SUPPLEMENTARY INFORMATION

Exploring Force-Driven Stochastic Folding Dynamics in Mechano-Responsive Proteins and Implications in Phenotypic Variation

Pritam Saha[1#], Vishavdeep Vashisht[1#], Ojas Singh[1], Gaurav Kumar Bhati[1], Surbhi Garg[1], and Sabyasachi Rakshit[1*]

[1]Department of Chemical Sciences, Indian Institute of Science Education and Research Mohali, Punjab, India

*Correspondence to Sabyasachi Rakshit

Email: srakshit@iisermohali.ac.in

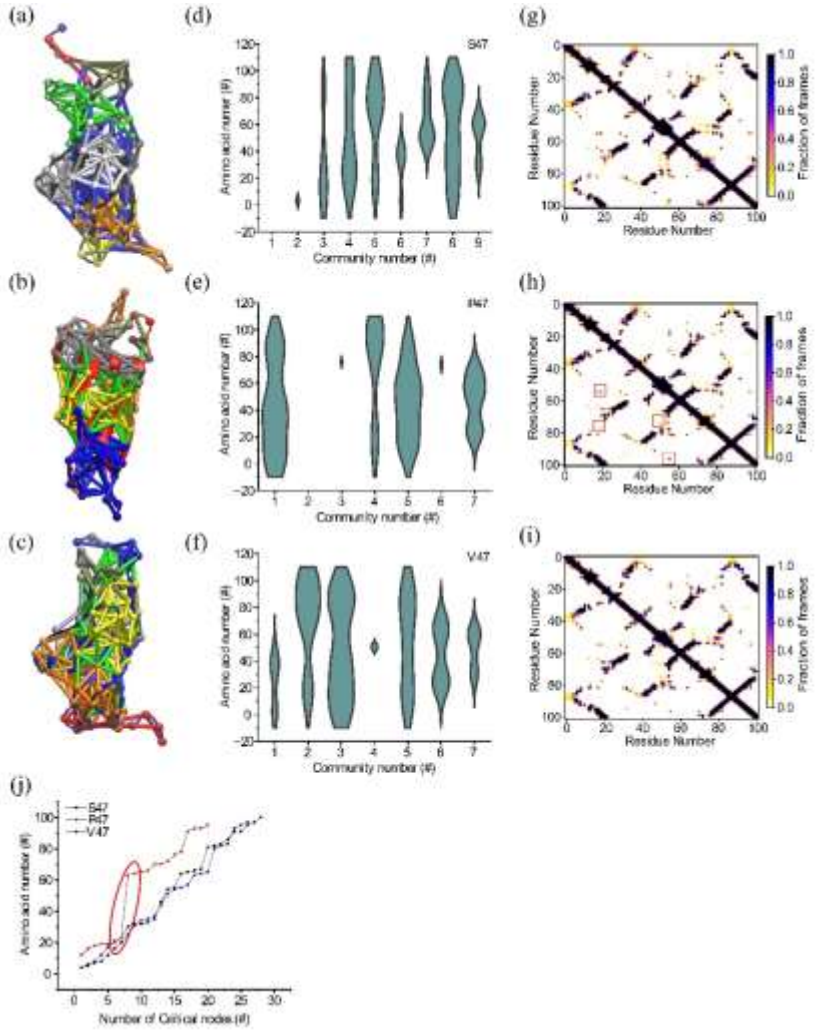
# Equal contribution

Supplementary figures

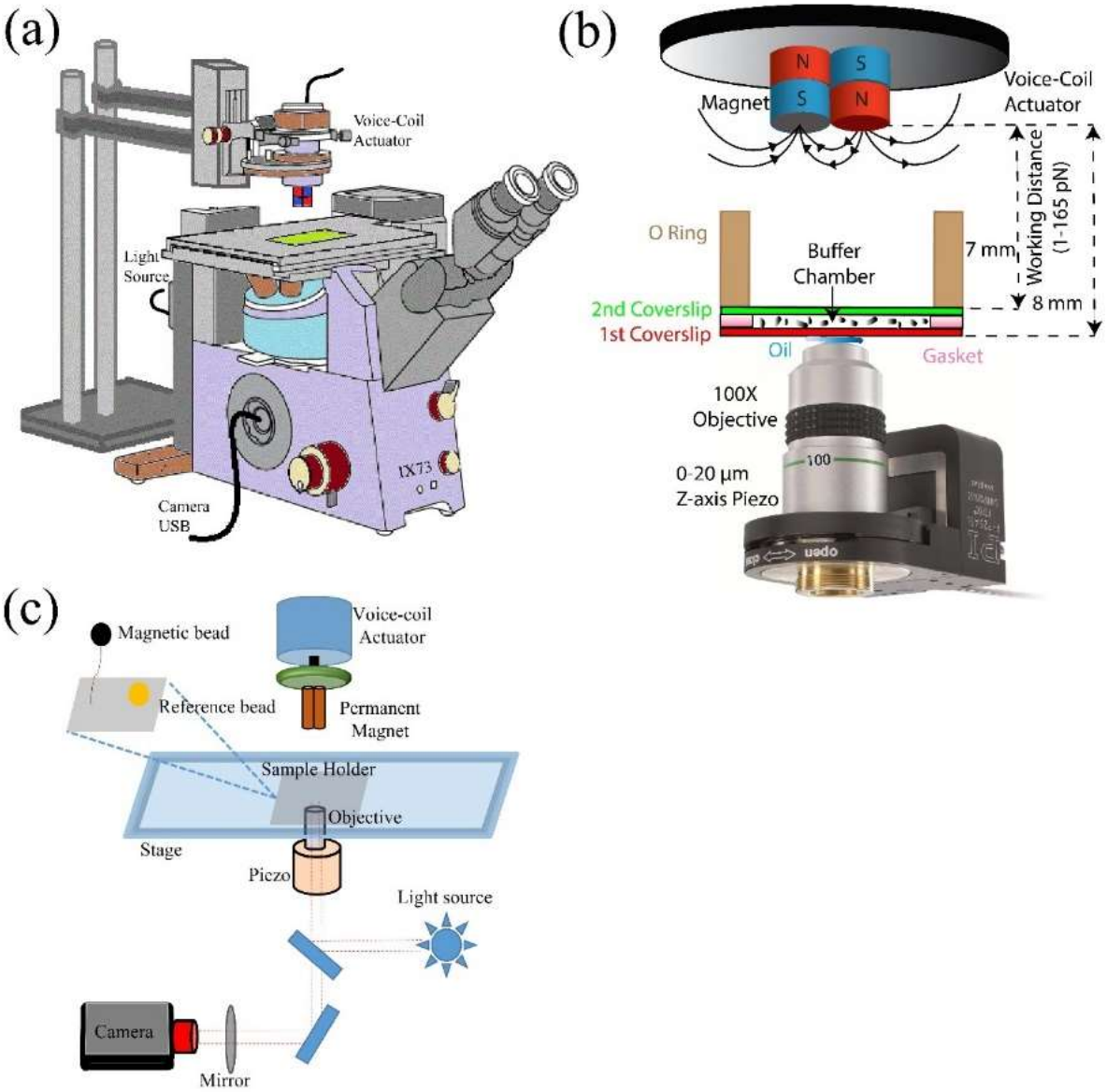
Fig 1	**Differential packing density and crankshaft motion from community analysis and anti-cross correlation of three variants**
Fig 2	**Schematics of magnetic tweezers Setup**
Fig 3	**Example of a trace of the highest acquired FPS**
Fig 4	**Flow-chart for Command Line Interface (CLI) based high fps data collection diagram**
Fig 5	**Effect of CLI in speeding up the frame-rate**
Fig 6	**100x zoomed in camera-eye view of the beads**
Fig 7	**Stack collection and image processing**
Fig 8	**Flow-diagram of noise filtering using haar-wavelet transformation method**
Fig 9	**Attachment protocol and unfolding characteristic of fingerprint molecule I_{27}**
Fig 10	**Increasing step-size in unfolding and refolding of protein-variants with increasing force**
Fig 11	**Showing force-response of three variants at different forces with three-state model representation**
Fig 12	**Experimental design of continuous force clamp for all three variants**
Fig 13	**Cluster analysis to resolve heterogeneity**
Fig 14	**Force calibration for the magnet law, using the DNA B-S overstretching**
Fig 15	**Force calibration verification using step-size measurement of MBP and I_{27}**
Fig 16	**Force range visualization with magnet position**
Fig 17	**Noise estimation from the voice-coil actuator**
Fig 18	**Example of non-specific or multi-tethered bead data**
Fig 19	**Identification of the specificity of the tether**
Fig 20	**Trace of instrumental drift in data from magnetic and non-magnetic bead**
Fig 21	**Process of cluster analysis using 10 nm bin**

