

RESEARCH ARTICLE

# Mathematical Model for Length Control by the Timing of Substrate Switching in the Type III Secretion System

Maulik K. Nariya<sup>1</sup>, Johnny Israeli<sup>1</sup>, Jack J. Shi<sup>1\*</sup>, Eric J. Deeds<sup>2,3,4\*</sup>

**1** Department of Physics and Astronomy, University of Kansas, Lawrence, Kansas, United States of America,

**2** Center for Computational Biology, University of Kansas, Lawrence, Kansas, United States of America,

**3** Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas, United States of America,

**4** Sante Fe Institute, Santa Fe, New Mexico, United States of America

\* [jshi@ku.edu](mailto:jshi@ku.edu) (JJS); [deeds@ku.edu](mailto:deeds@ku.edu) (EJD)



## Abstract

Type III Secretion Systems (T3SS) are complex bacterial structures that provide gram-negative pathogens with a unique virulence mechanism whereby they grow a needle-like structure in order to inject bacterial effector proteins into the cytoplasm of a host cell. Numerous experiments have been performed to understand the structural details of this nanomachine during the past decade. Despite the concerted efforts of molecular and structural biologists, several crucial aspects of the assembly of this structure, such as the regulation of the length of the needle itself, remain unclear. In this work, we used a combination of mathematical and computational techniques to better understand length control based on the timing of substrate switching, which is a possible mechanism for how bacteria ensure that the T3SS needles are neither too short nor too long. In particular, we predicted the form of the needle length distribution based on this mechanism, and found excellent agreement with available experimental data from *Salmonella typhimurium* with only a single free parameter. Although our findings provide preliminary evidence in support of the substrate switching model, they also make a set of quantitative predictions that, if tested experimentally, would assist in efforts to unambiguously characterize the regulatory mechanisms that control the growth of this crucial virulence factor.

## OPEN ACCESS

**Citation:** Nariya MK, Israeli J, Shi JJ, Deeds EJ (2016) Mathematical Model for Length Control by the Timing of Substrate Switching in the Type III Secretion System. *PLoS Comput Biol* 12(4): e1004851. doi:10.1371/journal.pcbi.1004851

**Editor:** Ashok Prasad, Colorado State University, UNITED STATES

**Received:** September 8, 2015

**Accepted:** March 6, 2016

**Published:** April 14, 2016

**Copyright:** © 2016 Nariya et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** The authors received no specific funding for this work.

**Competing Interests:** The authors have declared that no competing interests exist.

## Author Summary

The Type III Secretion System (T3SS) is a molecular needle that allows pathogenic bacteria (e.g. *Salmonella*) to inject proteins into host cells and control their behavior. Two mechanisms have been proposed to explain how bacteria regulate the length of the T3SS needle, but to date neither of these mechanisms has been subjected to any rigorous quantitative analysis. In this work we constructed a mathematical model for one of these mechanisms, namely length control via the timing of substrate switching. We showed that this model is quantitatively consistent with experimental data from *S. typhimurium*. In addition to

providing evidence for the substrate switching mechanism, our work provides a framework for future quantitative evaluation of length control in the T3SS.

## Introduction

The Type III Secretion System (T3SS) is a major virulence factor found in a large number of pathogenic bacteria, including prominent disease-causing organisms such as *Salmonella typhimurium*, *Shigella spp.* and *Yersinia pestis* [1]. The structure of this massive secretion machinery is homologous to that of the bacterial flagellum: it consists of a base complex that spans the inner and outer membranes of gram-negative bacteria and a long needle that protrudes into the surrounding environment [1–4]. A specific set of “tip” and “translocon” proteins at the end of the needle can form a pore in the plasma membranes of cells from the host organism (e.g. epithelial cells lining the large intestines in mammals), creating a narrow channel from the bacterial cytoplasm to that of the eukaryotic cell [1, 5–8]. The bacterium uses this channel to inject various “effector proteins” into the host cell; these proteins alter the behavior of the cell, in some cases leading to endocytosis of the bacterium and eventual invasion of the epithelial layer [1, 4]. Correct functioning of this “injectisome” apparatus is crucial for pathogenesis in many bacterial species, making the T3SS an attractive target for the development of antimicrobials and vaccines [1, 9, 10].

The structure of the T3SS system has been studied in detail, resulting in near-atomic resolution models of the base complex, the needle itself and the interaction of the needle with tip and translocon proteins [2–8, 11, 12]. Since functioning needle complexes are required for efficient pathogenesis, the assembly of this complex structure is highly regulated. One parameter of particular importance is the length of the needle: if the needle is too short, it will not emerge past the lipopolysaccharide (LPS) layer on the outer membrane and thus will not be able to engage eukaryotic cells. On the other hand, if the needle is too long, the efficiency of transport/injection may be reduced, and long needles can also fracture due to shear stress, which could render the T3SS inactive [1, 2, 5]. Fairly precise control of needle length is thus essential to injectisome function. The needle itself consists of a polymer of “needle” proteins, called PrgI in *Salmonella* and YscF in *Yersinia*; these proteins form a helical structure with a  $\sim 25$  Å pore in the middle [5–7]. The needle is assembled via export of these subunits from the base: as subunits are exported, they are incorporated into the end of the needle and the needle becomes longer. When the needle reaches approximately the correct length, this information is somehow relayed to the base complex, which ceases secreting needle proteins and begins to secrete the tip proteins and other factors needed for virulence and invasion [1].

Two mechanisms for controlling the length of the needle have been proposed. Cornelis and co-workers have posited that a dedicated protein might serve as a “molecular ruler.” The idea in this case is that the C-terminus of the ruler protein can interact with the base, while the N-terminus can interact with the tip of the growing needle. As a result of this interaction, the length of the ruler is then compared to the length of the needle, and when the ruler protein is stretched to its full length (say, adopting a fully extended  $\alpha$ -helical conformation), this information is relayed to the base. At that point the base begins secreting tip proteins and the needle complex becomes mature [1]. In *Yersinia spp.* and related organisms, the putative ruler protein has been identified as YscP: it has been shown that lengthening this protein by inserting amino acid sequences resulted in longer needles, while deletions that shortened the protein shortened the needles, consistent with the expected behavior of a molecular ruler [13, 14]. Ruler proteins have since been found to be involved in the length control of other structures, including the bacterial flagellum [1, 15–18].

