Biophysical *Journal*
Article

Biophysical Society

Robustness and the evolution of length control strategies in the T3SS and flagellar hook

Maulik K. Nariya,[1] Abhishek Mallela,[2] Jack J. Shi,[1] and Eric J. Deeds[3,4,*]
[1]Department of Physics and Astronomy, University of Kansas, Lawrence, Kansas; [2]Department of Biostatistics, University of Kansas Medical Center, Kansas City, Kansas; [3]Center for Computational Biology, University of Kansas, Lawrence, Kansas and [4]Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas

ABSTRACT Bacterial cells construct many structures, such as the flagellar hook and the type III secretion system (T3SS) injectisome, that aid in crucial physiological processes such as locomotion and pathogenesis. Both of these structures involve long extracellular channels, and the length of these channels must be highly regulated in order for these structures to perform their intended functions. There are two leading models for how length control is achieved in the flagellar hook and T3SS needle: the substrate switching model, in which the length is controlled by assembly of an inner rod, and the ruler model, in which a molecular ruler controls the length. Although there is qualitative experimental evidence to support both models, comparatively little has been done to quantitatively characterize these mechanisms or make detailed predictions that could be used to unambiguously test these mechanisms experimentally. In this work, we constructed a mathematical model of length control based on the ruler mechanism and found that the predictions of this model are consistent with experimental data—not just for the scaling of the average length with the ruler protein length, but also for the variance. Interestingly, we found that the ruler mechanism allows for the evolution of needles with large average lengths without the concomitant large increase in variance that occurs in the substrate switching mechanism. In addition to making further predictions that can be tested experimentally, these findings shed new light on the trade-offs that may have led to the evolution of different length control mechanisms in different bacterial species.

SIGNIFICANCE Bacteria produce many structures on their surfaces, such as the flagellar hook and the type III secretion system, that aid in crucial processes such as locomotion and pathogenesis. One critical aspect of these structures is their length: if they are too short or too long, they generally will not function optimally. Experimental work has suggested two competing mechanisms bacteria could use to regulate length: the ruler mechanism and the substrate switching mechanism. In this work, we make a mathematical model of the ruler mechanism and show that quantitative predictions from this model are consistent with available data. In addition to suggesting future experiments to resolve the debate surrounding these two mechanisms, our work provides insight into evolutionary trade-offs between them.

INTRODUCTION

Bacterial cells build a variety of structures on their exterior, e.g., flagella, pili, the type III secretion system (T3SS), etc., which help in carrying out important functions such as locomotion, DNA transfer, and pathogenesis (1). To ensure effective function and to optimize the efficiency of transport, bacteria need to control the length of these structures with high precision. This raises a natural question about the regulation of the assembly process: how does the bacterial cell "know" when to stop growing these structures, especially because the structures themselves are outside the cell? It is as if one was trying to construct a building that is precisely 30 stories high, from underground, with no way of directly looking at the structure as it is being built.

The solution to this fundamental problem has been perhaps best studied experimentally in two homologous model systems in gram-negative bacteria: the T3SS injectisome and the flagellar hook (1,2). The flagellum is an organelle that consists of a curved tubular structure known as the hook with one of its ends attached to a motor complex that terminates in the outer membrane of the cell and the other to

Submitted April 22, 2020, and accepted for publication May 20, 2021.

*Correspondence: deeds@ucla.edu

Maulik K. Nariya's present address is Department of Systems Biology, Harvard Medical School, Boston, Massachusetts.

Abhishek Mallela's present address is Department of Mathematics, University of California Davis, Davis, California.

Eric J. Deeds's present address is Department of Integrative Biology and Physiology, University of California Los Angeles, Los Angeles, California.

Editor: Ido Golding.

https://doi.org/10.1016/j.bpj.2021.05.032

a long flexible filament that "whips" as the motor turns the hook to provide locomotion. Even though it is the filament that facilitates the actual motion, the formation of the hook is a crucial step in the assembly process (2). The T3SS injectisome needle is homologous to the flagellar hook; it consists of a narrow needle that grows in the extracellular region and helps transport effector proteins into host cells during pathogenesis (3–6). Morphologically, both the T3SS injectisome and flagellum consist of a basal body that spans the inner and the outer membranes of the bacterial cell and has a narrow passage for the secretion of proteins. This basal body, which is often referred to as the base, along with a number of regulatory genes, controls the assembly of these structures (1). Several mechanisms have been proposed to explain length control for these systems, and among these, the two most popular ones are the substrate switching mechanism and the ruler protein mechanism.

In the substrate switching mechanism, there is an inner rod that spans the inner and outer membranes of the cell and is located inside the base. Initially the inner rod and the needle are assembled at the same time. Once the inner rod is completed, the base stops growing the needle and switches substrates, secreting tip proteins to form a mature injectosome. In *Salmonella*, PrgJ was posited to be the protein that forms the inner rod, and Marlovits et al. observed that overexpressing PrgJ resulted in shorter needles as compared with ones in wild type, whereas deleting PrgJ resulted in extremely long needles (7–9). In contrast, according to the ruler mechanism, there is a dedicated protein that acts a "molecular ruler." During the assembly of the needle, the base periodically secretes a ruler protein, and once the needle reaches an optimal length, the C-terminus of the ruler protein interacts with the base, signaling the base to stop growing the needle. Such ruler proteins have been identified not only in the T3SS (e.g., YscP in *Yersinia*) but also in the flagellar hook (e.g., FliK in *Salmonella*) (10,11). Furthermore, there is recent experimental evidence for the length control via the ruler mechanism in the *Salmonella* injectisome, leading to an ongoing controversy about the functions of some of the proteins involved in assembly (12,13). Experimental evidence for the ruler mechanism has primarily been provided by demonstrating that the average length of the needles/hooks increases linearly with increases in the length of the ruler protein itself.

Although there is thus qualitative agreement between available experimental data and these two alternative mechanisms, there has been comparatively little quantitative modeling of these two alternative mechanisms. One exception to this is the work of Keener and colleagues, who showed that a continuous, probabilistic model of the ruler mechanism was consistent with experimentally observed length distributions (14,15). Similarly, we recently developed a discrete stochastic model of the substrate switching mechanism and also found that it was consistent with data from the *Salmonella* T3SS (16). Although this work suggests that both models can explain observed length distributions, many questions remain unanswered. For instance, how does manipulating the concentrations of various proteins (rulers, needle/hook proteins, bases, etc.) influence the average and variance of the length distribution? What are the differences between the predictions of the ruler and substrate switching models that would allow us to unambiguously distinguish between the two?

