

SUPPLEMENTARY INFORMATION: Physical Picture for
Mechanical Dissociation of Biological Complexes: from Forces to
Free Energies

R. Tapia-Rojo, C. Marcuello, A. Lostao, C. Gómez-Moreno, J.J. Mazo and F. Falo

1 Force spectroscopy experiments

Functionalization. The typical random immobilization on AFM tips and samples leads to many of the immobilized protein molecules are unable to interact with their partners because their interacting surfaces are used to anchor them to the support, which makes binding and subsequent rupture only occurring in a small percentage of approaches. Herein, an implemented procedure for tip and sample functionalization optimizing the recognition ability was used. Each protein molecule has a few lysine residues in their structure. First, once the protein complexes were formed, Sulfo-LC-SPDP crosslinker bound randomly to the primary amino groups of superficial lysines, and then labeled proteins were separated chromatographically from each FNR complex. Later, proteins were reduced with DTT and left exposed reactive sulfhydryl groups reacted with maleimide-PEG tips, forming disulfide bonds, or reacted with the thiol protected PDP groups in the case of the mica substrates. Thus, both partners were immobilized exposing the interaction surface of one molecule towards the other [Ref. 41 in the ms].

This method achieved a large increase in successful rupture events with respect to the corresponding random labeling using the same procedure, as was analyzed previously in detail [Ref. 34 in the ms]. The results range from 5-23%, with random labeling, to 40-77% with Fld-tagged tips and 34-61% for Fd-tagged probes in efficiency ratio for "useable curves" showing specific unbinding events regarding the total approaches. The increase with regard the randomly functionalized samples oscillated between 3-13 and 2-4 times for Fld and Fd approaches, respectively, and the differences attributed to the type of complex [Ref. 34 in the ms].

Selection of the specific forces. The total adhesion peaks generated during each force-distance curve either originates from a specific interaction (formation of a FNR-Fd/Fld bond) or from a non-specific one of any other origin. The use of PEG spacers to attach protein ligands to the AFM tip in unbinding DFS studies increase the length and flexibility of the sensor, allowing the molecules to freely move favoring first recognition and later the identification of the specific forces at the scans. The feature-rich stretching profile in water presented by PEG tethers constitutes

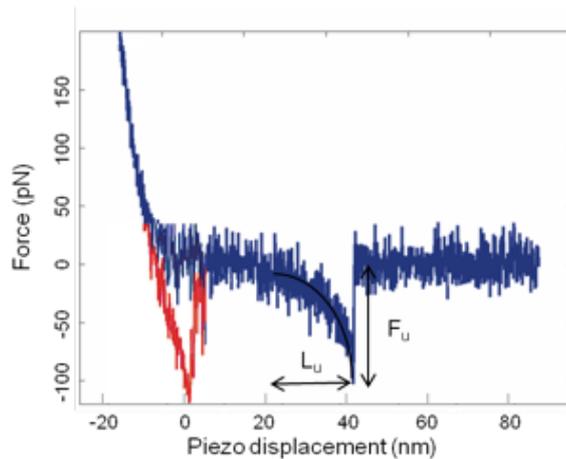


Figure 1: Representative experimental retraction force curve for a specific unbinding event corresponding to a single FNR-Fd complex. The F_z scan shows also a non-specific adhesion peak -in red to be distinguished- that follows the slope of the retraction. The unbinding event occurs at the unbinding length or tip-sample separation that is close to the length of the linker, around 20 nm, given by the piezo displacement encompassing the non-linear portion of the retraction curve before the rupture. The black line at this part represents the corresponding stretch of PEG according the WLC function. The shape of the force peak and the distance at which occurs ensure specificity. These considerations add certainty that measured rupture forces come from recognition events under study and not from artifacts or non-specific tip-sample adhesions.

