Identifying Dynamical Bottlenecks of Stochastic Transitions in Biochemical Networks

Identifying Dynamical Bottlenecks of Stochastic Transitions in Biochemical Networks

Christopher C. Govern,[1] Ming Yang,[1] and Arup K. Chakraborty[2,3,*]

[1]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
[2]Departments of Chemical Engineering, Chemistry, and Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
[3]Ragon Institute of MGH, MIT, and Harvard, 149 13th Street, Charlestown, Massachusetts 02129, USA

(Received 6 October 2011; published 30 January 2012)

In biochemical networks, identifying key proteins and protein-protein reactions that regulate fluctuation-driven transitions leading to pathological cellular function is an important challenge. Using large deviation theory, we develop a semianalytical method to determine how changes in protein expression and rate parameters of protein-protein reactions influence the rate of such transitions. Our formulas agree well with computationally costly direct simulations and are consistent with experiments. Our approach reveals qualitative features of key reactions that regulate stochastic transitions.

DOI: 10.1103/PhysRevLett.108.058102 PACS numbers: 87.16.Yc, 05.40.−a, 82.39.Rt, 87.16.A−

Biochemical reactions underlie the function of living cells. For example, they translate stimulatory external cues to functional responses. At the microscopic level, fluctuations in reaction rates and protein concentrations are ubiquitous [1]. Such noise can drive cells out of stable states (e.g., the ''resting'' state of unstimulated cells), leading to spurious responses [2,3]. Discovering the key mechanisms that are responsible for regulating such stochastic fluctuations in cellular reaction networks is important for uncovering design principles of biological signaling networks [4]. In particular, understanding how changes in protein concentrations or mutations that change rate parameters promote fluctuation-driven transitions to pathological cellular states (e.g., cancer [5]) can help identify key deleterious protein mutations and inform efforts to manipulate specific proteins (drug targets) that would reverse such aberrant regulation.

Toward obtaining such mechanistic pictures, we calculate how small changes in rate parameters and protein concentrations affect the average time it takes for a biologically meaningful stochastic transition to occur—either a (potentially undesirable) fluctuation out of a stable state to an unstable state with biological significance or a stochastic switch between different stable states. Shorter transition times correspond to more unstable networks.

Determining these sensitivities can be computationally complex for two main reasons. First, direct simulations to determine transition times can be costly because transitions away from a stable state can be rare and therefore hard to sample [6–8]. Second, realistic biochemical networks often involve many reaction rates and concentrations, each of which must be perturbed to determine if it significantly affects the transition.

Here, we develop a semianalytical technique that aims to overcome these challenges by exploiting ideas from large deviation theory [9–18]. The semianalytical expressions reveal qualitative features that characterize the key components that affect network stability. For a nontrivial biological model, we demonstrate that our technique gives results quantitatively consistent with trajectory-based simulation results. Computationally, the advantage of our approach is that it requires only a single deterministic simulation to determine the effect of perturbing all rate constants and species concentrations, as long as the transitions are rare and the small perturbations do not lead to phase transitions.

Consider a system of N_S different chemical species, whose copy numbers evolve stochastically according to a reaction network with N_R reactions, characterized by a stoichiometric matrix E (of dimension $N_S \times N_R$) and rate constants $\mathbf{k}$ (of dimension N_R) according to predefined rate laws (e.g., mass action). The system may have one or more stable steady states for species concentrations, corresponding to different stable cellular states. We assume that the volume V is sufficiently large to make large excursions away from any particular fixed point rare relative to relaxation to the fixed point, and derivatives with respect to copy numbers (or concentrations) are well-defined. In this limit, transitions away from a fixed point, specified by a vector of concentrations ($\mathbf{c}_A^{\mathrm{FP}}$) to another fixed point $\mathbf{c}_B^{\mathrm{FP}}$ or to any other biologically relevant (set of) states, can be described by a single rate K. The percentage change in K when a rate constant k_i is perturbed by a small percentage can be quantified by $\frac{\partial \ln K}{\partial \ln k_i}$. Similarly, the percentage change in K when the concentration of species i at some arbitrary time t, $c_i(t)$, is perturbed by adding molecules of species i to the system is $\frac{\partial \ln K}{\partial c_i}$. (The absolute change in c_i is used because there is no unique time-independent concentration scale.)