For instance, one of the key predictions of our previous model for substrate switching was that the variance in the needle lengths depends quadratically on the average length (16). We found this relationship to be independent of all but one underlying parameter of the model, which is the number of inner proteins required to complete the inner rod. By comparing our results with the available experimental data, we predicted the number of inner protein subunits for completing the inner rod to be around six. Recent structural studies have shown that PrgJ does not form an inner rod but rather mediates the interaction between the needle proteins and other components of the base complex (13). Interestingly, although PrgJ does not form an entire inner rod, structural and proteomic studies both confirm that there are six PrgJ subunits in the base in *Salmonella* (13,17). The quadratic relation between the variance and average length implies large variability in lengths, especially for longer needles. This makes substrate switching a potentially inefficient mechanism for growing longer structures. Furthermore, it is possible that different bacterial species have evolved distinct mechanisms for length regulation. The only way to test such hypotheses is by understanding the ruler mechanism in a similar quantitative framework and to compare and contrast the implications of the mathematical model with experimental data. Although previous models of the ruler mechanism have been extremely helpful, that work did not explicitly make predictions (such as the relationship between average length and variance) that could be directly tested to distinguish between the two competing mechanisms (14). Also, these continuous models only consider critical variables like ruler and needle/hook protein concentration implicitly, making it difficult to fully explore the set of future experiments that might be able to distinguish the ruler and substrate switching models from one another in any given bacterial system.

To overcome these challenges, here we developed a discrete, stochastic model for the ruler mechanism that uses essentially the same mathematical and computational framework as our previous model of substrate switching (16). Our initial model, in which the measurement of the needle/hook structure takes place deterministically with the help of an "exact" ruler, predicts that variance should be independent of needle length and depend only on number of ruler and needle proteins in the bacterial cell. Our analysis of experimental studies in which the authors obtained length distributions of these structures for a wide

range of ruler lengths revealed a linear relationship between the variance and the average length (10–12). A more realistic approximation of our model, in which we introduce uncertainty in the measurement process by the ruler protein, which we call the "error-prone" or "logistic" ruler, predicts a linear relationship between the variance in needles and the average needle lengths if the average length is increased by simply inserting additional amino acid residues to the ruler protein, which is what is done experimentally. Interestingly, if we allow the ruler protein length to covary along with other parameters (such as the ruler protein concentration), it is possible to increase the needle length without increasing the variance even in this error-prone ruler model, which is likely impossible in the substrate switching mechanism. Taken together, our findings suggest an interesting set of evolutionary trade-offs between these two mechanisms: the substrate switching mechanism allows for energy-efficient control of the length of smaller structures, whereas the ruler protein mechanism allows for robust control of the lengths of longer structures at a higher energetic cost. Our results also suggest a very straightforward experimental program for further investigating the length control mechanism in a given bacterial species.

RESULTS

The model

The assembly of macromolecular structures such as the flagellum and T3SS injectisome is a highly regulated process and involves interaction between a number of proteins (Fig. 1 *A*) (1,2,18). Evidence suggests that there are dedicated ruler proteins, such as YscP in *Yersinia* and its homolog FliK in *Salmonella*, that are secreted during the assembly process. When the injectisome or the flagellar hook achieves an appropriate length, the C-terminal domain of the ruler protein interacts with the gate protein (FlhB in *Salmonella* and YscU in *Yersinia*) (10,19,20). It is believed that this interaction causes the substrate specificity to switch, thus preventing the needle from growing any further (11). To obtain a quantitative description for the ruler model, we first constructed a mathematical model that considers the dynamical interplay between the base complex, the needle (or the hook) subunits, and a molecular ruler of fixed length.

Fig. 1 *B* depicts the mechanism of the needle growth in T3SS injectisome using ruler proteins as per our initial mathematical model. Note that the figure only refers to the T3SS needle, but all the results described below are valid for the flagellar hook as well. Imagine a base being produced at

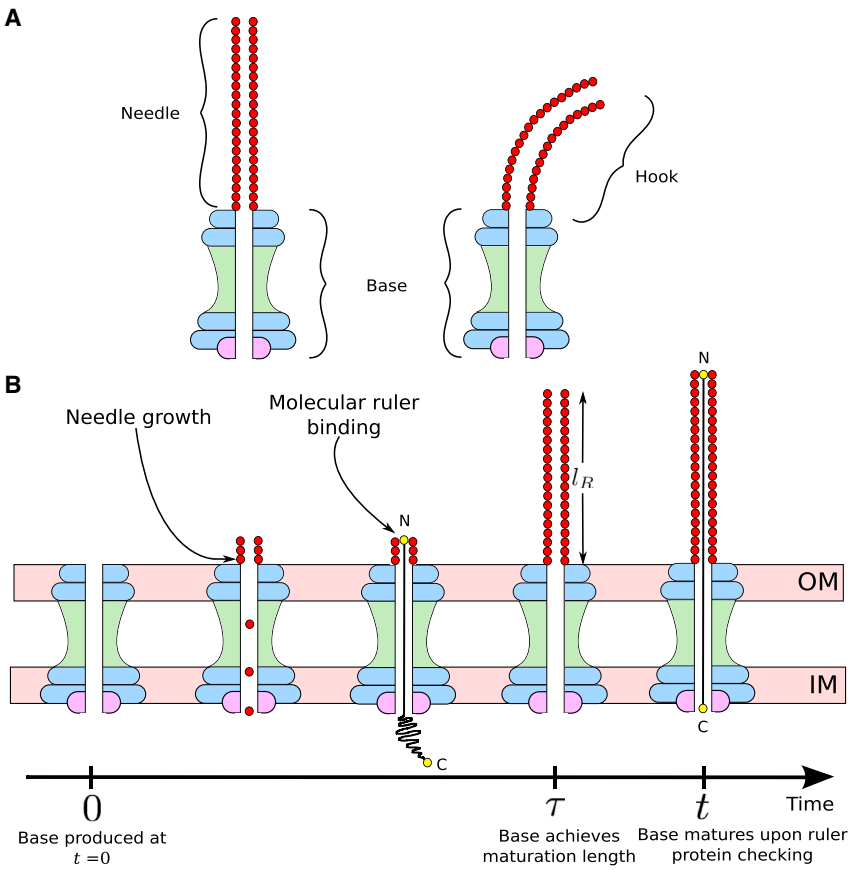

FIGURE 1 Schematic figures of the structure of the T3SS injectisome and flagellar hook (*top*) and the growth of the T3SS injectisome using an exact molecular ruler (*bottom*). (*A*) are schematic diagrams (not intended to represent structural accuracy) of the T3SS needle (*left*) and of the flagellar hook (*right*). (*B*) shows a base that is produced at time $t = 0$, binds needle proteins, and periodically checks for its needle length L with the effective ruler length l_R. At time τ, the needle achieves a length appropriate for maturation, i.e.,$L = l_R$, and the base matures at some later time t after exactly one ruler protein binding. To see this figure in color, go online.

