APS physics

CHORUS

This is the accepted manuscript made available via CHORUS. The article has been published as:

Statistical Mechanics of Viral Entry

Yaojun Zhang and Olga K. Dudko

Phys. Rev. Lett. **114**, 018104 — Published 6 January 2015

DOI: 10.1103/PhysRevLett.114.018104

Statistical mechanics for the invasion strategy of enveloped viruses

Yaojun Zhang and Olga K. Dudko
Department of Physics, University of California at San Diego, La Jolla, CA 92093
(Dated: December 12, 2014)

Viruses that have lipid-membrane envelopes infect cells by fusing with the cell membrane to release viral genes. Membrane fusion is known to be hindered by high kinetic barriers associated with drastic structural rearrangements – yet viral infection, which occurs by fusion, proceeds on remarkably short timescales. Here, we present a quantitative framework that captures the principles behind the invasion strategy shared by all enveloped viruses. The key to this strategy – ligand-triggered conformational changes in the viral proteins that pull the membranes together – is treated as a set of concurrent, bias field-induced activated rate processes. The framework results in analytical solutions for experimentally measurable characteristics of virus-cell fusion and enables us to express the efficiency of the viral strategy in quantitative terms. The predictive value of the theory is validated through simulations and illustrated through recent experimental data on influenza virus infection.

PACS numbers:

Many viruses – including influenza, HIV and West Nile virus – have lipid envelopes that enable them to enter host cells and help avoid detection by the host immune system. Viruses reproduce by hijacking the host replication machinery, which requires mixing viral genes into the cell – a task that is accomplished by the fusion of the viral and cell membranes [1]. However, membrane fusion is a process that is highly energetically unfavorable, as it involves high kinetic barriers due to extensive rearrangements in the membranes and at the membrane/water interface [2]. Various ways to induce membrane fusion *in vitro* have been developed [3–6]. For virus-cell fusion to occur *in vivo* on a physiologically viable timescale, viral "fusion proteins" serve the required role of a catalyst. Once exposed to a trigger (low pH or cell receptors), the proteins, anchored in the viral membrane, first unfold and insert into the target membrane, and then refold, harnessing the energy liberated by the refolding to pull the membranes together (Fig. 1).

This general strategy – a signal-triggered conformational change in fusion proteins that lowers the barriers to membrane fusion – is shared by all enveloped viruses despite their differences in structure and details of entry [7]. In addition to viral infection, these principles operate in rapid neuronal communication enabled by Ca^{2+}-triggered fusion of synaptic vesicles with a cell, as well as during fertilization, development and immune responses, all of which involve fusion between cells [8].

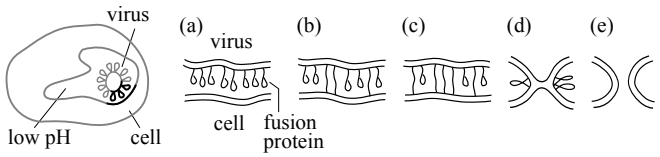

FIG. 1: Schematic of the invasion strategy of an enveloped virus. (Left) The virus at the onset of fusion with the membrane of the endosome, a low-pH compartment in the cell. The region in black is shown in (a-e) at different stages of fusion. (a) The virus is docked to the endosomal membrane. (b) In response to low pH, viral fusion proteins extend and insert into the target membrane, arresting the virus. (c) A sufficient number of neighboring proteins (here, three) are inserted to induce membrane curvature. (d) Proteins refold, pulling the membranes together to attain hemifusion. (e) Fusion pore. For a more detailed illustration, see [7].

The common principles that arise in the bewildering diversity of biological fusion phenomena suggest that these phenomena can be unified through a single physics problem [9]. Here, we formulate a predictive statistical-mechanical framework that attempts to capture – quantitatively – the common principles by which enveloped viruses overcome high kinetic barriers to infect host cells on a timescale of minutes. The theory establishes a direct connection with a series of recent single-particle assays capable of measuring the kinetics of viral infection [10, 11] and neuronal communication [12, 13]. The theory results in analytical expressions for the key quantities that have become accessible due to these experiments: the dwell time distributions for the stages of fusion as a function of an external signal. As we demonstrate through the application to influenza virus fusion measurements [10, 11], the expressions are suitable for a direct, global fit to experimental data. The parameters extracted from the fit yield, for any value of the external signal, the activation barriers, the kinetic rates and the required number of fusion proteins.

