Report on a problem studied at the UK Mathematics-in-Medicine Study Group Oxford 2005
< https://mmsg.mathmos.net/uk/2005/knockout-proteins/ >

Bacterial proteins that knock out cell defences: cruise missiles or cannonballs?

Alexander Fletcher, Gary Mirams, Jonathan Rougier,
Ilya Shvartsman, Alex Walter

Mathematics in Medicine Study Group, 2005

Abstract

Many bacterial pathogens are able to suppress and modulate host cell defences by 'injecting' effector proteins into the cytoplasm of host cells. These secretion systems are found in both animal and plant pathogens. It is currently not known how bacteria deliver effector proteins to their targets within a host cell, although several potential strategies have been identified. In this study group report, continuum and statistical models of effector protein invasion are considered. We use the models predict that effector proteins cannot rely on diffusion alone to reach their targets, but must also employ some co-localisation mechanism.

1 Introduction

Some bacteria are able to attack plant and animal cells by 'injecting' proteins known as 'effectors' through the cell membrane into the cytoplasm of host cells. To deliver these effector proteins, the bacteria use a secretion apparatus known as a Type III Secretion System (TTSS). This system, which is found both in plant and animal pathogens, consists of one or more injection needles or 'pili', each ∼2-3nm in diameter, that the bacteria uses to 'drill' through the host cell membrane. The area of contact between the bacteria and cell is $\sim 0.5\mu$m in diameter. Each pilus has a hollow core that is only 20Å in diameter. When a pilus has reached the cytoplasm, the bacteria releases proteins through it. Since a bacterial protein in its natural conformation is ∼5nm is diameter, these proteins are injected with only secondary structure (α-helices and β-sheets).

Upon entry into the host cell, these effector proteins fold up into their active form, and suppress host cell defences so that the bacteria are able to grow as parasites on, or in, the infected cell. Alternatively, the effector proteins may activate host cell death mechanisms so that the infected cell undergoes apoptosis; this can provide a large food source for the bacteria, as well as disabling all the host's defence mechanisms. Using type III secretion systems, bacteria $\sim 2\mu$m in diameter are able to disable eukaryotic cells $\sim 50\mu$m in diameter, around 100-1000 times their own size. Type III secretion systems inject into the host cells a relatively small number (10^3-10^4) of effector proteins compared with the number of proteins in a host cell (10^9-10^{10}). Effector proteins are diverse in their structure and function, attacking a wide range of host targets in many different areas of the cell including the nucleus, cell membrane, cytoplasm, mitochondria, chloroplasts (in plant cells) and membrane-bound organelles (see Figure 1). For an effector protein to deactivate or evade host defence mechanisms, it must find its target as fast as possible after entering the cell.

The most-well studied bacterial proteins are those that effectively target a plant cell's membrane. Each plant cell contains 40-400 mitochondria and 10^3-10^4 cytoplasmic targets for bacterial proteins. Cytoplasmic targets may be $\sim 0.1\mu$m in diameter. Unlike an animal cell, a plant cell also contains a large vacuole and 10-20 chloroplasts within its interior, as well as an rigid outer cell wall. Each bacterium injects 10-25 different types of proteins into animal cells. In plant cells the corresponding number is either 1-5 or more than 30, depending on the type of plant cell. Experiments usually involve the insertion of antibiotics into a bacterium attached to a host plant cell, which inhibit bacterial protein synthesis. This successfully stops the bacterium from invading the plant cell if it is carried out in the first few seconds after attachment, but not if it is carried out after 1-2 hours, since by that time the bacterium has produced and injected all the effector proteins necessary to overcome the host cell's defences. Real-time imaging of effector delivery by *Salmonella enterica*, for example, has shown that the injection of 6000 SipA proteins begins 10-90 seconds after the bacteria docks with a host cell, and continues for 100-600 seconds. It is known (by observing the physiological reaction of the host cell) that the cell detects these proteins and reacts within minutes; degradation pathways become more active, the cell wall is thickened, and protein synthesis is stepped up as defence-related genes are activated.

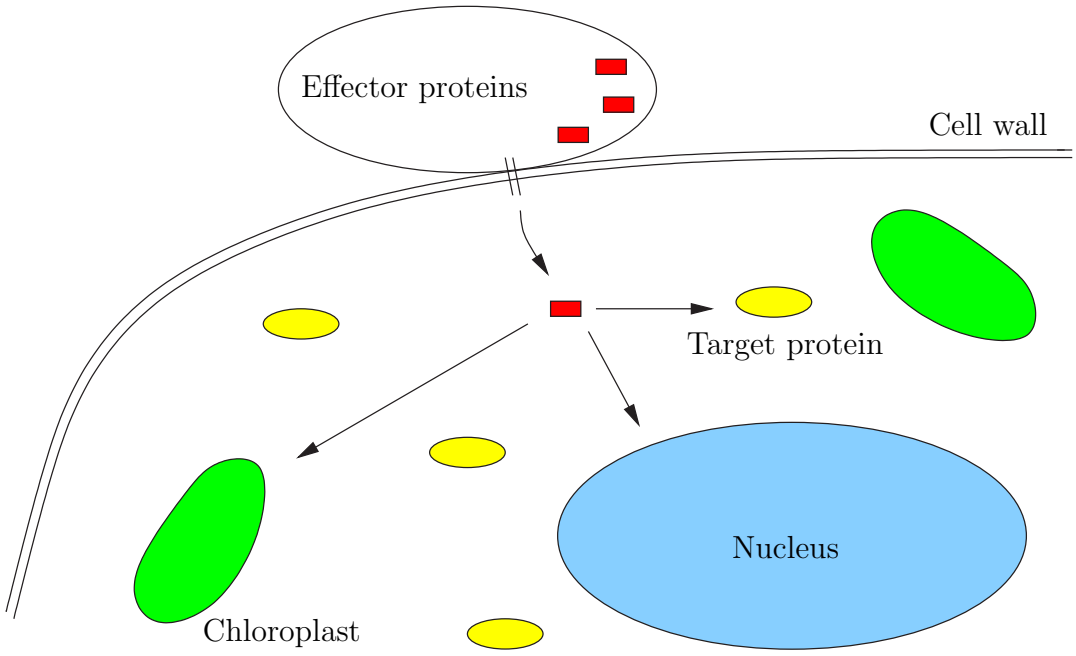


Figure 1: Schematic diagram of the injection of effector proteins by a type III secretion system into a plant cell.

