Estimating the Directional Flexibility of Proteins from Equilibrium Thermal Fluctuations

*Sanjoy Paul and Ravindra Venkatramani**

Department of Chemical Sciences, Tata Institute of Fundamental Research, Mumbai-400 005, India

Keywords: Variance, Energy Landscape, Protein Dynamics, Protein Stiffness, Directional Spring Constants, Molecular Dynamics Simulations, Elastic Network Models, Ubiquitin Family Proteins, Green Fluorescent Protein, GB1 Protein,

Corresponding Author

Ravindra Venkatramani (email: ravi.venkatramani@tifr.res.in)

Department of Chemical Sciences, Tata Institute of Fundamental Research, Dr. Homi Bhabha Road, Colaba, Mumbai 400005, Maharashtra, India

Abstract:

The directional spring constant of proteins is an equilibrium molecular property accessible to both experiment and computations. Single molecule force spectroscopy (SMFS) experiments can extract this metric to describe the mechanical anisotropy and directional flexibility of proteins. On the other hand, computational methods thus far have employed either indirect force based non-equilibrium simulations or coarse-grained Elastic Network Models (ENM) to predict protein directional spring constants. Here, we examine the ability of equilibrium atomistic Molecular Dynamics (MD) simulations to estimate the directional flexibility and mechanical anisotropy of proteins. We cross-correlate MD-derived directional spring constants with SMFS data for 5 different globular proteins. The sequence specificity of MD protein dynamics translates into computed spring constants which can distinguish the directional flexibility of ubiquitin (Ub) from structurally homologous Small Ubiquitin like Modifier (SUMO) isoforms. Further, average MD derived directional spring constants correlate with spring constants and unfolding forces obtained from SMFS experiments along 5 different directions of Green Fluorescence Protein (GFP) and 6 directions of immunoglobulin-binding B1 domain of streptococcal protein G (GB1). We predict two distinct classes of protein spring constants along the 7 lysine – C-term directions in Ub which are relevant for forming linkages with substrate proteins in the cellular context. Further, our studies reveal that the mechanical anisotropy of Ub is modified in a context sensitive manner by the binding of partner proteins (UBCH5A and UEV) which attach and recognize these biomolecular tag proteins. Based on equilibrium MD benchmarks, we critically examine the ability of ENM based computational schemes to predict directional flexibility of proteins and suggest modifications to improve these intuitive and scalable descriptions.

1. INTRODUCTION

The multidimensional potential energy surface (PES) of proteins exhibits a complex topology with a large number of equilibrium states and barriers connecting them. Single Molecule Force Spectroscopy (SMFS) provides an experimental avenue to probe the features of the PES utilizing the response of proteins under in vitro mechanical stress[1–4]. These techniques provide not only signatures of folding-unfolding transitions of a protein at a single molecule level but also access parameters of the underlying PES. For instance, constant velocity SMFS experiments yield directional spring constants which can quantify the stiffness of a protein along a bond vector (pulling direction) connecting the C_α atoms of a residue pair subjected to a pulling force[5–8]. Such directional spring constants essentially capture the curvature of the underlying PES around a native (equilibrium) protein state along the pulling reaction coordinate. For instance, SMFS experiments were able to distinguish the stiffness of structurally homologous ubiquitin family proteins (Fig-1A) along the C-terminus to N-terminus (N-C) pulling direction in terms of their spring constants[5]**.** Another important property accessed by SMFS experiments is protein mechanical anisotropy. Since proteins are large bio-molecules with complex contact topologies, it is expected that they will respond anisotropically to the uniaxial mechanical stress induced in SMFS experiments. Despite challenges in the experimental implementation[1], mechanical anisotropy has been explicitly demonstrated on several proteins using SMFS[7,9–11]. It has been shown, for instance, that Green Fluorescent Protein (GFP) when mechanically pulled along 5 different pulling directions exhibits spring constants in the range of 1-17 N/m (Fig-1B).[7] Mechanical unfolding forces in the range of 40-180 pN were observed along 6 pulling directions for the immunoglobulin-binding B1 domain of streptococcal protein G (GB1) which were attributed to differences in unfolding pathways along these directions (Fig-1C).[10]

The SMFS techniques, however, lack structural resolution and rely on computational methods to map the experimental stiffness measures to the underlying molecular structure. Further, computations can provide protein spring constants along directions which are not practically accessible to SMFS experiments. Such complementary computational techniques have included both non-equilibrium force-based and equilibrium fluctuation based approaches to examine protein directional stiffness.[12–17] In the former approach, mechanical forces are explicitly applied in dynamical simulations and the protein response is monitored to estimate the directional

flexibility. For instance, a Young's modulus like metric was used to estimate the stiffness of crystalline β-sheet rich peptide cross-links in silk fibres from atomistic constant velocity mechanical pulling simulations of the assembly.[14] In another study, mechanical forces were applied on residue pairs in coarse-grained Brownian dynamics simulations of GFP and directional spring constants were estimated using Hooke's law.[15] A drawback of these approaches is that the estimated stiffness is sensitive to the applied external forces or pulling speeds and is not a true equilibrium property. In the case of the latter approach, equilibrium thermal fluctuations derived from coarse-grained Elastic Network Model (ENM) normal modes were processed to generate directional spring constants in ubiquitin, GFP and other protein systems.[15,17] A limitation of standard implementations of ENM models is that they utilize simple harmonic inter-atomic interaction potentials which do not encode chemical sequence specificity, sidechain interactions and solvation effects. As a result, the standard coarse-grained ENM models[17] cannot resolve directional flexibilities of structurally homologous proteins wherein differential sidechain packing can give rise to distinct dynamics.[18] In their study of the mechanical anisotropy of GFP, Sacquin-Mora and Lavery employed a more detailed ENM model with 2-3 pseudodatoms per amino acid residue which accounts for the packing density of different amino acid sidechains.[15] Surprisingly, the authors found that directional ENM spring constants based on GFP crystal structures correlated poorly with those extracted from Brownian Dynamics simulations of GFP with explicit directional forces applied on the residues. Further, the best correlations with experimental spring constants which were obtained from explicit force simulations are not significantly better than those extracted by standard ENM models[17]. Thus, a deeper investigation into the estimation of directional spring constants from high resolution descriptions of equilibrium protein dynamics is required to examine the limitations of standard ENM models.

Interestingly, the ability of equilibrium atomistic molecular dynamics (MD) simulations to extract protein directional spring constants has not been explored. At first glance this appears surprising, as atomistic MD simulations provide one of the most realistic descriptions of thermal motions for solvated proteins available to date. Given an equilibrium distribution of bond vector lengths from protein MD trajectories, the directional spring constants along the vectors can be readily extracted using the classical equipartition theorem. However, assessing equilibrium in MD simulations has proved challenging due to the inherent complexity of the protein PES and the associated computational expense for sampling such a landscape.[19–22] We have recently made progress on

this front, by introducing the cumulative variance of coordinate fluctuations (CVCF) as an intuitive metric to assess both sampling and protein dynamics in MD simulations.[18] Features in the CVCF

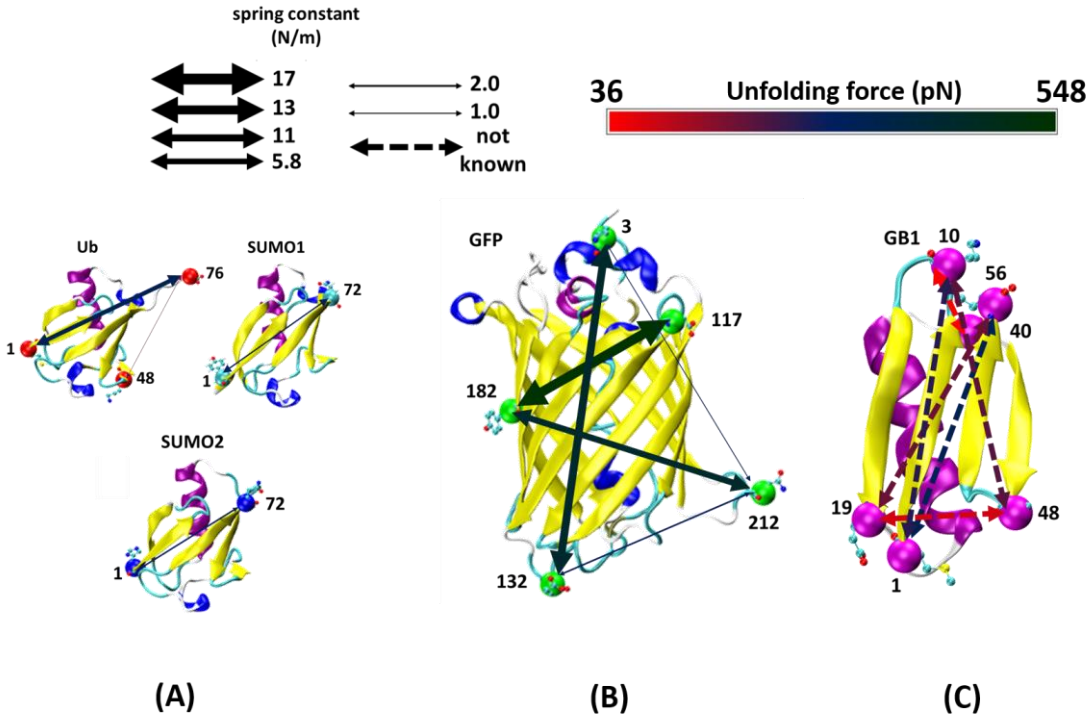

Figure 1: Backbone secondary structure and mechanical anisotropy in (A) Ubiquitin family protein (Ub, SUMO1 and SUMO2), (B) GFP and (C) GB1. SMFS derived mechanical unfolding forces and directional spring constants reported in the literature are shown using arrows drawn along bond vectors connecting C_α atoms of residue pairs on which forces were applied. The thickness of the arrows are proportional to the magnitude of directional spring constant values reported. The strength of mechanical unfolding force values are indicated using a color map (red-low to black-high) on the arrows.

trace when plotted as function of simulation time can provide an assessment of the curvature and ruggedness of the protein PES. Importantly, we have demonstrated that a CVCF based analysis can yield quasi-harmonic spring constants for subsets of protein atoms (for instance, backbone, sidechain, and interfacial atoms within protein complexes).[18] These developments coupled with the present abilities of MD to routinely access protein dynamics on the microsecond timescale provide a compelling motivation to examine the ability of MD to predict protein directional spring constants.

In this study, using MD simulations, we compute directional spring constants for the 5 proteins shown in Fig-1: ubiquitin (Ub), Small ubiquitin-like modifier proteins (SUMO1 and SUMO2),

GFP, and GB1. As discussed earlier, experimental spring constants and unfolding forces along the N-C pulling direction have been reported for these proteins [5,7,10]. Further, SMFS spring constants along multiple pulling directions are available for Ubq (2 directions)[9] and GFP (5 directions)[7]. For the GB1 system, SMFS derived PES parameters and unfolding force data has been reported for 6 directions.[10] We first carry out a CVCF-trace analysis of the protein backbone in sets of independent microsecond (μs) MD trajectories to assess both equilibration and the relative flexibility of structured and unstructured protein segments. The latter analysis enables us to discard flexible segments and map SMFS pulling directions onto structured segments which are expected to contribute to the mechanical resistance of the protein. We then proceed to show that the average directional spring constants computed from MD offer good correlations with the experimental SMFS data at much higher resolutions relative to coarse-grained ENM. We present experimentally testable predictions for the mechanical anisotropy of Ub along biologically relevant pulling directions and the expected changes in Ub anisotropy upon binding to two natural partner proteins (UBCH5A and UEV). Finally, using an extensive test set of directional bond vector fluctuations derived from MD trajectories of Ub, we propose improvements to an intuitive ENM framework proposed previously by Eyal and Bahar [17] to compute directional spring constants.