time $t = 0$. During the assembly process, the ruler proteins and the needle proteins are secreted one at a time. The needle proteins polymerize in the extracellular region of the cell, resulting in growth of the needle itself. The needle length is represented by L, which is the distance between the surface of the outer membrane and the tip of the needle. At some time τ, after a sufficient number of needle protein bindings, the needle achieves a length that is suitable for the switch in substrate specificity. In our model, we denote the effective length of the ruler protein by l_R, and when the needle reaches an appropriate length, this facilitates the interaction between the ruler protein and the gate protein, switching the substrate specificity. Because the needle length has to be greater than or equal to the ruler length for maturation, all the ruler protein binding events that occur before the time τ do not result in maturation of the base, and the ruler protein is simply excreted into the extracellular space. As shown in the figure, the next ruler protein binding event after the base reaches the appropriate length occurs at some (random) later time t, and it is at this time that the needle stops growing and the base becomes "mature." Because the maturation event occurs only when $L \geq l_R$, followed by a ruler protein checking event, for simplicity we ignored any uncertainties involved in the measurement process in our initial model, and as such we call it the exact ruler model. Below, we develop a model that incorporates measurement uncertainties, and we shall refer to it as a logistic or error-prone ruler model.

Mathematical derivation

To develop a mathematical model for the ruler mechanism, we considered the dynamics of the main constituent proteins of the system—namely, the base, the needle (or the hook) proteins, and the ruler protein. We represent the average numbers of the immature bases, needle proteins, and ruler proteins as B, O, and R, respectively. Because the flagellar hook as well as the injectisome needle assemble outside the cell, we use the same notation, O, to denote both the hook protein and the needle protein. The assembly of these structures involves the following key processes: synthesis or production of the constituent proteins in the bacterial cell, export of the needle subunits and the ruler proteins by the base, and dilution of these proteins from cell division. In our derivation, Q denotes the production or synthesis rates, whereas λ denotes the dilution rates of the corresponding proteins. β_O and β_R are the rate at which the needle and the ruler proteins bind to the base complex and are exported, respectively. Table 1 provides a summary of the parameters in this model as well as the observables that we track. As in our previous work (16), we use a system of ordinary differential equations to represent the dynamics of the constituent proteins in the system (see the Supporting material for more details). This allows us to relate the fundamental rate parameters of the system (i.e., Q, β, etc.) to the corresponding steady-state levels of B, O, and R.

TABLE 1 Model variables, parameters, and observables and their description

	Symbol	Description
Variables	B	immature bases
	O	needle (or hook) protein
	R	ruler protein
	L	needle (or hook) length
Parameters	$Q_i \;\; i = B,R,O$	production rates
	$\beta_i \;\; i = R,O$	binding rates
	$\lambda_i \;\; i = B,R,O$	dilution rates
	l_R	ruler length
Observables	$\langle L \rangle$	average needle (or hook) length
	σ^2	variance in needle (or hook) length

We assume that the individual stochastic processes involved in the formation of the needle (namely, the production, dilution, and binding of the constituent proteins) are statistically independent Poisson processes. Note that, in this case, we are assuming that the loss of proteins from the system is a first-order stochastic chemical reaction, which is a common approach to modeling the dilution of proteins due to cell division in a growing bacterial population (see Supporting material) (16,21,22). In this derivation, protein numbers are also assumed to be constant and set at their steady-state values; in other words, we ignore fluctuations in B, O, and R. When an O or R protein binds the base, there is a transport process that carries the protein through the growing structure and allows it either to bind to the growing needle or hook (in the case of an O protein) or make a length measurement (in the case of the R protein). This transport process, which is likely driven purely by random diffusion (23), takes time, but a simple analysis of the timescales indicates that diffusion along the growing structure is ~5 orders of magnitude faster than the other events considered in this model (see Supporting material, subsection "Diffusion of the outer and ruler proteins through the needle"). As such, we consider these transport and measurement processes to occur essentially instantaneously in this model.

When a ruler protein binds, the idea is that it diffuses along the length of the needle/hook and measures that length; this represents the "periodic" ruler model put forward in (1). In our initial model, we assume that the ruler is very precise, and every ruler has the same length l_R. When a ruler protein binds, if the length of the needle/hook associated with that particular base (which we denote L) is less than l_R, then the ruler protein is excreted into the extracellular space, and the base keeps growing. If, however, $L \geq l_R$, then the base always matures; the base stops exporting needle/hook proteins and switches to exporting the next set of proteins required to finalize construction of the structure. Because every ruler is identical, and always drives maturation at exactly the same length, we refer to this as the exact ruler model.

Keeping these assumptions in mind, imagine a base being produced at $t = 0$. The probability that this base stops

growing after achieving a certain needle length L can be written as

$$P(L) = \int_0^\infty p_{\text{stop}}(L,t)\mathrm{d}t, \tag{1}$$

where $p_{\text{stop}}(L,t)$ is the probability density that the needle is of length L and the base stops growing it at time t. This can happen in two ways: either the base remains intact until time t and achieves maturation exactly at time t or the base stays immature until t and is lost due to dilution at time t. Mathematically, this can be written as

$$p_{\text{stop}}(L,t) = P_{\text{undeg}}(t)\, p_{\text{mature}}(L,t) + p_{\text{deg}}(t)\, p_{\text{immature}}(L,t). \tag{2}$$

$p_{\text{deg}}(t)$ represents the probability density that the base is lost due to cell dilution at time t, whereas $P_{\text{undeg}}(t)$ represents the probability that the base remains undiluted at time t. Both $p_{\text{deg}}(t)$ and $P_{\text{undeg}}(t)$ depend on a single rate parameter λ_B and are given by

$$p_{\text{deg}}(t) = \lambda_B e^{-\lambda_B t},$$

$$P_{\text{undeg}}(t) = e^{-\lambda_B t}.$$

$p_{\text{mature}}(L,t)$ is a joint probability density of the following three processes: the base achieves the maturation length l_R at time $\tau < t$, there is one ruler protein binding event at time t, and the base continues to grow the needle until the ruler protein "checks" its length.

$$p_{\text{mature}}(L,t) = \int_0^t \underbrace{\left(\mathrm{e}^{-\beta_O O\tau}\frac{(\beta_O O)^{l_R}\tau^{l_R-1}}{(l_R-1)!}\right)}_{l_R \text{ needle proteins in time } \tau} \underbrace{\beta_R R \mathrm{e}^{-\beta_R R(t-\tau)}}_{1 \text{ ruler protein at } t} \underbrace{\left(\mathrm{e}^{-\beta_O O(t-\tau)}\frac{[\beta_O O(t-\tau)]^{L-l_R}}{(L-l_R)!}\right)}_{L-l_R \text{ needle proteins between } \tau \text{ and } t} d\tau.$$

