Supplemental Material

Stochastic Ratchet Mechanisms for Replacement of Proteins Bound to DNA

S. Cocco, J.F. Marko, R. Monasson

I. TRANSITION MATRICES FOR THE DIFFERENT MODELS

We show in Table I the transition matrices W for the four models of Fig. 1 in the main text. States are numbered from 1 to $2N$ in the following way (the unbound state U is omitted as it is an absorbing state). The initially bound protein has N possible bound states T_i, where i is the number of attached units of the protein to the DNA. The 'invading' protein has, for each state of the bound protein, only two possible configurations: on or off ($j = 0, 1$), at the site corresponding to where the $(i + 1)$th unit of the bound protein would be, which we call the 'zipping' site. There are thus $2N$ states, labelled by the index $S = 2\,i + j - 1$, for $i = 1, \ldots . N$, $j = 0, 1$. The transition matrices W have therefore $2N$ lines and columns. Empty squares correspond to zero entries. For $N > 3$ the boundaries (two first and two last lines and columns) are the same, and the 2×2 central blocks are repeated $N - 2$ times on the band diagonal.

$W^{NZ} =$

-ρ-c-1	ρ	ρ	ρ		
c	-2ρ				
1		-ρ-c-1	ρ	ρ	ρ
		c	-2ρ		
		1		-ρ-c	ρ
				c	-2ρ

,

$W^{Z-NS} =$

-ρ-c-1	ρ	ρ			
c	-2ρ		ρ		
1		-ρ-c-1	ρ	ρ	
		c	-2ρ		ρ
		1		-ρ-c	ρ
				c	-2ρ

,

$W^{Z-S-NSB} =$

-ρ-1		ρ			
	-2ρ		ρ		
1		-ρ-1		ρ	
	ρ		-2ρ		ρ
		1		-ρ-c	ρ
			ρ	c	-2ρ

,

$W^{Z-S-SB} =$

-ρ-1		ρ			
	-2ρ		ρ		
1		-ρ-1-c	ρ	ρ	
	ρ	c	-2ρ		
		1		-ρ	1
					-1

.

TABLE I: Transition matrices W for the NZ (top, left), Z-NS (top, right), Z-S-NSB (bottom, left), and Z-S-SB (bottom, right) models with $N = 3$ binding units. All columns sum up to zero due to probability conservation, except the first two columns as states 1 (T_1) and 2 (R_1) may decay into the unbound state U, not represented in the matrices above.

II. UNBINDING RATES VS. CONCENTRATION FOR DIFFERENT ϵ AND N

In Fig. S1, Fig. S2, and Fig. S3 we show the unbinding rate r as a function of the concentration c for the Z-NS, Z-S-NSB and Z-S-SB models, for $\epsilon = 2, 3$ and different sizes N.

III. FIT OF MODEL PARAMETERS FROM EXPERIMENTAL DATA

To fit the Fis-Fis and CueR-CueR data we have proceeded as follows. We have first considered three possible elementary length $a = 0.5$, 1 and 2 nm. In Table II we give the elementary time t_0 and the unit of concentrations c_0 obtained in the three cases. The experiments give us access to the values of three quantities: the spontaneous dissociation rate $r(0)$, the order of magnitude of the concentration c_R at which replacement dominates over spontaneous dissociation, and the exchange rate R. We have first fitted the value N of the number of units for the Z-NS and Z-S models from the leading contribution to R (formulas given in main text), and we have then adjusted N and ϵ to fit the

a (nm)	0.5	1	5
t_0 (s)	$2\ 10^{-10}$	$1.6\ 10^{-9}$	$2\ 10^{-7}$
c_0 (M)	8	1	0.008

TABLE II: Values of the time constant t_0 and of the unit of concentration c_0 for three different elementary lengths a.

spontaneous dissociation. The order of magnitude of c_R, given the fitted N and ϵ allows us to check the consistency of the model.