A paradigm for the viral infection mechanism is influenza, whose entry pathway is typical for many viruses [1]. To build the framework, we first summarize some key observations on influenza entry [1, 2, 7, 8, 10, 11]:

1) The virus enters the cell by endocytosis (engulfment by the cell) and moves to the acidic cellular compartment called endosome, where, exposed to low pH ($\lesssim 5.5$ [14]), fuses with the endosomal membrane to release viral genes.

2) Virus-endosome fusion is catalyzed by viral proteins hemagglutinins, which, subject to low pH, undergo a two-step irreversible conformational rearrangement. First, the proteins extend and insert into the membrane (Fig. 1b), arresting the virus. Second, the proteins refold, pulling the viral and endosomal membranes together (Fig. 1d).

3) Fusion proceeds through hemifusion, in which only the outer layers of the two bilayers have merged (Fig. 1d); the hemifused bilayer then opens into a fusion pore (Fig. 1e) delivering the viral contents into the cytosol.

4) Effects of mutations in fusion proteins indicate that the rate-limiting step of the hemifusion process is the protein conformational transition into the extended state in conjunction with the requirement that several neighboring proteins must undergo this transition. Once the required number of neighboring proteins have inserted into the target membrane, the protein refolding and membrane hemifusion occur very rapidly.

These observations help us identify the key states in the kinetic scheme for virus-cell fusion:

$$D \xrightarrow{Mk} I_1 \xrightarrow{(M-1)k} \dots I_{N_A} \dots I_{N_H} \xrightarrow{k_1} H \xrightarrow{k_2} F, \quad (1)$$

where $k = k(pH)$ is the rate of the conformational transition into the extended state for a single protein. The scheme describes a process initiated in the "docked" state (D), in which a viral membrane patch containing M fusion proteins is brought to proximity with the endosomal membrane. Subject to a pH drop, proteins within the patch independently undergo transitions into the extended state and insert into the target membrane (I_i for i insertions). N_A insertions anywhere in the patch arrest the virus (I_{N_A}), N_H neighboring insertions lead to hemifusion (H), followed by fusion (F).

Our goal is to derive a solution for an information-rich quantity amenable to experimental measurement: the distribution of times to reach hemifusion (Fig. 1d) in response to a pH drop. We accomplish the derivation in two steps. First, by solving the set of rate equations [15] corresponding to the kinetic scheme in Eq. (1), we arrive to a recurrence relation for the distribution of times for n viral proteins anywhere in the patch of M proteins to insert into the target membrane (Fig. 1b): $p(t|n) = (M-n+1)ke^{-(M-n+1)kt}\int_0^t e^{(M-n+1)kt'}p(t'|n-1)dt'$. We find that this relation can be solved analytically:

$$p(t|n,M;pH) = \frac{M!k}{(n-1)!(M-n)!}\left(e^{kt}-1\right)^{n-1}e^{-Mkt}. \quad (2)$$

When n is set to be equal to N_A, Eq. (2) describes the distribution of times from a pH drop to viral arrest with N_A viral proteins inserted into the membrane: $p_{D\to A}(t,pH)$. The result in Eq. (2) could have alternatively been obtained simply by realizing its connection to the probability of getting n successes, each with probability $1-e^{-kt}$, out of M Bernoulli trials, described by the binomial distribution. In the limit of $M \gg n$, Eq. (2) reduces to a gamma-distribution [10, 11].

