate for a finite value of the (effective) replication rate, at close values ($r_{eff} = 1.15\alpha$ vs. $r = 1.09\alpha$). When the replication is at fixed time intervals instead of a fixed replication rate, the maximum growth rate is higher, and it dips faster at increasing effective replication rate. Indeed, in the case of fixed replication rate, the distribution of durations between two replications is exponential, thus more spread. Close to the maximum, the presence of short replication intervals makes that there can be more cluster formation, and conversely, at higher replication rates, the presence of longer replication intervals results in more production of free bacteria.

Chain length distribution

We show here the main steps to calculate analytically an approximation for the chain length distribution, and more details are given in section 4 in appendix. We define $n_i(t)$ the number of chains of length i at t with t taken just before a replication. Assuming i even,

$$n_i(t+\tau) = \sum_{j=0}^{\infty} n_{i/2+j}(t) l(i+2j, 2j, \tau).$$

In the long time limit, $n_i(t) = Cp_i \exp(\lambda t)$, with λ the long term growth rate, that is such that $\exp(\lambda\tau) = \mathcal{N}$, with $\mathcal{N}$ the largest eigenvalue of the matrix of equation 7. Then previous equation leads to:

$$\mathcal{N} p_i = \sum_{j=0}^{\infty} p_{\frac{i}{2}+j} e^{-\alpha\tau(i-1+2j)} (e^{\alpha\tau} - 1)^{2j} \frac{2^{2j}}{(2j)!}.$$

We make the assumption that the first term of the sum is large compared to the rest of the sum (assumption discussed in appendix, section 4). Then,

$$p_i \simeq \frac{1}{\mathcal{N}} p_{\frac{i}{2}} e^{-\alpha\tau(i-1)}$$

and recursively, for $i = 2^k$, with k integer,

$$p_i \simeq p_1 i^{\frac{\alpha\tau - \log(\mathcal{N})}{\log(2)}} e^{-2\alpha\tau i}.$$

When $\alpha\tau \gg 1$, links typically break before the next replication, thus there is little impact of the clustering on the growth. Consequently, the growth will be close to its value in the absence of clustering, i.e. doubling every τ, and thus in this limit $\mathcal{N} = 2$:

$$p_i \simeq p_1 i^{\frac{\alpha\tau}{\log(2)} - 1} e^{-2\alpha\tau i} = p_1 i^{\frac{\alpha}{r_{eff}} - 1} 2^{-2i\alpha/r_{eff}}. \tag{8}$$

This rough approximation allows to explain the core of the observed distribution (figure 2F). There are bumps, due to the replication every τ (which in the absence of link breaking would results in chains of length 2^k only), which makes that chains of power-of-two length are overrepresented. Compared to the case with fixed replication rate, the distribution is much narrower.

Model with linear chains independent after breaking ($q > 0$)

Limit case : subchains always remain independent linear chains after breaking ($q = 1$)

In this model, when a chain breaks, the two resulting chains remain independent and can thus continue to participate in the dynamics of the system:

$$\frac{dn_i}{dt}(t) = (i-1)r\ n_{i-1}(t) + (-\alpha(i-1) - ir)\ n_i(t) + 2\alpha \sum_{j=i+1}^{\infty} n_j(t)$$

We recognize here the equation studied in[18], where they described chains of growing unicellular algae. As it has been shown, the steady state solution of the system is:

$$n_i(t) = C \exp(rt) \left(\frac{r}{\alpha + r}\right)^i,$$

with C a constant dependent on the initial state of the system. In the steady state, the growth rate is equal to the replication rate. Note that the resulting chain length distribution is then exactly equal to its approximation in the base model (4). The average chain length is $1 + \frac{r}{\alpha}$, which shows that, as expected, if the link breaking rate is high compared to the replication rate ($r/\alpha \ll 1$), the average length is close to one as no cluster has the time to form: all the bacteria remain free.

Intermediate case : chains can either be independent or trapped after breaking

More realistically, after breaking, chains will have some probability to either encounter each other and remain trapped in more complex clusters, or to escape and become independent. We will assume in the following that if a chain of length N breaks at a link at the extremity, releasing a chain of length $N-1$ and a free bacterium, then the free bacterium, smaller and likely more mobile, will escape in all cases; but that if the link that breaks is elsewhere, the probability for the new chains of lengths $N-k$ and k ($k > 1$) to escape and continue as

two independent linear chains will be q, and the probability that they bind and form a more complex cluster will be $1-q$, with q independent of k. We write the equations for the number $n_i(t)$ of chain of i bacteria:

$$\frac{dn_1}{dt} = -rn_1 + 2\alpha \sum_{j=2}^{\infty} n_j$$

$$\frac{dn_i}{dt} = -rin_i + r(i-1)n_{i-1} - \alpha(i-1)n_i + 2\alpha n_{i+1} + 2\alpha q \sum_{j=2}^{\infty} n_{i+j}.$$

In the long time, $n_i(t) \to Cp_i \exp(\lambda t)$ with λ the largest eigenvalue.

$$\lambda p_i = -rip_i + r(i-1)p_{i-1} - \alpha(i-1)p_i + 2\alpha p_{i+1} + \sum_{j=i+2}^{\infty} 2\alpha q p_j \tag{9}$$

This is valid for any i. We assume that p_i decreases fast enough with i such that the sum from $i+2$ to ∞ of the p_i is an order of magnitude less than ip_i. Then, the largest elements of equation (9) when i is large enough are the terms multiplied by i, and consequently:

$$0 \simeq -rp_i + rp_{i-1} - \alpha p_i$$

Leading to:

$$p_i \simeq \frac{r}{\alpha + r} p_{i-1}$$

If this is valid for any i, theproportion of chains of length k among linear chains and free bacteria is:

$$p_k = \frac{\alpha}{\alpha + r} \left(\frac{r}{\alpha + r} \right)^{k-1}. \tag{10}$$

Note that this expression does not depend on q and is the same as the approximation for the base model (4) and the exact expression for $q = 1$. We compare this approximation with the numerical results and they are in good agreement (figure 2H), except when both q is small and r/α is large, and even in this case it gives a reasonable approximation.

Replacing p_i by its expression (10), equation (9) simplifies to:

$$\lambda \simeq -r\left(1 + \frac{\alpha}{r}\right) + \alpha + 2\alpha \left(\frac{r}{\alpha + r} \right) + \sum_{j=2}^{\infty} 2\alpha q \left(\frac{r}{\alpha + r} \right)^j$$

which after simplifications leads to:

$$\lambda \simeq r \frac{\alpha + (2q-1)r}{\alpha + r}. \tag{11}$$

This approximation does not work for $q < 0.5$, but it works well for q close to 1, and gives the right slope for r/α large for $q > 0.5$ (figure 2G). We can observe that for $q > 0.5$, λ increases indefinitely when r increases; but it has an intermediate maximum for $q < 0.5$. Intuitively, if $q > 0.5$, when a chain breaks it leads to more than one independent linear chain, thus the population of linear chains and thus free bacteria may increase monotonically with r/α,

whereas if $q < 0.5$, chains that break lead to less than one independent chain on average, and thus the behaviour of the system is more determined by the fate of the chains, and thus closer to the results for $q = 0$, for which the growth rate λ has a maximum as a function of r.

The proportion of free bacteria relative to the total number of bacteria is proportional to $\exp(\lambda t)/\exp(rt)$, which, using approximation (11) tends to $\exp(-2r^2t(1-q)/(r+\alpha))$. The proportion of free bacteria thus decreases over time, and decreases with increasing replication rate r. Thus the proportion of bacteria trapped in clusters, which is 1 minus the proportion of free bacteria, increases with increasing replication rate even when $q > 0.5$.

Model with force-dependent breaking rate

Equations

What drives link breakage? The links could break if there was some process degrading the sIgA, but the sIgA are thought to be very stable[19]. Another possible explanation for link breaking is that the bound antigen can be extracted from the bacterial membrane, at a rate which may vary exponentially with the force[20][15]. The forces applied on the links are likely mostly due to the hydrodynamic forces exerted by the digesta flow on the bacterial chain. Taking the linear chain as a string of beads, as done for polymer chains, and in a flow with a constant shear rate, the force is predicted to be larger as the chain grows longer, and the largest at the center of the chain[17]. A more detailed discussion and the calculations can be found in section 5.1 in the appendix. Taking α as the breaking rate in the absence of shear, and β a constant expressing the strength of the coupling between hydrodynamic forces and link breaking, the resulting equations for this minimal model taking into account the forces are:

$$\frac{dn_1}{dt} = -rn_1 + 2\sum_{i=2}^{\infty} \alpha n_i \exp\left(\beta \frac{i-1}{2}\right)$$

and for i even,

$$\frac{dn_i}{dt} = -rin_i - \alpha n_i e^{\beta i^2/8}\left(1 + 2\sum_{j=2}^{i/2} e^{-(j-1)^2\beta/2}\right) + r(i-1)n_{i-1} + 2\alpha n_{i+1}e^{\beta i/2} \tag{12}$$

and for $i > 1$ odd,

$$\frac{dn_i}{dt} = -rin_i - 2\alpha n_i e^{\beta i^2/8}\sum_{j=1}^{(i-1)/2} e^{-(j-1/2)^2\beta/2} + r(i-1)n_{i-1} + 2\alpha n_{i+1}e^{\beta i/2} \tag{13}$$

Free bacteria growth rate as a function of the bacterial replication rate

The growth rate as a function of the replication rate has a qualitatively similar shape as for the base model (figure 2I), with a finite replication rate maximizing the growth rate. The limit $\beta \to 0$ corresponds well to the base model, as expected. When β increases, the replication rate maximizing the growth rate increases, as the effective breaking rate is higher. Numerically, we find (see

supplementary figure 7) that the replication rate maximizing the growth rate scales as $\alpha \exp(0.8\beta)$.

Chain length distribution

Similarly to the other models, for t long enough, $n_i \simeq C p_i \exp(\lambda t)$ (with λ the largest eigenvalue), and assessing which terms in equations (12) and (13) will be dominant, we ultimately obtain (details in supplementary section 5.3):

$$p_i \simeq \left(\frac{r}{\alpha}\right)^{i-1} \frac{(i-1)!}{Y^{floor(i/2)} Z^{floor((i-1)/2)}} \exp\left(-\frac{\beta}{8}\left(-1 + \frac{i + 3i^2 + 2i^3}{6}\right)\right), \tag{14}$$

with $Y = 1 + 2\sum_{j=1}^{\infty} \exp(-\beta j^2/2)$ and $Z = 2\sum_{j=1}^{\infty} \exp(-\beta(j-1/2)^2/2)$. This approximation works well, except for small β (figure 2J, and supplementary figure 8). Compared to the base model, the number of chains decreases much faster with their length. Indeed, the breaking rates for each link increase importantly with the chain length, thus larger chains are much less stable than in the base model.

