Identical Sequences, Different Behaviors: Protein Diversity Captured at the Single-Molecule Level

Rafael Tapia-Rojo,[*,†,‡] Alvaro Alonso-Caballero,[†,‡] Carmen L. Badilla,[†] and Julio M. Fernandez[†]

†*Department of Biological Sciences, Columbia University, New York, NY 10027, USA*

‡*Current address: Department of Physics, King's College London, UK*

E-mail: Rafa.T.Rojo@gmail.com

Abstract

The classical "one sequence, one structure, one function" paradigm has shaped much of our intuition of how proteins work inside the cell. Partially due to the insight provided by bulk biochemical assays, individual biomolecules are assumed to behave as identical entities, and their characterization relies on ensemble averages that flatten any conformational diversity into a unique phenotype. While the emergence of single-molecule techniques opened the gates to interrogating individual molecules, technical shortcomings typically limit the duration of these measurements to a few minutes, which prevents to completely characterize a protein individual and, hence, to capture the heterogeneity among molecular populations. Here, we introduce a magnetic tweezers design, which showcases enhanced stability and resolution that allows us to measure the folding dynamics of a single protein during several uninterrupted days with a high temporal and spatial resolution. Thanks to this instrumental development, we do a complete characterization of two proteins with a very different force-response: the talin R3^{IVVI} domain and protein L. Days-long recordings on the same single molecule accumulate several thousands of folding transitions sampled with sub-ms resolution, which

allows us to reconstruct their free energy landscapes and describe how they evolve with force. By mapping the nanomechanical identity of many different protein individuals, we directly capture their molecular diversity as a quantifiable dispersion on their force response and folding kinetics. Our instrumental development offers a new tool for profiling individual molecules, opening the gates to the characterization of biomolecular heterogeneity.

Introduction

Traditionally, it is assumed that a protein's sequence univocally dictates its structure, which consequently determines its function.[1] For example, the classical funnel model for protein folding offers an intuitive vision of how the myriad of energetically-disfavored conformations accessible to a given residue sequence eventually collapse to a unique free energy basin that defines the functional native state.[2,3] Bulk biochemical and structural techniques have supported this paradigm, as the experimental signal is an intrinsic ensemble average over a population of thermodynamic size. However, increasing evidence resulting from novel experimental approaches have challenged this dogma, suggesting that proteins are dynamic entities with an intrinsic diversity arising from more complex conformational landscapes.[4–8]

In this context, the emergence of single-molecule force spectroscopy techniques shifted the paradigm of biophysical experimentation by allowing to interrogate individual molecules under force to directly sample the distributions of their molecular properties without relying on pre-averaged signals.[9,10] While protein nanomechanics has been typically approached using atomic force miscroscopy (AFM) and optical tweezers, these two techniques suffer from large mechanical drift that limits the measurements to a few minutes, eventually requiring to interrogate several molecular samples to accumulate enough statistics.[11,12] By contrast, magnetic tweezers, an intrinsically more stable technique, has been mostly devoted to the study of nucleic acids, likely owing to the lower temporal and spatial resolution of the early instrumental designs, and the initial difficulty to obtain specific and long lasting

2

tethers.[10,12–16] However, recent instrumental advances and the development of novel anchoring strategies are now demonstrating the expediency of magnetic tweezers for measuring protein mechanics.[17–20] Magnetic tweezers have the crucial advantage of offering intrinsic force-clamp conditions, which afford direct control of the intensive variable (force), while the extensive variable (extension) is measured, providing the natural statistical ensemble for studying protein dynamics in equilibrium. In this sense, by pushing the instrumental boundaries to increase the stability and resolution limitations, magnetic tweezers could enable to fully portray the nanomechanical properties of a single protein and, hence, open the gates to capturing the diversity of protein phenotypes and directly characterizing their functional heterogeneity.

Here, we present a novel magnetic tweezers design that enables us to do a full nanomechanical characterization of a single protein through long-term ultra-stable recordings at high temporal and spatial resolution. Our setup highlights its simple and bespoke design from off-the-shelf and affordable components that make it a highly accessible instrument, which we describe with all details necessary for its installation in any lab in the world. Thanks to this instrumental development, we measure days-long recordings of the folding dynamics of two different proteins, the talin $R3^{IVVI}$ domain and protein L, obtaining a complete nanomechanical mapping using a single molecular individual on each case. Our data demonstrates that polymer physics is a dominant component of protein folding under force, which governs the refolding transition by dictating the shape of the unfolded basin and the evolution of the free energy landscape with force. To directly explore the nanomechanical diversity of these two proteins, we measure several protein individuals and characterize the force response of each molecule, which shows an extent of dispersion both on its folding equilibrium and kinetics while their elastic properties remain homogeneous, validating our calibration and fingerprinting. Our results strongly suggest that proteins have heterogeneous nanomechanical properties, previously veiled by the limited stability of the available techniques, and that could result in a spectrum of functional heterogeneity.

3

METHODS

Instrumental Design

In magnetic tweezers force spectroscopy, a single molecule is immobilized between a glass cover slide and a superparamagnetic bead. By controlling a magnetic field gradient, the molecule is subjected to pN-level forces, while its extension is measured by video-tracking the bead diffraction or interference patterns, depending on the illumination geometry.[17,18,21] Due to the overly compliant trap created by the magnetic field, magnetic tweezers offers passive force-clamp conditions that afford direct control of the pulling force, which permits to interrogate single molecules in equilibrium.

We present here a novel magnetic tweezers setup, assembled from off-the-shelf and affordable components. Figure 1 shows a three dimensional model of our setup, indicating each of its elements (see SI Appendix for the parts list). Our setup has a simple and dedicated microscope design, using an optical arrangement for a 160-mm objective (see Fig. S1 and SI Appendix for the description and schematics of the optical pathway). Briefly, we illuminate from below using a LED source (Thorlabs, element #12), and collimate the light with a plano-convex lens (Thorlabs, element #10) to focus it on the objective with a tube lens (Thorlabs, element #7), while the image is recorded with a CMOS camera (Ximea, element #14).

While most magnetic tweezers designs control the force by displacing a pair of magnets using a motor or a voice coil,[17,18] we use here a magnetic tape head (Brush Industries, element #17), an old piece of technology more commonly employed to read magnetic bands in music cassettes or credit cards.[19] This implementation offers enhanced control of the pulling force through the electric current supplied to the tape head, which can be modulated with high accuracy and speed that enables us to do swift force changes or to apply complex force signals such as mechanical noise or force oscillations.[19,22] To generate calibrated and reproducible forces, it is paramount to position the tape head over the magnetic probe with

4
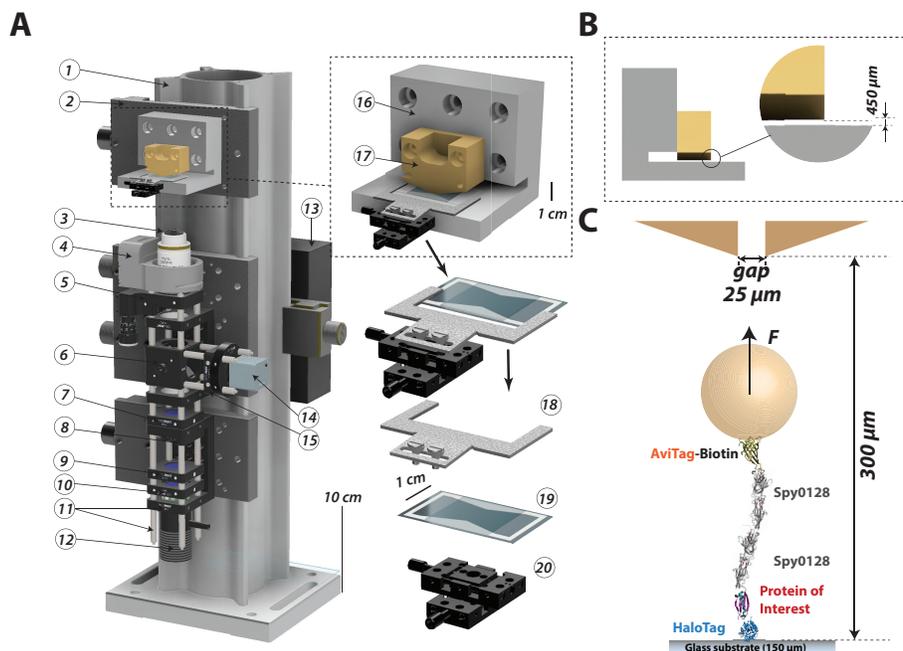
Figure 1: **Three-dimensional model of the magnetic tweezers setup:** (A) The setup comprises the following components: (1) 95-mm construction rail; (2) optical rail carriage; (3) 100x 160 mm objective; (4) piezo scanner; (5) Z flexor stage; (6) 50:50 beam splitter; (7) tube lens; (8) diaphragm; (9) filter; (10) plano-convex lens; (11) cage-rod system; (12) LED light source; (13) electronics box; (14) CMOS camera; (15) plano-convex lens; (16) mount piece (card-reader); (17) magnetic tape head; (18) manipulation fork; (19) flow cell; (20) XY linear stage. (B) Detail of the card reader piece. The tape head is maintained at a distance of 450 μm, which, when using 150 μm thick bottom glasses, positions the gap 300 μm away from the magnetic beads. (C) Schematics of our anchoring strategy for single-protein magnetic tweezers experiments. The protein of interest is inserted between a HaloTag and two Spy0128 repeats, following a biotilinated AviTag for tethering with streptavidin-coated superparamagnetic beads (M-270). This strategy provides a highly specific and stable anchor and an inextensible molecular handle.

micron precision, given that the magnetic field generated by the tape head changes over a length scale defined by the gap width, 25 μm in our case. To solve this problem, we introduce here a mount piece fabricated with high-precision CNC machining (element #16, see Fig. S2)—dubbed the "card-reader"—where we mount the tape head with a dowel-pin system. Thanks to the card-reader, the tape head gap is positioned precisely at 450 μm over the surface so that, when using 150 μm thick bottom glasses, the gap is at 300 μm over the magnetic probe (Fig. 1B). The tape head is maintained under electronic feedback with a PID current-clamp circuit, which showcases a bandwidth greater than 10 kHz (see Fig. S3 and SI Appendix). Importantly, the electronic circuit and the tape head is powered by a pair

5

of 12 V batteries—lead-acid 18 Ah motorbike batteries—which provide a clean and stable source of electric current, key to conduct swift force changes at very high rates.

The introduction of the card-reader piece is of capital importance for the ultra-stable conditions of our magnetic tweezers design, allowing us to apply highly reproducible and precise forces by simply controlling the electric current supplied to the tape head as:

$$F(I) = \left(2.79 \times 10^{-5}\right) I^2 + \left(1.64 \times 10^{-2}\right) I, \qquad (1)$$

where F is the pulling force in pN and I the electric current in mA. As described in,[19] the parameters in Eq. (1) are determined by the specifics of the tape head arrangement, including the vertical distance separating the gap and the bead, and the magnetic properties of the superparamagnetic beads, the M-270 Dynabeads in the present case. In our implementation, we can apply forces between 0 and 44 pN when using electric currents between 0 and 1,000 mA, and do force changes as small as $\sim 10^{-2}$ pN through electric current changes of ~ 1 mA. The force range can be increased by using magnetic beads of a larger diameter or higher magnetization. For instance, the implementation of the M-450 Dynabeads increases the maximum force to ~ 220 pN.[23]

The simplified dedicated microscope of this magnetic tweezers setup, combined with the capability of applying precise constant forces with the card-reader piece, results in an ultra-stable instrument compared to designs that implement standard microscopes, which permits doing continuous measurements during several days with negligible drift. In the SI Appendix, we include a practical guide with all details for reproducing our setup, including the parts list, electronic circuit diagram, and technical drawing for the card-reader piece.

Fluid Chamber Preparation

We conduct our single-molecule experiments in single-use fluid chambers (element #19) built by sandwiching two glass cover-slides (Ted Pella) with a custom-designed laser-cut parafilm

6

pattern. Using a 40x24 mm bottom glass and a 22x22 mm top glass, the fluid chambers shows two open wells that allow for buffer exchanges, useful, for instance, for altering in real-time the chemical conditions of the experiment.[24,25] Additionally, when conducting long-term experiments, the fluid chamber can be sealed using high-density white oil, which greatly stabilizes the buffer conditions and prevents evaporation. The bottom glass is functionalized to harbor the HaloTag O4 ligand, as previously described,[17] while the top glass is treated with repel silane to hydrophobize it and allow solution flow inside the chamber (see Fig. S4 and SI Appendix). The fluid chamber is clipped on a metal-stamped manipulation fork (element #18) using double-sided tape. The fork is mounted on an XY-stage (Newport, element #20) using two high-precision shoulder screws. The XY-stage is used to scout for beads over small 3x3 mm areas, but the fork can be loosened to displace it laterally and move to new regions in the fluid cell.

Molecular construct for single-molecule studies

To measure individual protein domains during extended periods of time, it is necessary to develop a long-lasting anchoring technique and molecular handles that interfere minimally with the signal of the protein of interest. Our strategy implements HaloTag chemistry for covalent and specific anchoring of the protein construct to functionalized glass cover slides, as previously discussed,,[11,17] while the other terminus is tethered to streptavidin-coated beads through a biotinilated AviTag (Fig. 1C). To measure single protein domains, we have here designed a molecular template comprising two repeats of the Spy0128 protein, which is inextensible under force owing to the isopeptide bonds that clamp its termini (Fig. 1C).[26,27] In this simple anchoring strategy, the protein of interest is inserted between the HaloTag enzyme at the N-terminus and the Spy0128 repeats, which provides a $\sim$15 nm-long stiff molecular handle that shows no mechanical response, hence, allowing the direct measurement of the dynamics of the protein of interest. While the biotin-streptavidin tethering at the C-terminus is stable enough to withstand typical folding forces ($<$15 pN) during extended

7

A

```
{"Pulse_Number_0":{
    "Meta_Data": {
        "Pulse Protocol": "7/10800/C;",
        "date (s)": "01/14/2020;",
    },
    "time": [
        758.685,
        758.725,
        (...)
    ],
    "z": [
        29.505,
        19.050,
        (...)
    ],
    "force": [
        7.190,
        7.396,
        (...)
    ],
},"Pulse_Number_1":{
    "Meta_Data": {
        (...)
}}
```

B
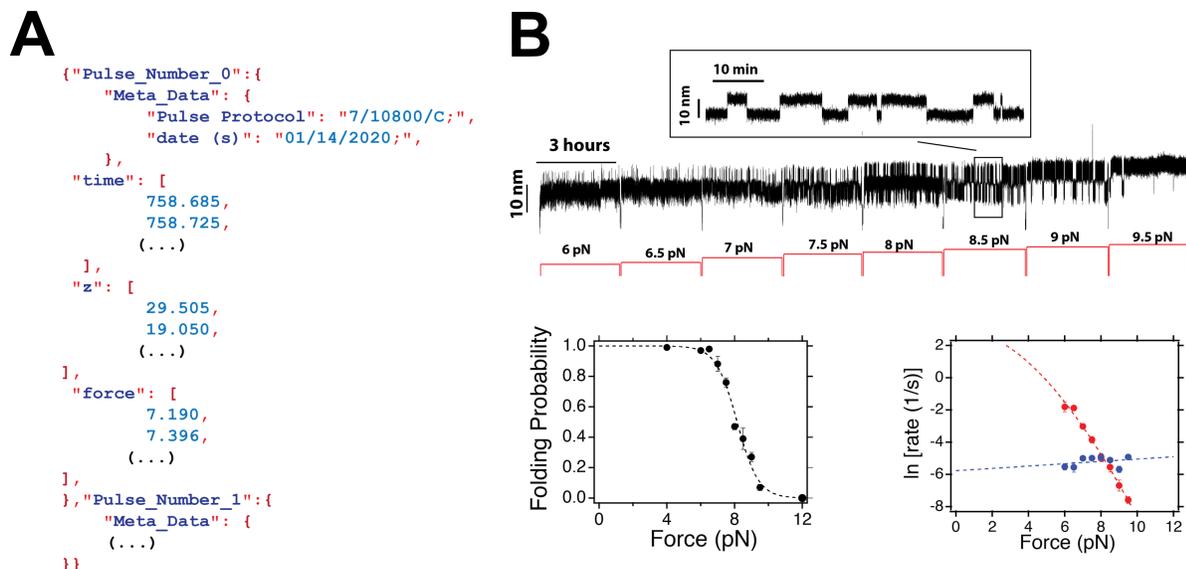
Figure 2: **A JSON-based file format for automatic single-protein data analysis**: (A) Architecture of the JSON-based single-molecule data format. The file is structured in attribute-values ("keys") according to the experimental force-pulse protocol. Each attribute contains the data as arrays with the time mark ("time"), protein extension ("z"), and force ("force"). (B) Force dependency of protein L folding dynamics measured through JSON-pulse protocols. One day-long recording, exploring different forces for 3 hours, following JSON-based pulses (upper panel). Given the rational organization of the data based on the utilized experimental protocol, it is possible to automatically access and analyze each fragment following the JSON keys. The folding probability and folding (red) and unfolding (blue) rates are automatically analyzed from the JSON file format (lower panel). This strategy enables the manipulation and analysis of large data sets with minimal human intervention.

periods of time, we have developed more enduring strategies that involve HaloTag anchors at both termini, which permit to pull at arbitrarily high forces.[23,25] We include in the SI Appendix (Fig. S5) a detailed vector map for cloning and detailed protocols for protein cloning, expression, and purification.

