



## OPEN Exploring the unfolding pathways of protein families using Elastic Network Model

Ranjan Kumar & Sandipan Dutta

We explore how a protein's native structure determines its unfolding process. We examine how the local structural features, like shear, and the global structural properties, like the number of soft modes, change during unfolding. Simulations are performed using a Gaussian Network Model (GNM) with bond breaking for both thermal and force-induced unfolding scenarios. We find that unfolding starts in areas of high shear in the native structure and progressively spreads to the low shear regions. Interestingly, analysis of single domain protein families (Chymotrypsin inhibitor and Barnase) reveal that proteins with distinct unfolding pathways exhibit divergent behavior of the number of soft modes during unfolding. This suggests that the number of soft modes might be a valuable tool for understanding thermal unfolding pathways. Additionally, we found a strong link between a protein's overall structural similarity (TM-score) and its unfolding pathways, highlighting the importance of the native structure in determining how a protein unfolds.

The dynamic process of protein folding, in which a linear polypeptide chain transforms into a functional three-dimensional structure is an important phenomenon in molecular biology. The unfolding of the protein, on the other hand, typically involves the loss of the structure through the disruption of various stabilizing interactions, such as hydrogen bonds, hydrophobic interactions, and disulfide bonds<sup>1–7</sup>. This is a multi-step process involving formation of many intermediate states before it completely unfolds into the linear chain. It has been observed for small monomeric proteins that the folding and unfolding pathways share striking similarities, hence studying protein unfolding can provide insights into the folding pathways<sup>8,9</sup>. Moreover, the study of unfolding pathways can also lead to a better understanding of the progression of diseases associated with protein misfolding<sup>10</sup>.

In recent years, considerable progress has been made through theoretical and experimental approaches that provide insights into the factors leading to denaturation. Spectroscopic technique such as circular dichroism has been used to study the unfolding and folding of proteins as a function of temperature<sup>11</sup>. Fluorescence Spectroscopy<sup>12</sup>, single-molecule techniques<sup>3,13–16</sup>, small angle X-ray scattering<sup>17</sup> and Nuclear Magnetic Resonance<sup>18</sup>, have contributed valuable insights into protein folding/unfolding dynamics. Theoretical investigations have predominantly relied on Molecular Dynamics (MD) and Monte Carlo (MC) simulations to study unfolding phenomena<sup>19,20</sup>. Probing the intricate nature of protein folding and unfolding both experimental and theoretical investigations have consistently revealed the wealth of information concealed within the native-state topology. The native structure plays a pivotal role in determining the speed and mechanisms of protein folding<sup>21–23</sup>. Proteins with similar folds but distinct sequences often exhibit comparable folding rates<sup>23</sup>. To simulate protein unfolding, researchers have devised a method that involves systematically breaking hydrogen bonds and salt bridges within the native structure, taking into consideration their relative energies. In the pursuit of unraveling protein unfolding, a graph-based approach has proven valuable for identifying rigid and flexible regions within proteins, shedding light on the dynamic changes in flexibility during the process<sup>24,25</sup>. This collective body of work strongly underscores the critical role of the native-state topology in governing protein folding and unfolding mechanisms.

In this work we delve into the structural changes proteins undergo during unfolding, focusing on both local and global transitions. While some previous research have explored local rigidity changes during unfolding<sup>26</sup>, a complete understanding of the interplay between the local and the global structural changes remains elusive. We introduce the shear part of the strain<sup>27–29</sup>, a readily calculable parameter, to identify “floppy” regions within the protein structure. These floppy regions are more prone to unfolding and could be potential drug targets in diseases associated with protein misfolding. Proteins belongs to a larger classes of materials called amorphous materials and the soft vibrational modes can be used to identify the soft regions in amorphous materials that are prone to rearrangement during shear deformation<sup>30–34</sup>. They also play a crucial role during glass transition<sup>35–38</sup>

Department of Physics, Birla Institute of Technology and Science, Pilani, Rajasthan 333031, India. email: sandipan.dutta@pilani.bits-pilani.ac.in

and the shear localization which can lead to global structural failure in the system<sup>39–42</sup>. Rader et. al.<sup>24</sup> showed that the global structural property, the number of floppy modes, can capture the rigidity transition in proteins during unfolding. They found a sharp first-order-like transition from rigid to floppy regime. Here we show that the number of floppy modes contains much more information than just the rigidity to floppy transition. Our unfolding simulations of the members of protein families show that the unfolding is a gradual process occurring through intermediate unfolding pathways and that the number of soft modes can effectively differentiate the unfolding pathways for various proteins. Our simulations also establish a strong connection between the native structure of a protein and its unfolding pathways.

In our work we simulate unfolding through a bond breaking algorithm based on GNM developed by Su et. al.<sup>43</sup>. GNM belong to a larger class of models called Elastic Network Models (ENM)<sup>43–57</sup> that are useful in the study of the elasticity properties of proteins. These models along with Normal Mode Analysis successfully capture the large scale functional dynamics of proteins about their native structure. The residues are represented by their  $C_\alpha$  atoms, while the detailed interactions between residues in their native states are approximated using springs. In the GNM based algorithm, the bonds experiencing the highest fluctuations under the influence of thermal or external forces are broken leading to a gradual unfolding of the protein<sup>51,58</sup>. Remarkably, the simplified model based on GNM accurately capture the bond breaking sequence of atomistic MD simulations<sup>43,59,60</sup>. Our focus would be on the families of monomeric proteins CI2 and Barnase for which the GNM method has been well tested<sup>43</sup>.

