Supplementary Information

Orthogonal fingerprinting for accurate and fast single-molecule mechanical profiling of proteins

Carolina Pimenta-Lopes[*,#], Carmen Suay-Corredera*, Diana Velázquez-Carreras, David Sánchez-Ortiz, Jorge Alegre-Cebollada*

Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), 28029 Madrid, Spain

*equal contribution

[#]current address: Department of Physiological Sciences, University of Barcelona, Spain

Correspondence to Jorge Alegre-Cebollada: jalegre@cnic.es

Supplementary Table 1. Number of unfolding events per AFM experiment analyzed in this study, following both TFP or OFP strategies.

$(C3)_8$, TFP	$(C3)_8$, OFP with $(C3\text{-}L)_4$	$(C3\text{-}L)_4$, TFP	$(C3\text{-}L)_4$, OFP with $(C3)_8$	$(C3\text{-}L)_4$, OFP with $(C3\text{-}SUMO1)_4$	$(C3\text{-}SUMO1)_4$, TFP	$(C3\text{-}SUMO1)_4$, OFP with $(C3\text{-}L)_4$
117	117	14	29	55	69	54
84	71	35	52	42	7	71
140	182	17	81	82	119	16
141	225	42	136	14	148	92
65	30	31	13	6	149	52
224		38		42	166	57
96				67	47	176
191				47	37	29
52				43		41
75				65		38
149				178		230
				66		84
				81		70
				85		33

Supplementary Table 2. Monte Carlo simulations considering different mechanical parameters. The table reports the value of RSD of the distribution of ΔmF_u, at different values of r_0 and Δx. The remaining simulation parameters were the same as in Figure 1B in the main text.

	Calibration Uncertainty (%)	r_0 *= 0.2 nm* Δx *= 0.01 s^-1^*	r_0 *= 0.2 nm* Δx *= 0.015 s^-1^*	r_0 *= 0.25 nm* Δx *= 0.01 s^-1^*	r_0 *= 0.25 nm* Δx *= 0.015 s^-1^*
OFP	3.6	0.033	0.034	0.025	0.028
	18	0.035	0.037	0.035	0.036
TFP	3.6	0.053	0.054	0.051	0.045
	18	0.238	0.232	0.227	0.231

Supplementary Text 1. Sequences of the proteins used in this report.

We highlight the sequence of the domains in different colors: C3, red; protein L, bold type; SUMO1, green). Linkers and extra amino acids are shown in regular black type.

(C3)$_8$

MRGSHHHHHHGSPVLITRPLEDQLVMVGQRVEFECEVSEEGAQVKWLKDGVELTREETFKYRFKKDGQRHHLIINEAMLEDAGHYALCTSGGQALAELIVQEKRSPVLITRPLEDQLVMVGQRVEFECEVSEEGAQVKWLKDGVELTREETFKYRFKKDGQRHHLIINEAMLEDAGHYALCTSGGQALAELIVQEKRSPVLITRPLEDQLVMVGQRVEFECEVSEEGAQVKWLKDGVELTREETFKYRFKKDGQRHHLIINEAMLEDAGHYALCTSGGQALAELIVQEKRSPVLITRPLEDQLVMVGQRVEFECEVSEEGAQVKWLKDGVELTREETFKYRFKKDGQRHHLIINEAMLEDAGHYALCTSGGQALAELIVQEKRSPVLITRPLEDQLVMVGQRVEFECEVSEEGAQVKWLKDGVELTREETFKYRFKKDGQRHHLIINEAMLEDAGHYALCTSGGQALAELIVQEKRSPVLITRPLEDQLVMVGQRVEFECEVSEEGAQVKWLKDGVELTREETFKYRFKKDGQRHHLIINEAMLEDAGHYALCTSGGQALAELIVQEKRSPVLITRPLEDQLVMVGQRVEFECEVSEEGAQVKWLKDGVELTREETFKYRFKKDGQRHHLIINEAMLEDAGHYALCTSGGQALAELIVQEKRSPVLITRPLEDQLVMVGQRVEFECEVSEEGAQVKWLKDGVELTREETFKYRFKKDGQRHHLIINEAMLEDAGHYALCTSGGQALAELIVQEKRSCC