In the second step of deriving the hemifusion time distribution we account for the requirement that the inserted proteins must be neighbors to afford the distortion of the membrane necessary for hemifusion (Fig. 1c-d) [10, 11]. This requirement can be expressed as a convolution of two distributions: the time distribution for n protein insertions [Eq. (2)] and the probability to find n insertions when N_H neighbors first appear among them, $p(t,pH) = \int_0^M p(t|n,M;pH)p(n|N_H,M)dn$. Since the subsequent protein refolding and membrane hemifusion are very fast, $k_1 \gg Mk$, the result of the convolution, $p(t,pH)$, is essentially equivalent to the hemifusion time distribution. For a sufficiently large M, $p(n|N_H,M)$ is a Gaussian with some mean μ and variance σ^2, and the above convolution can be solved analytically to yield the distribution of times between a pH drop and hemifusion:

$$p_{D\to H}(t,pH) = \frac{Mke^{-kt}e^{-\frac{\left[(M-1)(1-e^{-kt})+1-\mu\right]^2}{2[(M-1)(1-e^{-kt})e^{-kt}+\sigma^2]}}}{(2\pi)^{\frac{1}{2}}\left[(M-1)(1-e^{-kt})e^{-kt}+\sigma^2\right]^{\frac{1}{2}}}. \quad (3)$$

Describing the key quantity reported in single-particle measurements of viral fusion, Eq. (3) emerges as a natural fitting tool for analyzing experimental data. Simple scaling arguments [15] lead to the functional form of μ and σ^2, the mean and variance for the number of insertions when N_H neighbors first appear among them:

$$\begin{aligned}\mu(M,N_H) &= a_\mu(N_H)M^{1-\frac{1}{N_H}} + b_\mu(N_H),\\ \sigma^2(M,N_H) &= a_\sigma(N_H)M^{2-\frac{2}{N_H}+b_\sigma(N_H)}\end{aligned} \quad (4)$$

with the coefficients $a_{\mu,\sigma}(N_H)$, $b_{\mu,\sigma}(N_H)$ determined by the geometry of the protein arrangement in the viral membrane. With a hexagonal lattice as the natural choice of protein arrangement, the coefficients are accurately described by $a_\mu(N_H) = \left[0.73(N_H-3.3)^2+2.5\right]^{-\frac{1}{N_H}}$, $b_\mu(N_H) = 0.86N_H - 0.72$, $a_\sigma(N_H) = \left[2.8(N_H-3.7)^2+10\right]^{-1}$, $b_\sigma(N_H) = \left[0.0037(N_H-4.2)^2-0.0068\right]N_H^2$.

Equation (3) describes the distribution of times elapsed between the pH drop and hemifusion. Using the appropriate convolution, we derive the analogous distribution of times elapsed between arrest and hemifusion:

$$p_{A\to H}(t,pH) = \frac{(M-N_A)ke^{-kt}e^{-\frac{\left[(M-N_A-1)(1-e^{-kt})+1-\mu+N_A\right]^2}{2[(M-N_A-1)(1-e^{-kt})e^{-kt}+\sigma^2]}}}{(2\pi)^{\frac{1}{2}}\left[(M-N_A-1)(1-e^{-kt})e^{-kt}+\sigma^2\right]^{\frac{1}{2}}}. \quad (5)$$

The non-single-exponential time distributions for hemifusion evident from Eq. (3) highlight the fact that

attaining hemifusion is not a single-rate process. Nevertheless, the time distribution in Eq. (3) allows us to compute the mean time to reach hemifusion at a given pH and the corresponding variance:

$$\bar{t}_{D\to H}(pH) = \frac{\mu}{M-(\mu-1)/2}\frac{1}{k(pH)},$$
$$\overline{(t-\bar{t})^2}_{D\to H}(pH) = \frac{\mu+\sigma^2}{[M-(\mu-1)/2]^2}\frac{1}{k^2(pH)}. \quad (6)$$

In simple terms, the results for the mean and variance reveal that it takes on average μ protein insertions, each an exponential decay with the timescale $\sim [Mk(pH)]^{-1}$ and variance $\sim [Mk(pH)]^{-2}$, to attain hemifusion, and the term σ^2 arises from the fact that the insertion number is distributed around μ with variance σ^2.