## Methods

### Unfolding simulations

We perform unfolding simulations of proteins based on the GNM model of Su et. al.<sup>43</sup>. GNM is a coarse grained model where the protein is represented as a bead-spring network with residues as beads. The  $\alpha$ -Carbon atoms of the residues within a certain cutoff distance  $r_c = 7 \text{ \AA}$  are connected by springs of the same spring constant  $\kappa$ . The covalently bonded pairs along the chain backbone have a spring constant  $c\kappa$ . The parameters  $c$  and  $\kappa$  for the protein are obtained by fitting the B-factors with the experiments. Given that the springs are harmonic and the residue fluctuations are isotropic and Gaussian, the overall conformational potential of the structure can be written as

$$V = \frac{\kappa}{2} \sum_{i,j}^N [\Delta x_i \Gamma_{ij} \Delta x_j + \Delta y_i \Gamma_{ij} \Delta y_j + \Delta z_i \Gamma_{ij} \Delta z_j], \quad (1)$$

where  $\Delta x_i = x_i - x_i^0$  is the fluctuations of the  $x$ -coordinate of the  $i$ th residue about its equilibrium position  $x_i^0$ .  $N$  is the number of residues in the protein;  $\Gamma_{ij}$  are the elements of the Kirchhoff matrix also called connectivity in graph theory and in our case takes the values

$$\Gamma_{ij} = \begin{cases} -c & \text{if } |i - j| = 1 \\ -1 & \text{if } |i - j| > 1 \text{ and } r_{ij} < r_c \\ 0 & \text{if } |i - j| > 1 \text{ and } r_{ij} > r_c \\ -\sum_{i,j \neq i} \Gamma_{ij} & \text{if } i = j. \end{cases} \quad (2)$$

$r_{ij}$  is separation between the  $C_\alpha$  atoms of the  $i$ th and  $j$ th residues and  $r_c$  is cutoff distance. We use the same values of  $c = 9.3$  and  $\kappa = 0.493k_B T / \text{\AA}^2$  for all the proteins of the 2CI2 family, and  $c = 2.3$  and  $\kappa = 0.945k_B T / \text{\AA}^2$  for all proteins of the Barnase family in our work<sup>43</sup>.

The probability distribution of fluctuations  $\Delta \mathbf{R}_i = [\Delta x_i, \Delta y_i, \Delta z_i]$  about the native structure are Gaussian  $p(\{\mathbf{R}_i\}) \propto \exp(-\frac{\kappa}{2k_B T} V)$  and isotropic along  $x$ ,  $y$  and  $z$  directions.  $k_B$  and  $T$  are the Boltzmann constant and the temperature of the water environment. Using this, the cross-correlations fluctuation between  $i$ th and  $j$ th residue can then be shown to be

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle = \frac{3k_B T}{\kappa} [\Gamma^{-1}]_{ij}, \quad (3)$$

where  $\Gamma^{-1}$  is the inverse of  $\Gamma$ . From this we obtain the mean-square fluctuation in the distance vector  $R_{ij}$  between the residues  $i$ th and  $j$ th to be

$$\langle (\Delta R_{ij})^2 \rangle = \langle (\mathbf{R}_{ij} - \mathbf{R}_{ij}^0)^2 \rangle = \frac{3k_B T}{\kappa} ([\Gamma^{-1}]_{ii} + [\Gamma^{-1}]_{jj} - 2[\Gamma^{-1}]_{ij}). \quad (4)$$

$\mathbf{R}_{ij}$  and  $\mathbf{R}_{ij}^0$  are the instantaneous and equilibrium separation vectors between  $i$ th and  $j$ th residues.

For thermal unfolding, the temperature is increased and the bonds in the protein which undergo large fluctuations start breaking. All the forces are then redistributed among the other surviving bonds altering the fluctuations of the remaining bonds. Thus the unfolding of proteins is a highly nonlinear process. The following algorithm uses GNM for thermal unfolding of proteins

1. The mean square fluctuations between all bonds connecting the  $C_\alpha$  atoms from matrix  $\Gamma$  in Eq (4).

2. The bond with the largest fluctuation between residues  $i$  and  $j$  is broken, and matrix  $\Gamma$ , which represents a new topology during protein unfolding, is recalculated:  $\Gamma_{ij} \rightarrow \Gamma_{ij} + 1$ ,  $\Gamma_{ii} \rightarrow \Gamma_{ii} - 1$  and  $\Gamma_{jj} \rightarrow \Gamma_{jj} - 1$ .
3. The mean square fluctuations of the distance between all residue pairs are calculated using the new matrix  $\Gamma$ .
4. The above steps are repeated until all non-backbone bonds are broken. To study force-induced unfolding, we consider a protein being acted on by an external force  $f$ . Within the GNM we can obtain the  $x$  component of the displacements  $\mathbf{u}$  of the residues from the Hooke law

$$\mathbf{u}_x = \kappa^{-1} \Gamma^{-1} \mathbf{F}_x. \quad (5)$$

The  $i$ th element of the vectors  $\mathbf{u}_x$  and  $\mathbf{F}_x$  store the  $x$  component of the displacement and the force on  $i$ th residue respectively. Similar expressions for the Hooke law hold for the  $y$  and  $z$  components. Here we apply unit force along the  $x$  direction at the end residues,  $\mathbf{F}_x(1) = -1$  and  $\mathbf{F}_x(N) = 1$ .

We use the following algorithm for force-induced bond breaking

1. First the displacements of the residues are calculated from Eq (5) for a given applied force.
2. The bond with the largest stretching  $\Delta u_{ij} = |\mathbf{u}_i - \mathbf{u}_j|$  is broken and the connectivity matrix  $\Gamma$  is updated:  $\Gamma_{ij} \rightarrow \Gamma_{ij} + 1$ ,  $\Gamma_{ii} \rightarrow \Gamma_{ii} - 1$  and  $\Gamma_{jj} \rightarrow \Gamma_{jj} - 1$ .
3. The above steps are repeated until all the non-backbone bonds are broken. In this algorithm the network progressively becomes more flexible and the bond breaking is accelerated as they get stretched more under the fixed external force.

For all members of CI2 (Barnase) family, we stop both the thermal and the force-induced simulations after 110 (200) bonds are broken.