$p_{\text{immature}}(L,t)$ involves dilution at time t and incorporates two scenarios: one in which the length of the needle L is less than the required length for maturation at time t and a second in which the base achieves the maturation length at time τ and no ruler checking events occur between τ and t.

$$p_{\text{immature}}(L,t) = \underbrace{\mathrm{e}^{-\beta_O O t}\frac{(\beta_O O t)^L}{L!}}_{L \text{ needle proteins in time } t} \int_0^t \underbrace{\left(\mathrm{e}^{-\beta_O O\tau}\frac{(\beta_O O)^{l_R}\tau^{l_R-1}}{(l_R-1)!}\right)}_{l_R \text{ needle proteins in } t} \underbrace{\mathrm{e}^{-\beta_R R(t-\tau)}}_{0 \text{ ruler proteins between } \tau \text{ and } t} \underbrace{\left(\mathrm{e}^{-\beta_O O(t-\tau)}\frac{[\beta_O O(t-\tau)]^{L-l_R}}{(L-l_R)!}\right)}_{L-l_R \text{ needle proteins between } \tau \text{ and } t} d\tau.$$

The detailed mathematical derivation can be found in the Supporting material. The probability distribution for lengths in this model is

$$P(L) = \Theta(L-l_R)\left[\frac{z_O}{z_R}\frac{(1+z_R)}{(1+z_O)^{l_R}\left(1+z_O+\frac{z_O}{z_R}\right)^{L+1-l_R}}\right] + \Theta(l_R-L-1)\left[\frac{z_O}{(z_O+1)^{L+1}}\right], \tag{3}$$

where $z_O = \lambda_B/\beta_O O$ and $z_R = \lambda_B/\beta_R R$. The Heaviside theta functions ensure that the term associated with mature needles does not have lengths $L < l_R$ and vice versa.

For the bacteria to have most of its needles mature, it should be able to produce and bind enough needle proteins before it dilutes, which means $\beta_O O \gg \lambda_B$ or $z_O \ll 1$. Similarly, the rate at which the ruler protein checks the needles should also be sufficiently large or else dilution would dominate over maturation. At the same time, it would be energetically inefficient for the bacteria to produce a large number of ruler proteins to check the needle too many times, so $\beta_R R \not\gg \beta_O O$. This leaves us with $z_R \lesssim z_O$. We used Eq. 3 and the ignored contributions from terms $\mathcal{O}(z^2)$ or higher to obtain the average and variance in needle lengths (see the Supporting material):

$$\langle L \rangle \simeq l_R + \frac{\beta_O O}{\beta_R R}, \tag{4}$$

$$\sigma^2 \simeq \frac{\beta_O O}{\beta_R R} + \left(\frac{\beta_O O}{\beta_R R}\right)^2. \tag{5}$$

According to Eqs. 4 and 5, the average length increases linearly with the length of the molecular ruler, whereas

the variance is independent of the ruler protein length and only depends on the average number of ruler and needle/hook proteins present in the bacterial cell.

Comparison between the analytical model and stochastic simulation

In our derivation above, we made a number of simplifying assumptions (e.g., the lack of fluctuations in the values of B, O, and R) that might not hold true in a bacterial cell. To test these assumptions, we performed stochastic simulations using the Doob-Gillespie algorithm (24). In these simulations, the constituent proteins were treated as independent "agents" that interact with one another to form the needle (or the hook) (16,25–28). We model the synthesis of each type of protein (B, O, and R) as a 0 th-order chemical reaction with a constant production rate given by the relevant Q parameter. The loss of each protein from the system due to dilution is modeled as a first-order chemical reaction (with corresponding rate λ), and the binding events are represented as second-order chemical reactions (with the corresponding β-values as the rate constant). As such, total levels of B, O, and R fluctuate in these simulations due to the stochastic nature of each of these events. Each base has an arbitrary number of needle proteins associated with it. We keep track of the number of unbound constituent proteins, number of mature and immature bases, and the number of needle proteins attached to a given base. Further details of our simulation approach may be found in the Supporting material. The precise values of the parameters used in this model have not been determined experimentally, so for the purposes of comparison with our analytical results, we chose parameter values subjected to reasonable constraints; the specific values we chose for these parameters are noted in the inset and legend to Fig. 2. Note that, in our model, all lengths are measured in terms of number of proteins. We used the structural details in (29) to convert the lengths to nanometers.