Interestingly we have found that the choice of a does not change the values, at the leading order, of R and c_R. Indeed at the leading order they both depend on the product $c_0 \times t_0$, which is independent of the value of a. More precisely, using the experimental values for R^{Fis} and R^{CueR}, we have found:

$$N_{Fis}^{Z-NS} \approx -\log_2 (c_0\, t_0\, R_{Fis}) = 13.3 \ , \tag{1}$$

and

$$N_{CueR}^{Z-S-NSB} \approx (c_o\, t_o\, R_{CueR})^{-1} - 1 = 21 \ . \tag{2}$$

Given the approximate values of N above the concentration at which replacement start to dominates is given by

$$(c_R)_{Fis}^{Z-NS} \approx c_0\, t_0\, r_{Fis}(0)\ 2^N \approx 1\ nM \ , \tag{3}$$

and

$$(c_R)_{CueR}^{Z-S-NSB} \approx c_0\, t_0\, r_{CueR}(0)\ (N+1) \approx\ 17\ nM \ . \tag{4}$$

Then we have used the spontaneous dissociation rate $r(0)$ to estimate, given N, the order of magnitude of the binding evergy ϵ through

$$\epsilon \approx -\log(t_0\ r(0))/N \ , \tag{5}$$

for both Fis and CueR data. Note that the Z-NS model is not compatible with CueR data because we would obtain $N_{CueR}^{Z-NS} = 4$, which would corresponds to large binding energies of $\epsilon = 5.7,\ 5.2,\ 4\ k_B\,T$, for, respectively, $a = 0.5,\ 1,\ 5$ nm. Conversely, Z-S-NSB is not compatible with Fis data because we would obtain $N_{Fis}^{Z-S-NSB} = 10,000$.

We have then tuned the value of $N_{CueR}^{Z-S-NSB}$ and N_{Fis}^{Z-NS} to precisely fit R when the whole expression of the exchange rate is taken into account (see Eqs. (8) and (9) of main text), in particular the multiplicative factors in $1-\rho = 1-e^{-\epsilon}$. The final value for $N_{CueR}^{Z-S-NSB}$ ranges from 13 to 16 depending on the value of a (see Fig. 3 of the main text). There is less variability in the value of N_{Fis}^{Z-NS} , due to the exponential dependence of R upon N. The best fit is obtained in the three cases for $N = 14$; the best value for a is $a = 5$ nm. Finally, once N is fixed, a fine tuning of ϵ is easily done to reproduce the spontaneous dissociation rate.

We show in Fig. S4 the fits of the Fis-Fis, Fis-Hu, and CueR-CuerR experiments presented and cited in the main text. In each panel three fits are presented, corresponding to the three values of the elementary length $a = 0.5, 1, 5$ nm. The values of N and ϵ are given in the panels. The slope of the curves are the exchange constants R:

- The experimental Fis-Fis exchange constant $R_{Fis-Fis} = 6\,10^4 \pm 3\,10^3\ \mathrm{M^{-1}\,s^{-1}}$ has to be compared with $R_{Fis-Fis}^{Z-NS} = 4.9\,10^4, 5.1\,10^4, 6\,10^4\ \mathrm{M^{-1}\,s^{-1}}$ for, respectively, $a = 0.5, 1, 5$ nm. The replacement concentration is $(c_R)_{Fis}^{Z-NS} = 2, 2, 1.6$ nM for $a = 0.5, 1, 5$ nm.

- The experimental Fis-Hu exchange constant $R_{Fis-Hu} = 2.7\,10^3 \pm 5\,10^2\ \mathrm{M^{-1}\,s^{-1}}$ has to be compared with $R_{Fis-Hu}^{Z-NS} = 2.3\,10^3, 2.6\,10^3, 4\,10^3\ \mathrm{M^{-1}\,s^{-1}}$ for, respectively, $a = 0.5, 1, 5$ nm. The replacement concentration is $(c_R)_{Fis-Hu}^{Z-NS} = 400, 370, 180$ nM for $a = 0.5, 1, 5$ nm.

- The experimental CueR-CueR exchange constant is $R_{CueR-CueR} = 2.8\,10^7\ \mathrm{M^{-1}\,s^{-1}}$ has to be compared with $R_{CueR-CueR}^{Z-S-NSB} = 2.8\,10^7, 2.9\,10^7, 3.1\,10^7\ \mathrm{M^{-1}\,s^{-1}}$ for, respectively, $a = 0.5, 1, 5$ nm. We obtain the replacement concentration $(c_R)_{CueR-CueR} = 19, 16, 17$ nM for $a = 0.5, 1, 5$ nm.