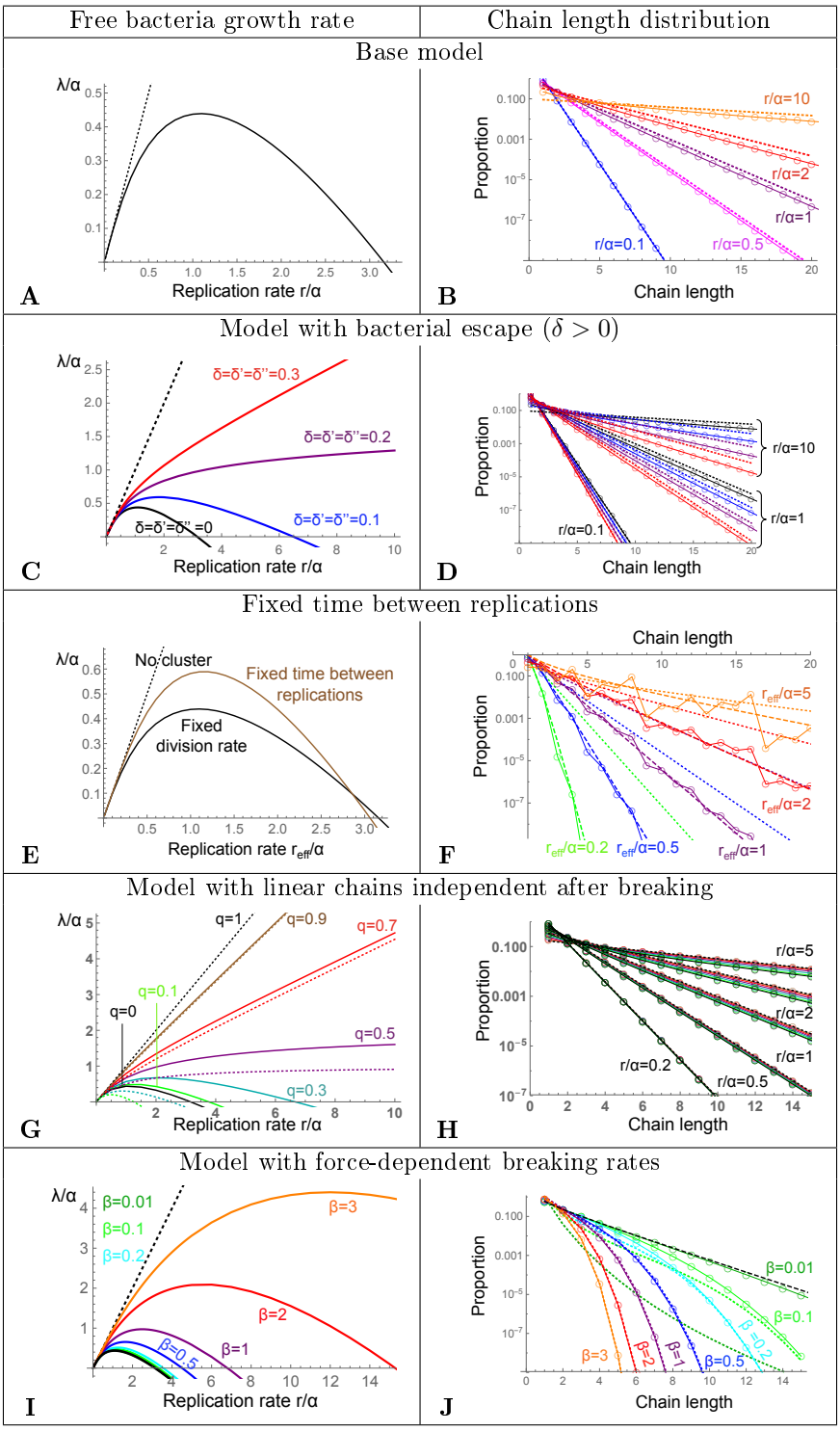


Figure 2 :
Left panels A, C, E, G, I: Growth rate λ of the free bacteria as a function

of the bacteria replication rate r, both in units of α. Numerical results (solid colored lines), and limit with no clusters ($\lambda = r$) (black dotted line). The base model is represented in black in all the left panel figures to ease comparison between models.

Right panels B, D, F, H, J: Chain length distribution. Open circles linked by solid lines: numerical results.

A, B: Base model, $n_{max} = 40$. B. dotted lines: approximation (4) (almost overlaid with the numerical results for $r/\alpha = 0.1$).

C, D: Model with bacterial escape. $\delta = \delta' = \delta'' = 0, 0.1, 0.2, 0.3$. $c = c' = 0$, $n_{max} = 40$. D. dotted lines: approximation (5).

E, F: Fixed time between replications. $r_{eff} = \log(2)/\tau$. $n_{max} = 32$. F. approximation (8) (dashed lines), numerical result in the base model (dotted lines). $r_{eff}/\alpha = 0.2, 0.5, 1, 2, 5$.

G, H: Model with linear chains independent after breaking. . G. The dotted black line is the case $q = 1$, for which $\lambda = r$, like in the absence of clusters. The colored dotted lines are the analytical approximation (11). $n_{max} = 200$. H. The dotted black lines are the approximate distribution (10) for each r/α, which is the exact distribution for $q = 1$. The colours represent the same q values than for the left panel. All curves are almost overlaid for small r. $n_{max} = 80$.

I,J: Model with force-dependent breaking rates. Each color represents a different β: $\beta = 0.01$ ($n_{max} = 20$), $\beta = 0.1$ ($n_{max} = 15$), $\beta = 0.2$ ($n_{max} = 15$), $\beta = 0.5$ ($n_{max} = 15$), $\beta = 1$ ($n_{max} = 15$), $\beta = 2$ ($n_{max} = 10$), $\beta = 3$ ($n_{max} = 10$). I. The black line is the numerical result for the base model, equivalent to $\beta = 0$. The curve for $\beta = 0.01$ (dark green) is almost overlaid with the curve for $\beta = 0$. J. Chain length distribution for $r/\alpha = 1$. The colored dotted lines the analytical approximation (14), and the black dashed line the numerical result for the base model.

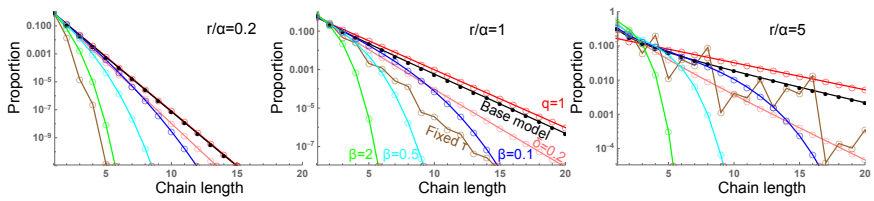


Figure 3: Comparison of the chain length distributions for the different models. The base model is represented in black. For the case with bacterial escape when a bacterium replicates, one numerical value is represented ($\delta = \delta' = \delta''$) (pink), the lower this value, the closer to the base model. The model with fixed replication time is represented in brown (for this model we choose $\tau = log(2)/r$). The model with linear chains independent after breaking ($q > 0$) is shown in red for $q = 1$, the most different from the base model. All intermediate q values are between the black and the red curves. The model with the force dependent breaking rate is represented for 3 values of β: 0.1 (blue), 0.5 (cyan), 2 (green). All the results are numerical, using n_{max} values as in figure 2, except for $q = 1$ which is an exact analytical result.

Comparison with experimental data

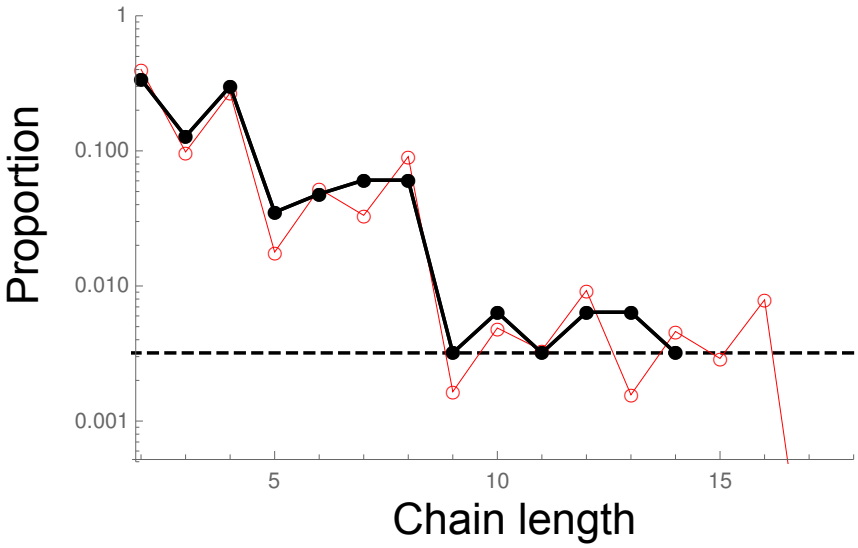


Figure 4: Comparison with experiments. Chain length distribution (proportions relative to the linear chains of length ≥ 2 are represented in log scale). The black dots and line are the experimental data. The horizontal dotted black line represents the case in which there is one chain of the given size. No chains longer than size 14 were detected in our experimental images. The red line and points are the numerical results for the model with fixed replication time. In this model, there is only one free parameter, $r_{eff}/\alpha = \log(2)/(\alpha\tau)$, which fitted value is 4.1 (see appendix 7 for more details).

We analyzed (see section 7 in the appendix) microscopy images of cecal content from vaccinated mice infected with *S.*Typhimurium, which were acquired for our previous study [12] . Most clusters are large, and of complex shape. But smaller clusters are linear, and we obtained the distribution shown by the black line and points in figure 4. The model with fixed division time is *a priori* more realistic. The best fit is obtained with this model, with the one adjustable parameter $r/\alpha = 4.1$ (red line and points). It seems however, though there is not enough data to quantitatively ensure it, that there are less long clusters than expected (see appendix section 7.3 for an expanded discussion). The data may be biased, as longer chains may not be fully in the focal plane. That the distribution is relatively narrow could also be compatible with force-dependent breaking rates.

Summary of results and discussion

We started from the recent finding[12] that the protection effect of sIgA, the main effector of the adaptive immune system in the gut, can be explained by enchained growth. Because sIgA are multivalent, they can stick identical bacteria together if they encounter each other. Early in infection, bacteria of the

same type are at low density, thus typical encounter times are very long, but when a bacterium replicates, the daughter bacteria are in contact and thus can remain enchained to each other by IgA. Bacteria in clusters are less motile than individual bacteria, and in particular, are not observed close to the epithelial cells. In the case of wild type *S.*Typhimurium, only free bacteria which can interact with the epithelial cells contribute to the next steps of the infection process. Despite the presence of sIgA, some free bacteria are observed. It could be that they escape at the moment of replication. But, along with the observation that clusters do not grow indefinitely, it could also be a sign that the links between bacteria break. It is also physically expected that the links have some finite breaking rate. If the typical time between two bacterial divisions is much larger than the typical time for the link to break, then there would be no cluster. Conversely, in the inverse case, bacteria will be very likely to be trapped in large clusters. Then, even if sIgA are produced against all bacterial types, the bacteria dividing faster will be disproportionately affected.

We investigated if this qualitative idea holds with more realistic models. We started from a base model in which: bacteria replicate at a fixed rate; remain enchained upon replication; until the link between them breaks at a given fixed breaking rate, identical for all links; and considering that, because of the way bacteria such as *Salmonella* or *E.coli* divide, the early clusters are linear chains of bacteria; when the chain breaks at an outermost link, we assumed the free bacteria will escape; but if the chain breaks elsewhere, we assumed that the two resulting sub-chains encounter each other quickly and form clusters of more complex shapes from which individual bacteria do not escape. We studied this base model with a combination of analytical and numerical approaches. We also tested the robustness of our findings by studying separately several variations of the base model: a probability of escaping upon replication, loss rates, fixed replication time, non-zero probability for the subchains to escape, and force-dependent breaking-rates. For each model, we studied how the growth rate of the free bacteria varies with the replication rate (which would be equal if there were no clusters), and the distribution of chain lengths.

We find that, except in the very specific case in which subchains always escape upon link ($q = 1$), the growth rate of the number of free bacteria is lower than the replication rate. And more spectacularly, in most of the models studied (but not if more than half the subchains escape upon link breaking, or if there is a significant probability for bacteria to escape enchainement upon replication), the growth rate of the number of free bacteria is non-monotonic with the replication rate : there is a finite replication rate which maximizes the growth rate of non-clustered bacteria. At very high replication rates, bacteria get trapped in more complex clusters and cannot contribute anymore to the free bacteria dynamics and thus to the next steps of the infection process. The replication rate maximizing the growth rate is of the order of the breaking rate, though its specific value depends on the details of the model. To summarize, except when $q = 1$, we always find that the higher the replication rate, the higher the proportion of bacteria trapped into clusters; and in many cases, the effect is even more dramatic, with the growth rate of free bacteria that may decrease with the replication rate.

