Heterogeneity in conformational state space enhances the force-tolerance of mechanosensory proteins

Pritam Saha
IISER MOHALI
Vishavdeep Vashisht
Ojas Singh
Gaurav Bhati
Surbhi Garg
Sabyasachi Rakshit (srakshit@iisermohali.ac.in)
Indian Institute of Science Education and Research Mohali https://orcid.org/0000-0002-8083-6857

Article

Keywords:

Posted Date: March 23rd, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2677234/v1

Abstract

β-strands in proteins undergo anti-cross correlated crankshaft-type motions and adapt to the input mechanical cues. However, a direct study to reveal the molecular relation of force-adaptability with crankshaft motions of β-strands is long-awaited. To elucidate, here we explore the differences in mechanical tolerance of a gating-spring protein in hearing, cadherin-23, with genotypic and phenotypic variations on a single residue. Though the variants possess comparable topology, differ in contact-orders. Higher contact-order induces higher crankshaft. We identified that the variants with higher crankshaft exhibit larger heterogeneity in the conformational state space and thus, higher force-tolerance. However, protein-variants with lower contact-orders possess higher folding-cooperativity and faster intrinsic-folding, though their folding-energy landscape is most prone to distortion under tension. Overall, our study provides a unique relation between the transition-cooperativity amongst the sparsely populated conformational states and the force-adaptations by β-rich proteins. The use of phenotype and genotype variants also help us to deduce the mechanical fingerprinting of healthy spring and malicious spring.

Introduction

β-strand-rich proteins, in physiology, are naturally selected against large mechanical forces[1]. Examples include silk proteins that surpass the extreme tensile strength of a steel[2], muscle protein titin[3], cadherins as gating-spring in hearing[4–7], and many more. β-strands in proteins undergo strongly anticorrelated, ultrafast-frequency (in THz) breathing motions[8,9]. Such global motions in β-strands are correlated with crankshaft motion due to their unique resemblance to the reciprocating motion of the vehicle shaft, driven by a series of cranks and crankpins continuously[10]. In mechanosensory proteins, the crankshaft-type motion is attributed to force-adaptation that helps proteins to regain conformations after mechanical perturbations[8,10,11]. Here we aim to quantitatively decipher the importance of crankshaft in β-rich proteins that assists proteins in adapting repetitive mechanical assaults of varying magnitudes.

Force-induced folding-unfolding dynamics of proteins have been extensively characterized using single-molecule force spectroscopy[12–16]. Tweezer-based force probes along ultrafast atomic force spectrometers, have identified several microstates in the protein-folding funnel[12–14, 16]. Theoretical models with simultaneous use of all-atom simulations have also provided the molecular resolution of the microstate structures and developed the microscopic model of protein-folding[17–21]. However, the physical parameters that correlate the viscoelasticity of proteins with the microscopic protein folding model are poorly understood. Along this line, here we use the mechanosensing protein, Cadherin-23 (Cdh23), that possesses numerous β-strand-rich extracellular domains[22]. Cdh23 is a tip-link protein that actively participates in the mechanotransduction of hearing (Fig. 1A,B) [6,7], receives force pulses of various intensities and frequencies during the entire lifetime[23,24]. Notably, the protein is also one of the loci for multifactorial age-induced / noise-induced hearing loss[23–28], implicating the temporal loss of protein viscoelasticity with aging and, thus the loss of sensory abilities. In this study, we use three structurally near identical variants, geneotypic and phenotypic, with different force-response and probe their force-

adaptations directly using our laboratory-made magnetic tweezer (MT). MT enabled us to capture the microstates in the folding energy landscape under little mechanical tensions ranging from about 4 pN – 40 pN, and at a high spatial resolution of nanometres (~ 5 nm) and temporal resolution of milliseconds. Further, we devised a covalent tethering of protein variants with the glass coverslips and the magnetic beads to perform the unfolding-refolding studies at variable tensile forces repetitively and monitor for an extended period. Overall, our study here deciphers that the cooperative interactions facilitate proteins to undergo dynamic and reversible switches between metastable microstates under mechanical clamps. Such reversible switches among metastable states help proteins to withstand a large range of mechanical forces and prolong the repetitive signal transduction.

Results

Three variants of Cdh23 as model protein. We chose the first extracellular (EC1) domain of Cdh23 as the model protein. Cdh23 EC1 consists of nine β-strands interconnected by reverse β turns, 3_{10} α-helix, and random coils (Fig. 1C)[29,30]. We used three variants, Cdh23 EC1 (S47), Cdh23 EC1(V47), and Cdh23 EC1(P47), with different native packing densities, long-range H-bond networks, and different cross-correlated crankshaft types motions among β-strands[30] (Fig. 1D). Cdh23 EC1(S47) is a wildtype variant conserved for a majority of the species, including Homo sapiens (Hs), whereas Cdh23 EC1 (V47) is another wildtype variant conserved in lower order vertebrates like Callorhinchus milii (fish), Gekko japonicus (reptiles), anser cygnoides domesticus (Swan goose), Alligator mississippiensis (crocodile reptile), Gallus gallus (ave)[30]. Evolutionarily these species may be in a lower order than sapiens. However, some of these species possess better hearing sensitivity at a lower frequency range than humans[31,32]. Further, some of these lower vertebrates require hearing at low-air pressure (at high altitude for swans) or under-water pressure (for alligators) where the noise threshold is high[33]. Moreover, the evolutionary trend for proteins may not correlate with the ranks of the expressors. In line, Cdh23 EC1(V47) possesses the highest number of native contacts, long-range H-bonds, and thus most robust cross-correlated motions among β-strands [30](Fig. 1D). The last variant is Cdh23 EC1(P47) which is a mutant-variant of Cdh23 EC1(S47) that features a progressive hearing loss (PHL) phenotype in mice[34]. PHL is an aggressive form of hearing-loss with aging where a patient suffers complete hearing loss at a very early age (less than 20 years in humans)[31,35]. Regarding native packing, Cdh23 EC1(P47) ranks last among the three variants (Fig. 1D,1E). Reportedly, V47 shows the highest resistance against thermal and chemical denaturants, whereas P47 is the least[30]. Serine to proline isn't unique for Cdh23, relatively abundant with phenotypes[36]. Even for mechanosensitive Titin proteins, Ser22 to Pro mutation is reported with a phenotype of cardiomyopathy[37].

Covalent tethering of proteins and MT. To monitor the responses of protein variants to small tensile forces, we force-clamped the chimeric polyprotein constructs using MT. A detailed description of the MT including hardware, Instrument software and analysis software are described in the methods and supplementary Figs. 1, 9–13. Two repeats of Cdh23 EC1 variants were sandwiched between I_{27}

constructs, a trimer of I_{27} domains at the C-terminus and a monomeric I_{27} domain at the N-terminus (Fig. 1). The trimer of I_{27} domains was covalently attached to the glass-coverslip, and the I_{27} monomer at the N-terminus was attached to the paramagnetic bead, both covalently using sortase-mediated enzymatic stapling as described previously [38] (methods). The I_{27} domains serve as a spacer, and a standard marker for specific pulling as the mechanical stability of I_{27} is significantly higher than all three variants of Cdh23 EC1 [14,15](Supplementary Fig. 3). The assymmetry in the constructs is primarily to ease out the DNA recombinant process. We clamped the protein variants to varying forces for 5 minutes from a resting force of 4 pN and monitored the change in lengths (ΔL) from the jumps in bead positions in real-time (Supplementary Fig. 4–5). Notably, all protein variants, featuring the behaviour of Worm-like chain (WLC) polymer, showed a gradual increase in the end-to-end extensions, x(F), with force [39] (Fig. 4D, Supplementary Fig. 4). x(F) defines the extent of unfolding and, thus marks the final denaturant states at the respective clamping forces (Supplementary Fig. 4–5).

Number of long-range interactions, more precisely electrostatic H-bond interactions, follow decreasing order from V47 to S47 to P47 variants (Fig. 1, supplementary Fig. 6). Reportedly, the chemical and the thermal stabilities also followed the decreasing order from V47 to P47 variant[30]. Accordingly, we measured the weakest force-resistance for Cdh23 P47 featuring force-induced extensions between 5 pN – 15 pN, followed by Cdh23 S47 between 13 pN – 25 pN, and strongest for Cdh23 V47 between 19 pN – 38 pN (Fig. 2, Supplementary Fig. 5, 7). Irrespective of the force- resistance, we identified multiple microstates for all the variants during unfolding-refolding transitions at small clamping forces. To characterize the microstates, we measured the length-change (ΔL) for each step and plotted them as distributions (Methods and Fig. 2). The histograms of ΔL for all variants followed comparable trend, narrow distributions at lower forces, wide at intermediate forces reaching saturations, and finally sharp distributions at very high forces. The widths of the distributions indicate the extent of reversible transitions among microstates under tension while the number of peaks in the distributions infers the number of microstates within the spatial limit of the MT.

To further highlight the heterogeneity in the microstates across variants, we plotted the fraction of states or probability of states as they survive at the clamping forces (Fig. 3). For simplicity, we used only 3-states, a native or folded state (with 0 nm $\leq \Delta L \leq$ 15 nm), a complete denatured or unfolded state (45 nm $\leq \Delta L$, maximum x(F) is obtained from Fig. 4 unfolding length data), and a cluster of intermediate states (15 nm $\leq \Delta L \leq$ x(F)). The probability of the states at respective force-clamps was estimated from the dwell time of each state, normalized to the total clamping time. To quantitatively estimate the dwell time of steps, we performed 'Autostepfinder', a widely used step-finder protocol that is based on the mean standard deviation model [40]. We noticed a sigmoidal transition for the probability of folded state with force for all variants. The transitions from native to unfolded states occurred at different critical forces (F_{crit}), lowest for P47 (8.8 ± 0.2 pN) and largest for V47 (25.7 ± 0.2 pN). However, strikingly we noticed different widths (ΔF) or slopes of the transitions across variants. We highlighted these regions with light green boxes (Fig. 3). The green boxes marked the co-existence of all 3 states, indicating the force-range at which the protein-variants undergo reversible transitions between numerous conformational states. While

the critical force-range (F_{crit}) of transition is a measure of force-resistance, the width of the transitions (ΔF) measures heterogeneity in the microstates during the transition from folded to unfolded states. More the width, more the heterogeneity, and hence more the shock-absorptivity or force-dissemination. Together, the critical force-range of transitions and the width measures the force-adaptation range. Clearly, the V47 variant is not only the most resistant to tensile forces among the three, it also possesses maximum heterogeneous states during phase transitions whereas P47 has least.