A single collision between a bacterial protein and a target may be enough to alter the host cell's gene expression within the nucleus; yet in general, many collisions are required for an effect to occur that can be observed experimentally. Since an effector requires a cytoplasmic chaperone to get to, and through, the nucleus, we would expect a large amount of time to be taken up by the bacterial protein actually getting to the right chaperone. This raises the question of how a limited number of bacterial proteins injected at a specific location in the periphery of the cell are able to encounter target proteins with sufficient speed and efficiency to disable the host cell defence mechanisms. In particular, scientists wish to know which transport mechanism(s) are used by effector proteins to locate their targets within minutes. A number of possibilities exist:

1. Targeted point of entry: bacteria dock with a host cell only at sites at which membrane target proteins are located.

2. Simple diffusion: effector proteins reach their targets by diffusing through a host cell's cytoplasm.

3. Membrane diffusion: effector proteins remain bound to the inside of the cell membrane of a host cell, and diffuse across it to membrane target proteins.

4. 'Shrapnel' diffusion: upon injection, effector proteins alter the function of small host proteins; these can diffuse around the host cell more quickly than the effector proteins themselves, in order to attack their targets.

5. Active targeting: effector proteins have evolved to include host cells' 'address labels', which ensure that effectors are rapidly transported by a host cell's active transport systems to specific targets.

The Study Group investigation examined the effect of cell architecture, spatial distribution and concentration of effectors and targets, and passive versus active transport processes on the likelihood of effector-target contact. In particular, we aimed to answer the following questions:

1. How important are active spatial localisation mechanisms in enabling bacterial proteins to connect with host targets?

2. What experiments must be performed in order for us to be able to describe and interpret protein targeting mechanisms in bacteria–host interactions?

Continuum and stochastic frameworks were employed to model different aspects of this problem. A continuum approach was used to simulate the concentrations of effector and target proteins and study how they may spread around the cell using diffusion, cytoplasmic streaming and protein interactions. A discrete statistical model was also developed, in order to evaluate the time it should take for well-mixed individual proteins to come into contact over different spatial scales.

2 Continuum approach

An important question in the investigation of this problem is how long a bacterial protein takes to meet its target within the a cell, as opposed to the average time for observational effects to occur. At present, it is known that the host cell's fate is played out within a couple of hours, but whether this is determined much earlier remains unclear. A continuum approach is therefore used to model the invasion of a population of effector proteins into a plant cell, in order to suggest some of the timescales over which some of the different transport mechanisms mentioned above may operate. In particular, we investigate the relative contributions of diffusion and convection to effector protein invasion.

To ensure that as much sunlight as possible is absorbed by every chloroplast in a plant cell, a convective flow termed 'cytoplasmic streaming' exists within the cytosol. Cytoplasmic streaming is not the result of external forces acting on the cytosol, but instead the result of the drag caused by tiny molecular motors, which ratchet along the filaments of the cytoskeleton within the cell. As might be expected, cytoplasmic streaming is reduced deep inside a plant. To model this effect on protein transport within a plant cell, we model the cytoplasm as a viscous fluid. To determine the relative contributions of inertial and viscous effects, it is necessary to evaluate the Reynolds number

$$Re = \frac{\rho U L}{\mu}, \tag{1}$$

for this system, where ρ is the cytosolic density, U is a characteristic flow speed of the cytoplasm, L is a characteristic length scale of the flow and μ is the coefficient of viscosity of the cytosol. Using approximate values $\rho \sim 1000\text{kg m}^{-3}$, $U \sim 10^{-6}\text{m s}^{-1}$, $L \sim 5 \times 10^{-5}$ m and $\mu \sim 10^{-3}$ Pa s gives

$$Re \sim 5 \times 10^{-5} \ll 1 \tag{2}$$

for cytoplasmic streaming. This indicates that viscous effects outweigh inertial effects and that the approximations of low Reynolds number flow are appropriate.

Most mature plant cells have a central membrane-bound vacuole, which often takes up more than 90% of the cell interior. This is shown schematically in Figure 2. The vacuole is filled with water, enzymes, inorganic ions and other substances. Besides being a plant cell's storage center, the main purpose of a vacuole is to maintain the cell's structural stability by providing a turgor pressure against the cell wall. This is why the vacuole is so comparatively large in the plant cell. Another function of the vacuole is to push all the cell's cytoplasm against the cellular membrane, and thus keeps the chloroplasts closer to light.

For simplicity, we model the plant cell as two-dimensional, and 'unwrap' the cell by treating it as a two-dimensional pipe with periodic boundary conditions, employing coordinates (x,y) as shown in Figure 3. We model the cell wall ($y = H$) and vacuole surface ($y = -H$) as rigid boundaries, at which we enforce no-slip conditions on the cytoplasmic flow. We impose periodic boundary conditions at the boundaries $x = 0$ and $x = L$ to close the system. Note that $H \sim 5 \times 10^{-6}$ m.

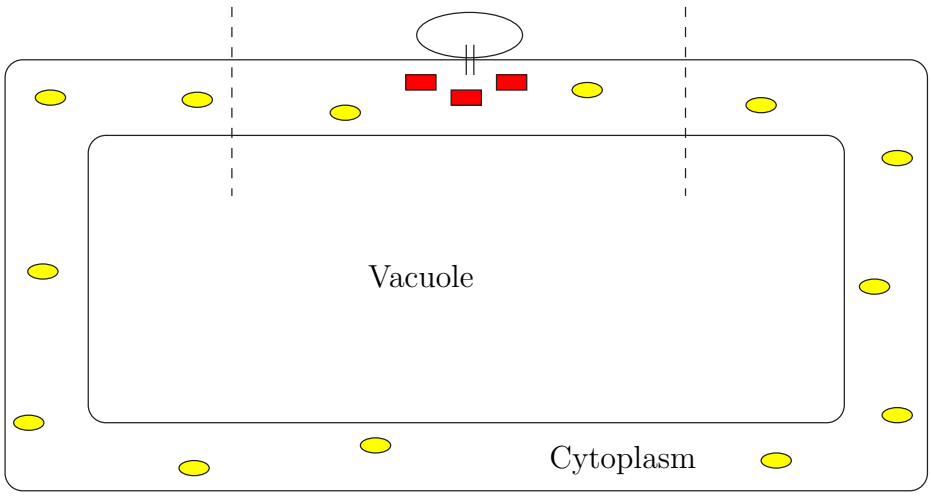


Figure 2: Schematic diagram of the injection of effector proteins by a type III secretion system into a plant cell. Effector proteins are represented as red rectangles and target proteins are represented by yellow ellipses. We simplify the problem by considering only the region bounded by the dotted lines and enforcing periodic boundary conditions.