The chain length distribution is dependent on the model (see figure 3). In most cases, the proportion of linear chains having length k decreases as γ^k, with γ some constant smaller than 1. When replication occurs at fixed time,

or when breaking rates are force-dependent, the proportion of longer chains decreases faster. There are models with different chain length distributions but qualitatively similar dependence of the growth rate on the replication rate, and the opposite is true too. This shows that large clusters have little importance for free bacteria production, what matters most is the small chains dynamics. It is reassuring, as we did not consider buckling, which would make long linear chains fold on themselves and produce more complex clusters, and may bias the linear chain distribution for very large lengths. It should also be noted that with fixed division time, not only the distribution is bumpy, as chains comprising a power of two number of bacteria are more frequent than others, but the distribution is also narrower.

We analyzed experimental data on clusters of *S.*Typhimurium in the cecum of vaccinated mice. The experimental chain length distribution is in line with the model of fixed replication time, which is indeed more realistic. There is however somewhat less large chains than expected. More data would be necessary to asses this more reliably. This could be because of possible bias in the data. This could be also compatible with force-dependent breaking rates. Additional experiments, for instance to measure the breaking rate, could help by giving additional independent information and constrain the fitting. To test the dependence of the growth rate with the replication rate, an ideal experiment would be to compare similar bacterial strains, but with differing replication rates, and compete them in the same individual. It is however very challenging to obtain bacteria that differ only by their replication rate, particularly in vivo.

sIgA-enchained bacterial clusters could be studied in vitro to measure how they break. However, using in vitro results to draw conclusions on in vivo systems is limited. First, there could be chemical or enzymatic components of the lumen that could facilitate or hinder link breaking, and the non-Newtonian viscosity of the digesta could play a role in the mechanic forces felt by the links, thus a simple buffer may not mimic well the real conditions. More crucially, the exact forces felt by particles of the size of bacterial clusters are not well characterized. Most studies of the flow characteristics in the digestive system rely either on external observations of the peristaltic muscles[21] or indirect measures of times for a marker to exit some section of the digestive track[22]. More quantitative study of the digestive flow at small scales is just beginning[23, 24, 25, 26, 27, 9, 10] and in the future it may give more clues to assess to which forces bacteria are subjected to in the digestive track.

The mechanism we propose is nevertheless plausible. The observation in vaccinated mice of the existence of single bacteria and small clusters, and particularly small linear chains with an odd number of bacteria, are pieces of evidence that clusters do break in these in vivo conditions. An alternative explanation could be that some bacteria escape enchainement upon replication. However, at higher bacterial densities, we have evidence of independent bacteria binding when they encounter[12], thus sIgA coated bacteria are adhesive. When two daughter bacteria divide, they are in contact, thus if sIgA is adhesive, escape is unlikely (see appendix section 2). Importantly, even though our results show that specific conditions are needed for the growth rate to decrease with high replication rates, we almost always find that the higher the replication rate, the higher the proportion of bacteria trapped in clusters. Thus, even when it does not reverse the relationship between the growth rate of the free bacteria and the replication rate, it is at least dampening this relationship, and can be a tool

both to control pathogenic bacteria, but also to maintain homeostasis of the gut microbiota. It is also interesting that there are other host effectors besides sIgA that bind bacteria together: neutrophil extracellular traps for instance[28], and there could also be an interplay between replication rates and the breaking of the links mediated by these other effectors, as the mechanism we propose here is generic.

As for any mechanism to fight against bacteria, how easily resistance can be evolved is crucial. On the one hand, the replication rate could evolve. But bacteria replicating slower would be less competitive with other bacteria in the absence of sIgA, and a slower growth leaves more time for further host response. On the other hand the typical link breaking time could evolve. On the host side, sIgA is thought to be mechanically very stable, and experiments about the bonding of cells by sIgA seem to point to the link failing because of the extraction of the antigen rather than because of sIgA breaking, and rather than the sIgA/antigen bond detaching[20][14]. In the case of IgA defficiency, there is more secretion of IgM, and microbiota is disturbed[29]: we may speculate that IgM being less powerful for microbiota homeostasis is related to these immunoglobulins being more protease-sensitive than IgA and thus cleaved on shorter time scales[30]. On the other side, bacteria could evolve surface antigens. It could be interesting to think that bacteria could produce decoy antigens with no functional value, but against which the immune system will mount an immune response, and that are more easily released from the bacteria, thus disabling the main sIgA mode of action (being easily evolvable would also be a benefit). Such decoys would however be a metabolic cost for the bacteria, and when breaking, may unmask other antigens corresponding to crucial functions of the bacteria. It could be argued that the capsule around bacteria such as *Salmonella* spp., and also common in pathogenic *E.coli*, may behave as a decoy, though it has also other functions. Evolving resistance to IgA-mediated enchainment would thus be costly.

Along the same lines, we may speculate whether mechanical aspects could be a reason why sIgA against some antigens are not efficient for protection. For instance, while anti-flagella sIgA aggregate very well *Salmonella* Enteriditis together, they are not efficient for protection[31]. A main reason could be that as *Salmonella* can switch flagella production on and off, then some *Salmonella* will always escape these sIgA, and seed the infection[32]. An additional possibility could be that flagella may more easily break, especially as distance between bacteria bound by flagella (long) is likely larger than for bacteria bound by O-antigens (on chains shorter than flagellas)[33], and thus the shear forces would be larger. Further, the mechanical properties of the outer sugar layer of the gram negative bacteria could vary, and thus could be used to tune interactions. However, it would add another constraint on bacteria, and the general result that the growth rate compared to the replication rate is at least dampened by the cluster formation would remain.

In the crowded environment of the gut, it is hard for the host to identify the good and the bad bacteria. That vaccination with dead bacteria is sufficient to produce sIgA and protection, shows that the host does not discriminate well against which bacteria they produce sIgA, as these dead bacteria do not harm. Linking the effect (here the clustering) of the immune effectors with a property directly relevant to the potential bacterial pathogeneicity (here the replication rate) avoids to make complex decisions about which bacteria to produce effectors

against.

Acknowledgments

The authors thank Anne-Florence Bitbol for careful reading of an earlier version of the manuscript, as well as other Laboratoire Jean Perrin members for useful discussions, in particular Raphaël Voituriez. We also thank Roger Lentle for giving useful indications about the flow in the mouse digestive system.

References

[1] Helander HF, Fändriks L. Surface area of the digestive tract–revisited. Scandinavian journal of gastroenterology. 2014;49(6):681–689. 2

[2] Casteleyn C, Rekecki A, Van der Aa A, Simoens P, Van den Broeck W. Surface area assessment of the murine intestinal tract as a prerequisite for oral dose translation from mouse to man. Laboratory animals. 2010;44(3):176–183. 2

[3] Sender R, Fuchs S, Milo R. Revised estimates for the number of human and bacteria cells in the body. PLoS biology. 2016;14(8):e1002533. 2

[4] Stecher B, Hardt WD. The role of microbiota in infectious disease. Trends in microbiology. 2008;16(3):107–114. 2

[5] Dishaw LJ, Cannon JP, Litman GW, Parker W. Immune-directed support of rich microbial communities in the gut has ancient roots. Developmental & Comparative Immunology. 2014;47(1):36–51. 2

[6] Donaldson GP, Lee SM, Mazmanian SK. Gut biogeography of the bacterial microbiota. Nature Reviews Microbiology. 2016;14(1):20. 2

[7] Williams R, Gibbons R. Inhibition of bacterial adherence by secretory immunoglobulin A: a mechanism of antigen disposal. Science. 1972;177(4050):697–699. 2

[8] Strugnell RA, Wijburg OL. The role of secretory antibodies in infection immunity. Nature Reviews Microbiology. 2010;8(9):656. 2

[9] Cremer J, Arnoldini M, Hwa T. Effect of water flow and chemical environment on microbiota growth and composition in the human colon. Proceedings of the National Academy of Sciences. 2017;114(25):6438–6443. 3, 21

[10] Logan SL, Thomas J, Yan J, Baker RP, Shields DS, Xavier JB, et al. The Vibrio cholerae type VI secretion system can modulate host intestinal mechanics to displace gut bacterial symbionts. Proceedings of the National Academy of Sciences. 2018;115(16):E3779–E3787. 3, 21

[11] Endt K, Stecher B, Chaffron S, Slack E, Tchitchek N, Benecke A, et al. The microbiota mediates pathogen clearance from the gut lumen after non-typhoidal Salmonella diarrhea. PLoS pathogens. 2010;6(9):e1001097. 3

[12] Moor K, Diard M, Sellin ME, Felmy B, Wotzka SY, Toska A, et al. High-avidity IgA protects the intestine by enchaining growing bacteria. Nature. 2017;544(7651):498–502. doi:10.1038/nature22058. 3, 4, 5, 7, 19, 21

[13] Kudoh K, Shimizu J, Ishiyama A, Wada M, Takita T, Kanke Y, et al. Secretion and excretion of immunoglobulin A to cecum and feces differ with type of indigestible saccharides. Journal of nutritional science and vitaminology. 1999;45(2):173–181. 3

[14] Tees DF, Coenen O, Goldsmith HL. Interaction forces between red cells agglutinated by antibody. IV. Time and force dependence of break-up. Biophysical Journal. 1993;65(3):1318–1334. doi:10.1016/S0006-3495(93)81180-9. 3, 22

[15] Evans EA, Calderwood DA. Forces and bond dynamics in cell adhesion. Science. 2007;316(5828):1148–1153. 3, 15

[16] McGrady E, Ziff RM. "Shattering"transition in fragmentation. Physical review letters. 1987;58(9):892. 3

[17] Odell J, Keller A. Flow-induced chain fracture of isolated linear macromolecules in solution. Journal of Polymer Science Part B: Polymer Physics. 1986;24(9):1889–1916. 3, 15

[18] Gherardi M, Amato A, Bouly JP, Cheminant S, Ferrante MI, d'Alcalá MR, et al. Regulation of chain length in two diatoms as a growth-fragmentation process. Physical Review E. 2016;94(2):022418. 13

[19] Brandtzaeg P. Role of secretory antibodies in the defence against infections. International Journal of Medical Microbiology. 2003;293(1):3–15. doi:10.1078/1438-4221-00241. 15

[20] Evans E, Berk D, Leung A. Detachment of agglutinin-bonded red blood cells. I. Forces to rupture molecular-point attachments. Biophysical Journal. 1991;59(4):838–848. doi:10.1016/S0006-3495(91)82296-2. 15, 22

[21] Hulls C, Lentle RG, de Loubens C, Janssen PW, Chambers P, Stafford KJ. Spatiotemporal mapping of ex vivo motility in the caecum of the rabbit. Journal of Comparative Physiology B. 2012;182(2):287–297. 21

[22] Janssen PWM, Lentle RG, Asvarujanon P, Chambers P, Stafford KJ, Hemar Y. Characterization of flow and mixing regimes within the ileum of the brushtail possum using residence time distribution analysis with simultaneous spatio-temporal mapping. The Journal of Physiology. 2007;582(3):1239–1248. doi:10.1113/jphysiol.2007.134403. 21

[23] Lentle R, Janssen P. Physical characteristics of digesta and their influence on flow and mixing in the mammalian intestine: a review. Journal of Comparative Physiology B. 2008;178(6):673–690. 21