IV. SCALING OF THE REPLACEMENT RATE R WITH N

We start by writing the average unbinding time as

$$t_{unb}(c) = (1 - P_R)\, t_{unb}(0) + P_R\, t_{unb-R}\ , \tag{6}$$

where P_R is the probability that unbinding occurs (at least partially) through the replacement pathway, $t_{unb}(0) = 1/r(0)$ is the average unbinding time through the thermal pathway only, and t_{unb-R} is the average unbinding time through the replacement pathway.

In the limit of small solution-phase protein concentrations, *i.e.* for small rates c, we expect P_R to scale linearly with c, $P_R \simeq p_R\, c$, and, thus, that $P_R \ll 1$. In addition, as unbinding through replacement is much faster than through thermal activation alone, we have $t_{unb-R} \ll t_{unb}(0)$. Taking the inverse of (6) we obtain

$$r(c) = \frac{1}{t_{unb}(c)} \simeq r(0) + p_R\, c\, r(0)\ . \tag{7}$$

Comparing with Eq. (3) in the main text we see that the replacement rate R coincides with

$$R = p_R\, r(0)\ . \tag{8}$$

To calculate p_R, which corresponds to the linear term in the expansion of P_R in powers of c, we consider 'replacement' paths with only one binding event for the invader. The scaling behavior of p_R with N can be simply guessed from the most likely scenario leading to dissociation through the replacement pathway, see main text and blue configurations sketched in Fig. 1 therein. Below, we consider all contributions to p_R to find back the exact expression for R given in Eqs. (7), (8) and (9) in the main text for the three Z- kinetic models; the unbinding time for the NZ model at small c is not significantly larger than the thermal rate, and the asymptotic dependence of the replacement rate with N is easy to determine, see main text.

We will use the following result, true for the thermal pathway alone ($c = 0$): the average time spent in state T_i (before the protein eventually unbinds) is

$$\tau_i = (1 - \rho)\, \rho^{N-i} \left(1 + O(\rho^N)\right)/r(0)\ , \tag{9}$$

where $r(0) = (1 - \rho)^2 \rho^N$ to the leading order in powers of ρ, see main text. The sum of τ_i over all states $i = 1, ..., N$ coincides with our expression for the thermal unbinding time, $t_{unb}(0) = 1/r(0)$.

A. Z-NS model

One possible scenario of unbinding due to replacement, considered in the main text, consists in reaching state R_N from state T_N. As the average time spent in state T_N is $\tau_N = (1-\rho)/r(0)$ the probability that the invader eventually binds is, to the lowest order in c, $\simeq c\,\tau_N$. Next, the probability to stay and go all the way up along the alternative replacement pathway is 2^{-N}, because each transition $\mathrm{R}_i \to \mathrm{R}_{i-1}$ has the same rate as the transition $\mathrm{R}_i \to \mathrm{T}_i$ (invader detachment). The contribution of this scenario to the replacement rate is

$$\tau_N\ 2^{-N}\ r(0) = (1 - \rho)\ 2^{-N} \tag{10}$$

Other replacement scenarios, with a single invader-binding event, are as follows: the system spends time in the thermal pathway, undergoes the transition $\mathrm{T}_k \to \mathrm{R}_k$ at some level (number of bound units) k, remains in the replacement pathway until level $\ell (\leq k)$, and goes back to the thermal pathway through $\mathrm{T}_\ell \to \mathrm{R}_\ell$, until thermal binding occurs (the scenario above corresponds to the case $k = N$, without back transition to the thermal pathway). The probability of climbing (without leaving) the replacement pathway from T_k to T_ℓ is $2^{-(k-\ell)}$.