a fingerprint of specificity, so specific force peaks show a nonlinear delay parabolic-like shape, which is characteristic of the stretching of a PEG linker, preceding the jump. The flexible tether sustains the increasing force until the complex dissociates, as indicated by a sudden jump to zero force. This occurs at a certain force value (unbinding force, F_U) and tip-sample distance (unbinding length, L_U). In contrast, in non-specific adhesions the contact curve extends towards negative values keeping the same slope, indicating that the bare tip, not the sensor, remains in contact with the surface. Figure 1 shows a representative Fz curve describing the two adhesion types. Control experiments were developed adding an excess of free ligand to block the FNR binding sites, at different R and for both type of complexes. In all the cases, they produced an important decrease in successful rupture events, giving very similar more probable rupture forces at the histograms, thus verifying the specificity and correctness of the measurements [Ref. 34 in the ms].

2 Potential Model and Simulation Details

We model the free energy profile for a mechanical rupture of a biological complex event as:

$$G(x_p) = D(1 - e^{-ax_p})^2 + Ue^{-(x_p - x^\dagger)^2/b} + F_0 [1 + \tanh w(x_p - s)]. \quad (1)$$

Each one of the three terms of this profile reflects one of the three characteristic regions of this system: 1) a Morse potential creating an equilibrium well for the bound state; 2) a Gaussian barrier for the first step barrier; 3) a tanh term for the smooth slope to the dissociation state, as an energetic plateau. In this sense, we can relate the free energy magnitudes with the parameters of the potential $\Delta G^\dagger \approx D+U$ and $\Delta G^0 = D+2F_0$. The remaining parameters set the qualitative shape of the profile, set such that the we have a steep first slope and a smooth second slope.

The exact set of parameters is : $D = 12\text{pNnm}$, $a = 3\text{nm}^{-1}$, $U = 24\text{pNnm}$, $x^\dagger = 0.5\text{nm}$, $b = 0.03\text{nm}^{-2}$, $F_0 = 24\text{pNnm}$, $w = 0.75\text{nm}^{-1}$ and $s = 4.0\text{nm}$, setting a free energy barrier $\Delta G^\dagger = 7.7k_B T$ and $\Delta G^0 = 14.7k_B T$.

We want to stress that the exact shape of this profile is not a critical aspect of the model, as long as we maintain the scale separation of the two slopes.

3 Simulation protocol

Numerical force spectroscopy experiment simulations are carried out by integrating the Langevin equation of motion,

$$m\ddot{x}_p = -m\eta\dot{x}_p - \nabla G(x_p) + F_{WLC}(\gamma - x_p) + \xi(t), \quad (2)$$

where m is the reduced mass of the complex, η the viscous damping and $\xi(t)$ white thermal noise. The term $-\nabla G(x_p)$ is the force derived from the movement in the free energy profile, while F_{WLC} is the force exerted by the polymer, modeled, by a Worm Like Chain model:

$$F_{WLC}(x) = \frac{k_B T}{P} \left[\frac{1}{4} \left(1 - \frac{x}{L} \right)^{-2} - \frac{1}{4} + \frac{x}{L} \right], \quad (3)$$

where P is the persistence length, $P = 0.37nm$ for the PEG polymer used here, and L its contour length ($L = 20nm$ in our case).

Regarding the involved coordinates, x_p is the coordinate of the particle moving in the free energy profile $G(x_p)$ while γ is the distance of this particle to the linear spring (AFM tip), such that $\gamma - x_p$ is the extension of the WLC linker (polymer linker), and thus $F_{WLC}(\gamma - x_p)$ is the

force exerted by the polymer. As the AFM moves at constant velocity (force-extension mode), $\lambda = Vt$ is a control parameter (not fluctuating), which can be expressed as $\lambda = \gamma + F_{WLC}/K_c$, being K_c the elastic constant of the linear spring. Here we consider that force equilibrium at the tip of the AFM, and thus $F_{WLC}(\gamma - x_p) = K_c \Delta z$, where Δz is the elongation of the linear spring. This is a reasonable assumption due to the scale separation between the AFM tip ($\sim \mu m$) and the polymer linker ($\sim nm$).