Next, we mapped the probability of the open state of all three variants and fit the rise in states to linear equation, excluding the force-value points with no unfolding probability. The slope gives quantitative estimation of protein's unfolding cooperativity. Conceptually, a completely cooperative unfolding will show a strict two-step transition with an infinite slope, whereas the completely non-cooperative transition will have slope of 0. Here, the slope value for P47 is 0.15 ± 0.01/pN, S47 is 0.09 ± 0.01/pN, and V47 is 0.05 ± 0.01/pN. The order of unfolding cooperativity is, P47 > S47 > V47. Overall, our data from force-clamp experiments indicate that Cdh23 V47, which contains the most densely packed network of intra-domain interactions, is mechanically the most stiff yet most disseminating of tensile forces. We, therefore, infer that higher crankshaft or contact order transforms β-rich proteins to more malleable under tensile forces, however, without causing any permanent damage. The inter β-strand interactions serve as shock-absorber and resist elongation from mechanical inputs.

Effect of intra-domain contacts in folding rates and force tolerance. Contact-orders contribute to the folding dynamics of proteins. Further, mechanoresponsive proteins undergo unfolding-folding periodically under tension in physiology. The folding dynamics of the mechanoresponsive proteins are, thus functionally important. We, therefore, set to decipher the effect of inter domain interactions in the folding and unfolding kinetics.

We performed a dual-step force-clamp spectroscopy with a monomer of Cdh23 EC1 domain parsed between I_{27} domains as -Nterm-$(I_{27})_3$-EC_1-I_{27}-Cterm (Fig. 5). The use of a monomeric domain enabled us to measure the kinetics for direct transitions from the native state to the complete denatured state and vice versa. The resting force was set to 4 pN. Subsequently, the protein variants were individually clamped at varying constant forces (19–38 pN for S47 variant, 11–19 pN for P47 variant, and 23–38 pN for V47 variant) for 30–60 seconds and monitored the survival time till complete unfolding. We excluded the extensions below 5 nm in the analysis due to the resolution limit in our measurements (Supplementary Fig. 2). After each force-clamp, we subjected the variants to a high tensile force of 38 pN for 15 seconds to ensure complete extension prior to refolding. For refolding, we clamped the proteins at low forces, 11 − 4 pN for S47 variant, 8 − 4 pN for P47 variant, and 19 − 9 pN for V47 variant, respectively. We completed the cycle by quenching the force to 4 pN. For each variant, we monitored a minimum of 30 different beads (Supplementary Table 1).

The survival probability of the folded and unfolded states was estimated from the dwell time (Fig. 5). We deduced the lifetimes of folding and unfolding from the single exponential decay fit to corresponding survival probabilities (Fig. 5B) (Supplementary Table 1) and obtained the force-induced on-rate and off-

rate data (Fig. 5D). Subsequently, the variation in rates with force were fit to Bell's model[41] and obtained the intrinsic kinetic parameters, $k_0^{f/uf}$ (intrinsic-rate at zero force, 'f' represents folding and 'uf' refers to unfolding) and $x_\beta^{f/uf}$ (the width of the potential energy barrier) (Fig. 5C and Supplementary Table 2).

Though intrinsic fold-rates at the 'no-force' condition may not be physiologically relevant for force-sensors, we measured the fastest unfolding rate for P47 and the slowest for V47 (Supplementary Table 2). In continuation to unfold trend, we measured the fastest refolding for P47 (52.45 ± 32.52 s^{-1}) and the slowest for V47 (0.21 ± 0.00 s^{-1}) among the three. We, thus, inferred that the inter β-strand interactions drive the refolding kinetics, a protein having the least intra-domain interactions forms the native state fastest. The susceptibility of proteins towards force-induced unfolding-folding, however, depends on the distance to transition states, $x_\beta^{f/uf}$. x_β^f is negative in sign as folding is opposite to the force vector (Supplementary Table 2). A higher x_β^f diminishes the folding probability under tension by increasing the barrier-height ($E(F) = E_A - \vec{F}.\vec{x}_\beta^f$), whereas a higher x_β^{uf} increases the unfolding probability by reducing the potential barrier height under tension. We measured the longest $x_\beta^{f/uf}$ for P47 ($x_\beta^f \sim$ -3.6 and $x_\beta^{uf} \sim$ 0.3) and the shortest for V47 ($x_\beta^f \sim$ -0.3 and $x_\beta^{uf} \sim$ 0.2), indicating that the mechanical tension is most vulnerable for P47 and least for V47 (Fig. 5D).

Time-dependent adaptation against fatigue while receiving repetitive force pulse:. The time-dependence of a mechanoresponsive protein is often correlated with the molecular-fatigue which arises from the repetitive stretch and release cycle of the protein[42,43]. Cdh23 too experiences repetitive force perturbations from the sound. It is thus interesting to monitor the molecular fatigue of Cdh23 and its relation to force-adaptibility. We, therefore, exposed the protein variants to a cycle of high force and low-force (4 pN), repetitively. The proteins were completely unfolded for 15 seconds at the respective critical unfolding forces (For S47 protein 23 pN, P47 protein 15 pN, and V47 protein 38 pN) and then refolded at a refolding force of 4 pN and waited for 15 seconds (Fig. 6A). The critical forces of unfolding and refolding were selected based on the data from the chevron plot of folding-rates and the probability of states plot. (Fig. 5C).

We noticed two distinct unfolding patterns, direct unfolding (unfolds directly along with the initial stretch) and delayed stepwise unfolding (goes through the metastable states before complete unfolding) (Fig. 6A). Counterintuitive to the folding-kinetics pattern, we observed the more frequent direct unfolding of P47 (20.7%) than S47 (11.3%) and V47 (6.25%), respectively. It is thus, logical to relate the direct unfolding events with the delay in folding in the previous force-cycle, a feature of molecular fatigue. Overall, we infer that proteins with low force- adaptibility, suffer faster mechanical fatigue. More importantly, a direct unfolding relates to 'no force-buffer' and conveys the input mechanical signal directly, exposing the sensory organs to potential damage.

Discussion

Mechanical cue is abundant in physiology. Mechanoresponsive proteins, abundant in nature too, have evolved to sense the different magnitudes of forces and fulfil functional need. The overarching objective of this work is to identify the parameters that outfit proteins in different mechanical environments. We used three variants of a mechanoresponsive protein, Cdh23, that reflect different phenotypes in physiology. Two wildtype variants of Cdh23 EC1, S47 and V47, serve as gating-spring and mediate normal hearing in different environments. The mutant variant, Cdh23 EC1 P47, however, reflects a detrimental effect with time and causes progressive hearing loss in mice. In short, both S47 and V47 variants are naturally selected for hearing to withstand and function under tension, whereas P47 is rejected. Though, all three variants possess comparable static structures, β-strand rich, with minuscule differences in the contact orders or inter β-strand interactions. The location of the 47th residue is either at the terminal of a β-strand or at the turn in-between two β-strands (Fig. 1C). P47 resides at the antiparallel β-turn whereas both S47 and V47 are at the terminal of the β-strand. Importantly, turns in proteins are the foundations of the β-strands[44] and key regulators of inter-strand interactions[45,46]. Pro brings restricted torsion angle in the turn and makes the β-strand packing comparatively weaker among the three, leading to a lesser anti-cross correlated crankshaft-type motion among β-strands[46–49]. How such trivial alterations in topology severely affect the mechanics of proteins leading to extreme phenotypes is the primary objective of this work. We, thus, clamped all three variants of Cdh23 EC1 under tension and deciphered their differences in overall force-adaptability.

We noticed maximum mechanical stiffness for the V47 variant with a critical transition force (F_{crit}) from folded to unfolded state of 25.7 ± 0.2 pN, followed by S47 with a F_{crit} of 17.5 ± 0.2 pN and lastly, P47 with a F_{crit} of 8.8 ± 0.2 pN. Notably, all three variants crossover multiple metastable states reversibly during the transitions between folded and unfolded states. The width of the transition (ΔF) from the folded to unfolded state measures the heterogeneity in the metastable states. We measured the largest heterogeneity for V47 variant and the least for the P47, indicating that the V47 has enough room in conformational state space to counteract the tensile forces and avoid any irreparable damages under loud sound. ΔF thus, is a direct measure of the force adaptation and the F_{crit}, where the protein attains the heterogeneity indicates the force tolerance.

We also noted that the Cdh23 V47 is not only the most heterogeneous, but it also ill-responsive to lower force-regime. As this variant is found in aquatic animals with higher thresholds of noise, we hypothesized that little or no response at a lower force-range may assist proteins to filter noise among mechanical cues. Cdh23 P47 on the other hand starts showing metastability at a very low force, which tends to make the protein malleable to forces. As the physiological range of hearing in humans range vast, the malleability of Cdh23 P47 reduces the adaptability against force in the long run and against loud noises, and possibly leads to a progressive hearing loss.