Supplementary tables

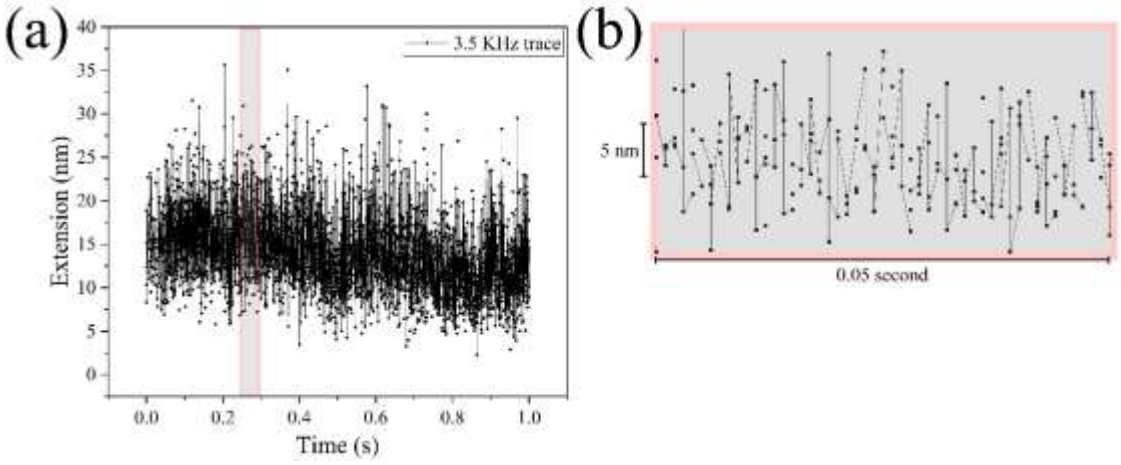
Table 1	**Force dependent refolding and unfolding lifetime data and the goodness of fitting**
Table 2	**Comparison of intrinsic refolding and unfolding rates and transition states distance**
Table 3	**Noise probability of the voice-coil actuator at different magnet position**



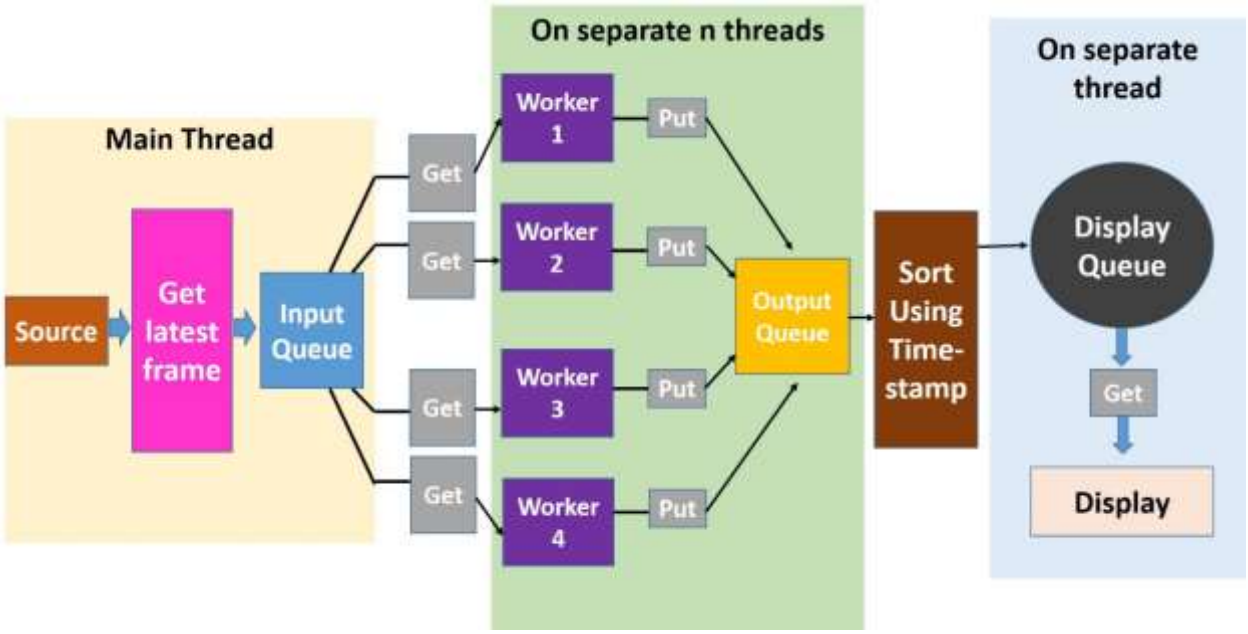
Supplementary Figure 1. Differential packing density and crankshaft motion from community analysis and anti-cross correlation of three variants. Here are the representative community maps of a) S47, b) P47 and c) V47. In amino acid residue map, higher the number of communities, higher the chance of homogeneous response while receiving perturbation. Moreover, global communities (wide-spread in 3D structure) are far more essential than local communities (restricted in 3D structure, occupying 1-3 amino acids in 1D). Amongst d) S47, e) P47 and f) V47, P47 has 4 global communities and 3 local communities, while V47 with 1 local and 6 global communities and S47 with 2 local and 7 global communities, are comparatively more interconnected. Inter-residue distance probability map of (g) S47, (h) P47 and (i) V47 Contact map provide detail view of residue-residue interaction pattern. Each dot in map represents an interaction within the pair of residues and its colour depends on frequency of occurrence of a particular interaction within the protein. In this scale, 0-1, represent the occurrence probability. The lost contacts have been highlighted with red square in P47. (j) This plot shows the number of critical nodes for all three variants. Critical nodes are the amino acid residues with has the highest impact on interaction network upon removal and gives the ability to survive against perturbation. Higher the number of critical nodes, the structure is more interconnected, thus having higher packing density. Here, both S47 (n=27) and V47 (n=26) has higher number of critical nodes than P47 (n=20). Not only that, but the absence of critical nodes near 30-60th amino acids in P47 is highlighted using red circle here.