### Soft modes

Anisotropic Network Model (ANM) is another kind of ENM<sup>57,61</sup> based on spring network similar to GNM. The potential of this network is  $\frac{\kappa}{2} \sum_{j < i} \Gamma_{ij} (R_{ij} - R_{ij}^0)^2$ . The Hessian  $H$  of this network is the second derivative of this potential with respect to the position of the residues. The Hessian matrix  $H$  is defined in terms of  $3 \times 3$  blocks

$$H_{ij} = \kappa \Gamma_{ij} \mathbf{n}_{ij} \mathbf{n}_{ij}^T, \quad (6)$$

for  $i \neq j$ , where  $\mathbf{n}_{ij} = \mathbf{R}_{ij}^0 / R_{ij}^0$  is the unit vector along the bond joining the  $C_\alpha$  atoms of  $i$ th and  $j$ th residues. The diagonal blocks of  $H$  are calculated as  $H_{ii} = -\sum_{j \neq i} H_{ij}$ .

The eigenvalues  $\lambda_i$  of the Hessian matrix  $H$  of ANM are positive or zero. The eigenvectors corresponding to the zero eigenvalues are called the floppy modes. In this work, we consider an eigenmode floppy if its eigenvalue is less than  $10^{-4}$ . The translation and rotational symmetries of the matrix  $H$  always result in six zero eigenvalues, while the presence of more floppy modes signifies that the protein is under-constrained.

### Calculation of shear

Consider a protein with the coordinates of residues in the crystal structure  $\mathbf{x}$  and of the deformed state  $\mathbf{x}' = \chi(\mathbf{x})$ . The local deformation of the coordinates  $\mathbf{x}$  is given by  $\mathbf{F} = \frac{\partial \mathbf{x}'}{\partial \mathbf{x}}$ . In the discrete form, the deformation matrix depends on the un-deformed relative positions of the  $m$ th residue and its neighboring  $n$ th residues  $\Delta \mathbf{x}^{mn} = \mathbf{x}^n - \mathbf{x}^m$ , the deformed positions  $\Delta \mathbf{x}'^{mn} = \mathbf{x}'^n - \mathbf{x}'^m$  and a weight function  $w_n$ . The results have no strong dependence on the weight functions provided the neighborhoods are chosen large enough. Here we choose a radial weight function

$$w_{mn} = \begin{cases} 1 & r_{mn} \leq 6 \text{ \AA} \\ 1 - \frac{1}{2}(r_{mn} - 6) & 6 \text{ \AA} < r_{mn} \leq 8 \text{ \AA} \\ 0 & r_{mn} \geq 8 \text{ \AA}, \end{cases} \quad (7)$$

where  $r_{mn} = |\Delta \mathbf{x}^{mn}|$ . It has the expression  $\mathbf{F}_m = \mathbf{A}_m \mathbf{D}_m^{-1}$ , where  $\mathbf{A}_m = \sum_n \Delta \mathbf{x}^{mn} \Delta \mathbf{x}^{mnT} w_{mn}$  and  $\mathbf{D}_m = \sum_n \Delta \mathbf{x}'^{mn} \Delta \mathbf{x}'^{mnT} w_{mn}$ . The Eulerian strain tensor is calculated as  $\epsilon_m = \frac{1}{2} [\mathbf{I} - (\mathbf{F}_m \mathbf{F}_m^T)^{-1}]$ <sup>27,62</sup>, where  $\mathbf{I}$  is the identity  $3 \times 3$  matrix. The strain tensor is then  $\gamma_m = \epsilon_m - \frac{1}{3} (\text{Tr} \epsilon_m) \mathbf{I}$ . The shear component of the elastic energy of deformation at the residue  $m$  is  $s_m = \sum_{i,j}^3 (\gamma_m)_{ij}^2$ . In our work we calculate the total shear  $s_m$  at residue  $m$  by summing over the contribution of the shear due to each soft modes i.e. choosing  $\mathbf{u}$  as a soft eigenmode calculated in Sec 2.2. If a fraction  $f$  of the  $3N$  modes are softmodes, the total shear is obtained by summing over these  $3fN$  softmodes. As the protein becomes more floppy during unfolding, the fraction  $f$  becomes larger.

### Correlation matrices

To determine whether two proteins unfold along similar pathways, we compare the sequence in which the bonds between their secondary structures break. For CI2 with an  $\alpha$ -helix, three  $\beta$ -strands and two loops, we obtain  $N_r = 5$  regions by cluster analysis of the contact map for the CI2 protein family corresponding to the interactions between the different secondary structures. To obtain a quantitative description of the bond breaking sequence,

we define  $N_r$  dimensional vectors  $v_r$  which has 1 at one position and zero everywhere else. For example, for CI2 we use  $v_1 = [1, 0, 0, 0, 0]$  for N-terminus- $\beta_3$ ,  $v_2 = [0, 1, 0, 0, 0]$  for  $\alpha$ -Loop<sub>2</sub>,  $v_3 = [0, 0, 1, 0, 0]$  for  $\beta_1$ - $\beta_2$ ,  $v_4 = [0, 0, 0, 1, 0]$  for  $\beta_2$ - $\beta_3$  and  $v_5 = [0, 0, 0, 0, 1]$  for the rest of the non-covalent bonds. The individual bond breaking events are categorized into region-wise vector representation. For each protein, the  $N_b$  bond breaking sequence is transformed into  $N_b \times N_r$  which is flattened to a  $N_b N_r \times 1$  vector. For  $N_p$  proteins in the family we obtain a matrix  $N_p \times N_b N_r$  from which the covariance matrix is calculated. For the Barnase family we get  $N_r = 30$  regions in the contact map based on the interactions between the N-terminus, the three  $\alpha$ -helices, the three  $\beta$ -strands and the two loops.