The trend in the intrinsic folding-rate is, however, reverse in trend than the force-tolerance, fastest for the P47 variant and slowest for V47. Weaker the local interactions, lesser time the protein requires to collapse and fold. However, intrinsic fold-rate may not be a correct parameter to evaluate the mechanoresponsive

proteins as these proteins are always under tension. Rather, the adapting-rates of folding/unfolding under mechanical tension are physiologically more relevant. We measured the longest x_{β}^{f} for the P47 variant, indicating that the force-induced increase in the free-energy barrier for folding is maximum for the P47 and least for the V47. Thus, a small force in physiology can hike the barrier-height for the P47 to a larger extent and practically block the refolding under tension. The changes in barrier-height of folding are significantly shallow for S47 and V47 than for P47. The trend in force-induced unfolding is, however, reverse. x_{β}^{uf} is longest for the P47 variant, indicating that the tensile forces reduce the free-energy barrier for unfolding to maximum extent for P47 and least for V47. Thus, P47 is most susceptible to unfold under little tensile forces and may not even serve as mechano-protein at ambient temperature.

Mechanoresponsive proteins receive repetitive tension for signal cues. It is thus important for mechano-sensing proteins to adapt to the periodic mechanical impulses faster than the pulse frequency. A delay in the response to force pulses is defined as mechanical fatigue, a measure of the extent of the fast adaptations to a periodic mechanical tension. From the repeated stretch-release activity in the force-pulse experiments, we noticed direct jumps to the unfolded state from the folded one for Cdh23 P47 more frequently than the two other variants. Although the etiology of this behaviour is not completely certain, it can be contributed to the loss of internal contacts during refolding and forming a quasi-folded closed conformation.

We correlated the force adaptation of all three variants with unfolding cooperativity[50]. Cooperativity measures the sharpness of configurational or conformational transition[51]. Higher the cooperativity, lower is the number of kinetic traps, thus lesser chance of misfolding, however, higher is the sensitivity to mechanical perturbations[52]. Thus, P47 possessing highest cooperativity, showed the highest unfolding probability, especially during repetitive force-pulses.

Conclusions

Using force-spectroscopy, we directly probed how mechano-sensitivity and mechano-adaptability of β-rich proteins depend on the anti-cross correlated motions among β-strands. In summary, our study opens up the possibility of looking into force adaptation by studying the transition dynamics using effective and robust single-molecule techniques.

Material And Methods

Magnetic Tweezers Setup

The tweezer is composed of a voice coil actuator (VCA, from BEI Kimco), a high-speed CMOS camera (Ximea xiQ USB 3.0 SuperSpeed), an inverted microscope (Olympus IX73), an objective piezoscanner (P-725.xDD PIFOC) and controller (PI- E-709), and a python-based algorithm for automation and data acquisition. VCA works in closed-loop with a z-resolution of 10 μm and is controlled by Ingenia Motion Controller (PLU-1/5-48C). We use VCA as a linear actuator to move permanent magnets (Neodymium

permanent ring magnets from K&J Magnetics, USA) and exert force on superparamagnetic beads (2.8 µm diameter, Fe_3O_4). Beads are covalently attached to a surface via the protein to be probed. The entire setup is in the optical microscope, as shown in SI Fig. 1. Bead movement in the z-direction under force is captured using the CMOS camera, and processed in real-time using a Python-based algorithm. The temporal z-displacement of the magnetic particle features the unfolding and refolding patterns of the probe protein under varying forces. A representative scheme and a picture of our home-built MT are shown in Figs. 2A and supplementary Fig. 1. Along with the magnetic beads, reference beads are also used for drift correction (See section 3 in SI).

Command Line Interface-based Custom Written Code For Image Capture And Analysis At High FPS

We plan to capture images containing beads at a very high speed, save and transfer those data to CPU, and simultaneously run image-processing algorithm to get real-time data. Alongside, we also run a visualization programme to monitor the fluctuations of beads in real time. To perform these high-end parallel computing, there are several approaches already exist. We have used multiprocessing to spawn different instances of tracking function (workers) continuously grabbing fresh frames from the queue and submitting the z-position data to another thread (recorder). This gives a significant boost in tracking fps compared to traditional sequential algorithms. We have used a 32-core (64 thread) AMD Ryzen Threadripper Pro 3975WX processor and were able to easily saturate USB bandwidth of the CMOS camera, thereby achieving the highest image acquisition rate (3.5 KHz) while performing heavy tracking computation in real time (supplementary Fig. 2 and supplementary Fig. 9).

Stack Collection And Image Processing

The z-positions of the magnetic beads determine the length change of the biomolecule extension. The magnetic beads are covalently and specifically attached to the protein variants. The nonmagnetic reference beads that are used for drift corrections are fixed by physisorption to surface non-specifically. We have used a 32-core CPU system with AMD Ryzen Threadripper processor to perform image processing and real-time bead tracking. To measure the real time bead position with continuous feedback, we have built a library of images with offset in z, which is referred here as stack. To create the stack, the piezoelectric objective scanner is moved up in z-direction to 2 µm at a step size of 10 nm. During a flight between steps, 100 snap frames of the two beads are taken for future averaging and error correction to counter the miniscule changes from the instruments. A 2D FFT for each of the 100 frames at every step was computed and an average frame based on the pixel intensity matrix was calculated. After that, a region of interest (ROI) of 128x128 pixels is selected from the 10% area square box from the central point because it is the region that is most sensitive to focus change. 2D FFT helps to eliminate the x-y fluctuations coming from the beads. The ROI intensity matrix is then multiplied by a factor of 20, to distinctly differentiate the position of the frames and increase the resolution. After this library of images

and their 2D FFT and ROI intensity matrix are computed, the radial profile of these images was calculated using pixel intensity value. During experiment, the z-position of the beads are determined by a correlation function, which is used to match the radial profile from the stack library (supplementary Fig. 12).

Filtering Methods

Acquiring high speed data comes with its pros and cons. Force spectroscopy data from the real-time experiment exhibit slowly changing patterns and oscillations with abrupt transitions. Except for frequency filters, all other existing filtering methods like adjacent averaging or Savitzky-Golay filter causes loss of pattern information and thus affect the goodness of the fit. However, the Fourier transformation does not represent abrupt changes or step-like transitions efficiently. It will not be a good filtering method since it represents data as a sum of various waveforms like sine wave, etc. which are not localized in time and frequency. That is where we applied the wavelet transformation method for better fitting of steplike function and for more robust filtering. The wavelet transformation technique can decompose a data signal into several lower resolution steps in terms of both time and frequency, simultaneously. It is computationally cheap and fast. Depending on the wavelet that we choose to use to convert the data into equally spaced samples, the filtering goodness increases. For magnetic tweezer related experimental data, we use the haar wavelet because of its nature for approximating sudden step-like jumps with unfolding and refolding behaviour of proteins. Haar wavelet, a rectangular shaped waveforms with varying amplitude, is used to fit to a low frequency signal after sequentially separating high frequency noise (supplementary Fig. 13).

Data analysis

Step-fitting analysis for force curves from short time experiments were carried out using custom written MATLAB program (using version of MATLAB 2015b and MATLAB2021a, MATLAB2022a). However, it took more computational time in case of high fps (frame per second) data from experiment of longer period. So, to counter this problem, we used open-source resource. Auto-stepfinder, a GUI based MATLAB programme developed from the lab of Prof. Chirlmin Joo with collaboration of Cees Dekker, was used to fit the step function with the raw data. This programme also gives us an estimation of goodness of fit by providing the S-value, which is basically the ratio of sum of variance between data points and the fitted line of a counter-fit vs. existing fit. Other fitting analysis like dwell time and folding rate analysis, has been done using OriginPro software. Linear fitting, Gaussian peak fitting, kernel density plot, and single-exponential decay fitting has been done using the same software. For data visualisation as well as to highlight the three states in the force-clamp results (supplementary Fig. 11), we used our own custom written programme in Python.

Force Calibration Using DNA

For the calibration of force using a pair of permanent neodymium-grade N52 magnets, a force ramp experiment was performed to observe the DNA B-S transition. For the sample, a 605 bp long DNA linker was cloned from λ-phage DNA (Thermo Scientific). The ends were modified with an amine group and biotin using standard PCR reaction protocol. The amplified DNA was identified in UV light (302 nm, 365 nm) and purified using the PCR clean-up protocol (Geneoid). DNA was incubated in PEG buffer with NHS-$(PEG)_2$-NHS modified surface for 4 hours, and then after washing the surface it was incubated with a Streptavidin-coated bead for 20 minutes. Later after setting up the fluid chamber and measuring the highest length of voice coil movement along the z-axis and we performed force ramp experiment to observe B-S overstretching. From the well-known magnet law modified with fitting parameters[15] we get- $F = F\left(B-S\right) * e^{b(MP(B-S)-MP)}$. We know the value of F_{B-S} to be 65 pN. We have also found a similar step-size distribution of maltose binding protein (MBP) at similar forces and the same step-size for $I_{27,}$ with respect to force (supplementary Fig. 15), so we have considered the fitting parameter value to be 0.9[53]. With these constants and after measuring the magnet distance for B-S overstretching, we have our final magnet law- $F\left(MP\right) = 65 * e^{0.9(4.31-X(MP))}$, where X is the magnet distance. Using this equation and magnet position we can get desired clamping force between 4 pN -165 pN (supplementary info 14, 16). The noise from the voice-coil actuator has also been taken into account for minute changes in the force during the experiment (supplementary Fig. 17, supplementary table 3)

Specific Tether Identification

Our chimeric construct isn't isotropic. We have a trimer and monomer of titin domains at different ends of the Cdh23 EC1 dimer. To maximise the success of recombinant reactions during cloning, we created this anisotropic design. MT experiments are successful only when a single-tethered magnetic bead is observed and not a multitethered single bead. To address this issue, we optimised the surface preparation protocol, with varying APTES concentrations of 5–0.5%. In the protocol with low concentration of APTES, we observe the least number of multi-tether events. We can differentiate nonspecific or multitethered events by quantifying the initial entropic extension (< 20 nm) at high force (> 20 pN) (supplentary Fig. 18). Specific events where we identify 3 or more distinct unfolding features of I_{27} at high force (> 100 pN) and the corresponding initial length jump shows a higher initial extension (> 50 nm), where the orientation of domains along the force directions also contributes to the initial unfolding entropic extension (supplementary Fig. 19). At very high force, the enthalpic extension contributes to overall distance and residue level entropic extension comes into play.