(C3-L)$_4$

MRGSHHHHHHGSPVLITRPLEDQLVMVGQRVEFECEVSEEGAQVKWLKDGVELTREETFKYRFKKDGQRHHLIINEAMLEDAGHYALCTSGGQALAELIVQEKRS**MEEVTIKANLIFANGSTQTAEFKGTFEKATSEAYAYADTLKKDNGEWTVDVADKGYTLNIKF**RSPVLITRPLEDQLVMVGQRVEFECEVSEEGAQVKWLKDGVELTREETFKYRFKKDGQRHHLIINEAMLEDAGHYALCTSGGQALAELIVQEKRS**MEEVTIKANLIFANGSTQTAEFKGTFEKATSEAYAYADTLKKDNGEWTVDVADKGYTLNIKF**RSPVLITRPLEDQLVMVGQRVEFECEVSEEGAQVKWLKDGVELTREETFKYRFKKDGQRHHLIINEAMLEDAGHYALCTSGGQALAELIVQEKRS**MEEVTIKANLIFANGSTQTAEFKGTFEKATSEAYAYADTLKKDNGEWTVDVADKGYTLNIKF**RSPVLITRPLEDQLVMVGQRVEFECEVSEEGAQVKWLKDGVELTREETFKYRFKKDGQRHHLIINEAMLEDAGHYALCTSGGQALAELIVQEKRS**MEEVTIKANLIFANGSTQTAEFKGTFEKATSEAYAYADTLKKDNGEWTVDVADKGYTLNIKF**RSCC

(C3-SUMO1)$_4$

MRGSHHHHHHGSPVLITRPLEDQLVMVGQRVEFECEVSEEGAQVKWLKDGVELTREETFKYRFKKDGQRHHLIINEAMLEDAGHYALCTSGGQALAELIVQEKRSMSDQEAKPSTEDLGDKKEGEYIKLKVIGQDSSEIHFKVKMTTHLKKLKESYCQRQGVPMNSLRFLFEGQRIADNHTPKELGMEEEDVIEVYQEQTGGRSPVLITRPLEDQLVMVGQRVEFECEVSEEGAQVKWLKDGVELTREETFKYRFKKDGQRHHLIINEAMLEDAGHYALCTSGGQALAELIVQEKRSMSDQEAKPSTEDLGDKKEGEYIKLKVIGQDSSEIHFKVKMTTHLKKLKESYCQRQGVPMNSLRFLFEGQRIADNHTPKELGMEEEDVIEVYQEQTGGRSPVLITRPLEDQLVMVGQRVEFECEVSEEGAQVKWLKDGVELTREETFKYRFKKDGQRHHLIINEAMLEDAGHYALCTSGGQALAELIVQEKRSMSDQEAKPSTEDLGDKKEGEYIKLKVIGQDSSEIHFKVKMTTHLKKLKESYCQRQGVPMNSLRFLFEGQRIADNHTPKELGMEEEDVIEVYQEQTGGRSPVLITRPLEDQLVMVGQRVEFECEVSEEGAQVKWLKDGVELTREETFKYRFKKDGQRHHLIINEAMLEDAGHYALCTSGGQALAELIVQEKRSMSDQEAKPSTEDLGDKKEGEYIKLKVIGQDSSEIHFKVKMTTHLKKLKESYCQRQGVPMNSLRFLFEGQRIADNHTPKELGMEEEDVIEVYQEQTGGRSCC

Supplementary Text 2. Estimation of uncertainty in calibration by the thermal fluctuations method.

Calibration of cantilevers by the thermal fluctuations method considers the cantilever as a harmonic oscillator and applies the equipartition theorem to calculate its spring constant (k_{sc}) according to:

$$k_{sc} = \frac{k_b \cdot T}{S^2 \cdot \langle z^2 \rangle} \quad \textit{Equation S1}$$

In Equation S1, $k_b$is the Boltzmann constant, T is the absolute temperature, S (deflection sensitivity, in nm/V) is the slope of the change in voltage detected by the photodetector for displacements of the surface while in contact with the cantilever, and $\langle z^2 \rangle$ is the mean squared displacement of the cantilever in units of V^2 [1]. Hence, cantilever calibration by the thermal fluctuations methdo requires estimation of S and $\langle z^2 \rangle$. Forces are then calculated from the deflection signal of the cantilever (A-B, in units of V) according to:

$$F = (A - B) \cdot S \cdot k_{sc} \qquad \textit{Equation S2}$$

Combining Equations S1 and S*2*, we obtain

$$F = (A - B) \cdot \frac{k_b \cdot T}{S \cdot \langle z^2 \rangle} \qquad \textit{Equation S3}$$

Equation S3 shows that S and $\langle z^2 \rangle$ contribute equally to the error in force. We measured experimental distributions of S and $\langle z^2 \rangle$ for a single cantilever and found that the RSD of S was 3.5% while the RSD for $\langle z^2 \rangle$ was 0.8% (Supplementary Figure 4). Hence, we conclude that inaccuracies in the determination of deflection sensitivity, which are 4-5 times larger than variations in $\langle z^2 \rangle$, are the main driver of interexperimental variation in spring constants of AFM cantilevers estimated using the thermal fluctuations method, as proposed before [2].

Considering error propagation, we propose that a reasonable value for the minimum calibration uncertainty in force is given by $\sqrt{0.035^2 + 0.008^2}$ = 3.6%. It is interesting to note that the uncertainty in the determination of k_{sc} is higher since it depends on the square of S: $\sqrt{(2 \cdot 0.035)^2 + 0.008^2}$ = 7.1%. This effect can be observed in the distributions in Supplementary Figure 4.