## Results and discussion

### Shear evolution and floppy modes analysis - local and global parameters

The weakest regions of a protein, that are more susceptible to unfolding, can be obtained from the calculation of strain<sup>27</sup>, which measures the local deformation in an elastic medium from the spatial gradient of displacement<sup>62</sup>. If the coordinates of an elastic medium is deformed by  $\mathbf{x} \rightarrow \mathbf{x}' = \mathbf{x} + \mathbf{u}$ , strain is calculated as

$$\varepsilon_{ij} = \frac{1}{2} \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} - \sum_k \frac{\partial u_k}{\partial x_i} \frac{\partial u_k}{\partial x_j} \right), \quad (8)$$

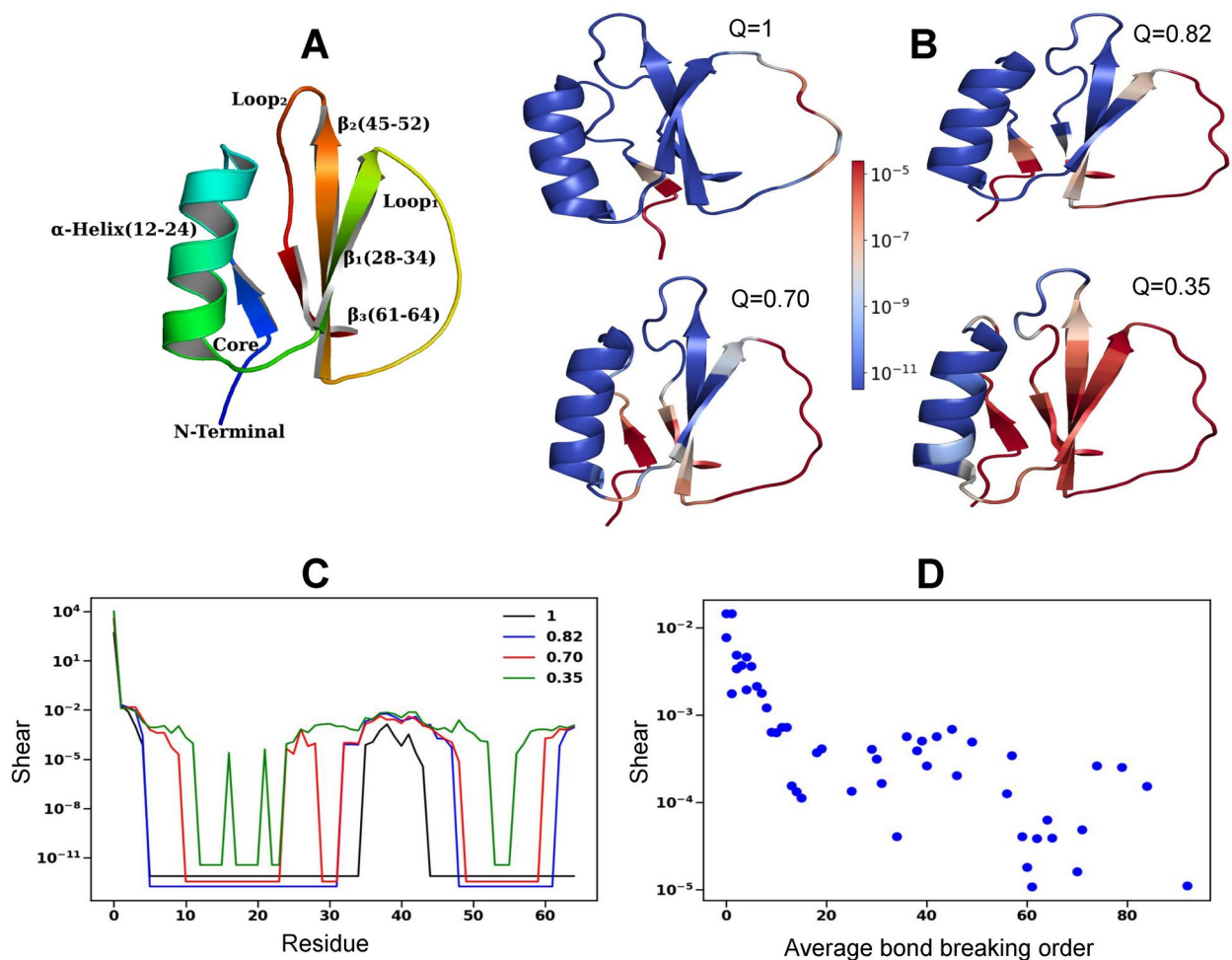
where  $u_i$  is the  $i^{\text{th}}$  component of the displacement vector  $\mathbf{u}$ . The strain tensor is the second rank tensor (similar to  $3 \times 3$  matrix) which has six distinct components that describe entirely the local stretching or compression, twisting, and shear phenomena occurring at a specific location within the structure. The shear tensor is obtained from the strain tensor using

$$\gamma = \varepsilon - \frac{1}{3}(\text{Tr}\varepsilon) \mathbb{I}, \quad (9)$$

where  $\mathbb{I}$  is the  $3 \times 3$  identity matrix. The shear energy is the sum of all elements of the shear matrix  $\sum_{i,j} \gamma_{ij}$  (Sec 2.3). The displacements from the floppy modes are calculated in Sec 2.2 and are used to obtain the total shear energy (Sec 2.3). From now on we call the total shear energy as the shear. The most flexible regions of the protein are regions with high shear.