[24] Kim HJ, Huh D, Hamilton G, Ingber DE. Human gut-on-a-chip inhabited by microbial flora that experiences intestinal peristalsis-like motions and flow. Lab on a Chip. 2012;12(12):2165. doi:10.1039/c2lc40074j. 21

[25] Lentle RG, Loubens C. A review of mixing and propulsion of chyme in the small intestine: fresh insights from new methods. Journal of Comparative Physiology B. 2015; p. 369–387. doi:10.1007/s00360-015-0889-5. 21

[26] Cremer J, Segota I, Yang Cy, Arnoldini M, Sauls JT, Zhang Z, et al. Effect of flow and peristaltic mixing on bacterial growth in a gut-like channel. Proceedings of the National Academy of Sciences. 2016;113(41):11414–11419. 21

[27] Wiles TJ, Jemielita M, Baker RP, Schlomann BH, Logan SL, Ganz J, et al. Host gut motility promotes competitive exclusion within a model intestinal microbiota. PLoS biology. 2016;14(7):e1002517. 21

[28] Fournier B, Parkos C. The role of neutrophils during intestinal inflammation. Mucosal immunology. 2012;5(4):354. 22

[29] Mirpuri J, Raetz M, Sturge CR, Wilhelm CL, Benson A, Savani RC, et al. Proteobacteria-specific IgA regulates maturation of the intestinal microbiota. Gut microbes. 2014;5(1):28–39. 22

[30] Richman LK, Brown WR. Immunochemical characterization of IgM in human intestinal fluids. The Journal of Immunology. 1977;119(4):1515–1519. 22

[31] Iankov ID, Petrov DP, Mladenov IV, Haralambieva IH, Mitov IG. Lipopolysaccharide-specific but not anti-flagellar immunoglobulin A monoclonal antibodies prevent Salmonella enterica serotype Enteritidis invasion and replication within HEp-2 cell monolayers. Infection and immunity. 2002;70(3):1615–1618. 22

[32] Diard M, Garcia V, Maier L, Remus-Emsermann MN, Regoes RR, Ackermann M, et al. Stabilization of cooperative virulence by the expression of an avirulent phenotype. Nature. 2013;494(7437):353. 22

[33] Liu B, Knirel YA, Feng L, Perepelov AV, Senchenkova SN, Reeves PR, et al. Structural diversity in Salmonella O antigens and its genetic basis. FEMS microbiology reviews. 2014;38(1):56–89. 22

Appendix

1 Order of magnitude of the encounter time between two bacteria

The typical time to find one target of radius a in a sphere of radius b by diffusion is of the order of $b^3/(Da)$, so the typical time when there are N bacteria in a volume V is of the order of $V/(NDa)$. For bacteria, a is in the micrometer range. Bacteria such as salmonella or E.coli typically swim at $10\mu m/s$, and change direction every second, which gives a diffusion coefficient of the order of $10^{-10}m^2/s$ [1, 2, 3] (The peristaltic motions of the digesta are large scale movement rather than local diffusion, so we assume they have a smaller effect on diffusion). The mouse's cecum has a volume of the order of $(1\text{cm})^3$. In experiments of [4], the smallest inoculum consists in $N = 10^5$ bacteria, which is already large compared to what could be a realistic number of pathogenic bacteria in food poisoning (10^5 is the typical number of *Salmonella* for food poisoning in humans [5], which are much larger than mice). With these numbers, the typical encounter time is of the order of $10^5 s$, i.e 30h, about 10 times longer than the typical digestion time in mice.

2 Argument for high enchainment probability upon replication

When a bacterium replicates, the time for septation is of the order of a few minutes. We intuitively think that this time is much larger than the time τ_k required for bacteria to stick together when they randomly meet. The aim of this section is to check this intuition by giving an overestimate of τ_k.

If the diffusion coefficient is high enough, the time for bacteria to stick to each other will be limited by which proportion of the time they spend in close vicinity, and the rate k at which bacteria stick to each other when they are in close vicinity, k being the inverse of τ_k. If the diffusion coefficient is smaller, then the time to first encounter will also play a role, but as we calculate an overestimate of τ_k, we can neglect this scenario.

We use the data on figure 1k of [4] about non-dividing bacteria (so the only sticking is from random encounters). The majority of them are aggregated after τ_{exp} up to 8 hours (from the inoculum ingestion to the sampling used for imaging) for a concentration of $10^7 - 10^8$ bacteria. As we will see, this estimate of τ_k is proportional to τ_{exp} and N, so to be conservative, as we will calculate an overestimate of τ_k, we take the highest concentration and the maximum experimental time, i.e. $N = 10^8$ bacteria in $V = 1cm^3$ (cecum volume) and $\tau_{exp} = 8$ hour.

The bacteria typical size is a few micrometers, we thus take $3\mu m$ as an overestimate of the maximum bacterial size. Thus to be in close contact, two bacteria must be at most at $a = 3\mu m$ away. Let us assume that then, the volume of possible contact is $4/3\pi a^3$, which is also an overestimate, because only certain orientations will allow bacteria to touch each other. Then, the proportion of time spent in close contact will be of the order of $(N4\pi a^3)/(3V)$.

Then the typical time to stick to each other will be $\tau_{exp} = \tau_k 3V/(N4\pi a^3)$. Then $\tau_k = \tau_{exp} N4\pi a^3/(3V)$. Numerically, we obtain about 5 minutes as an overestimate of τ_k.

Note that this is a large overestimate. Indeed, when bacteria get clumped to each other, their effective concentration decreases, thus it takes longer for the last bacteria to meet others, and thus the time for most bacteria to be clumped will be significantly larger than the inverse of the early clumping rate.

With all these highly conservative estimates, we find τ_k at the very most of the same order of magnitude as the septation time, and very likely much smaller. Hence the probability for bacteria to escape enchainment is small, which justifies that we take in general the limit of no escape.

3 Model with bacterial escape ($\delta > 0$) and differential loss ($c \neq c'$).

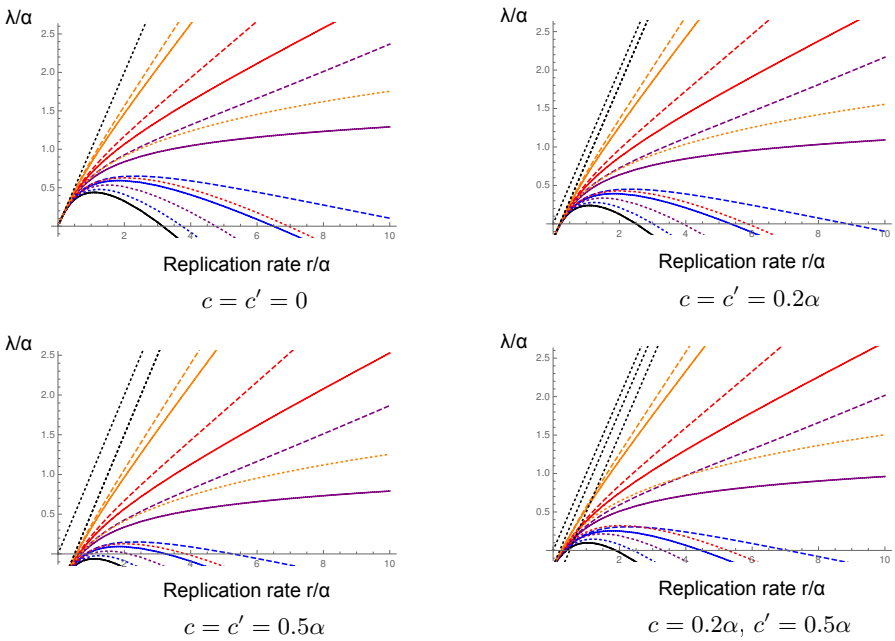


Figure S1: Growth rate λ as a function of the replication rate r, both in units of α. Numerical results (colors), with $\delta = \delta' = \delta''$ (solid lines), $\delta = \delta'$, and $\delta'' = 0$ (dashed lines), $\delta' = \delta'' = 0$ (dotted lines). $\delta = 0$, 0.1, 0.2, 0.3, 0.5. Note that for $\delta = 0$, solid, dashed and dotted lines collapse, as expected. The black dotted lines are either r/α, $(r-c)/\alpha$ or $(r-c')/\alpha$. As expected, if $c = c'$, the resulting growth rate are the same than when $c = c' = 0$, minus c. If $c \neq c'$, the results are closer for small r/α to the results if both c and c' had the c value. For the numerical results, $n_{max} = 40$.

Figure S1 shows how the growth rate depends on r for different δ, δ', δ'', c and c'.

Our numerical study of the system showed us that there is some critical value δ_c below which the behavior is qualitatively similar to the behavior of the system with $\delta = 0$, i.e. with a finite maximum of the growth rate of the free bacteria as a function of the replication rate; and above which the growth rate continues to increase with replication rate. Actually, for $\delta > 0.5$, the growth rate necessarily continues to increase with the replication rate. Indeed, upon replication, one free bacteria becomes two daughter bacteria, an average of 2δ of them staying free. Thus the net gain in free bacteria is $2\delta - 1$. Thus for $\delta > 0.5$, the growth rate of free bacteria is at minimum $r(2\delta - 1)$. Consequently, $\delta_c \leq 0.5$.

We detail here how to obtain the approximation for the chain length distribution. In the long time limit, the number of chains of length i is of the order of $Cp_i \exp(\lambda t)$, with λ the largest eigenvalue. Equation (8) of main text simplifies to:

$$\lambda p_i = r(2\delta' - i)p_i + rp_{i-1}(i - 1 - 2\delta' + 3\delta'' - i\delta'') - (i-1)p_i\alpha + 2\alpha p_{i+1} - c'p_i$$

Assuming that i is large,

$$p_i \simeq (1 - \delta'')\frac{r}{r+\alpha}p_{i-1}$$

is required. Using this approximation for all i, the proportion of chains of length k is:

$$p_k = \left(1 - (1-\delta'')\frac{r}{r+\alpha}\right)\left((1-\delta'')\frac{r}{r+\alpha}\right)^{k-1}$$

Free bacteria are released at a rate $2r\delta' + 2\alpha$ per chain. This rate is independent of the chain length. The direct contributions to the increase of free bacteria from chains of length i compared to all the larger chains will be (with $K = (1-\delta'')r/(r+\alpha)$):

$$\frac{\textit{contribution larger}}{\textit{contribution } i} = \frac{(2r\delta' + 2\alpha)\sum_{j=i+1}^{\infty}(1-K)K^j}{(2r\delta' + 2\alpha)(1-K)K^i} = \sum_{j=1}^{\infty} K^j = \frac{K}{1-K} = \frac{(1-\delta'')r}{\alpha + r\delta''}$$

If r is small compared to α (replication rate $\ll$ breaking rate), then this ratio is small. Thus the larger chains are quickly negligible. Indeed, in this regime, chains typically dislocate before new replications, so there are few larger chains.

Figures S2, S3, S4, S5 show how the chain length distribution depends on δ, δ', δ'', c and c'.