The Langevin equation of motion is integrated by a fourth-order Runge-Kutta stochastic algorithm. Choosing a certain pulling velocity $V = \lambda/t$, we run numerical experiments starting at $\lambda = 0$ and stopping at $\lambda = 40nm$ to ensure that the rupture event has taken place (polymer length is $L = 20nm$). This is looped for 10000 realizations for each pulling velocity.

We use normalized time and mass units $\tilde{m} = 1$ and $\tilde{t} = 1$, but pN units for force and nm for length, given that we are mainly interested in comparing forces, lengths and energy with the experiments. Simulations are carried out at room temperature $T = 4.1pNnm = k_B T$. The damping in the normalized time units is $\eta = 10$.

4 Calculation of the work performed over a force-extension trajectory

As discussed in [1], the proper definition for the non-equilibrium work performed on a system is $dW = Fd\lambda$, where λ is a control parameter (in contrast with fluctuating variables, which are stochastic variables) Here $d\lambda = Vdt$ - where V is the pulling velocity of the cantilever. According to Fig. 1 in main text, $\lambda = \gamma + \Delta z$, where Δz is the cantilever deflection, modeled as a linear spring, so $\Delta z = F/K_C$. The work accumulated along a non-equilibrium transition from $\lambda = 0$ to λ^\dagger , where λ^\dagger is a sufficiently large value of λ so that the rupture event has occurred¹, is defined as:

$$W = \int_0^{\lambda^\dagger} F(\lambda)d\lambda. \quad (4)$$

By changing the variable to γ (and neglecting the change in the coordinate x_p), we obtain

$$W = \int_0^{\gamma^\dagger} F_{WLC}d\gamma + \frac{1}{2} \frac{f^\dagger}{K_C}. \quad (5)$$

In this sense W depends only on the rupture force f^\dagger .

5 Fitting of experimental data from other complexes

Table I in main manuscript shows data for ΔG^\dagger and ΔG^0 for six different ligand-receptor complexes, including the FNR-Fd and FNR-Fld analyzed here, showing how $\Delta G^0 > \Delta G^\dagger$ is found. ΔG^0 values are taken from calorimetry experiments. ΔG^\dagger are determined from force spectroscopy measurements. In order to determine them we reinterpret the original data in the context of our model.

Dynamic force spectroscopy AFM experiments for LFA-1:ICAM-1 and LFA-1:ICAM-2 complexes expressed in Jurkat cells where originally published in [2], and interpreted within a two-barrier profile obtaining by a two-region fitting of Bell-Evans expression. As discussed in the

¹In the case of mechanical unbinding the exact value of λ^\dagger is not critical, as once the rupture event has occurred, the interaction disappears and thus $\langle F \rangle = 0$ from then on. This is different from stretching of biomolecules where, once denatured, and underlying polymer stretching is still present

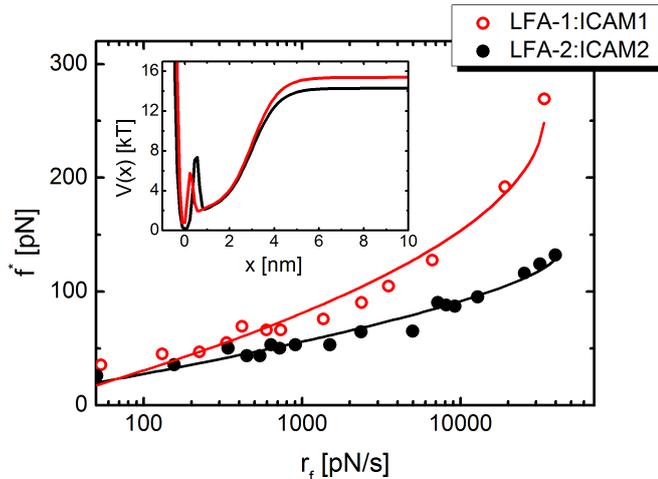


Figure 2: Fitting of dynamic force spectroscopy experiments for LFA-1:ICAM-1 and LFA-1:ICAM-2 complexes taken from [2] to Eq. (3) in main text. *Inset*: Proposed free energy profile for such complexes according to previous fitting.

main text, we consider more appropriate the proposed protocol, and thus we reinterpret the data in this context.