First, consider transitions between stable steady states $\mathbf{c}_A^{\mathrm{FP}}$ and $\mathbf{c}_B^{\mathrm{FP}}$. We describe a possible transition path by the actual reaction propensities at each time $\hat{\boldsymbol{\nu}}(t)$ (that is, the number of reactions of each type that occur per volume per time). Given a starting point $\mathbf{c}(0)$ (e.g., $\mathbf{c}_A^{\mathrm{FP}}$), $\mathbf{c}(t)$ is uniquely defined by $\hat{\boldsymbol{\nu}}(t)$ through Eq. (1):

$$\mathbf{c}(t) = \mathbf{c}(0) + E\int_0^t \hat{\boldsymbol{\nu}}(t')dt'. \quad (1)$$

Suppose a system is evolving according to a particular path $\hat{\boldsymbol{\nu}}(t)$ and that the system is therefore at a particular point $\mathbf{c}(t)$ at time t. Over the next differential time interval $[t, t+\Delta t]$, the actual number of reactions of type i that occur is $n_i = \hat{\nu}_i(t)V\Delta t$. However, the expected number of reactions of type i is $\lambda_i = \nu_i V\Delta t$, where ν_i, the deterministic propensities, are given by the predefined rate law as a function of the state $\mathbf{c}(t)$ and the rate constant k_i. The probability of observing n_i reactions of type i over the differential time Δt follows a Poisson distribution characterized by λ_i.

Hence, the probability of observing $\mathbf{n} = \hat{\boldsymbol{\nu}}(t)V\Delta t$ reactions is

$$\mathbb{P}(\mathbf{n}) = \prod_{i=1}^{N_R}\frac{\lambda_i^{n_i}}{n_i!}e^{-\lambda_i} = \prod_{i=1}^{N_R}\frac{(\nu_i V\Delta t)^{\hat{\nu}_i V\Delta t}}{(\hat{\nu}_i V\Delta t)!}e^{-\nu_i V\Delta t}. \quad (2)$$

The first equality holds, assuming the time interval is sufficiently small so that the species concentrations do not change significantly over the interval. By Stirling's approximation and in the continuum limit, the probability density of the path $\hat{\boldsymbol{\nu}}(t)$, not necessarily a transition path connecting the end points, with t from 0 to τ, is proportional to $\exp[-VS(\hat{\boldsymbol{\nu}}, \tau)]$, where

$$S(\hat{\boldsymbol{\nu}}, \tau) = \sum_i^{N_R}\int_0^\tau dt\left(\hat{\nu}_i \ln\frac{\hat{\nu}_i}{\nu_i} - \hat{\nu}_i + \nu_i\right). \quad (3)$$

$S(\hat{\boldsymbol{\nu}}, \tau)$ serves as the action (or the rate function) of the path $\hat{\boldsymbol{\nu}}$ over $[0, \tau]$. Equation (3) parses the action of a path in a complex reaction network into contributions from individual reactions (it has been derived in a different way by Liu [19]). Hence, it can potentially identify reactions that can most effectively alter rates of rare transitions.

The minimal action S^* for the transition and the corresponding most probable (least action) transition path $\hat{\boldsymbol{\nu}}^*$ can be found by minimizing Eq. (3) over all the paths that originate from $\mathbf{c}_A^{\mathrm{FP}}$ and reach $\mathbf{c}_B^{\mathrm{FP}}$ at time τ and then minimizing over τ: $S^* = \min_\tau \min_{\hat{\boldsymbol{\nu}}} S(\hat{\boldsymbol{\nu}}, \tau)$ [9,15].

The transition rate K of a rare event is, from large deviation theory (e.g., WKB approximation [9,10]), $K = A\exp(-VS^*)$, where A depends subexponentially on the volume and both A and S^* depend on the parameters describing the system (e.g., $\mathbf{k}$ and initial concentrations). In the large-V limit, a simple calculation obtains the sensitivity of K to different parameters to be $\frac{\partial \ln K}{\partial \ln k_i} \approx -V\frac{\partial S^*}{\partial \ln k_i}$ and $\frac{\partial \ln K}{\partial c_i} \approx -V\frac{\partial S^*}{\partial c_i}$.