2. Methods

2.1 MD Simulations:

Details of simulation conditions and parameters of ubiquitin family proteins (Ub, SUMO1 and SUMO2) and Ub:UBCH5A and Ub:UEV protein complexes were reported by us recently.[18] We used the 10 x 1μs trajectories for these three protein systems from reference 18 directly in our present analysis. Here, we carried out additional all atom MD simulations for GFP and GB1 using broadly the same protocol described in reference 18 with a few system specific modifications as described below. We carried out simulations using GROMACS[23] version 5.1.3 with the AMBER force field ff14SB[24] for GFP and the CHARMM36m force field[25] for GB1. In both cases, an explicit solvent TIP3P water model was employed.[25] Initial coordinates of GFP were derived from Protein Data Bank (PDB code 1GFL) and processed through the tleap program of AmberTools20[26] to add hydrogens. The protein was solvated using a TIP3P rectangular water box with a 10 Å padding around the protein along each of the X, Y and Z directions of the box. Further, 5 Na^+ ions were added to neutralize the system. The resultant structure and topology built in this way were

converted into GROMACS readable format using the ParmEd tool[26]. In case of GB1, initial coordinates were derived from the PDB (code 1PGA) and processed through the CHARMMGUI webserver[27,28] to add hydrogen atoms and solvate the system using a TIP3P rectangular water box. A solvent padding of 10 Å around the protein was created along each of the X, Y and Z direction with 4 Na^+ ions added to neutralize the system. Following these initial modelling steps, the systems were equilibrated and production runs were carried out employing Periodic Boundary Conditions (PBC) with the Particle Mesh Ewald (PME) method[29] for long-range electrostatic interactions. A 1.0 nm short-range van der Waals cut-off and a 1.0 nm short range electrostatic cut-off were used during the equilibration and production runs. During equilibration, we first optimized the solvated protein systems, while keeping protein heavy atoms fixed for 10000 steps using the steepest descent algorithm. Then, the systems were equilibrated at 300 K (thermal equilibration) with protein heavy atoms fixed for 100 ps using the Nose-Hoover[30,31] thermostat with a 0.5 ps relaxation constant. Optimization and thermal equilibration were repeated again with harmonic constraints of 25 kcal/(mol $Å^2$) on protein heavy atoms, followed by a 150 ps NPT equilibration at a temperature of 300 K and a pressure of 1 bar using the Parrinello-Rehman[32] isotropic pressure coupling scheme with a relaxation time constant of 2.0 ps. The NPT equilibration step was then repeated twice with harmonic constraints of 10 and 5 kcal/(mol $Å^2$) applied on protein heavy atoms successively. Finally, unconstrained NPT simulations were performed for 1 ns. The solvated protein coordinates and velocities obtained at the end of this step were used to seed 10 independent 1 μs NVT production runs for each protein system (GFP and GB1) with the aforementioned temperature coupling method. For all 5 protein systems, production trajectories sampled coordinates every 20 ps, which were analysed using Visual Molecular Dynamics (VMD) 1.9.1.[33]

2.2 Principal Component Analysis of MD Trajectories

Given rigid body rotation-translation free coordinates from an MD trajectory of length T_S for an N atom protein system, Principal Component Analysis (PCA)[34,35] yields a set of 3*N*-6 orthogonal principal components (PC eigenvectors) and their variances (the eigenvalues) by diagonalizing the 3*N*×3*N* variance-covariance matrix of atomic coordinates with elements $\sigma_{ij}^2(T_S) = \frac{\sum_{t=1}^{T_S}(r_i(t)-\langle r_i(t)\rangle)(r_j(t)-\langle r_j(t)\rangle)}{T_S}$. Here the brackets < > indicate an average over the MD trajectory and $r_{i/j}$ are a pair of atomic Cartesian coordinates. The set of PC eigenvectors $\boldsymbol{\xi} = \{\boldsymbol{\xi_1}, \boldsymbol{\xi_2} \ldots \boldsymbol{\xi_{3N-6}}\}$

are sorted in descending order of their corresponding eigenvalues (variances λ_i) : $\lambda_1 > \lambda_2 > \ldots\ldots \lambda_{3N-6}$. We carried out PCA of protein C_α atom fluctuations sampled in 10 independent 1 μs trajectories of free Ub. Rigid-body translations and rotations were eliminated by aligning the coordinates of the protein C_α atoms in each frame of the Ub trajectory to their positions in the first frame (t = 1 timepoint). PCs obtained using this procedure were processed further as described in Sec. 2.4.3 and Sec. 2.4.4 to yield directional spring constants for Ub.

2.3 ENM Calculations

Coarse-grained normal modes were derived using C_α atom coordinates from crystal structures of the proteins: Ub (PDB id: 1UBQ), SUMO1 (PDB id: 4WJQ), SUMO2 (PDB id: 1WM3), GFP (PDB:id: 1GFL) and GB1 (PDB id: 1PGA) as described by Eyal and Bahar.[17] For an N atom system, the ENM analysis yields a set of 3N-6 orthogonal normal modes (eigenvectors) and their spring constants (eigenvalues) by diagonalizing the 3N×3N Hessian matrix (H) of second derivatives of the PES constructed at a minimum energy configuration of the system.[36] ENM assumes a harmonic inter-atomic potential energy form with a minimum at the experimentally (X-ray/NMR) derived protein coordinates:

$$\mathrm{V} = \frac{1}{2}\sum_{a,b} \gamma_{a,b}(r_{a,b} - r^0_{a,b})^2 \qquad (1)$$

where, $r_{a,b}$ represents the distance between C_α atom pairs (nodes) of protein amino acid residues a and b, $r^0_{a,b}$ is the corresponding equilibrium distance (distance in the X-ray/NMR protein structure) of the atom pair. The spring constant $\gamma_{a,b}$ between the atom pairs is defined as $\gamma_{a,b} = \gamma$ if $r^0_{a,b} \leq r_C$ and 0 otherwise. In the previous study of Eyal and Bahar, the value of γ was set to 0.25 kcal/(mol Å^2) for GFP and 1.75 kcal/(mol Å^2) for Ub [17]. However, for the present study, we find that a uniform of choice of $\gamma = 0.8$ kcal/(molÅ^2) for all the protein systems provides good agreement between MD and CG-ENM derived directional spring constants (Section S.1 in ESI). It must be noted that the value of γ serves as a scaling factor (Fig-S1 A-C in ESI) for the ENM derived spring constants and does not alter the relative trends across different pulling directions and protein systems (Fig-S1 D-F in ESI). We set r_C to 13 Å for all protein systems in accordance with Eyal and Bahar.[17] Diagonalization of the H, yields a set of normal mode eigenvectors $\xi = \{\xi_1, \xi_2 \ldots. \xi_{3N-6}\}$ which were sorted in ascending order of their corresponding eigenvalues (spring

constants λ_i) : $\lambda_1 < \lambda_2 < \ldots\ldots \lambda_{3N-6}$. The normal modes were then processed as described below (Sec. 2.4.3 and 2.4.4) to obtain directional spring constants for all 5 proteins.

2.4 Calculation of Protein Directional Spring Constants:

Both MD simulations and ENM provide a means to extract equilibrium fluctuations of protein atomic coordinates. Here, we aim to examine the correspondence between SMFS derived directional spring constants and unfolding forces with equilibrium protein dynamics. Below we first present the reaction coordinates (pulling directions) that we use for our analysis, followed by a description of methods to compute protein spring constants along these coordinates from MD/ENM equilibrium fluctuations.

2.4.1 Reaction coordinates to Examine Directional Protein Flexibility

SMFS experiments apply a pulling force along the protein backbone to unfold proteins.[9,10,37] In essence, the directional spring constants are reported along vectors connecting the C_α atoms of residues on which the pulling force is applied. Thus, by analogy, equilibrium fluctuations of the magnitude of bond vectors $\overrightarrow{r_{a,b}}$ connecting C_α atom pairs of amino acid residues *a* and *b* could be used to estimate spring constants along these directions[15,17]. However, this description is an oversimplification as it may include flexible protein segments which extend without resistance to a pulling force[5,7,9,10]. For instance, SMFS experiments report spring constants for Ub along $\overrightarrow{r_{1,76}}$ as the pulling force is applied to the backbone of this residue pair. However, the terminal segment 73-76 in Ub is a disordered loop (Fig-2) that should extend without resistance and transmit the pulling force to the closest structured segment. Thus the mechanical resistance of Ub would primarily arise from the core protein segment 1-72 and equilibrium fluctuations of the bond vector $\overrightarrow{r_{1,72}}$ should correspond better with the SMFS directional spring constant. The inclusion of flexible segments in the definition of the pulling reaction coordinate is a major problem in MD where a diversity of conformations would be sampled for the segments which would be extended without resistance in experiments. The problem is somewhat mitigated in ENM calculations as the equilibrium fluctuations are restricted to thermal amplitudes around the crystal structures. Nevertheless, the conformations and interactions of the terminal loops are still not representative of SMFS experiments and the computed flexibilities may deviate from that of the core protein structure. A CVCF-trace analysis of MD trajectories of Ub, GFP, and GB1 reveals that the per

atom variance of loops can more than an order of magnitude greater than that for the structured protein core (Fig-S2 of ESI). On the other hand, the per atom variance of the short loop formed by residues 48-49 in GB1 overlaps with that of the structured protein core. Based on these observations, we modified the experimental pulling directions to exclude the distinctly flexible segments (relative to the structured core) in MD (examples in Fig-2) for Ub, GFP, and GB1. For these modifications, we adopted a uniform protocol wherein we modified the experimental directions by shifting the reference C_α atoms from flexible loops (based on the CVCF analysis in Fig-S2 of ESI) to the closest residue in the structured protein core. Some examples of modified directions for Ub, GFP, and GB1 are shown in Fig-2 and the full list of experimental pulling directions and corresponding MD bond vectors used in our analysis are provided in Table-S3 of ESI. Protein spring constants along these directions were calculated by multiple methods as indicated in the following subsections.

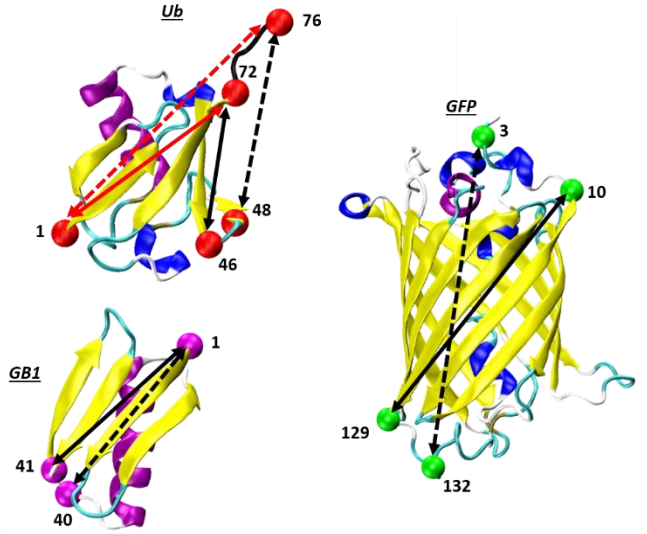

Figure 2: Bond vectors joining pairs of Cα atoms (red/purple/green spheres) representing SMFS pulling directions (dashed lines) and their corrections in MD simulations (solid lines) for Ub, GFP and GB1. For the corrections, the experimental bond vectors are redefined by shifting the terminal Cα atoms which lie in flexible segments to the closest residue which lies in the structured part of the protein.

2.4.2 Direct Calculation of Protein Directional Spring Constants from MD Trajectories:

MD simulation trajectories provide the evolution of atomic coordinates as a function of time. If $x_i(t)$ is the coordinate of the i^{th} atom at snapshot t along the MD trajectory then the Cumulative Variance of Coordinate Fluctuations (CVCF) over T_s snapshot/frames is defined as[18]

$$\sigma^2_{CVCF}(T_S) = \sum_{i=1}^{N} \frac{\sum_{t=1}^{T_S}(x_i(t)-\langle x_i(t)\rangle)^2}{T_S} = \sum_{i=1}^{N} \frac{1}{T_s}\sum_{t=1}^{T_s}\left(x_i(t) - \frac{1}{T_s}\sum_{t'=1}^{T_s} x_i(t')\right)^2 \qquad (2)$$

Where the index i runs over a subset of N protein atoms of interest (for instance, subset of all heavy, backbone and sidechain atoms, or interfacial and non-interfacial atoms in protein

complexes). While applying Eqn. 2 for a given subset of atoms, rigid-body translations and rotations should be eliminated by aligning the coordinates of the subset in each frame of the trajectory to their positions in the first frame (t = 0 timepoint). For individual MD trajectories, features of the CVCF-trace as a function of simulation time can help assess converged sampling (Boltzmann statistics) for a section of the PES. Further, a CVCF-trace analysis of multiple trajectories all aligned at their *t=0* timepoint can provide an estimation of the effective curvature and roughness of the protein PES in terms of the average CVCF ($<\sigma^2_{CVCF}>$) and its standard deviation (SD_{CVCF}) over the set of trajectories. In the present manuscript, we examine the CVCF of protein backbone atoms from MD trajectories as this is the relevant subset for extracting directional spring constants (*vide infra*). In the standard CVCF-trace analysis presented here, we examine $<\sigma^2_{CVCF}>$ and SD_{CVCF} from 10 independent μs MD simulation trajectories for each of 5 protein systems as a function of simulation time.

In order to assess the nature of the PES projected along SMFS pulling directions we follow the equilibrium fluctuations of bond vectors $\overrightarrow{r_{a,b}}$ connecting C_α atom pairs *a* and *b* as a function of simulation time T_S in terms of the directional CVCF:

$$\sigma^2_{CVCF}\left(T_S, \overrightarrow{r_{a,b}}\right) = \frac{\sum_{t=1}^{T_S}(r_{a,b}(t)-\langle r_{a,b}(t)\rangle)^2}{T_S} = \frac{1}{T_s}\sum_{t=1}^{T_s}\left(r_{a,b}(t) - \frac{1}{T_s}\sum_{t'=1}^{T_s} r_{a,b}(t')\right)^2 \qquad (3)$$

The indices *t* and t' run over the number of frames/snapshots sampled in the trajectory and $r_{a,b}(t)$ is the magnitude of $\overrightarrow{r_{a,b}}$ at time *t*. Since $\overrightarrow{r_{a,b}}$ represents an internal bond coordinate vector, no alignment of the MD trajectory snapshots to eliminate overall rotation and translation is required. The equipartition theorem then allows us to relate the thermal fluctuations along $\overrightarrow{r_{a,b}}$ in MD trajectories to a directional spring constant

$$k_{MD}\left(T_S, \overrightarrow{r_{a,b}}\right) = \frac{k_B T}{\sigma^2_{CVCF,direct}\left(T_S, \overrightarrow{r_{a,b}}\right)} \qquad (4)$$

Where *T* is the temperature (set to 300 K in this study) and k_B is the Boltzmann constant. In the directional spring constant analysis presented here, averages and standard deviations in $k_{MD}(\overrightarrow{r_{a,b}})$ were computed over 10 MD trajectories at the $T_S = 1\ \mu s$ timepoint based on the CVCF-trace analysis presented in Sec. 3.1.