Data Acquisition and Analysis

We have developed a custom-written software in C++/Qt for data acquisition. Briefly, at the start of an experiment, we build a z-stack library of the magnetic and reference beads taking 20 nm steps using the piezoelectric actuator with a total of 128 images. Then, the radial vector of each image is calculated from the Fourier Transform with a pixel-addressing algorithm.[17] In the course of an experiment, the z-extension of the molecule is computed in

8

real time by correlating the radial vectors of the magnetic and reference beads with the stack library and calculating the optimal Gaussian fit of the Pearson correlation coefficients using Caruana's algorithm.[28] The use of an analytical method instead of a minimum-square fit greatly accelerates the image processing algorithm, which, with our current implementation, can reach sampling rates of $\sim$1.7 kHz ($\sim$580 μs). The software writes in real time a sequential file containing the time mark, protein extension, and force stored in binary format. The file can be accessed using an independent software (we employ mostly Igor Pro 8, Wavemetrics) to visualize in real-time the course of the measurement without interfering with the data acquisition process. We include as part of our SI Appendix the source code for the data acquisition, which has been uploaded to the online repository GitHub.

The combination of stability and the high sampling density of our instrument results in large volumes of data, which, for instance, can easily reach $\sim$6$\times$10^8 data points for a single protein measured over a few days. Such large volume of data can be arduous to handle and analyze manually, raising the same kind of problems that the Big Data field deals with. To this aim, we have developed a JSON-based file format for single-molecule data, which enables the direct access to specific experiment fragments for its automatic analysis. JSON (JavaScript Open Notation) is an open source file format compatible with most programming languages and software—including Python, C++, Matlab, or Igor Pro— that organizes the data by attribute-value pairs or "keys", which enables to automatically load these data portions instead of sequentially accessing the whole file. In our case, the architecture of our file format uses keys defined by the pulse protocols employed to measure the molecule following a "F/T/flag" format, where F is the employed force in pN, T the duration of the pulse in seconds, and flag either "C" for constant force, or "L" for a linear force ramp. Under a protocol key, the data is stored hierarchically as arrays identified by their key, "time" (time mark in s), "z" (protein extension in nm), and "force" (applied force in pN). Figure 2A shows a representation of our JSON-based file format, which illustrates its architecture. With this implementation, days-long experiments are stored in a rational

format, which allows processing large amounts of data in an automatic way and with minimal human intervention.

As an illustration of the potential of the JSON format approach, Fig. 2B shows the fragment of a protein L recording, aimed at directly characterizing the force-dependency of its folding dynamics. This 24 hours-long recording is organized in eight 3 hours-long JSON pulses, each exploring a different force between 6 and 9.5 pN, covering the full range over which protein L undergoes reversible folding dynamics. By storing this recording in our JSON-file format, the experimental data can be automatically accessed and analyzed according to the employed pulse, in this case the explored force. Hence, the folding properties of protein L, portrayed by its folding equilibrium and folding kinetics (Fig. 2B, lower panel), can be automatically calculated by simply analyzing each pulse fragment.

Results

Our experimental technique permits the continuous measurement of protein folding dynamics during extended timescales and at high resolution. To illustrate our approach, we character-ize the folding dynamics of the talin R3^{IVVI} and protein L using a single molecular individual. This allows us to compare the nanomechanical performance of different molecular specimens and directly observe the extent of diversity between their nanomechanical performances.

Nanomechanics of the Talin R3^{IVVI} domain

To characterize the nanomechanics of the R3^{IVVI} domain, we collect a single 27 hour-long recording on the same single protein, employing the molecular construct discussed above (Fig. 3A). The fast folding dynamics of this protein allow accumulating over 70,000 folding and unfolding transitions along the whole experiment, and fully characterize the force-dependence of its folding equilibrium. Figure 3B shows 50 seconds-long fragments of R3^{IVVI} folding dynamics at three different forces. These transitions between the folded (low extension) and

10
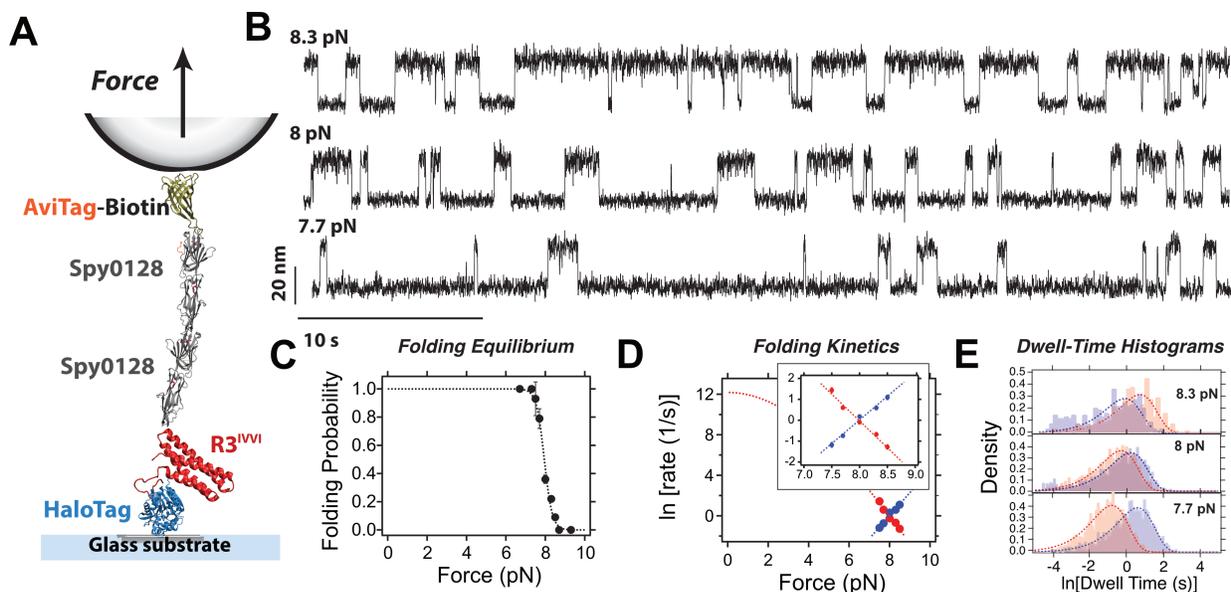
Figure 3: **Complete nanomechanical mapping of R3$^{\text{IVVI}}$:** (A) Schematics of the magnetic tweezers and anchoring strategy for R3$^{\text{IVVI}}$. (B) Fragments of the measured trajectory at three different forces, 8.3 pN, 8 pN, and 7.7 pN. R3$^{\text{IVVI}}$ transitions in equilibrium between the folded (low extension) and unfolded (high extension) over a narrow force range, giving rise to changes on its extension of ~ 20 nm. (C) Folding probability of R3$^{\text{IVVI}}$. The folding probability drops with force in a sigmoidal fashion over a force range of $\sim$1.5 pN. Fitting parameters $F_{1/2} = 7.95 \pm 0.03$ pN, $w = 0.20 \pm 0.02$ pN. (D) Folding (red) and unfolding (blue) rates of R3$^{\text{IVVI}}$ as a function of force. The scale employed highlights the narrow range that the rates span. (Inset) Detail of the folding rates over the 7-9 pN range. The folding rates follow a nearly exponential force dependence given the narrow force range that is explored. Fitting to Bell model yields $k_U^0 = (1.38 \pm 0.15) \times 10^{-8}$ s^{-1} and $x_U^\dagger = 9.29 \pm 0.99$ nm (unfolding) and $k_F^0 = (8.00 \pm 0.80) \times 10^8$ s^{-1} and $x_F^\dagger = 10.52 \pm 0.98$ nm. (E) Dwell-time histograms calculated with a logarithmic binning for 8.3 pN, 8 pN, and 7.7 pN (red, folded state; blue, unfolded state). We observe a single-exponential dependence at every force.

unfolded (high extension) states yield $\sim$20 nm changes on the protein extension, which gives rise to the characteristic "hopping" dynamics. As displayed in Fig. 3B, R3$^{\text{IVVI}}$ is highly sensitive to the pulling force, and changes of a fraction of a pN are sufficient to shift its folding equilibrium considerably.

We first characterize the force response of R3$^{\text{IVVI}}$ folding equilibrium and kinetics. Using 20 min-long JSON pulses between 7.3 pN and 8.7 pN, we calculate the folding probability (Fig. 3C) and the folding/unfolding rates (Fig. 3D). The folding probability drops steeply in a sigmoidal fashion, with a coexistence force of $F_{1/2}$=7.95$\pm$0.03 pN ($F_{1/2}$, force at which the folded and unfolded states are equally populated). Due to this sharp force response, the

11

folding/unfolding rates have a nearly-exponential dependence with force (Fig. 3D, inset), with very similar slopes. While the force dependence of the unfolding rates of a protein is usually well described by the Bell model,[29] the folding rates r_F must account for the entropic elasticity of the unfolded state, which dominates the refolding transition.[24,30,31] The force-dependence of the folding rates can thus be described by integrating over an energy profile dominated by the freely-jointed chain (FJC) model, leading to:[24,32]

$$r_F(F) = k_0 \exp\left[-\frac{\Delta L_c}{l_K}\ln\left|\sinh\left(\frac{Fl_K}{kT}\right)\right| - \ln\left(\frac{Fl_K}{kT}\right)\right], \tag{2}$$

where is k_0 the unfolding rate extrapolated to zero-force, ΔL_c the increment in contour length, and l_K the Kuhn length. In the case of R3^{IVVI}, due to the narrow range of forces over which it folds/unfolds in equilibrium, the elastic contribution is nearly negligible, as seen in Fig. 3D, and the simple exponential Bell theory is a good approximation.

We continue our recording by studying in detail the folding dynamics at three different forces, 7.7 pN, 8 pN, and 8.3 pN, measuring six-hours-long pulses at each force. Figure 3E shows the dwell-time histograms in the folded (red) and unfolded (blue) states calculated with a logarithmic binning. Under this representation, standard in ion-channel analysis, a single-exponential distribution is transformed into a peaked distribution as $p(x) = \exp[x - x_0 - \exp(x - x_0)]$, where $x = \ln(t)$, and $x_0 = \ln(\tau)$, being τ the average (folding or unfolding) time, which is the peak of the distribution.[19,33] This transformation is useful to systematically identify multiple underlying kinetic processes, which can be hidden when representing the dwell-time distributions with linear binning. In our case, the three histograms fit a single exponential distribution, which indicates that the observed dynamics correspond to a simple Poisson process governed by a single folding/unfolding rate.

Our instrumental design balances the ability to measure single proteins for extended times and sampling its dynamics at a high temporal resolution. This recording was measured at 1352 Hz, which is sufficient to capture how the protein diffuses along its free energy profile and reconstruct this landscape by resolving low occupation regions utilizing thousands of

12
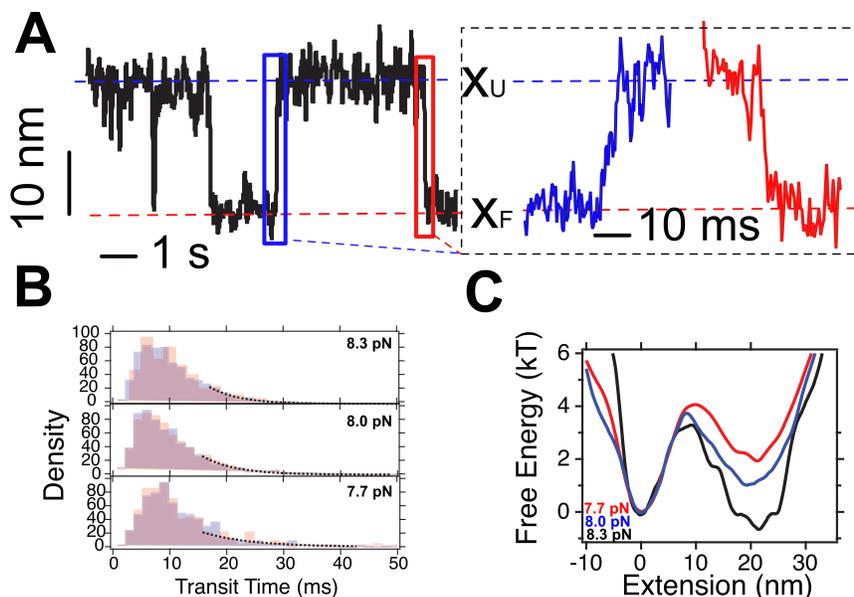
Figure 4: **Transition paths and free energy landscape reconstruction**: (A) Detail of R3$^{\mathrm{IVVI}}$ unfolding (blue) and folding (red) transition, with the unfolding (blue) and folding (red) transition paths highlighted. (B) Transition path time histograms at 8.3 pN, 8 pN, and 7.7 pN. The folded (red) and unfolded (blue) transit time histograms overlap, as expected from time reversal symmetry. The transition path times are insensitive to the magnitude of the pulling force ($\tau_{\mathrm{TP}} \approx 11$ ms), over the explored range. (C) Free energy landscape of R3$^{\mathrm{IVVI}}$ at 7.7 pN (red), 8 pN (blue) and 8.3 pN (black).

transitions sampled at a high rate. Figure 4A shows a detail of the folding transitions of the measured R3$^{\mathrm{IVVI}}$. While a protein in equilibrium spends the majority of its time dwelling over the folded and unfolded basins, the brief excursions to switch between these states contain the most relevant information about a protein's free energy landscape—dubbed the transition paths.[34,35] We extract the transition paths for the folding and unfolding reactions at each force by identifying the fragments of trajectory where the protein escapes from the folded/unfolded basin (defined by their extensions, x_F or x_U) to directly reach the unfolded/folded one (Fig. 4A, right). We measure the duration of the transition paths at each force (8.3 pN, 8 pN, and 7.7 pN) and calculate the histograms of transit times (Fig. 4B) for the folding (red) and unfolding (blue) transitions. At each force, the folding and unfolding transit time distributions are equal within error, as expected from the time-reversal symmetry.[36] Unlike the folding and unfolding rates, which depend exponentially on the free energy barrier height $\Delta G^{\dagger}$ and, hence, have a strong force-dependence, the transit times

13

are dominated by the diffusion coefficient D, and the effect of the force is negligible. Assuming a harmonic barrier in the high-barrier limit, the average transition path time τ_{TP} can be expressed as $\tau_{\text{TP}} \approx kT \ln \left(2e^{\gamma} \Delta G^{\dagger}/kT\right)/(D\kappa_b)$, where κ_b the stiffness of the barrier top (assuming a harmonic barrier), and $\gamma \approx 0.577...$ the Euler-Mascheroni constant.[34] As seen in Fig. 4B, we do not observe any appreciable change on the transition path time histograms with force, being the average transit time at every force $\tau_{\text{TP}} \approx 11$ ms. Following this approximation, the transition path time distribution $P_{\text{TP}}(t)$ decays exponentially as $P_{\text{TP}}(t) \approx \frac{2\omega_K \Delta G^{\dagger}}{kT} \exp\left(-\omega_K t\right)$, where $\omega_K = D\kappa_b$.[34,35] Both expressions provide a robust method to estimate the diffusion coefficient, given that the properties of the free energy landscape are known.