Figure 2 shows unbinding force data for LFA-1:ICAM-1 and LFA-1:ICAM-2 taken from [2] and fitted to Eq. (3) of main text. This allows us to obtain the free energy barrier height and position, yielding LFA-1:ICAM-1: $\Delta G^\ddagger = 8.57 \pm 0.42 k_B T$ and $x^\ddagger = 0.18 \pm 0.01 nm$ and LFA-1:ICAM-2: $\Delta G^\ddagger = 7.57 \pm 0.38 k_B T$ and $x^\ddagger = 0.40 \pm 0.01 nm$. Inset shows the proposed free energy profile according to our model and the obtained magnitudes.

Data for Biotin:Streptavidin and Biotin:Avidin complexes is taken from [3] although originally published in [4]. Here, due to the vast range of unfolding rates employed, one would see a clear deviation from Eq. (3) at lowest rates. As mentioned in the main text, this might be an effect of the second smooth slope, which starts to be "seen" by the experiment due to the extremely low pulling velocities.

6 Goodness of fit and model selection.

In the main text we show the results of the experimental data fit to equation (1) in three different cases: Bell-Evans model ($\nu = 1$ and two fitting parameters) and two more complex models ($\nu = 1/2$ and $\nu = 2/3$ and three fitting parameters). A first approach to the goodness of the fit is to evaluate the χ^2 parameter which is an usual output of main statistic programs. However, as both kind of models have different number of fitting parameters a better measure of its increasing complexity is needed in order to compare the performance of different models. In the table (2) of the main text we write down the *reduced* χ^2 which is defined as the ratio of the χ^2 and the number of fit degrees of freedom ($n_{dof} = n - k$, where n is the number of data and k the number of parameter of the model). Although this quantity takes into account the number of parameters k , other quantities have been proposed, like the Akaike (AIC) or Bayesian (BIC) information criterion, that rigorously discriminate between different models [5]. We have also computed and include in table (2) of the paper the result of the BIC of each model.

The BIC or Schwarz criterion is defined by [6]:

$$BIC = -2 \log L + (k + 1) \log n, \quad (6)$$

where L is the maximum likelihood of the fit. A simple commonly used approach to the calculus of the BIC is given by $BIC = n + n \log 2\pi + n \log(RSS/n) + (k + 1) \log n$ with RSS the residual sum of squares, which is directly proportional to the usually computed χ^2 . Thus maximum likelihood criterion corresponds to minimum χ^2 . BIC introduces a penalty of $\log n$ for each additional fitting parameter of the model. The model with lowest value of BIC is the best model from the data statistics point of view. BIC has been proved to be in many cases an effective approach to solve the model selection problem. However, a couple of caveats should be mentioned here: BIC is suitable for number of data n much larger than the number of parameters in the model k ; and BIC generally penalizes free parameters more strongly than other criteria.

7 Fit of numerical data to other theories

In the manuscript we have fitted the experimental data to Eq. (1) (main text) using three values for the exponent ν . This theory considers that the barrier decreases as $(1 - f/f_c)^{1/\nu}$, being f_c the critical force, $f_c = \Delta G^\ddagger/\nu x^\ddagger$. Here, $\nu = 2/3$ correspond to a cubic potential, which is a reasonable choice as any analytical potential can be expanded to a cubic polynomial next to the rupture force (for intermediate forces, as the ones found here).