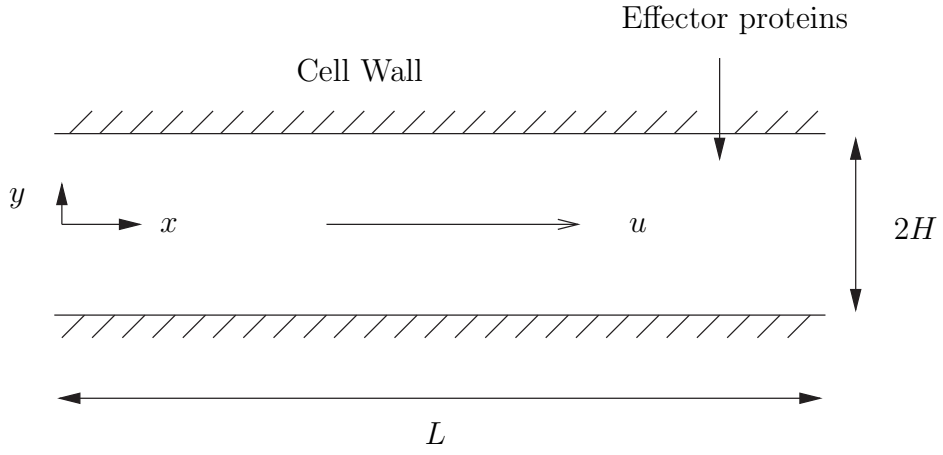


Figure 3: Schematic diagram of the model

We impose no-slip boundary conditions at the membrane of the vacuole and at the cell wall. We solve the Navier-Stokes equations under the small Re assumption and a constant pressure gradient $\Delta P = -dp/dx$ between the 'ends' of the cytosol caused by the molecular motors, finding that the cytoplasmic velocity field $\mathbf{u} = (u(y), 0)$ satisfies

$$u = U\left(1 - \frac{y^2}{H^2}\right) \tag{3}$$

in the region $0 \leq x \leq L$, $-H \leq y \leq H$, where $U = \Delta P H^2/2\mu$. This is a so-called Poiseuille flow.

Within the cytoplasm, we consider a population of effector proteins E and target proteins P, assuming a single species in each case for simplicity. Suppose that an effector protein suppresses a target protein by reacting with it to form a new complex C in the reversible reaction

$$E + P \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C,$$

and that each species diffuses through the cytoplasm with constant diffusivity. Denoting the concentrations of effector proteins, target proteins and complexes by $E(x, y, t)$, $P(x, y, t)$ and $C(x, y, t)$ respectively, and denoting their diffusivities by D_E, D_P and D_C respectively, we obtain the reaction-convection-diffusion equations

$$\frac{\partial E}{\partial t} + U\left(1 - \frac{y^2}{H^2}\right)\frac{\partial E}{\partial x} = D_E\nabla^2 E - k_1 EP + k_{-1}C, \tag{4}$$

$$\frac{\partial P}{\partial t} + U\left(1 - \frac{y^2}{H^2}\right)\frac{\partial P}{\partial x} = D_P\nabla^2 P - k_1 EP + k_{-1}C, \tag{5}$$

$$\frac{\partial C}{\partial t} + U\left(1 - \frac{y^2}{H^2}\right)\frac{\partial C}{\partial x} = D_C\nabla^2 C + k_1 EP - k_{-1}C. \tag{6}$$

We suppose that initially, there is a uniform distribution of target proteins within the cytosol, that there are no complexes present and that effector proteins are injected at the point $(x_0, 1)$:

$$P(x, y, 0) = P_0, \tag{7}$$

$$C(x, y, 0) = 0, \tag{8}$$

$$E(x, y, 0) = E_0\delta(x - x_0)\delta(y - H). \tag{9}$$

We also impose periodic boundary conditions on the system:

$$[E]_{x=0}^{x=L} = [P]_{x=0}^{x=L} = [C]_{x=0}^{x=L} = \left[\frac{\partial E}{\partial x}\right]_{x=0}^{x=L} = \left[\frac{\partial P}{\partial x}\right]_{x=0}^{x=L} = \left[\frac{\partial C}{\partial x}\right]_{x=0}^{x=L} = 0. \tag{10}$$

The Péclet number for each protein population within the cytoplasm is defined to be

$$Pe = \frac{UL}{D}, \tag{11}$$

where D is the diffusion coefficient in cytoplasm of the population. The Péclet number is the ratio of the characteristic time scales for diffusion to convection, and is a measure of the relative contributions of the two mechanisms to protein transport within the cytosol. For each of the populations, D may be estimated using the Stokes-Einstein formula

$$D = \frac{kT}{6\pi R\mu} \tag{12}$$

where $k = 1.38 \times 10^{-23}$ J K^{-1} is the Boltzmann constant, $T = 298$ K is the absolute temperature and $R \sim 2.5 \times 10^{-9}$ n m is the 'radius' of each protein (as stated earlier). This estimate gives

$$D \sim \frac{1.38 \times 10^{-23} \times 298}{6 \times 3.142 \times 10^{-3} \times 2.5 \times 10^{-9}} \sim 10^{-10}\,\mathrm{m^2\,s^{-1}}, \tag{13}$$

although it has been shown that the cytoskeletal ctructures within the cytosol increase the effective viscosity of the cyotsol to the motion of large particles, decreasing the effective diffusivity. We therefore take a lower estimate $D \sim 10^{-12}\,\mathrm{m}^2\,\mathrm{s}^{-1}$. Calculating the Péclet number for these values of U, L and D, we obtain $Pe \sim 50$.