Effct Of Drift On Clamping Force

Force measurement in magnetic tweezers cannot be calibrated directly, unlike AFM, so it relies on indirect measurements of the position of the beads. The force of the magnetic tweezers is determined by the distance between the magnet and the bead. We measured the extension due to drifting and from that distance how much the clamping force changes (supplementary Fig. 20). According to the formula for

force calibration, if the baseline force drift changes from 4 pN to 5 pN, the distance will have to change by more than 0.75 mm. The bi-directional repeatability of our translation piezo is less than 20 μm (18 μm to be exact). However, because of the focal range, we limit this movement to 2 μm only. Not only that, for a limited time measurements we see the drift of about 200–300 nm which is even lesser than the whole limit of piezo movement. Therefore, the force is almost intrinsic constant with error of less than 0.001 pN. The drift observed on extension (Supplementary Fig. 20) is mainly due to the focus drift of the microscope or buffer leakage from the MT chamber. Except for instrumental drift, there are two sources of error that could change the clamping force. Firstly, the z-positional error of the objective in mounting piezo, and secondly, the linear displacement due to the vibration of the voice-coil actuator, both of which have been neglected during force-calibration measurements.

Network Analysis

We performed the molecular dynamics simulation for all three proteins (S47, P47, and V47) using the QuickMD plugin in VMD[53,54]. The crystal structure solved for the S47 and P47 was taken from PDB database. But since the structural data for V47 were not available, we made the variant V47 using the structural data from WT S47 in VMD (Fig. 1). We used the TIP3P water system to solvate the molecules and the ions Na^+ and Cl^- were randomly placed at the final concentration of 15 mM. We used NAMD version 2.14 [55] and CHARMM36[56] force field to run the simulation. Initially, the system was minimised for 0.1 seconds and then increased the system temperature to 300K at the rate of 1 K for 600 steps. Followed by equilibration for 2 ns. Then we run the GAMD[57,58] for 300 ns. System pressure was maintained at 1 atm using the Noose-Hover method and long-range interactions were controlled by Particle Mesh Ewald (PME)[59,60]. We did the dynamic network analysis using the network view plugin into VMD[54]. A model for all three proteins was generated using a coarse-grained representation. In this model, we assign C_α atom of amino acids by a node, and nodes are connected to each other within a cut-off distance of 4.5 Å for 75% of the MD trajectory. Neighbouring C_α atoms were not considered for the analysis. The edge distance between two nodes is calculated from the pairwise correlation, which gives the probability of information transfer across the given edges. Edge's weight was given by the correlation coefficient, which is measured from the correlation matrix using the programme Carma[61] (supplementary Fig. 6).

Cluster Analysis

We characterised all transitions during continuous force clamp study and mapped those transitions in energy landscape using extension length as a proxy. The complete unfolding length of our two-domain construct is 70 nm. However, depending on the force, it can vary between 50–80 nm. Here, we have taken six clusters. C0 (0–10 nm), C1 (10–20 nm), C2 (20–30 nm), C3 (30–40 nm), C4 (40–50 nm), U (> 50 nm). First, we consider each and every step position under 10 nm and map the next step position in the cluster matrix. Since we have both unfolding and refolding and our resolution is about 5 nm, from the C0 cluster the protein can go to CO, C1, C2, C3, C4, U as an unfolding transition and can also come to C0 itself as a

refolding step. Similarly, from C1, it can go C1, C2, C3, C4, U as unfolding and come back to C1 and C0 as refolding. Same way we can map the number of all transitions in all possible position in cluster and their next transition direction. The transition line with value of 0 can later be removed for the sake of visualisation and thus the hetrogeneity can be mapped using the matrix clustering process. Needless to say, the bin size can be changed depending on the resolution limit as well as the proteins transition pattern. For smaller proteins, this size can be reduced to 3–5 nm and larger or multidomain proteins can use a bin size larger than 10 nm (supplementary Fig. 21).

Covalent Attachment Strategy

For performing long-term robust studies using magnetic tweezers all connections were prepared covalently.

1. Surface Preparation

Coverslips were kept in the plasma chamber for 30 seconds. Plasma treatment will expose the gas inside the chamber to an energy source like a microwave and make the air into a free radical, ion mixture. This will remove all organic contamination. The coverslips were then kept in piranha solution (75% H_2SO_4, 25% H_2O_2) for 1 hour to remove all remaining traces of organic components. These surfaces were sonicated in Milli-Q water 4–5 times to remove the traces of acid. Then, it was aged with 1 M KOH solution for 2 min[62]. Again, the sonication procedure was repeated with Milli-Q water. The water was changed after each wash. These surfaces were subjected to silanization in a solution mixture of 48 ml of acetone, 0.5 v/v (3-Aminopropyl) triethoxysilane (APTES) (0.5 ml), and 1.5 ml of Milli-Q water. It was incubated for 45 minutes. Then, they were washed with Acetone by sonicating for 5 minutes. The same process was repeated with Milli-Q water. After the last wash with acetone, the surfaces were kept in a vacuum oven at 110°C for 1 hour. This process will help expose the ends of the alkoxysilane molecules that were coated during APTES incubation. Pegylation was carried out on the surfaces with NHS-$(PEG)_2$-Maleimide, mPEG-SVA or whatever modification was necessary to attach proteins and beads in the presence of PEG buffer (100 mM $NaHCO_3$, 600 mM K_2SO_4, pH 8.3)[62]. This reaction will take 4 hours. After that, sonicate the coverslips in milli-Q for 5 minutes. Next, an incubation was performed on the surface with 100 µM tetra-glycine peptide (GGGGC) that will interact with maleimide from the PEG molecule, in SEC buffer (pH = 7.5) for 7 hours[63]. Then after washing is done, the incubation was done for the surface with protein and sortase A mixture (1 µM protein and 1–2 µM Sortase A) for 2 hours [64]. All of these reactions are performed at room temperature.

2. Bead Preparation

Two types of beads are used for performing the force spectroscopy experiments in magnetic tweezers: Magnetic beads and non-magnetic beads. Nonmagnetic beads, which were used as reference for correcting surface-related drift, were attached using adsorption for 30 minutes. However, two types of modification were done with M-270 amine-modified paramagnetic beads[7] of 2.8 µM diameter. Beads

were first incubated in NHS-$(PEG)_2$-NHS solution of PEG buffer for 7 hours. The PEG chemicals that were not used in the pegylation reaction were removed from the buffer to increase the efficiency and yield of the second reaction.To make a covalent bond with the protein of interest, these beads with N-hydroxysuccinimide ends were incubated in a LPETGSSC peptide solution of pH 7.5 for 7 hours. It was washed with the same buffer condition that will be used for the experiment. To make noncovalent interactions using a streptavidin-biotin combination, these beads were incubated with a streptavidin protein solution of pH 7.5 in the same buffer condition for the experiment. Later the non-reactive streptavidin that remained in the buffer was removed. All of these reactions are performed at room temperature with continuous shaking.

Declarations

Author Contributions

S.R. conceived the idea. S.R., P.S., V.V. build the MT instrument. O.S. wrote the code for image processing, filtering, and data acquisition from MT system. P.S. and V.V. designed and performed all the MT experiments. P.S. performed all the chemical modification and surface chemistry for tethering. P.S. and V.V. analysed the data. G.K.B. performed all the MD simulations. P.S. expressed and purified all the protein variants. S.G. cloned all the chimeric constructs. S.R. and P.S. made the figures. S.R. and P.S. wrote the manuscript.

Conflicts of interest

The authors declare no competing interests.

Acknowledgements

This work was supported by the Core Research Grant from SERB, Govt. of India.

SR acknowledges and thanks Dr. Raj K. Ladher, NCBS, for generously donating the Cadherin-23 plasmid. SR acknowledges the financial support by the DBT/Wellcome Trust India Alliance and Indian Institute of Science Education and Research Mohali (IISERM) for setting up the molecular biology facilities. PS is thankful to CSIR-India for providing fellowship. V.V., G.K.B., S.G. sincerely thank IISERM for fellowships.

References

1. Xiao, S., Xiao, S. & Gräter, F. Dissecting the structural determinants for the difference in mechanical stability of silk and amyloid beta-sheet stacks. *Physical Chemistry Chemical Physics* **15**, 8765–8771 (2013).
2. Vollrath, F. & Knight, D. P. Liquid crystalline spinning of spider silk. *Nature* **410**, 541–548 (2001).
3. Pfuhl, M. & Pastore, A. *Tertiary structure of an immunoglobulin-like domain from the giant muscle protein titin: a new member of the I set.* Structure **3**, 391-401 (1995).