Perturbation of $\mathbf{k}$.—The perturbation of rate constant k_i by Δk_i changes the minimal action S^*. As shown in the Supplemental Materials [20], the action of the new optimal path can be parsed in terms of that of the original optimal path under the new k_i, plus corrections for the fact that the two optimal paths are different (not only because their end points, which are steady states of the system, change with k_i). However, the corrections are of lower order, $O(\Delta k_i^2)$ (for proof, see the Supplemental Materials and Fig. S1 [20]). Hence, to compute $\frac{\partial S^*}{\partial \ln k_i}$, and therefore $\frac{\partial \ln K}{\partial \ln k_i}$, we only have to evaluate the change of action [Eq. (3)] along the unperturbed optimal path $\hat{\boldsymbol{\nu}}^*$ when k_i changes:

$$\frac{\partial S^*(\mathbf{k})}{\partial \ln k_i} = \int_0^{\tau^*}\frac{\partial \mathcal{L}}{\partial \nu_i}\frac{\partial \nu_i}{\partial \ln k_i}\Big|_{\hat{\boldsymbol{\nu}}^*}dt = k_i\int_0^{\tau^*}\left(1 - \frac{\hat{\nu}_i^*}{\nu_i^*}\right)\frac{\partial \nu_i^*}{\partial k_i}dt, \quad (4)$$

where $\mathcal{L} = \hat{\nu}_i \ln\frac{\hat{\nu}_i}{\nu_i} - \hat{\nu}_i + \nu_i$. If ν_i is linear with k_i (e.g., mass action), the sensitivities can be further simplified:

$$\frac{\partial \ln K}{\partial \ln k_i} \approx -V\frac{\partial S^*(\mathbf{k})}{\partial \ln k_i} = V\int_0^{\tau^*}(\hat{\nu}_i^* - \nu_i^*)dt. \quad (5)$$

For reactions that must occur more frequently than they would have deterministically in order to make the transition occur, the right-hand side of Eq. (5) is positive, suggesting that the transition rate K increases with the rate parameters of such reactions. For rate constants that participate in multiple reactions (e.g., dephosphorylations by the same phosphatase), the right-hand side of Eq. (5) will contain a summation over all such reactions.

Equation (5) provides a way to calculate the effect of perturbing rate constants on the transition time, given a single input, the unperturbed optimal path $\hat{\boldsymbol{\nu}}^*$. This input path can be determined numerically using the efficient geometric minimal action method (gMAM; [15]).

Perturbation of $\mathbf{c}$.—Adding molecules of a species not governed by conservation laws does not alter the rate of rare transitions (e.g., steady states do not change). For species governed by conservation laws [as specified by E in Eq. (1)], perturbing concentrations by adding molecules to the system can change the steady states $\mathbf{c}_A^{\mathrm{FP}}$ and $\mathbf{c}_B^{\mathrm{FP}}$, the optimal path $\hat{\boldsymbol{\nu}}^*$, and the deterministic propensities $\boldsymbol{\nu}^*$ along the optimal path. We show that, ignoring terms of $O(\Delta c_i^2)$ and higher, the change in minimal action is due to the change in deterministic propensities evaluated along the unperturbed optimal path $\hat{\boldsymbol{\nu}}^*$ (for proof, see the Supplemental Materials and Fig. S2 [20]). Arguments analogous to the perturbation in k_i lead to

$$\frac{\partial \ln K}{\partial c_i} \approx -V\sum_j^{N_R}\int_0^{\tau^*}\left(1 - \frac{\hat{\nu}_j^*}{\nu_j^*}\right)\frac{\partial \nu_j^*}{\partial c_i}dt, \quad (6)$$

where $\frac{\partial \nu_i^*}{\partial c_i}$ can be evaluated at each time point, given the rate law. For mass action kinetics, the right-hand side of Eq. (6) can be further simplified to be $V\sum_j^{N_R}\int_0^{\tau^*}E_{ij}\frac{\hat{\nu}_j^* - \nu_j^*}{c_i}dt$.

To study the escape from a stable state to a predefined condition other than a steady state (e.g., the number of molecules of a particular species acquires a specific value), Eqs. (5) and (6) still apply. The only modification is that the optimal path $\hat{\boldsymbol{\nu}}^*$ as the input into Eqs. (5) and (6) needs a further optimization over all the possible end states consistent with this condition, as described algorithmically in

[15]. Note that Eq. (6) does not directly apply, as derived, to the case where the perturbed species participates in the definition of the specified end condition (see the Supplemental Materials [20]).