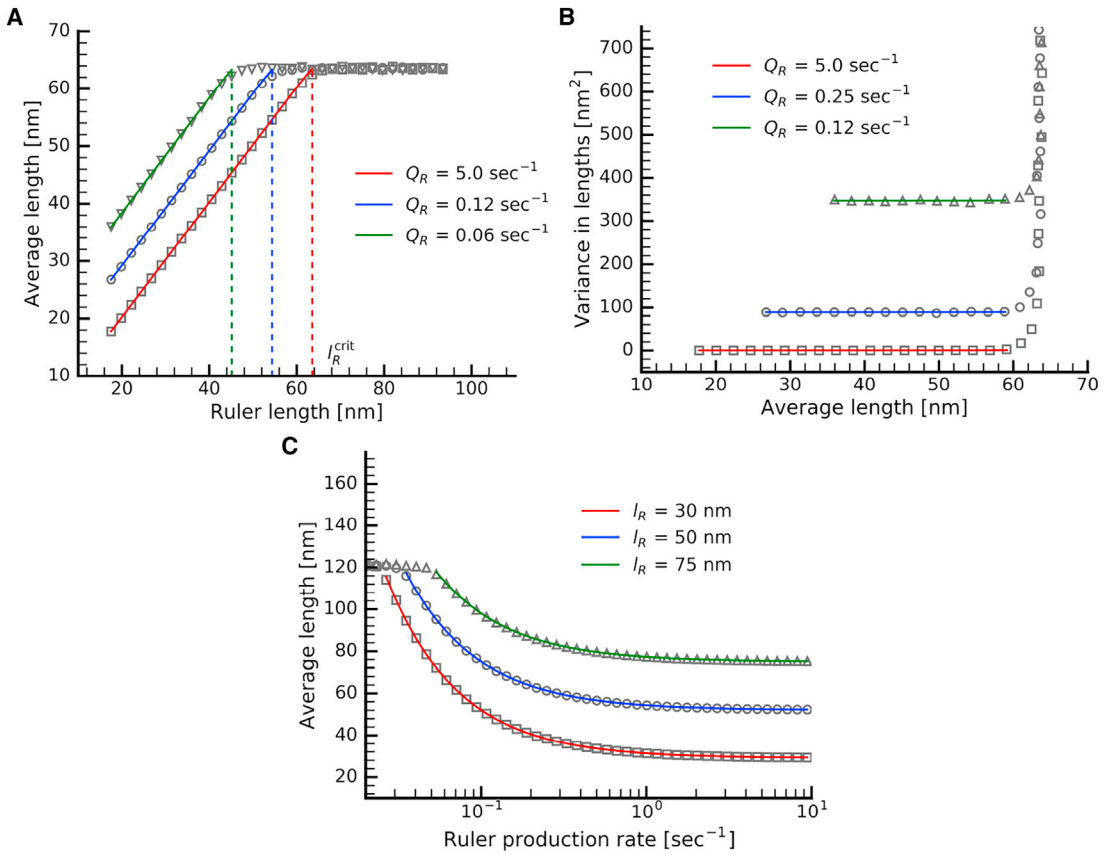

FIGURE 2 Comparison of the average and the variance in lengths obtained from the mathematical model and the stochastic simulations. (A) shows that average needle length increases linearly with the ruler length up to a certain value of ruler length l_R^{crit}, beyond which it remains constant and is approximately equal to Q_O/Q_B (in terms of number of proteins). In (B), we see that upon varying ruler length, the variance remains independent of the average needle length. In (A) and (B), the lines represent results from mathematical model for three different ruler production rates (red: $Q_R = 5.0$ s^{-1}, blue: $Q_R = 0.12$ s^{-1}, and green: $Q_R = 0.06$ s^{-1}), whereas the points represent the values from stochastic simulation for the corresponding ruler production rates. (C) shows how the average length varies upon varying the ruler production rate. In (C), the lines represent the results from mathematical model for three different effective ruler lengths (red: $l_R = 30$ nm, blue: $l_R = 50$ nm, and green: $l_R = 75$ nm). $\lambda = 5 \times 10^{-4}$ s^{-1}, and $\beta = 10^{-2}$ molecules^{-1}s^{-1} in all cases. To see this figure in color, go online.

Fig. 2 shows the comparison between the mathematical model and the stochastic simulation at steady state in terms of experimentally relevant quantities such as average length, variance in length, and ruler production rate. Note that in Fig. 2 we do not plot the solid lines for the entire range of parameters because the assumptions of our model are not valid in these regions (see Supporting material). In Fig. 2 *A*, we observe a linear dependence of the average needle length on the ruler protein length for different ruler production rates, predicted as per Eq. 4. We notice that the average needle length increases in response to increase in the ruler length only up to a certain critical value, l_R^{crit}, beyond which it remains constant. The constant length is a consequence of the fact that the bases go from being mostly mature; for instance, in Fig. 2 *A*, 100% of the bases are mature at $l_R = 10$ nm, whereas only 40% of the bases are mature beyond l_R^{crit} for $Q_R = 5.0$ molecules/s. The average length in this case is close to the ratio of production rates of the needle protein to that of the base (Q_O/Q_B) (in terms of number of proteins).

According to Eq. 5, the variance is independent of average length and depends only on the number of ruler and needle proteins. In Fig. 2 *B*, we can see that, upon varying l_R, the variance remains constant with respect to the average length for variety of ruler production rates. Fig. 2 *B* also shows exceedingly large variances for extremely long ruler lengths. When the ruler proteins are extremely long, the percentage of immature bases increases, giving needles of varying lengths. The fact that the variance is independent of average length when most of the bases are mature makes the ruler model a unique mechanism. In (16), we found that the variance depends quadratically with the average length in the substrate switching model, and because the number of inner-rod proteins required to complete the inner rod is more or less constant for any given species, there is no way to increase the length without entailing the quadratic increase in variance. However, in case of the (exact) ruler mechanism, the process of achieving the maturation length is completely deterministic. Thus, the variance does not depend on the value of l_R and depends only on the rate parameters, $\beta_O O$ and $\beta_R R$, making it possible to increase the average length by using longer rulers without changing the rate parameters of the ruler and needle proteins, leaving the variance unaltered (see Eqs. 4 and 5). Thus, the ruler mechanism can provide a much more tightly regulated needle length distribution than substrate switching.

Fig. 2 *C* shows the change in average length when the ruler protein production rate is varied. For large values of ruler protein concentration, the checking rate increases, and hence, the needles achieve maturation at a length close to the ruler length. Decreasing the ruler protein production reduces the frequency of ruler checking, thus allowing the needles to grow beyond the ruler length before they become mature. At very low production rates of the ruler protein, there are so few rulers compared with the number of bases that most bases remain immature and the average needle length in the population no longer depends on l_R.

Experimental data

In (10), the authors changed the ruler protein length by adding residues to the wild-type ruler protein YscP in *Yersinia pestis* and saw a linear relation between the ruler length and the average needle lengths. A similar study was conducted by examining the flagellar hook lengths when the length of the FliK was varied by adding chimeric residues in (11). More recently, researchers have identified InvJ (in *Salmonella enterica*), a protein homologous to YscP as well as FliK, and studied the effect of varying its length on the average length of T3SS needles (12). The key result of these experiments was that the average length increased linearly as the length of the molecular ruler, giving strong evidence for the ruler mechanism. We obtained the variance of the needle and hook lengths from the data in these studies. Fig. 3, *A and B* shows that

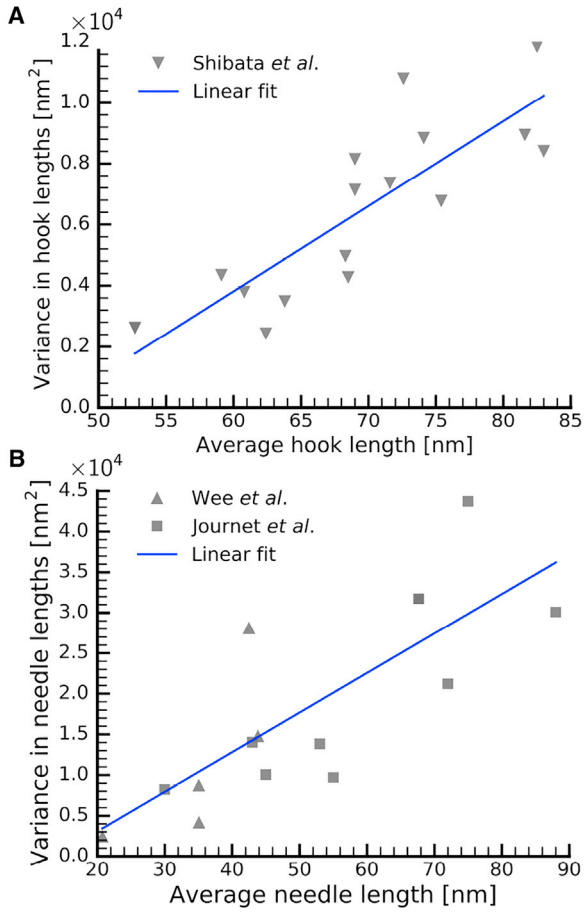

FIGURE 3 Dependence of variance on the average length in (*A*) and (*B*). Data were adapted from (10–12). Our regression analysis shows no statistical significance for a quadratic relation between the variance and the average length. The points denote the experimental data, whereas the lines represent the linear fits. To see this figure in color, go online.

variance is not independent of the average length but rather increases linearly with the average length. Our regression analysis confirms that there was no statistical significance for a quadratic relationship between these two quantities (see the Supporting material). Although a linear relationship between variance and needle length allows a tighter control over needle lengths as compared with quadratic relationship, these experimental data are not consistent with the exact ruler model.