2.4.3 Calculation of Directional Spring Constants from ENM and PCA

Eyal and Bahar proposed an intuitive coarse-grained ENM based scheme to extract directional spring constants from experimental protein structures.[17] This approach is appealing as it provides a rapid estimation of protein directional flexibility from X-ray/NMR derived protein structures. In this scheme, ENM normal modes are projected along bond vectors representing desired pulling directions and their contributions to the directional deformation are combined to estimate thermal fluctuations and directional spring constants. Projection of a normal mode *k* along a specific direction $\overrightarrow{r_{a,b}}$ for a pair of C_α atoms *a* and *b* can be computed as

$$cos\alpha_{ab}^{k} = \frac{\overrightarrow{r_{a,b}} \cdot \Delta\xi_k^{ab}}{|\overrightarrow{r_{a,b}}||\Delta\xi_k^{ab}|} \tag{5}$$

where, $\Delta\xi_k^{ab} = \xi_k^b - \xi_k^a$ represents the change in the bond distance between the C_α atom pairs *a* and *b* from its reference value due to the application of normal mode *k* with unit amplitude to the protein structure. Here, ξ_k^b and ξ_k^a are the coefficients of atoms *a* and *b* respectively in the k^{th} normal mode eigenvector. The contribution of the thermally activated k^{th} mode with eigenvalue λ_k to the overall deformation $d_{ab} = \sum_k d_{ab}^k$ is given by

$$d_{ab}^{k} = (k_B T/\lambda_k)^{1/2} cos\alpha_{ab}^{k} |\Delta\xi_k^{ab}| \tag{6}$$

and its contribution to the macroscopic force $F_{ab} = k_{direct}(\overrightarrow{r_{a,b}})d_{ab}$ that induces the deformation d_{ab} around the equilibrium state is

$$f_{ab}^{k} = \lambda_k d_{ab}^{k} \tag{7}$$

Force contributions from each normal mode can be summed up to balance the macroscopic force

$$F_{ab} = \sum_k f_{ab}^{k} \tag{8}$$

Which then leads to the following definition of spring constant from ENM modes as prescribed by Eyal and Bahar

$$k_{EB-ENM}(\overrightarrow{r_{a,b}}) = \frac{\sum_k \lambda_k d_{ab}^{k}}{\sum_k d_{ab}^{k}} \tag{9}$$

Although, this scheme was proposed for ENM modes, the spring constants can also be derived using quasiharmonic modes derived from PCA. In the manuscript, we will also apply Eqn. 9 to

derive spring constants from quasiharmonic modes ($k_{EB-PCA}(\overrightarrow{r_{a,b}})$). However, as we will show in the manuscript (Sec. 3.4), the definition of directional spring constant in Eqn. 9 does not reproduce MD thermal fluctuations and shows an unphysical dependence on high frequency modes of the system. We propose a modification to the original Eyal and Bahar formalism wherein the force balance condition in Eqn. 8 is replaced with an energy balance criterion. In this modified framework the potential energy contributions of each mode projected along $\overrightarrow{r_{a,b}}$ is summed up to obtain the total potential energy E_{ab} of the bond between atoms *a* and *b*:

$$E_{ab} == \frac{1}{2} k_{direct}(\overrightarrow{r_{a,b}}) d_{ab}^2 = \sum_k E_{ab}^k \tag{10}$$

Which then leads to a modified definition of the directional spring constant along $\overrightarrow{r_{a,b}}$ derived from ENM/PCA modes:

$$k_{ENM/PCA}(\overrightarrow{r_{a,b}}) = \frac{\sum_k \lambda_k (d_{ab}^k)^2}{\sum_k (d_{ab}^k)^2} \tag{11}$$

2.4.4 Covariance Propagation Scheme for Computing Directional Spring Constants from PCA and ENM

Here we describe a simple scheme to compute protein directional spring constants from a set of MD-PCs or ENM-normal modes based on the standard method of error propagation. For a function $f(x_1, x_2, x_3, \dots, x_n)$ of variables $x_1, x_2, x_3, \dots, x_n$ with variances $\sigma_{x_1}^2$, $\sigma_{x_2}^2$, $\sigma_{x_3}^2$, … $\sigma_{x_n}^2$ respectively and covariance $\sigma_{x_i y_{i'}}^2$ (i and i' run over 1 to n and $i \neq i'$) the variance in *f* can be generally written as:

$$\sigma_f^2 = \sum_{i=1}^n (\frac{\partial f}{\partial x_i})^2 \sigma_{x_i}^2 + \sum_{\substack{i=1 \\ i \neq i'}}^n \sum_{i'=1}^n \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_{i'}} \sigma_{x_i x_{i'}}^2 \tag{12}$$

The magnitude $r_{a,b}$ of the vector connecting C_α atom pairs is one such multivariate function of the constituent atomic coordinates of the vector:

$$r_{a,b} = \sqrt{\sum_{k=\{X,Y,Z\}} (k_b - k_a)^2} \tag{13}$$

Where the index $k_{a/b}$ runs over Cartesian coordinates *X*, *Y* and Z of the C_α atoms of residues *a* and *b*. In accordance with Eqn. 12, the variance of $r_{a,b}$ can be written by propagating the covariance matrix of the coordinates $k_{a/b}$:

$$\sigma^2_{CP}(\overrightarrow{r_{a,b}}) = \sum_{i,j=\{a,b\}} \sum_{k,k'=\{X,Y,Z\}} (\frac{\partial r_{a,b}}{\partial k_i})^2 \sigma^2_{k_i} + (\frac{\partial r_{a,b}}{\partial k'_j})^2 \sigma^2_{k'_j} + 2(\frac{\partial r_{a,b}}{\partial k_i})(\frac{\partial r_{a,b}}{\partial k'_j}) \sigma^2_{k_i k'_j} \quad (14)$$

Where the subscript CP stands for covariance propagation. After explicitly performing the differentiation with respect to $r_{a,b}$ we obtain:

$$\sigma^2_{CP}(\overrightarrow{r_{a,b}}) = \sum_{i,j=\{a,b\}} \sum_{k,k'=\{X,Y,Z\}} (k_i - k_j)^2 \left(\sigma^2_{k_i} + \sigma^2_{k'_j} - 2\sigma^2_{k_i k_j}\right) + 2(k_i - k_j)(k'_i - k'_j)(\sigma^2_{k_i k'_i} + \sigma^2_{k_j k'_j} - \sigma^2_{k_i k'_j} - \sigma^2_{k_j k'_i}) \quad (15)$$

where the indices *k* and *k'* run over the Cartesian coordinates, and indices *i* and *j* run over atoms *a* and *b*. In case of PC modes derived from an MD trajectory with T_S snapshots, k_i is the average k^{th} component of the coordinate of i^{th} atom from T_S snapshots. In case of ENM, k_i represents the k^{th} component of the coordinate of atom *i* in the crystal/NMR structure. Given an MD trajectory for *N*-atom protein system, the covariance matrix for these coordinates can be reconstructed from any subset of *M* PCs (M ≤ 3N-6) using the following transformation:

$$\left(\xi_1\ \xi_2 \ldots \xi_M\right)_{3N\times M} \begin{pmatrix} \lambda_1 & 0 & \ldots \\ 0 & \ddots & 0 \\ \ldots & 0 & \lambda_M \end{pmatrix}_{M\times M} \begin{pmatrix} \xi^{\mathrm{I}}_1 \\ \xi^{\mathrm{I}}_2 \\ \cdot \\ \cdot \\ \cdot \\ \xi^{\mathrm{I}}_M \end{pmatrix}_{M\times 3N} = (C)_{3\mathrm{N}\times 3\mathrm{N}} \quad (16)$$

$\boldsymbol{\xi} = \{\boldsymbol{\xi_1}, \boldsymbol{\xi_2} \ldots \boldsymbol{\xi_M}\}$ are the eigenvectors obtained from PCA on a MD trajectory of T_S snapshots and sorted based on their decreasing eigenvalues (variance) $\boldsymbol{\lambda_1} > \boldsymbol{\lambda_2} > \ldots\ldots \boldsymbol{\lambda_M}$. The reconstructed covariance matrix constructed from a subset of PCs can then be compared with that directly extracted from MD to reveal the dominant collective motions contributing to the protein directional flexibility. On the other hand, ENM yields normal mode eigenvectors and mode spring constants as corresponding eigenvalues. The spring constants can be converted into thermal mode variances using the equipartition expression in Eqn. 3. Hence, for an *N*-atom protein system a covariance matrix can be constructed from the top *M* (*M* ≤ 3*N*-6) normal modes (eigenvectors) and their corresponding spring constants (eigenvalues) as:

$$\left(\xi_1\ \xi_2\ \ldots\ \xi_M\right)_{3N\times M}\begin{pmatrix}\frac{k_BT}{\lambda_1} & 0 & \ldots\\ 0 & \ddots & 0\\ \ldots & 0 & \frac{k_BT}{\lambda_M}\end{pmatrix}_{M\times M}\begin{pmatrix}\xi_1^{\dagger}\\ \xi_2^{\dagger}\\ \cdot\\ \cdot\\ \cdot\\ \xi_M^{\dagger}\end{pmatrix}_{M\times 3N} = (C)_{3N\times 3N} \qquad (17)$$

$\boldsymbol{\xi} = \{\boldsymbol{\xi_1}, \boldsymbol{\xi_2} \ldots \boldsymbol{\xi_M}\}$ are the eigenvectors obtained from ENM on the crystal structure of the protein and sorted based on their increasing eigenvalues (decreasing variance of each normal mode) $\boldsymbol{\lambda_1} < \boldsymbol{\lambda_2} < \ldots\ldots \boldsymbol{\lambda_M}$. Elements of the covariance matrix C derived from a subset of *M* MD-PCA/ENM modes can be used in Eqn. 15 to obtain directional variances $\sigma^2_{CP-PCA}(\overrightarrow{r_{a,b}}, M)$ / $\sigma^2_{CP-ENM}(\overrightarrow{r_{a,b}}, M)$ and corresponding directional spring constants $k_{CP-PCA}(\overrightarrow{r_{a,b}}, M)$ / $k_{CP-ENM}(\overrightarrow{r_{a,b}}, M)$. The MD-PCA derived spring constants also depend on the length of the trajectory (T_S) and we assume that the PCA is carried out on an equilibrated trajectory as determined by the CVCF-trace analysis. On the other hand, the ENM derived spring constants depend on the cut-off radius r_C used for interatomic interactions (Sec. 2.3) and we set this parameter to an optimal value based on the studies by Bahar and co-workers[17,38]. In general, the isomorphism in the covariance propagation scheme (Fig-S4 of ESI) for computing spring constants using normal modes and PCs should be useful to quantitatively compare the fluctuations extracted from ENM and MD trajectories with varying cut-off radii and/or timescales respectively. As discussed in the results section (Sec. 3.4), such comparisons enable us to improve the description of ENM formalisms to extract directional spring constants.

3. Results

3.1 CVCF-trace analysis reveals locally equilibrated MD sampling over a 1μs timescale

Recently we presented an analytical framework to assess sampling in MD trajectories using CVCF.[18] Briefly, the features of the CVCF-trace when plotted as a function of MD simulation time reveals sections of the trajectory which are locally equilibrated (CVCF plateaus), discovering new minima (rise in CVCF), or converging towards a local equilibrium (decreasing CVCF-traces which eventually lead to plateaus). The criterion for assessment of equilibration is rigorous and plateaus of the CVCF essentially ensure Boltzmann population statistics for the section of PES sampled by

the MD trajectory. We also showed that by aligning CVCF-traces from multiple trajectories originating from the same phase point (same coordinates and velocities), the effective curvature and ruggedness of the PES around the originating phase point can be estimated in terms of the average CVCF ($<\sigma^2_{CVCF}>$) and associated standard deviation (SD_{CVCF}) across the set of trajectories respectively. In this section, we first apply this framework to examine equilibration of the protein backbone in our MD simulations and then the nature of the protein PES as projected along the pulling reaction coordinate $\overrightarrow{r_{a,b}}$ for Ub, SUMO1, SUMO2, GFP and GB1.