To test the accuracy of the methods developed above, we apply them to a biochemical reaction network that is present in several cell-signaling networks. Upon external stimulation, in many cell types, Ras proteins are converted from their inactive, guanosine diphosphate (GDP)-bound state to an active guanosine triphosphate (GTP)-bound state. Active Ras can stimulate several downstream pathways, and aberrant regulation of Ras activation underlies many cancers. We study a particular version of a simple model for Ras activation described previously [21–23]. This model has 26 reactions (and associated rate constants), 14 species governed by 5 conservation equations, and in our study has about 400 molecules involved. The copy number of individual species is as small as 10 in our model studies. The dynamics do not observe detailed balance.

The main feature of the model is that Ras can be activated via two pathways mediated by proteins called Ras guanyl nucleotide-releasing protein (RasGRP) and Son of Sevenless (SOS) and deactivated by Ras GTPase-activating protein (RasGAP). The activation by SOS is governed by a positive feedback loop: SOS's catalytic activity increases significantly by the binding of the active form of Ras, RasGTP, to its allosteric site [denoted by GTP-SOS; in this Letter, we use x-SOS (SOS-x) to denote the species x bound to the SOS allosteric (catalytic) pocket, where $x =$ GTP or GDP, with "Ras" omitted]. This enables the system to exhibit bistability when SOS concentration is at an intermediate level [21]. Meanwhile, at a low SOS level, only one stable state exists, characterized by a low level of RasGTP. Thus, the model is rich enough to investigate how cells control stochastic transitions between multiple stable states (at an intermediate SOS level) and suppress fluctuations from a single stable state (at a low SOS level), preventing spurious activation to an undesired state with a high level of active Ras. We used Eqs. (5) and (6) to predict the sensitivities of the transition time for both these situations. To obtain the unperturbed optimal path $\hat{\boldsymbol{\nu}}^*$, we implemented the geometric minimal action method [15].

To test the accuracy of Eqs. (5) and (6), we explicitly calculated the transition time under the unperturbed parameters and under perturbed values of each parameter (1% increase in rate constants; 1 molecule increase in each conserved species). We chose forward flux sampling (FFS; see [24]) as the trajectory-based simulation method and the RasGTP level as its thresholding parameter.

As shown in Figs. 1(a) and 1(b), Eqs. (5) and (6) predict a correctly ordered list of key reactions and species that influence the transition time. Additionally, the predictions are in remarkable quantitative agreement with the direct simulation results. Our technique correctly identifies the key reactions for regulating Ras activation to be RasGAP catalytic activity, GTP-SOS catalytic activity, and RasGRP