Ruler length with an error-prone ruler

One of the assumptions in our initial model was that the interaction of the C-terminus of the ruler protein with the gate protein occurs at a fixed length. In reality, this interaction depends on conformational changes that arise as a consequence of the folding of the ruler protein, resulting in uncertainties in the measurement process. To incorporate this effect in our simulation, we allowed the length of the ruler involved in every binding interaction to fluctuate in such a way that maturation probability follows a logistic function (see Supporting material) (14). This variability in the effective ruler length makes it possible for the base to achieve maturation for a needle length that might have been too short for maturation in the exact ruler model. This effect is not straightforward to incorporate in the analytical calculation for the probability distribution of needle lengths, and thus, in the analysis that follows, we focus on simulation results. Our simulations for the error-prone ruler are essentially identical to those for the exact ruler, but in this case, when a ruler protein binds a base, the probability of maturation is not 0 or 1 (depending on whether $L < l_R$ or $L \geq l_R$) but is rather a logistic function of L. Further details on these simulations may be found in the Supporting material.

Fig. 4 shows the average and variance in lengths for a logistic ruler with a fixed error in length. In Fig. 4 *A*, we observe that varying the length of the error-prone ruler

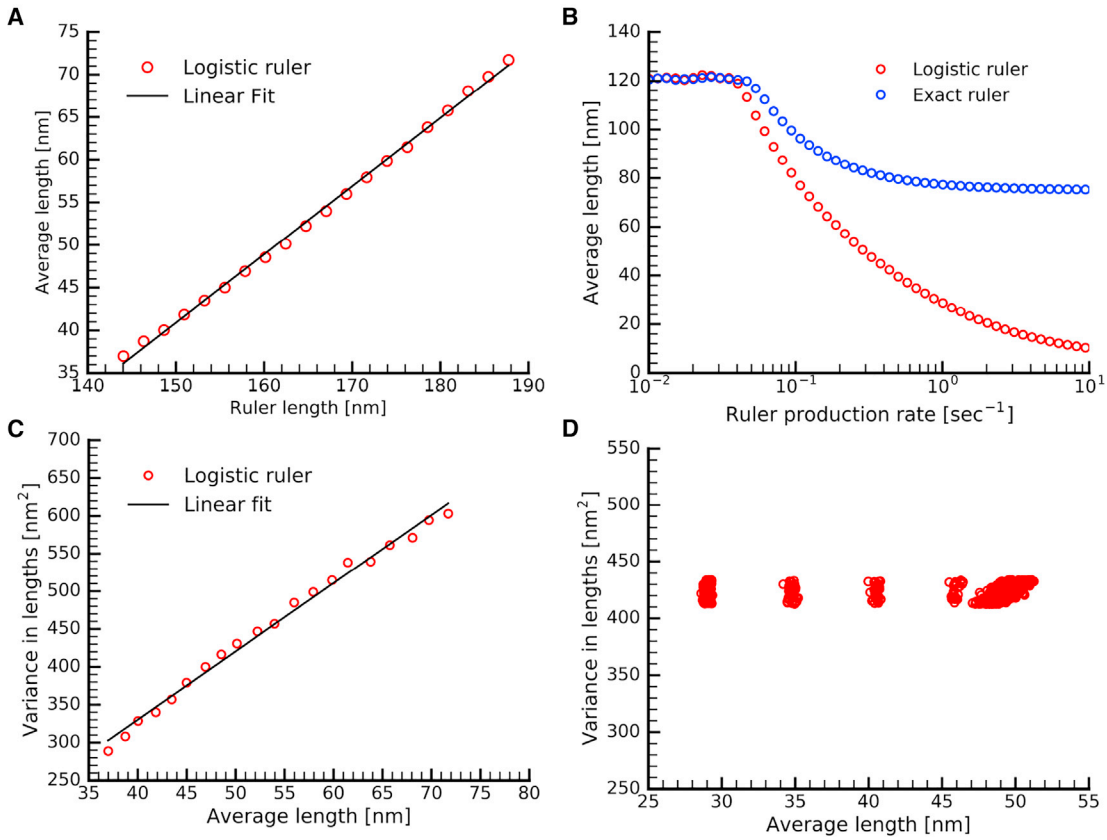

FIGURE 4 Average and variance in lengths using a logistic or error-prone ruler. (*A*) shows that the average length increases linearly with the ruler length. (*B*) shows that the average length of the needle is approximately equal in case of exact and the logistic ruler for relatively lower ruler production values. However, for larger ruler production values, even though the probability of having a short effective ruler length is small, the frequency of ruler binding is so high that the chances of a needle with a short length achieving maturation increases dramatically. In (*C*), we see that variance in length increases linearly with the average length when the ruler length is increased. Because the logistic ruler allows maturation of relatively smaller needles, this skews the needle length distribution, increasing the variance in lengths. The ruler lengths were varied from 120 to 200 nm with a fixed error of $\sim$ 30 nm in length, and $Q_R = 7.5$ and $Q_O = 3.75$. (*D*) shows that, as per a logistic ruler, it is possible to vary the ruler length without increasing the variance. Here, $l_R = 65$–100 nm with the error same as in (*C*), $Q_R = 0.25$–2.75, and $Q_O = 1.0$–3.25. Note that the variance remains roughly constant over a range of ruler lengths for different Q_R and Q_O values. To see this figure in color, go online.

(keeping all other parameters fixed) leads to a linear relationship between the average length and the ruler length, which is what was observed in case of the exact ruler as well. In Fig. 4 *B*, we can see the change in average lengths when the ruler production rate is varied. For lower production rates, only a few rulers are being produced in comparison with the number of bases available, giving us the same average lengths as obtained using the exact ruler. But when we increase the ruler production rates, the average lengths in case of a logistic ruler decrease more rapidly than they did in case of an exact ruler. This is because, at higher production rates, even though maturation at small values of L is less likely to occur, there are enough ruler checking events to ensure maturation even for needles with L significantly less than l_R. Note that this effect is prominent even at moderate values of Q_R. In Fig. 4 *C*, we observe that varying the length of the error-prone ruler (same conditions as in Fig. 4 *A*) leads to a linear relationship between the variance and the average length. A logistic ruler allows for maturation events at much smaller needle lengths, giving a large number of needles whose lengths are less than l_R. This skews the distribution of needle lengths, resulting in larger variances. Note that variance obtained through the simulation is smaller than that observed in the experiments, which could be due to a number of experimental uncertainties that were not incorporated in our simulations.