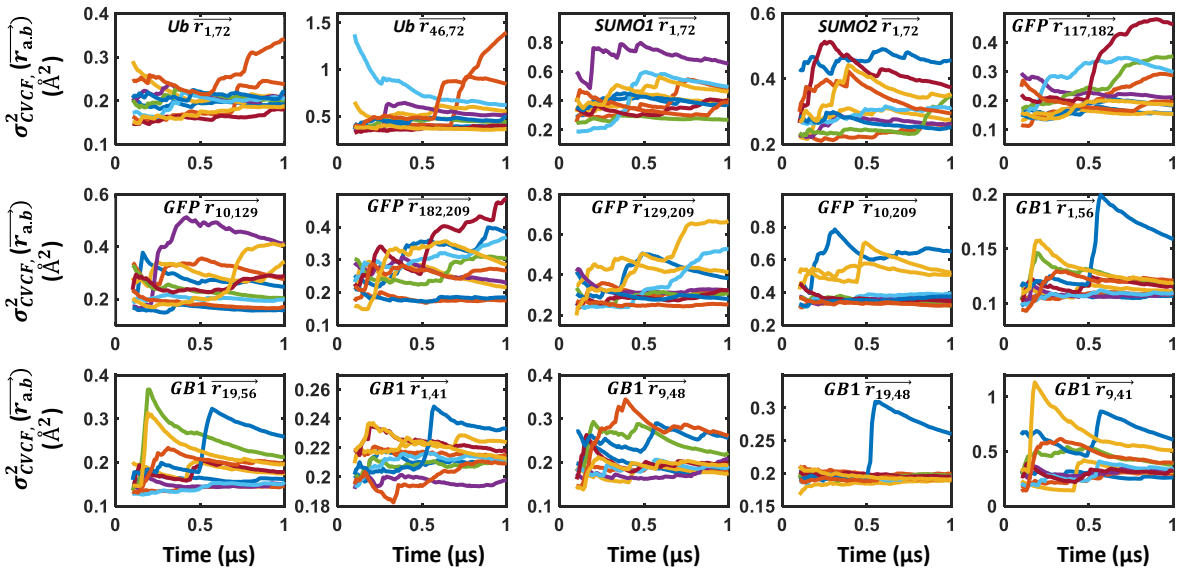

Figure 3: CVCF-trace analysis for bond vector fluctuations in 10 × 1μs MD trajectories of Ub, SUMO1, SUMO2, GB1, and GFP. The panels show directional CVCF traces corresponding to different pulling directions (Sec. 2.4.1 and Table-S3 of ESI) for which SMFS data is available.

To first examine equilibration, we aligned 10 independent μs trajectories of the 5 proteins at their initial (common) *t*=0 timepoint to plot the evolution of the backbone CVCF-traces of the proteins (Fig-S5 of ESI). While, the data reveal heterogeneity in traces, CVCF plateaus at 1μs are seen for a significant fraction of the trajectories for each of the five protein systems. Specifically, the fraction of trajectories showing convergence to plateaus varies from 0.6 (GFP) to 1.0 (SUMO1 and GB1). Thus, the sampling of the protein backbone has achieved near Boltzmann statistics in a significant fraction of trajectories for all 5 proteins. The statistics gets even better, when the directional CVCF-traces along the pulling reaction coordinates $\overrightarrow{r_{a,b}}$ (Column 4 in Table-S3 of ESI) are examined for the proteins (Fig-3). Here, the fraction of trajectories which show near Boltzmann

statistics at the 1 μs timepoint ranges from 0.8-1.0, indicating that these directions for the proteins are mostly equilibrated around their native state over this timescale. The trends in directional CVCF across multiple trajectories can be effectively combined in terms of a $<\sigma^2_{CVCF}(\overrightarrow{r_{a,b}})>$ trace plotted with associated SD_{CVCF} as a function of time (Fig-S6). Based on the average CVCF trace, PES features[18] and protein directional spring constants (Eqn. 4) can be extracted from the set of trajectories for all 5 proteins at the $T_s = 1$ μs timepoint.

In Table-S7, we report the $<\sigma^2_{CVCF}(\overrightarrow{r_{a,b}})>$ and SD_{CVCF} of the protein PES projected along the different pulling directions along with separation (ΔX_U) between native and transition states (potential width) for protein unfolding along those directions provided by SMFS experiments (Table-S7). The square root of $<\sigma^2_{CVCF}(\overrightarrow{r_{a,b}})>$ provides a measure of the PES range (MD equilibrium amplitude) that is sampled around the native minima in equilibrated simulation trajectories. Comparing ΔX_U with the MD equilibrium amplitude, we find that MD simulations sample a diverse fraction (~8-42 %) of the PES slice along the pulling coordinate lying between the native state and the transition state. In case of GFP, the stiffest direction ($\overrightarrow{r_{117,182}}$) is sampled the most (~42 %) and the most flexible direction ($\overrightarrow{r_{3,212}}$) is sampled the least (14 %). The lowest PES fractions sampled (~8-19 %) are along the different pulling directions for the GB1 protein. The structurally homologous ubiquitin family proteins show a relatively broad distribution of sampling fractions (~12-31%) wherein the PES fraction sampled along the stiffest direction (Ub: $\overrightarrow{r_{1,72}}$) is the most and that along the most flexible direction (SUMO1: $\overrightarrow{r_{1,72}}$) is the least. Within each protein fold (ubiquitin family, GFP, GB1), the MD thermal amplitudes and experimentally derived ΔX_U and spring constants correlate (Table-S7). With only a single exception (GB1: $\overrightarrow{r_{19,48}}$), we find a direct (inverse) relationship between MD thermal amplitudes in 1 μs trajectories and experimental ΔX_U (spring constants) for a given protein fold. While the fluctuations along $\overrightarrow{r_{19,48}}$ appear to be locally equilibrated (Fig-3) around the native state, the fraction of the PES between native and transition states sampled along this direction is the lowest (~8 %). Based on these observations, within a given protein fold, we expect systems which sample a greater fraction of the PES between the native and transition state to show stronger correlations with the SMFS experimental data. On the other hand, the data across different protein folds are not correlated and highlight difficulties in comparing mechanical properties across protein folds in both experiments and simulations. Nevertheless, relative trends in mechanical properties for a given protein fold can

be computationally determined and cross validated by experiments (see next section). Finally, in Table-S7 we also report the roughness of the PES slice projected along each pulling direction extracted from the CVCF-trace analysis of $10 \times 1\mu s$ MD trajectories for each protein system. The data show that the PES roughness along different directions is also diverse, ranging from ~1-26 % of MD equilibrium amplitudes, dictating the resolution of spring constants obtained from MD.

3.2 MD-derived average directional spring constants correlate with directional protein flexibility and unfolding forces from SMFS experiments for a given protein fold

In this subsection, we examine the ability of MD to estimate directional spring constants (k_{MD}) of Ub, SUMO1, SUMO2, GFP and GB1. Experimental spring constants k_{SMFS} and unfolding forces (F_{U-SMFS}) along the computationally examined directions have been reported (Fig-1 and Table-S7) and we compare these with k_{MD} extracted from the CVCF-trace analysis of $10 \times 1\ \mu s$ MD trajectories (Eqn. 4). Given the difficulties in comparing mechanical properties of proteins with different folds (see previous section), we examine the correlations between computed and experimental data for each protein fold separately. For ubiquitin family proteins (Fig-4A), the $k_{MD}(\overrightarrow{r_{1,72}})$ correlates well with that obtained from SMFS experiments[5]. Note that although SMFS experiments report directional spring constants and unfolding forces along $\overrightarrow{r_{1,76}}$ in Ub and $\overrightarrow{r_{1,72}}$ in case of SUMO1 and SUMO2, k_{MD} is computed along $\overrightarrow{r_{1,72}}$ for all the three protein systems. The unstructured protein tail formed by residues 73-76 in Ub (Fig-2) is expected to elongate upon pulling with negligible mechanical resistance. Such modifications to experimental distances in the MD analysis are discussed in the methods (Sec. 2.4.1) and in the remainder of this section we refer to the data only in terms of the modified MD reaction coordinates (column 4 in Table-S3 of ESI). Both computations and experiments predict Ub to be significantly stiffer than the SUMO isoforms along $\overrightarrow{r_{1,72}}$ and SUMO1 and SUMO2 to have comparable stiffness values. Mechanical unfolding forces have also been reported along $\overrightarrow{r_{46,72}}$ for Ub (pulling speeds of 0.3 μm/s) and along $\overrightarrow{r_{1,72}}$ for Ub, SUMO1 and SUMO2 (pulling speeds of 0.4 μm/s).[9] We find that the computed k_{MD} also correlates with the reported experimental unfolding forces (Fig-4D). The experimental unfolding

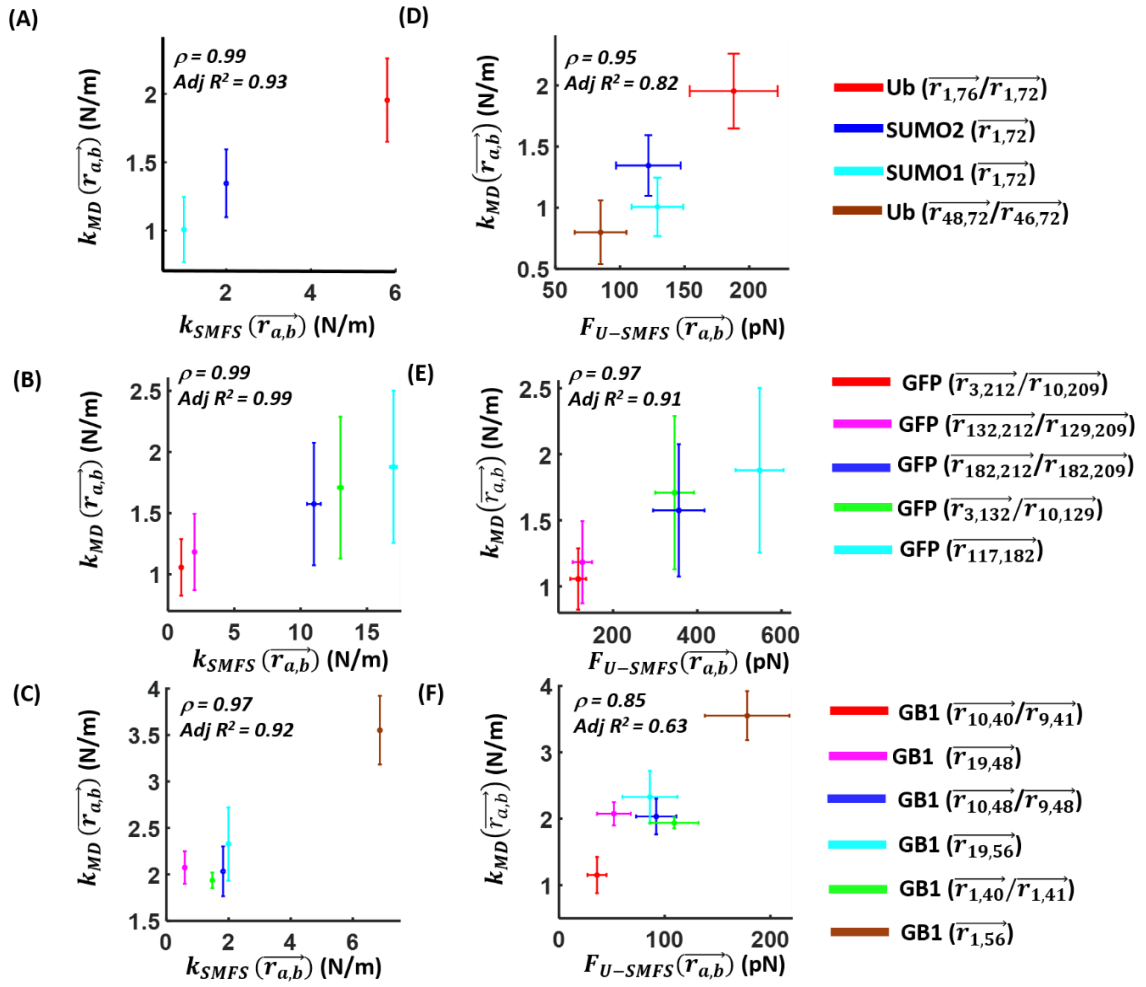

Figure 4: Correlation between derived spring constants with from SMFS data. (A-C) Scatter plots of spring constants $k_{MD}(\overrightarrow{r_{a,b}})$ computed from MD simulations (Eqn, 4) and that from SMFS experiments ($k_{SMFS}(\overrightarrow{r_{a,b}})$) for ubiquitin family proteins, GFP and GB1. (D-F) Scatter plots of $k_{MD}(\overrightarrow{r_{a,b}})$ and SMFS mechanical unfolding forces ($F_{U-SMFS}(\overrightarrow{r_{a,b}})$) for ubiquitin family proteins (pulling speeds of 0.3-0.4 μm/s), GFP (0.4 μm/s), and GB1 (0.4 μm/s). Data points for different pulling directions are color coded as indicated. For cases where the experimental and MD pulling directions are the same a single direction is provided otherwise the pair of directions (Expt/MD) are given (see Sec. 2.4.1). Average values and standard deviations in $k_{MD}(\overrightarrow{r_{a,b}})$ are derived from 10 MD trajectories of the proteins at the 1 μs timepoint. Experimental error bars where available are also indicated. Each panel shows the correlation coefficient (ρ) and Adjusted R-square ($Adj\ R^2$) between the computed and experimental datasets.

