Stochastic Ratchet Mechanisms for Replacement of Proteins Bound to DNA

S. Cocco,[1] J. F. Marko,[2] and R. Monasson[3]
[1]Laboratoire de Physique Statistique de l'Ecole Normale Supérieure, CNRS and Université Pierre and Marie Curie, 75005 Paris, France
[2]Department of Molecular Biosciences and Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208, USA
[3]Laboratoire de Physique Théorique de l'Ecole Normale Supérieure, CNRS and Université Pierre and Marie Curie, 75005 Paris, France

(Received 20 December 2013; published 11 June 2014)

Experiments indicate that unbinding rates of proteins from DNA can depend on the concentration of proteins in nearby solution. Here we present a theory of multistep replacement of DNA-bound proteins by solution-phase proteins. For four different kinetic scenarios we calculate the dependence of protein unbinding and replacement rates on solution protein concentration. We find (1) strong effects of progressive "rezipping" of the solution-phase protein onto DNA sites liberated by "unzipping" of the originally bound protein, (2) that a model in which solution-phase proteins bind nonspecifically to DNA can describe experiments on exchanges between the nonspecific DNA-binding proteins Fis-Fis and Fis-HU, and (3) that a binding specific model describes experiments on the exchange of CueR proteins on specific binding sites.

DOI: 10.1103/PhysRevLett.112.238101 PACS numbers: 87.15.kj, 02.50.-r, 87.15.rs

Recent single-molecule experiments have revealed that exchange processes between proteins bound to DNA and proteins in solution can occur at rates in excess of spontaneous dissociation, increasing with protein solution concentration [1–4]. This effect is likely important to determining rates of turnover of proteins *in vivo*, where protein concentrations may be very high, and to understanding regulation of DNA transcription, replication, repair and packaging. Canonical models of protein-DNA interactions generally assume unbinding pathways dependent on a single rate-limiting step characteristic of the interaction of *one* protein molecule with its DNA substrate [5], and cannot explain this effect. In this paper, we introduce kinetic models aimed at describing the *multistep* dynamics of biomacromolecule interactions [6,7]. This addition leads naturally to concentration-dependent *competition* between bound and 'invading' molecules for substrate, and concentration-dependent replacement processes.

In [1] a single DNA was stretched out, and spontaneous dissociation of fluorescently labeled Fis (a bacterial chromosomal organization protein) was observed to be slow (about 90% of initially bound protein remained bound for 30 minutes). However, when nonfluorescent protein solution (either Fis or another DNA-binding protein, e.g., HU) was added, the fluorescent protein unbound rapidly (solution-phase Fis at 50 nM leads to 50% dissociation after 3 minutes). The concentration dependence is described by a replacement rate $R \simeq 6 \times 10^4\ M^{-1}\ s^{-1}$ and $R \simeq 2.7 \times 10^3\ M^{-1}\ s^{-1}$ for Fis-Fis and HU-Fis replacement. Reference [2] studied a copper-ion (Cu^+) dependent transcription factor, CueR, which exists in a Cu^+-bound ($CueR^+$) and a Cu^+-free ($CueR^-$) conformations. Both conformations compete for a specific binding site to regulate genes protecting cells from copper-induced stress. Spontaneous dissociation of $CueR^+$ at a rate $k_{off} = 0.5\ s^{-1}$ was sped up by $CueR^-$ in solution, with replacement rate $R \simeq 2.8 \times 10^7\ M^{-1}\ s^{-1}$ (Fig. 4 of [2]). In Ref. [3] it has been shown that fluorescently labeled polymerases in solution are recruited close to the replication fork, at a solution-concentration-dependent rate, and replace the DNA synthesizing polymerase at rate $k_{exc} \simeq 0.018\ s^{-1}$.

These experiments show that off-rates of proteins from DNA can depend on solution-phase concentrations of proteins competing for the same DNA. Here we describe a stochastic ratchet model of this competition. We propose that, due to thermal fluctuations, part of the bound protein releases from DNA, allowing a solution-phase protein to take its place. Rebinding of the released binding domain cannot occur because it requires thermal opening of the newly bound protein. Iterating this for a series of binding interactions allows gradual replacement. Through this process a solution-phase protein can replace a bound one far faster than if complete dissociation of the initially bound protein was required for its replacement. Here, we introduce four distinct models of kinetic pathways for protein replacement. We then compute concentration-dependent dissociation rates to determine which pathways best describe specific experiments.