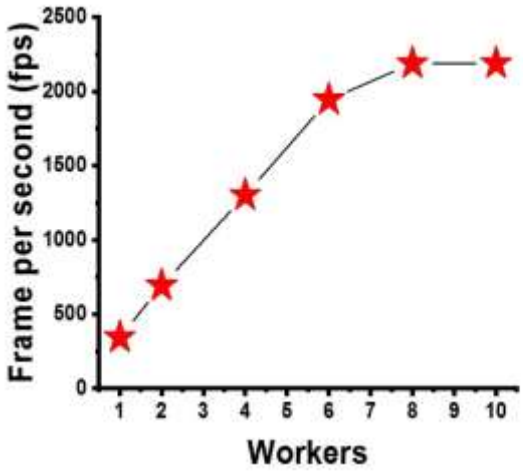
Supplementary Figure 2: Schematics of magnetic tweezer setup. a) Representative diagram of magnetic tweezer setup is shown. b) Working distance of permanent magnet is shown using zoomed in area of objective mounted piezo, oil droplet, sample chamber and magnetic field vector lines. c) Cartoon-diagram of the schematics of a magnetic tweezer components and light-path.



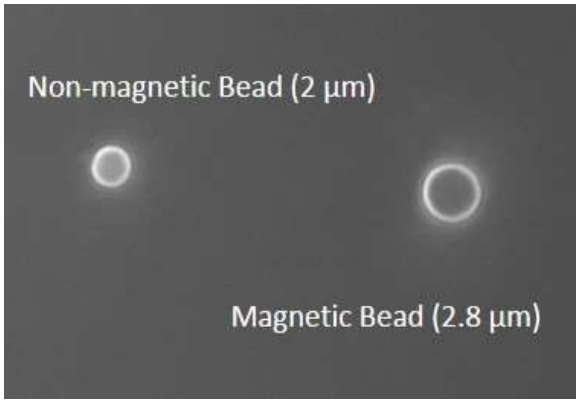
Supplementary Figure 3: Example of a trace of the highest acquired FPS. a) Plot of a 1-second trace recorded at 3.5 KHz FPS. A total of 3480 data points is here. b) Zoomed in part is from 0.25-0.3 second time of the data. Both spatial and temporal resolution are highlighted in the plot. A total of 174 data points is here with about 10-15 nm peak-to-peak distance, and 5 nm spatial resolution.



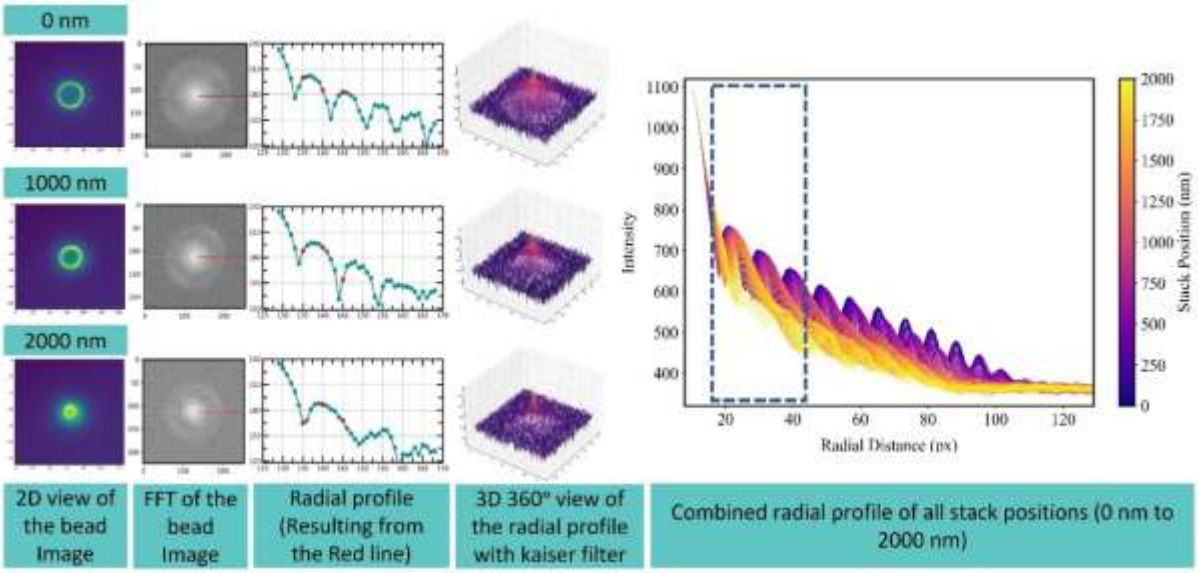
Supplementary Figure 4: Flow-chart for Command Line Interface (CLI) based high fps data collection. Concept diagram of using workers in hyper-threading to increase efficiency of computational processing power. The workers help the CPU in micro processing simultaneously to fully unlock the potential possibly of CPU cores to withstand heavy computational load.



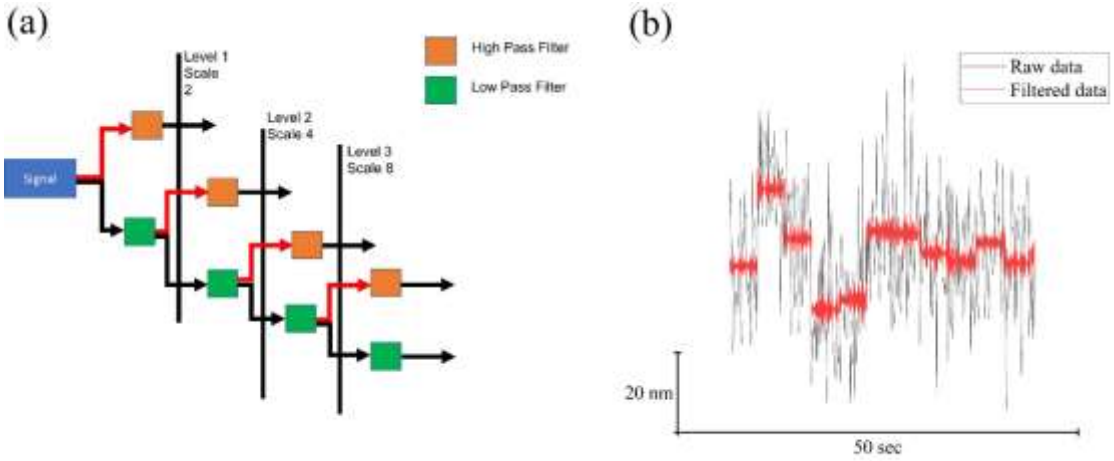
Supplementary Figure 5: Effect of CLI in speeding up the frame-rate. FPS of real time tracking of a 128x128 pixel of bead image at low exposure vs the number of workers. The graph shows that fps increases with an increasing number of workers. It also depicts that after a certain number of workers, the fps does not increase due to the threshold from other instruments, limiting the result. The highest fps of 3.5 kHz is achieved by decreasing the exposure to the minimum requirement.