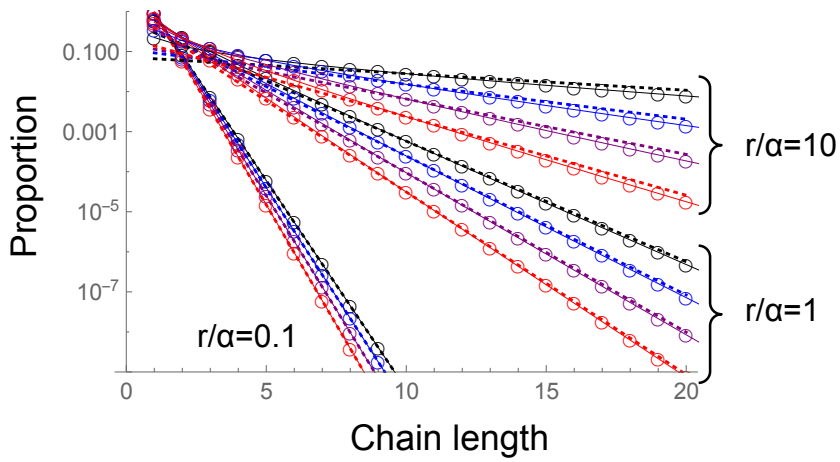
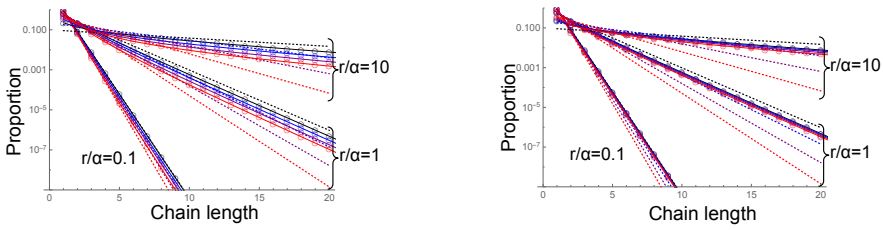
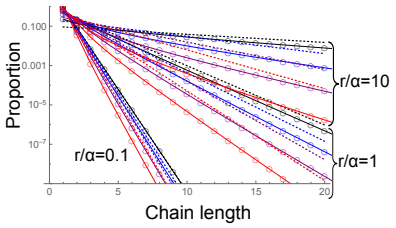


Figure S2: Chain length distribution. All as in main figure 2D (model with bacterial escape) (numerical results: points linked by solid lines), except that the approximation (9) of main text (dotted lines) is rescaled by the numerical value at $n = 10$, i.e. instead of representing $\log(p_{i,approx})$, what is represented is $\log(p_{i,approx} * p_{10,numeric}/p_{10,approx}) = \log(p_{i,approx}) + \log(p_{10,numeric}/p_{10,approx})$. This shows that the approximation captures well the length distribution of large chains. We do this because the base for the analytical approximation is the ratio p_{i+1}/p_i, for which we get a limit expression valid for large i. To calculate the whole distribution p_i, we assume that this limit expression for the ratio is valid for any i, whereas this will not be the case for small i. If the limit expression for p_{i+1}/p_i is correct for large i but not for small i, the slope in log scale plot will be correct, but with some offset dependent on how wrong we got the small i case. Making this renormalization enables to check more easily whether the slope is correct. $\delta = \delta' = \delta'' = 0$ (black), 0.1 (blue), 0.2 (purple), 0.3 (red). $n_{max} = 40$.

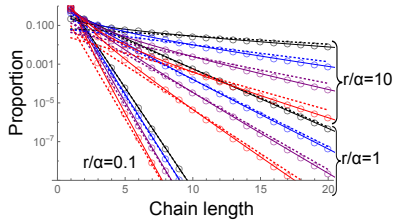


$\delta = \delta' \geq 0$, $\delta'' = 0$ $\qquad$ $\delta \geq 0$, $\delta' = \delta'' = 0$

Figure S3: Chain length distribution. All as in figure 2D (model with bacterial escape) (numerical results: points linked by solid lines), except the values of δ' and δ'': $\delta = 0$, 0.1, 0.2, 0.3. Dotted lines: approximation (5) of main text. Approximation (9) of the main text predicts that the distribution should depend only on δ'', and not δ nor δ'. In these figure where $\delta'' = 0$ but δ (and in the left pannel δ') have non-zero values, we do observe that the distribution, in particular its slope, is closest to the result for $\delta = \delta' = \delta'' = 0$. $c = c' = 0$, $n_{max=40}$.



Dotted lines : approximation (9) of main text



Dotted lines: approximation (9) of main text, rescaled by the numerical value at $n = 10$ similarly to figure S2.

Figure S4: Chain length distribution, for $\delta = \delta' = 2\delta''$. Other parameters as in figure 2D of main text : $\delta = 0$, 0.1, 0.2, 0.3. $c = c' = 0$, $n_{max} = 40$. Dotted lines: approximation (9) of main text. The approximation does not work as well as when $\delta = \delta' = \delta''$.

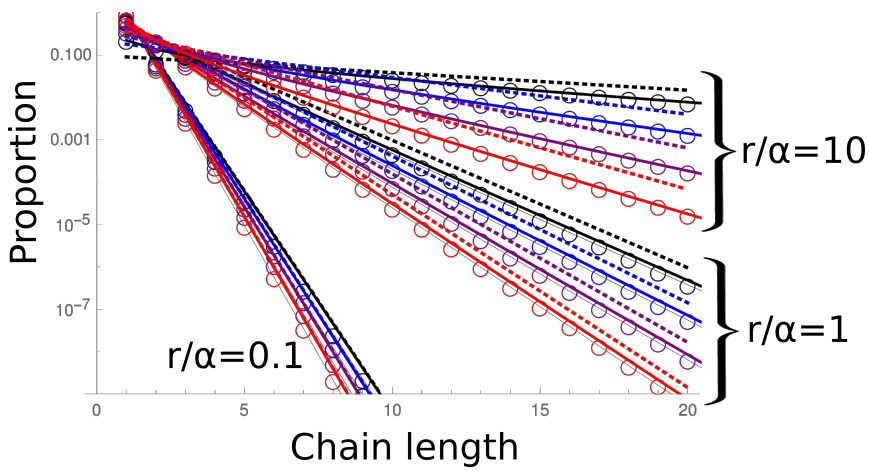


Figure S5: Chain length distribution. Similar to figure 2D of main text. $\delta = \delta' = \delta'' = 0$, 0.1, 0.2, 0.3. Thick solid lines with no markers are for $c = c' = 0$, Thin solid lines with $\circ$ are for $c = 0.2\alpha$, $c' = 0.5\alpha$. Dotted lines: approximation (9) of main text. There is very little change in the chain length distribution. $n_{max} = 40$.

4 Chain length distribution with a fixed replication time - approximation

Below, we present in details the assumptions and calculations to obtain the approximation of the chain length distribution when bacteria replicate every τ.

We define $n_i(t)$ the number of chains of length i at t with t taken just before a replication. Assuming i even,

$$n_i(t+\tau) = \sum_{j=0}^{\infty} n_{i/2+j}(t) l(i+2j, 2j, \tau).$$

This is because just before a replication, there are $n_{i/2+j}(t)$ chains of length $i/2+j$. Then, just after the replication, these chains are of length $i+2j$. Time $t+\tau$ is just before the next replication. With probability $l(i+2j, 2j, \tau)$, these chains of length $i+2j$ have lost $2j$ bacteria on their edges and are now chains of length i. We sum over all the possible j. In the long time, $n_i(t) = Cp_i \exp(\lambda t)$, with λ the long term growth rate, that is such that $\exp(\lambda\tau) = \mathcal{N}$, with $\mathcal{N}$ the largest eigenvalue of the matrix. Replacing $l(i+2j, 2j, \tau)$ by its expression as in equation (11) of the main text, the previous equation leads to:

$$\mathcal{N} p_i = \sum_{j=0}^{\infty} p_{\frac{i}{2}+j} e^{-\alpha\tau(i-1+2j)} (e^{\alpha\tau} - 1)^{2j} \frac{2^{2j}}{(2j)!}.$$

We compare the 1st term of the sum to the rest of the sum. The first term is$p_{\frac{i}{2}} e^{-\alpha\tau(i-1)}$, the rest of the sum is:

$$\sum_{j=1}^{\infty} p_{\frac{i}{2}+j} e^{-\alpha\tau(i-1+2j)} (e^{\alpha\tau} - 1)^{2j} \frac{2^{2j}}{(2j)!}.$$

We divide both by $e^{-\alpha\tau(i-1)}$. Then this is equivalent of comparing $p_{i/2}$ with:

$$S = \sum_{j=1}^{\infty} p_{\frac{i}{2}+j} e^{-2j\alpha\tau} (e^{\alpha\tau} - 1)^{2j} \frac{2^{2j}}{(2j)!}.$$

When $\alpha\tau$ is large, links typically break before the next replication, so there is little cluster formation, and it is thus expected that the chain length distribution decreases fast with i, so that for $j > 0$, $p_{\frac{i}{2}+j} \ll p_{i/2}$. When $\alpha\tau$ is small, replication is fast compared to the typical time for one link to break. However, for a chain of length $i/2$, τ has to be compared to $(i/2-1)/\alpha$, the typical first link breaking time, thus we expect n_i to decrease with i for i large enough, thus $p_{\frac{i}{2}+j} \lesssim p_{i/2}$ for $j > 0$. We define B such as $p_{\frac{i}{2}+j} \leq B, \forall j > 0$. For $\alpha\tau$ large, $B \ll p_{i/2}$, and for $\alpha\tau$ small, if i is large enough, $B \lesssim p_{i/2}$. Then:

$$S \leq \sum_{j=1}^{\infty} B(1 - e^{-\alpha\tau})^{2j} \frac{2^{2j}}{(2j)!} = B\left(cosh\left(2(1 - \exp(-\alpha\tau))\right) - 1\right)$$

For $\alpha\tau$ large, $(cosh\left(2(1 - \exp(-\alpha\tau))\right) - 1) \simeq cosh(2) - 1 \simeq 2.7$. For $\alpha\tau$ small, $(cosh\left(2(1 - \exp(-\alpha\tau))\right) - 1) \simeq 2(\alpha\tau)^2 \ll 1$.

Thus in the case of $\alpha\tau$ large, S is small relative to $p_{i/2}$ because S is smaller than a few units times B, with B much smaller than $p_{i/2}$. In the case of $\alpha\tau$ small, S is small relative to $p_{i/2}$ because S is of the order of $(\alpha\tau)^2 B$, with B of the order of $p_{i/2}$. Then this justifies the assumption that only the first term of the sum matters:

$$p_i \simeq \frac{1}{\mathcal{N}} p_{i/2} \exp\left(-\alpha\tau\left(i-1\right)\right).$$

We assume $i = 2^k$, with k an integer. This is obviously true only for a very restricted set of i, but however this still yields an approximation for how the distribution depends on i for large i. Then, by recursion,

$$p_i \simeq \frac{1}{\mathcal{N}^k} p_1 \exp\left(-\alpha\tau\left(i(1+1/2+1/2^2+...+1/2^{k-1})-k\right)\right).$$

If i is large enough, $1+1/2+1/2^2+...+1/2^{k-1} \simeq 2$. Remembering that k was defined as $i = 2^k$, the result is:

$$p_i \simeq p_1 i^{\frac{\alpha\tau - \log(\mathcal{N})}{\log(2)}} \exp\left(-2\alpha\tau i\right).$$

When $\alpha\tau \gg 1$, links typically break before the next replication, thus there is little impact of the clustering on the growth. Consequently, the growth will be close to its value in the absence of clustering, i.e. doubling every τ, and thus in this limit $\mathcal{N} = 2$:

$$p_i \simeq p_1 i^{\frac{\alpha\tau}{\log(2)} - 1} \exp\left(-2\alpha\tau i\right).$$

This rough approximation allows to explain the core of the observed distribution.

5 Model with force-dependent breaking rate

5.1 Model and equations

A link between bacteria may consist of several sIgA bonds, and the number of bound sIgA may not be exactly the same from one inter-bacteria link to the next, but as sIgA are likely well mixed, many per bacteria and that bacteria are similar to each other, let us assume that link heterogeneity is negligible. The links could break if there was some process degrading the sIgA, but the sIgA are thought to be very stable[6]. Another possible explanation for link breaking is that the antigen get extracted from the bacterial membrane, at a rate which may depend exponentially with the force applied on the link[7][8]. If the forces are produced by the bacteria themselves (such as by flagella rotation), there are likely to fluctuate on timescales which are short compared to the time between two bacterial replications, and their distribution is likely to be the same for all links, so it would be appropriate to model their effect as a fixed breaking rate, the same for all the links. Another force is the hydrodynamical force exerted by the flow on the bacterial chain.