Kinetic schemes for the four models are shown in Fig. 1. We assume that binding of proteins to the DNA occurs by at most N units ($N = 3$ in Fig. 1). Units bind to putative anchoring sites; values of N and bond energy ϵ will be discussed below. Each model contains two unbinding pathways: a spontaneous unbinding pathway where the

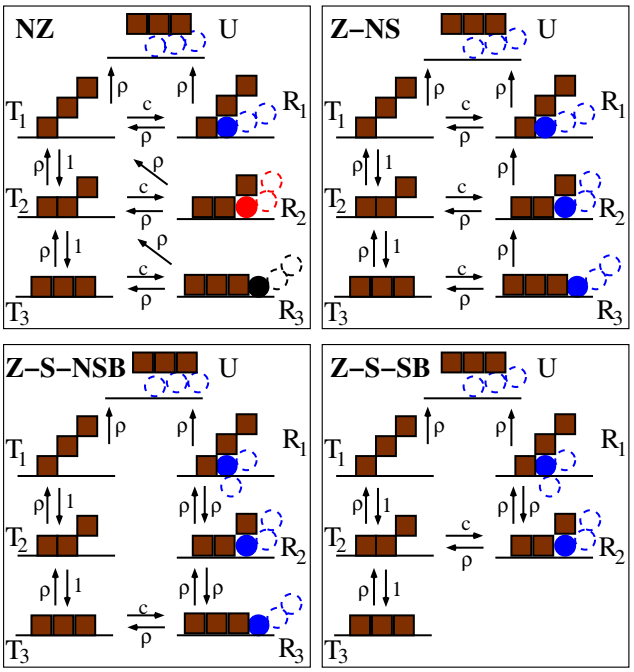

FIG. 1 (color online). The four proposed unbinding pathways. Brown squares show N (= 3 here) units of a protein bound to DNA (dark horizontal line). Circles show units of the invader proteins, with different colors corresponding to different proteins. Filled circles show units occupying the zipping site. The most likely replacement scenario at small concentration is shown with the blue invader protein. Parameters entering the rates are (i) the mean number c of solution-phase proteins per binding site, in units of the elementary concentration $c_o = 1/a^3$, of one particle per binding site, where a is a length scale associated with the linear dimension of a binding site (for $a = 1$ nm, $c_o = 1$ M) and (ii) the ratio of the unbinding and binding rates for one unit: $\rho = e^{-\epsilon}$, where ϵ is the binding energy in k_BT units. Time is expressed in terms of the time scale t_o, equal to the self-diffusion time for one unit of the protein: $t_o = 2\pi\eta a^3/k_BT \approx 1.6 \times 10^{-9}$ s, for $a \approx 1$ nm, $\eta = 0.001$ Pa s and $k_BT = 4 \times 10^{-21}$ J. In units of $1/t_0$ the zipping rate of a protein unit on a free binding site is equal to one.

units of the bound protein dissociate one after the other with rate $\rho = e^{-\epsilon}$ (states $T_1, \ldots, T_N$ to the left of each box in Fig. 1) and a replacement pathway (right) where the bound protein is replaced by invading proteins (states $R_1, \ldots, R_N$). In the nonspecific (NS) pathways, the invading protein may bind on each site left unoccupied by the bound protein, while in the specific (S) pathways, binding occurs only from one extremity of the bound protein. In the zipping models (Z) the invader, once attached to DNA, binds adjacent sites successively (zips) as they are released by the bound protein. In the nonzipping (NZ) model (top left panel, Fig. 1), when one bound unit of the protein detaches, this site is left unoccupied, corresponding to transitions $R_i \to T_{i-1}$. By contrast, in the Z models, as soon as the bound unit bound detaches, the invading protein occupies the vacant site, making transitions $R_i \to R_{i-1}$ possible.

For the NZ and zipping nonspecific (Z-NS) models (top panels of Fig. 1), the invading proteins can bind with rate c proportional to the concentration of solution proteins (transitions $T_i \to R_i$), and unbind with rate ρ (transitions $R_i \to T_i$) on each possible anchoring site, i.e., whatever the value of i. We consider two zipping specific (Z-S) models (bottom panels of Fig. 1). In the nonspecific binding (Z-S-NSB) case, the invader may bind next to the fully DNA-bound protein (state R_N), which represents a nonspecific nearby site. For the specific binding (Z-S-SB) case, the invader can only bind specifically to the first available site after one unit of the bound protein detaches (state R_{N-1}), after which zipping of the invader may then proceed.