force (and computed directional spring constant) for Ub along $\overrightarrow{r_{1,72}}$ is well resolved and higher than that along $\overrightarrow{r_{46,72}}$ (Fig-4D). Both computations and experiments predict the $\overrightarrow{r_{46,72}}$ direction

in Ub to be slightly more mechanically pliant relative to the $\overrightarrow{r_{1,72}}$ direction in SUMO1/SUMO2. To summarize, we find a strong linear correlation between computed and experimental directional spring constants/unfolding force values of ubiquitin family proteins (Fig-4A and Fig-4D). Next, we examined the mechanical anisotropy in k_{MD} for GFP by computing the spring constant along the $\overrightarrow{r_{10,209}}$, $\overrightarrow{r_{129,209}}$, $\overrightarrow{r_{182,209}}$, $\overrightarrow{r_{10,129}}$ and $\overrightarrow{r_{117,182}}$ directions and find good correlations with SMFS experimental data (Fig-4B and Fig-4E).[37] However, the resolution of computed spring constants is poorer than that reported for experimental spring constants and unfolding forces. SMFS experiments identify three distinct classes of spring constants/unfolding forces for GFP, showing the $\overrightarrow{r_{117,182}}$ direction as the stiffest, the $\overrightarrow{r_{10,209}}$/$\overrightarrow{r_{129,209}}$ directions as the most flexible, and the $\overrightarrow{r_{182,209}}$/$\overrightarrow{r_{10,129}}$ directions as having intermediate flexibility. However, the computed k_{MD} distributions for these directions have significant overlaps. Nevertheless, we again find a strong linear correlation for the average computed k_{MD} with experimental spring constants/average unfolding forces. Finally, in case of GB1, only experimental unfolding forces (pulling speed of 0.4 μm/s) along $\overrightarrow{r_{1,56}}$, $\overrightarrow{r_{10,40}}$, $\overrightarrow{r_{19,48}}$, $\overrightarrow{r_{10,48}}$, $\overrightarrow{r_{19,56}}$ and $\overrightarrow{r_{1,40}}$ have been reported.[39] However, spring constants can be derived (Table-S7) from reported values of potential widths (ΔX_u) and unfolding rate constant at zero force (α_0). Here, we find a somewhat poorer correlation (relative to the other protein systems) between the computed k_{MD} and the experimental spring constants/unfolding force with the former failing to predict the mechanical anisotropy along $\overrightarrow{r_{19,48}}$ (Fig-4C and Fig-4F). This is because the computed k_{MD} overestimates the mechanical resistance for the former direction. Excluding the $\overrightarrow{r_{19,48}}$ data point, we find that fairly good correlations between experimental directional spring constant (and unfolding forces) and computed k_{MD} (Fig-4C and Fig-4F). Both experiments and computations show that the $\overrightarrow{r_{1,56}}$ and $\overrightarrow{r_{9,41}}$ directions possess the strongest and weakest mechanical resistances respectively. The mechanical resistance of $\overrightarrow{r_{1,41}}$, $\overrightarrow{r_{19,56}}$ and $\overrightarrow{r_{9,48}}$ directions are found to be of intermediate strength. The computed k_{MD} is also able to distinctly resolve the directions with the strongest and intermediate/lowest mechanical resistances correctly. Overall, we still find a reasonable correlation (including the $\overrightarrow{r_{19,48}}$ data point) of the average computed k_{MD} with experimental spring constants/average unfolding forces.

While the average k_{MD} values are lower than the experimentally measured k_{SMFS} in almost every case, relative trends in the mechanical anisotropy are still correctly predicted by computations in all the protein systems investigated here with only a single exception (the $\overrightarrow{r_{19,48}}$ direction for GB1).

As noted in the previous section, this is likely related to the lower coverage of the PES section along the pulling directions for GB1 in MD simulations. We note that computed $k_{MD}(\overrightarrow{r_{a,b}})$ with unmodified experimental pulling directions show poor correlations with experimental data (Fig-S8 and Fig-S9). As discussed in section 2.4.2, this is because in some cases the pulling force in SMFS experiments is applied on residues which belong to flexible loops and do not contribute to the mechanical resistance of the protein. Arbitrary modifications to the directions may produce a diversity of mechanical anisotropy trends (Fig-S8 and Fig-S9). To produce consistent results we propose a modification of the experimental direction by replacing the flexible loop residue with the closest neighbour in the structured part of the protein. Such modifications produce average $k_{MD}(\overrightarrow{r_{a,b}})$ trends in good agreement with experiment (Fig-4). Since our model for extracting directional spring constants (k_{MD}) is validated against SMFS experimental results here, we employ it further to predict the mechanical anisotropy of Ub along various functionally relevant lysine to C-terminal directions in the next section.

3.3 Mechanical anisotropy in Ub and its modulation upon complexation with partner proteins

The Ub protein possesses 7 lysines (Fig-5A) in its sequence which in paired combination with the protein C-terminus act as tethering points for substrate or poly-Ub chain linkages in the cell.[40,41] The lysine tethering point determines both the architecture of poly-Ub chains on the tagged cargo and the subsequent biological fate of the cargo.[42] We therefore examined Ub flexibility along directions connecting the C-terminus to each of the 7 lysines ($\overrightarrow{\boldsymbol{r}_{6,72}}$, $\overrightarrow{\boldsymbol{r}_{11,72}}$, $\overrightarrow{\boldsymbol{r}_{27,72}}$, $\overrightarrow{\boldsymbol{r}_{29,72}}$, $\overrightarrow{\boldsymbol{r}_{33,72}}$, $\overrightarrow{\boldsymbol{r}_{48,72}}$ and $\overrightarrow{\boldsymbol{r}_{63,72}}$) and find that these directions possess two distinct classes of spring constant values (Fig. 5B). Ub is significantly stiffer (k_{direct} is almost an order of magnitude larger) along $\overrightarrow{\boldsymbol{r}_{6,72}}$, $\overrightarrow{\boldsymbol{r}_{27,72}}$, $\overrightarrow{\boldsymbol{r}_{29,72}}$, and $\overrightarrow{\boldsymbol{r}_{63,72}}$ than along $\overrightarrow{\boldsymbol{r}_{11,72}}$, $\overrightarrow{\boldsymbol{r}_{33,72}}$, and $\overrightarrow{\boldsymbol{r}_{48,72}}$. Previous ENM based computational studies have proposed that the $\overrightarrow{\boldsymbol{r}_{11,76}}$ and $\overrightarrow{\boldsymbol{r}_{48,76}}$ directions in Ub are 2-fold more flexible relative to the $\overrightarrow{\boldsymbol{r}_{29,76}}$ and $\overrightarrow{\boldsymbol{r}_{63,76}}$.[17] Further, previous Steered Molecular Dynamics (SMD) simulations of Ub showed lower unfolding forces along $\overrightarrow{\boldsymbol{r}_{48,76}}$ relative to $\overrightarrow{\boldsymbol{r}_{1,76}}$ and even lower unfolding forces (< 50 pN) along $\overrightarrow{\boldsymbol{r}_{11,76}}$.[43] Our results here further predict two distinct (differing by about an order of magnitude) classes of spring constant values for Ub along the 7 lysine:C-term

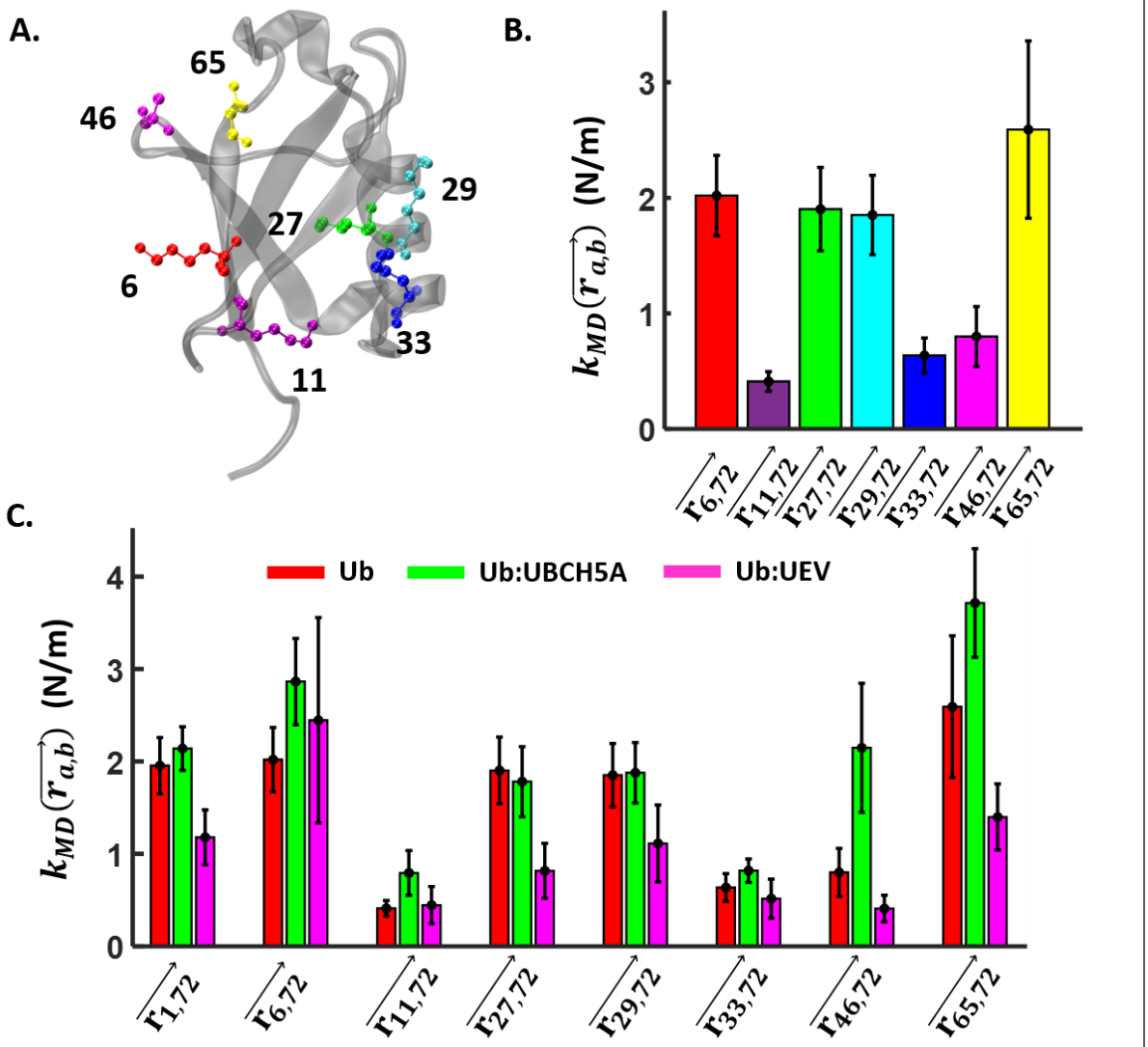

Figure 5: Mechanical Anisotropy in Ub and its alteration upon complexation with partner protein. (A) Positions of the 7 lysine residues (Lys6, Lys11, Lys27, Lys29, Lys33, Lys48, and Lys63) present in Ub which along with the C-terminal serve as tethering points for Mono- or poly-ubiquitination of substrate proteins. (B) Computed $k_{MD}\left(\overrightarrow{r_{a,b}}, T_S = 1\mu s\right)$ along 7 lysine to C-terminal (residue 72) pulling directions for Ub. Two of the directions containing Lys48 and Lys63, the direction was modified to $\overrightarrow{r_{46,72}}$ and $\overrightarrow{r_{65,72}}$ respectively as these residues fall on flexible loops (Sec. 2.4.1) (C) Changes in $k_{MD}(\overrightarrow{r_{a,b}}, T_S = 1\mu s)$ of free Ub upon complexation with two partner proteins UBCH5A and UEV. We find that the trends in mechanical anisotropy and its change upon complexation to be unaltered if the $\overrightarrow{r_{48,72}}$ and $\overrightarrow{r_{63,72}}$ directions are retained (Sec. S10 of ESI).

directions which can be tested by SMFS experiments. The set of high spring constant values (~ 2 N/m) are comparable to the N-C term spring constant of Ub (Fig-4A). Our recent work showed

that Ub changes its overall flexibility in a context sensitive manner upon binding to two protein partners (the UEV domain of TSG101 and UBCH5A) that have distinct functional consequences.[18] While the TSG101 UEV domain is a Ub recognition motif, recognizing ubiquitylated cargo, UBCH5A is an E2 ligase, which attaches Ub to protein substrates. Interestingly, by analyzing the trajectories of Ub:UBCH5A and Ub:Uev complexes, we find that these protein-protein interactions can also selectively and in a context sensitive manner alter the mechanical anisotropy of Ub relative to its free form. Upon binding to UBCH5A, Ub directional flexibility increases selectively along $\overrightarrow{\boldsymbol{r}_{6,72}}$, $\overrightarrow{\boldsymbol{r}_{48,72}}$ and $\overrightarrow{\boldsymbol{r}_{11,72}}$ but remains unchanged along other directions (Fig-5C). In contrast, binding to UEV selectively decreases the Ub directional flexibility along $\overrightarrow{\boldsymbol{r}_{1,72}}$, $\overrightarrow{\boldsymbol{r}_{27,72}}$, $\overrightarrow{\boldsymbol{r}_{29,72}}$, and $\overrightarrow{\boldsymbol{r}_{48,72}}$ while preserving the flexibility along other directions. Upon examining other lysine directions, we find that UBCH5A and UEV prefer to access Ub directions which have different flexibility scales and alter protein flexibility in contrasting ways (Fig. 5C). While UBCH5A tends to selectively decrease directional protein flexibility, UEV tends to selectively increase directional protein flexibility.