We choose the timescale to be the timescale for diffusion of effector proteins across the cytosol and non-dimensionalize equations (4)-(6) using the scalings

$$x = L\tilde{x}\,, \quad y = H\tilde{y}\,, \quad t = \frac{H^2}{D_E}\tilde{t}\,, \quad E = E_0\tilde{E}\,, \quad P = P_0\tilde{P}\,, \quad C = E_0\tilde{C}\,, \tag{14}$$

and, dropping tildes, obtain

$$\frac{\partial E}{\partial t} + \epsilon^2\, Pe\,(1-y^2)\frac{\partial E}{\partial x} = \epsilon^2\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} - \tilde{k}_1 EP + \tilde{k}_{-1}C\,, \tag{15}$$

$$\frac{\partial P}{\partial t} + \epsilon^2\, Pe\,(1-y^2)\frac{\partial P}{\partial x} = \tilde{D}_P\,\epsilon^2\frac{\partial^2 P}{\partial x^2} + \tilde{D}_P\,\frac{\partial^2 P}{\partial y^2} - \lambda\tilde{k}_1 EP + \lambda\tilde{k}_{-1}C\,, \tag{16}$$

$$\frac{\partial C}{\partial t} + \epsilon^2\, Pe\,(1-y^2)\frac{\partial C}{\partial x} = \tilde{D}_C\,\epsilon^2\frac{\partial^2 C}{\partial x^2} + \tilde{D}_C\,\frac{\partial^2 C}{\partial y^2} + \tilde{k}_1 EP - \tilde{k}_{-1}C\,, \tag{17}$$

where we have defined the non-dimensional parameters

$$\epsilon = \frac{H}{L}\,, \quad Pe = \frac{UL}{D_E}\,, \quad \tilde{k}_1 = \frac{k_1 P_0 H^2}{D_E}\,, \quad \tilde{k}_{-1}\frac{k_{-1}H^2}{D_E}\,,$$

$$\tilde{D}_P = \frac{D_P}{D_E}\,, \quad \tilde{D}_C = \frac{D_C}{D_E}\,, \quad \lambda = \frac{E_0}{P_0}\,. \tag{18}$$

Since a typical value for ϵ is 0.1, $\epsilon^2 Pe \sim 1$ and the left–hand side of equation (15) indicates that convection of effector proteins in the direction of cytoplasmic streaming is balanced by lateral diffusion.

Unfortunately, the system described above cannot be solved analytically. Although we may use the assumption $\epsilon \ll 1$ to neglect certain terms in the equations, there remains little data on the rates of reaction between effector proteins and target proteins. Several numerical simulations were performed to investigate the behaviour of this system for certain parameter values. Typical results are shown in Figure 4, in which the fraction of the cytosol occupied by protein complexes is plotted as a function of time. This fraction is obtained by numerically estimating the area of the domain

$$\{(x,y)\,:\,0 \le x \le 1,\ -1 \le y \le 1,\ C(x,y,t) > 0\} \tag{19}$$

and dividing this value by the area of the cytosol, which is 2 in nondimensionalized

units. The resulting fraction indicates the effective reach of the effector proteins into the cytosol. As Figure 4 shows, the presence of cytoplasmic streaming enhances the invasion of effector proteins into the cytosol, with total infiltration occuring sooner than in the case of diffusion only.

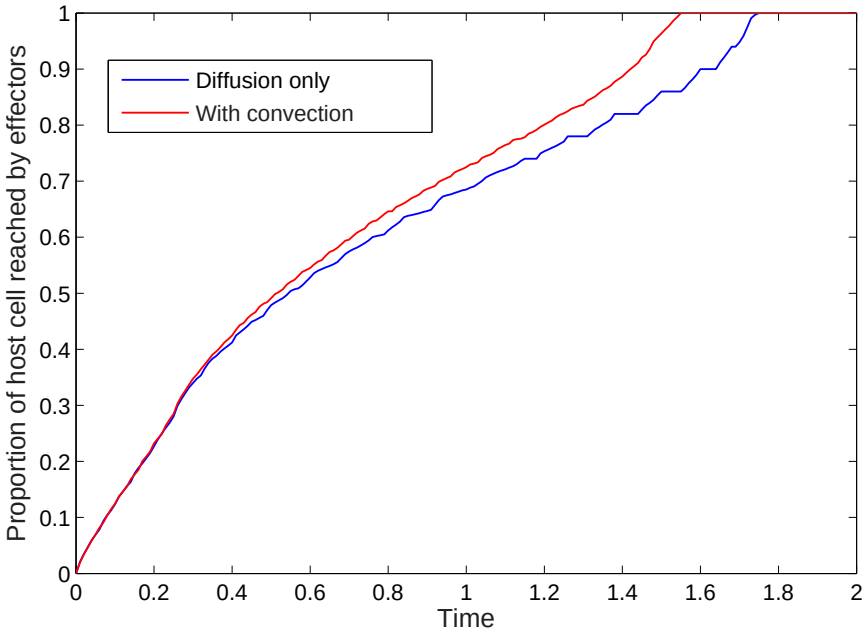


Figure 4: Typical numerical simulation of the model. The following parameter values were used for this simulation: $\epsilon = 0.1$, $Pe = 50$, and $\tilde{k}_1 = \tilde{k}_{-1} = \tilde{D}_P = \tilde{D}_C = \lambda = 1$. These parameter values were chosen for illustration purposes only.