The probability $P_S(t)$ that the system is in state S at time t is described by the master equation

$$\frac{dP_S(t)}{dt} = \sum_{S'} W(S \leftarrow S') P_{S'}(t). \tag{1}$$

The system is initially in state T_N, corresponding to a fully-bound protein, with no invader present, and eventually reaches the unbound state U (Fig. 1). Off-diagonal, nonzero elements of the transition matrix W are given for the four models in Fig. 1 and its caption (see also Supplemental Material [8]). Diagonal elements conserve probability, $W(S \leftarrow S) = -\sum_{S' (\neq S)} W(S' \leftarrow S)$.

Given the W matrix, the average occupancy time, or equivalently the inverse of the binding rate $r(c)$ of the protein, is

$$\frac{1}{r(c)} = \int_0^\infty dt \sum_{S \neq U} P_S(t) = -\langle B | W^{-1} | T_N \rangle, \tag{2}$$

where W^{-1} is the inverse matrix of W, and $|B\rangle$ denotes the sum of all $2N$ bound states $|T_i\rangle$ and $|R_i\rangle$, with $1 \le i \le N$. The unbinding rate $r(c)$ is plotted as a function of c in Fig. 2 for the four models of Fig. 1.

Without solution-phase protein ($c = 0$) the unbinding rate is $r(0) = (1-\rho)^2 \rho^N$ (up to $O(\rho^{2N})$ corrections) for all four models: pure thermal unbinding is exponentially slow in N. The unbinding rate at small concentration $c > 0$ can be studied perturbatively. Using the linear dependence of W on c we write $W = W_o + cW_1$, where spontaneous dissociation is described by W_o and invasion-zipping is described by W_1. We have $W^{-1}(c) = W_o^{-1} - cW_o^{-1}W_1W_o^{-1} + 0(c^2)$. The mean unbinding rate is therefore approximately

$$r(c) \simeq r(0) + Rc + O(c^2), \tag{3}$$

where the replacement rate R is

$$R \equiv \left.\frac{dr}{dc}\right|_{c=0} = -r(0)^2 \langle B | W_o^{-1} W_1 W_o^{-1} | T_N \rangle. \tag{4}$$

We define the replacement concentration c_R as the concentration at which the unbinding rate is twice its zero-concentration value,

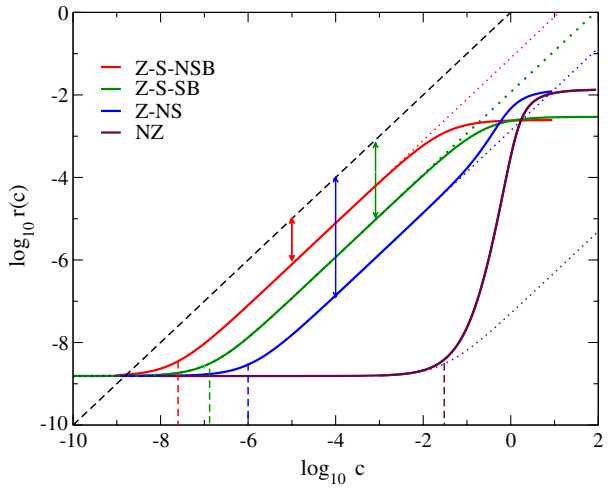

FIG. 2 (color online). Unbinding rates $r(c)$ (full lines) from Eq. (2) versus dimensionless concentration c and for parameters $N = 10$, $\epsilon = 2$. The zero-concentration rate, $r(0) = 1.8 \times 10^{-9}$ (units of $1/t_0$), is the same for all four models. The concentrations c_R at which replacement starts to dominate over pure thermal unbinding, i.e., $r(c)$ starts to vary linearly with c, are indicated by the vertical dashed colored lines for the four models; $c_R = 3 \times 10^{-2}$ for NZ, 2×10^{-6} for Z-NS, 2×10^{-8} for Z-S-NSB, and 1.3×10^{-7} for Z-S-SB (units of c_0). The offsets between the linear regimes of the rate curves [dotted lines, from Eq. (3)] and the $r = c$ (dashed black) line are $\log R$ (log of replacement rate, double arrow vertical lines); R is approximatively equal to 10^{-1} for Z-S-NSB, 10^{-2} for Z-S-SB and 10^{-3} for Z-NS. The dotted lines are only visible when the linear approximation breaks down.

$$c_R = \frac{r(0)}{R}. \tag{5}$$

The unbinding rate increase at small concentration, Rc, is the rate at which unbinding-replacement involves essentially one invading protein. The most likely unbinding scenario is indicated by the sequence of blue invader configurations in Fig. 1, providing an approximation sufficient to understand the scaling of the replacement rate R with the number of binding units, N (see Supplemental Material [8]).