To this aim, we reconstruct the free energy landscape $G(x)$ at 7.7 pN, 8 pN, and 8.3 pN from the distribution probability of the molecular extension $p(x)$ and its time derivative $\dot{x}$ as:

$$G(x) = -kT \ln p(x) - \frac{kT}{D} \int_{x_1}^{x_2} \dot{x}(x)dx \tag{3}$$

being x_1 and x_2 the range of x. This equation accounts for the equilibrium contribution (first term, Boltzmann inversion) and the non-equilibrium contribution (second term).[37,38] While this latter term is crucial when employing non-equilibrium trajectories, in our case, given the symmetry between the folding and unfolding transition paths, it is negligible. In practice, as it has been extensively discussed,[39,40] the measured dynamics do not correspond to the real molecular dynamics, but include the convolved instrumental influence, so the direct application of Eq. (3) on the measured signal does not result in the actual molecular free energy landscape. In such a case, the molecular free energy landscape is calculated by estimating an instrumental point-spread-function and using a deconvolution algorithm to remove the instrumental effect.[39] Here, and unlike the case of optical tweezers and AFM, we do not have a direct effect of the force probe due to the negligible stiffness of the magnetic trap ($\ll 10^{-6}$ pN/nm) that results in the hallmark force-clamp conditions of magnetic

14

tweezers. However, the image analysis algorithm we employ to infer the molecular extension introduces instrumental noise to our recordings that smears the actual protein dynamics. To deconvolve this effect, we estimate the point-spread-function of our experiment by measuring long recordings of two non-magnetic beads, which are tightly attached to the glass surface and do not respond to force, hence, mimicking the noise introduced by our image analysis algorithm. As shown in Fig. S6, the point-spread-function is a Gaussian distribution with a standard deviation of 1.7 nm.

Figure 4C shows the free energy profiles at 7.7 pN, 8 pN, and 8.3 pN calculated with the described procedure. Force tilts the free energy landscapes towards the unfolded state, hence, stabilizing this conformation. Due to the small range of forces explored by $R3^{\text{IVVI}}$, we only observe a small increase in the extension change upon unfolding (19.5 nm at 7.7 pN, and 20.3 nm at 8.3 pN). However, we readily observe the differences in the width of the unfolded basin that arise from the gain in contour length, leading to a very different stiffness, $\kappa_F = 0.28 \pm 0.02$ pN/nm (folded) and $\kappa_U = 0.08 \pm 0.01$ pN/nm (unfolded) (at 8 pN, see SI Appendix for the free energy landscapes properties at all forces). In a simple two-state scheme, to a first approximation, the force is assumed to act linearly with a $-F \cdot x$ term. The landscapes estimated for $R3^{\text{IVVI}}$ are in agreement with this approximation, as subtracting the landscapes at 8.3 pN and 7.7 pN recovers this linear contribution, which has a slope of ~ 0.6 pN, the difference in force between these two landscapes (Fig. S7).

From the properties of the free energy landscapes, we can estimate the diffusion coefficient of the measured dynamics in two different ways. First, from the average transition time and since the stiffness of the free energy barrier is $\kappa_b = 0.37 \pm 0.04$ pN/nm and the barrier height $\Delta G_F^{\dagger} = 3.72 \pm 0.05$ kT (using the data at 8 pN), we obtain a first estimation of the diffusion coefficient $D \approx 2870 \pm 690$ nm^2/s. Second, by fitting the tail of the transition path time distributions (knowing $\Delta G_F^{\dagger}$), we obtain that the fundamental frequency is $\omega_K = 273.3 \pm 4.2$ s^{-1}, which leads to $D \approx 3032 \pm 331$ nm^2/s. In this sense, both methods provide very similar estimates for the diffusion coefficient. With this value of the diffusion coefficient

15

and the properties of the free energy landscape, we can estimate the folding and unfolding rates using Kramers' reaction rate,[41] which yields $r_U \approx 0.86$ s^{-1} and $r_F \approx 1.28$ s^{-1}, in excellent agreement with the measured rates (Fig. 3D), which supports the consistency of the estimated free energy landscapes and values of the diffusion coefficient. Notably, the diffusion coefficient of a superparamagnetic bead is over two orders of magnitude larger than the diffusion coefficient of R3$^{\text{IVVI}}$ dynamics, which suggest that the protein folding dynamics are rate-limiting, as previously suggested.[22]

Nanomechanics of protein L

Protein L—a superantigen bacterial domain with a topology that comprises four β strands packed against an α helix—is a common protein folding model, and has been used as a molecular template to calibrate single-molecule instrumentation.[16,17,19] Protein L has slow folding dynamics between forces of 4 and 10 pN, which arise from the hysteresis of its folding and unfolding forces—unfolding quickly at forces >40 pN, while collapsing and folding fast below 4 pN.[42] Owing to the slow kinetics of protein L folding dynamics, its nanomechanics have been mostly characterized using chimeric polyprotein constructs that amplify the signal. To characterize in detail the folding dynamics of a single protein L individual, we measure one protein L monomer for five uninterrupted days (Fig. 5A). Our full recording comprises initial short pulses at high force to measure its unfolding kinetics and long JSON-pulses ($\sim$8 hours) at low force to characterize its folding properties. Figure 5B shows short fragments of protein L dynamics at 7.5 pN, 6 pN, and 4.5 pN, which highlights the much slower folding kinetics compared to R3$^{\text{IVVI}}$. Since protein L exhibits reversible folding dynamics over a wide force range, we can readily observe the impact of polymer elasticity on its folding mechanism as the force scaling of its extension changes (step sizes), being $\sim$5 nm at 4.5 pN, while $\sim$8.5 nm at 7.5 pN.

Similar to the scheme developed in Fig. 3, we characterize first the force-response of protein L through the folding probability (Fig. 5C) and the folding/unfolding rates (Fig.

16
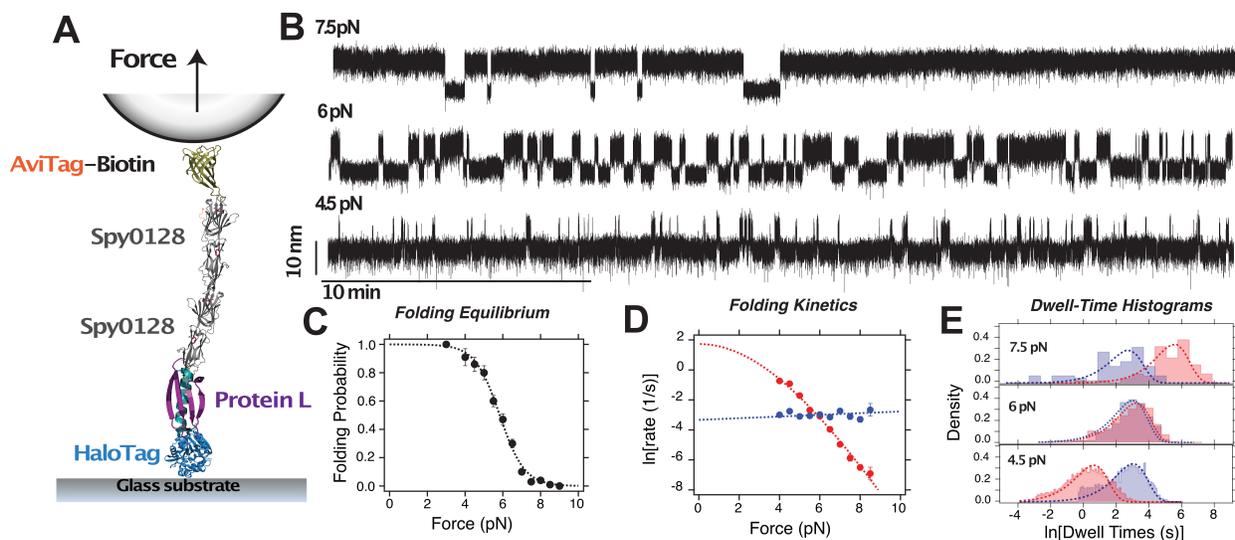
Figure 5: **Complete nanomechanical mapping of a single protein L domain:** (A) Schematics of the magnetic tweezers and anchoring strategy for our protein L experiments, identical to the one employed for R3$^{\text{IVVI}}$. (B) Fragments of the measured trajectory at three different forces, 7.5 pN, 6 pN, and 4.5 pN. In contrast to R3$^{\text{IVVI}}$, protein L folding dynamics explore a much wider range of forces with slower kinetics, that lead to completely different nanomechanical properties. (C) Folding probability of the protein L molecule. The folding probability drops with force following a sigmoidal dependence, showing reversible folding dynamics between 4 pN and 9 pN, approximately. Fitting parameters $F_{1/2} = 5.83 \pm 0.04$ pN, $w = 0.65 \pm 0.04$ pN (D) Folding (red) and unfolding (blue) rates of the protein L molecule as a function of force. The force dependency of the rates is strongly asymmetric; the folding rates drop quickly with force, while the unfolding rates have a nearly flat force-dependency. Additionally, the folding rates show a non-exponential behavior due to the contribution of the elasticity of the unfolded state on the refolding transition. Fitting the folding rates to Eq. (2) leads to $k_0^F = 4.97 \pm 1.00$ s^{-1}, $L_c = 14.82 \pm 6.4$ nm, and $l_K = 0.97 \pm 0.52$ nm, these latter two parameters in great agreement with the known FJC properties of protein L. Fitting the unfolding rates to the Bell model yields $k_0^U = 0.036 \pm 0.008$ s^{-1} and $x^\dagger = 0.23 \pm 0.02$ nm. (E) Dwell-time histograms calculated with a logarithmic binning for 7.5 pN, 6 pN, and 4.5 pN (red, folded state; blue, unfolded state). We observe a single-exponential dependence at every force.

5D). The folding probability also has a sigmoidal dependence, but with a shallower force dependence, showing a coexistence force of $F_{1/2}$=5.83 ± 0.04 pN. The difference between the nanomechanics of R3$^{\text{IVVI}}$ and protein L is distinctly reflected in the force-dependency of its folding and unfolding rates that, as shown in Fig. 5D, are highly asymmetric. The unfolding rates (blue) are nearly flat over the folding range, which, accounting for its rates measured at high forces (not shown), leads to a distance to the transition state of $x^\dagger = 0.23 \pm 0.02$ nm, assuming Bell model. By contrast, the folding rates (red) drop very quickly with force, following a non-exponential shape as predicted by Eq. (2), readily suggesting that the folding

17
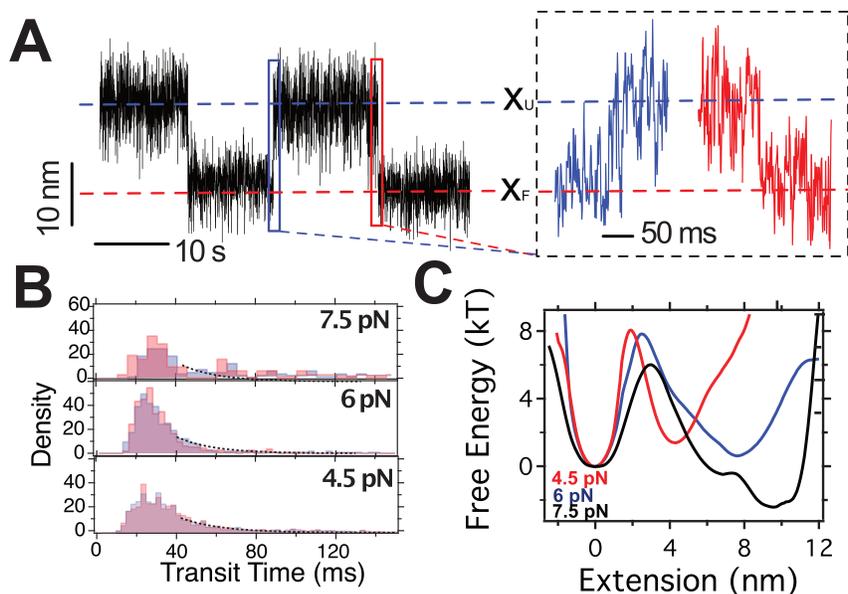
Figure 6: **Transition path times and free energy reconstruction of protein L:** (A) Detail of protein L unfolding (blue) and folding (red) transitions. (B) Transition path time histograms at 7.5 pN, 6 pN, and 4.5 pN. The folded (red) and unfolded (blue) histograms overlap, as expected from time reversal symmetry. The average transition time is nearly 3 times higher than those of R3$^{\text{IVVI}}$ ($\tau_{TP} \approx 33$ ms). (C) Free energy landscapes at 4.5 pN (red), 6 pN (blue) and 7.5 pN (black). The shape of the free energy landscapes and how they are modulated by force illustrate that the folding mechanics of protein L are not well-understood with a simple two-state model. The unfolded state position shifts with force following the non-linear FJC model, and also widening significantly with force. This indicates that the effect of the force on the folding landscape cannot be ascribed to a simple linear tilt.

transition is dominated by the entropic elasticity of the extended polypeptide chain. The dwell-time histograms (Fig. 5E) are exponentially distributed, with the dwell-time in the folded state (blue) remaining nearly unchanged, while the time spent in the unfolded state (red) increases very quickly with force.

Figure 6 shows a detail of protein L folding dynamics. Although these dynamics appear similar to those observed for R3$^{\text{IVVI}}$, the folding mechanism of protein L strongly differs. To unveil this process, we extract the transition paths, and calculate the transition path time histograms (Fig. 6B), which also fulfil the time-reversal conditions and maintain an apparent insensitivity with force. However, the average transition times increase considerably for protein L ($\tau_{\text{TP}} \approx 33$ ms) compared to R3$^{\text{IVVI}}$, which could suggest slower diffusive dynamics.