To make Eqs. (3-6) applicable in the entire range of the experimental pH values, we must establish the pH-dependence of the rate of a transition of a single protein into the extended state, $k(pH)$. While proton binding may be fast relative to protein bond fluctuations [16, 17], proton binding to a folded protein can be slowed down significantly due to the buried groups [17]. This makes the number of protons, Q, that can contribute to the protein conformational transition through binding, a slow variable, rendering Q a reasonable reaction coordinate for the transition. The combined Gibbs free energy $G(Q, pH)$ of a fusion protein and the proton reservoir at a given pH can be related to the combined free energy $G_0(Q)$ at physiological pH_0 as [18]:

$$G(Q, pH) = G_0(Q) + \ln 10\; k_BTQ(pH - pH_0). \quad (7)$$

The expression for the rate $k(pH)$ can now be derived as Kramers' rate of a thermally activated reaction, $k = D\left[\int_{well} dQe^{-G/k_BT}\int_{barrier} dQe^{G/k_BT}\right]^{-1}$ [19], on the potential in Eq. (7). With a generic shape of $G_0(Q)$ featuring a well and a barrier, this leads to [15]:

$$k(pH) = k_0\left(1-\frac{2}{3}\ln 10\;(pH_0-pH)\frac{k_BTQ^\ddagger}{\Delta G^\ddagger}\right)^{\frac{1}{2}}\times$$
$$\exp\left\{\frac{\Delta G^\ddagger}{k_BT}\left[1-\left(1-\frac{2}{3}\ln 10\;(pH_0-pH)\frac{k_BTQ^\ddagger}{\Delta G^\ddagger}\right)^{\frac{3}{2}}\right]\right\}, (8)$$

where k_0 is the rate at the physiological pH_0, and $\Delta G^\ddagger$ and $Q^\ddagger$ are the height and location of the activation barrier at pH_0. We note the close analogy between the pH-dependent rate in Eq. (8) and the force-dependent rate in [20].

To validate the developed theory, we apply it to the time distributions, generated from simulations in the pH range 4 <pH< 6, for two stages of the fusion process: from the pH-drop to viral arrest (Fig. 2, left), and from the pH-drop to hemifusion (Fig. 2, right). The simulations model insertions of proteins, arranged on a hexagonal lattice with M sites, as transitions with the rate

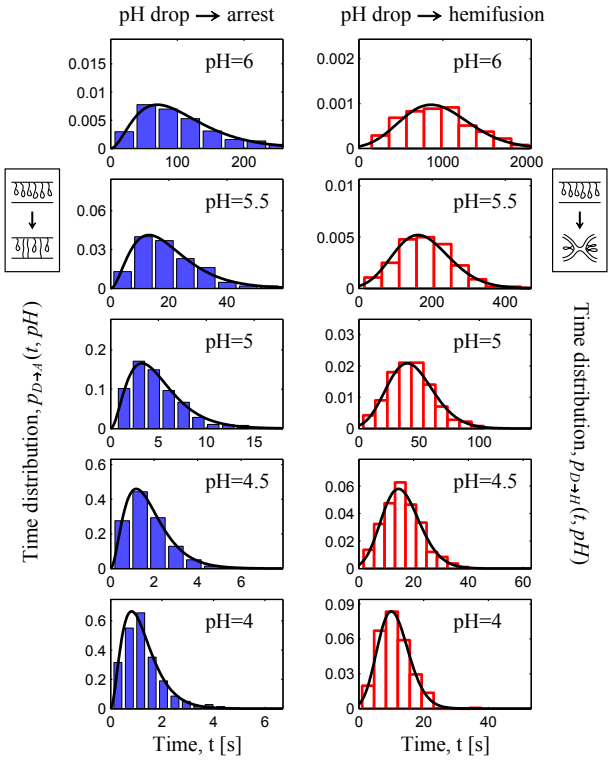

FIG. 2: Time distributions for attaining arrest (left) and hemifusion (right) at different values of pH from numerical simulations (histograms) and a global fit (curves) to the analytical expressions in Eq. (2) (left) and Eqs. (3,4) (right).