For the NZ model (brown curve in Fig. 2) we find a replacement rate, Eq. (4), of

$$R^{NZ} = \rho^{N-1} \frac{(N - 1 - N\rho - \rho)(1 - \rho)}{2} + O(\rho^{2N-1}). \tag{6}$$

The most probable unbinding scenario with replacement at small c is that, from state T_1, occupied with probability ρ^{N-1}, the invader binds with rate c (transition $T_1 \to R_1$ in Fig. 1, top and left). The protein is then equally likely to dissociate ($R_1 \to U$) or to come back to the thermal pathway ($R_1 \to T_1$). We therefore obtain $R^{NZ} \propto \rho^{N-1}$, in agreement with (6). As R^{NZ} is exponentially small in N, the replacement concentration $c_R^{NZ} \sim 2\rho/N$ can become large; for a binding energy $\epsilon = 2k_BT$, $N = 10$ protein units, and $c_0 = 1\ M$, we find $c_R^{NZ} = 3 \times 10^{-2}\ M$ (Fig. 2), well above that experimentally observed (tens of nM [1,2]). The concentration range where the linear approximation, Eq. (3), holds is very narrow. Contrary to experiments, the unbinding rate of the NZ model shows a highly nonlinear concentration dependence, $r(c) \sim c^N$ for $c \gtrsim c_R$ (Fig. 2).

For the Z-NS model (blue curve in Fig. 2) the replacement rate,

$$R^{Z\text{-}NS} = \frac{(1 - \rho)^2}{(1 - 2\rho)^2 2^N} + O(\rho^{N-2}), \tag{7}$$

is also exponentially small in N, but decays less quickly with N than in the NZ model, as ρ is generally smaller than $\frac{1}{2}$. In the most likely unbinding scenario (blue configurations, Fig. 1, top and right), the invader attaches through the transition $T_N \to R_N$ from the frequently occupied, fully bound state T_N. The scenario follows the replacement pathway all the way up with probability 2^{-N}, until the protein is released and gives the scaling $R^{NZ} \propto 2^{-N}$ (7). Figure 2 shows that the linear approximation (3,7) describes the exact unbinding rate $r(c)$ (blue curve; see Supplemental Material [8] for results obtained for varied N and ϵ). At large concentrations the unbinding rate grows $\approx c^N$ as in the NZ model, since invading proteins can attach and attempt replacement at every site.

Figure 3 compares the Z-NS model to experiment for Fis-Fis and Fis-HU replacement. We fit N and ϵ (see Fig. 3

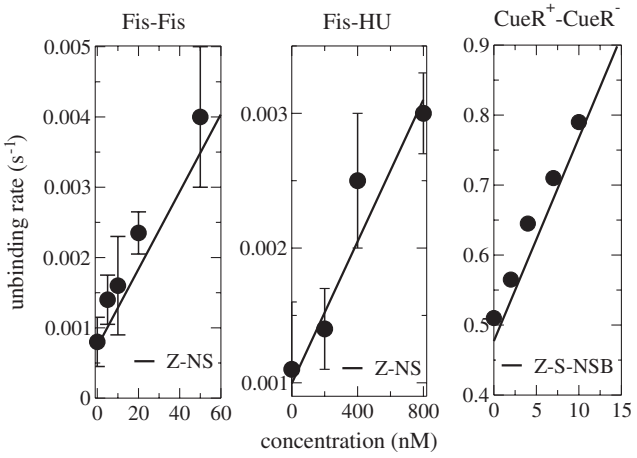

FIG. 3. Fit of concentration-dependent unbinding rates of Fis bound to DNA in the presence of Fis (left) and HU (middle) proteins in solution [1], and for CueR dissociation rates as a function of CueR concentration in solution (right) [2], using $a = 1$ nm [9], and N = 14, $\epsilon = 1.95$ for Fis-Fis, N = 19, $\epsilon = 1.4$ for Fis-HU, and N = 15, $\epsilon = 1.36$ for CueR-CueR. For replacement rates (slopes of unbinding rates versus concentration) we find $R^{\text{Fis-Fis}} = 5 \times 10^4\ M^{-1}\,\text{s}^{-1}$, $R^{\text{Fis-}HU} = 2.6 \times 10^3\ M^{-1}\,\text{s}^{-1}$, and $R^{\text{CueR-CueR}} = 2.9 \times 10^7\ M^{-1}\,\text{s}^{-1}$, in agreement, considering the error bars, with experimental fits. Replacement concentrations found with the Z-NS model are $c_R = 2$ nM for Fis-Fis and $c_R = 370$ nM for Fis-HU, while we find $c_R = 16$ nM for CueR-CueR with the Z-S-NSB model.