Following the deconvolution procedure, we reconstruct the free energy landscape at the

18

measured forces. Figure 6C shows the free energy landscapes along the molecular extension at 4.5 pN (red), 6 pN (blue) and 7.5 pN (black), which manifest the great differences between $R3^{IVVI}$ and protein L folding under force. First, the reconstructed free energy landscape captures both the asymmetry of the landscape and the notable shift in the position of the unfolded state, which scales non-linearly following the FJC, with ~ 5 nm (4.5 pN) and ~ 9 nm (7.5 pN). In this regard, it is clear that a linear approximation does not account for the actual effect of the force on the free energy landscape, and subtracting the landscapes leads to non-linear contributions that do not resemble the classical linear tilt (SI Appendix, Fig. S7). When estimating the diffusion coefficient D, we obtain surprisingly low values of $D \approx 71.9 \pm 4.7$ nm^2/s (from τ_{TP}), and $D \approx 67.4 \pm 3.5$ nm^2/s (from the fits to P_{TP}). Using these estimations, the Kramer's rate at 6 pN is $r \approx 0.0057$ s^{-1}, while the measured ones are nearly 10 times higher. This discrepancy arises likely due to protein L folding mechanism, which is dominated by the entropic elasticity of the unfolded state, eventually leading to the failure of theories developed for simple analytic two-state landscapes.

Characterizing single-protein nanomechanical diversity

While single-molecule approaches enable to interrogate individual biomolecules, it is typically necessary to assess several molecular specimens in order to accumulate enough data for a robust statistical significance. This approach is intrinsically based on the ergodic hypothesis, which states that the time average of a physical observable equals its ensemble average.[43] Here, we have undertaken this first approach, and thoroughly mapped the nanomechanics of two different proteins by measuring their dynamics for long timescales and time-averaging their properties. However, an inherent assumption of the ergodic hypothesis is that all the individual components of the ensemble are identical entities so that all trajectories employed for the ensemble average transverse the same phase space. Here, we can directly test this hypothesis and compare the nanomechanical performance of different single molecules to look into their functional heterogeneity.

19
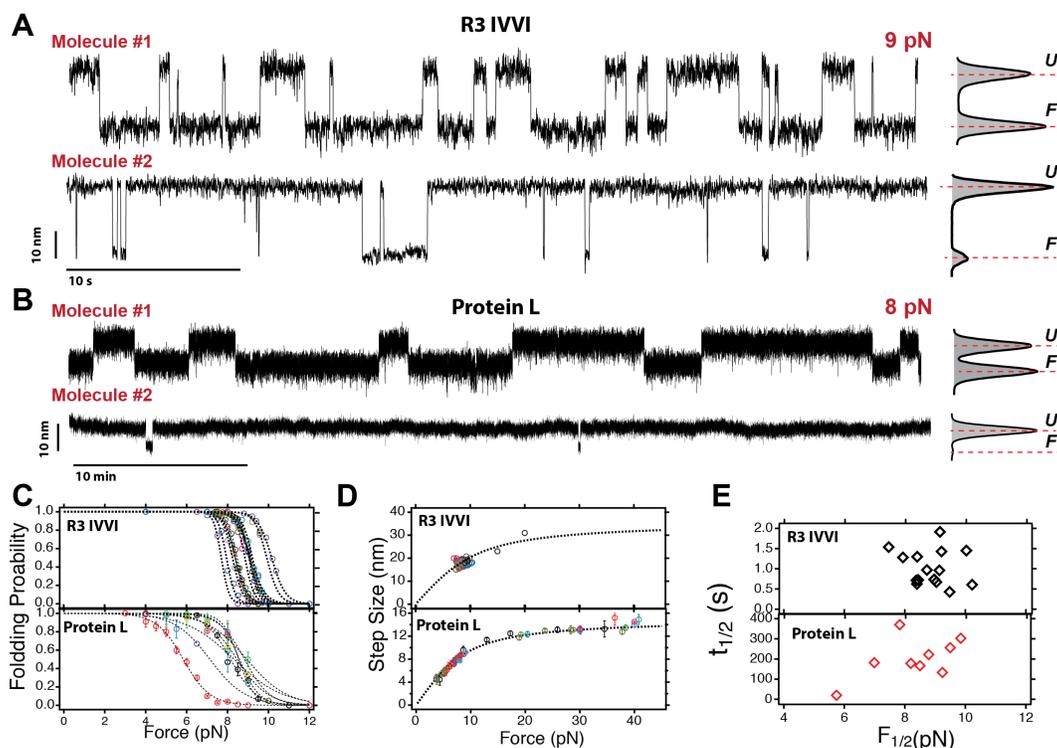
Figure 7: **Characterization of the nanomechanical identity among protein individuals:** (A) Fragment of R3$^{\text{IVVI}}$ folding mechanics close to $F_{1/2}$ measured for two different R3$^{\text{IVVI}}$ molecules. (B) Fragment of protein L folding mechanics close to $F_{1/2}$ measured for two different protein L molecules. (C) Folding probabilities for R3$^{\text{IVVI}}$ (upper) and protein L (lower) measured over different protein individuals (each color/symbol is a different molecule). We observe in both cases a dispersion in the folding probability force dependency, which is more evident in the case of protein L. (D) Step-sizes as a function of the pulling force for R3$^{\text{IVVI}}$ (upper) and protein L (lower). While it is only possible to measure the step-sizes of R3$^{\text{IVVI}}$ over a narrow range of forces, all the protein L molecules have step-sizes that fall on a master FJC curve, which indicates that our setup applies well-calibrated and reproducible forces, which do not explain the dispersion in the nanomechanical properties we observe. (E) Scatter plot of the coexistence force $F_{1/2}$ and average coexistence dwell-times $t_{1/2}$ for R3$^{\text{IVVI}}$ (black) and protein L (red).

We measure several individuals of R3$^{\text{IVVI}}$ and protein L and compare their response to force. Figures 7A and B show the folding dynamics of R3$^{\text{IVVI}}$ (A) and protein L (B) for two different protein specimens subject to the same pulling force. Interestingly, each protein individual shows a remarkably different behavior at the same force, populating differently the folded/unfolded states and exhibiting different transition kinetics. Hence, the comparison of these recordings readily suggests a degree of dispersion in the nanomechanical properties of each protein L and R3$^{\text{IVVI}}$ molecule.

To further characterize this observation, we measure the folding probability of several

20

R3$^{\text{IVVI}}$ (16 molecules) and protein L (9 molecules) (Fig. 7C), which readily shows the heterogeneity in the mechanics of each protein individual. In the case of R3$^{\text{IVVI}}$, the folding probabilities have a sigmoidal dependence with $F_{1/2}$ that spans between ~ 8 pN and $\sim$10 pN, while for protein L the dispersion in $F_{1/2}$ spans between $\sim$6 pN and $\sim$10 pN. Notably, the differences in the folding probabilities do not correspond to simple shifts of the sigmoidal curve, but their shape also differs amongst them, showing a different force-dependency over the full folding range. To demonstrate that these differences do not arise from a poor calibration of our setup or from the size-dispersion in the magnetic beads (M-270 beads have a size dispersion of $\sim$2 %), we measure the step sizes of R3$^{\text{IVVI}}$ and protein L. This molecular observable depends on the contour length of the molecules, this is, the number of residues in the protein sequence, and, as previously demonstrated, is an extremely force-sensitive metric that can be employed as a calibration strategy.[17,19] Figure 7D shows the step sizes for R3$^{\text{IVVI}}$ (upper) and protein L (lower) as a function of force. Unfortunately, R3$^{\text{IVVI}}$ explores a minimal range of forces, and its step sizes can only be ascertained over a narrow force range, since R3$^{\text{IVVI}}$ unfolding becomes too fast at forces above 15-20 pN to resolve the event. By contrast, protein L, owing to its slow unfolding kinetics, allows us to measure its step sizes between 4 and 40 pN, and fully account for their force dependency. In contrast to the folding probability, the step sizes of every protein L molecule fall on the same curve given by the FJC model with the previously determined properties, $\Delta L_c = 16.3$ nm and $l_K = 1.1$ nm, readily indicating that our force calibration and bead-to-bead variations allow us to apply forces in an accurate and reproducible way and that the observed dispersion is an intrinsic property of the measured molecules.

To portray the nanomechanical heterogeneity among different R3$^{\text{IVVI}}$ and protein L individuals, we measure their folding/unfolding kinetics and determine their coexistence force $F_{1/2}$ and their dwell times at this force $t_{1/2}$. To accurately estimate these values, we obtain them from the intersection point between the folding/unfolding rates by fitting their force-dependency (SI Appendix, Figs. S8 and S9). Figure 7E shows a scatter plot of $F_{1/2}$

21

and $t_{1/2}$ that illustrates the heterogeneity on the folding mechanics of each measured protein. In both cases, we observe a significant dispersion in the values of $F_{1/2}$ and $t_{1/2}$, more pronounced in the case of protein L. R3$^{\text{IVVI}}$ molecules have an average coexistence force of $\langle F_{1/2} \rangle \approx 8.7$ pN and $\langle t_{1/2} \rangle \approx 1.3$ s, both values in agreement with previous work,[44] while the dispersion of these averages (standard-deviations) is $\delta F_{1/2} \approx 1$ pN, $\delta t_{1/2} \approx 0.75$ s. For protein L, we measure $\langle F_{1/2} \rangle \approx 8$ pN, and $\langle t_{1/2} \rangle \approx 200$ s, with a dispersion of $\delta F_{1/2} \approx 2.4$ pN, $\delta t_{1/2} \approx 150$ s. Interestingly, the average properties of protein L are in agreement with measurements of the protein L octamer, which could suggest that the use of polyproteins offers a "pre-averaged" signal, that reduces the apparent nanomechanical heterogenity.[42] Overall, while the step sizes, an intrinsic property of the polypeptide chain, are homogeneous among different molecules, the properties related to their folding landscape show a high degree of heterogeneity.

Discussion

The inception of single-molecule techniques opened the gates to reconstructing distributions of molecular properties, and appreciate the dynamic and heterogeneous nature of biomolecular behaviors. These advances offered valuable insight on the functional heterogeneity in diverse biomolecular systems, such as enzymatic reactions, ligand binding, or molecular motors.[45–47] Force-clamp atomic force microscopy (AFM) experiments evidenced the existence of protein heterogeneity in poly-ubiquitin molecules, which exhibited non-exponential unfolding kinetics likely due to a distribution of free energy barrier heights—a manifestation of the so-called static disorder in the Zwanzig's terminology.[5] More recently, a systematic study of several single-molecule force spectroscopy data-sets suggested a widespread mechanical heterogeneity in different protein, nucleic-acid, and complex-ligand systems.[48] While these works offered indirect evidence of functional heterogeneity among single proteins, instrumental limitations have prevented to directly address this question, which would require a direct

comparison between the properties obtained from time averages over long trajectories and those that arise from ensemble averages over several molecules; this is, to test if the ergodic hypothesis holds in protein nanomechanics.[43]

Our instrumental development opens up the gates to profile the nanomechanics of an individual protein by collecting extensive recordings of its dynamics and accumulating enough events to fully portray its molecular properties. Due to this capability, we have directly captured the mechanical diversity of two protein domains, the talin R3[IVVI] domain and protein L. This heterogeneity is manifested as a dispersion of their coexistence forces $F_{1/2}$ and their folding/unfolding kinetics. While the potential impact of such protein diversity in a physiological setting is unclear, it is remarkable to note how the functional heterogeneity of RNA molecules has been widely reported, while this observation on single proteins has remained far more elusive.[49–51] Hence, it seems evident that the diversity of RNA phenotypes is far more obvious, compared to that of proteins. It is enticing to think that RNA is a more "immature" and ephemeral biomolecule (from the evolutionary perspective), while proteins have an optimized structure and their synthesis undergoes further proofreading process, which could narrow the distribution of functional molecules. However, a protein's response to a stretching force is a key determinant of many biological processes. For example, in the case of talin, its folding status regulates the binding of ligand partners such as vinculin.[44,52] More generally, the mechanical stability of proteins has been suggested to tune its import rate through the nuclear pore complex, the gateway of nuclear translocation.[53] Hence, it could be argued that uniform mechanical properties are a desirable molecular property, specially in the case of mechanosensing proteins. However, it remains an open question whether the observed mechanical diversity also applies to proteins operating in the cell, or biological control processes can further filter protein populations by its mechanical properties.

From the physical perspective, the exact origin for such nanomechanical dispersion is uncertain. As some theoretical and computational protein folding models have explored, biomolecular free energy landscapes can be highly rugged, containing a degenerate ground

23

state with multiple functional minima separated by large barriers.[5,48,54] In such a case, while bulk techniques would measure the average conformation of such ensemble of states, single-molecule approaches would sample independently biomolecules "frozen" at each individual state. However, the observed dynamics are robust and each individual protein maintains, within statistical significance, a "nanomechanical" identity. This would imply that, if the rugged landscape model is correct, such multiple states are separated by extremely high barriers, inaccessible even over timescales of several days. Additionally, a protein is a dynamic and evolving molecule, which is likely to undergo multiple and random chemical modifications that can alter in a more or less significant way its free energy landscape. As we previously demonstrated, mechanical stretching exposes cryptic residues that can be irreversibly oxidized, which eventually blocks protein folding and leaves an unstructured and labile polypeptide.[42,55] By contrast, these random chemical modifications on residues natively exposed to the solvent are less detrimental on the protein's fold, while, however, could still render a modified, yet functional, landscape.[56] Importantly, our data clearly indicates that the length of the protein sequence is a constant among the different molecular specimens, as demonstrated by the high homogeneity of the polymer elasticity properties of the unfolding protein L domains, which also verifies our ability to apply calibrated forces and the unambiguous fingerprint for our tethers.

Our data further demonstrates the key role of polymer physics on protein folding under force. The entropic elasticity of the unfolded polypeptide dominates the refolding transition, owing to the mechanical work that protein folding must do in order to collapse against the pulling force.[24,30,57] The most obvious manifestation of this physical property is the change in the step-sizes with force following standard polymer physics models such as the FJC.[58] Another prediction from the FJC that our recordings readily show is the increase in the magnitude of the end-to-end length fluctuations upon unfolding, due to the gain in contour length that results in a wider free energy basin. This property is captured now in the reconstructed free energy landscapes (Figs. 4 and 6), where the stiffness of the unfolded

24

state is reduced by up to 70% for R3$^{\text{IVVI}}$ (contour length increment $\Delta L_{\text{c}} \sim 40$ nm) and 55% for protein L (contour length increment $\Delta L_{\text{c}} \sim 16.3$ nm). The recognition of the fundamental role of polymer physics on a protein's folding landscape has often remained disregarded, likely since most of the studied proteins explore narrow force-ranges which make valid two-state-like theories based on linear approximations of the effect of force. In this sense, these properties are clearly reflected when comparing the nanomechanics of R3$^{\text{IVVI}}$ and protein L. Similar to nucleic-acid hairpins or proteins with fast hopping kinetics, R3$^{\text{IVVI}}$ shows reversible folding dynamics over a range of ~ 1 pN, which result in a negligible change on its step sizes. Its folding/unfolding rates have an exponential dependence following the Bell model and resulting in a good agreement of its properties with theories developed for simple bistable potentials, which indeed resemble its free energy landscape (Fig. 4). In this sense, R3$^{\text{IVVI}}$ folding is likely driven by the hydrophobic collapse of the unfolded polypeptide, with minimal contribution of short-range bonds, resulting in a compliant free energy landscape. By contrast, protein L, akin to other proteins with Ig-like or α/β folds, has a very cooperative and brittle fold, as reflected by a very short distance to transition state for unfolding (<1 nm). While the unfolding rates increase very slowly with force, the refolding rates have a very steep and non-exponential dependence with force, which can be well explained assuming that this transition is dominated by the free energy surface dictated by the FJC, as our fit to Eq. (2) suggests. As our free energy reconstruction demonstrates, the free energy landscape of protein L is indeed highly asymmetric, which readily demonstrates the non-linear effect of force in the shift of the unfolded state position due to the contribution of stretching an unfolded polypeptide. As a consequence of this contribution, the theories based on two-state-like descriptions of protein folding fail to explain protein L folding dynamics. In this sense, thanks to the compromise between high-resolution and stability of our magnetic tweezers setup, we have been able to reconstruct the nanomechanics of two different proteins and directly demonstrate the inarguable contribution of polymer physics in protein folding under force.