We then need to calculate the probability q_ℓ that the system reaches back the thermal pathway in state T_ℓ. Let us start by estimating the probability μ_ℓ that, once in state ℓ, the protein will unbind quickly, *i.e.* in a time much lower than the average time t_{unb} in the thermal pathway. The following physical picture is useful to compute μ_ℓ. A particle is undergoing a biased random random walk on discrete sites $i \geq 0$ along the 1D semi-infinite line; $i = 0$ is an absorbing site. The random walk is biased by a force F pushing the particle towards the right (large i), with $F = -\log \rho$. μ_ℓ is the probability that, starting from state $i = \ell$, the particle with reach the absorbing state, rather than being attracted towards $i \to \infty$. We obviously have

$$\mu_\ell = \frac{1}{1+\rho}\, \mu_{\ell+1} + \frac{\rho}{1+\rho}\, \mu_{\ell-1}\ , \quad \forall \ell = 2, \ldots, N\ , \tag{11}$$

with the boundary conditions $\mu_0 = 1$ and $\mu_\infty = 0$, which gives

$$\mu_\ell = \rho^\ell \ , \tag{12}$$

for all $\ell \geq 0$. Formally speaking, μ_ℓ is the probability that the system reaches the U state in a finite time (when $N \to \infty$), conditioned to the starting state, T_ℓ. The probability q_ℓ that the system reaches back the thermal pathway in state T_ℓ conditioned to fast unbinding is, according to Bayes' rule,

$$q_\ell = \frac{\mu_\ell}{\sum_{i\geq 0} \mu_i} = (1-\rho)\,\rho^\ell \ . \tag{13}$$

The sum of the contributions corresponding to all pairs (k, ℓ) gives the replacement rate

$$R^{Z-NS} = \sum_{\ell \leq k \leq N} \tau_k \, 2^{-(\ell-k)} \, q_\ell \, r(0) = (1-\rho)^2 \sum_{\ell \leq k \leq N} \rho^{N-k+\ell} \, 2^{-(k-\ell)} \tag{14}$$

$$= (1-\rho)^2 \sum_{0 \leq m \leq N} (N+1-m)\,\rho^{N-m}\, 2^{-m} = \frac{(1-\rho)^2}{(1-2\,\rho)^2\, 2^N} \ , \tag{15}$$

to the leading order in N, in agreement with Eq. (7) in the main text.

B. Z-S-NSB model

In the Z-S-NSB model the alternative scenario to thermal unbinding starts with a transition $\mathrm{T}_N \to \mathrm{R}_N$. Later the system changes states in the replacement pathway, until unbinding occurs. We need to estimate the probability Q that this scenario occurs, rather than the systems reaches back state T_N from R_N, and unbinding occurs through the thermal pathway. Let $\mu_{i,N}$, with $1 \leq i \leq N$ the probability that the system never leaves the replacement pathway (and unbinding eventually occurs), starting from state T_i; we want to calculate $Q = \mu_{N,N}$. According to Fig. 1 in the main text (bottom & left panel), the rates of the forward ($\mathrm{R}_i \to \mathrm{R}_{i-1}$) and backward ($\mathrm{R}_i \to \mathrm{R}_{i+1}$) transitions are identical (away from the boundaries). Hence the $\mu_{i,N}$s fulfill the recurrence equations

$$\mu_{i,N} = \frac{1}{2}(\mu_{i-1,N} + \mu_{i+1,N}) \ , \quad \forall i = 2, \ldots, N \ . \tag{16}$$

The above equation can be extended to the states $i = N$ and $i = 1$, with the prescriptions $\mu_{N+1,N} \equiv 0$ (which expresses the fact that the system cannot re-enter the replacement pathway after having left it), and $\mu_{0,N} \equiv 1$ (where the subscript 0 stand for the unbound state U here). The solution to these equations is

$$\mu_{i,N} = \frac{N+1-i}{N+1} \ . \tag{17}$$

We deduce that the replacement rate is given by

$$R^{Z-S-NSB} = \tau_N \, \mu_{N,N} \, r(0) = \frac{1-\rho}{N+1} \ , \tag{18}$$

in perfect agreement with Eq. (8) in the main text.

C. Z-S-SB model

The replacement rate for the Z-S-SB model is calculated in the same way as for the Z-S-NSB model. The only differences are: the replacement pathway is entered through the $\mathrm{T}_{N-1} \to \mathrm{R}_{N-1}$ transition, and the 'length' (number of states) of this replacement pathway is now $N-1$ instead of N. We readily obtain, using (9) and (17),

$$R^{Z-S-SB} = \tau_{N-1} \, \mu_{N-1,N-1} \, r(0) = \frac{\rho(1-\rho)}{N} \ , \tag{19}$$

in agreement with Eq. (9) in the main text.