Regarding the numerical data, in Fig. 5 of the manuscript only results for the fitting to the $\nu = 2/3$ exponent are shown. However, in order to study the validity of the the different theories we can compare predictions for ΔG^\ddagger and x^\ddagger to the correct values for a well-known, although non trivial, potential profile, as the one described in the Eq. (1) of this SI. This analysis is done in the table and in Figure 3. It shows small differences between ($\nu = 2/3$ and $\nu = 1/2$, though the first theory describes in a better way the barrier dependence $\Delta G(F)$). Note that Bell-Evans theory, $\nu = 1$, does not allow for obtaining a prediction for ΔG^\ddagger .

Being the force barrier dependence the main factor in the theory, we show in Figure 3 the dependence of the barrier with the force as numerically obtained for the proposed model and predicted for the two theories. It can be seen that the $\nu = 2/3$ is better for a wide range of force values.

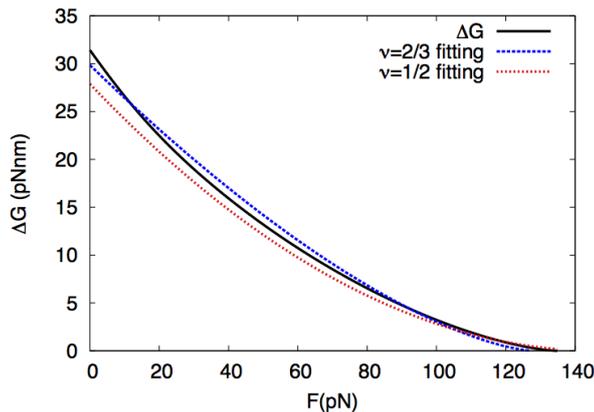


Figure 3: Barrier decreasing as a function of the applied force for our model potential [Eq. (1)], and for DFS theory with $\nu = 2/3$ and $\nu = 1/2$.

Table 1: Potential parameters ΔG^\ddagger and x^\ddagger and reduced χ^2 and R parameters.

| | $\Delta G^\ddagger(k_B T)$ | $x^\ddagger(nm)$ | χ^2/R^2 |
|------------------|----------------------------|------------------|--------------|
| Model | 7.70 | 0.50 | - |
| Fit: $\nu = 2/3$ | 7.28 ± 0.2 | 0.35 ± 0.08 | 0.06/0.996 |
| Fit: $\nu = 1/2$ | 6.80 ± 0.14 | 0.38 ± 0.03 | 0.08/0.997 |

8 Validation of the analysis protocol

For the sake of consistency, we probe our analysis protocol on the physical model using four different parameter sets. Figure 4 (left) shows the profiles for each of the four chosen parameter sets. In order to proof that our analysis procedure allows obtaining both ΔG^\ddagger and ΔG^0 , the profiles have four different dissociation free energies ΔG^0 but just two different barrier heights ΔG^\ddagger . This fact guaranties that the obtention of both free energy magnitudes from the same force data is completely independent.