Numerical simulations of our model provide evidence that cytoplasmic streaming enhances the transport of effector proteins through the cytosol of a host cell. This is because as an effector protein diffuses further through the cytosol from the cell wall, the magnitude of the convective flow that it experiences increases, causing it to travel further around the cell. In order to compare the time scales of protein invasion suggested by this model with that of other models, we suggest a comparison of the time taken for a certain percentage (e.g. 50%) of the effector proteins to become complexed with target proteins within the cytosol. This model could be compared to 'shrapnel' diffusion, for example, by removing cytoplasmic streaming and introducing a cascade of protein interactions

$$\begin{aligned} E + P_1 &\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C_1, \\ C_1 + P_2 &\underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} C_2, \end{aligned}$$

and so on, with each successive protein species have a larger diffusion coefficient. The role of defence proteins within the host cell could also be considered. However, the

Quantity	Units	Value	Symbol
♯ Effector proteins		6000	n_e
♯ Target proteins		1000	n_t
♯ Defence proteins		10^4	n_d
Total ♯ particles		10^{11}	N
Cell length	μm	50	ℓ
Protein ('large') diameter	μm	0.005	d_e
'Small' particle diameter	μm	0.002	d_s
Proportion of small particles in N	%	80	p_s
Typical particle velocity	$\mu\,\mathrm{m\,s^{-1}}$	1	V
Prob(join | bump)		0.667	γ
Prob(cell survives attack)		0.9	κ

Table 1: Typical values and notation

lack of experimental data on reaction rates prohibits any further prediction using this modelling approach.

3 The cell as a random environment

The problem is still to analyse the way in which a bacterium subverts a cell through its effector proteins. However, in this section we treat the cell as a random environment in which mixing is fully developed, so that encounter rates between the various particles to be found in the cell are spatially and temporally constant. The particles we concentrate on are (i) effector proteins injected by the bacterium, (ii) target proteins, and (iii) defence proteins that can tag and destroy the effector proteins. The latter two species 'belong' to the cell. The values we use in our numerical illustrations are shown in Table 1. At this stage it is an open question whether we can treat the problem accounting only for these particles and effectively relegating the remainder to the background. Initially we thought this would be reasonable, but now it seems much less certain that this is possible, and so we leave this as something to be deduced within a treatment of all particles in the cell. Therefore Table 1 also contains the quantity N, the total number of particles (or particle–equivalents) in the cell.

We consider the amount of time necessary for the effectors to overwhelm the cell, where 'overwhelm' in this case denotes achieving a given number ω of successful collisions between effectors and target proteins. A successful collision is one in which an effector collides *and* binds with a target protein. We divide this calculation into two parts. First, how many collisions are required before there are ω successes?

Second, what is the mean time between collisions for an effector in the cell? The product of these two values gives us an idea of the time-scale on which the cell is subverted by the bacterium.

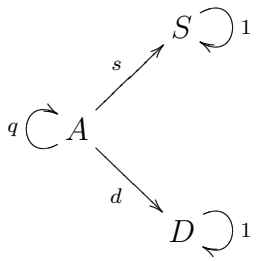
Our calculation proceeds on the basis that the cell represents a random environment. Therefore, if we find that the time–scale is very much longer than that typically observed, we may conclude (subject to our various modelling choices) that the bacterium does not rely on chance alone in subverting the cell. In other words, this provides support for the hypothesis that some type of co–localisation mechanism is responsible for increasing the success–rate of the effectors above that commensurate with simply random encounters.

3.1 Number of collisions until ω successes

We track all n_e effectors individually, but this is only possible if we treat them as operating independently and identically. More precisely, we treat the path of each effector through the cell interior as a discrete Markov process from collision to collision, and we treat the transition probabilities of this process as fixed. That is to say, no matter where the effector is, no matter what its path has been, and no matter what has happened to the other effectors, the chance of an effector which is still active colliding with any given type of particle depends only on the type. The probabilities by type are determined simply by the proportions of each type among all particles in the cell.

This treatment requires that mixing in the cell is fully developed: if mixing were incomplete, then initially effectors would mostly collide with other effectors, since they are all injected at a similar location in the cell membrane. But since this incomplete mixing would delay the time until any given effector collides with a target protein, our computed value for the number of collisions will be a conservative estimate (i.e. too low). This treatment also requires that successful collisions between effectors and target proteins, and unsuccessful (from the point of view of the effector) collisions between effectors and defence proteins, have a negligible effect on the relative proportions of the particles in the cell; this seems assured for our calculation as long as ω is not an unduly large proportion of $\min\{n_e, n_t, n_d\}$.

Any given effector can be in one of three states $\{A, S, D\}$. States S and D are absorbing states. State S indicates that the effector has collided and successfully joined with a target protein; state D indicates that the effector has (unsuccessfully) collided and joined with a defence protein (technically, been tagged and ubiquitinated by a defence protein). The transition probabilities for the effector from one collision to the next are



$$s \triangleq \gamma \frac{n_t}{N}$$
$$d \triangleq \gamma \frac{n_d}{N}$$
$$q \triangleq 1 - s - d$$

where γ is the probability of binding given a collision.

For a typical effector, there is a probability s of success in the first collision; a probability qs of success in the second; $q^2 s$ of success in the third, and so on. The probability that a typical effector is ultimately successful is the sum of all of these:

$$\sum_{r=0}^{\infty} q^r s = s\,(1-q)^{-1} = s\,(s+d)^{-1}. \tag{20}$$

Similarly, the probability that a typical effector is ultimately unsuccessful is $d\,(s+d)^{-1}$. For the next part of the calculation, we need the probability that our effector is successful in exactly k collisions *given that it is ultimately successful*. Let $\tau = k$ represent this event, then this probability is given by

$$\mathrm{Prob}\big(\tau = k\big) = \frac{q^{k-1}s}{s\,(1-q)^{-1}} = (1-q)\,q^{k-1}. \tag{21}$$

The distribution function for τ is therefore

$$\mathrm{Prob}\big(\tau \le k\big) = \sum_{j=1}^{k} (1-q)\,q^{j-1} = 1 - q^k. \tag{22}$$

Now we consider the situation in which exactly v of the n_e effectors are successful, for the moment treating v parametrically. We can compute the distribution function of $\tau_{(\omega)} \mid v$, which is the ω^{th} order statistic for the collection of quantities $\tau_1, \ldots, \tau_v$. This uncertain quantity represents the number of collisions before ω of the v effectors have been successful. Treating our effectors as independent, we have

$$\begin{aligned}
\mathrm{Prob}\big(\tau_{(\omega)} \le k \mid v\big) &= \sum_{j=\omega}^{v} \binom{v}{j} \mathrm{Prob}\big(\tau \le k\big)^j \,\big[1 - \mathrm{Prob}\big(\tau \le k\big)\,\big]^{v-j} \\
&= \sum_{j=\omega}^{v} f_{\mathrm{B}}\big(j; 1-q^k, v\big) \\
&= 1 - F_{\mathrm{B}}\big(\omega - 1; 1-q^k, v\big)\ ,
\end{aligned} \tag{23}$$

where $f_{\mathrm{B}}\big(i; p, n\big)$ denotes the binomial probability function for i successes out of n trials with a success probability of p, and $F_{\mathrm{B}}\big(i; p, n\big)$ is the distribution function of the same.