The native structure of the extensively studied Chymotrypsin inhibitor-2 (CI2) (PDB code 2ci2), consists of an  $\alpha$ -helix and three-strand  $\beta$ -sheet as shown in Fig 1-A. In the native structure, the shear peaks at the hinge site (at residue 38 in the black curve in Fig 1-C in the native state) as well as at the N-terminus and the C-terminus and can be used to identify the weakest regions of the proteins. At these regions, there are low density of bonds. As the protein unfolds the  $Q$  value, the fraction of native contacts still intact, decreases and the protein loses rigidity. The shear which initially peaks at the few flexible residues, gradually becomes delocalized over the entire protein. In the native state ( $Q = 1$  in Fig 1-B), the shear peaks at N-terminus and Loop<sub>1</sub> with N-terminus becoming floppy. When 30 bonds are broken we find the shear at N-terminus,  $\beta_2$  and  $\beta_3$  increases ( $Q = 0.82$  in Fig 1-B) suggesting breaking of the bonds between N-terminus- $\beta_3$  and  $\beta_2$ - $\beta_3$ . Later ( $Q = 0.7$  in Fig 1-B) both  $\beta_1$  and  $\beta_2$  become floppy suggesting their loss of contact. All throughout the bonds of  $\alpha$ -helix remain stable. Fig 1-C shows that the thermal unfolding initiates at these weakest region of the protein like nucleation, then gradually it spreads to regions next to them and finally the shear is delocalized over most of the protein signaling that the entire protein becomes floppy. We explore if the shear in the native structure ( $Q = 1$ ) can be used to predict the bond breaking sequence during unfolding. For each residue, we record the order in which the bonds of the residue with its neighbors are broken during unfolding. For each residue we take the average of the bond breaking order and find that there is a strong negative Pearson correlation of -0.79 between the logarithm of the shear in the native structure and the average bond breaking order in Fig 1-D, implying that the regions of high shear are more prone to bond breaking and that shear can be used to predict the bond breaking sequence or high shear regions have low average bond breaking order. The unfolding process involves the disruption of bonds, leading to the reduction in the average coordination number  $\langle z \rangle$  per residue and the increase in floppy modes for thermal unfolding shown in red in Fig 2-A. Force-induced unfolding is shown in blue and is detailed in the next paragraph. For both cases, in instances of high  $\langle z \rangle$ , there are flat regions where alterations in  $\langle z \rangle$  do not affect the number of floppy modes. Initially, the protein is overly constrained due to an abundance of redundant bonds, and breaking these bonds does not generate new floppy modes. At later stages of unfolding  $\langle z \rangle$  less than 3, not many redundant bonds are left which lead to rapid increase of floppy modes. The changes in the number of floppy modes are gradual as compared to the sudden jump during the rigidity transition as observed in Ref<sup>24</sup>.

Alternatively, we unfold CI2 by pulling at their ends along x-axis direction in opposite directions as detailed in Sec 2.1. A question naturally arises whether a pulling force can trigger similar unfolding pathways as thermal unfolding. Pulling the residues at the N-terminus and C-terminus of 2CI2 causes uneven tension in the protein resulting in the bonds near both the termini becoming floppy ( $Q = 0.82$  in Fig 2-B). Gradually the bonds between the N-terminus region with the  $\beta_3$  and the bonds between the  $\alpha$ -helix and Loop<sub>1</sub> get ruptured ( $Q = 0.35$  in Fig 2-B). The bonds within  $\alpha$ -helix and  $\beta_1$  with its neighbor break making them flexible. The initial unfolding pathway is similar to the pathway observed during thermal unfolding. However, we find from the comparison of the shear in Fig 1-B and Fig 2-B that in thermal unfolding the protein becomes floppy from the hinge and in the force-induced unfolding from the N-terminus. This is a consequence of the fact that the  $\alpha$ -helix near the N-terminus unfolds first due to the proximity of the point of application of the force leading to unfolding predominantly from the N-terminus side. Our shear analysis shows that the unfolding pathways are different. These differences



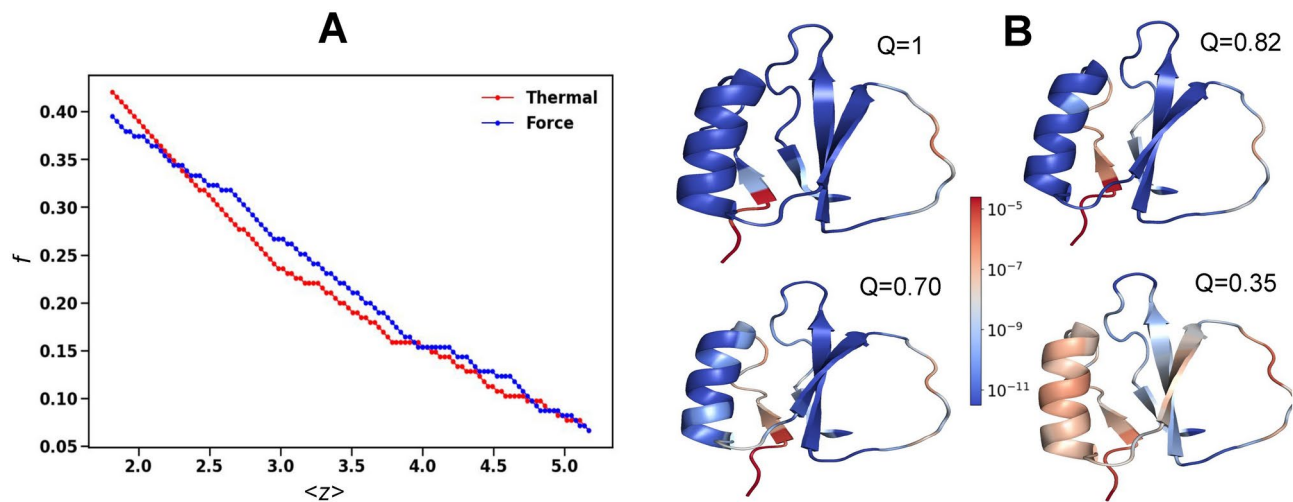
**Fig. 1.** Thermal unfolding of CI2. (A) The secondary structures in the native state of CI2 (PDB ID: 2ci2). (B) Shear for different  $Q$  values, which is the fraction of native contacts, at the different stages of unfolding. Regions of high (low) shear are shown in red (blue). (C) The shear at each residue at different  $Q$  values. (D) The scatter-plot of the shear at the residues in the native structure ( $Q = 1$  in (B)) vs corresponding the average order of bonds breaking. The average bond-breaking order for each residue was calculated by averaging the order at which its individual bonds to neighboring residues are broken during a simulation of 110 bond ruptures within the protein.

also show up in the behavior of the number of floppy modes with the coordination number in Fig 2-A. For force-induced unfolding also, we find a high correlation of 0.68 between log shear and the average bond breaking order (not shown). But the correlation is lower than that of the thermal unfolding.