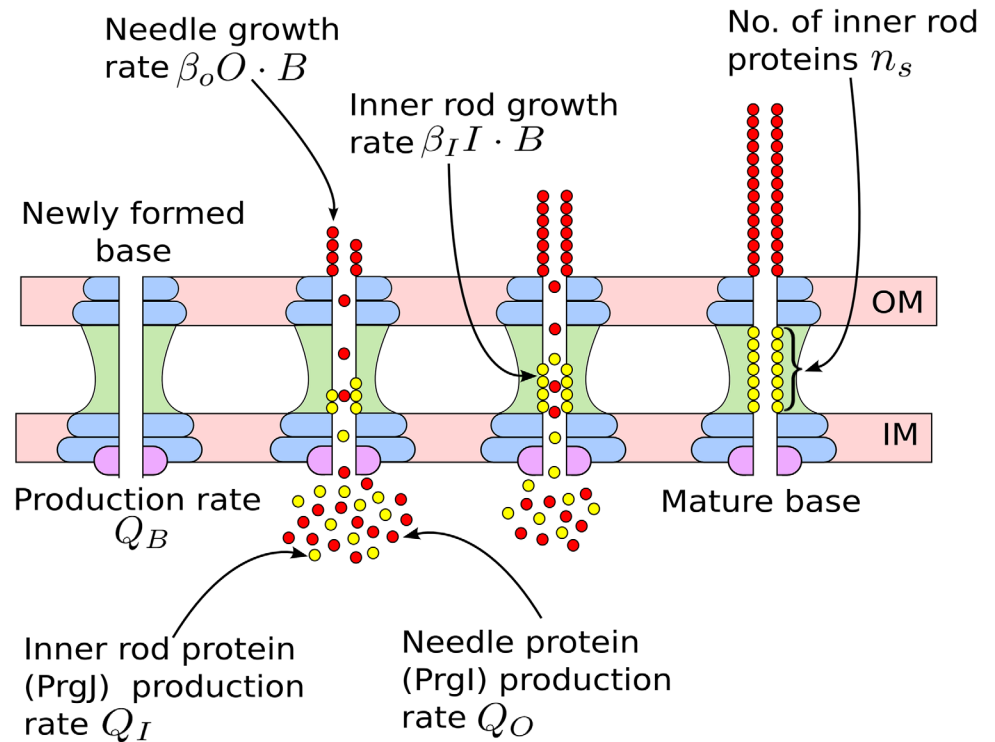
In *Salmonella* and *Shigella*, however, there is an “inner rod” complex that is distinct from the outer needle complex; this inner rod spans the inner and outer membranes inside the base complex and is composed of a different protein from the needle itself [1, 2]. The existence of this inner rod led to the proposal of an alternative model for needle length control. In this model, during the first phase of needle assembly, the inner rod and needle are assembled at the same time. Completion of the inner rod leads to “substrate switching:” the base stops exporting outer needle proteins and begins to secrete tip proteins and create a mature injectisome. In *Salmonella* the inner rod is composed of a protein called PrgJ, and Marlovits et al. found that over-expressing PrgJ resulted in *shorter* needles, while deleting it resulted in needles that were incredibly long [2], indicating that substrate switching might indeed control the length of needles in this case. Recent work from Hughes and co-workers, however, suggests that the protein InvJ (a homolog of YscP) may serve as a molecular ruler in *Salmonella*, and it is currently unclear which mechanism actually controls the length of *Salmonella* needles [14].

While both the ruler and substrate switching models are qualitatively consistent with various experimental findings, to date there has been no attempt to quantitatively predict how changing crucial parameters in the system (such as the concentration of PrgI, PrgJ or InvJ) would influence the distribution of needle lengths. In this work, we developed a straightforward mathematical model for the substrate switching mechanism in *Salmonella*, which we validated using stochastic simulation. In the case where the majority of needles present on a cell at steady-state are functional (*i.e.* mature), analysis of the model revealed that there is a fixed relationship between the average and variance in needle length. Comparison of our results with the experimental data in Marlovits et al. indicates that a total of around six PrgJ proteins must bind to the inner rod in order to induce substrate switching [2]. The distributions predicted by the model provide excellent agreement with experimental data with only a single free parameter. These findings provide quantitative support for the substrate switching model in *Salmonella*, and also suggest a set of straightforward experiments that would provide stringent tests of the mechanism in future work. Interestingly, the average lengths of needles in *Yersinia* are considerably longer than those observed in *Salmonella*, and the substrate switching model predicts that *Yersinia* needles should have a *much larger* variance in length than has been empirically observed. These results suggest that *Yersinia* and related species may have evolved a different mechanism (namely a ruler protein like YscP) to allow for more precise control over needle length when the needles themselves must be longer. Our work thus provides a mathematical framework for understanding the evolutionary pressures that have shaped length control mechanisms for this critically important virulence factor in bacteria.

## Results

### Ordinary differential equations for the system

To construct a mathematical model of the substrate switching mechanism, we examined the dynamics of the three important constituents of the system, namely the bases, inner rod proteins and needle proteins [1, 2]. We represent the average number of immature bases, inner rod proteins and “outer” needle proteins as  $B$ ,  $I$  and  $O$ , respectively. Each of these molecules is synthesized by the cell, and each can be lost from the system due to dilution from cell division or active degradation processes. When a needle (or inner rod) protein binds any given base, this will increase the length of the needle (or inner rod) associated with that base. A schematic of the model is shown in Fig 1. In order to relate the rates of these underlying processes (*i.e.* the  $Q$ 's and  $\beta$ 's in Fig 1) to the distribution of needle lengths one would obtain, we first generated a deterministic system of Ordinary Differential Equations (ODE's) to calculate the expected



**Fig 1. Various stages in the assembly of a needle complex.** This schematic shows various stages in a needle complex from base synthesis until maturation. The base spans the inner and the outer membranes ('IM' and 'OM' respectively) of the bacterial cells. The number of *immature* bases in the cell is represented by the variable  $B$ , the number of needle proteins by  $O$  and the inner rod proteins by  $I$ . Molecules are synthesized or produced at the indicated constant rates  $Q$  ( $Q_B$ ,  $Q_O$ , etc.). Binding of needle proteins to bases occurs with rate  $\beta_O O \cdot B$  and binding of the inner rod proteins to bases occurs at a rate  $\beta_I I \cdot B$ . A binding event increases the length of the needle (or the inner rod) and reduces the concentration of the relevant protein within the cell. The base reaches maturation when the number inner rod proteins attached reaches a particular value  $n_s$ , and a mature base can no longer interact with inner rod or needle proteins. The figure does not show the loss of proteins from the cell due to dilution cell division or active degradation. The rates of these processes form the basis of the system of ODEs (eqs (1)–(3)) that allow us to calculate the average values of  $B$ ,  $O$  and  $I$  at steady state, as well as our statistical model for the corresponding needle length distribution (eqs (7)–(14)).

doi:10.1371/journal.pcbi.1004851.g001

average steady-state values of  $B$ ,  $I$  and  $O$ . These averages were then used as the basis for a statistical model of the needle length distribution, as described below.

We constructed our ODE model of the dynamics based on the following considerations:

1. **Production:** The individual components in the system are synthesized in the cell at rates specified by parameters  $Q_B$ ,  $Q_I$  and  $Q_O$ . This results in an overall increase in the number of proteins *i.e.*  $dI/dt \sim Q_I$  (we would have a similar relation for the needle proteins and the bases as well). We assume that the  $Q$ 's are independent of time. Note that, while the molecular assembly of the base is a complex process in and of itself, in this model we represented the net effect as a constant production of bases.
2. **Degradation:** Dilution from cell division (and any additional active degradation processes) results in effective loss of proteins. Following common convention [19, 20], degradation/dilution is assumed to be first-order and depends on the rates  $\lambda_I$ ,  $\lambda_O$  and  $\lambda_B$ . This results in an overall decrease in the number of proteins *i.e.*  $dI/dt \sim -\lambda_I I$ . Note that the loss of a base would result in the loss not only of the base complex, but also any inner rod or outer needle proteins associated with it.