In summary, we have here introduced a novel magnetic tweezers development, which combines a simple and affordable design with improved resolution and stability that allows the full characterization of individual proteins. Our approach is grounded on our previous developments, implementing the use of a tape head for enhanced force control and stability, [19] and the use of HaloTag chemistry to achieve highly specific and long-term tethers. [17] The introduction of a high-precision mounting piece for the tape head results in an enhanced and reliable control of the pulling force that involves no moving parts, which further increases the instrumental stability. Furthermore, the simplified and dedicated microscope design, without relying on standard multi-functional microscopes, results in an increased signal-to-noise ratio and negligible drift. These properties combined enable the acquisition of long recordings that accumulate a high number of molecular events, and resolving at the same time the nuances of the free energy landscape. Importantly, our setup highlights its bespoke and affordable design, which facilitates its assembly and installation following the details and instructions of the SI Appendix. Hence, we anticipate that our magnetic tweezers approach has the potential to become a widespread tool to study biomolecules under force, and provide insights on the relevance of mechanical regulation in biological processes including cellular adhesion, mechanotransduction, or tissue mechanics.

Acknowledgement

References

(1) Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science* **1973**, *181*, 223–230.

(2) Leopold, P. E.; Montal, M.; Onuchic, J. N. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proceedings of the National Academy of Sciences* **1992**, *89*, 8721–8725.

(3) Dill, K. A.; MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science* **2012**, *338*, 1042–1046.

(4) James, L. C.; Tawfik, D. S. Conformational diversity and protein evolution – a 60-year-old hypothesis revisited. *Trends in Biochemical Sciences* **2003**, *28*, 361 – 368.

(5) Kuo, T.-L.; Garcia-Manyes, S.; Li, J.; Barel, I.; Lu, H.; Berne, B. J.; Urbakh, M.; Klafter, J.; Fernández, J. M. Probing static disorder in Arrhenius kinetics by single-molecule force spectroscopy. *Proceedings of the National Academy of Sciences* **2010**, *107*, 11336–11340.

(6) van Oijen, A. M.; Blainey, P. C.; Crampton, D. J.; Richardson, C. C.; Ellenberger, T.; Xie, X. S. Single-Molecule Kinetics of Exonuclease Reveal Base Dependence and Dynamic Disorder. *Science* **2003**, *301*, 1235–1238.

(7) English, B. P.; Min, W.; van Oijen, A. M.; Lee, K. T.; Luo, G.; Sun, H.; Cherayil, B. J.; Kou, S. C.; Xie, X. S. Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited. *Nature Chemical Biology* **2006**, *2*, 87–94.

(8) Ciftci, D.; Huysmans, G. H. M.; Wang, X.; He, C.; Terry, D.; Zhou, Z.; Fitzgerald, G.; Blanchard, S. C.; Boudker, O. Single-molecule transport kinetics of a glutamate transporter homolog shows static disorder. *Science Advances* **2020**, *6*.

(9) Neuman, K. C.; Nagy, A. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *5*, 491–505.

(10) Mora, M.; Stannard, A.; Garcia-Manyes, S. The nanomechanics of individual proteins. *Chem. Soc. Rev.* **2020**, *49*, 6816–6832.

(11) Popa, I.; Berkovich, R.; Alegre-Cebollada, J.; Badilla, C. L.; Rivas-Pardo, J. A.; Taniguchi, Y.; Kawakami, M.; Fernandez, J. M. Nanomechanics of HaloTag Tethers. *Journal of the American Chemical Society* **2013**, *135*, 12762–12771.

(12) Bustamante, C.; Alexander, L.; Maciuba, K.; Kaiser, C. M. Single-Molecule Studies of Protein Folding with Optical Tweezers. *Annual Review of Biochemistry* **2020**, *89*, 443–470, PMID: 32569525.

(13) Fernandez, J. M.; Li, H. Force-Clamp Spectroscopy Monitors the Folding Trajectory of a Single Protein. *Science* **2004**, *303*, 1674–1678.

(14) Gosse, C.; Croquette, V. Magnetic Tweezers: Micromanipulation and Force Measurement at the Molecular Level. *Biophysical Journal* **2002**, *82*, 3314 – 3329.

(15) De Vlaminck, I.; Dekker, C. Recent Advances in Magnetic Tweezers. *Annual Review of Biophysics* **2012**, *41*, 453–472, PMID: 22443989.

(16) Liu, R.; Garcia-Manyes, S.; Sarkar, A.; Badilla, C. L.; Fernández, J. M. Mechanical Characterization of Protein L in the Low-Force Regime by Electromagnetic Tweezers/Evanescent Nanometry. *Biophysical Journal* **2009**, *96*, 3810 – 3821.

(17) Popa, I.; Rivas-Pardo, J. A.; Eckels, E. C.; Echelman, D. J.; Badilla, C. L.; Valle-Orero, J.; Fernández, J. M. A HaloTag Anchored Ruler for Week-Long Studies of Protein Dynamics. *Journal of the American Chemical Society* **2016**, *138*, 10546–10553.

(18) Löf, A.; Walker, P. U.; Sedlak, S. M.; Gruber, S.; Obser, T.; Brehm, M. A.; Benoit, M.; Lipfert, J. Multiplexed protein force spectroscopy reveals equilibrium protein folding dynamics and the low-force response of von Willebrand factor. *Proceedings of the National Academy of Sciences* **2019**, *116*, 18798–18807.

(19) Tapia-Rojo, R.; Eckels, E. C.; Fernández, J. M. Ephemeral states in protein folding

under force captured with a magnetic tweezers design. *Proceedings of the National Academy of Sciences* **2019**, *116*, 7873–7878.

(20) Choi, H.-K.; Min, D.; Kang, H.; Shon, M. J.; Rah, S.-H.; Kim, H. C.; Jeong, H.; Choi, H.-J.; Bowie, J. U.; Yoon, T.-Y. Watching helical membrane proteins fold reveals a common N-to-C-terminal folding pathway. *Science* **2019**, *366*, 1150–1156.

(21) Lipfert, J.; Hao, X.; Dekker, N. H. Quantitative Modeling and Optimization of Magnetic Tweezers. *Biophysical Journal* **2009**, *96*, 5040 – 5049.

(22) Tapia-Rojo, R.; Alonso-Caballero, Á.; Fernández, J. M. Talin folding as the tuning fork of cellular mechanotransduction. *Proceedings of the National Academy of Sciences* **2020**, *117*, 21346–21353.

(23) Alonso-Caballero, A.; Tapia-Rojo, R.; Badilla, C. L.; Fernandez, J. M. Magnetic tweezers meets AFM: ultra-stable protein dynamics across the force spectrum. *bioRxiv* **2021**,

(24) Eckels, E. C.; Haldar, S.; Tapia-Rojo, R.; Rivas-Pardo, J. A.; Fernández, J. M. The Mechanical Power of Titin Folding. *Cell Reports* **2019**, *27*, 1836 – 1847.e4.

(25) Alonso-Caballero, A.; Echelman, D. J.; Tapia-Rojo, R.; Haldar, S.; Eckels, E. C.; Fernandez, J. M. Protein folding modulates the chemical reactivity of a Gram-positive adhesin. *Nature Chemistry* **2020**,

(26) Alegre-Cebollada, J.; Badilla, C. L.; Fernández, J. M. Isopeptide bonds block the mechanical extension of pili in pathogenic Streptococcus pyogenes. *The Journal of biological chemistry* **2010**, *285*, 11235–11242.

(27) Rivas-Pardo, J. A.; Badilla, C. L.; Tapia-Rojo, R.; Alonso-Caballero, Á.; Fernández, J. M. Molecular strategy for blocking isopeptide bond formation in nascent pilin proteins. *Proceedings of the National Academy of Sciences* **2018**, *115*, 9222–9227.

(28) Guo, H. A Simple Algorithm for Fitting a Gaussian Function. *IEEE SIGNAL PRO-CESSING MAGAZINE* **2011**, *10.1109/MSP.2011.941846*, 134–137.

(29) Bell, G. Models for the specific adhesion of cells to cells. *Science* **1978**, *200*, 618–627.

(30) Berkovich, R.; Garcia-Manyes, S.; Urbakh, M.; Klafter, J.; Fernandez, J. M. Collapse Dynamics of Single Proteins Extended by Force. *Biophysical Journal* **2010**, *98*, 2692–2701.

(31) Berkovich, R.; Garcia-Manyes, S.; Klafter, J.; Urbakh, M.; Fernández, J. M. Hopping around an entropic barrier created by force. *Biochemical and Biophysical Research Communications* **2010**, *403*, 133 – 137.

(32) Dudko, O. K.; Hummer, G.; Szabo, A. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proceedings of the National Academy of Sciences* **2008**, *105*, 15755–15760.

(33) Sigworth, F.; Sine, S. Data transformations for improved display and fitting of single-channel dwell time histograms. *Biophysical Journal* **1987**, *52*, 1047 – 1054.

(34) Chung, H. S.; McHale, K.; Louis, J. M.; Eaton, W. A. Single-Molecule Fluorescence Experiments Determine Protein Folding Transition Path Times. *Science* **2012**, *335*, 981–984.

(35) Neupane, K.; Foster, D. A. N.; Dee, D. R.; Yu, H.; Wang, F.; Woodside, M. T. Direct observation of transition paths during the folding of proteins and nucleic acids. *Science* **2016**, *352*, 239–242.

(36) Neupane, K.; Manuel, A. P.; Woodside, M. T. Protein folding trajectories can be described quantitatively by one-dimensional diffusion over measured energy landscapes. *12*, 700–703.

(37) Zhang, Q.; Brujić, J.; Vanden-Eijnden, E. Reconstructing Free Energy Profiles from Nonequilibrium Relaxation Trajectories. *Journal of Statistical Physics* **2011**, *144*, 344–366.

(38) Berkovich, R.; Fernandez, V. I.; Stirnemann, G.; Valle-Orero, J.; Fernández, J. M. Segmentation and the Entropic Elasticity of Modular Proteins. *The Journal of Physical Chemistry Letters* **2018**, *9*, 4707–4713.

(39) Woodside, M. T.; Anthony, P. C.; Behnke-Parks, W. M.; Larizadeh, K.; Herschlag, D.; Block, S. M. Direct Measurement of the Full, Sequence-Dependent Folding Landscape of a Nucleic Acid. *Science* **2006**, *314*, 1001–1004.

(40) Cossio, P.; Hummer, G.; Szabo, A. On artifacts in single-molecule force spectroscopy. *Proceedings of the National Academy of Sciences* **2015**, *112*, 14248–14253.

(41) Kramers, H. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **1940**, *7*, 284 – 304.

(42) Valle-Orero, J.; Tapia-Rojo, R.; Eckels, E. C.; Rivas-Pardo, J. A.; Popa, I.; Fernández, J. M. Proteins Breaking Bad: A Free Energy Perspective. *The Journal of Physical Chemistry Letters* **2017**, *8*, 3642–3647.

(43) Moore, C. C. Ergodic theorem, ergodic theory, and statistical mechanics. *Proceedings of the National Academy of Sciences* **2015**, *112*, 1907–1911.

(44) Tapia-Rojo, R.; Alonso-Caballero, A.; Fernandez, J. M. Direct Observation of a Coil-to-Helix Contraction Triggered by Vinculin Binding to Talin. *bioRxiv* **2019**,

(45) Lu, H. P.; Xun, L.; Xie, X. S. Single-Molecule Enzymatic Dynamics. *Science* **1998**, *282*, 1877–1882.

(46) Kim, J.; Zhang, C.-Z.; Zhang, X.; Springer, T. A. A mechanically stabilized receptor–ligand flex-bond important in the vasculature. *Nature* **2010**, *466*, 992–995.

(47) Liu, B.; Baskin, R. J.; Kowalczykowski, S. C. DNA unwinding heterogeneity by RecBCD results from static molecules able to equilibrate. *Nature* **2013**, *500*, 482–485.

(48) Hinczewski, M.; Hyeon, C.; Thirumalai, D. Directly measuring single-molecule heterogeneity using force spectroscopy. *Proceedings of the National Academy of Sciences* **2016**, *113*, E3852–E3861.

(49) Herschlag, D. RNA Chaperones and the RNA Folding Problem. *Journal of Biological Chemistry* **1995**, *270*, 20871–20874.

(50) Treiber, D. K.; Williamson, J. R. Exposing the kinetic traps in RNA folding. *Current Opinion in Structural Biology* **1999**, *9*, 339 – 345.

(51) Solomatin, S. V.; Greenfeld, M.; Chu, S.; Herschlag, D. Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature* **2010**, *463*, 681–684.

(52) del Rio, A.; Perez-Jimenez, R.; Liu, R.; Roca-Cusachs, P.; Fernandez, J. M.; Sheetz, M. P. Stretching Single Talin Rod Molecules Activates Vinculin Binding. *Science* **2009**, *323*, 638–641.

(53) Infante, E.; Stannard, A.; Board, S. J.; Rico-Lastres, P.; Rostkova, E.; Beedle, A. E. M.; Lezamiz, A.; Wang, Y. J.; Gulaidi Breen, S.; Panagaki, F.; Sundar Rajan, V.; Shanahan, C.; Roca-Cusachs, P.; Garcia-Manyes, S. The mechanical stability of proteins regulates their translocation rate into the cell nucleus. *15*, 973–981.

(54) Tapia-Rojo, R.; Arregui, S.; Mazo, J. J.; Falo, F. Mechanical unfolding of a simple model protein goes beyond the reach of one-dimensional descriptions. *The Journal of Chemical Physics* **2014**, *141*, 135102.

(55) Valle-Orero, J.; Rivas-Pardo, J. A.; Tapia-Rojo, R.; Popa, I.; Echelman, D. J.; Haldar, S.; Fernández, J. M. Mechanical Deformation Accelerates Protein Ageing. *Angewandte Chemie International Edition* **2017**, *56*, 9741–9746.

32

(56) de Graff, A. M.; Hazoglou, M. J.; Dill, K. A. Highly Charged Proteins: The Achilles' Heel of Aging Proteomes. *Structure* **2016**, *24*, 329 – 336.

(57) Eckels, E. C.; Tapia-Rojo, R.; Rivas-Pardo, J. A.; Fernández, J. M. The Work of Titin Protein Folding as a Major Driver in Muscle Contraction. *Annual Review of Physiology* **2018**, *80*, 327–351, PMID: 29433413.

(58) Doi, M.; Edwards, S. *The Theory of Polymer Dynamics*; Oxford Science Publications, 1986.