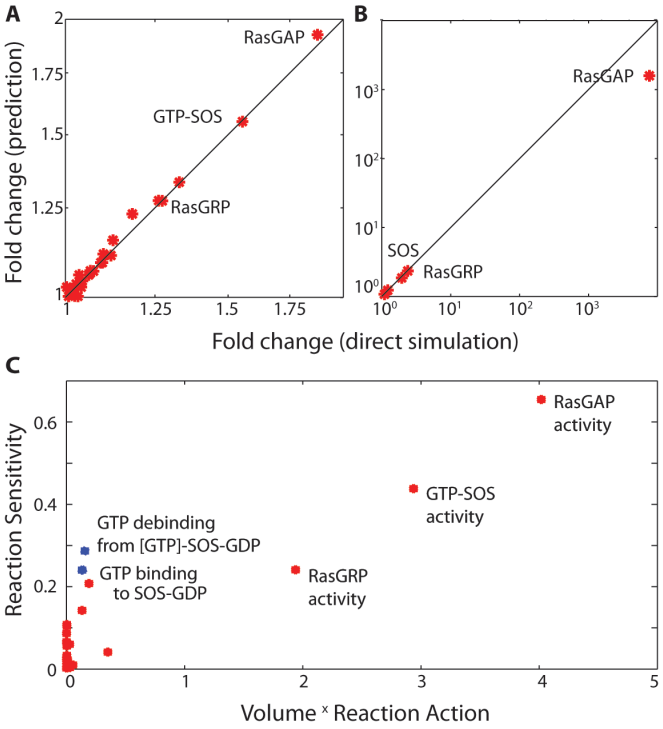

FIG. 1 (color). (a),(b) Quantitative consistency between predictions and direct simulation results at SOS = 50 (intermediate). The fold change in the transition rate (i.e., a larger rate divided by a smaller one) (a) when each rate constant is increased by 1% and (b) when one molecule of each type is added to the system. Each red dot represents the perturbation of (a) a different rate constant and (b) a different species concentration. The direct simulation results are the averages of ten independent FFS runs, and, for each FFS run, 10^3 points were stored on each surface. Only the results for the key reactions, namely, RasGAP catalytic activity, GTP-SOS catalytic activity, and RasGRP catalytic activity, are labeled in (a), and only the results for the key species, namely, RasGAP, SOS, and RasGRP, are labeled in (b). Similar calculations are done at SOS = 15 (low) and reported in Fig. S3 [20]. (c) Reaction action is a dominant but not sole factor for sensitivity. Each dot represents the sensitivity and action of one reaction at SOS = 50. The rankings of reactions by action and by sensitivity are similar. However, note that the binding (unbinding) reaction of GTP to (from) SOS-GDP, indicated by the blue dot, has small action but large sensitivity. Only the reactions with relatively large sensitivities are labeled.

catalytic activity. The most important species are RasGAP, SOS, and RasGRP, which are, as expected, associated with these key reactions. These predictions are consistent with recent experimental results in T-cell cancers [23], where the mutation of Ras that abrogates RasGAP activity is observed to significantly change active Ras level; i.e., spurious Ras activation is very sensitive to RasGAP activity. When the SOS level decreases, the relative importance of RasGRP increases {compare Figs. 1(a) and 1(b) to Figs. S3A and S3B [20]}. This is consistent with the experimental finding that RasGRP overexpression is a major cause for human cancer when little stimulus is present [25].

The quantitative discrepancies between predictions from Eqs. (5) and (6) and direct simulations [Fig. 1(a) and 1(b)] are due to the finite size of the simulation system, the finite change in rate parameters and initial concentrations, and statistical uncertainties in FFS results. Note that, for RasGAP, the species in Fig. 1(b) to which the transition time is the most sensitive, an increase of one molecule represents a relatively large (10%) change in concentration, which explains the deviation between the prediction and the direct simulation. Additional calculations (data not shown) suggest that, for this model, our method can predict the qualitative order of reactions or species (but not the quantitative values of their sensitivities), even when larger perturbations are applied, as long as the perturbations do not lead to phase transitions.

Both direct simulation and our method require one simulation for the unperturbed transition: for direct simulation, to determine the unperturbed transition rate K, and, for our method, to determine the unperturbed optimal path $\hat{\boldsymbol{\nu}}^*$. To compute the sensitivity, direct simulation requires additional simulations to explore all possible perturbations of the parameters. For the simple model we studied, this corresponds to 31 additional simulations. Our method requires no additional simulations.

Furthermore, the semianalytical forms of Eqs. (3) and (5) enable us to draw qualitative conclusions about key reactions responsible for network stability. It might be expected that the network will be the most sensitive to reactions that require the rarest fluctuations to drive the transition. Note that a reaction's contribution to the rarity of the transition is quantified by reaction action $S_i^* = \int_0^{\tau^*}(\hat{\nu}_i^* \ln\frac{\hat{\nu}_i^*}{\nu_i^*} - \hat{\nu}_i^* + \nu_i^*)dt$ [as seen from Eq. (3)]. By ranking reactions by their sensitivities and by their actions, we see that rarity has a dominant effect on sensitivity but that it is not the only effect [Fig. 1(c)].

Equations (3) and (5) show that (at least) three factors determine the sensitivity to a particular reaction: (a) the rarity of its required fluctuation, as measured by S_i^*; (b) the frequency of the reaction, as measured by its deterministic propensity ν_i^*; and (c) the uniformity of the fluctuations distributed over the time course of the transition. Holding the other two factors constant, the reaction sensitivity increases with each of these factors.