3. **Binding:** As mentioned in ref. [2], both the inner rod proteins and the outer needle proteins can bind to an immature base because substrate switching has not occurred yet. This binding is assumed to be a second-order reaction between the proteins and the base; the rate of this reaction is proportional to the product of the number of inner rod proteins (or outer needle proteins) and the number of bases and results in an overall loss in number of proteins:  $dI/dt \sim -\beta_I \cdot I \cdot B$ . The binding of an inner rod protein increases the length of the inner rod and the binding of a needle protein increases the length of the needle. Note that we do not explicitly consider the transport process for either inner rod or needle proteins. Also, we do not consider “unproductive” transport events, where, say, a needle protein might dissociate from the tip of the needle before the next needle protein binds. Since including this type of event would simply require a rescaling of the parameters in the model, we neglect these events without loss of generality. We also do not consider the possibility of a dissociated protein re-binding the needle, since the extracellular volume is likely so large that re-binding is highly unlikely (see the Supporting Information for further details).
4. **Substrate switching:** Substrate switching occurs when the number of inner rod proteins inside a base reaches a particular value  $n_s$ . Once substrate switching occurs, the base is no longer immature, which means that it cannot attach any more inner rod or outer needle proteins. For effective pathogenesis, it is crucial that the base is mature, since this is the only form that can secrete the tip and effector proteins. Note that rate of association of inner rod proteins to an immature base is given by  $\beta_I \cdot I \cdot B$  and for every  $n_s$  occurrences of association of inner rod proteins there is one substrate switching event, on average. In our simplified ODE model we thus approximate the maturation rate of bases by  $dB/dt \sim -\beta_I \cdot I \cdot B/n_s$

Based on the above considerations, the ordinary differential equations (ODE's) for the three species is given by:

$$\frac{dO}{dt} = Q_O - \lambda_O O - \beta_O O \cdot B \tag{1}$$

$$\frac{dI}{dt} = Q_I - \lambda_I I - \beta_I I \cdot B \tag{2}$$

$$\frac{dB}{dt} = Q_B - \lambda_B B - \frac{1}{n_s} \beta_I I \cdot B. \tag{3}$$

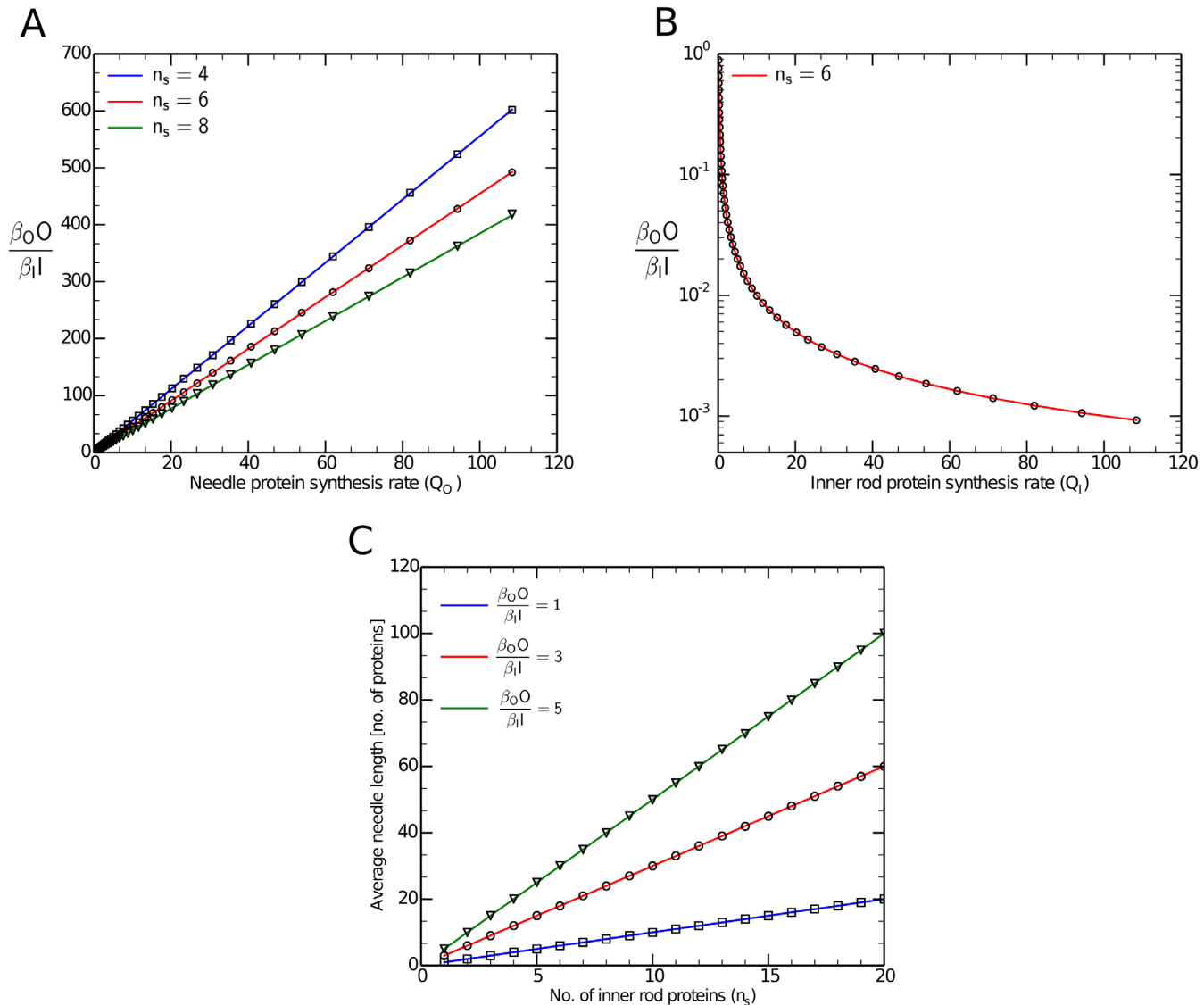
The steady-state values of the number of the inner-rod proteins, outer needle proteins and the immature bases can be solved from eqs (1), (2) and (3) as:

$$\bar{I} = \frac{1}{2\beta_I \lambda_I} \left( \sqrt{C^2 + 4n_s \beta_I \lambda_I \lambda_B Q_I} - C \right) \tag{4}$$

$$\bar{O} = \frac{Q_O (n_s \lambda_B + \beta_I \bar{I})}{n_s \beta_O Q_B + \lambda_O (n_s \lambda_B + \beta_I \bar{I})} \tag{5}$$

$$\bar{B} = \frac{n_s Q_B}{n_s \lambda_B + \beta_I \bar{I}}. \tag{6}$$

where  $C = \beta_I (n_s Q_B - Q_I) + n_s \lambda_I \lambda_B$ . While these equations allow us to calculate  $\bar{I}$ ,  $\bar{O}$  and  $\bar{B}$ , we should note that interpreting the above equations in terms of the average number of these proteins in a cell requires that the fluctuations in protein numbers be uncorrelated; that is,



**Fig 2. Comparison between theory and simulation.** The ratio of the rate parameters for outer needle to that for inner rod proteins controls the average length for a given  $n_s$ . Figure 2A shows that the ratio  $\beta_{O}O/\beta_{I}I$  increases linearly with the synthesis rate for outer needle proteins  $Q_O$ , whereas in Figure 2B the ratio has an inverse relation with the synthesis rate for inner rod proteins  $Q_I$ . The number of inner rod proteins required for substrate switching in Figures 2A and 2B was set to  $n_s = 6$ . Figure 2C shows the dependence of the average needle length on  $n_s$ . In each of the figures, the lines are predictions from the statistical model and the points represent values obtained from stochastic simulation, keeping all other parameters fixed.

doi:10.1371/journal.pcbi.1004851.g002

$\langle O \cdot B \rangle = \langle O \rangle \cdot \langle B \rangle$  (and similarly for the interaction of  $I$  and  $B$ ). As described below, comparison of the predictions from eqs (4)–(6) with the results of stochastic simulations indicate that this assumption holds good when the majority of bases are mature at steady state (see Fig 2). This assumption might well break down in parameter regimes not considered here, however, which would necessitate the consideration of correlated fluctuations in the model. We leave consideration of those regimes to future work.