$k(pH)$; the times t_A and t_H for the arrest and hemifusion are recorded whenever N_A and N_H of such transitions take place [15]. We find that Eqs. (2) and (3,4) reproduce the simulated time distributions under all pH conditions while accurately recovering the input parameters (Table I). We note that the six-parameter fit is robust due to its super-global nature: not only are the distributions for a given process fit with a single expression (Eq. (2) for pH-drop → arrest and Eq. (3) for pH-drop → hemifusion) for all values of pH, but the fit also involves a single set of parameters unifying these two processes.

Next, we apply the theory to recent experimental data on influenza fusion [11] (Fig. 3). The stochastic errors in the data made it necessary to fix one parameter (the number of proteins in a patch, M) for a robust fit. The fixed value $M = 150$ is based on the virus-cell contact area and protein density on a virus [11]. We find that the experimental time distributions are well fit by Eq. (2) and Eqs. (4,5) under all pH conditions (Fig. 3, left and middle panels). The fit extracts the number of proteins at the arrest and hemifusion, which agree with the previously reported values $N_A = 3.4 \pm 0.2$ and $N_H = 3.7 \pm 0.6$ at $pH = 5.5$ [11], as well as the energy barrier and the protein insertion rate at physiological pH_0 (Table I). The extracted parameters at pH_0 can now be used to predict the barriers and timescales at any pH, as shown below.

TABLE I: Parameters returned from a fit of the simulated data to theory (Fig. 2) in comparison to the input parameters; and parameters extracted from the fit of experimental data [11] (Fig. 3).

	M	N_A	N_H	k_0 ($\sec^{-1}$)	$\Delta G^{\ddagger}(k_BT)$	$Q^{\ddagger}$
Simulation input	169	3	3	3×10^{-6}	10	2
Simulation fit	197±35	3.01±0.06	2.93±0.12	$(2.62\pm0.46)\times10^{-6}$	9.96±0.04	1.99±0.01
Experiment fit	fixed at 150	2.85±0.31	2.78±0.15	$(3.29\pm0.95)\times10^{-6}$	9.95±0.29	2.03±0.08

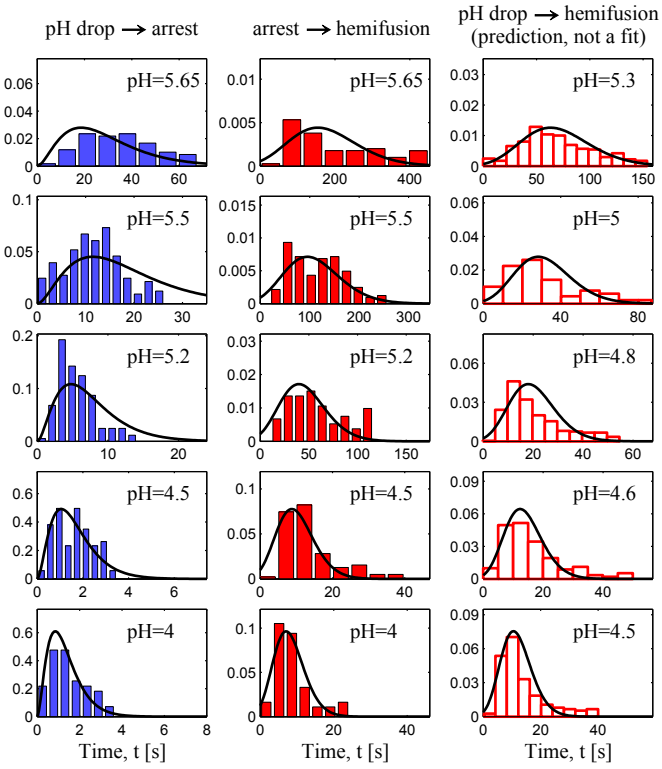

FIG. 3: Experiment *vs.* theory: time distributions for attaining different stages of influenza virus fusion under varying pH conditions. Stages of fusion are indicated above each data set. Experimental data from [11] (histograms) in the left and middle panels were fit globally to Eq. (2) and Eq. (4,5), respectively, with the same set of 5 free parameters used in both fits in the entire pH range (curves, Table I). Remarkably, with the parameters extracted from this fit, the theory is able to reproduce an independent data set [10] that was not used in the fit (Eq. (3,4), right panel).