4. Howard, J. & Hudspeth, A. J. *Compliance of the Hair Bundle Associated with Gating of Mechanoelectrical Transduction Channels in the Bullfrog's Saccular Hair Cell*. *Neuron* **1**, 189-199 (1988).
5. Hudspeth, A. J., J. H. J. Allum,M. Hulliger, Mechanoelectrical transduction by hair cells of the bullfrog' sacculus. in *Progress in Brain Research* **80**, 129–135 (1989).
6. Siemens, J. *et al.* Cadherin 23 is a component of the tip link in hair-cell stereocilia. *Nature* **428**, 950–955 (2004).
7. Kazmierczak, P. *et al.* Cadherin 23 and protocadherin 15 interact to form tip-link filaments in sensory hair cells. *Nature* **449**, 87–91 (2007).
8. Piazza, F. & Sanejouand, Y. H. Discrete breathers in protein structures. *Phys Biol* **5**, 026001-026014 (2008).
9. Fadel AR, Jin DQ, Montelione GT, Levy, *Crankshaft motions of the polypeptide backbone in molecular dynamics simulations of human type-co transforming growth factor*. *Journal of Biomolecular NMR* **6**, 221-226 (1995).
10. Fenwick, R. B., Orellana, L., Esteban-Martín, S., Orozco, M. & Salvatella, X. Correlated motions are a fundamental property of β-sheets. *Nat Commun* **5**, 4070-4079 (2014).
11. Garg, S. *et al.* Weakening of interaction networks with aging in tip-link protein induces hearing loss. *Biochemical Journal* **478**, 121–134 (2021).
12. Schönfelder, J., Perez-Jimenez, R. & Muñoz, V. A simple two-state protein unfolds mechanically via multiple heterogeneous pathways at single-molecule resolution. *Nat Commun* **7**, 11777-11784 (2016).
13. Hong, H. *et al.* Two energy barriers and a transient intermediate state determine the unfolding and folding dynamics of cold shock protein. *Commun Chem* **4**, 156-163 (2021).
14. Löf, A. *et al.* 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* **116**, 18798–18807 (2019).
15. Popa, I. *et al.* A HaloTag Anchored Ruler for Week-Long Studies of Protein Dynamics. *J Am Chem Soc* **138**, 10546–10553 (2016).
16. Dahal, N., Sharma, S., Phan, B., Eis, A. & Popa, I. Mechanical regulation of talin through binding and history-dependent unfolding. *Sci Adv* **8**, eabl7719 (2023).
17. Beauchamp, K. A., McGibbon, R., Lin, Y.-S. & Pande, V. S. Simple few-state models reveal hidden complexity in protein folding. *Proceedings of the National Academy of Sciences* **109**, 17807–17813 (2012).
18. Jacobs, W. M. & Shakhnovich, E. I. Structure-Based Prediction of Protein-Folding Transition Paths. *Biophys J* **111**, 925–936 (2016).
19. Voelz, V. A., Bowman, G. R., Beauchamp, K. & Pande, V. S. Molecular Simulation of ab Initio Protein Folding for a Millisecond Folder NTL9(1–39). *J Am Chem Soc* **132**, 1526–1528 (2010).

20. Galzitskaya, O. v & Finkelstein, A. v. A theoretical search for folding/unfolding nuclei in three-dimensional protein structures. *Proceedings of the National Academy of Sciences* **96**, 11299–11304 (1999).
21. Gopi, S., Aranganathan, A. & Naganathan, A. N. Thermodynamics and folding landscapes of large proteins from a statistical mechanical model. *Curr Res Struct Biol* **1**, 6–12 (2019).
22. Jaiganesh, A. *et al.* Zooming in on Cadherin-23: Structural Diversity and Potential Mechanisms of Inherited Deafness. *Structure* **26**, 1210-1225 (2018).
23. Johnson, K. R., Erway, L. C., Cook, S. A., Willott, J. F. & Zheng, Q. Y. A major gene affecting age-related hearing loss in C57BL/6J mice. *Hear Res* **114**, 83–92 (1997).
24. Bovo, R., Ciorba, A. & Martini, A. Environmental and genetic factors in age-related hearing impairment. *Aging Clin Exp Res* **23**, 3–10 (2011).
25. Chang, S. J., Chen, C. J., Lien, C. H. & Sung, F. C. Hearing loss in workers exposed to toluene and noise. *Environ Health Perspect* **114**, 1283–1286 (2006).
26. Morata, T. C. Chemical Exposure as a Risk Factor for Hearing Loss. *J Occup Environ Med* **45**, 676-682(2003).
27. Kurien, M., Thomas, K. & Bhanu, T. S. Hearing threshold in patients with diabetes mellitus. *J Laryngol Otol* **103**, 164–168 (1989).
28. Lautermann, J., Crann, S. A., McLaren, J. & Schacht, J. Glutathione-dependent antioxidant systems in the mammalian inner ear: effects of aging, ototoxic drugs and noise. *Hear Res* **114**, 75–82 (1997).
29. Sotomayor, M., Weihofen, W. A., Gaudet, R. & Corey, D. P. Structural Determinants of Cadherin-23 Function in Hearing and Deafness. *Neuron* **66**, 85–100 (2010).
30. Garg, S. *et al.* Weakening of interaction networks with aging in tip-link protein induces hearing loss. *Biochemical Journal* **478**, 121–134 (2021).
31. Mizutari, K. *et al.* High prevalence of CDH23 mutations in patients with congenital high-frequency sporadic or recessively inherited hearing loss. *Orphanet J Rare Dis* **10**, 60-69 (2015).
32. Bierman, H. S. & Carr, C. E. Sound localization in the alligator. *Hear Res* **329**, 11–20 (2015).
33. Somero, G. N. Adaptations to High Hydrostatic Pressure. *Annu Rev Physiol* **54**, 557–577 (1992).
34. Han, F. *et al.* A new mouse mutant of the Cdh23 gene with early-onset hearing loss facilitates evaluation of otoprotection drugs. *Pharmacogenomics J* **12**, 30–44 (2012).
35. Seldran, F. *et al.* Relationship Between Age of Hearing-Loss Onset, Hearing-Loss Duration, and Speech Recognition in Individuals with Severe-to-Profound High-Frequency Hearing Loss. *Journal of the Association for Research in Otolaryngology* **12**, 519–534 (2011).
36. Molnár, J., Szakács, G. & Tusnády, G. E. Characterization of Disease-Associated Mutations in Human Transmembrane Proteins. *PLoS One* **11**, e0151760- (2016).
37. Zuo, J., Zhan, D., Xia, J. & Li, H. Single-Molecule Force Spectroscopy Studies of Missense Titin Mutations That Are Likely Causing Cardiomyopathy. *Langmuir* **37**, 12128–12137 (2021).

38. Srinivasan, S., Hazra, J. P., Singaraju, G. S., Deb, D. & Rakshit, S. ESCORTing proteins directly from whole cell-lysate for single-molecule studies. *Anal Biochem* **535**, 35–42 (2017).
39. Bouchiat, C. *et al.* Estimating the Persistence Length of a Worm-Like Chain Molecule from Force-Extension Measurements. *Biophys J* **76**, 409–413 (1999).
40. Loeff, L., Kerssemakers, J. W. J., Joo, C. & Dekker, C. AutoStepfinder: A fast and automated step detection method for single-molecule analysis. *Patterns* **2**, 100256-100269 (2021).
41. Bell, G. I. Models for the Specific Adhesion of Cells to Cells. *Science (1979)* **200**, 618–627 (1978).
42. Mártonfalvi, Z. *et al.* Low-force transitions in single titin molecules reflect a memory of contractile history. *J Cell Sci* **127**, 858–870 (2014).
43. Kellermayer, M. S. Z., Smith, S. B., Bustamante, C. & Granzier, H. L. Mechanical Fatigue in Repetitively Stretched Single Molecules of Titin. *Biophys J* **80**, 852–863 (2001).
44. DuPai, C. D., Davies, B. W. & Wilke, C. O. A systematic analysis of the beta hairpin motif in the Protein Data Bank. *Protein Science* **30**, 613–623 (2021).
45. Marcelino, A. M. C. & Gierasch, L. M. Roles of β-turns in protein folding: From peptide models to protein engineering. *Biopolymers* **89**, 380–391 (2008).
46. Lahiri, P., Verma, H., Ravikumar, A. & Chatterjee, J. Protein stabilization by tuning the steric restraint at the reverse turn. *Chem Sci* **9**, 4600–4609 (2018).
47. Reiersen, H. & Rees, A. R. The hunchback and its neighbours: proline as an environmental modulator. *Trends Biochem Sci* **26**, 679–684 (2001).
48. Basu, C., Ganguly, H. K. & Basu, G. Accommodation of Gly-cisPro in a designed β-hairpin. *Chemical Communications* **57**, 7172–7175 (2021).
49. Alderson, T. R., Lee, J. H., Charlier, C., Ying, J. & Bax, A. Propensity for cis-Proline Formation in Unfolded Proteins. *ChemBioChem* **19**, 37–42 (2018).
50. Bhatia, S. & Udgaonkar, J. B. Heterogeneity in Protein Folding and Unfolding Reactions. *Chem Rev* **122**, 8911–8935 (2022).
51. Chan, H. S., Shimizu, S. & Kaya, H. Cooperativity Principles in Protein Folding. in *Methods in Enzymology* **380**, 350–379 (2004).
52. Dill, K. A. & Chan, H. S. From Levinthal to pathways to funnels. *Nat Struct Biol* **4**, 10–19 (1997).
53. Ribeiro, J. V *et al.* QwikMD — Integrative Molecular Dynamics Toolkit for Novices and Experts. *Sci Rep* **6**, 26536 (2016).
54. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J Mol Graph* **14**, 33–38 (1996).
55. Nelson, M. T. *et al.* NAMD: a Parallel, Object-Oriented Molecular Dynamics Program. *The International Journal of Supercomputer Applications and High Performance Computing* **10**, 251–268 (1996).
56. Best, R. B. *et al.* Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ1 and χ2 Dihedral Angles. *J Chem Theory Comput* **8**, 3257–3273 (2012).