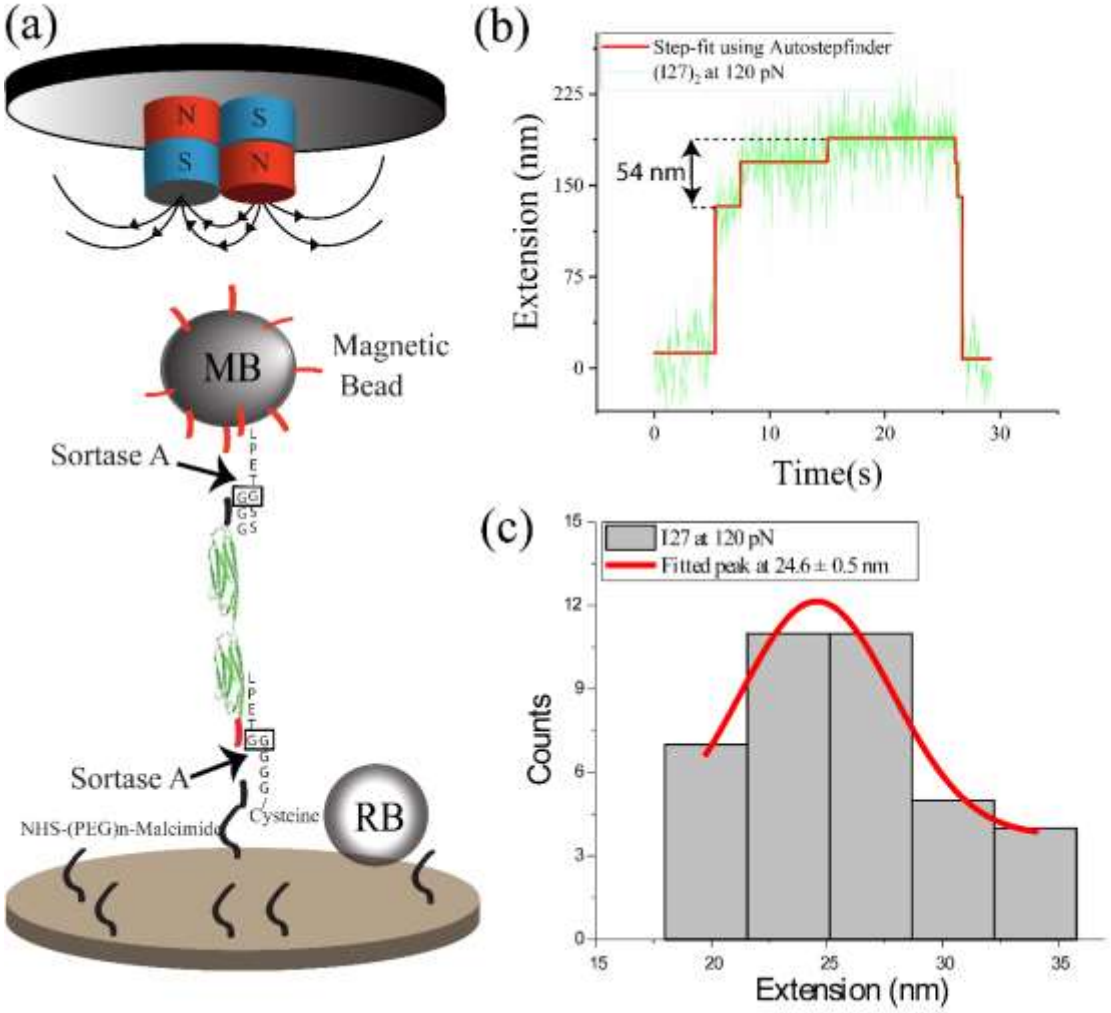
Supplementary Figure 6: 100x zoomed in camera-eye view of the beads. Snapshot from the camera with the 100x objective in IX73 microscope. Except the size, non-magnetic bead can also be identified because it has very less fluctuations as compared to magnetic bead. Also, magnetic bead can be differentiated by the subtle yellowish colour it has, compared to white coloured reference beads.



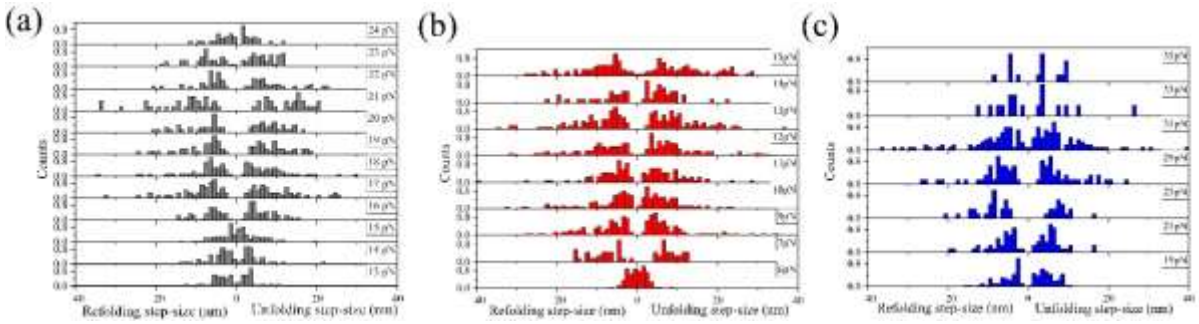
Supplementary Figure 7: Stack collection and image processing. From the left to right, 2D image of the magnetic bead, FFT of the bead images with red line for visualizing the radial profile, radial profile is drawn from the intensity matrix along the red line, 3D 360° view of intensity matrix with a kaiser filter [1], all combined radial profiles of the bead images at all z-positions spanning from 0 nm to 2000 nm. Bead images were snapped in a region of interest (ROI) and 2D FFT for each image was produced. That ROI gives a 128x128 pixel matrix, which with a kaiser filter gives this 3D radial profile of the bead image. The red line also shows the drop of the intensity as we select outer area with higher noises and X-Y fluctuations. This process is repeated for a range of positions, which combined gives the radial profile for all positions and produce stack.



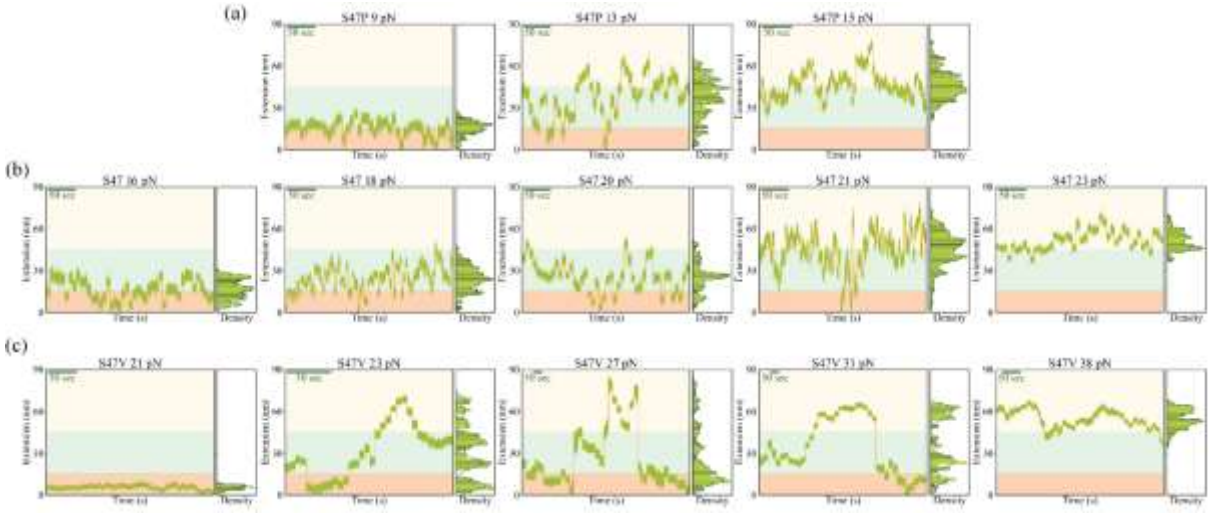
Supplementary Figure 8: Flow-diagram of noise filtering using haar-wavelet transformation method. a) Theoretical depiction of wavelet transformation technique and how it decomposes signal from noise using high and low pass frequency filter. b) P47 dimer 15 pN force-curve raw trace (in Black) and overlayed filtered data (in Red) shown together.