3.4 A Scalable Scheme to Compute Directional Spring Constants Based on Quasiharmonic/ Normal Modes

The previous sections show that protein directional spring constants $k_{MD}(\overrightarrow{r_{a,b}})$ can easily be computed along desired directions from equilibrated MD simulation data. A deeper understanding of the directional protein flexibility in relation to the protein structure can be obtained by examining the contributions to k_{MD} from collective protein motions. For this PCA of equilibrium MD trajectories can be used which provides a basis of quasiharmonic modes to express protein dynamics. Further, it would be desirable to extract directional flexibilities from cheaper computational schemes such as coarse-grained ENM analysis which provide a harmonic normal mode basis to express protein flexibility.[17,44] In both PCA and ENM, the objective is to compute directional spring constants from a set of orthogonal eigenvectors and their eigenvalues. In this subsection, using Ub as a model protein system, we demonstrate a simple covariance-propagation (CP) scheme (Fig-S4 and Sec 2.4.4) to extract directional protein flexibility from an eigenvector basis. First, we consider PC modes obtained by performing C_α only PCA (Sec. 2.2) on 10 x 1 μs MD trajectories of Ub. For each trajectory, PCs along with their corresponding eigenvalues are

processed (Sec. 2.4.4) to extract directional variances $\sigma^2_{CP-PCA}(\overrightarrow{r_{a,b}})$ which are converted into $k_{CP-PCA}(\overrightarrow{r_{a,b}})$ using Eqn. 4. The $k_{CP-PCA}(\overrightarrow{r_{a,b}})$ thus computed can be validated against $k_{MD}(\overrightarrow{r_{a,b}})$ obtained directly from distance distributions in the corresponding MD trajectories using Eqns. 3 and 4 in Sec. 2.4.2. The same scheme is transferable and can be employed to compute $k_{CP-ENM}(\overrightarrow{r_{a,b}})$ from coarse-grained ENM normal modes (Fig-S4 and Sec. 2.4.4) derived from the Ub crystal structure.

3.4.1 Validation against MD Trajectories

First, in order to validate the CP scheme, we compute $k_{MD}(\overrightarrow{r_{a,b}}, T_s = 1\mu s)$ from each of the 10 MD trajectories of Ub (C_α atoms only) for all possible $\overrightarrow{r_{a,b}}$ excluding nearest neighbour residue pairs in the core 72 residue sequence of Ub, a total of $\left(\binom{72}{2} - 71 = 2485\right)$ directions. The $\sigma^2_{CP-PCA}(\overrightarrow{r_{a,b}})$ computed using PC modes reproduces most (80 %) of the directional variances from direct MD distributions (Fig-6A). Further, the agreement is also clearly visible in the scatter plot for $k_{MD}(\overrightarrow{r_{a,b}})$ and $k_{CP-PCA}(\overrightarrow{r_{a,b}})$ computed using the two schemes (Fig-6B). However, it is also apparent from these plots that there are some directions (20 %) for which the variance in distributions is not well reproduced by the CP scheme. A deeper investigation reveals that the CVCF-traces along these directions has not converged to a plateau indicating non-equilibrium sampling over the 1μs timescale (Fig-S11 of ESI) with non-Gaussian distance distributions. For instance, in Fig-S11 of ESI the computations of the directional variance using the CP scheme (using all modes) and direct MD are compared for three directions: $\overrightarrow{r_{8,10}}$, $\overrightarrow{r_{24,33}}$, and $\overrightarrow{r_{41,60}}$ in a specific trajectory of Ub. The CP scheme reproduces $\sigma^2_{CVCF}(\overrightarrow{r_{24,133}})$ and $\sigma^2_{CVCF}(\overrightarrow{r_{41,60}})$ from MD distance distributions perfectly (Fig-S11A) as these directions exhibit converged CVCF plateaus (Fig-S11B and Fig-S11C) and distributions which can be described by single Gaussian fits (Fig-S11E-F). In contrast, for the $\overrightarrow{r_{8,10}}$ direction which shows a bimodal distance distribution (Fig-S11D) and no CVCF-trace plateau at 1 μs, the $\sigma^2_{CVCF}(\overrightarrow{r_{8,10}})$ from direct MD is poorly reproduced by the CP scheme (Fig-S11A). We also examined the ability of the method developed by Eyal and Bahar [17] for protein spring constants, to reproduce the $k_{MD}(\overrightarrow{r_{a,b}})$ computed from MD distance distributions. While the original description of the method was based on coarse-grained ENM normal modes, we follow the prescription by Eyal and Bahar to compute $k_{EB-PCA}(\overrightarrow{r_{a,b}})$ for all the

2485 directions of Ub from (C_α-atom) PCA on each of the 10 μs trajectories of the protein (Eqn. 9). From the scatter plot in Fig-6C it is apparent that the formalism of Eyal and Bahar using PCA fares relatively poorly in reproducing the spring constants of Ub from direct MD. The scatter plot of $k_{MD}(\overrightarrow{r_{a,b}})$ and $k_{EB-PCA}(\overrightarrow{r_{a,b}})$ shows a large spread, is shifted below the *x=y* line, and deviates from a linear fit at higher spring constant values. We note, that a fundamental difference between the CP framework proposed in Sec 2.4.4 and the prescription of Eyal and Bahar lies in the definition of the directional spring constants. While the former derives the directional spring constant from the equipartition principle (Eqn. 10) which balances thermal energy with cumulative mode energy, the latter uses Eqn. 8 which balances thermal force with the cumulative mode deformation force.

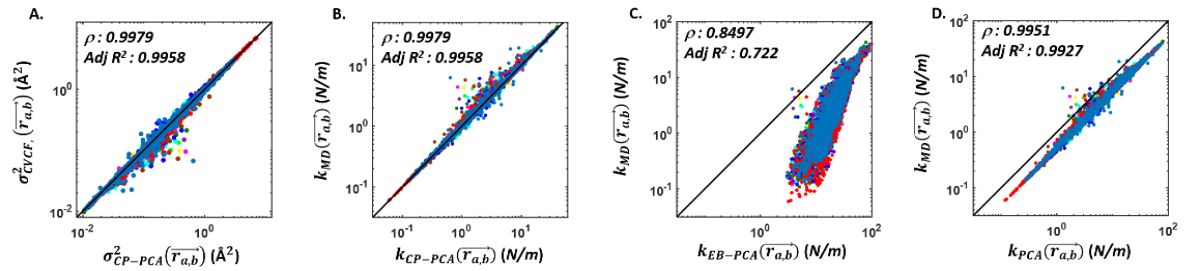

Figure 6: Correlations between Ub spring constants from MD trajectories and those computed using PC modes using three different methods: (A) Correlation between $\sigma^2_{CVCF}(\overrightarrow{r_{a,b}})$ for Ub obtained from distance distributions in MD trajectories (Eqn. 3)s and $\sigma^2_{CP-PCA}(\overrightarrow{r_{a,b}})$ obtained using the method of covariance propagation and C_α atom PCA with the full set of PC modes (Eqn. 16) (B) Correlation between $k_{MD}(\overrightarrow{r_{a,b}})$ and $k_{CP-PCA}(\overrightarrow{r_{a,b}})$ obtained from the variances in panel A using Eqn. 4. (C) Correlation between $k_{MD}(\overrightarrow{r_{a,b}})$ and $k_{EB-PCA}(\overrightarrow{r_{a,b}})$ obtained using the formalism of Eyal and Bahar on PC modes (Eqn. 9). (C) Correlation between $k_{MD}(\overrightarrow{r_{a,b}})$ and $k_{PCA}(\overrightarrow{r_{a,b}})$ obtained by using the modified formalism of Eyal and Bahar on PC modes (Eqn. 11). Each panel shows the correlation coefficient (ρ) and Adjusted R-square (*Adj* R^2) between the datasets. The diagonal line depicts *x=y*.

As, we will show in the next subsection the definition adopted by Eyal and Bahar also leads to an anomalous dependence of the computed spring constants on high frequency modes of the system. We therefore, modified the Eyal-Bahar formalism to adopt an equipartition based definition ($k_{PCA}(\overrightarrow{r_{a,b}})$ from Eqn. 11) of the spring constant. The correlation of the modified spring constant $k_{PCA}(\overrightarrow{r_{a,b}})$ with $k_{MD}(\overrightarrow{r_{a,b}})$ is significantly improved and matches that obtained for the CP scheme, except for a systematic underestimation of values relative to that derived directly from MD (Fig-

6D). Thus, having established the reliability of our CP scheme to compute protein spring constants from PCA and ENM, we examine the ability of scheme to reveal mode contributions to the protein directional spring constants in the next section.

3.4.2Collective PC and ENM mode contributions to the directional spring constant

PCA as applied to MD trajectories highlights the prominence of low frequency collective vibrational modes in protein dynamics. PCs sorted by decreasing order of their eigenvalues (variances) show a sharp drop in terms of their contributions to the total variance produced in MD trajectories (Fig-S12 of ESI). Specifically, more than 80% of the total fluctuations originate from top ten lowest frequency modes in globular proteins examined here (Fig-S12). Since the overall spring constant is inversely proportional to the variance of bond vector fluctuations, it is expected to decay as a function of mode index cumulatively when modes are sorted based on decreasing fluctuations. Here, we investigate the trends in directional variances and spring constants computed using the CP framework of Sec 2.4.4 and those using both the original and modified Eyal-Bahar methods[17] as a function of the number of PC modes *M* used in the computations. Specifically in Fig-7, we investigate the convergence of these measures along $\overrightarrow{r_{1,72}}$ to the σ^2_{CVCF} and k_{MD} limits at $T_s = 1\ \mu s$ in 10 MD trajectories of Ub. We perform PCA (Sec. 2.2) on each of the MD trajectories considering C_α atoms only and use the CP scheme (Sec 2.4.4) to compute $\sigma^2_{CP-PCA}(\overrightarrow{r_{1,72}}, M)$ and $k_{CP-PCA}(\overrightarrow{r_{1,72}}, M)$ as a function of the number of PC modes sorted based on their decreasing eigenvalues. These values are also compared with the total variance/spring constants of the bond vector $\overrightarrow{r_{1,72}}$ obtained directly from MD (horizontal coloured lines in Fig-7). As expected, in each of the 1 μs MD trajectories of Ub, $\sigma^2_{CP-PCA}(\overrightarrow{r_{1,72}}, \mathrm{M})$ increases with number of PC contributions and saturates around $M = 50$ indicating that the last ¾ high frequency modes have negligible contributions to the directional variance in MD trajectories (Fig-7A). Moreover, $\sigma^2_{CP-PCA}(\overrightarrow{r_{1,72}}, M)$ converges to within 10 % of $\sigma^2_{CVCF}(\overrightarrow{r_{1,72}})$ upon including the top 50 PCs. Accordingly, $k_{CP-PCA}(\overrightarrow{r_{1,72}}, M)$ also decays rapidly with *M* and saturates to a value close to k_{MD} (Fig-7B).

We find similar trends for the convergence of $k_{CP-ENM}(\overrightarrow{r_{1,72,}}, \overrightarrow{M})$ computed from Ub crystal structures (Fig-S13 of ESI). However, convergence of $\sigma^2_{CP-ENM}(\overrightarrow{r_{1,72}}, \mathrm{M})$ and $k_{CP-ENM}(\overrightarrow{r_{1,72}})$ with number of modes is more gradual as compared to their PCA counterparts (Fig-S13 A-B).