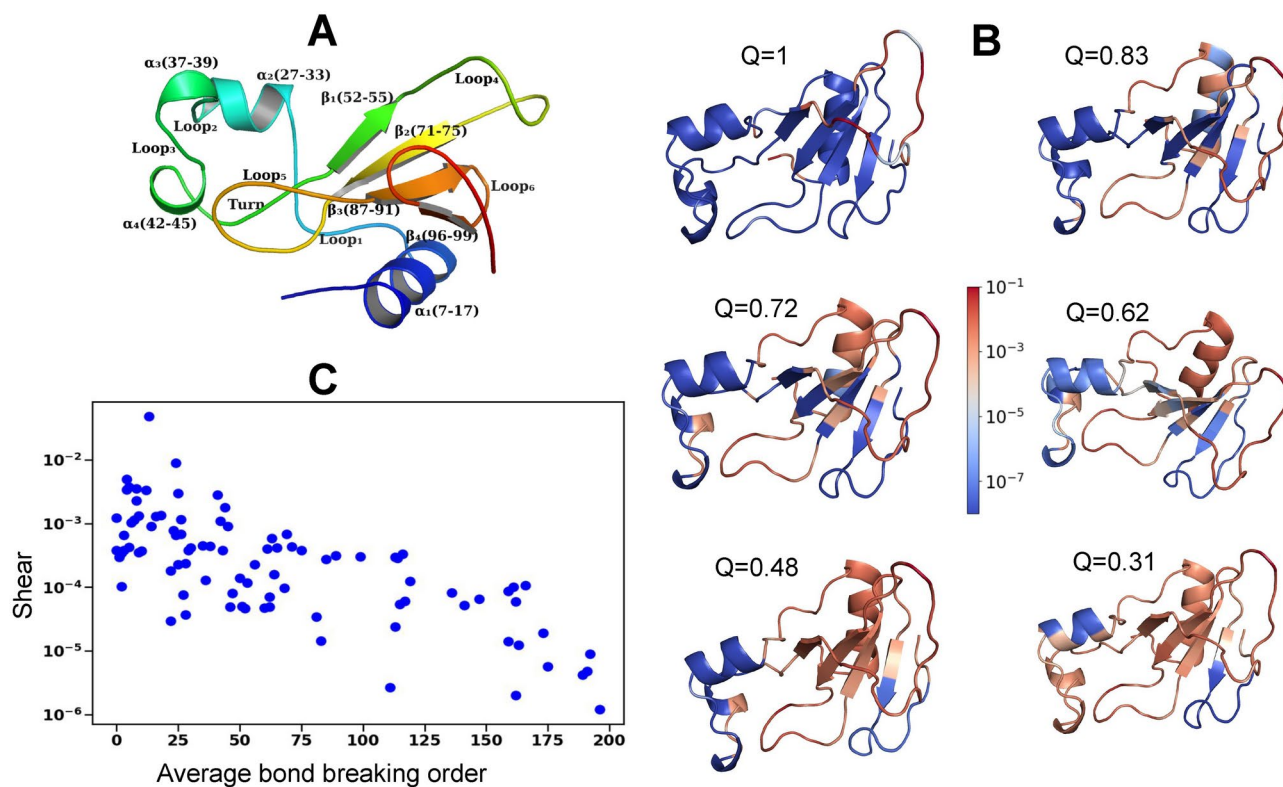
We study the thermal unfolding and the corresponding change in shear for another protein, Barnase (PDB code 1a2p) is a ribonuclease with four  $\alpha$ -helices and a four-stranded  $\beta$ -sheet (Fig 3-A) forming three hydrophobic cores. Fig 3-B shows that in native state  $Loop_4$  ( $Q = 1$ ) and N-terminus have high shear and are floppy. Next  $Loop_1$ ,  $Loop_4$ ,  $Loop_5$  and  $Loop_6$  all become flexible at  $Q = 0.83$  in Fig 3-B and further at  $Q = 0.72$  and  $Q = 0.62$ ,  $\alpha_1$ -helix opens up. At  $Q = 0.48$  and  $Q = 0.31$  the remaining helices,  $\beta_1$ ,  $\beta_2$  and  $\beta_3$  and at the end  $\beta_4$  become flexible resulting in the unfolding. We again find a strong correlation of 0.72 between the logarithm of the shear of the residues in the native state of Barnase and the average order at which the bonds with their neighbors are broken in Fig 3-C. Our results suggest that shear can be used to predict the bond breaking sequence for thermal unfolding but less so for the force-induced unfolding.

### Results on CI2 thermal unfolding

We simulate the unfolding pathways of 36 proteins of CI2 protein family with available PDB structures obtained from the PFAM database. Based on a coarse grained correlation (Section 2.4) we compare the pathways for the member proteins and quantify the relationship between their native structures, the unfolding pathways and the number of soft modes. Fig 4 shows the sequence alignment of the members of the CI2 family with coloring based on the secondary structure. There are three types of proteins with similar secondary structures except with a 3-10 helix, a  $\beta$ -strand or lacking both at the N-terminus. Helices and  $\beta$ -strands that are part of the hydrophobic core are mostly conserved. We apply the thermal unfolding algorithm outlined in Section 2.1 and we use a vector representation of the bond breaking sequence (more details in Sec 2.4) corresponding to the five clusters of bonds



**Fig. 2.** Force-induced unfolding of CI2. (A) The number of floppy modes,  $f$ , as a function of the average coordination number of the residues,  $\langle z \rangle$  for thermal and force unfolding of CI2. (B) The shear at each residue during unfolding for different Q values. Regions of high (low) shear are shown in red (blue).