To obtain errors (ε) in force for simulated AFM experiments, we drew random values from a normal distribution centered in 100 whose RSD corresponds to the % uncertainty being considered. Then:

$$\varepsilon = \frac{random\ value}{100} \qquad \textit{Equation S4}$$

This error is considered when calculating the probability of unfolding (Equation 3 in the main text).

Supplementary Text 3. A model for propagation of calibration errors to ΔmF_u.

We consider that each value of unfolding force ($F_u^{measured}$) is affected by a systematic error coming from uncertain force calibration:

$$F_u^{measured} = F_u^{real} + \delta \qquad \textit{Equation S5}$$

Hence, the value of mF_u that results from averaging unfolding data from *m* experiments, each one with a certain number of events *n*, can be estimated as:

$$mF_u^{measured} = \frac{\sum_{i=1}^{n_1}(F_{u,i}^{real}+\delta_{1,i})+\cdots+\sum_{i=1}^{n_m}(F_{u,i}^{real}+\delta_{m,i})}{n_1+\cdots+n_m} \qquad \textit{Equation S6}$$

We can consider that for every experiment j, $\sum_{i=1}^{n_j} F_{u,i}^{real} = n_j \cdot mF_{u,j}^{real}$, where n_j is the number of events in experiment j and $mF_{u,j}^{real}$ is the mean unfolding force that would have been measured in experiment j if there was no error in calibration. Similarly, considering that $\sum_{i=1}^{n_j} \delta_{j,i} = n_j \cdot \bar{\delta}_j$ is the average error per experiment, and $n_{events} = n_1 + \cdots + n_m$, we obtain:

$$mF_u^{measured} = mF_u^{real} + \frac{\sum_{j=1}^{m} n_j \cdot \bar{\delta}_j}{n_{events}} \qquad \text{Equation S7}$$

Equation 1 in the main text is derived from Equation S7 considering comparison between two proteins under the OFP assumption that $\bar{\delta}_j$ is the same for both proteins when these are measured in the same AFM experiment.

Supplementary Text 4. Interpretation of relative improvement in RSD by OFP with respect to TFP.

Figure 1E shows that the degree of improvement in RSD by OFP relative to TFP increases with the number of events per experiment and remains insensitive to the total number of experiments. Here, we provide a qualitative explanation for this observation.

The RSD of the distribution of ΔmF_u derives from two independent factors: the calibration uncertainty (Figure 1F) and the limited number of unfolding events defining the distribution of unfolding forces (Supplementary Figure 5B). In the case of symmetric datasets obtained in OFP, the contribution of calibration uncertainty is zero as predicted from Equation 1 (see main text).

What is the effects of increasing number of events per experiment in the RSDs for TFP and OFP measurements?

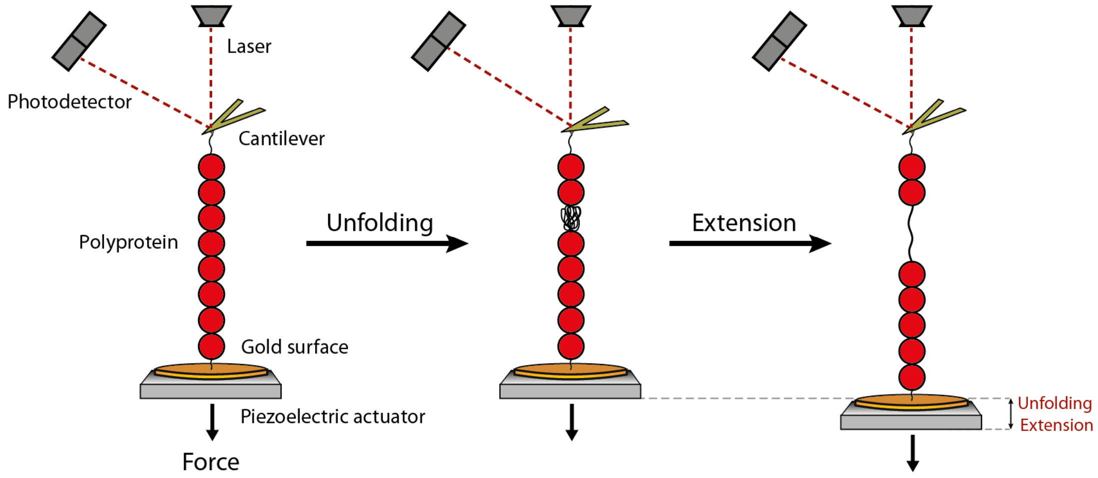
- In TFP, an increase in the number of events leads to better definition of the distribution of unfolding forces, but has no impact in the error associated to calibration uncertainty, which sets the value of RSD at high number of events (Supplementary Figure 5B).

- In OFP, the effect is the same, but in relative terms it is more prominent since the RSD associated to calibration uncertainty is already zero.