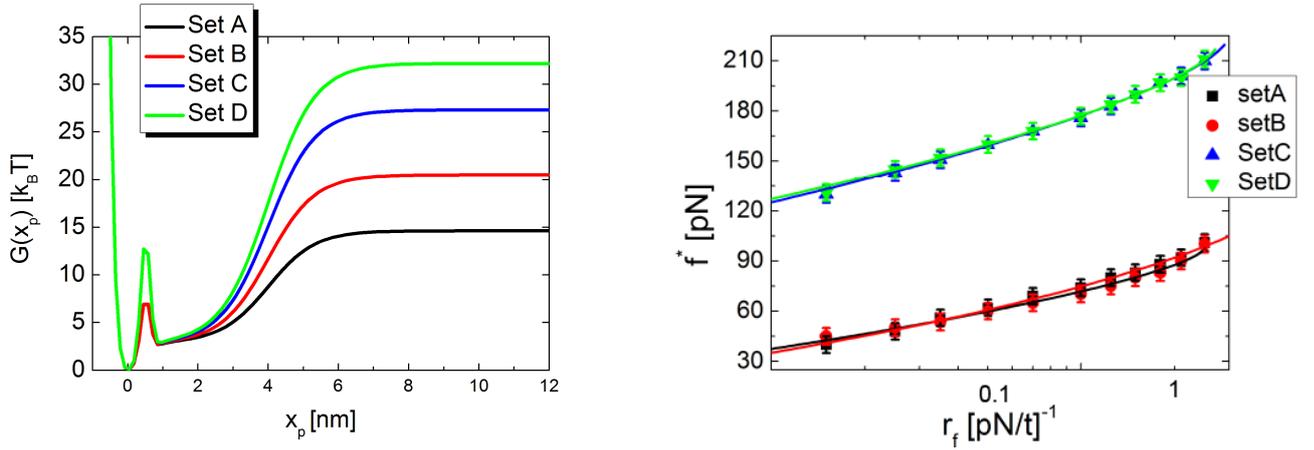


Figure 4: Left: Free energy profiles for the four chosen free energy profiles. All show different ΔG^0 values but barriers are equal in sets A-B and C-D respectively. Right: Typical rupture force f^* versus pulling rate r_f for the four parameter sets. Solid lines are fittings to Eq. (3) in main text.

We plot in Fig. 4 (right) the typical rupture forces versus the pulling rate for each of the four parameter sets. We can check how the curves corresponding to sets A-B and C-D superimpose respectively, as the profile has the same barrier height. The four data sets fit perfectly to expression (3) in the main text. Fitted parameters ΔG^\ddagger are shown in Table I below.

Figure 5 shows the Jarzynski estimator ΔG_J^0 as a function of the inverse of the pulling rate r_f^{-1} . Dashed lines indicate the dissociation free energy ΔG^0 set on each of the four parameter sets. We can see clear convergence of each data set to their respective ΔG^0 values, revealing that Jarzynski equality allows recovering the dissociation free energy successfully.

Table I gathers the ΔG^\ddagger and ΔG^0 values set for the four profiles together with the estimations from Jarzynski equality ΔG_J^0 and force spectroscopy theory ΔG_f^\ddagger respectively. Indicated ΔG_J^0 is the average of the last three values shown in Fig. 5 for each parameter set.

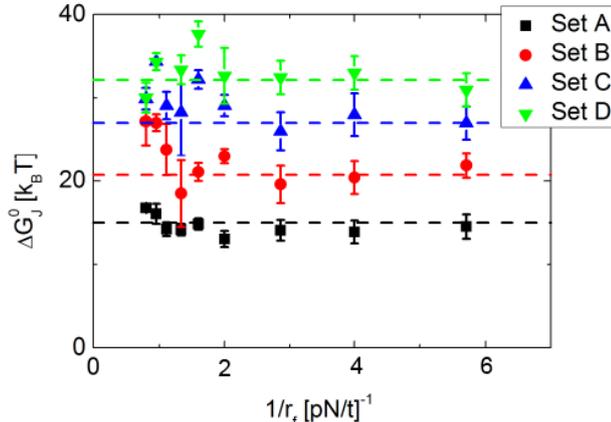


Figure 5: Jarzynski estimator ΔG_J^0 obtained from simulation performed for each parameter set. Dashed lines indicate the values set for each parameter set ΔG^0 .

Table 2: Free energy magnitudes ΔG^0 and ΔG^\dagger set for each parameter set and estimation according to our analysis protocol ΔG_J^0 from Jarzynski equality and fitted ΔG^\dagger .

| Parameter Set | $\Delta G^0(k_B T)$ | $\Delta G^\dagger(k_B T)$ | $\Delta G_J^0(k_B T)$ | $\Delta G_f^\dagger(k_B T)$ |
|---------------|---------------------|---------------------------|-----------------------|-----------------------------|
| A | 14.6 | 7.7 | 13.9 ± 0.5 | 7.3 ± 0.3 |
| B | 20.5 | 7.7 | 20.3 ± 1.0 | 6.7 ± 0.5 |
| C | 27.3 | 14.1 | 24.6 ± 0.3 | 13.2 ± 0.6 |
| D | 32.2 | 14.1 | 32.7 ± 1.5 | 12.5 ± 0.4 |