57. Miao, Y., Feher, V. A. & McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J Chem Theory Comput* **11**, 3584–3595 (2015).
58. Pang, Y. T., Miao, Y., Wang, Y. & McCammon, J. A. Gaussian Accelerated Molecular Dynamics in NAMD. *J Chem Theory Comput* **13**, 9–19 (2017).
59. Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *J Chem Phys* **101**, 4177–4189 (1994).
60. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J Chem Phys* **103**, 4613–4621 (1995).
61. Koukos, P. I. & Glykos, N. M. Grcarma: A fully automated task-oriented interface for the analysis of molecular dynamics trajectories. *J Comput Chem* **34**, 2310–2312 (2013).
62. Hazra, J. *et al.* Broken force dispersal network in tip-links by the mutations induces hearing-loss. *Biochem J* **0**, 614610 (2019).
63. Srinivasan, S., Hazra, J. P., Singaraju, G. S., Deb, D. & Rakshit, S. ESCORTing proteins directly from whole cell-lysate for single-molecule studies. *Anal Biochem* **535**, 35–42 (2017).
64. Garg, S., Singaraju, G. S., Yenghkom, S. & Rakshit, S. Europe PMC Funders Group Tailored-polyproteins using sequential staple and cut. *Bioconj Chem* **29**, 1714–1719 (2018).

Figures

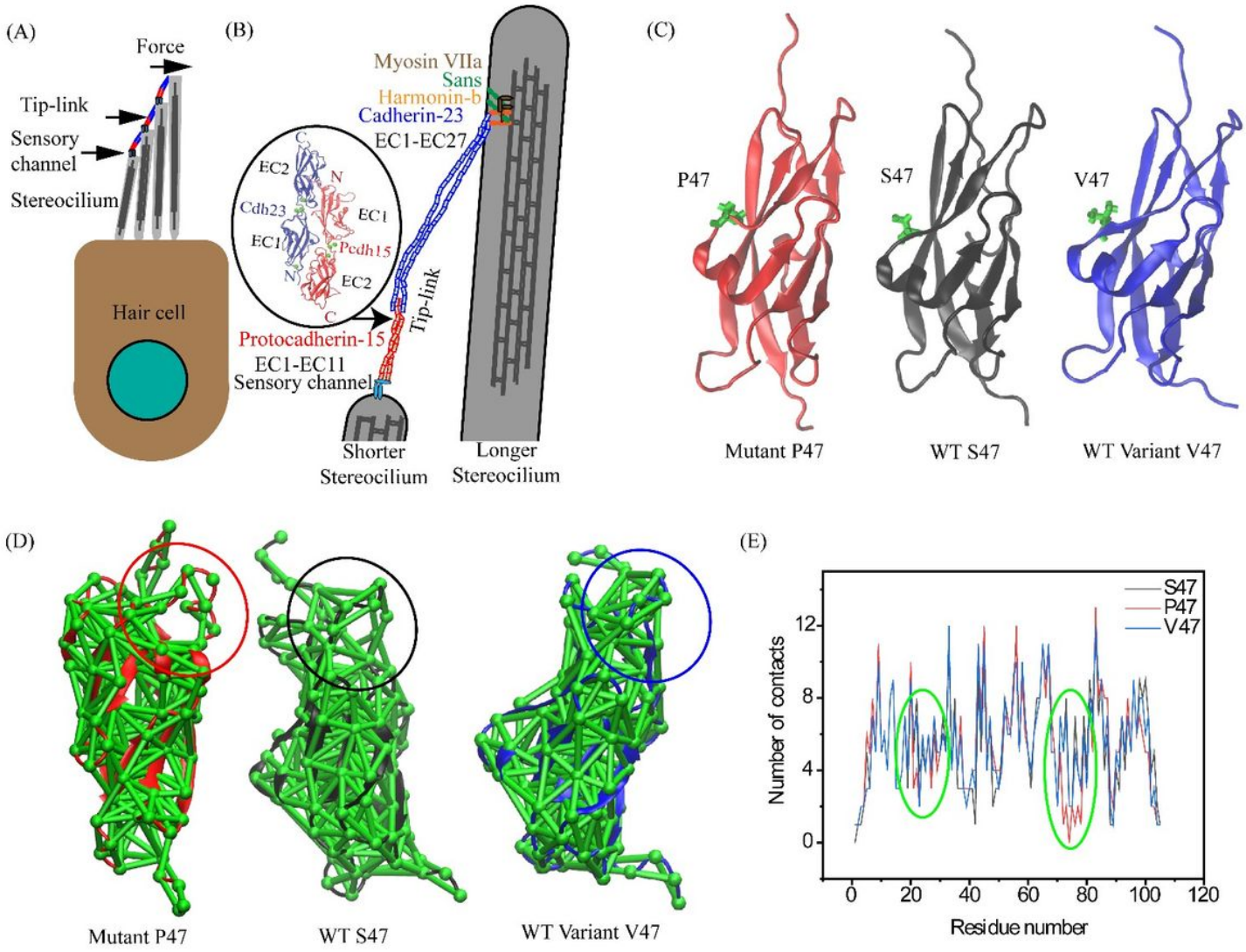

Figure 1

Schematic depiction of all three protein variants showing differential packing density. A) Schematic depiction of a hair-cell with stereocilia in staircase pattern. B) Representation of tip-link structure between two nearest-neighbouring stereocilium, where Cadherin-23 (in blue) and Protocadherin-15 (in red) forms a trans-heterotetrametric complex. The interacting domain of Cdh-23 EC1-EC2 and Pcdh-15 EC1-EC2 is zoomed in and highlighted in ribbon structures. C) Ribbon representation of three variants of Cdh23 EC1, S47 (in black), P47 (in red), and V47 (in blue). Structures of P47 and V47 variants are modelled using alphafold 2.0. D) Contact maps from the 300 ns MD simulations are shown for all three variants, P47 (red), S47 (black), and V47 (blue). Green balls represent the node, and solid lines represent the edges. Reduction in the packing density is highlighted with circles. E) Number of contacts for all the residues of Cdh23 EC1 S47 (Black), P47 (Red), and V47 (Blue) are plotted. Differences in number of contacts among P47, S47, and V47 variants in residues near 22nd and 76th are highlighted in green circles.

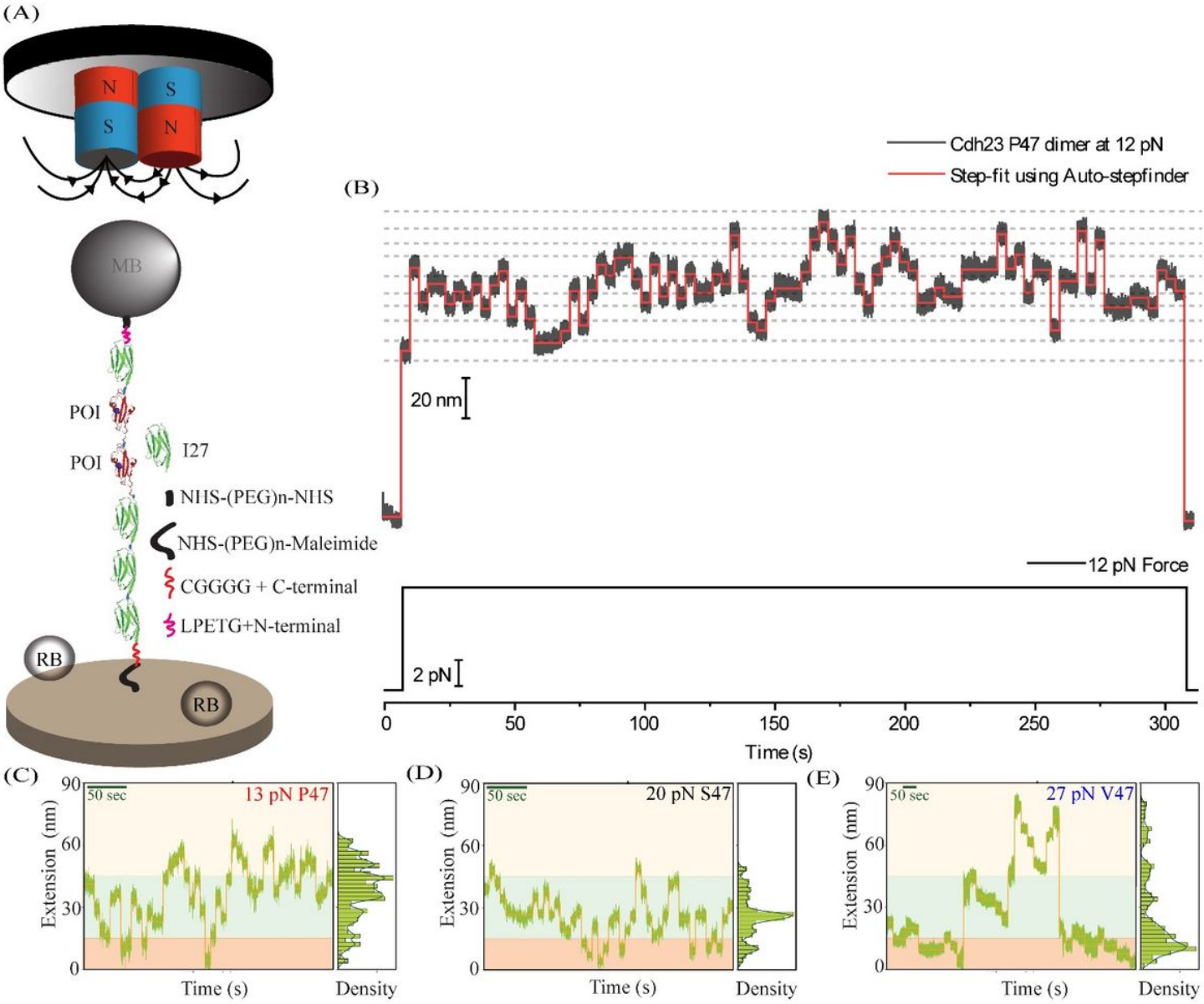

Figure 2

Schematic depiction of the continuous force-clamp experiment using a Magnetic Tweezer. A) Schematics of a typical magnetic tweezer with a chimeric polyprotein. The chimeric polyprotein, the protein of interest (POI in red) sandwiched between I27 (green), is covalently attached to the coverslips and M-270 super-paramagnetic beads following sortagging chemistry. Reference beads (RB) are non-magnetic and physiosorbed to surface. B) A representative extension of Cdh23 (EC1 P47)$_2$ at a clamping force of 12 pN. The dark grey line represents the data (collected at 580 Hz) and the red line maps the steps obtained from Autostepfinder. A number of conformations are seen between the closed and open states of the protein. (C-E) The time-trace of the protein-extension at constant clamping forces are shown for (C) P47 at 13 pN, (D) S47 at 20 pN, and (E) V47 at 27 pN. Orange solid lines guide the steps resulted from the Autostepfinder. Each force-curve is manually segregated in three-regions: peach box represents closed/native state (15 nm from the lowest step value), sky box marks the partially unfolded states (15-45 nm extension from the lowest step value), and the yellow box segregates the open/fully unfolded