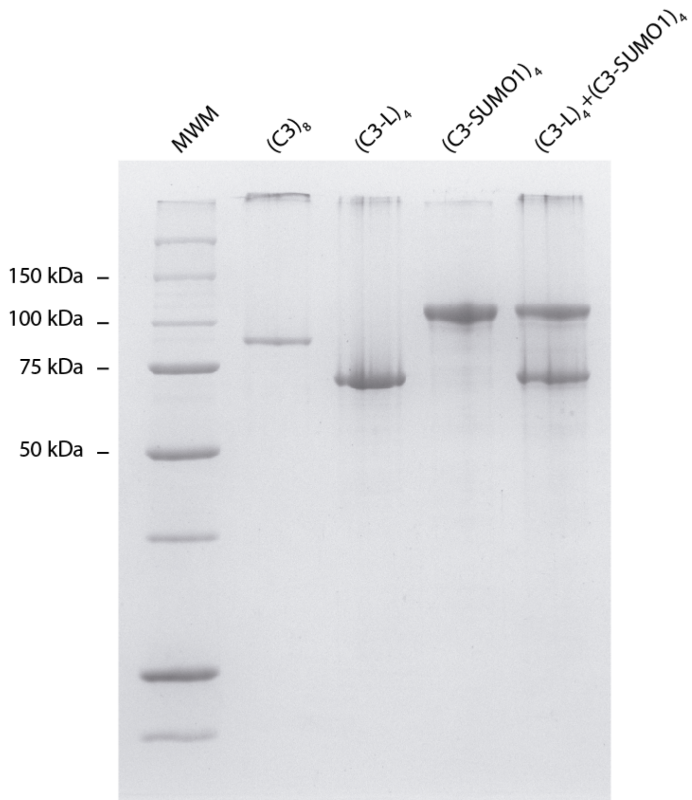
Increasing the number of experiments minimizes the impact of calibration uncertainty and also leads to an increase in the total number of events, both of which contribute to make the RSD of the distribution of ΔmF_u smaller in TFP experiments (Figure 1D). In OFP, since there is no error associated to calibration uncertainty, RSD decreases only as a consequence of higher number of events. However, as explained above, the impact of higher number of events on the RSD of ΔmF_u distributions is more pronounced for OFP experiments. We interpret that this differential impact of increasing number of events in TFP and OFP is behind the observation that the relative RSD between OFP and TFP remains fairly constant with increasing number of experiments (Figure 1E).

Supplementary Text 5. Validity of linear approximation in the kinetic Monte Carlo procedure used to obtain distributions of unfolding forces.

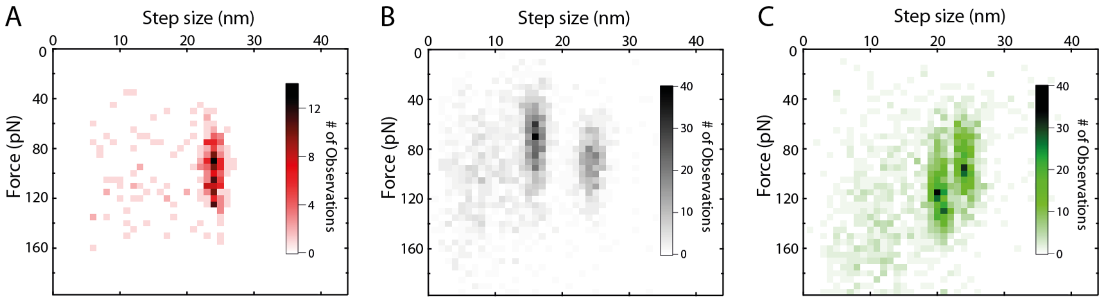
In our Monte Carlo simulations, we approximate the instantaneous probability of unfolding to the linear regime (Equation 3 in the main text), which is valid at low values of $n \cdot r \cdot \Delta t$ [3]. Considering that the mF_u in our simulations is around 100 pN, we can estimate the maximum number of events at the midpoint of the unfolding distribution that still satisfy $n \cdot r \cdot \Delta t < 0.05$, which according to Equation 3, is 385 (r_0 = 0.01 $s^{-1}$and Δx = 0.2 nm). Hence, we always kept the number of simulated unfolding events below 2 x 385.