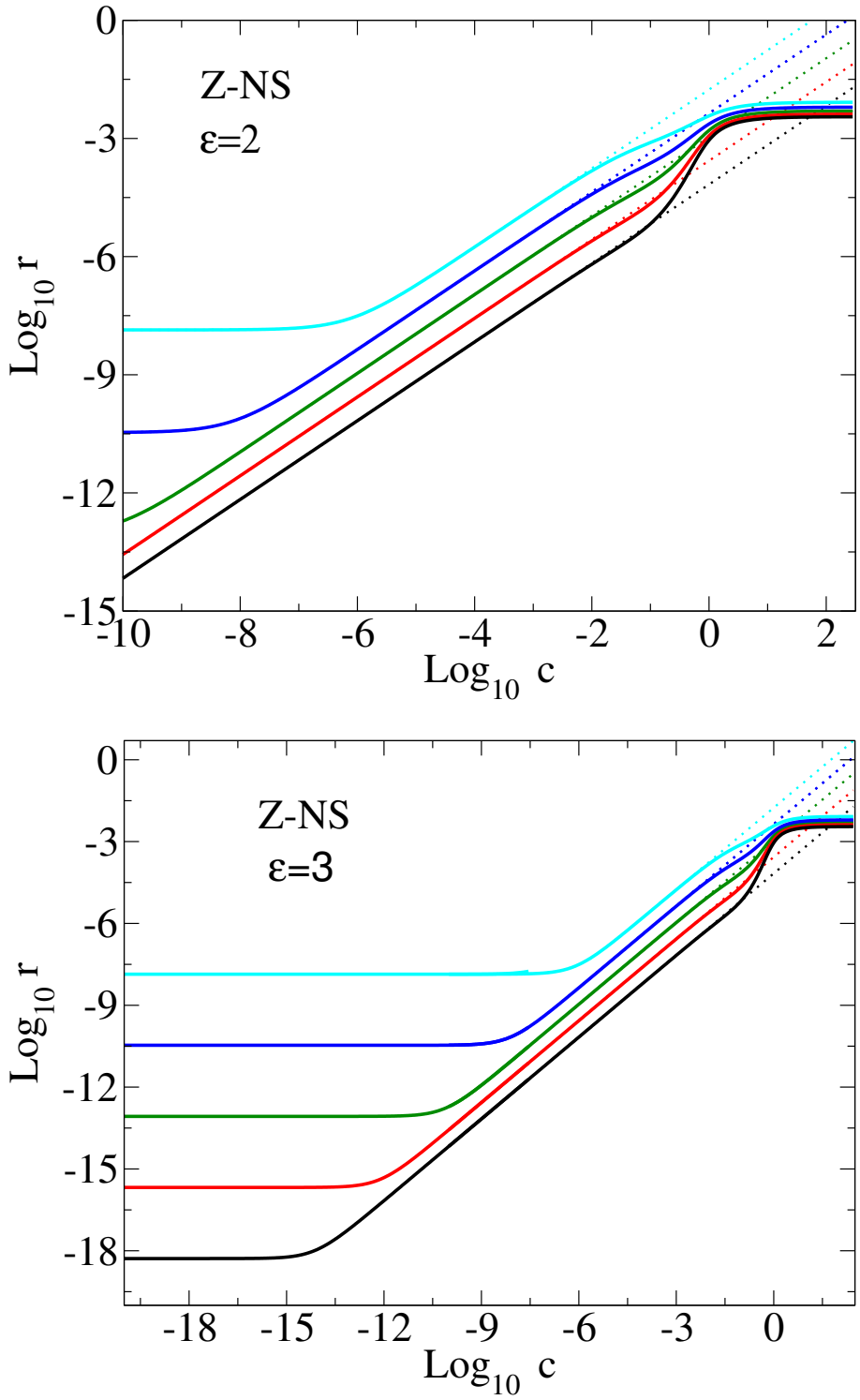

FIG. S1: Unbinding rate as a function of the concentration for the Z-NS model, and for different protein sizes N ranging from 6 to 14, and monomer binding energy $\epsilon = 2$ and 3 (in units of k_BT) in log-log scale. Full lines: numerical calculation of the unbinding rate. Dotted lines: linear approximations to the unbinding rate, see main text.