states (45 nm and above from the lowest step value). The rightmost side of each force-extension curve shows the distributions of the corresponding protein-extensions using bars and black solid lines are the kernel density estimates.

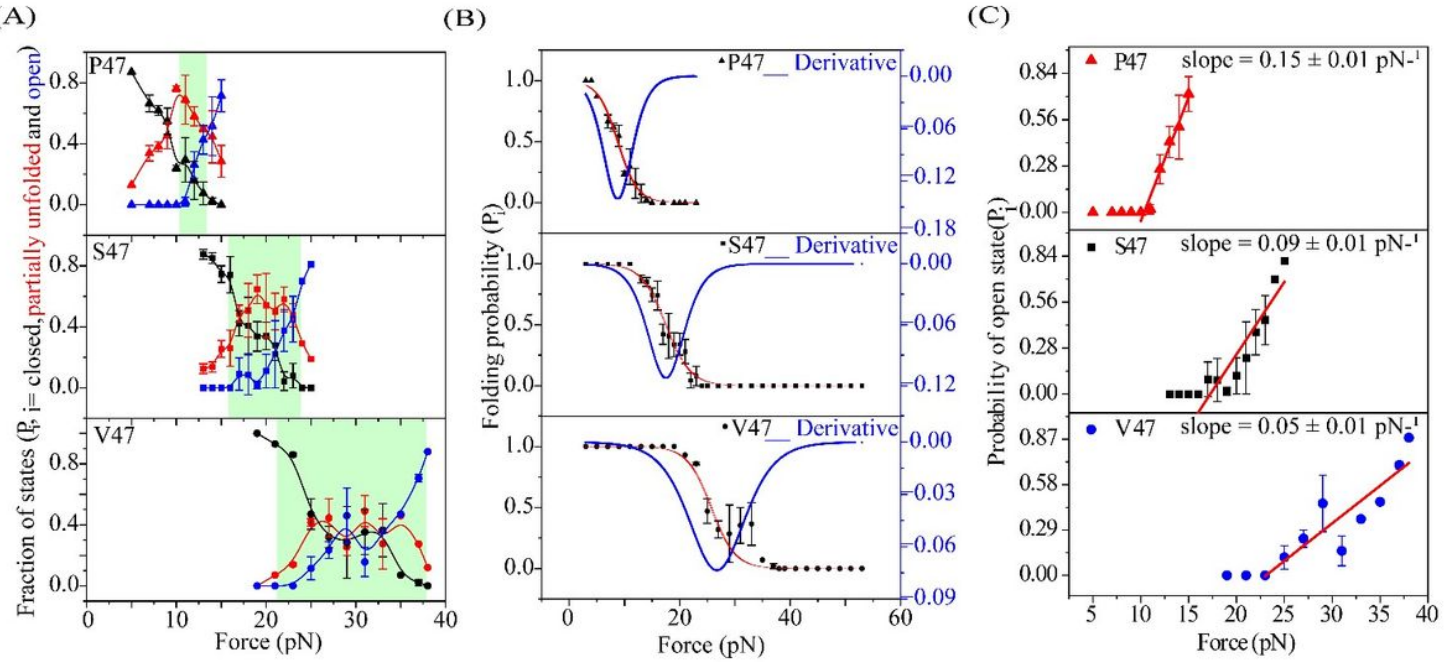

Figure 3

Estimating heterogeneity from the protein folding probability and cooperativity. A) Variations in the fraction of three states (native or closed in black, partially unfolded in red, fully unfolded, or open in blue) with clamping forces, estimated from the dwell-time analysis of force-clamp experiments, are shown for P47 (top), S47 (middle), and V47 (bottom). The green boxes qualitatively highlight the region where all three states co-exist. B) Probability of the closed state with clamping forces are shown for all three variants. To estimate the width of the transitions, we approximated the folded to unfolded transitions as two-state model and fit to Boltzmann equation (red dots) (see Methods). We next plotted the derivative of the fit (blue solids) to obtain the critical equilibrium force from the peak-position and the extent of force-tolerance from the FWHM. C) Probability of the open state with clamping forces are shown for all three variants from top to bottom. We fit (solid red line) the transition regime to linear equation and estimate the unfolding cooperativity from the slope.

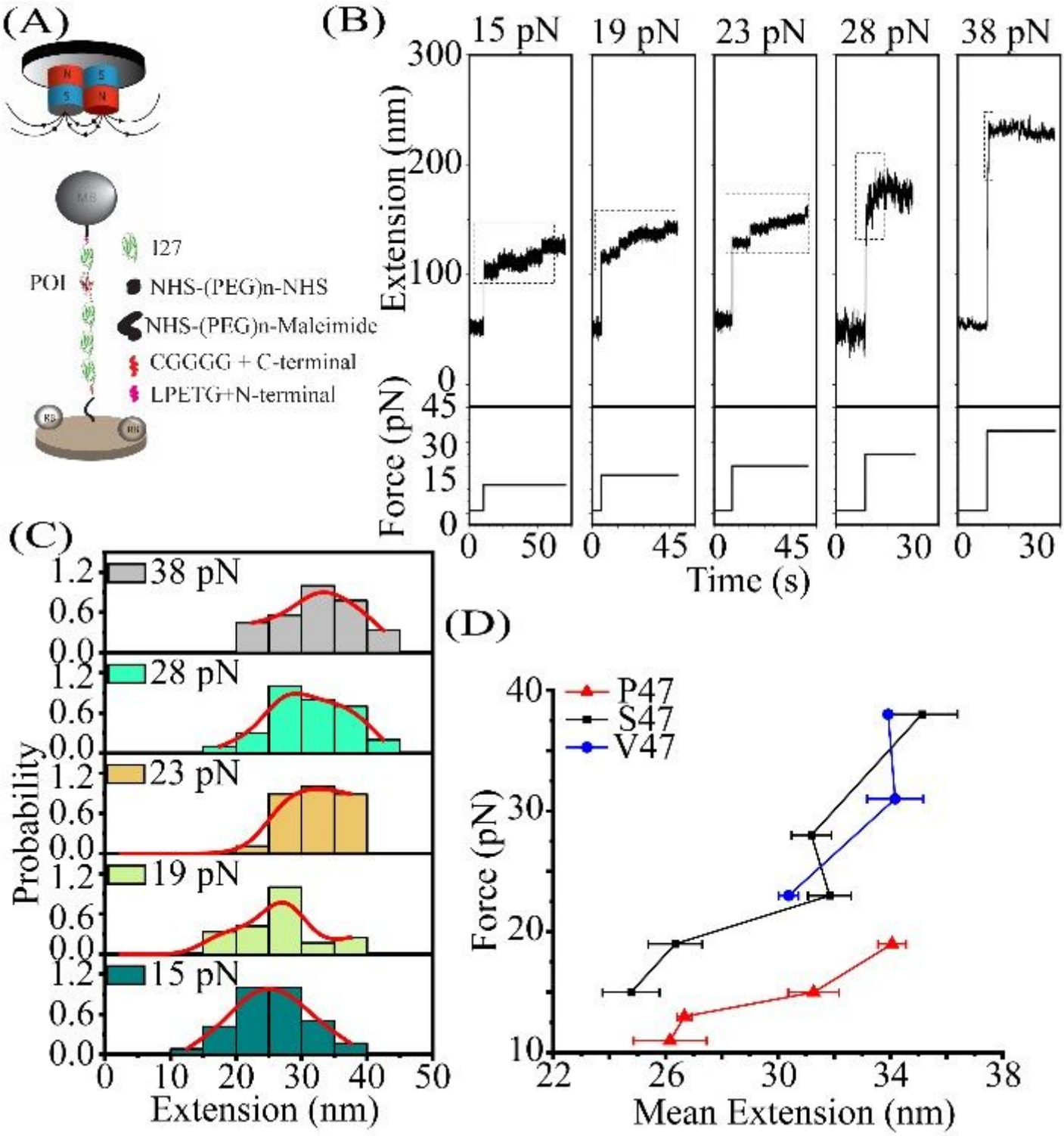

Figure 4

Extension (DL) of protein variants depends on clamping force. A) Schematics of the magnetic tweezer force-clamp setup with a single unit of Cdh23 (red) sandwiched between I27 domains (green). B) From left to right, representative force-extension data of Cdh23 EC1 S47 with increasing clamping forces (from 15 pN - 38 pN). We noted a gradual increase in the extension of protein with force, as expected from polymer chain theory. Further, the increase in the initial entropic jump with force is signatory to MT based force-clamp measurements. The small variations in the thermal noise appears from different trapped

beads. C) Distributions of extensions of Cdh23 EC1 S47 at varying clamping forces are shown. The red solid lines are the Kernel density estimators (KDE). Bandwidths of the kernels are estimated using Freedman-Diaconis rule ($h = 2\,IQR(x)/n^{1/3}$, where h is the bin width, n is the number of datapoints). D) The most-probable extension obtained from the distributions are plotted at different clamping forces for all three variants.