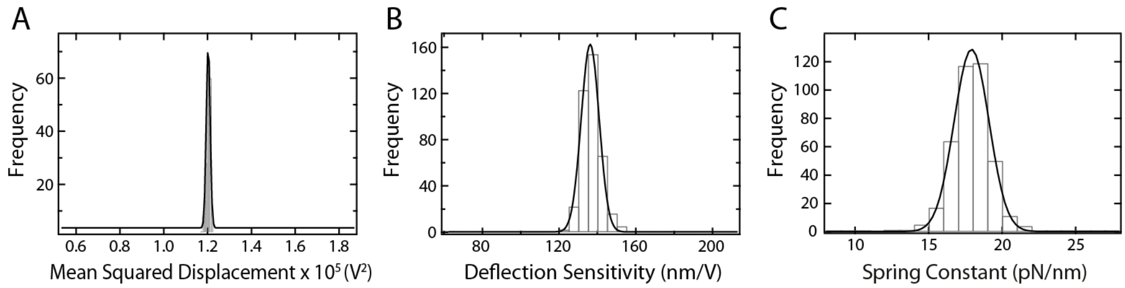
Supplementary Figure 1. Schematic representation of a single-molecule force-clamp AFM experiment. A single polyprotein is tethered between an AFM cantilever and a gold-coated surface. *Left*: The piezoelectric actuator moves away from the cantilever, which results in a pulling force applied to the polyprotein. The magnitude of the pulling force is calculated from the difference in voltage between the two regions of a split photodetector that is reached by the laser beam. *Middle*: When a domain unfolds, the force relaxes momentarily, changing the laser deflection. *Right*: To recover the programmed force set point in force-clamp experiments, the piezoelectric actuator is displaced, stretching the polyprotein and producing an unfolding step in experimental recordings (Figure 1A).



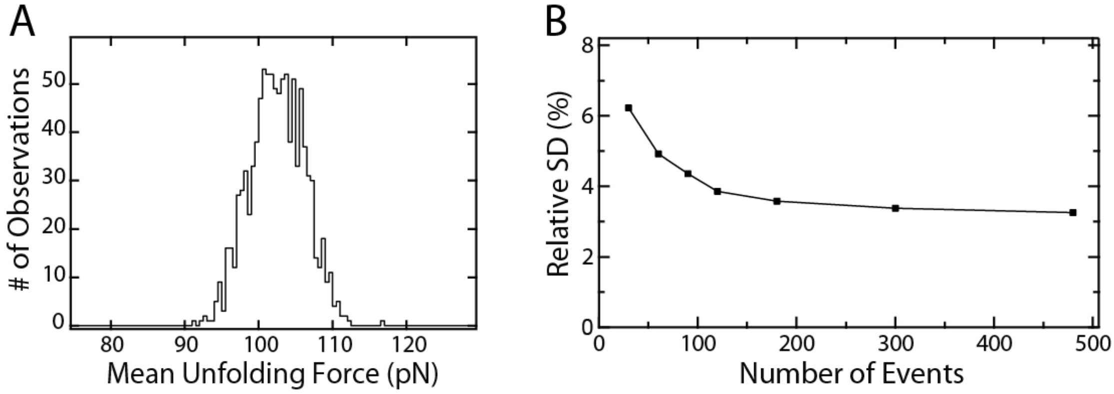
Supplementary Figure 2. 12% SDS-PAGE analysis of the purified proteins used in this report. We note that the C3-containing polyproteins have a tendency to show high-molecular weight aggregates that do not enter the resolving gel. We interpret this effect as an artifact of the electrophoresis, since equivalent aggregates do not appear in the void volume during size-exclusion chromatography. MWM: Precision Plus Protein Unstained Standards (Bio-Rad). The last lane shows results from simultaneous purification of (C3-L)$_4$ and (C3-SUMO1)$_4$



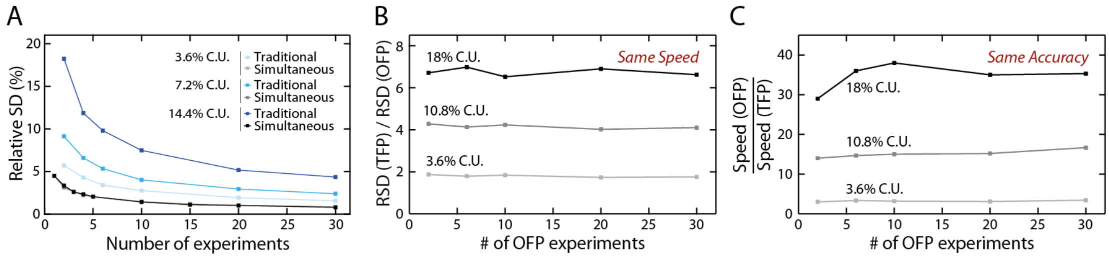
Supplementary Figure 3. **Definition of the unfolding lengths of C3, protein L and SUMO1.** Polyproteins were pulled at a rate of 40 pN/s. All traces containing at least two events of the same step size were included in the analysis. The bidimensional histograms show the frequency of the size and force of all steps in the selected traces. In all cases, the main unfolding lengths are clearly identified from a background of steps that correspond to non-specific interactions. For subsequent analysis, only traces showing the fingerprinting lengths were considered. **(A)** Results obtained for the mechanical unfolding of $(C3)_8$ (n = 293 steps), showing a single population of unfolding events at 24 ± 1 nm and around 90 pN. **(B)** Results for mechanical unfolding of $(C3\text{-}L)_4$ (n = 2555 steps). The two well-defined populations correspond to unfolding of L domains (step size 16 ± 1 nm at around 70 pN) and to unfolding of C3 (step size 24 ± 1 nm at around 90 pN). **(C)** Results for mechanical unfolding of $(C3\text{-}SUMO1)_4$ (n = 1998 steps). The two well-defined populations correspond to unfolding of SUMO1 (step size 20 ± 1 nm at around 115 pN) and C3 (step size 24 ± 1 nm at around 90 pN).