### Statistical model

To translate the average, “mean field” values for the number of inner rod, outer needle proteins and immature bases from eqs (4)–(6) into the corresponding needle length distribution, we

constructed a statistical model of the growth of the inner rod and needle. In this model, it is assumed that the system is already at the steady-state and each of the processes involved (production, binding and degradation) may be treated as uncorrelated sequential random events. For notational convenience we shall omit the bars (e.g.  $\bar{B}$ ) and write  $I$ ,  $O$  and  $B$  to indicate these steady-state averages below.

Imagine a base is synthesized at time  $t = 0$ . The probability that this particular base finishes its assembly such that the needle will achieve a length  $L$  (i.e. number of outer needle proteins bound) once it stops growing is:

$$P(L) = \int_0^\infty P(L|t)P_{\text{stop}}(t)dt, \tag{7}$$

where  $P(L|t)$  is a conditional probability that the needle contains “ $L$ ” proteins at time  $t$ , and  $P_{\text{stop}}(t)$  is the probability density that the needle stops growing at time  $t$ . Note that the rate at which the needle proteins bind to any given immature base is  $\beta_O O$ , and for a needle of length  $L$ , the number of consecutive protein bindings required to achieve this length is also  $L$ . Keeping this in mind and realizing that the process of an outer needle protein binding to an immature base is a Poisson process, we have:

$$P(L|t) = \frac{(\beta_O O t)^L e^{-\beta_O O t}}{L!}. \tag{8}$$

Note that we take  $I$  and  $O$  in this case represent the “mean field” values from eqs (4) and (5), respectively.

To calculate  $P_{\text{stop}}(t)$ , we must consider the two independent mechanisms by which growth of a particular outer needle might cease. Any given base might be degraded *before* its inner rod completes; alternatively, the inner rod might fully assemble (i.e. reach  $n_s$  proteins) before the base is degraded. The probability density of stopping at time  $t$  is simply the sum of these probabilities:

$$P_{\text{stop}}(t) = P_{\text{base}}(\text{undeg}|t)P_{\text{inner}}(\text{complete}|t) + P_{\text{base}}(\text{deg}|t)P_{\text{inner}}(\text{incomplete}|t), \tag{9}$$

where  $P_{\text{base}}(\text{undeg}|t)$  and  $P_{\text{inner}}(\text{incomplete}|t)$  denote the cumulative probabilities of the base remaining undegraded and the inner rod remaining incomplete until time  $t$ , respectively, and  $P_{\text{base}}(\text{deg}|t)$  and  $P_{\text{inner}}(\text{complete}|t)$  denote the instantaneous probability densities that the base degraded and the inner rod is completed at time  $t$ , respectively. Base degradation occurs at rate  $\lambda_B$ , so the probability of the base remaining undegraded until time  $t$  is  $P_{\text{base}}(\text{undeg}|t) = e^{-\lambda_B t}$ . Similarly, the probability density that the base is degraded precisely at time  $t$  is  $P_{\text{base}}(\text{deg}|t) = \lambda_B e^{-\lambda_B t}$ . Once again, the process of each inner rod protein binding to an immature base is a Poisson process with rate  $\beta_I I$ . The density of the *waiting times* until the  $n_s^{\text{th}}$  Poisson event is thus given by an Erlang distribution:

$$P_{\text{inner}}(\text{complete}|t) = \frac{e^{-\beta_I I t} (\beta_I I)^{n_s} t^{n_s-1}}{(n_s - 1)!}, \tag{10}$$

and the probability that the inner rod is incomplete is the cumulative probability that it remains incomplete until time  $t$ :

$$P_{\text{inner}}(\text{incomplete}|t) = e^{-\beta_I I t} \sum_{k=0}^{n_s-1} \frac{(\beta_I I t)^k}{k!}. \tag{11}$$

Using eqs (8)–(11) in Eq (7) and integrating over  $t$  we obtained an analytical expression for the probability distribution of needle lengths as:

$$P(L) = \frac{y_0^L}{L!} \left( \frac{(n_s + L - 1)!}{(n_s - 1)!} y_I^{n_s} + \epsilon \sum_{k=0}^{n_s-1} \frac{(k + L)!}{k!} y_I^{k+1} \right), \quad (12)$$

with  $y_0 = \frac{\beta_O O}{\beta_O O + \beta_I I + \lambda_B}$ ,  $y_I = \frac{\beta_I I}{\beta_O O + \beta_I I + \lambda_B}$  and  $\epsilon = \frac{\lambda_B}{\beta_I I}$ .

Interestingly, the last of these dimensionless parameters,  $\epsilon$ , plays a large part in determining the number of bases that are mature at steady state, since it compares the rate of base degradation/dilution ( $\lambda_B$ ) to the frequency with which any given immature base will bind inner rod proteins ( $\beta_I I$ ). When  $\lambda_B \gg \beta_I I$  (equivalently  $\epsilon \gg 1$ ), degradation dominates over inner rod protein binding, and most bases are degraded before they have enough time to bind  $n_s$  inner rod proteins. As a result, in this regime, the vast majority of bases are *immature* at steady state. Conversely, when  $\lambda_B \ll \beta_I I$  ( $\epsilon \ll 1$ ), then inner rod protein binding is much faster than degradation, and the majority of bases are *mature* at steady state. Injectisomes are only functional when they are mature, and since each base complex represents a massive investment of energy in protein synthesis by the cell, we expect that the system has evolved towards a parameter regime where the majority of the bases are mature (and thus functional) at steady state. As such, we take  $\epsilon \ll 1$  to be the relevant parameter regime for WT pathogenic bacteria.

Ignoring higher ordered terms in  $\epsilon$  when  $\epsilon \ll 1$ , the average needle length follows:

$$\langle L \rangle \simeq n_s \left( \frac{\beta_O O}{\beta_I I} \right). \quad (13)$$

Similarly, ignoring higher ordered terms in  $\epsilon$  in the second moment of  $L$  we obtained the variance in the needle lengths as:

$$\sigma^2 \simeq \langle L \rangle + \frac{\langle L \rangle^2}{n_s}. \quad (14)$$

According to eq (14), the variance in the needle lengths should depend quadratically on the average needle length. Furthermore, given a distribution of needle lengths we can predict the number of inner rod proteins required for substrate switching for any given bacterial species. Using eqs (13) and (14) in eq (12) and ignoring the term in  $\epsilon$  (since  $\epsilon \ll 1$ ), the probability distribution of needle lengths becomes the negative binomial distribution:

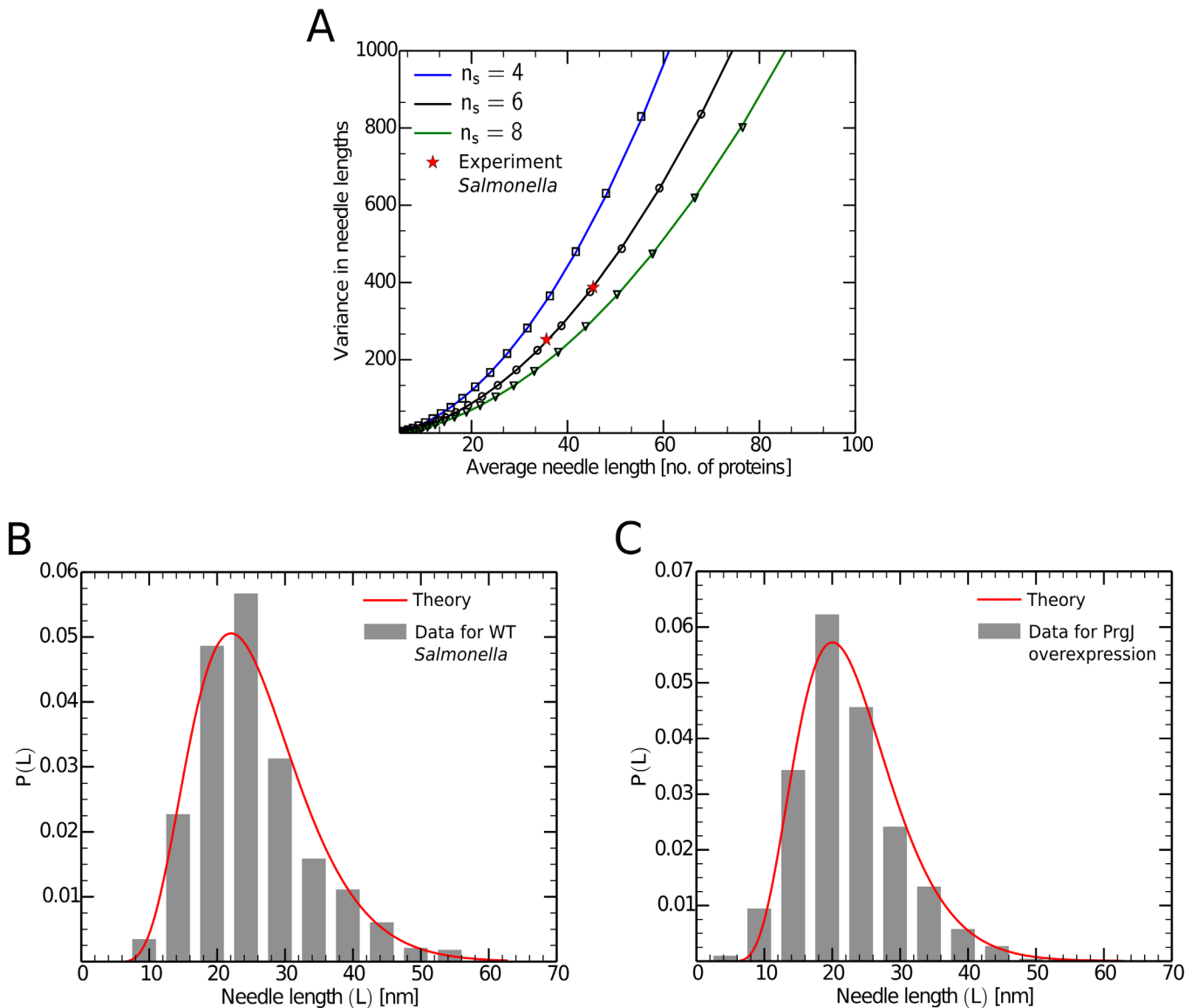
$$P(L) \simeq \frac{(n_s + L - 1)!}{L!(n_s - 1)!} \left( \frac{\langle L \rangle}{\sigma^2} \right)^{n_s} \left( 1 - \frac{\langle L \rangle}{\sigma^2} \right)^L \quad (15)$$

The predictions of eqs (13)–(15) are compared against both numerical simulations and experimental predictions in Figs 2 and 3.

### Comparison with stochastic simulation

The results discussed above are a consequence of two very important assumptions about the model: (a) the term  $\beta_I I \cdot B/n_s$  in eq (3), which approximates the fact that, for every  $n_s$  inner rod protein bindings there is one substrate switching event that deletes a base from the pool of immature bases and adds it to the pool of mature bases; (b) the fluctuations in the steady-state values of  $I$ ,  $O$  and  $B$  are uncorrelated with each other and would have no effect on the distribution of needle lengths. In order to test the validity of these assumptions, we performed numerical simulations of the assembly process using the Gillespie algorithm [21]. In our simulations





**Fig 3. Needle length distributions and comparison with experiment.** Figure 3A shows variance in needle lengths versus the average needle lengths. The lines represent predictions from the mathematical model, the points represent results obtained from stochastic simulation for parameter values specified in supplement. The stars are experimental values from data for *Salmonella* in ref. [2]. Figure 3B and 3C are normalized probability distributions of needle lengths for *Salmonella*. Figure 3B is the distribution for “Wild Type” whereas Figure 3C is for a variant with over-expressed inner needle proteins (PrgJ). The histograms represent the experimental values from ref. [2] and the lines represent the values expected according to the statistical model (eq 15), using eq (16) to convert the lengths to nanometers.

doi:10.1371/journal.pcbi.1004851.g003

the inner rod proteins, outer needle proteins and bases are treated as “agents” that can interact with one another to form needle complexes [22–25]. Each base has an arbitrary number of inner rod and outer needle proteins associated with it, and we track both the pool of mature and immature bases (depending on whether or not substrate switching has occurred for a given base). Note that here each base becomes mature when *precisely*  $n_s$  inner rod proteins have associated with it. We also have a pool of free inner rod and outer needle proteins.

The exact value of the parameters in our model have not been determined experimentally, but they can be subjected to a set of reasonable constraints. For instance, the total number of

bases associated with a bacterial cell is  $\sim 100$ , therefore  $Q_B/\lambda_B \simeq 100$ . Since “degradation” in this case is assumed to occur largely due to dilution from cell division, the decay parameter  $\lambda$  should be of the same for all proteins. Assuming  $\epsilon \ll 1$  also constrains the relative values of  $\lambda_B$  and  $\beta_I$ . Keeping these constraints in mind, we chose a set of reasonable values (see the Supporting Information) of  $Q$ 's,  $\lambda$ 's and  $\beta$ 's, which, along with initial values for the number of bases, inner rod and outer needle proteins, determine the “activities” of synthesis, degradation and binding in our Gillespie simulations. Note that these parameters were chosen simply for comparison between our analytical results and simulations, and were not used for comparisons with experimental data (see below). For every synthesis event, a protein (of the corresponding species) is spontaneously added to the pool of free proteins, whereas in case of a degradation event the protein is deleted from this pool. For binding of an inner rod (or outer needle) protein, a base is chosen randomly and the number of inner rod (or outer needle) proteins associated with it is increased by one. When substrate switching occurs for a particular base, it is deleted from the pool of immature bases and added to the pool of mature bases. For a given set of parameter values, we ran these simulations until the system reached steady-state.