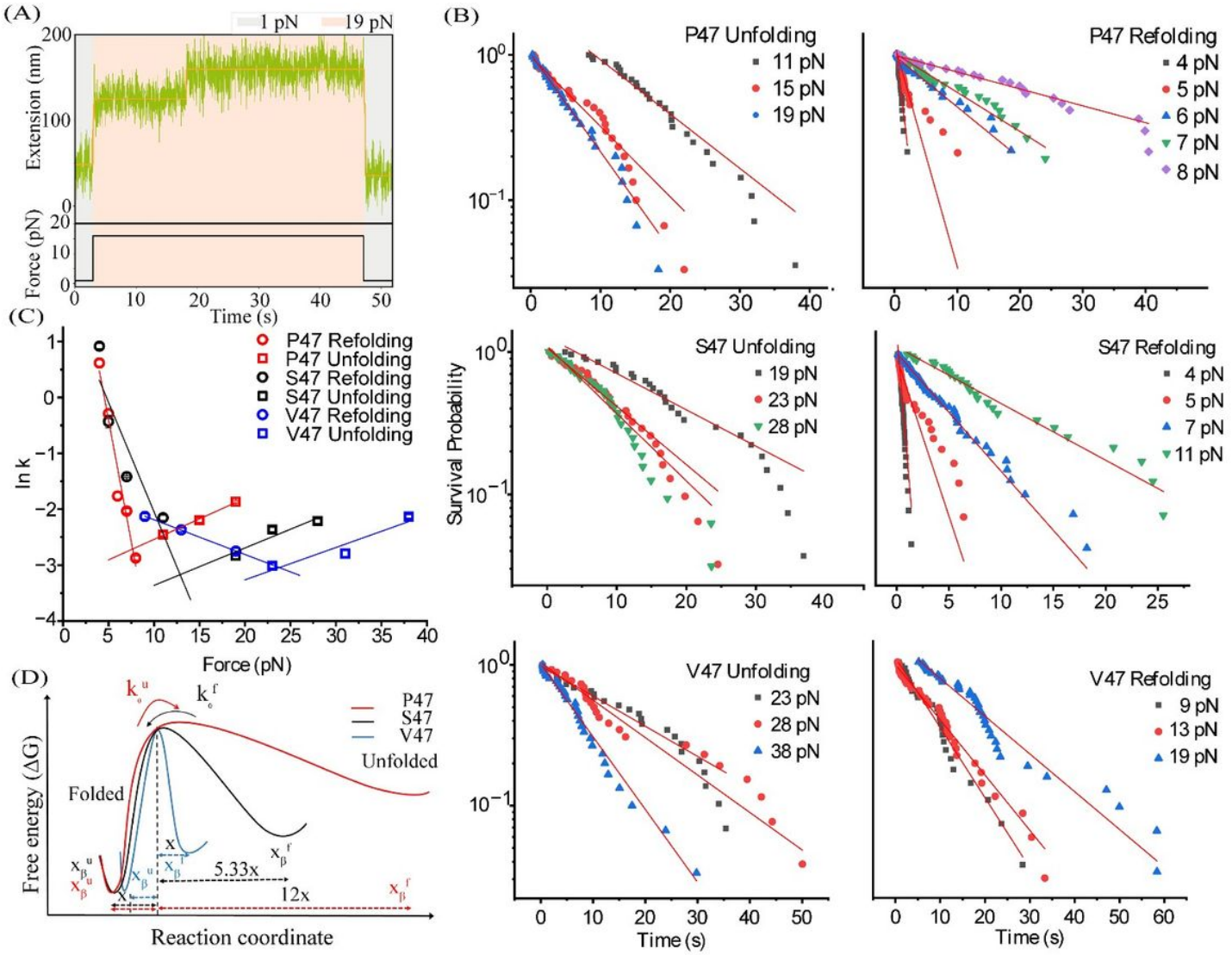

Figure 5

Folding dynamics of protein-variants under tensile force. A) A representative time trace curve of extension of Cdh23 S47 monomer under a clamping force of 19 pN. Dwell-time and step-height are collected from the step-fitting using autostepfinder. B) Normalized survival probability of unfolding and refolding are shown here for all three variants, Cdh23 P47 (top), S47 (centre), and V47 (bottom). Red solid lines are the fit to single-exponential decay function. C) Force-induced unfolding rates (dot) and refolding rates (open box) for S47 (Black), V47 (Blue) and mutant P47 (Red) are shown in the Chevron plot. The corresponding solid lines are the fit to the Bell's equation. D) Representative 1D potential-energy profiles for all three variants, relatively scaled according to the kinetic parameters obtained from the Chevron plot.

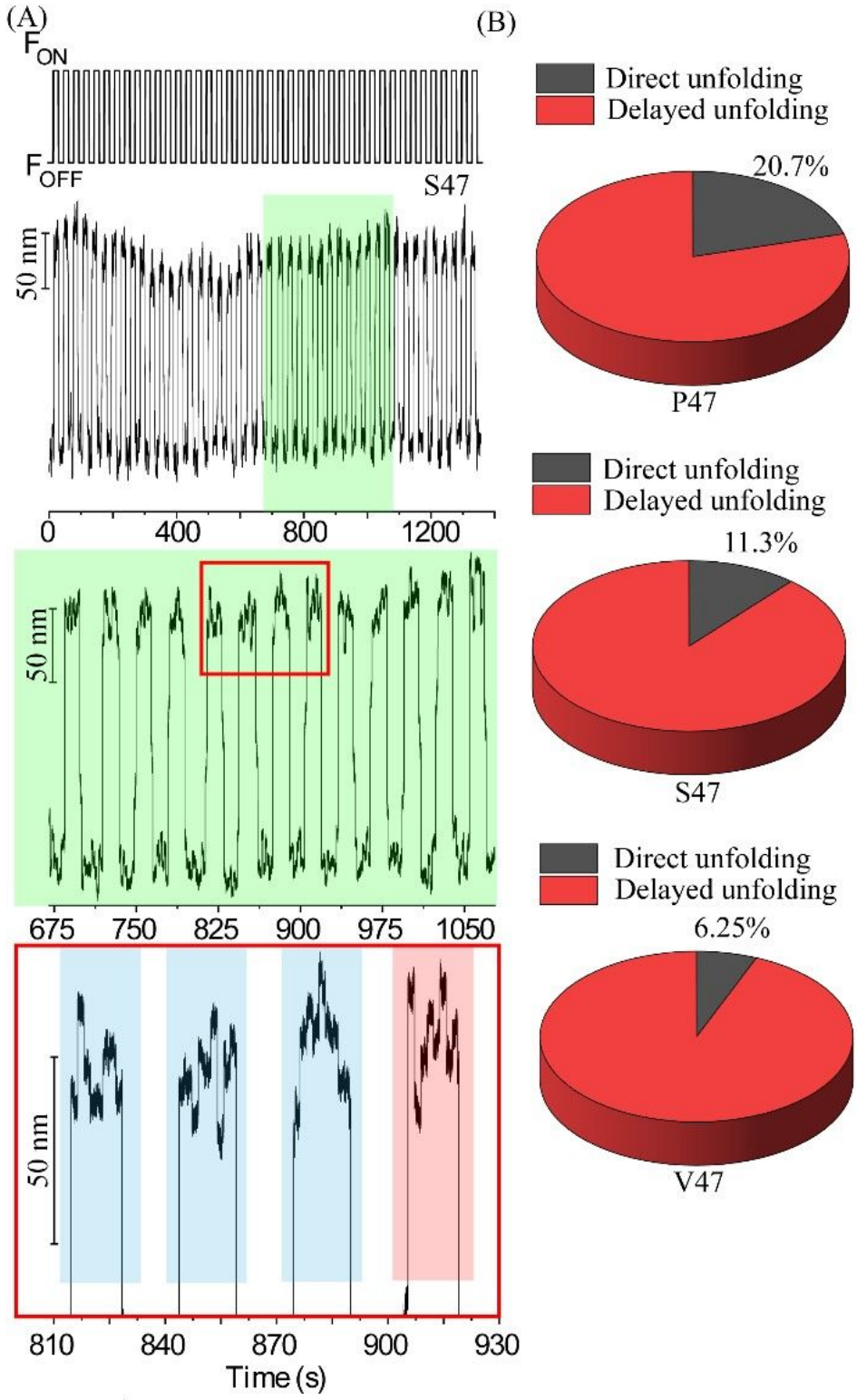

Figure 6

Repetitive force-pulse measurements show higher direct unfolding probability for the mutant variant (P47) than WT variants. A) The representative extensions of Cdh23 EC1 S47 from repetitive force-pulse measurements. The green shade and the red outlined box have been zoomed-in in two consecutive figures below, respectively. Bottom panel of figure (A) is shaded into three colours; red shade marks the transition of proteins to direct open state along with the initial elongation, and blue shades indicate the

delayed or indirect unfolding of proteins B) 3D pie representation to show the percentage of direct unfolding (black) and delayed unfolding (red) for all three variants, S47 WT, P47 mutant, and V47 WT variant. To get statistically significant number, we repeated this cycle for at least 30 times (For S47 protein N=44, P47 protein N=45, and V47 protein N=34).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- communicationchemistrysupplementary.docx