The flow in the digestive system is complex and not precisely characterized. Longer bacterial chains may also bend and their shape have complex interactions with the flow. Here, we present the simplest model taking into account the forces exerted by the flow on the link breaking rate. We aim to capture the main plausible effects of the flow when the link breaking rate is force-dependent.

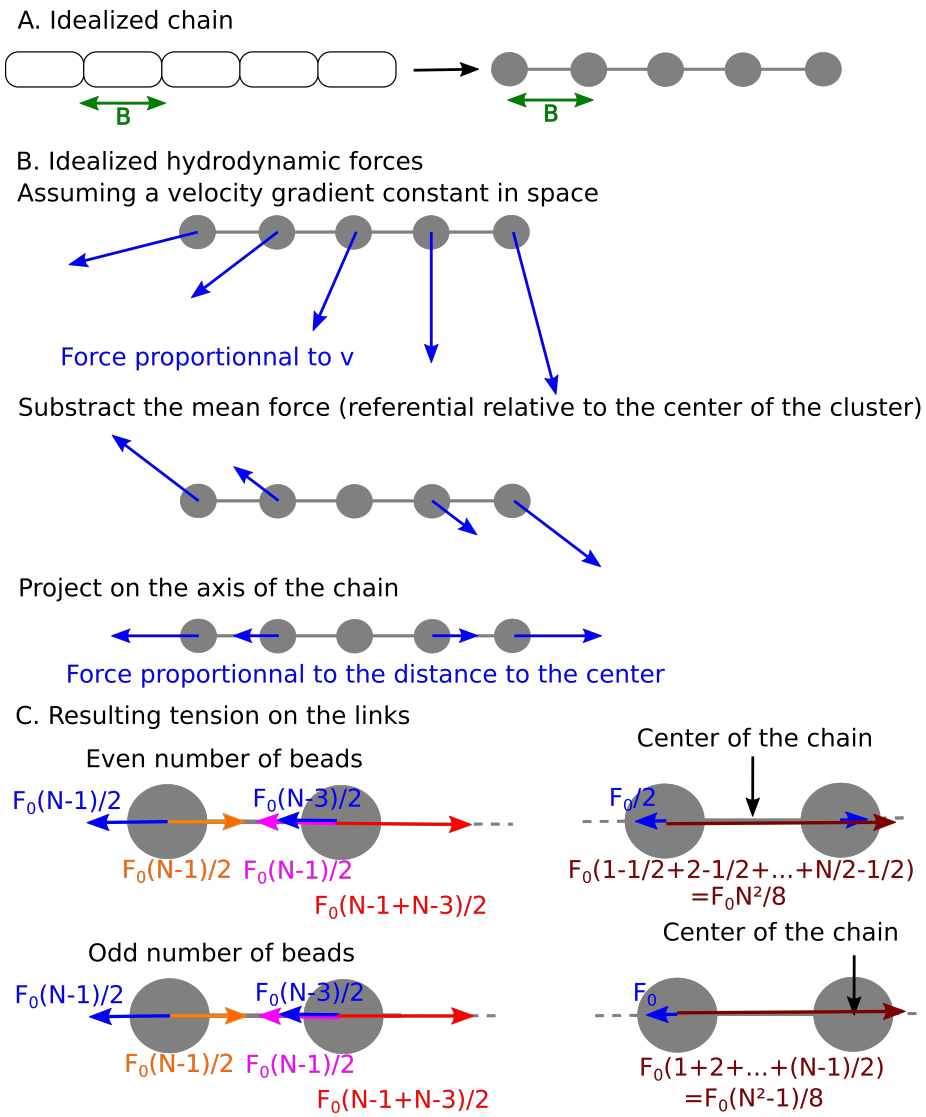


Figure S6: Schematic of the forces applied to the chain. **A** We assume a straight chain of beads with no hydrodynamic interactions between them. **B** We substract the average force to put ourselves in the referential of the center of the chain, as the total force will translate the whole chain and not impact on the forces on the links. We focus on the forces parallel to the chain that will impact the tension between the links. **C** Sum of the forces on each bead, for chains with even and odd number of beads.

Let us take a linear chain of N bacteria, each of length B. Let us approximate it by a rigid chain with beads linked by straight rods of length B (pannel A of figure S6). Let us assume that the rods are infinitely thin so they do not interact with the flow, and let us neglect the hydrodynamical interaction between the beads, so they each are subject to the same frictional force for a given fluid

velocity, and, given that the typical Reynolds numbers in the digestive tract are relatively low[9], then the viscous force on each bead is proportional to the flow velocity.

Then, let us assume that the velocity gradient in the fluid is constant around the chain. The rationale for this approximation is that the typical scales of the flow are of the order of the centimeter / millimeter (for instance in a mouse, the cecum typical size is in the cm range), much larger than typical bacterial chains (the length of one bacteria is about $2\mu m$, so even chains of dozens of bacteria remain small compared to the typical flow scale), thus we take a linear approximation of the velocity field in the vicinity of a bacterial chain.

Then, if we take the sum of the forces on the whole chain, it will be equal on mN multiplied by the acceleration of the center of mass of the chain, with m the mass of each bead. When all the beads move together, there is no force on the links, thus let us take the referential relative to the center of the chain, and subtract the mean force on each bead (panel B of figure S6). Then, there remain forces perpendicular to the axis of the chains, and forces parallel to the axis of the chain. The forces perpendicular to the axis of the chain will make it rotate, and as they are perpendicular, they have no effect on the tension on the rods. Then, let us consider only the forces parallel to the chain.

In the example portrayed here, the chain is elongated. The reverse could happen, but in this case, the chain would likely buckle, and the force applied on the links would be small. The flow varies considerably in time, due to peristaltic motions[10][9]. There would be moments with no force and little breaking, and moments with larger forces and more breaking. The flow due to peristaltic motions changes on time scales short compared to the typical bacterial division time, thus we will assume that periods of low breaking and high breaking rates will be equivalent to an average effective breaking rate. Then let us consider the case of elongation only, as portrayed here.

As we assume here that the velocity gradient is constant, the relative fluid velocity grows linearly with the distance from the center of mass of the chain. Then the force on each bead is equal to F_0 multiplied by the distance to the center divided by B. We assume, following [7][8], that the breaking rate is dependent on the force. Thus, we define α and β such that the breaking rate of a link is $\alpha \exp(\beta F/F_0)$ if a force F is applied to the link. In the limit of small force, the breaking rate will be α, the same for all links, as in the base model. β is some constant caracterizing how much the stability of the link is force-dependent.

We can write the force on each bead (pannel C of figure S6). Then, here, because the chain is rigid and straight, the sum of the forces on each bead has to be zero. The tension on the outermost link will simply be equal to the flow force on the outermost bead, i.e. F_0 multiplied by its distance to the center divided by B, i.e. $(N-1)/2$ (both for chains of odd and even number of beads). On the next link, the tension has to compensate for the flow force on the second bead, plus the tension applied by the outermost link. Thus the tension on this link is $F_0((N-1)/2 + (N-1)/2 - 1)$, and so forth (this is analogous to modelling of breaking of polymer chains in elongational flows, as in[11]).

For N even, the force on the jth link starting from the outermost link will

be:

$$F_{jth\ link,N\ even} = F_0 \sum_{k=N/2-j+1}^{N/2} (k-1/2)$$

Using $\sum_{i=1}^{n} i = n(n+1)/2$, it can be rewritten as:

$$F_{jth\ link,N\ even} = F_0 \left(\frac{N^2}{8} - \frac{(N-2j)^2}{8} \right).$$

There are two links j^{th} away from the extremities, for j from 1 to $N/2-1$, and one central link, for which $j = N/2$. The breaking rate of a given link is $\alpha \exp(\beta F/F_0)$ with F the total force applied to the link. Then the total breaking rate of one chain of length N even is:

$$\alpha \exp\left(\frac{\beta N^2}{8}\right) \left(1 + 2 \sum_{k=2}^{N/2} \exp\left(-\frac{\beta}{2}(k-1)^2 \right) \right). \tag{S1}$$

An outermost link of a chain of length $N+1$ (with N even, $N+1$ is odd) breaks at rate $\alpha \exp(\beta N/2)$. There are two such links for each chain. This and equation (S1) lead to equation (30) of the main text:

$$\frac{dn_i}{dt} = -rin_i - \alpha n_i \exp\left(\frac{\beta i^2}{8}\right) \left(1 + 2 \sum_{k=2}^{i/2} \exp\left(-\frac{\beta}{2}(k-1)^2 \right) \right) + r(i-1)n_{i-1} + 2\alpha n_{i+1} \exp\left(\frac{\beta i}{2}\right)$$

For N odd, the force on the jth link starting from the outermost link will be:

$$F_{jth\ link,N\ odd} = F_0 \sum_{k=(N-1)/2-j+1}^{(N-1)/2} k.$$

Simiarly to the N even case, we can rewrite:

$$F_{jth\ link,N\ odd} = F_0 \left(\frac{N^2}{8} - \frac{(N-2j)^2}{8} \right)$$

Because of the two sides, there are two links j for each chain, for j from 1 to $(N-1)/2$. The breaking rate of a given link is $\alpha \exp(\beta F/F_0)$ with F the total force applied to the link. Then the total breaking rate of one chain of length N odd is:

$$2\alpha \exp\left(\frac{\beta N^2}{8}\right) \sum_{k=1}^{(N-1)/2} \exp\left(-\frac{\beta}{2} \left(k - \frac{1}{2} \right)^2 \right) \tag{S2}$$

An outermost link of a chain of length $N+1$ (with N odd, $N+1$ is even) breaks at rate $\alpha \exp(\beta N/2)$. There are two such links for each chain. Then, this and equation (S2) lead to equation (31) of the main text for the evolution in time of the mean number of chains of odd length i:

$$\frac{dn_i}{dt} = -rin_i - 2\alpha n_i \exp\left(\frac{\beta i^2}{8}\right) \sum_{k=1}^{(i-1)/2} \exp\left(-\frac{\beta}{2} \left(k - \frac{1}{2} \right)^2 \right) + r(i-1)n_{i-1} + 2\alpha n_{i+1} \exp\left(\frac{\beta i}{2}\right).$$

5.2 Additional figure for the force-dependent model: replication rate maximizing the growth rate as a function of β

Figure S7 shows that the rate of replication maximizing the growth rate of free bacteria increases exponentially with β, which represent the strength of the dependence of the breaking rate on the force applied to the link.

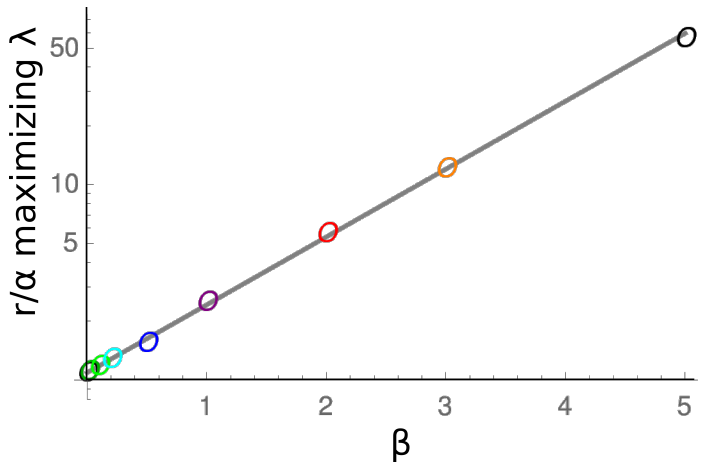


Figure S7: Log of the value of r/α maximizing the growth rate in the force-dependent breaking rate model as a function of β. The points are numerical maximums, the line is $1.09 \times \exp(0.8\beta)$. 1.09 is the value of (r/α) maximizing the growth rate for the base model (i.e. for $\beta \to 0$).