As discussed earlier, for an exact ruler, the variance in length is independent of the ruler length itself and depends only on the needle and ruler protein concentrations and export rates. We should note that Fig. 4, *A and C* represents a typical experimental situation in which the ruler protein length is changed via deletion of amino acid residues in the protein or insertion of new residues into a specific region (10–12). In this scenario, we find that the variance increases linearly with the average needle length, which is also what is observed experimentally (Figs. 3 and 4 *C*). The comparison here between our simulations and the experimental data is somewhat indirect; in other words, we are not directly fitting the computational model to the data. This is driven in part by the fact that we do not have an analytical form for the relationship between average and variance in the logistic ruler model, so we cannot simply fit the experimental data to a particular function. Although it is, in theory, possible to fit the parameters of the stochastic simulations to the experiment, the simulations in question are computationally expensive, especially because estimating the average and variance requires many simulations for any given set of parameters. As such, directly fitting the model to the data is beyond the scope of the current work. That being said, the fact that both the logistic ruler model and the data generate such a linear relationship suggests that the experimental findings are at least consistent with an error-prone ruler model.

Although only the length of the ruler protein is typically altered in experimental studies, during the course of evolution it would be possible for both the length of the ruler protein and the synthesis rates of the ruler and needle proteins to vary simultaneously. We thus tested whether it would be possible to find parameter sets in which the increase in variance seen in Fig. 4 *C* could be offset by varying other parameters. Interestingly, we found that it is indeed possible to increase the average length of the needles without significantly changing the variance (Fig. 4 *D*). Although more complex than the case with an exact ruler, it is thus nonetheless possible for an error-prone ruler to exhibit an independence of the average and variance in the length distribution.

DISCUSSION

The T3SS and flagellar hook are large and complex nanomachines that are crucial to bacterial cell function, pathogenesis, and adaptation to changing environments. One key aspect of these homologous structures is the large extracellular channel that must be constructed for both of them. The lengths of these channels have to be extremely precise; if they are too short or too long, that may negatively influence the corresponding function (30). For instance, in the case of T3SS needles, if the needles are too short, then they would be unable to get past the lipopolysaccharide layer present in the extracellular region of the bacterium, thus failing to inject the effector proteins into the host cell. On the other hand, if the needles are too long, then they will likely break from shear stress or lead to energetically inefficient transport of proteins through an excessively long channel. Similarly, a flagellar hook of an inappropriate size might lead to poor mechanical properties for locomotion.

The substrate switching and ruler mechanisms are the two most popular and well-studied explanations for length control in these structures. The substrate switching mechanism requires an inner rod present inside the base to be completed to form a mature needle, whereas according to the ruler mechanism, the bacteria "measures" the length of these structures with a dedicated ruler protein. Although there is some experimental evidence for these mechanisms, until recently, several important questions remained unanswered. For instance, is there a quantitative consistency of the proposed mechanism with experiments? What are the advantages, from an evolutionary perspective, to adapt different length control mechanisms? What future experiments could be done to further investigate these mechanisms?

In our previous work, we constructed a mathematical model for length control as per the substrate switching mechanism and compared our predictions with the available experimental data for *Salmonella*. Our model was able to explain the length distribution in the needle lengths obtained in (8), and we were also able to predict the number of PrgJ proteins required to complete the inner rod and found that our prediction was consistent with data from mass spectrometry (17). Similarly, previous continuous models had

shown that the ruler mechanism was broadly consistent with the length distributions of these structures observed experimentally in several bacterial species (14,15). Although interesting, these previous models did not investigate how changing certain key parameters, such as the concentration of needle/hook proteins or the production rate of the ruler protein, would influence the length distribution.

In the current work, we developed a discrete, stochastic model for the ruler mechanism. Interestingly, we found that an exact ruler, which is perhaps most commonly envisioned for length measurement according to the ruler mechanism, does not explain the experimental data and that a logistic ruler, which has a measurement uncertainty associated with it, is consistent with the available experimental data (10–12). Interestingly, an error-prone ruler allows bacterial cells to regulate the variance to a desired value by adapting the ruler length and ruler and needle protein synthesis rates simultaneously. This allows the ruler mechanism to exert a greater degree of control over the length of the structure than the substrate switching mechanism, particularly when the structures themselves are long. This higher degree of control, however, comes at a fairly high energetic cost. In particular, once ruler proteins are secreted, they are lost, and obtaining a low variance requires many unsuccessful ruler protein measurements for every base that becomes mature. In contrast, the substrate switching model involves a constant number of inner-rod proteins for each base, and if this number is low (around six or so, as predicted by our model and observed in experimental data (16,17)), then the energetic cost of control in this case is relatively low. We thus might expect the substrate switching mechanism to be favored as a low-energy control strategy for short structures, whereas the ruler protein mechanism might be favored for longer structures in which the lower cost of the substrate switching mechanism would be outweighed by the massive increase in variance that it entails as the needles or hooks become longer.

In addition to providing insight into the evolutionary trade-offs of various control strategies, our findings also suggest that the properties of the length distribution for an error-prone ruler depend critically on how those errors themselves are made. For instance, if error is constant (say, on average ± 1 nm), then in general one does not obtain the experimentally observed linear relationship between average length and the variance in length (Figs. 3 and 4). Instead, our model predicts that the error is proportional to the ruler length. The biophysical details of the measurement process, and how it might lead to a certain probability of maturation given L, are currently completely unknown. Biophysical simulations using molecular dynamics or more coarse-grained modalities (23,31) will likely be critical to determining exactly how the ruler protein measures the length of the growing structure, especially as more structural details of these complexes become available. The model presented here should allow the results of these simulations to be connected and compared with the length distributions observed experimentally.

Finally, our quantitative models suggest several experiments that could provide further insight into the control mechanisms themselves. For instance, researchers have studied the effect of varying the length of ruler proteins such as YscP (in *Yersinia*) and FliK (in *Salmonella*) on the length of needles and flagellar hooks, respectively, so an experiment in which one changes the concentration of the ruler proteins by means of titratable promoters for the wild type as well as for ruler length variants of these ruler proteins would help find further evidence for the ruler mechanism (see Figs. 2 *C* and 4 *C*). A similar experiment in which one changes the concentration of the inner-rod protein PrgJ in *Salmonella* could provide evidence for the substrate switching mechanism (16). In addition to the inner-rod protein PrgJ, researchers have also identified a homolog of ruler protein, InvJ, in *Salmonella* (12). By varying the concentration of InvJ as well as PrgJ, one could determine whether the substrate switching or the ruler mechanism is operational for that species. It is clear that a combination of quantitative models and experiments will be crucial for building a complete understanding of the assembly and regulation of these massive extracellular structures.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2021.05.032.