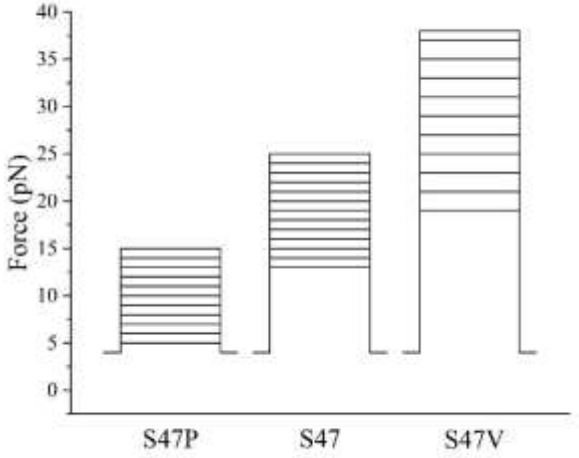
Supplementary Figure 9: Attachment protocol and unfolding characteristic of fingerprint molecule I_{27}. a) Schematic representation of the strategy for tethering the I_{27} dimer b) Extension vs. time-trace of the I_{27} dimer tether resulting in 54 nm at 120 pN. Raw data was recorded at 590 Hz. The grey force curves are the result of a 5-point window average-adjacent smoothening filter. The red line indicates step-fitting of the raw data by using step-finder algorithm from Autostepfinder c) Distributions of unfolding length at 120 pN clamping force are obtained and fitted using single peak Gaussian function, which gave the value of 24.6 ± 0.5 nm. Bin width was calculated using Freedman-Diaconis rule, $h = \frac{2\,\mathrm{IQR(x)}}{\sqrt[3]{n}}$, where h is the bin width, and n is the total number (n=34) of data in x dataset.



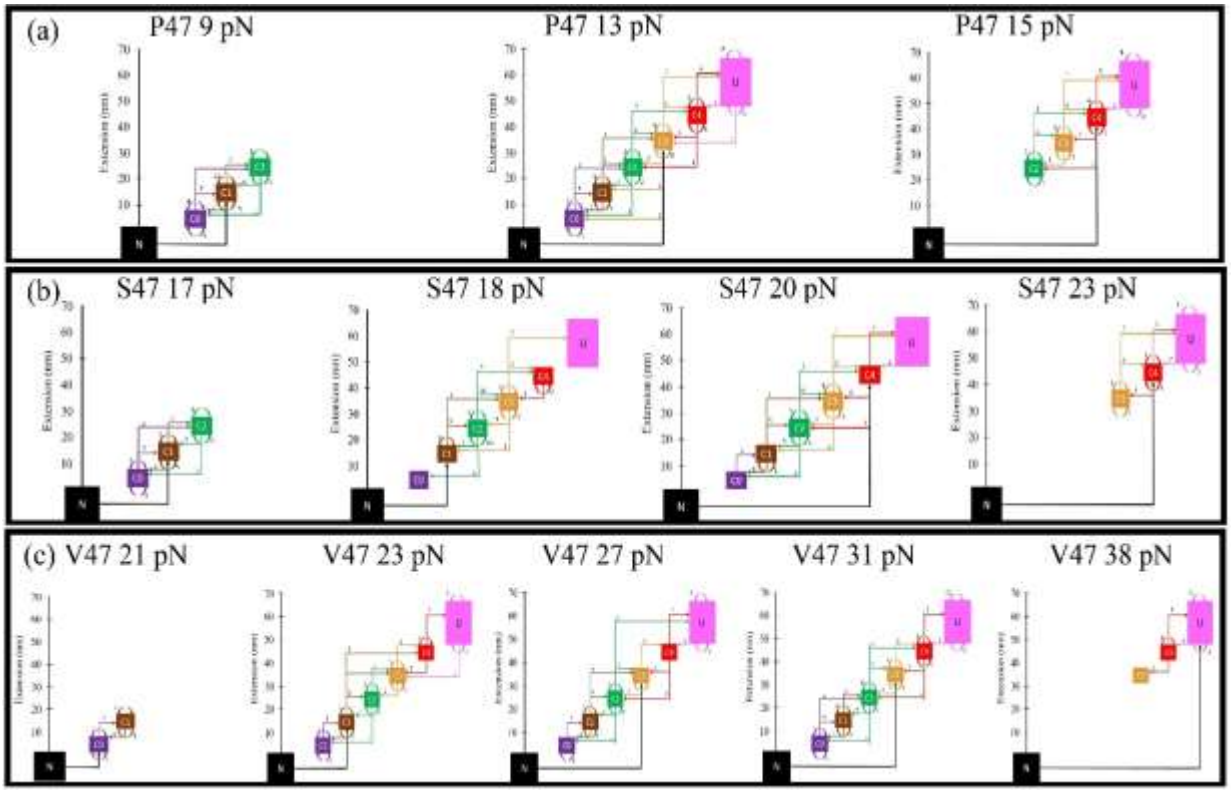
Supplementary Figure 10: Increasing step-size in unfolding and refolding of protein-variants with increasing force (in support of figure 2). a) Histogram of the step-size for S47 at 13-24 pN (bottom to top), b) for P47 variant at 6, 7, 9, 10, 11, 12, 13, 14 and 15 pN (bottom to top), and c) V47 variant at 19, 21, 25, 29, 31, 33, and 35 pN (bottom to top). We noticed a gradual increase in the width of the histograms with increasing forces indicating longer step jumps with force. However, the inter-state transition probability associated to long step-jumps is low at very high force-regime and thus, the width of histogram decreases. For each clamping force of each variants the number of steps were statistically significant ($n > 30$). In most time-traces the dataset is close of 100.



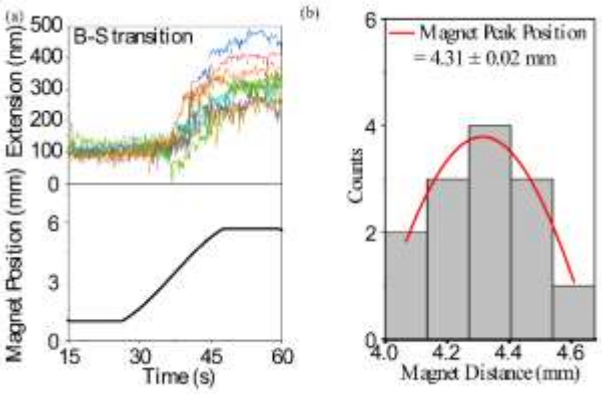
Supplementary Figure 11: **Showing force-response of three variants at different forces with three-state model representation (in support of figure 2).** Extension vs. time traces of long-term force-clamp is shown here. a) The top three are for P47 mutant at 9, 13 and 15 pN (left to right), b) in the middle there are four plots representing S47 variant at 16, 18, 20, 21 and 23pN (left to right), and c) bottom five are for V47 variant at 21, 23, 27, 31 and 38 pN (left to right). Raw data has been collected at 580 Hz. A haar wavelet filtering has been used to remove higher frequency noises and shown using dark green trace. (See methods) Step-fitting has been done using Autostepfinder of the filtered data (See methods). Filtering parameters are kept the same for all the traces (bead number, n=3, total clamping time, t = 15 minutes). However, the noise level is not the same for all. That is due to bead-to-bead variation and differential force-response of the tether. Each force-curve has been shown using these three-state models. The peach-coloured box represents closed state (15 nm from the lowest step value). It represents completely folded and almost completely folded state. The blue-coloured box represents intermediate state (15-45 nm from the lowest step value). It represents the partially folded and unfolded, differently sized conformations. Here, the intermediate state definition is not exactly similar to the popular notion of the rate-limiting step neither it can be described as off-pathway errors. The yellow-coloured box represents open state (45 nm and above from the lowest step value). It represents completely and nearly complete unfolded populations. In the respective low-force range, the protein stays folded and shows mostly closed state. As the force increase the transitions amongst the three states or variable energy conformations increases and at a certain respective high force it becomes completely unfolded showing mostly open state. The heterogeneity of these three variants differs with respect to force.