The importance of (b) can be seen more vividly by noticing that the second-order approximation of S_i^* yields $S_i^* \approx \int_0^{\tau^*}(\hat{\nu}_i^* - \nu_i^*)^2/2\nu_i^* dt$; hence, given the same action S_i^*, the reaction with larger ν_i^* has a larger sensitivity. This explains why, in Fig. 1(c), the binding (unbinding) reaction of RasGTP to (from) the allosteric pocket of SOS with RasGDP bound to its catalytic pocket (i.e., SOS-GDP) has small action but relatively large sensitivity, compared to RasGRP activity. Our calculation shows that the deterministic propensities of these two reactions are about 30 times bigger than that of RasGRP activity. For (c), note that S_i^* is convex with respect to $\hat{\nu}_i^*/\nu_i^*$ (a similar functional form also appears in the Boltzmann H theorem) while the reaction sensitivity is linear; we can show that, by Jensen's inequality, given two reactions with the same action and the same frequency, the reaction with sustained modest fluctuations, rather than a short large burst, has a larger sensitivity (for proof, see the Supplemental Materials [20]). The action also vanishes faster than the sensitivity because of the above difference in scaling. This may be why we see a cluster of reactions with moderate sensitivities but small actions in Fig. 1(c).

We have applied a large deviation theory method to a finite system to obtain accurate results, which offers substantial computational advantages compared to direct simulation. It also provides qualitative insights into the features of biochemical reactions important for regulating fluctuation-driven transitions. These features should enable studies of real biological systems. For example, we could seek to determine how spurious transitions caused by cancerous mutations may be quenched by drugs that target certain species involved in key reactions.

M. Y., C. C. G., and A. K. C. conceived the idea, did the analysis, and wrote the paper. M. Y. and C. C. G. collaborated on the theory and simulation.

We thank Chong Wang for helpful discussions. We would like to acknowledge financial support from NCI (PSOC) and P01 AI09158001.

*Corresponding author.
arupc@mit.edu

[1] A. Raj and A. van Oudenaarden, Cell **135**, 216 (2008).
[2] N. Q. Balaban *et al.*, Science **305**, 1622 (2004).
[3] M. Thattai and A. van Oudenaarden, Genetics **167**, 523 (2004).
[4] I. Lestas, G. Vinnicombe, and J. Paulsson, Nature (London) **467**, 174 (2010).
[5] P. B. Warren, Phys. Rev. E **80**, 030903 (2009).
[6] A. Dickson and A. R. Dinner, Annu. Rev. Phys. Chem. **61**, 441 (2010).
[7] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, Annu. Rev. Phys. Chem. **53**, 291 (2002).
[8] W. E and E. Vanden-Eijnden, Annu. Rev. Phys. Chem. **61**, 391 (2010).
[9] H. Touchette, Phys. Rep. **478**, 1 (2009).
[10] M. I. Dykman, E. Mori, J. Ross, and P. M. Hunt, J. Chem. Phys. **100**, 5735 (1994).
[11] R. S. Maier and D. L. Stein, Phys. Rev. E **48**, 931 (1993).
[12] E. Aurell and K. Sneppen, Phys. Rev. Lett. **88**, 048101 (2002).
[13] M. Heymann and E. Vanden-Eijnden, Phys. Rev. Lett. **100**, 140601 (2008).
[14] E. Vanden-Eijnden and M. Heymann, J. Chem. Phys. **128**, 061103 (2008).
[15] M. Heymann, Ph.D. thesis, New York University, 2007.
[16] M. I. Dykman, I. B. Schwartz, and A. S. Landsman, Phys. Rev. Lett. **101**, 078101 (2008).
[17] M. Khasin and M. I. Dykman, Phys. Rev. Lett. **103**, 068101 (2009).

[18] M. Assaf, E. Roberts, and Z. Luthey-Schulten, Phys. Rev. Lett. **106**, 248102 (2011).
[19] D. Liu, J. Comput. Phys. **227**, 8672 (2008).
[20] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.108.058102 for proofs of Eqs. (4) and (6), additional numerical results at a low SOS level, and proofs of qualitative features of key reactions [point (c)].
[21] J. Das *et al.*, Cell **136**, 337 (2009).
[22] M. J. Riese *et al.*, J. Biol. Chem. **286**, 5254 (2011).
[23] C. Hartzell *et al.* (unpublished).
[24] R. J. Allen, P. B. Warren, and P. R. ten Wolde, Phys. Rev. Lett. **94**, 018104 (2005).
[25] J. O. Lauchle *et al.*, Nature (London) **461**, 411 (2009).