9 Effective stiffness

DFS analysis is based in Eq. (1) in the main text which is obtained after several approaches. Thus, this equation is obtained in the so-called weak spring limit of the system where $K_M \gg K_L, K_C$. Deviations from this case have been previously studied in [7, 8] for instance. Another approach concerning the stiffness of the system is to assume a constant K_{eff} , thus neglecting the force dependence of the effective stiffness of the system (which comes mostly from the K_L term, the PEG is not a linear spring) [9]. With these caveats Eq. (1) can be used to analyze data and a well defined pulling rate $r_f = K_{\text{eff}}V$ is defined. Note here that from a theoretical perspective it would be natural to use the well controlled pulling velocity V as control parameter of the system instead of the pulling rate r_f . However, following the tradition in the literature on the subject, we have decided to present our results in terms of r_f . Thus, in our work, we use a constant value of $K_{\text{eff}} = 10\text{pN/nm}$. We could question the validity of this approximation, then we should point two crucial facts (i) Theory is *validated* by our numerical simulations. We use the same approach to study the numerical simulations. There, the barrier parameters obtained by the DFS analysis show good agreement to the correct values. It shows the degree of robustness and accuracy of the employed theories. (ii) The error associated to neglecting the $K_{\text{eff}}(F)$ dependence is similar (same order of magnitude) to the one introduced by two other unavoidable approximations under the theory: the assumption of a specific form for the force dependence of barrier and prefactor in the Kramers rate expression, which are at the core of Eq. (1) result. To finish, this is the price to pay in order to have a simple useful result as Eq. (1).

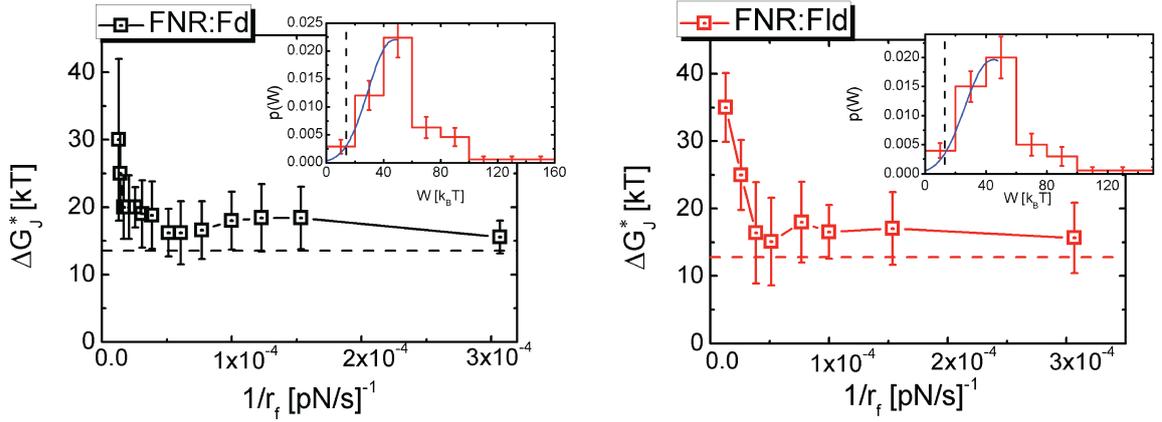


Figure 6: Corrected Jarzynski estimators by the bias estimation obtained fitting the work distributions (blue solid lines in the insets).