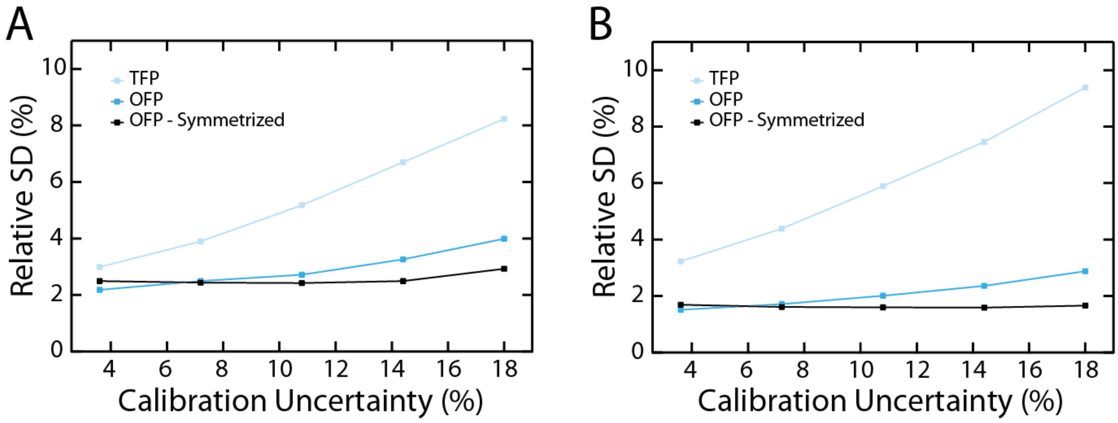
Supplementary Figure 4. Estimation of calibration uncertainty from multiple rounds of calibration of the same cantilever. (A) Distribution of 250 measurements of mean squared displacement ($\langle z^2 \rangle$) for a single cantilever ($1.2\cdot10^{-5} \pm 9.2\cdot10^{-8}$ V^2). **(B)** Distribution of 388 measurements of the deflection sensitivity (S) for a single cantilever (136.7 ± 4.9 nm/V). **(C)** Considering $\langle z^2 \rangle = 1.2\cdot10^{-5}$ V^2, we show the distribution of spring constants that arise from the values of deflection sensitivity in panel B (17.9 ± 1.3 pN/nm). Solid lines are Gaussian fits to the data. Mean ± SD of the distributions are indicated. In the three panels, the relative range of the x axes with respect to the mean value is the same.



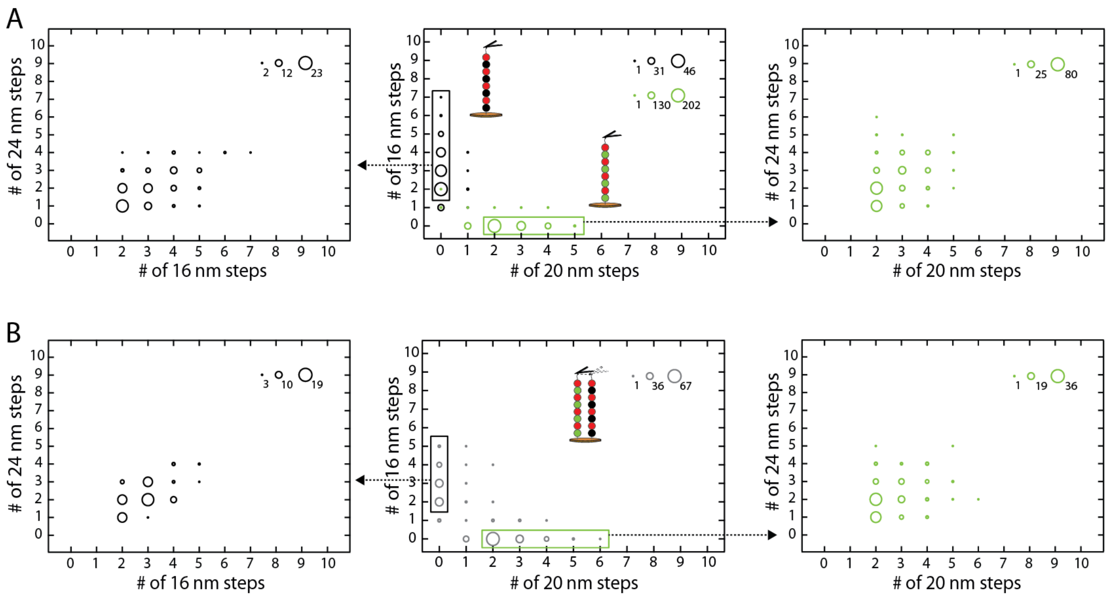
Supplementary Figure 5. Spread in mean unfolding force for an individual protein. (A) Distribution of mF_u corresponding to one traditional experiment (200 events) estimated from Monte Carlo simulations. **(B)** Dependence of the RSD in the distribution of mean unfolding force with the number of events, obtained for a single protein in an AFM traditional experiment. We considered a 3.6% calibration uncertainty. At low number of events, the distribution of mF_u is not well defined and the RSD is high. At higher number of events, distributions of mF_u are better defined, and the major contributor to RSD is the calibration uncertainty.