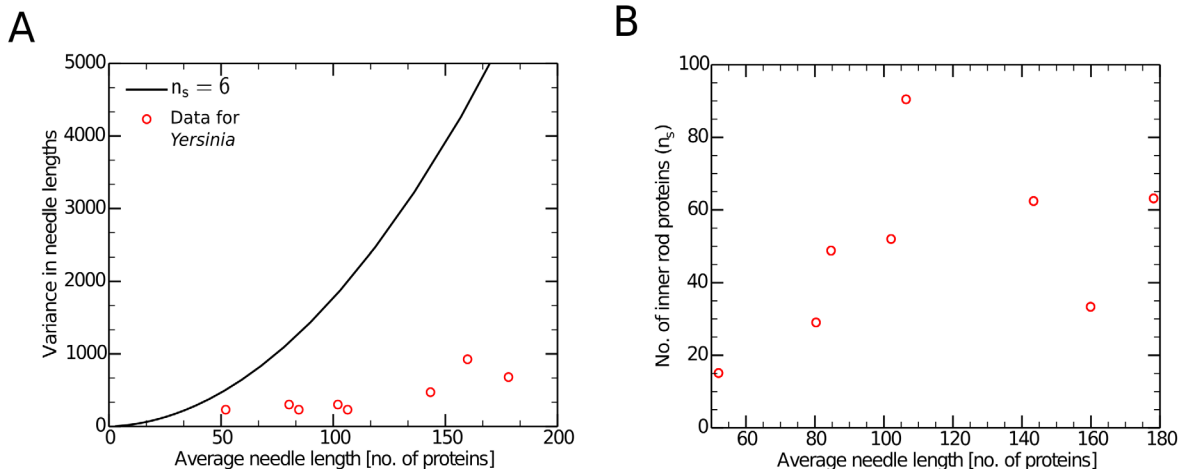
In order to compare our analytical results with the ones obtained from simulation, we focused on the quantity  $\beta_O/\beta_I I$ , which, according to eq (13), should control the average needle length. For the chosen set of parameters, we obtained the predicted steady-state values  $I$ ,  $O$  and  $B$  using eqs (4)–(6). We examined the effect of changing the synthesis rate for inner rod and outer needle proteins on  $\beta_O/\beta_I I$ , and, as shown in Fig 2A and 2B, the predictions of our model agree with the results obtained from simulation. We also used the predicted steady-state values in eq (13) to obtain the theoretical values of average length and made a similar comparison between theory and simulation examining the effect of changing the number of inner rod proteins required for substrate switching on the average needle lengths (see Fig 2C). As mentioned earlier, our statistical model predicts that the variance in needle lengths should depend quadratically on the average needle length. In order to test this prediction, we varied the synthesis rate of outer needle proteins and obtained average and variance in needle length for each case. This exercise was repeated for different  $n_s$  values and as shown in Fig 3A there is very good agreement between the theoretical model and simulation.

## Comparison with experimental data

To further validate our findings, we compared them with experimental results obtained from ref. [2]. In this work, the authors were able to obtain two different distribution of needle lengths in *S. typhimurium*, one for “Wild Type” and the other for a mutant variant where the inner rod proteins (PrgJ) were overexpressed. We calculated the average needle length and the variance in the needle lengths from these distributions. Note that in the experiment, the needle lengths were measured in nanometers (nm), whereas in our work these lengths are represented in number of proteins, and hence we need a transformation that relates the two. As per the structural model of the needle in ref. [7], the height of the first subunit attached to the base is 6.04 nm and the increase in the height of the needle for a every outer needle (PrgI) binding is 0.46 nm, which gives us:

$$L [\text{in nm}] = 0.46 \cdot L [\text{in number of proteins}] + 6.04. \quad (16)$$

We then compared the predictions for average and variances in needle length to those observed for the two experimental distributions (Fig 3A). Interestingly, both of these points lie on the same curve predicted by our model, indicating that the data for *Salmonella* is consistent with six inner rod proteins required for substrate switching (*i.e.*  $n_s = 6$ ). Substituting  $n_s = 6$  in eq (15), we obtained the analytical probability distribution of needle lengths and compared these



**Fig 4. Data from *Yersinia* is not consistent with substrate switching.** Figure 4 shows a comparison between the model and experimental data for *Yersinia* taken from ref. [13]. Figure 4A shows that the variance and the average lengths for *Yersinia* needles does not have the expected quadratic relationship for  $n_s = 6$ . Figure 4B shows that the number of inner rod proteins required for substrate switching in *Yersinia* is generally not consistent with eq (13), since a wide variety of values of this parameter (ranging between 10 and 100) would be required to explain the data.

doi:10.1371/journal.pcbi.1004851.g004

with the experimental length distributions from ref. [2] (See Fig 3B and 3C). Note that we obtained excellent agreement between theory and experiment with only one free parameter ( $n_s$ ).

As mentioned above, *Y. pestis* appears to have a molecular ruler (YscP) that regulates its needle lengths. The authors of ref. [13] added residues to this ruler protein and obtained needles of different lengths. We observed that the variances in these *Yersinia* needles are much smaller than the predictions for  $n_s = 6$  according to eq (14) (see Fig 4A). Indeed, not only do the variance and average length not satisfy the quadratic relationship with  $n_s = 6$ , there is no single  $n_s$  value with which the data is consistent. Applying eq (14) to these data, we observe a variety of  $n_s$  values scattered between  $\sim 10$  to  $\sim 100$  (Fig 4B). Unfortunately, we could not obtain data for the full  $P(L)$  distributions for these YscP length mutants, and further tests of the model against this data (e.g. comparisons to the distribution in eq (15)) will likely require the collection of new experimental data from *Yersinia*.

## Discussion

Competing models have been proposed to explain how the of the length of the T3SS needle is controlled. While these models are generally considered to be in qualitative agreement with various experimental data sets [1, 2, 13, 14], there has been no attempt to use quantitative approaches to obtain and test predictions from these models in a rigorous way. In this work, we considered the substrate switching mechanism that has been proposed for length control in *Salmonella*, and constructed a mathematical model based on a straightforward interpretation of the interactions thought to be important for this mechanism [1, 2]. The analytical results we obtained from this model are consistent not only with stochastic simulations, but also with available experimental data for *S. typhimurium* [2].

A major prediction of our model (for  $\epsilon \ll 1$ ) is that, for a given population of needles, the variance in the needle lengths should scale *quadratically* with the average needle lengths, parametrized by a consistent value of  $n_s$  (i.e. the number of inner rod proteins required for substrate switching, see eq 14). This finding implies that the substrate switching mechanism may not be optimum for bacteria such as *Yersinia pestis*, which need to grow longer needles. If the length of the needles is regulated by substrate switching, larger values of the average

length will result in very large variability, potentially impairing pathogenesis [13]. The quadratic relationship between average and variance predicted by our model also suggests a natural set of experiments to provide further evidence for the substrate switching mechanism in *Salmonella* and other bacterial species. One could readily vary the synthesis rate of either PrgI or PrgJ (by using, say, “titratable” promoters whose activity can be controlled by the concentration of an exogenous ligand), measure the average and variance of the resulting needle lengths, and compare those results with the quadratic function predicted by our model (e.g. Fig 3 with  $n_s = 6$ ). Performing such an experiment would provide a very stringent test of the substrate switching mechanism.