**Fig. 3.** Thermal unfolding of Barnase. (A) The native structure of Barnase (PDB ID : 1a2p) showing the various secondary structures. (B) The shear at each residue during unfolding with regions of high (low) shear are shown in red (blue). (C) The shear for the residues in the native structure ( $Q = 1$ ) vs the average order in which the bonds break for Barnase in a total of 200 bond breaking.

in the contact map of CI2 in Fig 1-A. The covariance matrix, whose  $(i, j)$  th element quantifies the correlations of the bond breaking sequence between protein  $i$  and  $j$ , is calculated using the method in Sec 2.4 and is plotted in Fig 5-A. This is superimposed with a measure of structural similarity between two proteins, the TM-score described in Ref<sup>63</sup> in Fig 5-A. TM-score is a measure of topological similarity of two proteins. It takes values between 0 and 1, with 1 indicating that the two structures perfectly match. The contour plot of the TM-score and the correlations between the unfolding pathways exhibit a qualitative resemblance. Five proteins 1tin, 1mit, 3ci2,

```

2ci2      SSVEKKPEGVNTGAGDRHNLKTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVTMEYRIDRVRLFVDKLDNIA-EVPRVG
2sni      SSVEKKPEGVNTGAGDRHNLKTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVTMEYRIDRVRLFVDKLDNIA-EVPRVG
5ffn      TGAGDRHNLKTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVTMEYRIDRVRLFVDKLDNIA-QVPRVG
3ci2      HNLKTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVTMEYRIDRVRLFVDKLDNIA-QVPRVG
6qiy      DLKTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVTMEYRIDRVRLFVDKLDNIA-EVPRVG
1coa      MKTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVTMEYRIDRVRLFVDKLDNVA-EVPRVG
7alh      MKTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVTMEYRIDRVRLFVDKLDNIA-QVPRVG
7aok      MKTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVTMEYRIDRVRLFVDKLDNIA-QVPRVG
7aon      MKTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVTMEYRIDRVRLFVDKLDNVA-QVPRVG
1tm1      MKTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVTMEYRIDRVRLFVDRLDNIA-QVPRVG
1ypc      MKTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVTMEYRIDRVRLFVDKLDNIA-QVPRVG
1y3b      KTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVTMSYRIDRVRLFVDRLDNIA-QVPRVG
1y34      MKTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVTMEYRIDRVRLFVDRLDNIA-QVPRVG
1ypa      MKTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVTMEYRIDRVRLFVDKLDNIA-QVPRVG
1y1k      MKTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVAMEYRIDRVRLFVDRLDNIA-QVPRVG
1tmg      MKTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVTMEYRIDRVRLFVDRLDNIA-QVPRVG
1y3d      MKTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVTMEYRIDRVRLFVDRLDNIA-QVPRVG
1y3c      MKTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVTMEYRIDRVRLFVDRLDNIA-QVPRVG
1ypb      MKTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVTMEYRIDRVRLFVDKLDNIA-QVPRVG
1y33      MKTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVPMEYRIDRVRLFVDRLDNIA-QVPRVG
1to2      KTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVTKEYRIDRVRLFVDRLDNIA-QVPRVG
1tm3      KTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVTKEYRIDRVRLFVDRLDNIA-QVPRVG
1y3f      KTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVTMEYRIDRVRLFVDRLDNIA-QVPRVG
1to1      KTEWPELVGKSVVEAKKVILODKPEAQIIVLPVGTIVTMEYRIDRVRLFVDRLDNIA-QVPRVG
1vbw      SRCQKSSWPQLVGSFGAAAKVIERENPRVRAVIKVGSGATKDFRCNRVRIWVNWKRGLVV-SPPRIG
1mit      GSSCPGKSSWPHLVGSGSVAKAIIERQNPVVKAVILEEGTPVTKDFRCNRVRIWVNWKRGLVV-SPPRIG
1tin      SSCPGKSSWPHLVGSGSVAKAIIERQNPVVKAVILEEGTPVTKDFRCNRVRIWVNWKRGLVV-SPPRIG
2sec      TEFGSELKSFPEVVGKTVDAQAREYFTLHYPQYDVYFLPEGSPVTLDLRYNRVRFVYFNPGTNIVVNHVPHVG
1acb      TEFGSELKSFPEVVGKTVDAQAREYFTLHYPQYDVYFLPEGSPVTLDLRYNRVRFVYFNPGTNIVVNHVPHVG
1sbn      TEFGSELKSFPEVVGKTVDAQAREYFTLHYPQYDVYFLPEGSPVTRDLRYNRVRFVYFNPGTNIVVNHVPHVG
1sib      TEFGSELKSFPEVVGKTVDAQAREYFTLHYPQYDVYFLPEGSPVTLDLRYNRVRFVYFNPGTNIVVNHVPHVG
4b2c      GTEFGSELKSFPEVVGKTVDAQAREYFTLHYPQYDVYFLPEGSPVTKDLRYNRVRFVYFNPGTNIVVNHVPHVG
4b2b      GTEFGSELKSFPEVVGKTVDAQAREYFTLHYPQYDVYFLPEGSPVTKDLRYNRVRFVYFNPGTNIVVNHVPHVG
4b1t      TEFGSELKSFPEVVGKTVDAQAREYFTLHYPQYDVYFLPEGSPVTKDLRYNRVRFVYFNPGTNIVVNHVPHVG
4b2a      SELKSFPEVVGKTVDAQAREYFTLHYPQYDVYFLPEGSPVTKDLRYNRVRFVYFNPGTNIVVNHVPHVG