Finally, with the parameters extracted from this fit, the theory [Eqs. (3,4)] is able to predict an independent data set [10] for the distributions of time between pH-drop and hemifusion (Fig. 3, right panel) – which is remarkable, as this data was not used in the fit.

The information extracted with the developed theory helps us quantitatively appreciate the remarkable efficiency of the viral invasion strategy (Fig. 4). In the absence of a catalyst, hemifusion of two membranes is precluded by an energy barrier as high as $100k_BT$ [21, 22]. This presents a challenge to an enveloped virus, which must fuse with the cell membrane in order to infect. The solution to the high-barrier problem, adopted by the viruses, is illustrated in Fig. 4: rather than directly conquering this $100k_BT$-high barrier, the virus first overcomes a few ($N_H \approx 3$, see Table I) small barriers in order to insert the proteins into the target membrane. Each such barrier governs the rate of one insertion as given by Eq. (8), where the expression in the exponential tells us how small each of these barriers is at a given pH: $\Delta G(pH) = \Delta G^{\ddagger}(1-\frac{2}{3}\ln 10(pH_0-pH)\frac{Q^{\ddagger}}{\Delta G^{\ddagger}})^{\frac{3}{2}}$, which gives a mere $4k_BT$ at the endosomal pH=5.5. Once these small barriers are crossed, a new source of energy becomes available to the virus from the energetically favorable protein refolding. The refolding corresponds to the downhill motion on the protein potential, releasing enough free energy ($\sim 30k_BT$ from each of the N_H proteins [7, 23]) to nearly compensate for the barrier on the membrane potential and to readily complete hemifusion as described by Eq. (3).

To further appreciate the efficient strategy of the viruses that detect the acidic environment, we make use of Eq. (6) to compare the times elapsed between the pH drop and hemifusion under two acidic conditions:

$$\frac{\bar{t}_{D\to H}(pH)}{\bar{t}_{D\to H}(pH+1)} \approx 10^{Q^{\ddagger}\left[1-\frac{2}{3}\ln 10(pH_0-pH-\frac{1}{2})\frac{Q^{\ddagger}}{\Delta G^{\ddagger}}\right]^{\frac{1}{2}}}. \quad (9)$$

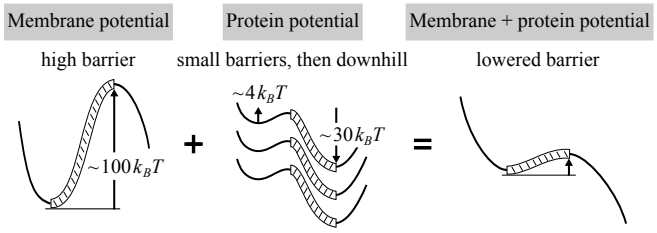

FIG. 4: Viral invasion strategy as a manipulation of the free energy barriers. *Left:* A high barrier (striped) precludes viral and cell membranes from spontaneous fusion. *Middle:* The viral strategy is to first cross the few small barriers required for protein unfolding and insertion; the subsequent protein refolding corresponds to the downhill region (striped) on the protein potential, which nearly compensates the uphill region on the membrane potential. *Right:* The compensation significantly lowers the barrier to membrane fusion.

For the endosomal pH value, $\bar{t}_{D\to H}(pH = 5.5)/\bar{t}_{D\to H}(pH = 6.5) \approx 46$, i.e. only a one-unit decrease in pH provides more than an order-of-magnitude acceleration of hemifusion.

The final stage ($H \to F$ in Eq. (1)) of the fusion process is marked by the formation of a pore (Fig. 1e) and occurs as a single-rate transition in influenza entry [10]: $p_{H\to F}(t) = k_2 e^{-k_2 t}$. The reported timescale $k_2^{-1} \approx 18s$ [10] of this step is nearly an order of magnitude faster than the timescale of hemifusion at the endosomal pH, $\bar{t}_{D\to H} \approx 120s$ (Fig. 3). Because hemifusion-to-fusion step has no considerable pH-dependence, membrane fusion studies under applied pressure [6, 24] could prove useful for extracting the parameters of this step.