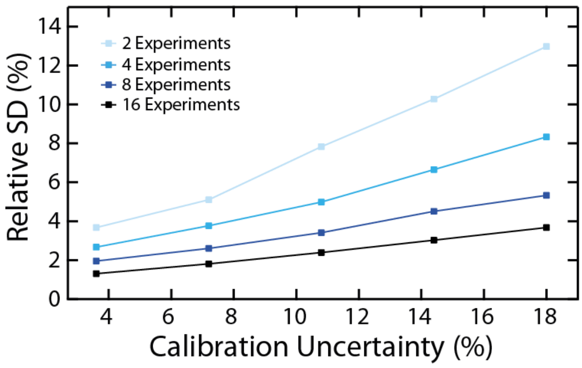
Supplementary Figure 6. Increased accuracy and speed of data acquisition by simultaneous AFM measurements. (A) RSD of the distribution of ΔmF_u estimated from Monte Carlo simulations at different calibration uncertainties (C.U.) and number of experiments. Note that in the case of simultaneous measurement, the RSD values overlap for the three calibration uncertainties. **(B)** Relative increase in accuracy achieved by OFP with respect to traditional AFM (at equivalent speed of data acquisition), at different calibration uncertainties and number of experiments. **(C)** Relative increase in throughput achieved by OFP with respect to traditional AFM (at equivalent accuracy), at different calibration uncertainties and number of experiments. Speed(OFP)/Speed(TFP) was calculated as the ratio between the number of TFP and OFP experiments that are needed to achieve the same RSD. All simulations considered 100 unfolding events per protein and experiment.



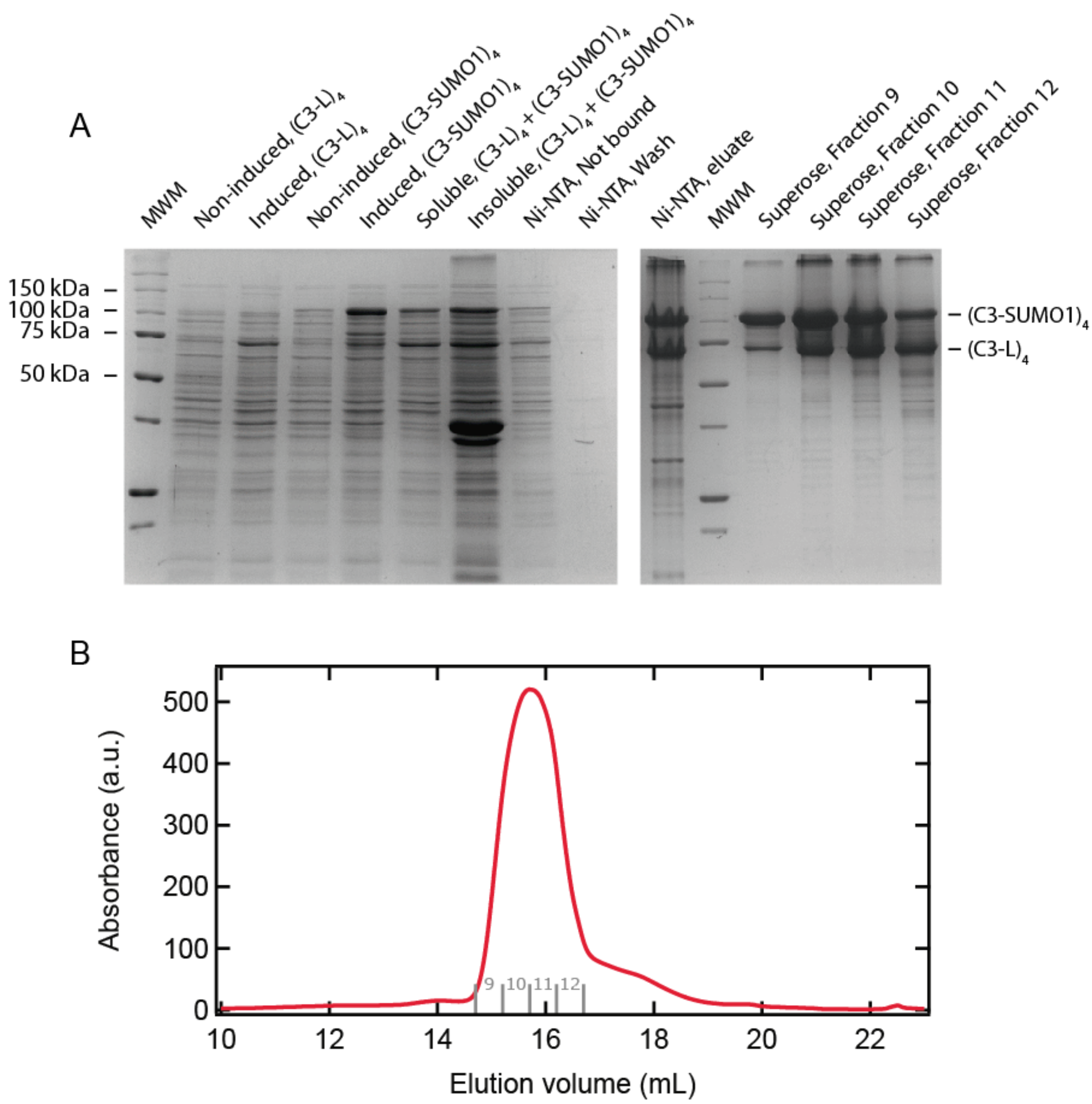
Supplementary Figure 7. Application of Monte Carlo simulations to estimate RSD associated to real datasets. (A) Monte-Carlo-estimated RSD of the distributions of ΔmF_u from the AFM experiments that compare the mF_u of C3 in the context of $(C3)_8$ and $(C3\text{-}L)_4$ (see Figure 2 in the main text). Simulations are fed with the actual number of unfolding events measured experimentally (Supplementary Table 1). Simulations were also run with symmetrized OFP data, by removing data to equal the number of events per protein within an OFP experiment. **(B)** Monte-Carlo-estimated RSD of the distributions of ΔmF_u from the AFM experiments that compare the mF_u of C3 in the context of $(C3\text{-}L)_4$ and $(C3\text{-}SUMO1)_4$ (see Figure 3 in the main text). Simulations are fed with the actual number of unfolding events measured experimentally (Supplementary Table 1). Simulations were also run with symmetrized OFP data, by removing data to equal the number of events per protein within an OFP experiment.