5.3 Force-dependent model: approximation for the chain length distribution.

We start from equations (30) and (31), and assume that for t long enough, $n_i \simeq Cp_i \exp(\lambda t)$ (with λ the largest eigenvalue). Then,

$$\lambda p_i = -rip_i - \alpha p_i \exp(\beta i^2/8)X + r(i-1)p_{i-1} + 2\alpha p_{i+1} \exp(\beta i/2) \qquad \text{(S3)}$$

with $X = 1 + 2\sum_{j=1}^{i/2-1} \exp(-\beta j^2/2)$ (i even) or $X = 2\sum_{j=1}^{(i-1)/2} \exp(-\beta(j - 1/2)^2/2)$ (i odd). Let us now determine which terms dominate in this expression.

For i large enough, $\lambda \ll ri$. Thus λp_i is negligible relative to rip_i.

For both i even and odd, X is a converging sum which tends to a finite number when i increases. Let us denote its limit $Y = 1+2\sum_{j=1}^{\infty} \exp(-\beta j^2/2) = \theta_3(0, \exp(-\beta/2)))$ in the even case, and $Z = 2\sum_{j=1}^{\infty} \exp(-\beta(j-1/2)^2/2) = \theta_2(0, \exp(-\beta/2)))$ in the odd case, with θ_i the Jacobi Theta functions. Thus, because β is positive, for i large enough, $ri \ll \alpha \exp(\beta i^2/8)X$.

The remaining main terms in equation (S3) are:

$$-\alpha p_i \exp(\beta i^2/8)X + r(i-1)p_{i-1} + 2\alpha p_{i+1} \exp(\beta i/2) \simeq 0.$$

The first term is negative, the two others are positive. Then we have to determine which of $r(i-1)p_{i-1}$ and $2\alpha p_{i+1}\exp(\beta i/2)$ dominates. If $2\alpha p_{i+1}\exp(\beta i/2)$ dominates, $\alpha p_i \exp(\beta i^2/8)X \simeq 2\alpha p_{i+1}\exp(\beta i/2)$, thus $p_{i+1}/p_i \simeq \exp(\beta i(i/8 - 1/2))X$, which for i large enough means that the long the chain, the more of it, which would diverge and does not make sense in this system. Thus $\alpha p_i \exp(\beta i^2/8)X \simeq r(i-1)p_{i-1}$,

$$\frac{n_i}{n_{i-1}} \to \frac{p_i}{p_{i-1}} \simeq \frac{r}{\alpha}\frac{i-1}{X}\exp\left(-\beta\frac{i^2}{8}\right). \tag{S4}$$

This approximation is valid for large chain sizes. We assume that it is valid for any chain length. As this expression is small and decreasing quickly with increasing i, p_1 will be close to 1. Then, as:

$$\frac{p_i}{p_1} = \Pi_{j=2}^{i}\frac{p_j}{p_{j-1}},$$

and using the known expression for the sum of the squares $\sum_{j=1}^{i} j^2 = (n + 3n^2 + 2n^3)/6$ and expression (S4):

$$p_{i,even} \simeq \left(\frac{r}{\alpha}\right)^{i-1}\frac{(i-1)!}{Y^{i/2}Z^{i/2-1}}\exp\left(-\frac{\beta}{8}\left(-1+\frac{i+3i^2+2i^3}{6}\right)\right)$$

$$p_{i,odd} \simeq \left(\frac{r}{\alpha}\right)^{i-1}\frac{(i-1)!}{Y^{(i-1)/2}Z^{(i-1)/2}}\exp\left(-\frac{\beta}{8}\left(-1+\frac{i+3i^2+2i^3}{6}\right)\right)$$

These two equations can be combined, and ultimately lead to:

$$p_i \simeq \left(\frac{r}{\alpha}\right)^{i-1}\frac{(i-1)!}{Y^{floor(i/2)}Z^{floor((i-1)/2)}}\exp\left(-\frac{\beta}{8}\left(-1+\frac{i+3i^2+2i^3}{6}\right)\right)$$

5.4 Additional figure for the force-dependent model: chain length distribution for other values of r/α

In panel 2J of the main text, we represented the distribution of chain lengths in the model with force-dependent link breaking rate for $r/\alpha = 1$. In figure S8 we represent the distributions for different values of r/α. Overall, the shapes are similar, and the smaller r/α is (as well as the larger β is), the better the analytical approximation works.

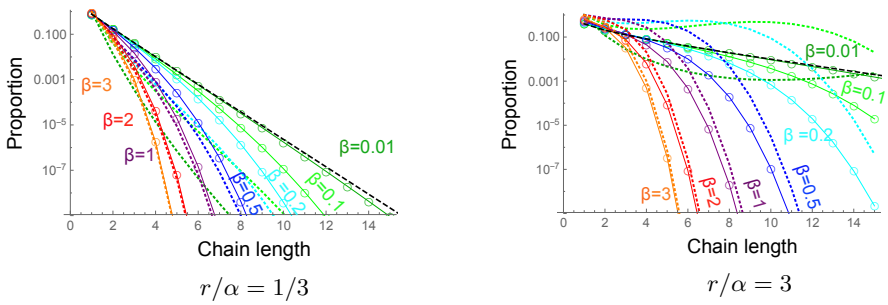


Figure S8: Chain length distribution, as in figure 2J, except for the value of r/α: model with force-dependent breaking rates. Each color represents a different β: $\beta = 0.01$ ($n_{max} = 20$), $\beta = 0.1$ ($n_{max} = 15$), $\beta = 0.2$ ($n_{max} = 15$), $\beta = 0.5$ ($n_{max} = 15$), $\beta = 1$ ($n_{max} = 15$), $\beta = 2$ ($n_{max} = 10$), $\beta = 3$ ($n_{max} = 10$). The black dashed lines are the numerical results for the base model, equivalent to $\beta = 0$. The curves for $\beta = 0.01$ (dark green) are almost overlaid with the curves for $\beta = 0$. The colored dotted lines the analytical approximation (equation (18) of the main text), and the black dotted line the approximation for the base model (equation (5) of the main text).

6 Zoomed-in distribution for all models

Figure S9 represents the zoomed-in chain length distributions of the right panels of figure 2 of the main text.

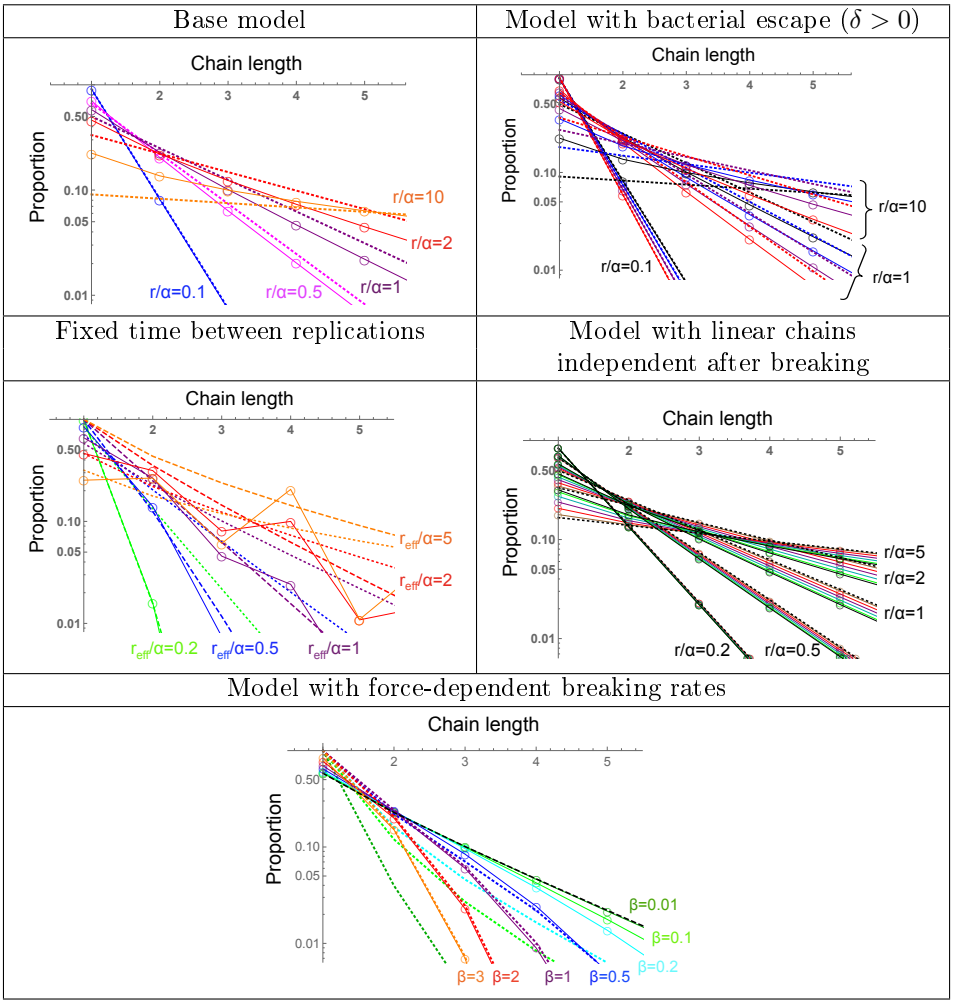


Figure S9: Zoomed-in chain length distributions: everything as in the right panels of figure 2 of the main text, but centered on the head of the distributions. Solid lines and open circles: numerical results. **Base model**: $n_{max} = 40$, dotted lines: approximation (5) of the main text (almost overlaid with the numerical results for $r/\alpha = 0.1$). **Model with bacterial escape**: $\delta = \delta' = \delta'' = \mathbf{0}$, 0.1, 0.2, 0.3, $c = c' = 0$, $n_{max} = 40$. Dotted lines: approximation (9) of the main text. **Fixed time between replications**: $r_{eff} = \log(2)/\tau$, $n_{max} = 32$. Approximation (18) (dashed lines), numerical result in the base model (dotted lines). $r/\alpha = 0.2$, 0.5, 1, 2, 5. **Model with linear chains independent after breaking**: $n_{max} = 100$. The dotted black lines are the approximate distribution (27) of the main text for each r/α, which is the exact distribution for $q = 1$. The colors represent $q = \beta = 0$, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, **1**. All curves are almost overlaid for small r. **Model with force-dependent breaking rates**: chain length distribution for $r/\alpha = 1$. Each color represents a different β: $\beta = 0.01$ ($n_{max} = 20$), $\beta = 0.1$ ($n_{max} = 15$), $\beta = 0.2$ ($n_{max} = 15$), $\beta = 0.5$ ($n_{max} = 15$), $\beta = 1$ ($n_{max} = 15$), $\beta = 2$ ($n_{max} = 10$), $\beta = 3$ ($n_{max} = 10$). The colored dotted lines the analytical approximation (18) of the main text, and the black dotted line the approximation (5) of the main text for the base model.

7 Experimental data

7.1 Methods

We perform a new analysis on images that were produced for [4]. We briefly describe below the experiments from which the images were produced, and describe our analysis.