SI Appendix: Identical Sequences, Different Behaviors: Protein Diversity Captured at the Single-Molecule Level

Rafael Tapia-Rojo,[1, *] Alvaro Alonso-Caballero,[1] Carmen L. Badilla,[1] and Julio M. Fernandez[1]

[1]*Department of Biological Sciences, Columbia University, New York, NY 10027, USA*

MAGNETIC TWEEZERS; SETUP DETAILS

Our magnetic tweezers design (Fig. 1 main text) is assembled from off-the-shelf and custom-made components for an affordable cost. This instrument is based on our previous designs that included a regular inverted microscope. Here, we design a simple dedicated microscope that greatly simplifies the setup and reduces its cost. We include here a guide of assembly, including details about the microscope design, the magnetic tape head implementation, and the electronics.

Microscope design

For the optics of our setup, we use a 160 mm objective mounted on a piezo electric actuator (element #4, P-725, PI), which is on top of the Z-stage (element #5, Thorlabs). While the Z-stage can be used to do coarse movements of the objective (for example, after mounting the fluid chamber to find the focal plane), the piezo electric actuator is used to focus at high-precision, allowing to track the position of the magnetic bead with nm resolution. Figure 1 shows a diagram for the optical pathway of our microscope design, separating the illumination pathway (yellow) and the image (orange). The colors of the beams are arbitrary. The sample is illuminated from below using a white LED that provides a point light source (element #12, Thorlabs). The light beam is then collimated using a plano-convex lens (element # 10, Thorlabs) and then focused on the 160 mm objective (element #3, Zeiss) using a tube lens (element #7, Thorlabs), after passing through a green filter (element #9, Thorlabs) and an iris diaphragm (element #8, Thorlabs). The filter helps in producing sharper interference rings, while removes high frequencies that increase the production of reactive oxygen species in the protein solution, while the iris modulates the aperture of the microscope. The virtual image created by the objective is directed to the CMOS camera using a 50:50 beam splitter (element #6, Thorlabs), and converted to a parallel image using a plano-convex lens (element #15).

The extension of the protein is tracked in real time using the image analysis algorithm previously described [1]. Briefly, at the start of the experiment, we build a stack library of the radial vectors of the chosen magnetic and reference beads with a 2,560 μm travel with 20 nm steps (128 radial vectors) using the piezo actuator. The radial vector is calculated from the Fourier transform of images of beads, by summing over the intensity of the pixels at equal radial distance (removing low frequencies to account only for the beads' rings). During the course of an experiment, the real-time radial vector of the magnetic and reference beads is compared to the stack library by calculating the Pearson cross-correlation. To obtain the change in extension of the protein (height of the magnetic bead compared to the reference one), we fit a Gaussian to the correlation vectors using the analytical Caruana's algorithm [2], and infer the extension from the difference in position of the peaks as obtained from the fit. Every 1,000 frames the drift is corrected by refocusing the objective to keep the reference bead on focus following its radial vector at the start of the experiment.

Implementation of the magnetic tape head

Our magnetic tweezers design implements a magnetic tape head (element #17), which allows generating pN-level forces with a resolution and control superior to other implementations of electromagnets or permanent magnets. As previously described, the magnetic field generated by the tape head can be described analytically, which allows to develop a theory for the pulling force as a function of the distance from the head gap to the magnetic probe and the electric current supplied to the head [3]. In our novel magnetic tweezers setup, we maintain the tape head at a fixed distance, and control the force only through the electric current. The pulling force changes over a length scale of the gap size, 25 μm, which requires an accurate positioning in order to generate calibrated and reproducible forces. To this aim, we designed a mount piece that is fabricated with high-precision CNC (element #16), which maintains the tape head fixed 300 μm away from the magnetic probe, greatly increasing the reliability and stability of our setup, since no moving parts are required. Figure 2 shows a technical drawing indicating the dimensions of the piece (in mm), designed for the rail-carriages and tape head we employ (see Table I). The design includes the holes for mounting it on the optical rail carriages (element #2) and maintaining the tape head gap aligned with the optical axis of our microscope. The tape head is mounted with a dowel-pin system that fixes it position. In this assembly, the gap is exactly 450 μm away over the surface so that, when using 150 μm-thick bottom glasses (Ted Pella), the gap is 300

μm from the magnetic beads. When using the M-270 beads, the force law is [3]:

$$F(I) = (2.79 \times 10^{-5})I^2 + (1.64 \times 10^{-2})I, \tag{1}$$

where I is the electric current in mA. The tape head saturates when using electric currents over 1 A. Hence, we can generate mechanical forces between 0 pN, and 44 pN. We have recently developed an implementation of the larger M-450 magnetic probes which pushes the upper force limit to $\sim$220 pN [4].

Electronic circuit

The tape head is maintained under electronic feedback, which enables to do swift force changes and apply complex force signals [3, 5]. We designed a custom PI circuit (proportional+integral) to maintain the electric current (force) under feedback (see Fig. 3 for its scheme). Since the tape head is an inductance impedance, we use a high-precision 2 Ω resistor (R17, Fig. 3) in series with the head and maintain the voltage drop across it under feedback, and, hence, the electric current through the head.

The circuit has two independent inputs (commands) from two separate DAQS, one accounting for the DC force (signals-3-4), and one for the AC component (signals-5-6). This implementation allows applying complex force signals such as oscillations or mechanical noise, as in [5]. The use of this second DAQ is optional, and the system can work only with a single command if no time-dependent force signals are to be applied. Both command signals are summed by a summing amplifier (A0) that feeds the error amplifier (A2). The other input is the voltage drop across the high-precision resistor (R17), which is also read as an output with a follower amplifier (A1). The error signal (difference between the command and the actual signal) goes to the PI circuit (proportional A3 and integral A4), calibrated for the used tape head. The signal has an output gain (A5) and is finally amplified by a power amplifier (LM3886), which supplies the desired current to the head. We use a pair of 12 V batteries as power supply, which provide a clean and fast source electric current, crucial for doing fast force changes. This implementation allows changing the force over $\sim$40 μs, only limited by the slew-rate of the power amplifier.

We include in Fig. 3 the board layout for surface mounting and a picture of the circuit once assembled.

Fluid chamber assembly and manipulation

Our fluid chambers (element #19) are custom fabricated using two glass cover slides (Ted Pella; 40x24 mm bottom, 22x22 top), which are functionalized as described below. The cover slides are sandwitched with a laser-cut parafilm pattern, and the assembly is melted with a hot plate with weight on top. The fluid chamber is clipped to a custom-design manipulation fork (element #18) fabricated by metal-stamp using two stripes of double-sticky tape. The fork is mounted on a XY stage (element #20) using two high-precision shoulder screws that allow horizontal and vertical displacement over a $\sim$3x3 mm region. The explored region can be changed by displacing horizontally the fork using its slit.

List-of-parts

We include in Table I a complete list of parts, following the numbering of Figure 1 (main text). All components can be either purchased from standard scientific retailers (Thorlabs, Newport), or are custom-designed, in which case we include the necessary files for producing them.

DESIGN OF MOLECULAR CONSTRUCTS AND FUNCTIONALIZATION PROTOCOLS

Fluid chamber functionalization

Our fluid chamber design comprises three components: 1) a 22x22 mm top glass; 2) a laser cut parafilm spacer; 3) a 40x24 mm bottom glass (Fig. 4). Similar to the protocol described previously [1], the bottom glass is functionalized to harbor the chloro-alkane O4 HaloTag ligand (element iii, Fig. 4). This is done through a three-step protocol, that

Number in Fig. 1	Element
1	95 mm Construction Rail, 500 mm height. XT95-500, (Thorlabs).
2	Optical Rail Carriage, M-CXL95-80, (Newport).
3	Objective, 100 X Plan Apo f=160 mm (Zeiss).
4	P-725.CDD PIFOC High Dynamics Piezo Scanner, (PI).
5	Z-flexor stage. SM1Z - Z-Axis Translation Mount (Thorlabs).
6	50:50 Beam Splitter, BSM10R-25x36 (Thorlabs).
7	Filter Green Filter, $\lambda \sim$550 nm, (Thorlabs)
8	Cage System Iris Diaphragm, CP20S, (Thorlabs)
9	30 mm Cage Plate with 1" Double Bore, CP35-30, (Thorlabs).
10	Plano Convex lens (Thorlabs)
11	Cage System U-Bench M6 and M4 Tapped Holes, CB1/M, (Thorlabs) (x2)
12	LED source, MCWHL5-C4-6500 K, (Thorlabs).
13	Custom cage for electronics.
14	CMOS Camera, MQ013MG-ON, (Ximea)
15	Plano-convex lens, (Thorlabs)
16	Mount-piece for tape-head (CNC fabricated).
17	Magnetic Tape Head, 902836 (Brush Industries)
18	Manipulation fork (Metal stamp, custom-made).
19	Fluid Chamber (custom-made)
20	XY Linear stage, miniature, MS-125-XY (NewPort).
Not shown	DAQ control card, USB-6341, (National Instruments).
Not shown	12 Volts batteries, 18 Ah).
Not shown	Battery charger.
Not shown	Piezo controller (all PI) E-515.0X (chassis); E-505.00 (amplifier); E-509.01 (servo).

TABLE I. Parts list for the magnetic tweezers setup

after cleaning the glass surface, involves its silanization to incorporate an amino-group to the glass surface (element ii, Fig. 4). After silanization, the chamber is assembled by sandwiching the three components on a hot plate with weight on them. Prior to that, the top glasses are functionalized with repel silane to make them hydrophobic (Fig. 4, a). Once the fluid chamber is assembled, the glutaldehyde group is incorporated (ii), which plays a double role: first, to allow the anchoring and immobilization of non-magnetic amine beads, used as reference beads; second, reaction with the O4 HaloTag ligand, which allows the covalent binding of HaloTagged proteins.

Molecular Biology Protocols

All the reagents employed in this research were purchased from Sigma-Aldrich, unless otherwise specified. Both R3$^{\text{IVV}}$ and protein L monomers were cloned into and expressed from a modified pFN18a-HaloTag T7 Flexivector (Fig. 5), that contains two copies of the Spy0128 gene followed by an AviTag(pFN18a-HaloTag-TEV site-(Spy0128)2-AviTag). All protein constructs contain a 6xHisTag for purification before the AviTag. All the cloning steps were done following the strategy previously described [6]. Briefly, the genes encoding the proteins contain a 5' BamHI and a 3' BglII restriction sites. The pFN18a plasmid has a BamHI restriction site between the TEV site and the first copy of the Spy0128 gene, and the digestion with BamHI generates compatible ends with BglII. The digested gene (BamHI/BglII) encoding the protein of interest can be inserted and ligated into the dephosphorilated pFN18a (BamHI). Due to the lack of directionality, the gene can be inserted in the correct (5' to 3') or incorrect (3' to 5') orientation in the plasmid. This is checked with a subsequent digestion with BamHI and BlpI (91 bp beyond the stop codon in the 3' end). In

the correct orientation, this digestion produces a fragment of 3900 bps (pFN18a-HaloTag-TEV site-BamHI) and a fragment of 1900+protein of interest sequence bp that accounts for BamHI-(protein of interest)-(Spy0128)2-AviTag-BlpI. In the case of an antisense ligation, the digestion produces the fragments pFN18a-HaloTag-TEV site-protein of interest sequence (3900 bps + sequence of gene bp) and a fragment of 1900 bp (BamHI-(Spy0128)2-AviTag-BlpI). All the cloning and amplification steps were conducted on the Escherichia coli XL10-Gold strain (Agilent Technologies). Finally, we sequence the construct to corroborate the correct insertion orientation of the gene and the fidelity of the sequence

E. coli ERL strain cells (gift of R. T. Sauer) were transformed with the plasmids containing the chimeric proteins. Protein expression inductions and purification protocols were done as previously described [7]. Briefly, the transformed cells were grown at 37 C with 250 rpm constant shaking until reached OD600 0.6. 1 mM IPTG was used to induce expression at 25 C with 250 rpm constant shaking for 12-16 h. Protein extraction from cells was done through mechanical lysis with a French press (Sim-Aminco), and the protein isolation from the lysate was done with the His60 Ni SuperflowResin (Clontech). Further purity was achieved with a size exclusion chromatography step (Superdex 200 FPLC column, GE Healthcare). Proteins were eluted in 10 mM Hepes (pH 7.2), 150 mM NaCl, 1 mM EDTA and stored with 10% v/v of glycerol at -80°C until use.

CALCULATION OF THE FREE ENERGY LANDSCAPES

To estimate the free energy landscape, we use long time-series of the measured extension $x(t)$, and estimate a first "naive" free energy landscape as:

$$G(x) = -kT \ln p(x) - \frac{kT}{D} \int_{x_1}^{x_2} \dot{x}(x)dx, \tag{2}$$

where $p(x)$ is the probability density function of $x(t)$, kT is the thermal energy, D the diffusion coefficient, and $x_2 - x_1$ the upper and lower boundaries of $x(t)$. Equation (2) accounts for the equilibrium contribution as the simple Boltzmann inversion (first term), and a second non-equilibrium contribution that weights the directional derivatives of the trajectory (second term). In the case of long equilibrium dynamics, as it is our case, this second term has a negligible contribution to $G(x)$. In force spectroscopy experimental techniques, the instrument influences the molecular dynamics so that the measured trajectory $x(t)$ is a convolution of the actual molecular trajectory $\xi(t)$ and the instrumental effect $\Xi(t)$. The actual molecular free energy landscape $G(\xi)$ must be obtained by a deconvolution process.

In optical tweezers and AFM, the main instrumental effect arises from the stiffness of the force probe, either the optical trap or the AFM cantilever. In the case of magnetic tweezers, the magnetic trap stiffness is negligible, which gives rise to its intrinsic force-clamp conditions. However, the differential bead height is inferred using an image analysis algorithm that might introduce noise that is convoluted with the actual molecular dynamics. In this regard, we calculate the free energy landscape of R3$^{\text{IVV}}$ and protein L through a deconvolution process that requires to estimate the point-spread function of our image analysis algorithm. To this aim, we measure a long recording of two reference beads, not affected by the pulling force and firmly anchored to the substrate, so that their Brownian fluctuations are negligible, and estimate our point spread function from the extension histogram (Fig. 6A), which is a Gaussian distribution with a standard deviation of ~1.7 nm. Using this point spread function, we estimate the actual molecular free energy landscape by spectral deconvolution using the Jansson algorithm [8]. Figure 6B compares the "naive" free energy landscape (grey) and the deconvoluted one (black) for R3$^{\text{IVV}}$ at 8 pN. The deconvolution procedure is key to recover the actual characteristics of the free energy landscape, especially in the barrier region, which typically is more susceptible to be smeared by instrumental effects. With this procedure, we are able to fully characterize the free energy landscapes of protein L and R3$^{\text{IVV}}$ at different forces and describe the effect of the pulling force. Tables II and III show the properties of the free energy landscapes in terms of the free energy barrier from the folded $\Delta G_F^{\dagger}$ and unfolded $\Delta G_U^{\dagger}$ state, and the stiffness at the folded κ_F and unfolded κ_U basin, and the barrier top κ_b.

Property	Velue
$\Delta G_F^\dagger$	15.3±0.2 pNnm
$\Delta G_U^\dagger$	11.1±0.4 pNnm
κ_F	0.28±0.02 pN/nm
κ_b	0.37±0.04 pN/nm
κ_b	0.08±0.01 pN/nm

TABLE II. Properties of the free energy landscape of R3$^{\text{IVV}}$ at 8 pN.