10 On the convergence of Jarzynski estimator

The Jarzynski estimator for a finite number of N irreversible work measurements W_i can be written as:

$$\Delta G_J = -\log \left[\frac{1}{N} \sum_{i=1}^N \exp \left(-\frac{W_i}{k_B T} \right) \right], \quad (7)$$

where k_B is the Boltzmann constant and T the absolute temperature. Although it is an unbiased estimator when $N \rightarrow \infty$, in practice it suffers from poor convergence and it is highly biased when the system is too far from equilibrium. Efforts have been made on developing analytical theories to correct the bias for finite N . In particular, under conditions of high dissipation, Ref. [10] proposes the following expression for the bias of the Jarzynski estimator:

$$B_N = \mu + \log N - \Omega (\log N)^{1/\delta} - \lambda^{(1-\delta)/\delta} \left[\gamma + \frac{1-\delta}{\delta} \log (\log N) + \log \frac{q}{\delta} \right], \quad (8)$$

where $\gamma_E = 0.5772 \dots$ is the Euler-Mascheroni constant and Ω and δ are parameters given by fitting the work distribution left tail to the following expression:

$$P(W) = \frac{q}{\Omega} \exp \left[- \left(\frac{|W - W_M|}{\Omega} \right)^\delta \right], \quad (9)$$

where

$$\mu = (\delta - 1) \left(\frac{\Omega}{\delta} \right)^{\delta/\delta-1}, \quad \lambda = \log N \left(\frac{\delta}{\Omega} \right)^{\delta/\delta-1}, \quad (10)$$

and q is a normalization constant. Then the corrected Jarzynski estimator is simply $\Delta G_J^* = \Delta G_J - B_N$.

In the case studied here, as mentioned in the main text and shown in Fig. 4, higher pulling rates lead to an overestimation of the free energy, due to the mentioned issues. Nonetheless, lowest rates meet the agreement between being close enough to equilibrium and having enough statistics to reach a well-converged Jarzynski estimator.

However, we calculate now the estimation of the bias shown in Eq. (8) to the work distributions showing high dissipation and thus being intrinsically biased. Figure 6 shows the

corrected Jarzynski estimator ΔG^* with an example of two work distributions fitted to Eq. (9) (insets), where the dashed vertical lines are the respective free energy values as obtained in the calorimetry experiments. Comparing with Fig. 4 in the main text, the free energy estimation is improved, specially in the intermediate rate region. Highest rates have a very poor estimation of the left tail (lower work values). Therefore, the fits to Eq. (9) are not meaningful, so is not the calculation of B_N .

References

- [1] Mossa A, de lorenzo S, Huguet JM, Ritort F (2009) Measurement of work in single-molecule experiments. *J Chem Phys* 130: 234116.
- [2] Wojcikiewicz W, Abdulreda M, Zhang X, Moy VT (2006) Force spectroscopy of LFA-1 and its ligands, ICAM-1 and ICAM-2. *Biomacromolecules* 7:3188-3195.
- [3] Hyeon C, Thirumalai D (2012) Multiple barriers in forced rupture of protein complexes. *J Chem Phys* 137:055103.
- [4] Merkel R, Nassoy P, Leung A, Ritchie K, Evans E (1999) Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature* 397:50.
- [5] Wit E, van den Heuvel E, Romeijn J-W (2012) ‘All models are wrong ...’: an introduction to model uncertainty. *Statistica Neerlandica* 66, 217-236.
- [6] Schwarz G, (1978) Estimating the dimension of a model. *The annals of Statistics* 6, 461-464.
- [7] Tshiprut Z, Klafter J, Urbakh M (2008) Single-molecule pulling experiments: when the stiffness of the pulling device matters. *Biophys. J.* 95: L42.
- [8] Maitra A, Arya G (2010) Influence of pulling handles and device stiffness in single-molecule force spectroscopy. *Phys Chem Chem Phys* 13:1836.
- [9] Ray C, Brown JR, Akhremitchev BB (2007) Rupture Force Analysis and the Associated Systematic Errors in Force Spectroscopy by AFM. *Langmuir* 2007, 23, 6076-6083.
- [10] Palassini M.; Ritort F. *Phys. Rev. Lett.* **2011**, *107*, 060601.