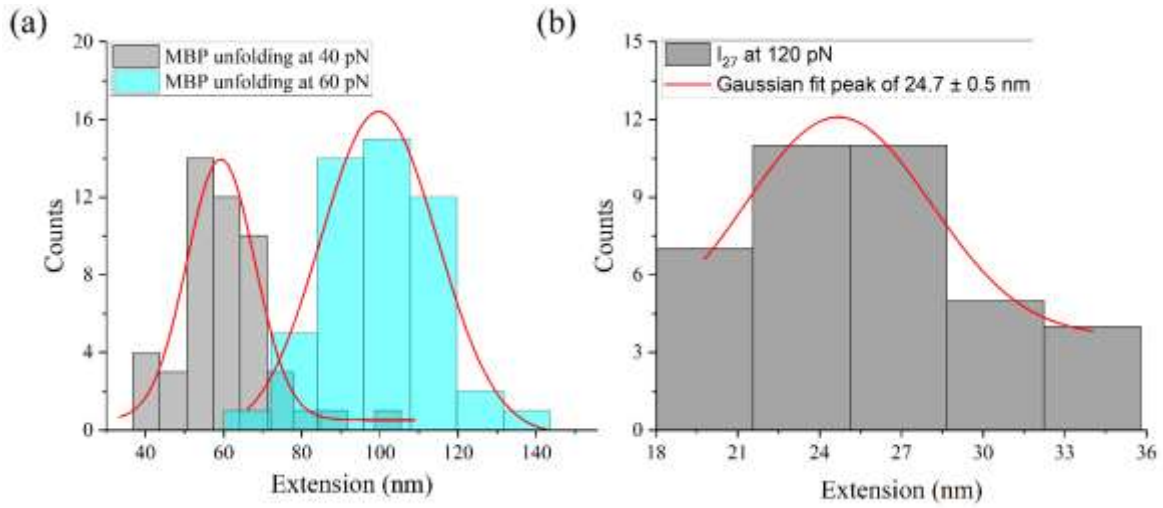
Supplementary Figure 12: Experimental design of continuous force clamp for all three variants (in support of figure 2). Constant-force experiment scheme for all three protein variants has been shown. Stair-cases represent variations in the forces: P47 (5-15 pN), S47 (13-25 pN) and V47 (19-38 pN).



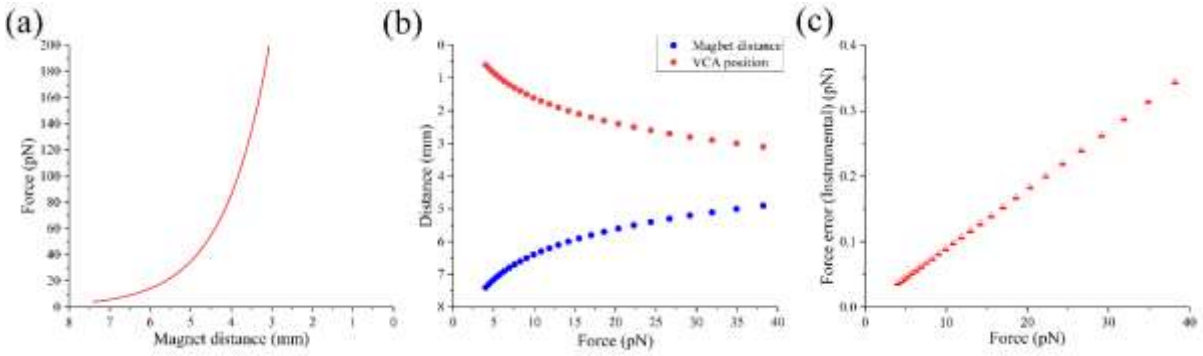
Supplementary Figure 13: **Cluster analysis to resolve heterogeneity (in support of figure 2).** Representative cluster analysis for all three variants. a) The top three are for P47 mutant at 9, 13 and 15 pN (left to right), b) in the middle there are four plots representing S47 variant at 17, 18, 20 and 23pN (left to right), and c) bottom five are for V47 variant at 21, 23, 27, 31 and 38 pN (left to right). In the respective low-force range, the protein stays folded and shows a mostly closed state. As the force increases the transitions amongst the three states or variable energy conformations increase and at a certain respective high force it becomes completely unfolded showing mostly open state. The heterogeneity of these three variants differs with respect to force, which can be quantified using this clustering method with different bin sizes. The size of the bin is kept at 10 nm because of the resolution limit, however lowering the bin size to 5 nm or increasing it to higher values than 10 nm will not change the pattern in heterogeneity.



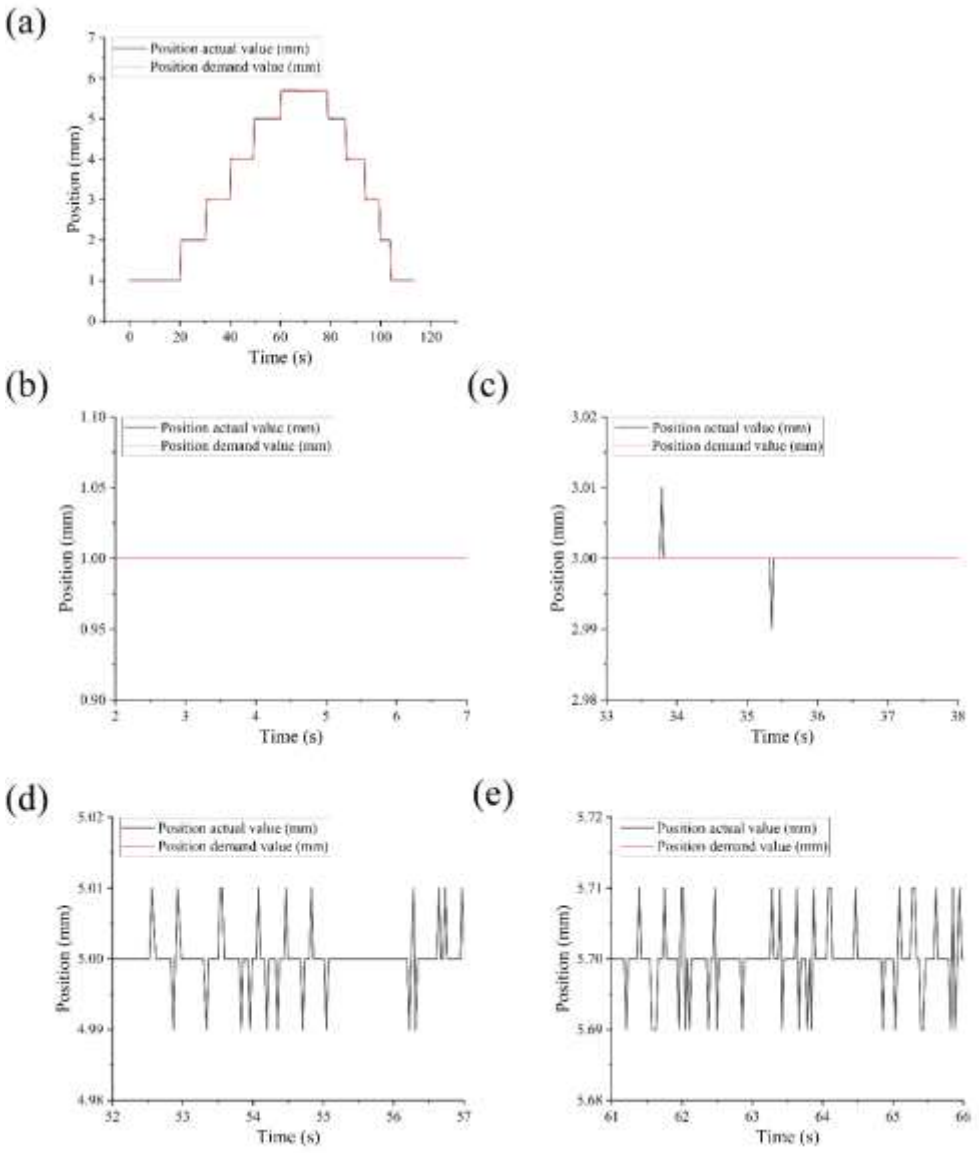
Supplementary Figure 14: Force calibration for the magnet law, using the DNA B-S overstretching a) A DNA construct composed of 605 bp DNA segment is tethered between a glass slide and a paramagnetic bead. The B−S transition is observed (n = 13) marking the 65 pN point. b) Histogram of magnet position values where the B−S transition is measured. The line is a Gaussian fit with magnet distance = 4.31 ± 0.02 mm. Bin width was calculated using Freedman-Diaconis rule, $h = \frac{2\,\mathrm{IQR(x)}}{\sqrt[3]{n}}$, where h is the bin width, and n is the total number of data in x dataset (n=13).