In general v is not known, so we must integrate it out. Taken over all n_e effectors, v itself has a binomial distribution with success probability $s\,(s+d)^{-1}$, so we have

$$\begin{aligned}
&\mathrm{Prob}\big(\tau_{(\omega)} \le k \mid v \ge \omega\big) = \\
&\qquad (1-\kappa)^{-1} \sum_{v=\omega}^{n_e} f_{\mathrm{B}}\big(v; s\,(s+d)^{-1}, n_e\big)\ \big[1 - F_{\mathrm{B}}\big(\omega - 1; 1-q^k, v\big)\,\big],
\end{aligned} \tag{24a}$$

where

$$\kappa \triangleq \text{Prob}\big(v < \omega\big) \equiv F_{\text{B}}\big(\omega - 1; s/(s+d), n_e\big)\,. \tag{24b}$$

Note that we still condition on $v \geq \omega$ because the condition $\tau_{(\omega)} \leq k$ is otherwise not well-defined.

In this model, the only way in which a cell can avoid being subverted is if fewer than ω effectors are successful. Therefore κ represents the probability that the cell survives the attack by the bacterium. If κ were close to zero, there would be little point in the cell investing in defence proteins, since in this case they could only delay but not prevent the ultimate subversion of the cell. On the grounds that the cell has evolved defence proteins for a reason, we take their presence to imply that κ is substantially greater than zero. In other words, we believe that the cell has a reasonable chance of surviving the attack. We do not have any clear idea about the appropriate value for ω, but we can infer it from (24b) if we have a value for κ; our inferred value will also depend on the values of n_e, n_t, n_d and γ. Our values in Table 1 give $\omega = 575$, which is about 10% of the total number of effectors.

Using equation (24) and our values in Table 1, we compute the distribution function for $\tau_{(\omega)}$ when $\omega = 575$. In fact, this can be tricky to compute exactly when n_e is large, but in this case it can be reasonably well–approximated by using the gaussian approximation to the binomial distribution. The result is shown in Figure 5. The median value for $\tau_{(\omega)}$ in this case is about 6×10^7 collisions.

3.2 Time between collisions

The second half of the calculation in this section is to quantify the time between collisions, so that we can approximate the time necessary for a bacterium to subvert the cell. From the previous subsection, we will take the number of collisions for subversion to be 6×10^7, noting that we are accounting for all possible collisions in a cell containing 10^{11} particles of various types, the vast majority of which are not relevant to our analysis, except insofar as they get in the way.

We can use a simple argument similar to that used in statistical mechanics, modified to allow for different sizes of particle in the cell. We will use two sizes of particle, small and large.

Effector proteins have diameter d_e, which we also take to be the standard diameter of 'large' particles in the cell. Any other large particle whose centre lies within $d_e/2 + d_e/2 = d_e$ of the centre of an effector protein may be assumed to be in contact with it. Over the course of 1 second, our effector sweeps out a volume $\pi(d_e)^2 \times V$ in which collision with another large particle is possible. If we multiply this by the density of large particles per unit volume in the cell, $(1 - p_s)N/\ell^3$, then we have an order of magnitude number of collisions per second between effectors and large particles. We can make the same calculation with effectors and small particles, for which the swept volume per second is $\pi(d_e/2 + d_s/2)^2 \times V$. Adding the two together

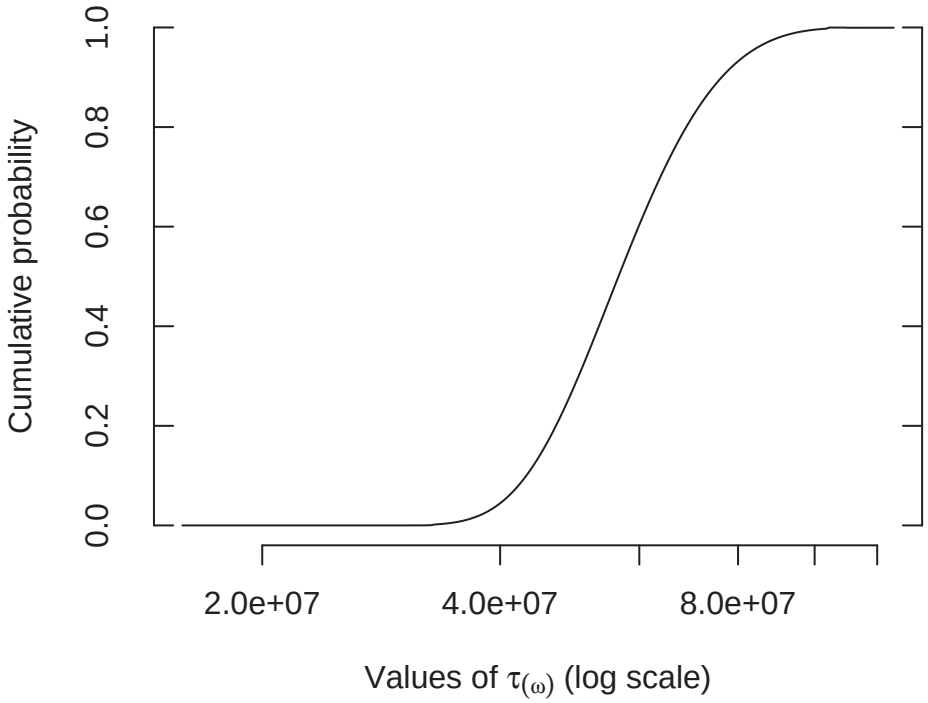


Figure 5: Cumulative distribution function for $\tau_{(\omega)}$ with $\omega = 575$.

gives us the number of collisions per second between effectors and either small or large particles, and then inverting gives us the number of seconds per collision, $\hat{t}$, where

$$\hat{t}^{-1} \equiv \pi \, (d_e)^2 \, V \, \frac{(1 - p_s)N}{\ell^3} + \pi \, \frac{(d_e + d_s)^2}{4} \, V \, \frac{p_s \, N}{\ell^3}.$$

It should be noted that this is an approximate result; statistical mechanics introduces further corrections. From the values in Table 1, we deduce

$$\hat{t}^{-1} \approx 13 + 25 \Longrightarrow \hat{t} \approx 0.03$$

or, in terms of its reciprocal, any particular particle experiences about 37 collisions per second. This estimate does not seem unreasonable.