The presented theory yields analytical expressions for experimentally measurable characteristics of viral entry and enables a description of the viral invasion strategy in quantitative terms. Experiments on force- and pH-induced transitions in single fusion proteins, control of the fusion kinetics via protein mutations, and high-resolution imaging of protein-membrane systems could provide independent tests of the extracted parameters. Due to the universality of principles behind the ubiquitous phenomenon of membrane fusion, the theory can be extended to other contexts, such as fusion between vesicles and cells, and between cells themselves.

We are grateful to Robert Best, Antoine van Oijen, Alex Groisman and Christopher Pierse for valuable discussions and Debarati Chatterjee for the involvement in the initial stages of this work. This research was supported by the National Science Foundation grant MCB-1411884.

[1] B. Alberts et al., Molecular Biology of the Cell, (Garland Science, New York, 2008).
[2] L.V. Chernomordik and M.M. Kozlov, Annu. Rev. Biochem. **72**, 175 (2003).
[3] D.E. Leckband, C.A. Helm, and J. Israelachvili, Biochemistry **32**, 1127 (1993).
[4] T. Kuhl, Y. Guo, J.L. Alderfer, A.D. Berman, D. Leckband, J. Israelachvili, and S.W. Hui, Langmuir **12**, 3003 (1996).
[5] S.A. Safran, T.L. Kuhl, and J.N. Israelachvili, Biophys. J. **81**, 659 (2001).
[6] M. H. Abdulreda and V. T. Moy, Biophys. J. **92**, 4369 (2007).
[7] S.C. Harrison, Nat. Struct. Mol. Biol. **15**, 690 (2008).
[8] S. Martens, T. McMahon, Nat. Rev. Mol. Cell Biol. **9**, 543 (2008).
[9] W. Bialek, Biophysics: Searching for Principles, (Princeton University Press, 2012).
[10] D.L. Floyd, J.R. Ragains, J.J. Skehel, S.C. Harrison, and A.M. van Oijen, Proc. Natl. Acad. Sci. U.S.A. **105**, 15382 (2008).
[11] T.Ivanovic, J.L. Choi, S.P. Whelan, A.M. van Oijen, S.C. Harrison, eLife 2:e00333 (2013).
[12] T.-Y. Yoon, X. Lu, J. Diao, S.-M. Lee, T. Ha, Y.-K. Shin, Nat. Struct. Mol. Biol. **15**, 707 (2008).
[13] M. Kyoung, A. Srivastava, Y. Zhang, J. Diao, M. Vrljic, P. Grob, E. Nogales, S. Chu, A.T. Brunger, Proc. Natl. Acad. Sci. U.S.A. **108**, 11737 (2011).
[14] A. Sorkin and M. von Zastrow, Nat. Rev. Mol. Cell Biol. **3**, 600 (2002).
[15] See Supplemental Material at ... for details on the derivations, simulations, and fitting procedure.
[16] Z. Luz and S. Meiboom, J. Am. Chem. Soc. **86**, 4768 (1964)
[17] A. R. Fersht, Structure and Mechanism in Protein Science, (W.H. Freeman, New York. 1999).
[18] M. Schaefer, M. Sommer, and M. Karplus, J. Phys. Chem. B **101**, 1663 (1997)
[19] H. Kramers, Physica **7**, 284 (1940).
[20] O. K.Dudko, and G. Hummer, and A. Szabo, Phys. Rev. Lett. **96**, 108101 (2006).
[21] L.V. Chernomordik, J. Zimmerberg, and M.M. Kozlov, J. Cell Biol. **175**, 201 (2006).
[22] P.I. Kuzmin, J. Zimmerberg, Y.A. Chizmadzhev, and F.S. Cohen, Proc. Natl. Acad. Sci. U.S.A. **98**, 7235 (2001).
[23] Y. Gao, S. Zorman, G. Gundersen, Z. Xi, L. Ma, G. Sirinakis, J.E. Rothman, and Y. Zhang. Science **337** 1340 (2012).
[24] A. Grafmuller, J. Shillcock, and R. Lipowsky, Phys. Rev. Lett. **98** 218101 (2007).