Our results also suggest that further structural studies could be helpful in determining the relative feasibility of the substrate switching model. EM structures of the base indicate that it is about 265 Å from top to bottom; the PrgJ portion of the inner rod takes up about 40%, or ~105 Å, of this distance [3]. If we assume that PrgJ forms a helical structure identical to that of the outer needle, then according to eq (16) it would take ~9 proteins to span that distance, which is not much larger than the  $n_s = 6$  prediction made by our model. In order to fill that distance with six monomers, the PrgJ portion of the inner rod would have to adopt a somewhat more extended conformation than the needle complex, with ~0.68 nm/monomer compared to 0.46 nm/monomer. Unfortunately, we currently do not have a higher-resolution structure of PrgJ nor is there a model of the inner rod structure akin to that of the needle [7]. Further structural studies could thus shed light onto whether the  $n_s = 6$  prediction of the model is actually consistent with the number of PrgJ monomers found in the inner rod. Also, it is currently unclear how the information that the inner rod is complete is relayed to the other components of the base complex that ultimately decide which proteins to export and thus drive injectisome maturation. Atomic-resolution structures of more components of this complex will ultimately be necessary in order to understand the allosteric basis of the substrate switching mechanism.

As discussed above, in the alternative ruler protein mechanism, the eponymous ruler (e.g. YscP in *Yersinia*) is thought to anchor its C-terminus to the beginning of the needle in the secretion channel and measure the length of the needle; since the secretion channel is very narrow, the ruler will almost certainly block the secretion of other proteins during the time it is bound [5, 7]. This measurement is likely repeated several times during the course of needle growth, until the needle is long enough for the C-terminus of the ruler protein to be in proximity of the needle’s end, which induces needle maturation and the end of needle polymerization [1, 13, 14]. Original support for this model came from studies in *Yersinia*, where it was shown that changes in the sequence length of YscP was strongly correlated with the average length of the resulting needles obtained from mutant cells [13]. Interestingly, Wee and Hughes recently provided evidence that changes in the length of the *Salmonella* homolog of YscP, InvJ, also causes concomitant increases in average length [14]. They argued that previous needle-length changes from PrgJ overexpression and deletion mutants in *Salmonella* were the result of changes in the frequency with which the length measurement is made by the ruler protein [2, 14]. Unfortunately, we were unable to extract reliable length distribution data from ref. [14], but inspection of their results indicates it is unlikely that their findings are consistent with a single value of  $n_s$ , as with the data for YscP mutants in *Yersinia* (Fig 4B) [13, 14].

Developing a full understanding of how parameters like the frequency of length measurement affects the length distribution in the ruler model will require a complete quantitative model similar to substrate switching model described above, which is beyond the scope of the current work. That being said, it is clear that changes in the expression of YscP or other ruler proteins will have distinct effects compared to changing the concentration of the inner rod protein. If the ruler protein is at very high concentration, then the frequency of length checks should also be high, and we would expect that most needles would have lengths quite close to

the minimum set by the length of the ruler protein itself, since bases with significantly shorter needles will not mature in that model [1]. As the concentration of the ruler decreases, however, we would expect some needles to grow too long due to the (relative) decrease in measurement frequency, leading to an asymmetric distribution of needle lengths. In contrast, overexpression of PrgJ or another inner rod protein leads to progressively shorter needles in the substrate switching model, since  $\langle L \rangle$  decreases as  $1/I$  (eq (13)). So, while the detailed derivation of  $P(L)$  for the ruler model awaits future study, it is clear that the two models would produce very different distributions based on changes in a similar parameter (namely the concentration of the relevant “length control” protein). In any case, our work emphasizes the role that mathematical models can play in explaining assembly processes and suggesting further experiments that can distinguish between competing mechanisms [19, 26–29].

## Methods

### Numerical simulations

All simulations were performed using a purpose-built simulator that implements the standard Doob-Gillespie approach for the exact stochastic simulation of systems of chemical reactions. Further details on the numerical calculations may be found in the Supporting Information.

### Supporting Information

**S1 Text. Supplement.** This file contains additional information about the mathematical details of the model and details of the stochastic simulation.  
(PDF)

### Author Contributions

Conceived and designed the experiments: MKN JI JJS EJD. Performed the experiments: MKN JI. Analyzed the data: MKN JI JJS EJD. Wrote the paper: MKN JI JJS EJD.

### References

1. Cornelis G R (2006) The type III secretion injectisome *Nat Rev Microbiol* 4(11):811–825. doi: [10.1038/nrmicro1526](https://doi.org/10.1038/nrmicro1526) PMID: [17041629](https://pubmed.ncbi.nlm.nih.gov/17041629/)
2. Marlovits T C, Kubori T, Lara-Tejero M, Thomas D, Unger V M, Galán J E (2006) Assembly of the inner rod determines needle length in type III secretion injectisome *Nature* 441:637–640. doi: [10.1038/nature04822](https://doi.org/10.1038/nature04822) PMID: [16738660](https://pubmed.ncbi.nlm.nih.gov/16738660/)
3. Marlovits T C, Kubori T, Sukhan A, Thomas D R, Galán J E, Unger V M (2004) Structural insights into the assembly of the type III secretion needle complex *Science* 306(5698):1040–1042. doi: [10.1126/science.1102610](https://doi.org/10.1126/science.1102610) PMID: [15528446](https://pubmed.ncbi.nlm.nih.gov/15528446/)
4. Kubori T, Matsushima Y, Nakamura D, Uralil J, Lara-Tejero M, Sukhan A, Galán J E, Aizawa S (1998) Supramolecular Structure of the *Salmonella Typhimurium* Type III Protein Secretion System *Science* 280(5363):602–605. doi: [10.1126/science.280.5363.602](https://doi.org/10.1126/science.280.5363.602) PMID: [9554854](https://pubmed.ncbi.nlm.nih.gov/9554854/)
5. Rathinavelan T, Zhang L, Picking W L, Weis D D, De Guzman R N, Im W (2010) A repulsive electrostatic mechanism for protein export through the type III secretion apparatus *Biophys J* 98:452–461. doi: [10.1016/j.bpj.2009.10.030](https://doi.org/10.1016/j.bpj.2009.10.030) PMID: [20141759](https://pubmed.ncbi.nlm.nih.gov/20141759/)
6. Blocker A J, Deane J E, Veenendaal A K J, Roversi P, Hodgkinson J L, Johnson S, Lea S M (2008) What’s the point of the type III secretion system needle? *Proc Natl Acad Sci USA* 105(18):6507–6513. doi: [10.1073/pnas.0708344105](https://doi.org/10.1073/pnas.0708344105) PMID: [18458349](https://pubmed.ncbi.nlm.nih.gov/18458349/)
7. Loquet A, Sgourakis Nikolaos G, Gupta R, Giller K, Riedel D, Goosmann C, Griesinger C, Kolbe M, Baker D, Becker S, Lange A. (2012) Atomic model of the Type III Secretin System Needle *Nature* 486(7402):276–279. doi: [10.1038/nature11079](https://doi.org/10.1038/nature11079) PMID: [22699623](https://pubmed.ncbi.nlm.nih.gov/22699623/)
8. Rathinavelan T, Lara-Tejero M, Lefebvre M, Chatterjee S, McShan A C, Guo D, Tang C, Galán J E, De Guzman R N (2014) NMR model of PrgI-SipD interaction and its implications in the needle-tip assembly