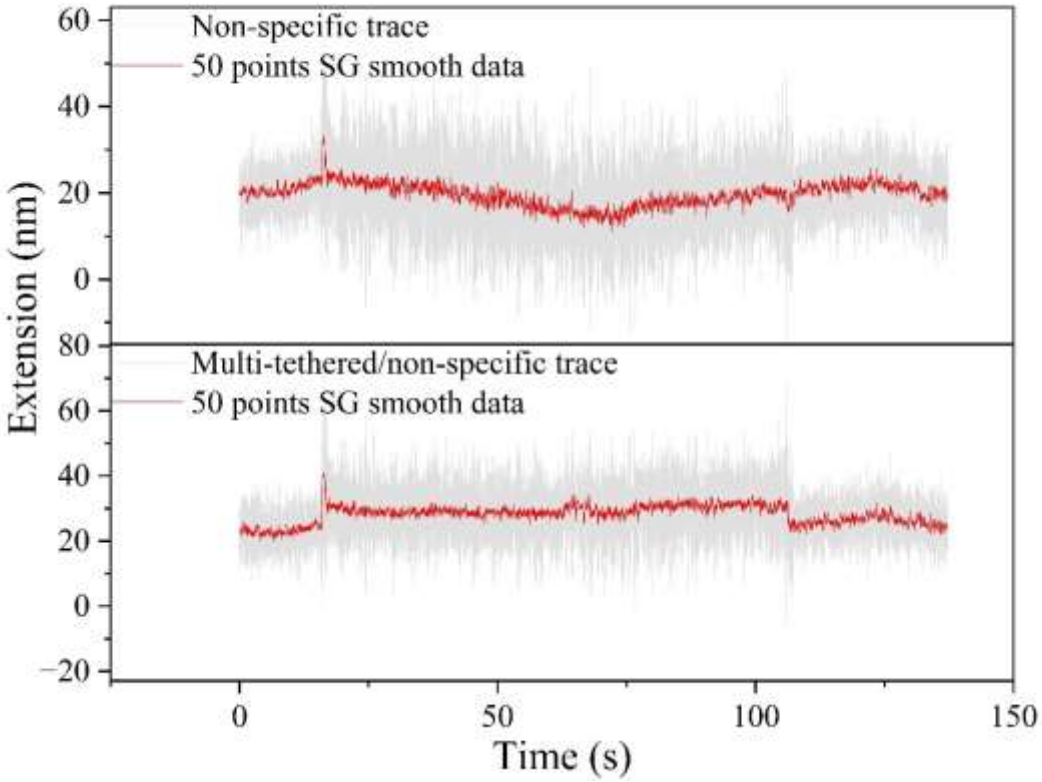
Supplementary Figure 15: Force calibration verification using step-size measurement of MBP and I_{27}. a) Distribution of unfolding length is obtained and fitted using single peak Gaussian function at 40 pN (n=49) and 60 pN (n=50). b) Distributions of unfolding length are obtained and fitted using single peak Gaussian function at 120 pN, which gave the value of 24.7 ± 0.5 nm (n=38). Bin width was calculated using Freedman-Diaconis rule, $h = \frac{2\,\mathrm{IQR(x)}}{\sqrt[3]{n}}$, where h is the bin width, and n is the total number of data in x dataset.



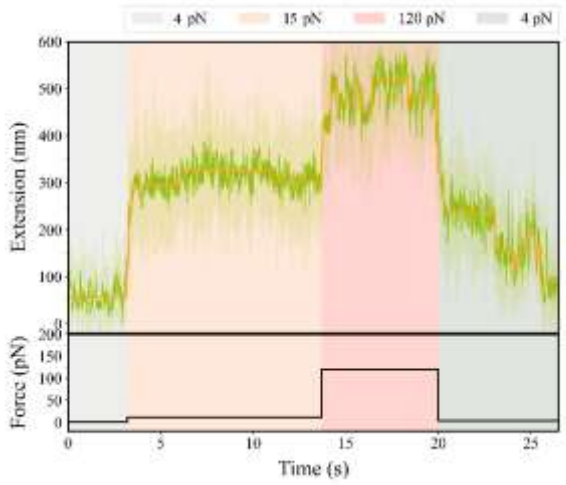
Supplementary Figure 16: Force range visualization with magnet position. a) Magnet distance (7.4 mm - 3 mm) versus force (~4 - 200 pN) depiction after finalizing the fitting parameters. b) Magnet distance and voice-coil magnet position has been shown with respect to force (4 - 40 pN). c) Force error due to instrumental positional error resulting due to voice-coil vibration versus force (4 - 40 pN) is plotted.



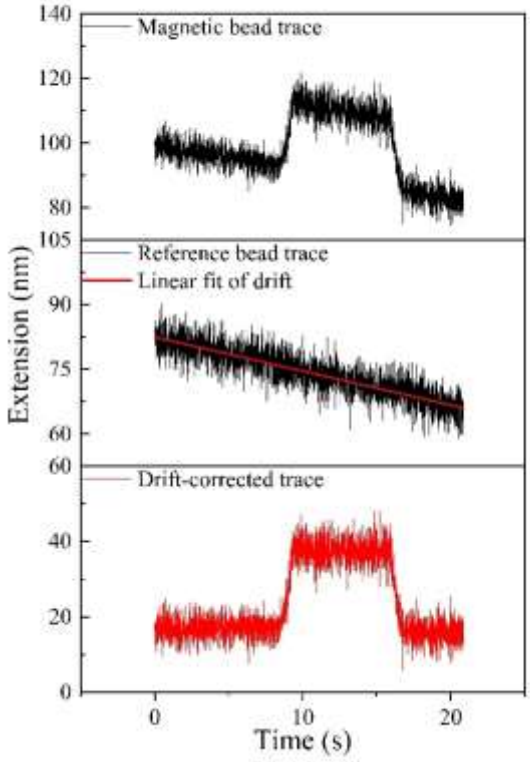
Supplementary Figure 17: Noise estimation from the voice-coil actuator. a) Representation of actual position value and target position value with time at different magnet position. b) Zoomed in region at magnet position 1 mm, the lowest tracking position of VCA, for 5 seconds. c) Zoomed in region at magnet position 3 mm for 5 seconds which shows 2 noise spikes with a spatial amplitude of 0.01 mm and temporal amplitude of 0.06 second. d) Zoomed in region at magnet position 5 mm for 5 seconds with 20 noise spikes with same spatiotemporal amplitude as in plot c, e) Zoomed in region at 5.7 mm, the highest tracking position of VCA, for 5 seconds with 32 noise spikes.