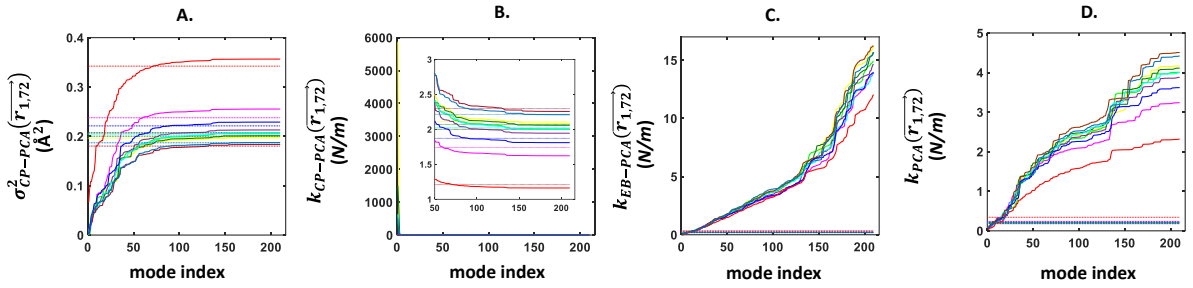

Figure 7: Convergence of variance and spring constants as a function of PC modes. (A) $\sigma^2_{CP-PCA}(\overrightarrow{r_{1,72}})$ plotted cumulatively as a function of vibrational modes obtained from PCA followed by method of error propagation on 10 independent 1 μs MD trajectories represented by different colors. The dotted horizontal lines represent the $\sigma^2_{CVCF}(\overrightarrow{r_{1,72}})$ obtained at end of the corresponding MD trajectories (Eqns. 2 and 3 in Sec. 2.4.2). (B) $k_{CP-PCA}(\overrightarrow{r_{1,72}})$ plotted cumulatively as a function of PC mode index (solid line) obtained from $\sigma^2_{CP-PCA}(\overrightarrow{r_{1,72}})$ using Eqn. 4 in Sec. 2.4.2. Also shown are the $k_{MD}(\overrightarrow{r_{1,72}})$ values obtained from distance distributions of the corresponding MD trajectory (dashed line). A zoomed in version for the PC modes 50-210 is also shown in inset. (C) $k_{EB-PCA}(\overrightarrow{r_{1,72}})$ from Eqn. 9 plotted cumulatively as a function of PC modes (solid line) shows a large deviation from $k_{MD}(\overrightarrow{r_{1,72}})$ values obtained directly from MD trajectories (dashed line) and does not converge/saturate with number of modes. (D) $k_{PCA}(\overrightarrow{r_{1,72}})$ from Eqn. 11 plotted cumulatively as a function of PC modes (solid line) also shows a large deviation from $k_{MD}(\overrightarrow{r_{1,72}})$ values obtained directly from MD trajectories (dashed line) but converges/saturates with increasing mode index.

While contributions to $\sigma^2_{CP-PCA}(\overrightarrow{r_{1,72}}, \mathrm{M})$ and $k_{CP-PCA}(\overrightarrow{r_{1,72}})$ from modes M > 50 are negligibly small, modes ~ 50-100 provide a significant contribution to $\sigma^2_{CP-ENM}(\overrightarrow{r_{1,72}}, \mathrm{M})$ and $k_{CP-ENM}(\overrightarrow{r_{1,72}})$ (Fig-S13 A-B). In stark contrast, $k_{EB-ENM/PCA}(\overrightarrow{r_{1,72}})$ values do not decay but instead show a rise which does not saturate even upon including all modes in the spring constant calculation (Fig-7C). Clearly, high frequency local modes contribute significantly to the directional flexibility $k_{EB-PCA/ENM}(\overrightarrow{r_{1,72}})$ of Ub along $\overrightarrow{r_{1,72}}$ which appears unphysical and inconsistent with the behaviour observed for the directional variance of the protein (Fig-7A). Moreover, there is a large discrepancy between $k_{MD}(\overrightarrow{r_{1,72}})$ and the spring constants obtained using the Eyal and Bahar framework (Fig-7C). Upon incorporating the modified description of the spring constant (Eqn. 11) as discussed in the previous section, the contribution of high frequency normal modes to $k_{ENM}(\overrightarrow{r_{1,72}})$ is lowered which leads to a weak convergence of the spring constant values at large *M* (Fig-7D).

3.5 Correspondence of ENM and MD derived spring constants

In this subsection, we apply our CP formalism to compute $k_{CP-ENM}(\overrightarrow{r_{a,b}})$ from coarse-grained ENM derived normal modes (Sec. 2.4.4). In previous sections, we investigated a few selected $\overrightarrow{r_{a,b}}$ and found strong correlations between computed $k_{MD}(\overrightarrow{r_{a,b}})$ and that from SMFS experiments. Here, we examine the ability of coarse-grained ENM to reproduce MD derived directional spring constants for Ub, SUMO, SUMO2, GFP and GB1.

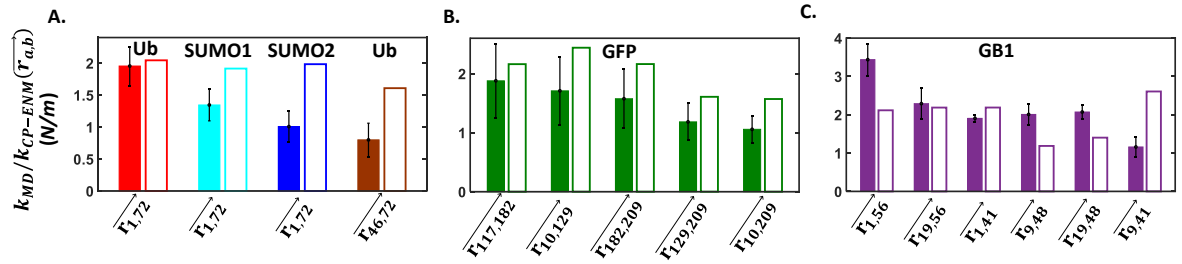

Figure 8: Protein directional spring constants as viewed by coarse-grained ENM and MD simulations. Computed $k_{MD}(\overrightarrow{r_{a,b}})$ (solid bar) and $k_{CP-ENM}(\overrightarrow{r_{a,b}})$ (hollow bar) for Ub family protein along $\overrightarrow{r_{1,72}}$ and $\overrightarrow{r_{48,72}}$ in case of Ub (A), for GFP along $\overrightarrow{r_{117,182}}$, $\overrightarrow{r_{10,128}}$, $\overrightarrow{r_{182,212}}$, $\overrightarrow{r_{132,212}}$ and $\overrightarrow{r_{3,212}}$ (B) and for GB1 along $\overrightarrow{r_{1,56}}$, $\overrightarrow{r_{10,40}}$, $\overrightarrow{r_{21,48}}$, $\overrightarrow{r_{10,48}}$, $\overrightarrow{r_{19,56}}$ and $\overrightarrow{r_{1,41}}$ (C). Average and standard deviation in computed k_{MD} (in solid bar) is taken over the set of 10 MD trajectories for each protein at 1 μs timescale.

For Ub family proteins, we have showed that spring constants from both SMFS experiments and MD simulations can resolve the flexibilities of Ub from its SUMO isoforms. In contrast, the $k_{CP-ENM}(\overrightarrow{r_{a,b}})$ computed from Ub, SUMO1 and SUMO2 crystal structures are virtually indistinguishable (Fig-8A). The poor resolution of ENM spring constants relative to MD is expected since the ubiquitin family proteins are structurally homologous with superimposable backbones and coarse-grained ENM is a C_α based method. However, ENM is able to resolve the Ub mechanical anisotropy along $\overrightarrow{r_{1,72}}$ and the $\overrightarrow{r_{46,72}}$ directions showing significantly higher spring constants along the former direction.. In case of GFP, ENM can correctly distinguish the flexibility along $\overrightarrow{r_{117,182}}$, $\overrightarrow{r_{10,129}}$, and $\overrightarrow{r_{182,212}}$, from that along $\overrightarrow{r_{129,209}}$ and $\overrightarrow{r_{10,209}}$ (Fig-8B) as captured in MD and SMFS experiments. However, ENM overestimates the $k_{CP-ENM}(\overrightarrow{r_{a,b}})$ along $\overrightarrow{r_{10,128}}$ and $\overrightarrow{r_{182,209}}$ relative to that along $\overrightarrow{r_{117,182}}$. In GB1, however, ENM fares poorly, failing to predict the significantly high directional spring constant along $\overrightarrow{r_{1,56}}$ relative to the other directions and the

least mechanically resistant direction along $\overrightarrow{r_{9,41}}$ (Fig-8C). Both of these trends are picked up by MD.

4. Discussion

SMFS experiments use an effective 2-state (folded/unfolded) model to map unfolding force data back to the underlying PES section along the pulling reaction coordinate.[5,7,10] Typically, Monte-Carlo simulations of the 2-state kinetics are performed to estimate the distance between the native and transition states (the potential width Δx_U) and the equilibrium unfolding rate constant (α_0), which then together provide the directional spring constant ($k_{SMFS}(\overrightarrow{r_{a,b}})$). Here, we show that atomistic MD simulations provide a means to assess the correspondence of the parameters extracted from SMFS to the underlying PES accessed by proteins under equilibrium sampling conditions. The assessment of equilibrium sampling conditions was based on a CVCF-trace analysis which also provides a means to estimate protein flexibility.[18] By running multiple trajectories, MD simulations can provide an assessment of the effective curvature and roughness of the section of PES sampled over a fixed timescale in terms of $<\sigma^2_{CVCF}(\overrightarrow{r_{a,b}})>$ and SD_{CVCF} respectively. We note that while both effective curvatures, extracted from MD (k_{MD}) and SMFS (k_{SMFS}**),** are based on the equipartition theorem, and their correlation is not immediately apparent. For instance, the former depends only on the thermal amplitude $\sigma_{CVCF}(\overrightarrow{r_{a,b}})$ and is related by a simple inverse relationship to the variance. In contrast, the latter is given in terms of the activation barrier ($\Delta G^{\ddagger}$) for unfolding (obtained from α_0) and the potential width ($k_{SMFS} = \frac{2\Delta G^{\ddagger}}{(\Delta X_U)^2}$). Our analysis here reveals that microsecond MD trajectories, in 5 different globular proteins, produce thermal amplitudes which cover only about 7-42% of the potential width along the pulling reaction coordinates (Table-S7). Thus it is encouraging to note that average k_{MD} values show strong correlations (ρ ~ 0.97-0.99 with *Adj* R^2 ~ 0.92-0.99) with the MD data. Our data show that MD underestimates the mechanical flexibility relative to that predicted from SMFS for directions ($\overrightarrow{r_{19,48}}$ for GB1) where the thermal amplitude is small relative to the potential width. A second interesting observation from MD is that roughness of the PES section along the pulling coordinates appears to be significant. The heterogeneity in sampling impacts the resolution of the computed k_{MD} which in general appears to be lower than experiments (Fig-4). The higher resolution of the SMFS data may be due to biased sampling conditions along the pulling coordinate in experiments

as opposed to the free thermal sampling conditions in MD. Importantly, however, these observations, along with excellent correlations seen for average spring constants, do suggest that directional flexibilities which are well resolved in MD should be also well resolved in SMFS experiments. A third crucial point touched upon in our analysis is the definition of the pulling reaction coordinate $\overrightarrow{r_{a,b}}$ for which spring constants are computed in MD. We show that a simplistic choice of $\overrightarrow{r_{a,b}}$ based on $C\alpha$ atom positions of residue pairs (a and b) on which the pulling force is directly applied in SMFS experiments may severely overestimate computed directional flexibilities (Fig-S8 of ESI) when either one or both residues of the pair lie on flexible protein segments. Such segments are expected to extend with pulling force without contributing to the mechanical resistance of the protein fold. In this case, we propose that the $\overrightarrow{r_{a,b}}$ definitions should be modified to use $C\alpha$ atoms of residue pairs from the structured protein core (Sec 2.4.1 and Table-S3). We show that a CVCF-trace analysis of the relative flexibility of candidate protein segments (Fig-S2 of ESI) with respect to the structured protein core can provide objective choices of $\overrightarrow{r_{a,b}}$ for computing directional protein flexibilities. To summarize, our studies show that an atomistic MD-based analysis of directional protein flexibilities and anisotropies can provide useful insights for mapping the reaction coordinates and PES parameters from SMFS experiments to equilibrium protein dynamics.

We have applied our MD-based computational framework to study the mechanical anisotropy of Ub in free form and in complexes with protein partners. Our computations reveal intriguingly that the directional flexibility of free Ub associated with conserved lysines (Fig-5A) fall into two distinct scales (Fig-5B) with the directions $\overrightarrow{r_{6,72}}$, $\overrightarrow{r_{27,72}}$, $\overrightarrow{r_{29,72}}$, and $\overrightarrow{r_{63,72}}$ an order of magnitude more stiffer than the directions $\overrightarrow{r_{11,72}}$, $\overrightarrow{r_{33,72}}$, and $\overrightarrow{r_{48,72}}$. The mechanical isotropy of Ub may be relevant for understanding mono- or poly-ubiquitination reactions in the cellular context. For instance, the E2 ligase UBCH5A innately prefers linking poly-Ub chains using lysine-11, lysine-48 and lysine-63 residues.[45] It is known that lysine-11 and lysine-48 linked chains trigger substrate protein degradation more frequently than other modifications[46] as the tethers are less likely to produce chain branches which can impede degradation[47]. In contrast, atypical poly-Ub chains linked through other lysines (residues 6,27,29,33, and 63) can create a range of molecular signals including proteasomal degradation.[48–50] Previous studies have employed structural arguments (proximity to the E2 catalytic cysteine) to rationalize lysine-11 linked poly-Ub chain formation.[45]

Our analysis adds a fresh dynamical perspective by predicting that two out of three flexible directions in free Ub, $\overrightarrow{r_{11,72}}$ and $\overrightarrow{r_{48,72}}$, become distinctly rigid upon binding to UBCH5A (Fig-5C). Additionally, one of the four rigid directions, $\overrightarrow{r_{6,72}}$ is further rigidified (Fig-5C). Interestingly, the binding to the Ub recognition domain UEV has exactly the opposite effect to that of the UBCH5A ligase, that of softening Ub directional flexibility (Fig-5C). Our calculations predict UEV binding to preferentially impact the rigid directions, reducing the spring constants of Ub along $\overrightarrow{r_{1,72}}$, $\overrightarrow{r_{27,72}}$, $\overrightarrow{r_{29,72}}$, and $\overrightarrow{r_{63,72}}$. Additionally, one the three flexible directions of Ub, $\overrightarrow{r_{46,72}}$ shows a distinctly lower spring constant in the bound form. These are clear predictions, well resolved in MD, for the mechanical anisotropy in free Ub and its contrasting changes upon binding to functionally distinct protein partners which can be tested using SMFS experiments.