- of the *Salmonella* type III secretion system *J Mol Biol* 426(16):2958–2969. doi: [10.1016/j.jmb.2014.06.009](https://doi.org/10.1016/j.jmb.2014.06.009) PMID: [24951833](https://pubmed.ncbi.nlm.nih.gov/24951833/)
9. Heine S J, Franco-Mahecha O L, Chan X, Choudhari S, Blackwelder W C, van Roosmalen M L, et al. (2015) Shigella IpaB and IpaD displayed on *L. lactis* bacterium-like particles induce protective immunity in adult and infant mice *Immunol Cell Biol* 93(7):641–652 doi: [10.1038/icb.2015.24](https://doi.org/10.1038/icb.2015.24) PMID: [25776843](https://pubmed.ncbi.nlm.nih.gov/25776843/)
  10. McShan A C, De Guzman R N (2015) The bacterial type III secretion system as a target for developing new antibiotics *Chemical Biology & Drug Design* 85(1):30–42. doi: [10.1111/cbdd.12422](https://doi.org/10.1111/cbdd.12422)
  11. Bergeron J R C, Worrall L J, Sgourakis N G, DiMaio F, Pfuetzner R A, Felise H B, Vuckovic M, Yu A C, Miller S I, Baker D, Strynadka N C J (2013) A Refined Model of the Prototypical *Salmonella* SPI-1 T3SS Basal Body Reveals the Molecular Basis for Its Assembly *PLoS Pathog* 9(4):e1003307. doi: [10.1371/journal.ppat.1003307](https://doi.org/10.1371/journal.ppat.1003307) PMID: [23633951](https://pubmed.ncbi.nlm.nih.gov/23633951/)
  12. Kubori T, Sukhan A, Aizawa S, Galán J E (2000) Molecular Characterization and Assembly of the Needle Complex of the *Salmonella Typhimurium* Type III Protein Secretion System *Proc Natl Acad Sci USA* 97(18):10225–10230. doi: [10.1073/pnas.170128997](https://doi.org/10.1073/pnas.170128997) PMID: [10944190](https://pubmed.ncbi.nlm.nih.gov/10944190/)
  13. Journet L, Agrain C, Broz P, Cornelis G R (2003) The needle length of bacterial injectisome is determined by a molecular ruler *Science* 302(5651):1757–1760. doi: [10.1126/science.1091422](https://doi.org/10.1126/science.1091422) PMID: [14657497](https://pubmed.ncbi.nlm.nih.gov/14657497/)
  14. Wee D H, Hughes K T (2015) Molecular ruler determines needle length for *Salmonella* Spi-1 injectisome *Proc Natl Acad Sci USA* 112(13):4098–4103. doi: [10.1073/pnas.1423492112](https://doi.org/10.1073/pnas.1423492112) PMID: [25775540](https://pubmed.ncbi.nlm.nih.gov/25775540/)
  15. Katsura I, Hendrix R (1984) Length Determination in Bacteriophage Lambda Tails *Cell* 39(3 Pt 2):691–698. doi: [10.1016/0092-8674\(84\)90476-8](https://doi.org/10.1016/0092-8674(84)90476-8) PMID: [6096021](https://pubmed.ncbi.nlm.nih.gov/6096021/)
  16. Marshall W F, Rosenbaum J L (2001) Intraflagellar Transport Balances continuous Turnover of the Outer doublet Microtubules: Implications for Flagellar Length Control *J Cell Biol* 135(3):405–414. doi: [10.1083/jcb.200106141](https://doi.org/10.1083/jcb.200106141)
  17. Shibata S, Takahashi N, Chevance F F V, Karlinsey J E, Hughes K T, Aizawa S (2007) FliK Regulates Flagellat Hook Length as an Internal Ruler *Mol Microbiol* 64(5):1404–1415. doi: [10.1111/j.1365-2958.2007.05750.x](https://doi.org/10.1111/j.1365-2958.2007.05750.x) PMID: [17542929](https://pubmed.ncbi.nlm.nih.gov/17542929/)
  18. Oda T, Yanagisawa H, Kamiya R, Kikkawa M (2015) A Molecular Ruler Determines the repeat length in Eukaryotic Cilia and Flagella *Science* 346(6211):857–860. doi: [10.1126/science.1260214](https://doi.org/10.1126/science.1260214)
  19. Deeds E J, Bachman J A, Fonatana W (2012) Optimizing Ring Assembly Reveals the Strength of Weak Interactions *Proc Natl Acad Sci USA* 109(7):2348–2353. doi: [10.1073/pnas.1113095109](https://doi.org/10.1073/pnas.1113095109) PMID: [22308356](https://pubmed.ncbi.nlm.nih.gov/22308356/)
  20. Mangan S, Alona U (2003) Structure and function of the feed-forward loop network motif *Proc Natl Acad Sci USA* 100(21):11980–11985. doi: [10.1073/pnas.2133841100](https://doi.org/10.1073/pnas.2133841100) PMID: [14530388](https://pubmed.ncbi.nlm.nih.gov/14530388/)
  21. Gillespie D T (1977) Exact stochastic simulation of coupled chemical reactions *J Phys Chem* 81:2340–2361. doi: [10.1021/j100540a008](https://doi.org/10.1021/j100540a008)
  22. Danos V, Laneve C, (2003) Formal Molecular Biology *Theor Comput Sci* 35(1):69–110.
  23. Danos V, Feret J, Harmer R, Krivine J (2007) Rule-based modelling of cellular signalling *Lect Notes Comput Sc* 4703:17–41. doi: [10.1007/978-3-540-74407-8\\_3](https://doi.org/10.1007/978-3-540-74407-8_3)
  24. Deeds E J, Krivine J, Feret J, Danos V, Fontana W (2012) Combinatorial complexity and composition drift in protein interactions networks *PLoS One* 7:e32032. doi: [10.1371/journal.pone.0032032](https://doi.org/10.1371/journal.pone.0032032) PMID: [22412851](https://pubmed.ncbi.nlm.nih.gov/22412851/)
  25. Suderman R, Deeds E J (2013) Machine vs. ensembles: Effective MAPK signaling though heterogeneous sets of proteins complexes *PLoS Comp Bio* 9:e1003278. doi: [10.1371/journal.pcbi.1003278](https://doi.org/10.1371/journal.pcbi.1003278)
  26. Bray D, Lay S (1997) Computer-based Analysis of the Binding Steps in Protein Complex Formation *Proc Natl Acad Sci USA* 94(25):13493–13498. doi: [10.1073/pnas.94.25.13493](https://doi.org/10.1073/pnas.94.25.13493) PMID: [9391053](https://pubmed.ncbi.nlm.nih.gov/9391053/)
  27. Siaz L, Vilar J M G (2006) Stochastic Dynamics of Macromolecular-assembly Networks *Mol Sys Biol* 2 2006.0024.
  28. Xie L, Smith G R, Feng X, Schwartz R (2012) Surveying Capsid Assembly Pathways through Simulation-Based Data Fitting *Biophys J* 103(7):1545–1554. doi: [10.1016/j.bpj.2012.08.057](https://doi.org/10.1016/j.bpj.2012.08.057) PMID: [23062347](https://pubmed.ncbi.nlm.nih.gov/23062347/)
  29. Wani P S, Rowland M A, Ondracek A, Deeds E J, Roelofs J (2014) Maturation of the Proteasome Core Particle Induces an Affinity Switch that Controls Regulatory Particle Association *Nat Commun* 6: 6384–6395. doi: [10.1038/ncomms7384](https://doi.org/10.1038/ncomms7384)