caption) by matching experimental exchange and spontaneous unbinding rates to R^{Z-NS} in Eq. (7) and $r(0)$. Fis-Fis replacement dominates thermal unbinding as soon as $c \gtrsim c_R \simeq (2\rho)^N \simeq 2$ nM, a value of concentration compatible with experimental observation [9]. Heterotypic Fis-HU replacement dominates at concentrations of hundreds of nM [1], in agreement with experiment. In both cases the binding energy ϵ is of the order of $\simeq 2k_BT$. However, for CueR$^+$-CueR$^-$ data [2] we obtain a binding energy per site of $\epsilon \simeq 5k_BT$, which is too large even for specific binding. The Z-NS model can reasonably describe Fis-Fis and Fis-HU exchange, but not CueR$^+$-CueR$^-$ exchange.

For DNA polymerase [3] the observed exchange rate following recruitment is very small, $\approx 2 \times 10^{-2}$ s^{-1}, with no concentration dependence since the polymerase is recruited and exchanged in separate kinetic steps. However, we can still proceed by setting the dimensionless concentration to $c = 1$. Using the Z-NS model with $a = 1$ nm we have $t_o = 1.6 \times 10^{-9}$ s which gives $N = 35$. This large N value is consistent with the large DNA-binding surface of DNApol.

The replacement rate for the Z-S-NSB model is

$$R^{Z-S-NSB} = \frac{1-\rho}{N+1} + O(\rho^{N-1}). \tag{8}$$

It decreases only algebraically with N and is much larger than its Z-NS counterpart which scales as 2^{-N}. In the most probable replacement pathway the invader attaches at the last zipping site (T_N in Fig. 1, bottom and left) with rate c. The probability that the system continues along the replacement pathway until the bound protein is released, and never reaches T_N again, scales as $1/(N+1)$, giving the $R^{Z-S-NSB}$ scaling in (8). The linear approximation for $r(c)$ (3, 8) is valid over a large range of concentration (red curve in Fig. 2). The Z-S-NSB model allows us to fit the replacement rate of the CueR$^+$-CueR$^-$ experiment, which is about 1000 times larger than the one observed for Fis-Fis replacement, and the corresponding replacement concentration $c_R \simeq (N+1)\rho^N$, with reasonable parameter values $N = 13$–16 and $\epsilon = 1.2$–1.4, giving a total binding energy of the order of 20 k_BT (Fig. 3). The Z-NS model is inappropriate to describe CueR$^+$-CueR$^-$ replacement, as it requires $N \sim 10^4$ to generate the observed exchange rate. For the Z-S-SB model the replacement rate reads

$$R^{Z-S-SB} = \frac{\rho(1-\rho)}{N} + O(\rho^N), \tag{9}$$

with replacement concentration $c_R \simeq N\rho^{N-1}$. The scaling with N is similar to that of Z-S-NSB, with $1/N$ instead of $1/(N+1)$ due to the shorter replacement pathway (Fig. 1). The Z-S-SB model does not reproduce the CueR$^+$-CueR$^-$ replacement rate. Indeed the replacement rate is smaller by a factor $1/\rho = e^{+\epsilon} \simeq 3$–$4$ with respect to the rate obtained with the Z-S-NSB model for $\epsilon = 1$–$1.4k_BT$ and N ranging from 13 to 16 as in Fig. 2 (green curve).