Finally, we have demonstrated the utility of the MD-based directional flexibility analysis in terms of examining and tuning the performance of coarse-grained ENM schemes. These scalable methods have shown great promise in previous applications for determining the directional flexibility of proteins.[15,17] However, given the sparse experimental dataset on protein mechanical anisotropy and the difficulty in combining directional spring constant data for different proteins, it has been difficult to quantitatively evaluate the predictive ability of ENM-based methods. Using a large dataset of bond vector fluctuations derived from MD and a transferrable CP framework (Sec. 2.4.3) we were able to demonstrate some limitations in a ENM scheme for computing directional spring-constants (Sec. 3.4) and propose modifications to overcome these issues. Calculations on ubiquitin family proteins, GFP and GB1 show both abilities and limitations of coarse-grained ENM models to resolve directional flexibilities of proteins (Fig 8). ENM based spring constants were found to be able to resolve the mechanical anisotropy of Ub along two directions, but were unable to resolve the N-C term stiffness for structurally homologous ubiquitin family proteins. For GFP, the ENM models were able to distinguish between the lowest two and highest three spring constants, but not the trends within each class. In GB1, ENM models completely fail to predict even major trends in protein mechanical anisotropy. Preliminary calculations wherein the coarse-grained ENM interaction potential was reparametrized with spring constants derived from MD simulations (data in Fig-5B) show that the ENM resolution can be improved to approach that of MD in reproducing directional spring constants for all proteins considered here (Fig-S14). Parameterization of ENM models with MD has been proposed before[51] and holds great promise for improving the predictive abilities of these scalable models.

5. Conclusions

In this paper we have demonstrated that equilibrium atomistic MD simulations can be used to estimate the directional flexibility of proteins and produce average spring constant values which correlate with SMFS experimental data. The MD-based framework to estimate protein directional spring constants is promising as it enables the extraction of the full mechanical anisotropy of the protein considering all residue pairs and present well-resolved directions for SMFS experiments to investigate. Further, as demonstrated here insights into the features of the native PES can be extracted from a CVCF-trace analysis which can be correlated to experimental data. We have also demonstrated that the MD-based framework can be useful to evaluate computationally cheaper and scalable coarse-grained ENM models and improve them. More experimental data on protein mechanical anisotropy is desirable to develop these methods into powerful predictive tools.

Acknowledgements

S. P., and R.V. acknowledge the support of the Department of Atomic Energy, Government of India, under Project No. 12-R&D-TFR-5.10-0100. The authors would like to thank Prof. A. S. R. Koti for stimulating discussions on SMFS techniques and data.

References

(1) Hoffmann, T.; Dougan, L. Single Molecule Force Spectroscopy Using Polyproteins. *Chemical Society Reviews*. July 21, 2012, pp 4781–4796.

(2) Clausen-Schaumann, H.; Seitz, M.; Krautbauer, R.; Gaub, H. E. Force Spectroscopy with Single Bio-Molecules. *Current Opinion in Chemical Biology*. Current Biology Ltd October 1, 2000, pp 524–530.

(3) Puchner, E. M.; Gaub, H. E. Force and Function: Probing Proteins with AFM-Based Force Spectroscopy. *Current Opinion in Structural Biology*. Elsevier Current Trends October 1, 2009, pp 605–614.

(4) Fisher, T. E.; Oberhauser, A. F.; Carrion-Vazquez, M.; Marszalek, P. E.; Fernandez, J. M. The Study of Protein Mechanics with the Atomic Force Microscope. *Trends Biochem. Sci.* **1999**, *24* (10), 379–384.

(5) Kotamarthi, H. C.; Sharma, R.; Rama, S.; Ainavarapu, K. Single-Molecule Studies on

PolySUMO Proteins Reveal Their Mechanical Flexibility. *Biophysj* **2013**, *104*, 2273–2281.

(6) Rief, M.; Fernandez, J. M.; Gaub, H. E. Elastically Coupled Two-Level Systems as a Model for Biopolymer Extensibility. *Phys. Rev. Lett.* **1998**, *81* (21), 4764–4767.

(7) Dietz, H.; Berkemeier, F.; Bertz, M.; Rief, M. Anisotropic Deformation Response of Single Protein Molecules. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (34), 12724–12728.

(8) Kotamarthi, H. C.; Yadav, A.; Koti Ainavarapu, S. R. Small Peptide Binding Stiffens the Ubiquitin-like Protein SUMO1. *Biophys. J.* **2015**, *108* (2), 360–367.

(9) Carrion-Vazquez, M.; Li, H.; Lu, H.; Marszalek, P. E.; Oberhauser, A. F.; Fernandez, J. M. The Mechanical Stability of Ubiquitin Is Linkage Dependent. *Nat. Struct. Mol. Biol.* **2003**, *10* (9), 738–743.

(10) Li, Y. D.; Lamour, G.; Gsponer, J.; Zheng, P.; Li, H. The Molecular Mechanism Underlying Mechanical Anisotropy of the Protein GB1. *Biophys. J.* **2012**, *103* (11), 2361–2368.

(11) Brockwell, D. J.; Paci, E.; Zinober, R. C.; Beddard, G. S.; Olmsted, P. D.; Smith, D. A.; Perham, R. N.; Radford, S. E. Pulling Geometry Defines the Mechanical Resistance of a β-Sheet Protein. *Nat. Struct. Biol.* **2003**, *10* (9), 731–737.

(12) Sułkowska, J. I.; Cieplak, M. Mechanical Stretching of Proteins - A Theoretical Survey of the Protein Data Bank. *J. Phys. Condens. Matter* **2007**, *19* (28), 60.

(13) Schoeler, C.; Bernardi, R. C.; Malinowska, K. H.; Durner, E.; Ott, W.; Bayer, E. A.; Schulten, K.; Nash, M. A.; Gaub, H. E. Mapping Mechanical Force Propagation through Biomolecular Complexes. *Nano Lett.* **2015**, *15* (11), 7370–7376.

(14) Xiao, S.; Stacklies, W.; Cetinkaya, M.; Markert, B.; Gräter, F. Mechanical Response of Silk Crystalline Units from Force-Distribution Analysis. *Biophys. J.* **2009**, *96* (10), 3997–4005.

(15) Sacquin-Mora, S.; Lavery, R. Modeling the Mechanical Response of Proteins to Anisotropic Deformation. *ChemPhysChem* **2009**, *10* (1), 115–118.

(16) Dietz, H.; Rief, M. Elastic Bond Network Model for Protein Unfolding Mechanics. *Phys. Rev. Lett.* **2008**, *100* (9), 098101.

(17) Eyal, E.; Bahar, I. Toward a Molecular Understanding of the Anisotropic Response of Proteins to External Forces: Insights from Elastic Network Models. *Biophys. J.* **2008**, *94* (9), 3424–3435.

(18) Paul, S.; Ainavarapu, S. R. K.; Venkatramani, R. Variance of Atomic Coordinates as a Dynamical Metric to Distinguish Proteins and Protein–Protein Interactions in Molecular Dynamics Simulations. *J. Phys. Chem. B* **2020**, *124* (21), 4247–4262.

(19) Sawle, L.; Ghosh, K. Convergence of Molecular Dynamics Simulation of Protein Native States: Feasibility vs Self-Consistency Dilemma. *J. Chem. Theory Comput.* **2016**, *12* (2), 861–869.

(20) Nemec, M.; Hoffmann, D. Quantitative Assessment of Molecular Dynamics Sampling for Flexible Systems. *J. Chem. Theory Comput.* **2017**, *13* (2), 400–414.

(21) Grossfield, A.; Zuckerman, D. M. Quantifying Uncertainty and Sampling Quality in Biomolecular Simulations. *Annu. Rep. Comput. Chem.* **2009**, *5*, 23–48.

(22) Zuckerman, D. M. Equilibrium Sampling in Biomolecular Simulations. *Annu. Rev. Biophys.* **2011**, *40* (1), 41–62.

(23) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.

(24) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713.

(25) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; De Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2016**, *14* (1), 71–73.

(26) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev,

A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *Journal of Computational Chemistry*. John Wiley & Sons, Ltd December 1, 2005, pp 1668–1688.

(27) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29* (11), 1859–1865.

(28) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **2016**, *12* (1), 405–413.

(29) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092.

(30) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31* (3), 1695–1697.

(31) Nosé, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52* (2), 255–268.

(32) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52* (12), 7182–7190.

(33) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38.

(34) Daidone, I.; Amadei, A. Essential Dynamics: Foundation and Applications. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2* (5), 762–770.

(35) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential Dynamics of Proteins. *Proteins Struct. Funct. Genet.* **1993**, *17* (4), 412–425.

(36) Lezon, T. R.; Shrivastava, I. H.; Yang, Z.; Bahar, I. Elastic Network Models For Biomolecular Dynamics: Theory and Application to Membrane Proteins and Viruses; 2009; pp 129–158.

(37) Dietz, H.; Berkemeier, F.; Bertz, M.; Rief, M. Anisotropic Deformation Response of

Single Protein Molecules. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (34), 12724–12728.

(38) Eyal, E.; Chennubhotla, C.; Yang, L.-W.; Bahar, I. Anisotropic Fluctuations of Amino Acids in Protein Structures: Insights from X-Ray Crystallography and Elastic Network Models. *Bioinformatics* **2007**, *23* (13), i175–i184.

(39) Li, Y. D.; Lamour, G.; Gsponer, J.; Zheng, P.; Li, H. The Molecular Mechanism Underlying Mechanical Anisotropy of the Protein GB1. *Biophys. J.* **2012**, *103* (11), 2361–2368.

(40) Komander, D.; Rape, M. The Ubiquitin Code. *Annu. Rev. Biochem.* **2012**, *81* (1), 203–229.

(41) Yau, R.; Rape, M. The Increasing Complexity of the Ubiquitin Code. *Nat. Cell Biol.* **2016**, *18* (6), 579–586.

(42) Komander, D. The Emerging Complexity of Protein Ubiquitination. *Biochem. Soc. Trans.* **2009**, *37* (5), 937–953.

(43) Li, P.-C.; Makarov, D. E. Simulation of the Mechanical Unfolding of Ubiquitin: Probing Different Unfolding Reaction Coordinates by Changing the Pulling Geometry. *J. Chem. Phys.* **2004**, *121* (10), 4826–4832.

(44) Tirion, M. M. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys. Rev. Lett.* **1996**, *77* (9), 1905–1908.

(45) Bosanac, I.; Phu, L.; Pan, B.; Zilberleyb, I.; Maurer, B.; Dixit, V. M.; Hymowitz, S. G.; Kirkpatrick, D. S. Modulation of K11-Linkage Formation by Variable Loop Residues within UbcH5A. *J. Mol. Biol.* **2011**, *408* (3), 420–431.

(46) Komander, D.; Rape, M. The Ubiquitin Code. *Annu. Rev. Biochem.* **2012**, *81* (1), 203–229.

(47) Hyoung, T. K.; Kwang, P. K.; Lledias, F.; Kisselev, A. F.; Scaglione, K. M.; Skowyra, D.; Gygi, S. P.; Goldberg, A. L. Certain Pairs of Ubiquitin-Conjugating Enzymes (E2s) and Ubiquitin-Protein Ligases (E3s) Synthesize Nondegradable Forked Ubiquitin Chains Containing All Possible Isopeptide Linkages. *J. Biol. Chem.* **2007**, *282* (24), 17375–

17386.

(48) Jin, L.; Williamson, A.; Banerjee, S.; Philipp, I.; Rape, M. Mechanism of Ubiquitin-Chain Formation by the Human Anaphase-Promoting Complex. *Cell* **2008**, *133* (4), 653–665.

(49) Kwon, Y. T.; Ciechanover, A. The Ubiquitin Code in the Ubiquitin-Proteasome System and Autophagy. *Trends in Biochemical Sciences*. Elsevier Ltd November 1, 2017, pp 873–886.

(50) Ikeda, F.; Dikic, I. Atypical Ubiquitin Chains: New Molecular Signals. "Protein Modifications: Beyond the Usual Suspects" Review Series. *EMBO Rep.* **2008**, *9* (6), 536–542.

(51) Sankar, K.; Mishra, S. K.; Jernigan, R. L. Comparisons of Protein Dynamics from Experimental Structure Ensembles, Molecular Dynamics Ensembles, and Coarse-Grained Elastic Network Models. *J. Phys. Chem. B* **2018**, *122* (21), 5409–5417.