AUTHOR CONTRIBUTIONS

E.J.D., J.J.S., and M.K.N. designed the research. M.K.N. performed the research. M.K.N. and A.M. contributed to analytical tools. M.K.N. and E.J.D. wrote the article. M.K.N., E.J.D., A.M., and J.J.S. provided suggestions and edited the manuscript.

REFERENCES

1. Cornelis, G. R. 2006. The type III secretion injectisome. *Nat. Rev. Microbiol.* 4:811–825.
2. Macnab, R. M. 2003. How bacteria assemble flagella. *Annu. Rev. Microbiol.* 57:77–100.
3. Van Gijsegem, F., C. Gough, …, C. Boucher. 1995. The hrp gene locus of *Pseudomonas solanacearum*, which controls the production of a type III secretion system, encodes eight proteins related to components of the bacterial flagellar biogenesis complex. *Mol. Microbiol.* 15:1095–1114.
4. Fields, K. A., G. V. Plano, and S. C. Straley. 1994. A low-Ca2+ response (LCR) secretion (ysc) locus lies within the lcrB region of the LCR plasmid in *Yersinia pestis*. *J. Bacteriol.* 176:569–579.
5. Woestyn, S., A. Allaoui, …, G. R. Cornelis. 1994. YscN, the putative energizer of the Yersinia Yop secretion machinery. *J. Bacteriol.* 176:1561–1569.
6. Abby, S. S., and E. P. C. Rocha. 2012. The non-flagellar type III secretion system evolved from the bacterial flagellum and diversified into host-cell adapted systems. *PLoS Genet.* 8:e1002983.

7. Marlovits, T. C., T. Kubori, …, V. M. Unger. 2004. Structural insights into the assembly of the type III secretion needle complex. *Science.* 306:1040–1042.

8. Marlovits, T. C., T. Kubori, …, J. E. Galán. 2006. Assembly of the inner rod determines needle length in the type III secretion injectisome. *Nature.* 441:637–640.

9. Lefebre, M. D., and J. E. Galán. 2014. The inner rod protein controls substrate switching and needle length in a *Salmonella* type III secretion system. *Proc. Natl. Acad. Sci. USA.* 111:817–822.

10. Journet, L., C. Agrain, …, G. R. Cornelis. 2003. The needle length of bacterial injectisomes is determined by a molecular ruler. *Science.* 302:1757–1760.

11. Shibata, S., N. Takahashi, …, S. Aizawa. 2007. FliK regulates flagellar hook length as an internal ruler. *Mol. Microbiol.* 64:1404–1415.

12. Wee, D. H., and K. T. Hughes. 2015. Molecular ruler determines needle length for the *Salmonella* Spi-1 injectisome. *Proc. Natl. Acad. Sci. USA.* 112:4098–4103.

13. Hu, J., L. J. Worrall, …, N. C. J. Strynadka. 2019. T3S injectisome needle complex structures in four distinct states reveal the basis of membrane coupling and assembly. *Nat. Microbiol.* 4:2010–2019.

14. Keener, J. P. 2010. A molecular ruler mechanism for length control of extended protein structures in bacteria. *J. Theor. Biol.* 263:481–489.

15. Erhardt, M., H. M. Singer, …, K. T. Hughes. 2011. An infrequent molecular ruler controls flagellar hook length in *Salmonella enterica*. *EMBO J.* 30:2948–2961.

16. Nariya, M. K., J. Israeli, …, E. J. Deeds. 2016. Mathematical model for length control by the timing of substrate switching in the type III secretion system. *PLoS Comput. Biol.* 12:e1004851.

17. Zilkenat, S., M. Franz-Wachtel, …, S. Wagner. 2016. Determination of the stoichiometry of the complete bacterial type III secretion needle complex using a combined quantitative proteomic approach. *Mol. Cell. Proteomics.* 15:1598–1609.

18. Deng, W., N. C. Marshall, …, B. B. Finlay. 2017. Assembly, structure, function and regulation of type III secretion systems. *Nat. Rev. Microbiol.* 15:323–337.

19. Minamino, T., Y. Saijo-Hamano, …, K. Namba. 2004. Domain organization and function of *Salmonella* FliK, a flagellar hook-length control protein. *J. Mol. Biol.* 341:491–502.

20. Wiesand, U., I. Sorg, …, D. W. Heinz. 2009. Structure of the type III secretion recognition protein YscU from *Yersinia enterocolitica*. *J. Mol. Biol.* 385:854–866.

21. Mangan, S., and U. Alon. 2003. Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. USA.* 100:11980–11985.

22. Ray, J. C. J., and O. A. Igoshin. 2010. Adaptable functionality of transcriptional feedback in bacterial two-component systems. *PLoS Comput. Biol.* 6:e1000676.

23. Rathinavelan, T., L. Zhang, …, W. Im. 2010. A repulsive electrostatic mechanism for protein export through the type III secretion apparatus. *Biophys. J.* 98:452–461.

24. Gillespie, D. T. 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81:2340–2361.

25. Danos, V., and C. Laneve. 2004. Formal molecular biology. *Theor. Comput. Sci.* 325:69–110.

26. Danos, V., J. Feret, …, J. Krivine. 2007. Rule-based modelling of cellular signalling. *In* CONCUR 2007 – Concurrency Theory: 18th International Conference, CONCUR 2007, Lisbon, Portugal, September 3–8, 2007, Proceedings L. Caires and V. T. Vasconcelos, eds.. Springer, pp. 17–41.

27. Deeds, E. J., J. Krivine, …, W. Fontana. 2012. Combinatorial complexity and compositional drift in protein interaction networks. *PLoS One.* 7:e32032.

28. Suderman, R., and E. J. Deeds. 2013. Machines vs. ensembles: effective MAPK signaling through heterogeneous sets of protein complexes. *PLoS Comput. Biol.* 9:e1003278.

29. Loquet, A., N. G. Sgourakis, …, A. Lange. 2012. Atomic model of the type III secretion system needle. *Nature.* 486:276–279.

30. Mota, L. J., L. Journet, …, G. R. Cornelis. 2005. Bacterial injectisomes: needle length does matter. *Science.* 307:1278.

31. Michalski, P. J., and L. M. Loew. 2016. SpringSaLaD: a spatial, particle-based biochemical simulation platform with excluded volume. *Biophys. J.* 110:523–529.