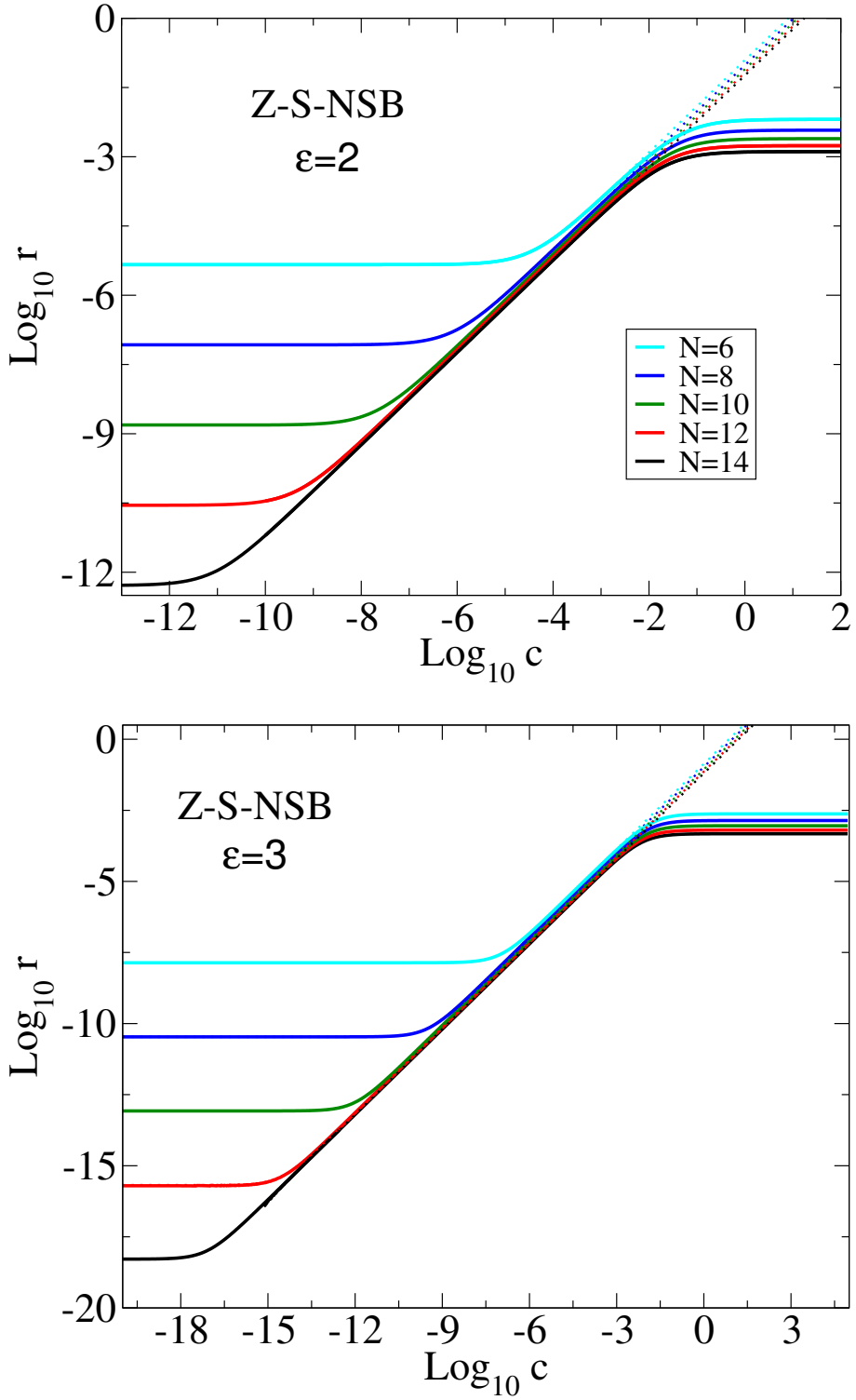

FIG. S2: Unbinding rate as a function of the concentration for the Z-S-NSB model, and for different protein sizes N ranging from 6 to 14, and monomer binding energy $\epsilon = 2$ and 3 (in units of k_BT) in log-log scale. Full lines: numerical calculation of the unbinding rate. Dotted lines: linear approximations to the unbinding rate, see main text.