Supplementary Figure 8. Protein gating in dual-marker orthogonal fingerprinting. (A) Traditional fingerprinting. **(B)** Orthogonal fingerprinting. $(C3\text{-}L)_4$ and $(C3\text{-}SUMO1)_4$ unfolding traces were first classified according to the number of marker 16 and 20 nm steps (middle panels). Sorted traces were further classified according to the number of 24 nm steps, which correspond to C3 unfolding events (left and right panels). Results show that highly similar distributions are obtained in traditional and orthogonal fingerprinting experiments, providing further support to the gating protocol to sort traces coming from mixtures of proteins.



Supplementary Figure 9. Variation in accuracy of OFP due to asymmetry of datasets depends on the number of experiments. Monte-Carlo-estimated RSD of the distributions of ΔmF_u taking into account increasing number of OFP experiments. We have simulated OFP asymmetric datasets with alternating 50/150 events for each protein in each experiment.



Supplementary Figure 10. Simultaneous purification of (C3-SUMO1)$_4$ and (C3-L)$_4$. (A) 12% SDS-PAGE to monitor the steps of the expression and purification. *E. coli* cells containing the expression plasmid for (C3-SUMO1)$_4$ or (C3-L)$_4$ proteins are induced separately. The expression of the proteins is detected in the induced samples. Cells expressing both proteins are lysed together, and the soluble fraction is loaded in a Ni-NTA column. The fractions of highest protein concentration are subject to size-exclusion chromatography in an FPLC system using a Superose 6 Increase 10/300 GL column. Fractions 9-12 contain different proportions of a mixture of (C3-SUMO1)$_4$ and (C3-L)$_4$. OFP experiments can be set with these fractions directly. The preferred fraction gives similar number of unfolding events for both proteins. MWM: Precision Plus Protein Unstained Standards (Bio-Rad). **(B)** Chromatogram from the FPLC purification shows that (C3-SUMO1)$_4$ and (C3-L)$_4$ are not resolved and co-elute in fractions 9-12.

1

Supplementary References

1.Popa, I. *et al.* Force dependency of biochemical reactions measured by single-molecule force-clamp spectroscopy. *Nat Protoc* **8**, 1261-1276. (2013).

2.Ohler, B. Practical Advice on the Determination of Cantilever Spring Constants. Veeco Instruments Incorporated, Technical Report., (2007).

3.Shi, J. *et al.* Simulated data sets for single molecule kinetics: some limitations and complications of data analysis. *European biophysics journal : EBJ* **35**, 633-645. (2006).