We now put this value together with the number of successful collisions necessary to overwhelm the cell. The result is that in our random environment, if the bacterium is successful, then it subverts the cell in about $6 \times 10^7 \times 0.03 \approx 2 \times 10^6$ seconds, i.e. in about 20 days. This is about three orders of magnitude greater than the observed time–scale, providing evidence in favour of the existence of some kind of co–localisation mechanism.

3.3 Further observations

These calculations confirm that N, the total number of particles (or particle–equivalents) is relevant in the calculation: typical values for $\tau_{(\omega)}$ are not linear in N, and consequently N does not cancel in a product such as $\tau_{(\omega)} \times \hat{t}$. However, the role of N appears to be minor. For example, setting $N = 2 \times 10^4$, roughly the number of involved particles in the cell, we compute a time–scale of about 14 days. Therefore uncertainty about N is unlikely to be a major contributory factor to our uncertainty about the time–scale in the random environment.

There is a feeling among the group that $V = 1\,\mu\,m\,s^{-1}$ represents the wrong kind of velocity, because it represents aggregate circulatory motion (convection) in the cell, rather than the random motion of individual particles. Seperate considerations from the continuum modelling and from the density of particles in the cell suggest that a value one or two orders of magnitude smaller might be better. It would then follow that our value for the time–scale of subversion is one or two orders of magnitude bigger, strengthening our conclusion about the need for a co–localisation mechanism.

4 Second random environment model

In this section, we use a different statistical model to try to estimate the time for invasion. We introduce the following model of a cell. The cell, which we assume for simplicity to be cubical, is divided into

$$N = \left(\frac{d}{l}\right)^3$$

compartments, where d is a diameter of a particle and l is the length of a cell. Each target protein and defence protein are assumed to occupy exactly one compartment. For simplicity we assume that the size of an effector is negligibly small compared to the size of other involved particles, so that more than one effector may end up in the same compartment. We also introduce time discretization $t_n = n\Delta t$, where

$$\Delta t = \frac{\text{average diameter of a particle}}{\text{average velocity of a particle}}.$$

At time t_0, we assume that particle species are uniformly distributed among the compartments. If an effector is in a compartment that is also occupied by a target or a defence protein, the effector binds with it with probability γ. At time t_1, the particles are 'reshuffled', but the number of effectors, targets and defence proteins is decreased: the decrease in effectors is equal to the total decrease in target/defence proteins, which is equal to the number of bindings at time t_0. This process is then repeated for t_2, and so on. Our goal is to estimate the critical time t_K when the number of targets is reduced by a given factor (say, is halved) or to a given number ω.

At each timestep, the probability that a given effector will end up in the same compartment with a target and bind to it is equal to $\gamma n_t/N$, where n_t is the current

number of 'free' target proteins. Similarly, the probability that a given effector will end up in the same compartment with a defence protein and bind to it is equal to $\gamma n_d/N$, where n_d is the current number of defence proteins.

Let us introduce the following notation:
X_n is the number of free effectors in the cell at time t_n;
Y_n is the number of free target proteins in the cell at time t_n;
Z_n is the number of free defence proteins in the cell at time t_n;
ξ_n^Y is the number of effectors which bind to targets at time t_n;
ξ_n^Z is the number of effectors which bind to defence proteins at time t_n;

Here, values for X_0,Y_0 and Z_0 may be estimated from experimental data.

Clearly, we have an identity

$$Y_n + Z_n - X_n = \text{const} = Y_0 + Z_0 - X_0 =: c_0.$$

The dynamics of the number of free effectors are governed by the equation:

$$X_{n+1} = X_n - \xi_n^Y - \xi_n^Z. \tag{25}$$

We would like to take the mathematical expectation of both sides of equation (25). The conditional random variables $\xi_n^Y|X_n, Y_n$ and $\xi_n^Z|X_n, Z_n$ have binomial distributions with the number of trials equal to X_n and probabilities of success $\gamma Y_n/N$ and $\gamma Z_n/N$ respectively. Therefore

$$\begin{aligned} E[\xi_n^Y|X_n, Y_n] &= \frac{\gamma}{N} X_n Y_n, \\ E[\xi_n^Z|X_n, Z_n] &= \frac{\gamma}{N} X_n Z_n. \end{aligned}$$

Thus equation (25) implies

$$\begin{aligned} E[X_{n+1}] &= E[X_n] - E[E[\xi_n^Y|X_n, Y_n]] - E[E[\xi_n^Z|X_n, Z_n]] \\ &= E[X_n] - \frac{\gamma}{N} E[X_n \underbrace{(Y_n + Z_n)}_{c_0 + X_n}] \\ &= \left(1 - \frac{\gamma c_0}{N}\right) E[X_n] - \frac{\gamma}{N} E[X_n^2] \\ &= \left(1 - \frac{\gamma c_0}{N}\right) E[X_n] - \frac{\gamma}{N} \mathrm{V}ar\,[X_n] - \frac{\gamma}{N} (E[X_n])^2. \end{aligned} \tag{26}$$

We now conjecture that $\mathrm{V}ar\,[X_n] \leq n\mathrm{V}ar\,[X_1 - X_0]$. To show this, observe that

$$\begin{aligned} \mathrm{V}ar\left[X_n\right] &= \left(\mathrm{V}ar\left[(X_n - X_{n-1}) + \cdots + (X_1 - X_0) + X_0\right]\right. \\ &= \sum_{k=0}^{n-1} \mathrm{V}ar\left[X_{k+1} - X_k\right] + \text{covariances.} \end{aligned}$$

If one can show that the covariances are negative, then our conjecture is proved, and this does seem to be the case. This is because the greater $X_{k+1} - X_k$ (the number of bindings with targets at time k) is, the smaller the number of bindings at time $m > k$ is likely to be, since less targets and effectors will be remaining.