Property	Velue
$\Delta G_F^\dagger$	32.2±0.3 pNnm
$\Delta G_U^\dagger$	29.7±0.3 pNnm
κ_F	3.40±0.04 pN/nm
κ_b	7.83±0.40 pN/nm
κ_b	1.55±0.04 pN/nm

TABLE III. Properties of the free energy landscape of protein L at 6 pN.

EFFECT OF THE MECHANICAL PERTURBATION ON THE EVOLUTION OF A PROTEIN FREE ENERGY LANDSCAPE

To a first approximation, the effect of the pulling force on a biomolecular free energy landscape is often modeled as a linear perturbation $-F \cdot x$ that tilts the landscape towards the unfolded state. However, an unfolded protein is a polypeptide chain extended by force, and its equilibrium extension is dictated by standard polymer physics models such as the FJC. In this sense, the position of the unfolded basin displaces with force in a non-linear way. However, when the biomolecule explores only a narrow range of forces, this non-linear effects can be negligible. To test this property, we subtract here the energy landscapes for R3$^{\text{IVV}}$ and protein L and recover the effect of the pulling force on the tilt of these landscapes.

Figure 7A shows the landscape at 7.7 pN (red) and 8.3 pN (black), together with the algebraic difference (green). As shown, the difference between both landscapes is approximately linear, with a slope of -0.6 pN, in agreement with the difference in force between both landscapes. In this sense, the effect of the pulling force on the explored force range for R3$^{\text{IVV}}$ is well approximated as a linear tilt, likely due to the narrow range of forces. By contrast, in the case of protein L, the explored range of forces is much broader, and the effect of the polymer extension are readily observed in the shift of the unfolded basin position, which establishes the step-size dependence with force. Here, the subtraction of the landscapes at 4.5 pN (red) and 7.5 pN (black) has a nonlinear shape, indicating a more complex effect of the pulling force.

CHARACTERIZATION OF THE DATA FOR THE PROTEIN INDIVIDUALS MEASURED IN FIGURE 5

We show here the properties of the molecules of R3$^{\text{IVV}}$ and protein L measured for Figure 5. Figures 8 and 9 show the unfolding (blue) and refolding (red) rates as function of force. From the intersection of the fits to these data, we estimate the values of $F_{1/2}$ and $r_{1/2}$. Even not all proteins explore the full range of forces, we can estimate the coexistence forces and rates from the fits to the dependence of their rates on a few forces.

Tables IV and V show the date at which each protein was measured, the total duration of the experiment, and the values for $F_{1/2}$ and $r_{1/2}$.

Date	Total time	$F_{1/2}$ (pN)	$r_{1/2}$ (1/s)
26-Feb-2020	2 h 5 min	9.1	1.0
25-Feb-2020	2 h 17 min	8.4	1.4
11-Mar-2020	21 h 24 min	7.9	0.8
25-Oct-2019	6 h 23 min	10.0	0.7
8-Nov-2019	9 h 43 min	10.2	1.7
14-Oct-2019	1 h 47 min	8.7	1.0
31-Oct-2019	5 h 21 min	9.0	1.5
21-Mar-2019	0 h 22 min	8.9	1.4
14-Ap-2019	1 h 18 min	8.4	1.4
16-Ap-2019	3 h 22 min	7.4	0.7
30-Sep-2019	0 h 24 min	8.4	0.8
30-Sep-2019	0 h 27 min	9.2	0.7
19-Ap-2019	0 h 48 min	9.1	0.5
19-Ap-2019	1 h 12 min	8.4	1.6
19-Ap-2019	2 h 03 min	9.5	2.3

TABLE IV. List of measured R3$^{\text{IVV}}$ molecules, indicating the date of measuring, total time of experiment, and their coexistence force $F_{1/2}$ and coexistence rate $r_{1/2}$.

Date	Total time	$F_{1/2}$ (pN)	$r_{1/2}$ (1/s)
14-Jan-2020	121 h 28 min	8.2	5.6×10^{-3}
14-Jan-2020	118 h 12 min	5.9	5.0×10^{-2}
21-Jan-2020	21 h 25 min	7.0	5.0×10^{-3}
24-Jan-2020	1 h 57 min	7.8	2.7×10^{-3}
31-Jan-2020	3 h 10 min	9.2	7.5×10^{-3}
5-Feb-2020	41 h 45 min	8.8	4.5×10^{-3}
5-Oct-2020	12 h 22 min	8.5	6.0×10^{-3}
6-Oct-2020	1 h 01 min	9.9	3.3×10^{-3}
6-Oct-2020	1 h 20 min	9.5	3.0×10^{-3}

TABLE V. List of measured protein L molecules, indicating the date of measuring, total time of experiment, and their coexistence force $F_{1/2}$ and coexistence rate $r_{1/2}$.

* rafa.t.rojo@gmail.com

[1] I. Popa, J. A. Rivas-Pardo, E. C. Eckels, D. J. Echelman, C. L. Badilla, J. Valle-Orero, and J. M. Fernandez, "A halotag anchored ruler for week-long studies of protein dynamics," *Journal of the American Chemical Society*, vol. 138, no. 33, pp. 10546–10553, 2016. PMID: 27409974.

[2] H. Guo, "A simple algorithm for fitting a gaussian function," *IEEE SIGNAL PROCESSING MAGAZINE*, vol. 10.1109/MSP.2011.941846, pp. 134–137, 2011.

[3] R. Tapia-Rojo, E. C. Eckels, and J. M. Fernández, "Ephemeral states in protein folding under force captured with a magnetic tweezers design," *Proceedings of the National Academy of Sciences*, vol. 116, no. 16, pp. 7873–7878, 2019.

[4] A. Alonso-Caballero, R. Tapia-Rojo, C. L. Badilla, and J. M. Fernandez, "Magnetic tweezers meets afm: ultra-stable protein dynamics across the force spectrum," *bioRxiv*, 2021.

[5] R. Tapia-Rojo, Á. Alonso-Caballero, and J. M. Fernández, "Talin folding as the tuning fork of cellular mechanotransduction," *Proceedings of the National Academy of Sciences*, vol. 117, no. 35, pp. 21346–21353, 2020.

[6] M. Carrion-Vazquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke, and J. M. Fernandez, "Mechanical and chemical unfolding of a single protein: A comparison," *Proceedings of the National Academy of Sciences*, vol. 96, no. 7, pp. 3694–3699, 1999.

[7] J. Alegre-Cebollada, C. L. Badilla, and J. M. Fernández, "Isopeptide bonds block the mechanical extension of pili in pathogenic streptococcus pyogenes," *Journal of Biological Chemistry*, vol. 285, no. 15, pp. 11235–11242, 2010.

[8] P. A. Jansson, "Deconvolution of images and spectra," 1997.
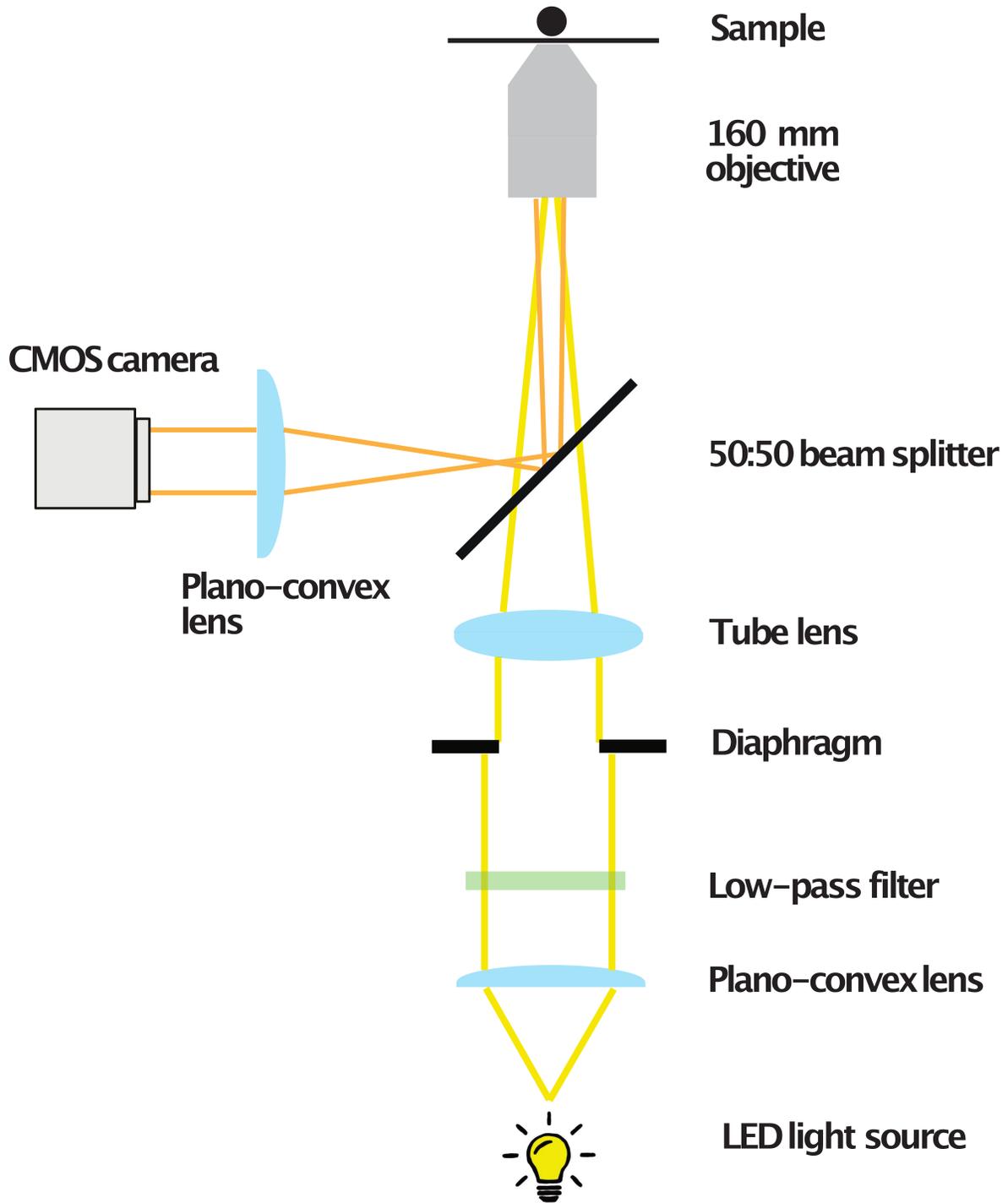
FIG. 1. **Schematics of the optical path**: Light (yellow beam) from a LED source is focused to a 160 mm objective using two plano convex lenses. The image (orange beam) is directed to a camera using a 50:50 beam splitter and plano convex lens to convert the 160 mm beam to a parallel one.

Card-reader mount piece designed for the Brush 902838 tape head and X95 Newport Carriages (metric). The tape head is mounted with a dowel-pin (lower holes) and screws (top holes), need taps. With this assembly, the tape head gap is positioned at 450 um from the surface. Milling on the objective hole might be needed.

Ø6.5
Ø3.16
Ø10.5
Ø3.8
29
12
5

Created by
Fernandez Lab 10/26/20
Approved by
Document type
Document status
Title
Card-reader v10
DWG No.
Technical reference
Dept.
Rev. Date of issue
Sheet
1/1