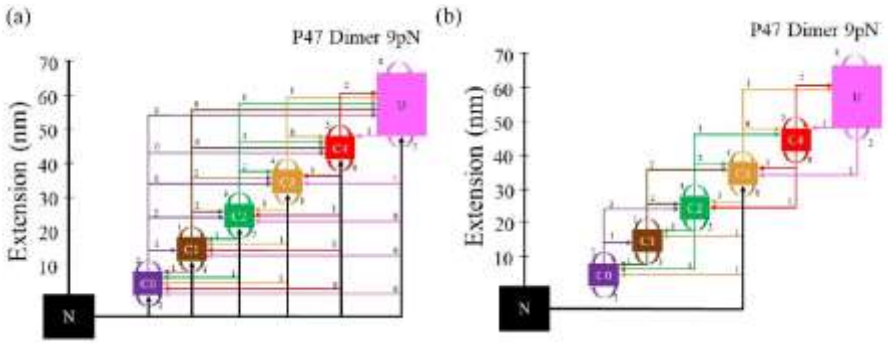
Supplementary Figure 18: Example of non-specific or multi-tethered bead data. Grey coloured force curves are the raw data tracing from drift corrected magnetic data. We use 50 points Savitzky-Golay filter to highlight the absence or very short initial entropic jump around 20^{th} seconds at 20 pN clamping force, which is well under 20 nm. These types of force-curves are considered to be the result of non-specific attachment or multi-protein tethering to a single bead.



Supplementary Figure 19: Identification of the specificity of the tether. A specific bead data is plotted to show the unfolding of Cdh23 EC1 P47 dimer at 15 pN and then I_{27} tetramer unfolding at 120 pN. Moreover, in the beginning there is a big entropic jump from 4 pN (>100 nm) and the refolding at 4 pN shows about 150 nm extension towards the native structure, which is the combined contour length of six-domains. All these features help distinguish the specificity of the tether which is then acted on for future experiments. For each experiment these specification study was performed in the very beginning.



Supplementary Figure 20: Trace of instrumental drift in data from magnetic and non-magnetic bead. Black coloured force curves are the raw data tracing from magnetic bead and non-magnetic bead. As there is a linear drift, we measure drift slope for baseline reference from non-magnetic bead. Red line fit of the linear drift is shown. If the drift is non-linear, in both magnetic or non-magnetic which is the case happens often when performing for a long-term clamp; we use 500 points adjacent averaging filter to quantify the drift pattern which later on helps in baseline subtraction.



Supplementary Figure 21: Process of cluster analysis using 10 nm bin (in support of figure 2). Here, P47 dimer is clamped at 13 pN force and divided into six-cluster using 10 nm bin size. C0 (0-10 nm, mean 5 nm ± 3 nm), C1 (10-20 nm, mean 15 nm ± 3 nm), C2 (20-30 nm, mean 25 nm ± 3 nm), C3 (30-40 nm, mean 35 nm ± 3 nm), C4 (40-50 nm, mean 45 nm ± 3 nm) and C0 (≥50 nm, mean 58 nm ± 7 nm). All transitions have been shown and since the resolution of our system in around 5 nm, we can see intra-cluster transitions too. a) In this figure, we map every refolding and unfolding on the energy landscape using extension as a proxy, which gives us the estimation of heterogeneity and several metastable populations. b) In this figure, all transition which show 0, are removed.

Clamping Force (pN)	Number of events (N)	Lifetime (τ)	Expdec1 fitting R^2
		P47 Refolding	
4	26	0.54±0.03	0.95458
5	27	1.34±0.1	0.9377
6	25	5.84±0.33	0.97052
7	33	7.64±0.3	0.98092
8	26	17.7±0.6	0.98614
		P47 Unfolding	
11	28	11.63±0.5	0.97164
15	30	9±0.37	0.97565
19	30	6.46±0.13	0.9942
		S47 Refolding	
4	49	0.4±0.02	0.94786
5	28	1.54±0.17	0.91622
7	53	4.13±0.06	0.99501
11	27	8.61±0.4	0.97387
		S47 Unfolding	
19	27	16.99±0.86	0.95993
23	31	10.6±0.4	0.97562
28	26	9.16±0.37	0.97051
		V47 Refolding	Expdec1
9	31	8.41±0.36	0.96025
13	39	10.68±0.35	0.98143
19	35	15.65±0.73	0.97417
		V47 Unfolding	
23	28	20.31±0.83	0.97429
31	26	16.35±1.04	0.95925
38	30	8.42±0.29	0.98204

Supplementary Table 1 : Force dependent refolding and unfolding lifetime data and the goodness of fitting (in support of figure 5). The survival probability of each plot has been fitted with single exponential decay. We have also fitted these to bi-exponential decay fitting, which gives higher error in many cases, because the fitting does not converge and a mutual dependency exits between the parameters. This suggests a certain degree of overparametrizing the fitting function for expdec2. However, no such error occurs for expdec1 function, $y = A1 * exp^{\left(-\frac{x}{\tau 1}\right)} + y0$, where we consider the offset y0 as 0. So, we choose the single-exponential decay function to fit all plots instead of bi-exponential.

Protein Variants	Fold-rate (k_0^f) (s^{-1})	Force-tolerance of folding, x_β^f (nm)	Unfold-rate (k_0^u) (s^{-1})	Force-tolerance of unfolding, x_β^u (nm)
P47	52.45±32.52	-3.6±0.4	0.04 ± 0.00	0.3±0.0
S47	6.49±5.84	-1.6±0.5	0.02 ± 0.01	0.3±0.1
V47	0.21±0.00	-0.3±0.0	0.01 ± 0.01	0.2±0.1

Supplementary Table 2 : Comparison of intrinsic refolding and unfolding rates and transition states distance (in support of figure 5). Values are calculated from $\ln \mathbf{K}_f = \ln \mathbf{K}_0 + \mathbf{F} * (\mathbf{X}_\beta / \mathbf{K}_B.T)$**.** From, Fig 4 (c), the intercept and slope are known which helps calculate the intrinsic rate of folding and unfolding and transition state distance, respectively. From the transition state distance from folded and unfolded state, X_β of unfolding and folding gives estimation of force tolerance, V47 > S47 > P47.

Magnet position from Voice-coil actuator (i)	Deviation count (Number of pulses) (N)	Unit deviation time (Δt)	Total target position dwell time (t)	Probability of error (W_i * 100) ($W_i = N*\Delta t/t$)
1 mm	1	0.06	30	0.2%
2 mm	2	0.06	15	0.8%
3 mm	15	0.06	15	6%
4 mm	15	0.06	15	6%
5 mm	96	0.06	17	33.88235%
5.7 mm	140	0.06	18	46.66667%

Supplementary Table 3 : Noise probability of the voice-coil actuator at different magnet position. We have calculated the error probability in VCA position. Although, the amplitude of the noise is almost always same (0.01 mm) but the frequency varies at various position of the voice-coil. As it is a mechanical device, having higher load automatically induces higher spatial error. We calculated the frequency of the noise at 1 mm, 2 mm, 3 mm, 4 mm, 5 mm and 5.7 mm and calculated the probability of error, respectively.

References

[1] I. Popa, J. A. Rivas-Pardo, E. C. Eckels, D. J. Echelman, C. L. Badilla, J. Valle-Orero, and J. M. Fernández, *A HaloTag Anchored Ruler for Week-Long Studies of Protein Dynamics*, J Am Chem Soc **138**, 10546 (2016).