Note that

$$\mathrm{V}ar\left[X_1 - X_0\right] = \mathrm{V}ar\left[\xi_0^Y + \xi_0^Z\right] = X_0 p_0 (1 - p_0), \tag{27}$$

where $p_0 := \gamma(Y_0 + Z_0)/N$. Therefore equation (26) implies

$$E[X_{n+1}] \geq \left(1 - \frac{\gamma c_0}{N}\right) E[X_n] - \frac{n\gamma}{N} X_0 p_0 (1 - p_0) - \frac{\gamma}{N} (E[X_n])^2. \tag{28}$$

Dividing both sides of inequality (28) by $E[X_n]$, we obtain

$$\frac{E[X_{n+1}]}{E[X_n]} \geq 1 - \frac{\gamma c_0}{N} - \frac{\gamma(1 - p_0)}{N} \cdot \frac{n p_0 X_0}{E[X_n]} - \frac{\gamma}{N} E[X_n]. \tag{29}$$

It turns out that the third term on the right–hand side of equation (29) is negligibly small compared to other terms (see calculation at the end). Since $E[X_n] < X_0$, we obtain from equation (29) neglecting the aforementioned term

$$\frac{E[X_{n+1}]}{E[X_n]} \geq 1 - \frac{\gamma c_0}{N} - \frac{\gamma}{N} X_0, \tag{30}$$

hence

$$E[X_n] \geq \left(1 - \frac{\gamma}{N}(c_0 + X_0)\right)^n X_0 = \left(1 - \frac{\gamma}{N}(Y_0 + Z_0)\right)^n X_0. \tag{31}$$

When the number of effectors decreases from X_0 to X_n, we expect that the number of targets will decrease to

$$E[Y_n] = Y_0 - \frac{Y_0}{Y_0 + Z_0}(X_0 - E[X_n]). \tag{32}$$

We now estimate the value of n such that $E[Y_n]$ is half of Y_0. We have

$$Y_0 = 2E[Y_n] = 2\left(Y_0 - \frac{Y_0}{Y_0 + Z_0}(X_0 - E[X_n])\right), \tag{33}$$

therefore

$$E[X_n] = X_0 - \frac{Y_0 + Z_0}{2}. \tag{34}$$

Note that the cell is not expected to die if $Y_0 + Z_0 > 2X_0$, i.e. if the number of targets or defence proteins to be hit is too large.

From equations (31)–(34) we obtain

$$\left(1 - \frac{\gamma}{N}(Y_0 + Z_0)\right)^n X_0 \leq X_0 - \frac{Y_0 + Z_0}{2}\,, \tag{35}$$

from which we obtain a lower bound for n:

$$n \geq \frac{\ln\left(1 - \frac{Y_0+Z_0}{2X_0}\right)}{\ln\left(1 - \gamma\frac{Y_0+Z_0}{N}\right)} \approx -\frac{N}{Y_0 + Z_0} \ln\left(1 - \gamma\frac{Y_0 + Z_0}{2X_0}\right). \tag{36}$$

Here we have used the fact that $Y_0 + Z_0 \ll N$ in our problem data.

A numerical simulation was conducted in order to verify the validity of equation (36). It predicts the value of n when the number of effectors is halved with a high precision for the problem data within 2–3 orders of magnitude of the given numbers.

Let us evaluate the right–hand side of equation (36) for the 'standard initial data' $N = 10^9$, $X_0 = 6 \times 10^3$, $Y_0 = 10^3$, $Z_0 = 10^4$, $\gamma = 2/3$, $v = 10^{-6}$, $d = 5 \times 10^{-9}$ and $l = 5 \times 10^{-6}$. In this case, we have

$$n \geq -\frac{10^9}{1.1 \times 10^4} \ln\left(1 - \frac{2}{3}\frac{1.1 \times 10^4}{1.2 \times 10^4}\right) \approx 10^5.$$

In this case, the numerical simulation provided the result $n \approx 1.2 \times 10^5$.

The time between 'reshufflings' is

$$\Delta t = \frac{d}{v} = \frac{5 \times 10^{-9}}{10^{-6}} = 5 \times 10^{-3},$$

so the real time it takes to hit 50% of the targets is of the order of at most $10^5 \times 5 \times 10^{-3} = 500$ seconds.

Note that to validate the above analysis, we must verify that $np_0X_0/(E[X_n])^2 \ll 1$. This is indeed the case, since $np_0X_0/(E[X_n])^2 \approx 10^{-3}$.

5 Conclusions

In this report we have developed continuum and probabilistic models of the invasion of bacterial effector proteins into a host plant cell. The continuum model considered the simplified two species case of effector proteins and target proteins binding in a reversible reaction, with binding representing the subversion of target proteins. The

model was used to suggest that cytoplasmic streaming aides effector proteins in their conquest of the host cell.

The first probabilistic model also considered the host cell's defence proteins and considered the case of a 'well–mixed' cell. We were able to show that the assumption of random collisions between effector proteins and their targets leads to an estimate of several weeks for the time for complete subversion of the host cell to occur. Since this far exceeds the observed time–scale, the model provides further evidence in favour of the existence of a co-localisation mechanism exploited by effector proteins. Further evidence for this was provided by the second probabilistic model.

These models could be extended in order to investigate in more detail which particular transport mechanisms are most likely to be used by effector proteins to locate their targets in a host cell within minutes. However, further development (of the continuum model in particular) is limited by experimental data on the diffusivities of different proteins and their rates of reaction.

6 Participants

The following people were involved with this problem during the study group: Ruth Baker, Guler Ergun, Alexander Fletcher, Sara Lombardo, Gary Mirams, Gail Preston, Jonathan Rougier, Ilya Shvartsman, Chris Stott, Alex Walter.