Mice, which were previously vaccinated with a peracetic-acid inactivated S.Typhimurium strain (PA-S.Tm), were pretreated with 0.8g/kg ampicillin sodium salt in sterile PBS. 24h later, mice received 10^5 CFU of a 1:1 mix of mCherry-(pFPV25.1) and GFP-(pM965) expressing attenuated S. Tm M2702. For imaging, cecum content was diluted gently 1:10 w/v in sterile PBS containing 6μg/ml chloramphenicol to prevent growth during imaging. 200μl of the suspension were transferred to an 8-well Nunc Lab-Tek Chambered Coverglass (Thermo Scientific) and imaged at 100x using the Zeiss Axiovert 200m microscope. To determine the distribution of bacteria in aggregates, n=25 high power fields per mouse were randomly selected and imaged for mCherry and GFP fluorescence. For some mice, sequential sampling was done, these mice were terminally anaesthetised and artificially respirated cecum content was sampled by tying off part of the cecum each hour for 3h. More details about the experimental procedures can be found in [4].

We analyzed all the images for the early data points (4 and 5 hours) of experiments starting from a low inoculum (10^5), to minimize the clustering from random encounters. Only the linear chains were counted. Images are for the red and green fluorescence, so complex clusters with two colors were not counted. The data were analyzed manually. The images are available as supplementary materials:

- images4h.zip contains the images for 3 of the mice only sampled at 4h.
- images4h_others.zip contains the images for the other 4 mice only sampled at 4h.
- images5h.zip contains the images of the mice only sampled at 5h
- imagesseq4h.zip contains the images at 4h of mice sampled sequentially
- imagesseq5h.zip contains the images at 5h of mice sampled sequentially

.

7.2 Results

For linear chains, we obtained the length distribution detailed in table 1 and shown on figure 4 of the main text.

Given the bumpy shape of the experimental distribution, we chose to fit the data with the fixed replication time model for the figure 4 of the main text. In this model, the only adjustable parameter is r_{eff}/α. For a given r_{eff}/α, we obtain the theoretical chain length distribution p_i by numerical resolution of the equations. As the data concerns only chains of length 2 and longer, we renormalize this distribution as $p_i/(1-p_1)$. Then, given that we observe a total of $N_{exp} = 313$ chains, for the theoretical process the probability to observe k

chain length	4h PI o (7 mice)	4h PI s (2 mice)	5h PI o (4 mice)	5h PI s (2 mice)	total
2	21	30	17	38	106
3	22	4	9	5	40
4	51	9	25	9	94
5	7	0	1	3	11
6	5	3	3	4	15
7	10	1	5	3	19
8	12	0	4	3	19
9	1	0	0	0	1
10	1	0	0	1	2
11	1	0	0	0	1
12	0	0	1	1	2
13	0	0	1	1	2
14	0	0	1	0	1

Table 1: Table of the linear chains counted on the images from several experiments, either with mice sampled once (o), or with mice sampled sequentially (s).

chains of length i is $Poisson(k, \lambda = N_{exp}p_i/(1-p_1)) = \lambda^k \exp(-\lambda)/k!$. Naming k_i the number of observed chains of length i, we then take the value of r_{eff}/α which maximizes:

$$\log(likelihood) = \sum_{i=2}^{N_{max}} \log\left(\frac{1}{k_i!}\left(\frac{N_{exp}p_i}{1-p_1}\right)^{k_i} \exp\left(-\frac{N_{exp}p_i}{1-p_1}\right)\right)$$

Given that we compute numerically the theoretical distribution, we have to use a finite N_{max}, whereas in theory it should be taken as infinite. In practice, the theoretical values for the proportions of long chains become quickly very small, and thus the probability to observe chains of such lengths is very small for any r_{eff}/α. Thus in order to fit r_{eff}/α, N_{max} does not matter, provided that it is taken large enough. We have tried $N_{max} = 16$ and 32, and we obtain the same fitted value $r_{eff}/\alpha = 4.1$. If we were to look for a value with more significant digits, this choice for N_{max} would matter more.

To quantify our impression that there are fewer long chains observed than expected, we performed the following calculations. Taking $r_{eff}/\alpha = 4.1$ and $N_{exp} = 313$, the expected number of chains of length 15 or longer is 3.7, whereas none is observed, which with a Poisson assumption has a probability $\simeq 0.025$ to occur. This probability seems low. Either this is a low probability but still happened (and if we look at a bit shorter chains, the expected average number of chains of length 9 and longer is 11.7, and 9 of them are actually observed, which is relatively close); or there is some process limiting the number of long chains. There are two main possibilities for the number of long chains to be limited: there could be an experimental bias limiting the observation of long chains (see discussion below); or there could be some force-dependence of the breaking rates, which would effectively act as a cut-off for the chain length (see figure 3 of main text), as in this case, breaking rates increase considerably with

chain length.

7.3 Discussion

The data may be biased. The mass of one bacterium is about one pg, and its density is about 10% more than the water density[12, 13], the thermal energy at ambient temperature is of the order of $4.10^{-21}J$, and gravity g is of the order of $10m/s^2$, thus thermal fluctuations will lift an individual bacterium by typically 4 μm higher than the bottom. Thermal fluctuations will have two effects:

- The average height of the center of gravity of chains will decrease with their length. This is confocal microscopy, which typical optical section is less than 1μm, focused close to the cover slip. This may bias the distribution by missing smaller chains.
- Longer chains are not rod-like, their shape fluctuate. It is apparent on the microscopy images that parts of long chains may get out of focus. The longer the chain, the less likely that it is entirely in the focus, and thus chains will look smaller than they are.

We focus on the chain length distribution because this quantity is more easily accessible by experimental measurements, at the end of an experiment. Comparing models and experiments enables to check whether the data is compatible with a process of growing and breaking of clusters; and determine which specific model is closest to the data. However, some models cannot be distinguished, no matter how much data is available for the chain length distribution. For example in the model with bacterial escape and the model where chains can remain independent after breaking, there are two parameters to fit (r/α and δ, or r/α and q). It is likely that fitting would mainly select a value for r/α, since the distribution does not depend much on the second parameter in both cases. These models could not be distinguished from the base model. On the other hand, models with different distribution shapes – either in the force-dependent model or in the fixed division time one– could be distinguished, provided that the bias can be overcome, and that more data can be collected.
We could fit the fixed replication time model to the data, and this strengthened our hypothesis that the chains are generated by a process of enchained growth and link breaking. However, there is somewhat less long chains observed than expected (especially in the range of lengths 14 to 16). One possibility could be that the breaking rate is force dependent. If we had 10 times more of unbiased data, we could answer whether there really is a deficit of longer chains. If there is indeed a deficit of longer chains, then we should combine the model of force-dependent breaking rates with the model with fixed replication time, to be able to make quantitative comparisons. To be more effective, comparison would likely require more data, as there would be two free parameters, r/α and β. If there is no deficit of longer chains in the range up to 16, then the simple model with fixed replication time predicts that we would get access to the distribution up to length 24 (and be at the limit for lengths 28 and 32) with 100 times more data than in current experiments (see figure S10). Thus overall we would need at least 10 times and likely 100 times more data for a more quantitative assessment.

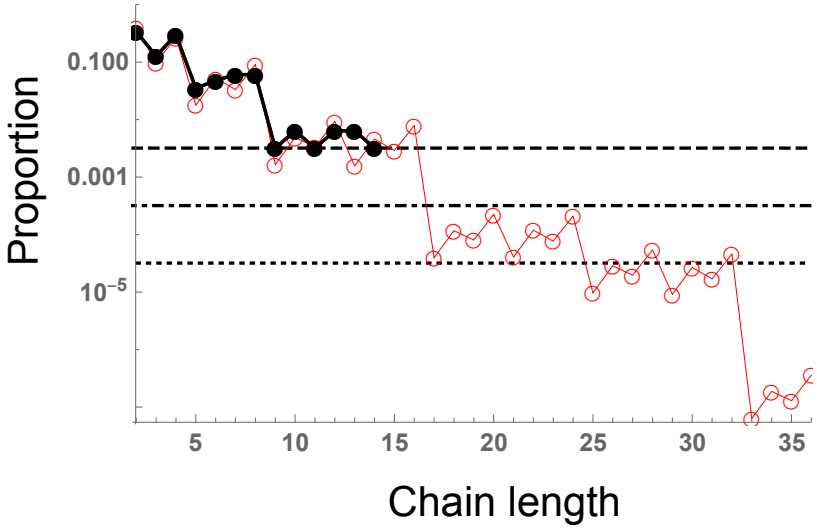


Figure S10: Comparison of chain length distribution for the experimental data (black) and the model with fixed replication time (red). This figure is as figure 4 of main text, but extended up to length 36. The black dashed line represents limit of one chain in the data. The black dot-dashed line represents what this limit would be if there was 10 times more data, and the black dotted line represents what this limit would be if there was 100 times more data. $r_{eff}/\alpha = 4.1$ (with $r_{eff}/\alpha = \log(2)/(\alpha\tau)$).

Increasing the amount of data would not necessarily require to sacrifice more mice, but merely to take more images for each cecum content. The challenge would be to do so with no bias, and with very standardized conditions so that the images are taken in conditions close enough so as to automate the chain detection and length count.

It would be also very useful if there would be ways to estimate the breaking rate in independent experiments, for instance injecting (without breaking them nor perturbing the system) chains of non-replicating bacteria of controlled length, and measuring how the length distribution changes over time. Then, as the replication rate can be estimated by other measures (dilution of non-replicating plasmids), we could get an estimate of the replication rate over the breaking rate, which would considerably constrain the fitting of the chain length distribution, and thus give more strength to the conclusions achieved.

References

[1] Berg HC. Random walks in biology. Princeton University Press; 1993. 1

[2] Berg HC. E. coli in Motion. Springer Science & Business Media; 2008. 1

[3] Elgeti J, Winkler RG, Gompper G. Physics of microswimmers—single particle motion and collective behavior: a review. Reports on progress in physics. 2015;78(5):056601. 1

[4] Moor K, Diard M, Sellin ME, Felmy B, Wotzka SY, Toska A, et al. High-avidity IgA protects the intestine by enchaining growing bacteria. Nature. 2017;544(7651):498–502. doi:10.1038/nature22058. 1, 16

[5] World Health Organization & Food and Agriculture Organization. Risk assessments of Salmonella in eggs and broiler chickens. vol. http://www.who.int/foodsafety/publications/micro/salmonella/en/; 2002. 1

[6] Brandtzaeg P. Role of secretory antibodies in the defence against infections. International Journal of Medical Microbiology. 2003;293(1):3–15. doi:10.1078/1438-4221-00241. 7

[7] Evans E, Berk D, Leung A. Detachment of agglutinin-bonded red blood cells. I. Forces to rupture molecular-point attachments. Biophysical Journal. 1991;59(4):838–848. doi:10.1016/S0006-3495(91)82296-2. 7, 9

[8] Evans EA, Calderwood DA. Forces and bond dynamics in cell adhesion. Science. 2007;316(5828):1148–1153. 7, 9

[9] Lentle R, Janssen P. Physical characteristics of digesta and their influence on flow and mixing in the mammalian intestine: a review. Journal of Comparative Physiology B. 2008;178(6):673–690. 9

[10] Hulls C, Lentle RG, de Loubens C, Janssen PW, Chambers P, Stafford KJ. Spatiotemporal mapping of ex vivo motility in the caecum of the rabbit. Journal of Comparative Physiology B. 2012;182(2):287–297. 9

[11] Odell J, Keller A. Flow-induced chain fracture of isolated linear macromolecules in solution. Journal of Polymer Science Part B: Polymer Physics. 1986;24(9):1889–1916. 9

[12] Bratbak G, Dundas I. Bacterial dry matter content and biomass estimations. Applied and environmental microbiology. 1984;48(4):755–757. 18

[13] Baldwin WW, Myer R, Powell N, Anderson E, Koch AL. Buoyant density of Escherichia coli is determined solely by the osmolarity of the culture medium. Archives of microbiology. 1995;164(2):155–157. 18