In conclusion we have introduced four kinetic models to describe replacement processes between DNA-bound proteins. We have solved the models at three levels: numerically, using a small–concentration expansion providing analytical formulae for the replacement rate (R_c), and with a "dominant pathway" approximation, which gives the correct scaling of R_c with N. The importance of zipping for efficient bound-invader exchange is illustrated by the inability of the NZ model to replace bound proteins at low concentrations. The Z-NS model is appropriate to describe Fis-Fis and Fis-HU exchanges, presumably due to the nonspecific nature of Fis- and HU-DNA interactions [1]. Moreover the Z-NS model is appropriate to describe polymerase replacements, since DNA-pol is able to interact with any DNA sequence. Z-S-NSB better describes CueR-CueR exchange, which is sensible since CueR interacts with a specific binding site [2]. Fits for N range from 14 to 35 depending on the protein, and the fit values of binding energy per unit are $\approx \epsilon = 1 - 2k_BT$, consistent with individual noncovalent interactions. DNA-binding proteins interact in complex ways with their substrate [10], and the number N of units used here can be thought of as an effective number of separate bonds. For the heterogeneous Fis-HU replacement the fit value of ϵ is smaller than for homogeneous Fis-Fis replacement; this is consistent with the larger DNA binding affinity of Fis compared to HU [1]. A more general (and precise) model would describe the invading protein through a binding energy ϵ' different from the binding energy ϵ for the bound protein in case of heterogeneous replacement, as well through its number of bound units, rather than the on/off description used here. Binding and unbinding of small DNA fragments (oligonucleotide) on a DNA under force [11] and exchange of DNA-binding oligonucleotides in DNA hybridization assays [12–14] are likely described by the Z-S-SB model. Including sequence specificity (dependence of ϵ on the sites) could help in modeling such experiments [15].

While revising this Letter, two papers have appeared providing further evidence for protein exchange [16,17], suggesting the generality of the replacement process. Luo and collaborators [16] have seen displacement of transcription factors by nucleosomes on DNA; see [18] for a related theory. Gibb and collaborators [17] have demonstrated exchange of replication protein A (RPA) and the recombinase Rad51 and exchange of RPA with single-stranded-binding (SSB) protein on single-stranded DNA; the kinetic scheme in Fig. 6 of [17] corresponds to our Z-NS model.

We thanks V. Croquette for useful discussions. J. F. M. acknowledges NSF Grants No. MCB-1022117 and No. DMR-1206868, and NIH Grant No. 1U54CA143869-01 (NU-PS-OC).

[1] J. S. Graham, R. C. Johnson, and J. F. Marko, Nucleic Acids Res.. **39**, 2249 (2011).
[2] C. P. Joshi, D. Panda, D. J. Martell, N. M. Andoy, T. Y. Chen, A Gaballa, J. D. Helmann, and P. Chen, Proc. Natl. Acad. Sci. U.S.A. **109**, 15121 (2012).
[3] J. J. Loparo, A. W. Kulczyk, C. C. Richardson, and A. M. van Oijen, Proc. Natl. Acad. Sci. U.S.A.. **108**, 3584 (2011).
[4] T. Ha, Cell **154**, 723 (2013).
[5] K. A. Dill and S. Bromberg, *Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience* (Garland Science, New York, 2002).
[6] C. Kittel, Am. J. Phys. **37**, 917 (1969).
[7] S. Cocco, R. Monasson, and J. F. Marko, Eur. Phys. J. E **10**, 153 (2003).
[8] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.112.238101 for additional information on the rate matrix W, the dominant-pathway approximation for R, the dependence of the unbinding rate on N and ϵ, and the fit procedure.
[9] The value of the elementary length a is not crucial, as R and c_R do not depend, at the leading order, on a; see Supplemental Material [8].
[10] R. Rohs, X. Jin, S. M. West, R. Joshi, B. Honig, and R. S. Mann, Annu. Rev. Biochem.. **79**, 233 (2010).
[11] F. Ding, M. Manosas, M. M. Spiering, S. J. Benkovic, D. Bensimon, J. F. Allemand, and V. Croquette, Nat. Methods **9**, 367 (2012).
[12] J.-C. Walter, K. M. Kroll, J. Hooyberghs, and E. Carlon, J. Phys. Chem. **115** (20), 6732 (2011).
[13] C. Trapp, M. Schenkelberger, and A. Ott, BMC Biophysics **4**, 20 (2011); A. Ott (private communication).
[14] N. F. Dupuis, E. D. Holmstrom, and D. J. Nesbitt, Biophys. J. **105**, 756 (2013).
[15] F. Ding, J. Ouellet, H. Gouet, J. F. Allemand, D. Bensimon, S. Cocco, M. Manosas, and V. Croquette (to be published).
[16] Y. Luo, J. A. North, S. D. Rose, and M. G. Poirier, Nucleic Acids Res. **42**, 3017 (2014).
[17] B. Gibb, L. F. Ye, S. C. Gergoudis, Y. H. Kwon, H. Niu, P. Sung, and E. C. Greene, PLoS One **9**, e87922 (2014).
[18] C. Chen and R. Bundschuh (to be published).