60
29.13
14.1
44.93
10

Ø27
Ø10
Ø1.27
65
35
7
5.08
70
26.9

29.71
52
34
25.4
70
25
10
10
25
60

FIG. 2. Technical drawing of the tape head mounting piece
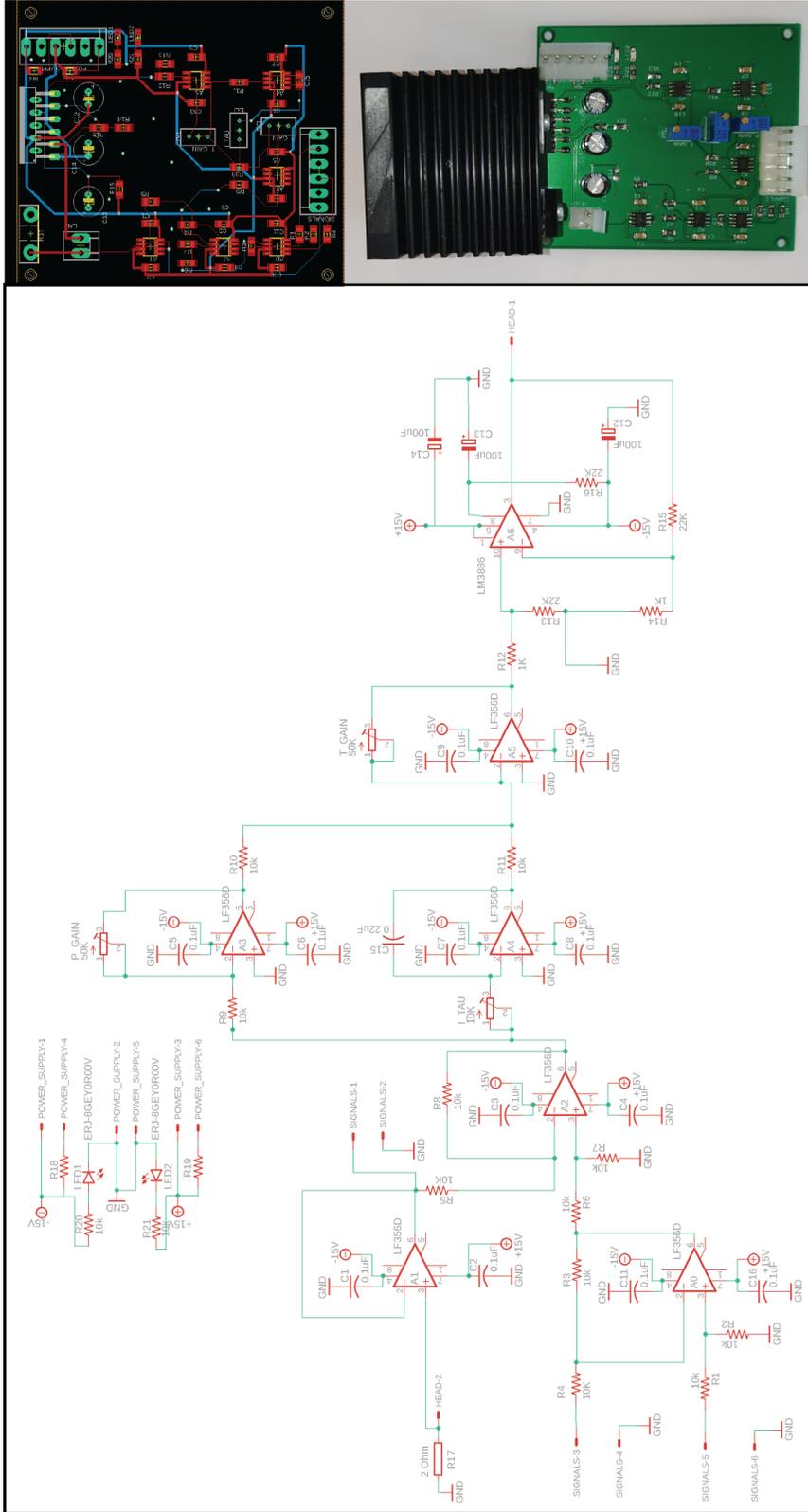
FIG. 3. **PI Control Circuit of the Magnetic Tape Head.** (Left) Schematic diagram of the PI circuit design. (right, up) Layout for the board for surface mount. (right, down) Picture of the assembled circuit. Eagle file included as SI File.
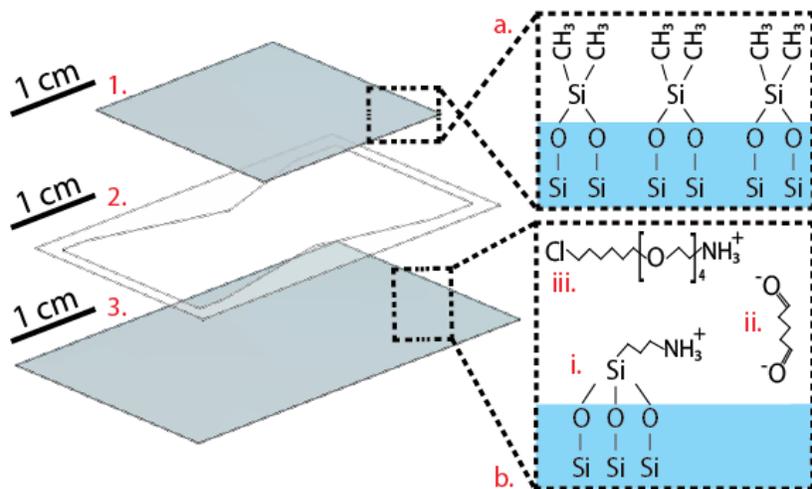
FIG. 4. **Schematics of the fluid chamber assembly and manipulation**
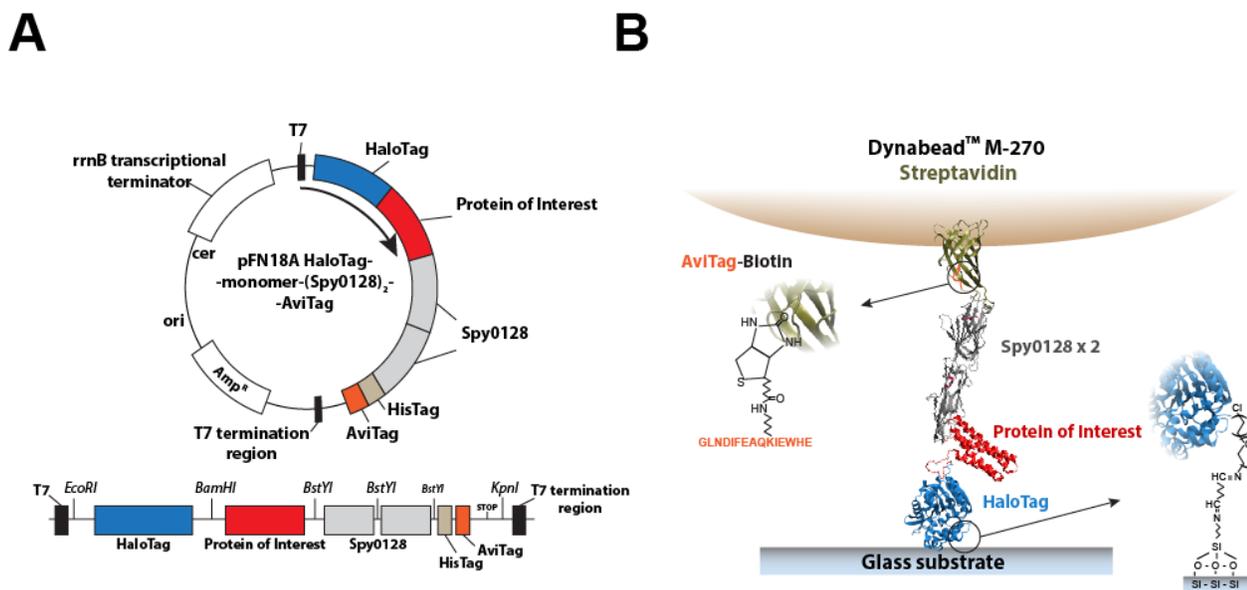
FIG. 5. **Expression vectors and single-molecule magnetic tweezers tethering.** A) Modified pFN18a HaloTag T7 Flexi vector. The T7 promoter regulates the expression of the protein HaloTag-monomer(here, talin R3 domain IVVI)-(Spy0128)2-HisTag-AviTag. Below is shown the 5'-3' scheme of the expression cassette with the restriction enzyme sites formed after subcloning. B) Magnetic tweezers single-molecule tethering. The N-terminal HaloTag binds covalently to a glass surface functionalized with the HaloTag ligand. The HaloTag ligand is covalently crosslinked to the surface through a glutaradehyde molecule that bridges the amino group of the ligand with the amino group of an APTES molecule. On the C-terminus of the protein, the biotinylated AviTag is recognized by the streptavidin molecules that coat the superparamagnetic bead (DynabeadM-270 Streptavidin). The two Spy0128 modules act as an inextensible molecular spacer that minimizes unspecific bead-surface interactions. Here the protein of interest is the R3$^{\mathrm{IVV}}$ domain from mouse talin-1, but any single domain protein or polyprotein can be inserted in this vector.
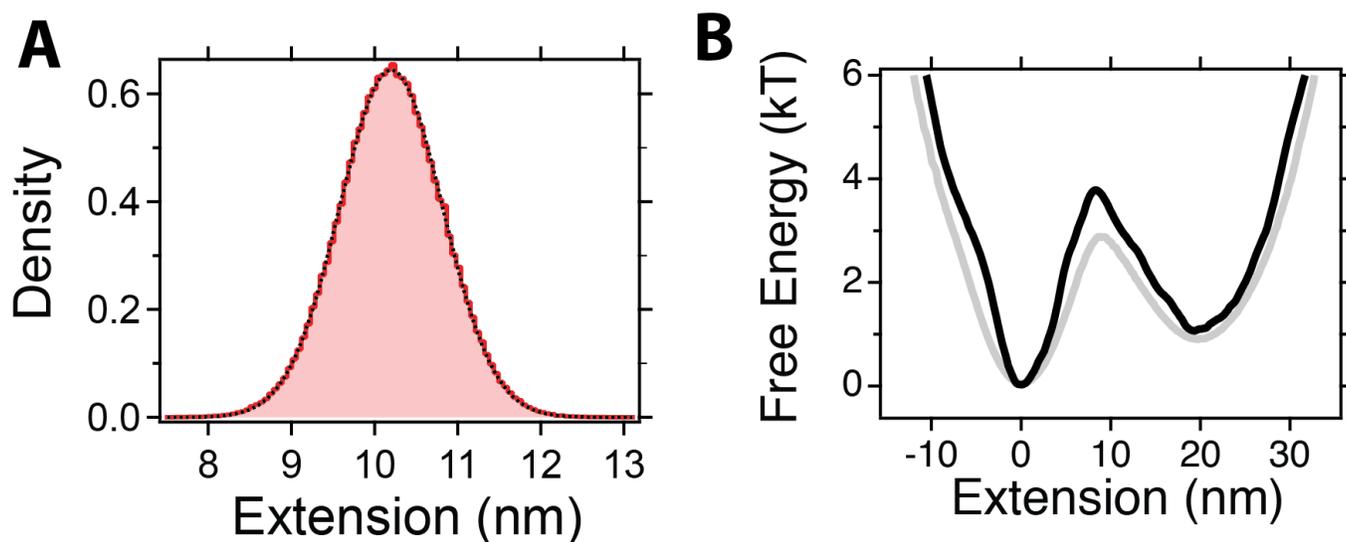
FIG. 6. **Point-spread-function:**
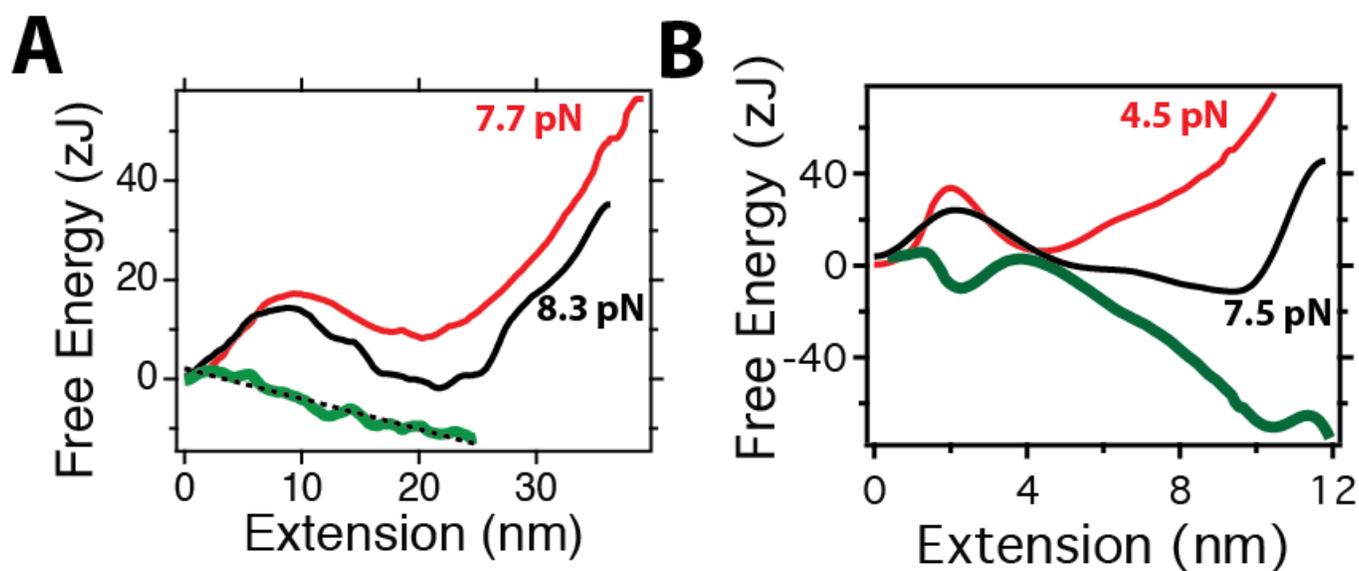
FIG. 7. **Difference between the free energy landscapes at different forces to infer the mechanical perturbation**. (A) Difference in the landscapes at 7.7 pN (red) and 8.3 pN (black) for R3$^{\mathrm{IVV}}$. The mechanical perturbation has roughly linear dependence, which a slope of $\sim$0.6 pN, which successfully recovers the perturbation between both landscapes. (B) Difference in the landscapes at 4.5 pN (red) and 7.5 pN (black) for protein L. The difference between both landscapes has a complex non-linear form, that indicates that the mechanical perturbation cannot be understood here with a simple $-Fx$ effect.
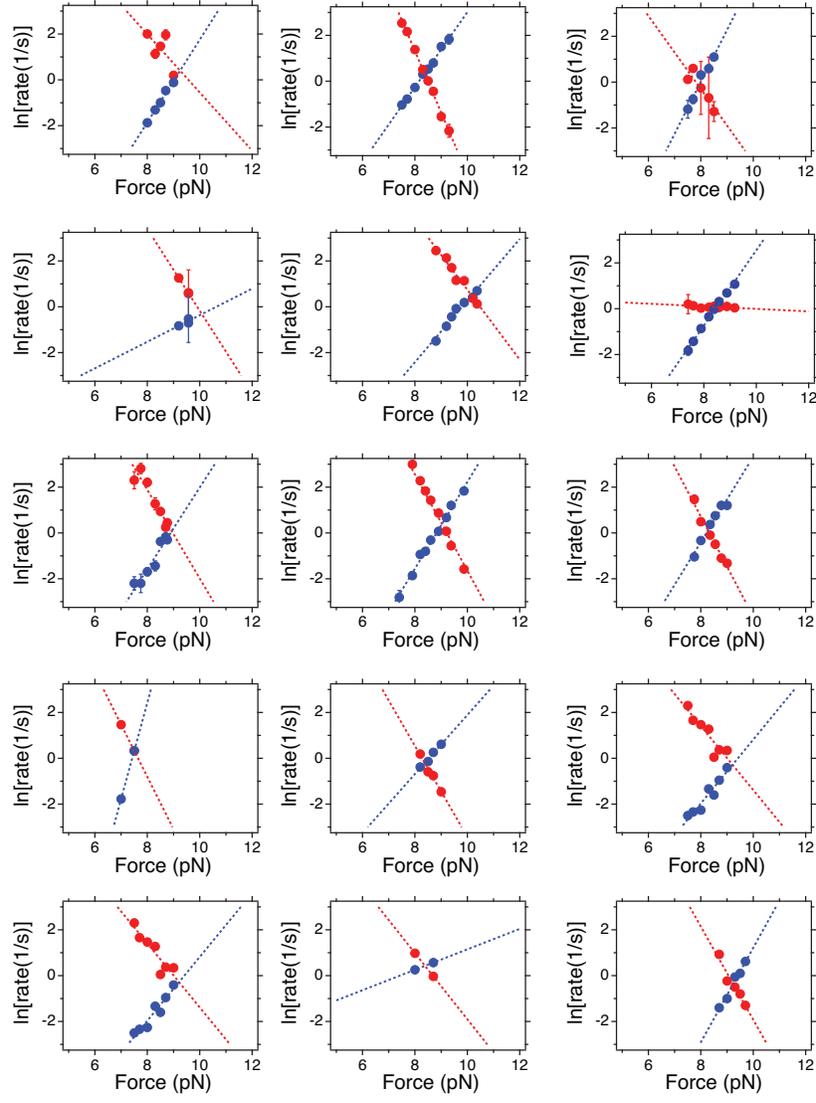
FIG. 8. **Folding (red) and unfolding rates for the measured R3$^{\text{IVV}}$ molecules**: The values of $F_{1/2}$ and $r_{1/2}$ employed for Fig. 4 (main text) are obtained from the intersection of the fits to this data (dotted lines).
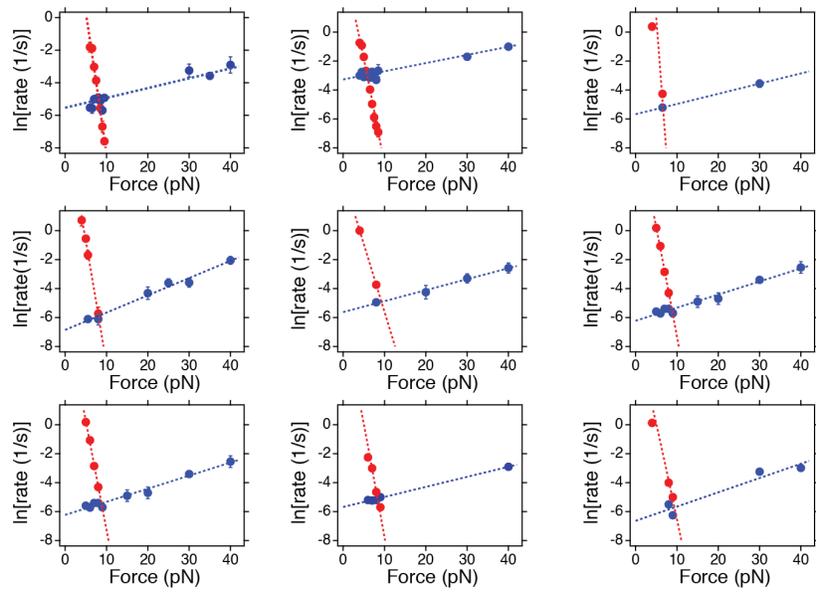
FIG. 9. **Folding (red) and unfolding rates for the measured protein L molecules**: The values of $F_{1/2}$ and $r_{1/2}$ employed for Fig. 4 (main text) are obtained from the intersection of the fits to this data (dotted lines).