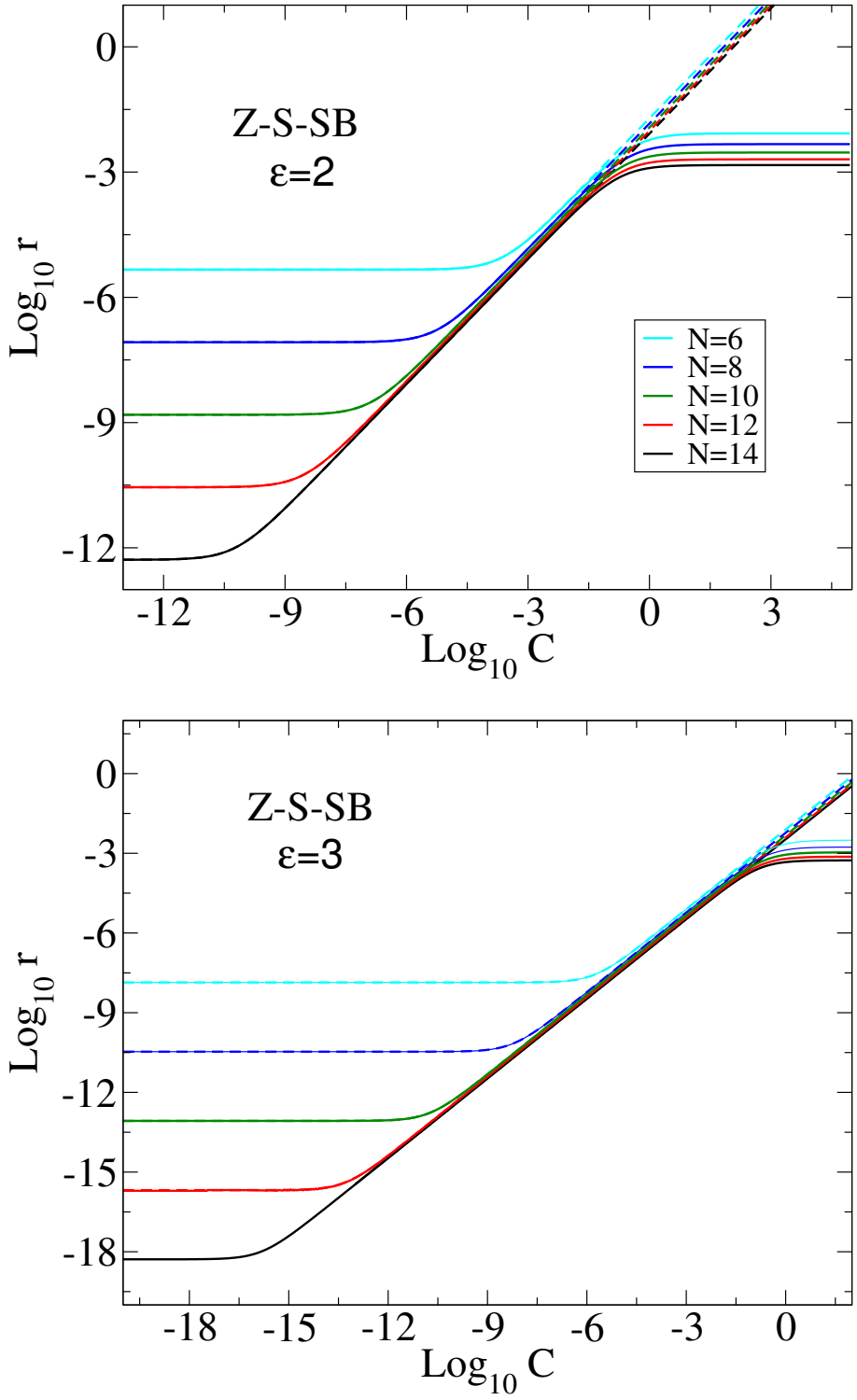

FIG. S3: Unbinding rate as a function of the concentration for the Z-S-SB model, and for different protein sizes N ranging from 6 to 14, and monomer binding energy $\epsilon = 2$ and 3 (in units of k_BT) in log-log scale. Full lines: numerical calculation of the unbinding rate. Dotted lines: linear approximations to the unbinding rate, see main text.

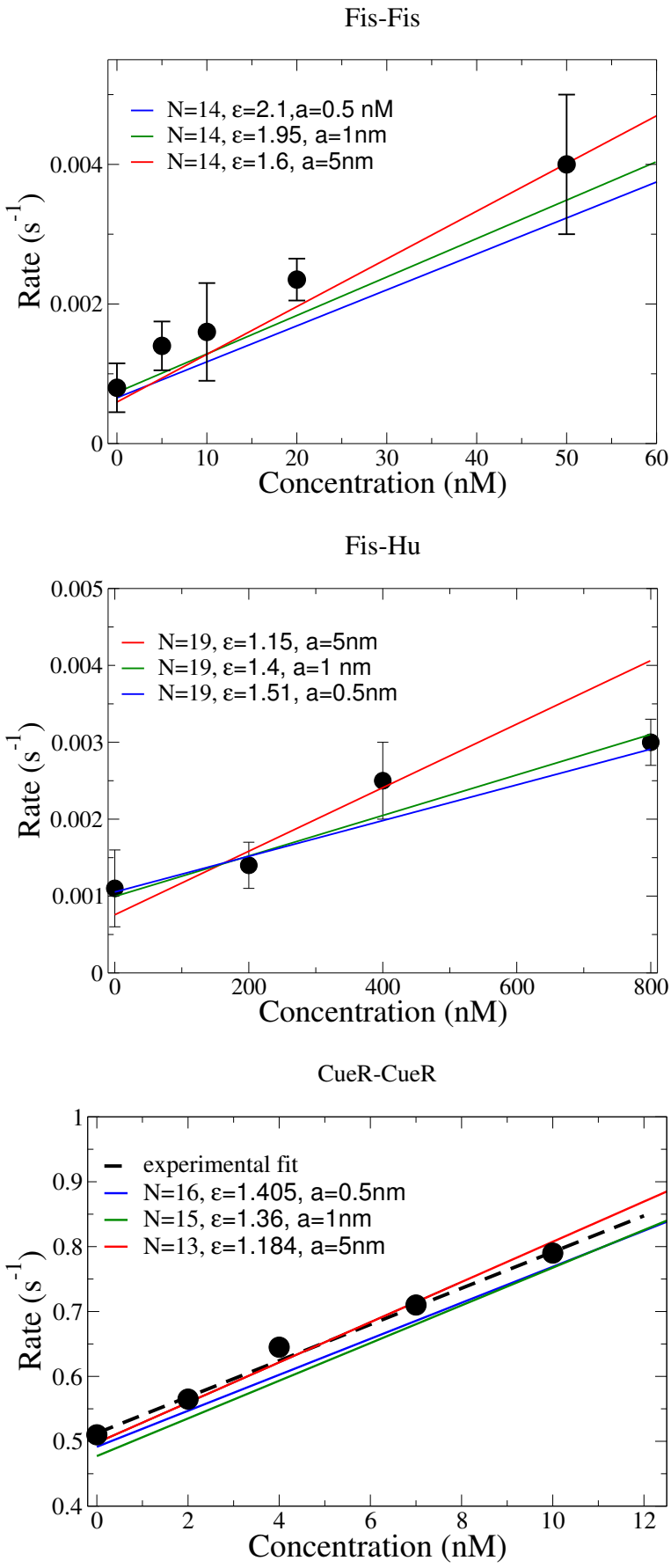

FIG. S4: Fit of the experimental data on concentration dependent dissociation rates of Fis proteins bounded on DNA as a function of concentration of Fis proteins (top) and Hu proteins (middle) in solution with the Z-NS model. Bottom: fit of CueR dissociation rates as a function of CueR concentration in solution with the Z-S-NSB model. See text for details on the parameter values.