```

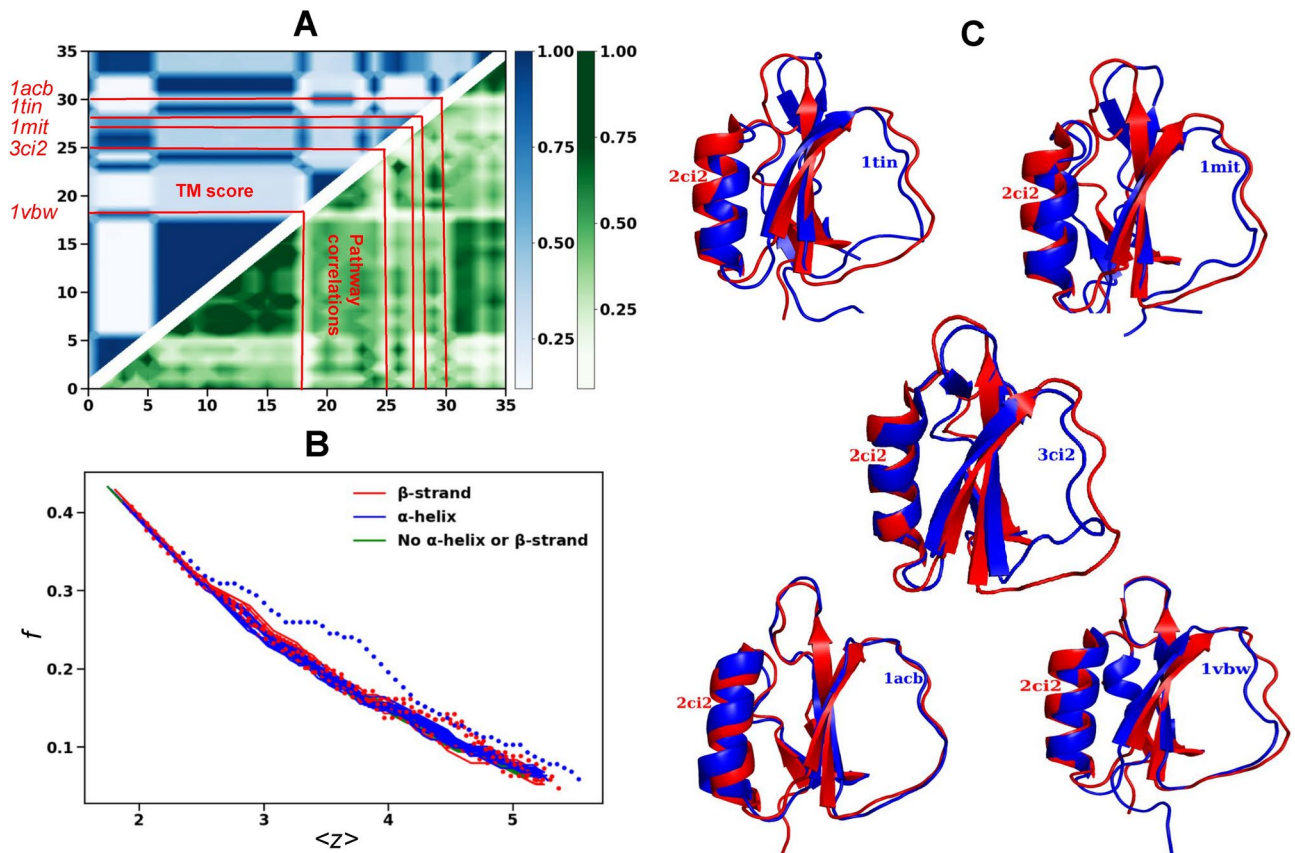
**Fig. 4.** CI2 family sequence. The sequence alignment and the color coding of the secondary structures of the 36 proteins of the CI2 family (PDB IDs labeled on the left) aligned using protein BLAST. The color red denotes  $\alpha$ -helix, blue denotes  $\beta$ -strands and black denotes loops. The gray regions are regions lacking any secondary structures (disordered regions).

1acb and 1vbw, with divergent floppy curves, not only have low TM-score but also have low correlations between the unfolding pathways implying that the proteins with different native structures will have different unfolding pathways. The well-established principle that the native structure of a protein governs its folding pathway<sup>64,65</sup> is confirmed by this finding. In Fig 5-B we find that the behavior of the number of floppy modes is most divergent initially when the coordination number is high, however after the coordination number drops below 3.7 their pathways become more convergent. The number of floppy modes depends on the entire protein and hence it is a global measure of the structural feature of the protein. The inflection point of the number of floppy modes has been used by Rader et. al.<sup>24</sup> to obtain a sharp first-order-like rigid to floppy transition. However, in our case Fig 5-B shows more of a gradual behavior for these curves. The proteins with most diverging floppy curves in Fig 5-B are same proteins in Fig 5-A with low TM-scores and low pathway correlations: 1tin, 1mit, 3ci2, 1acb and 1vbw. The proteins with more coinciding or similar (dissimilar) floppy modes curves in Fig 5-B, have more similar (dissimilar) unfolding pathways implying that the floppy mode curves can be used to characterize the order of bond breaking between the secondary structures of proteins that leads to their unfolding. All these results imply that the floppy curves can serve as a tool to differentiate the unfolding pathways of proteins particularly the ones that are structurally different.

We examine the structural characteristics of proteins to identify the factors leading to the most distinct pathways with minimal correlations in Fig 5-C. The superposition of the two proteins, 2ci2 and 3ci2 shows the structural differences between the two. 3ci2 has an extra  $\beta$ -strand in Loop<sub>2</sub> between  $\beta_2$  and  $\beta_3$  that causes strong interactions between  $\alpha$ -helix and Loop<sub>2</sub>, and  $\beta_2$ - $\beta_3$  resulting in the bonds to break later. 1vbw has smaller  $\beta_2$  and  $\beta_3$  and has an  $\alpha$ -helix at the N-terminus unlike 2ci2. The bonds of  $\alpha$ -helix and  $\beta_1$ - $\beta_2$  break first which is the reverse of 2ci2. 1acb has larger  $\beta_3$  and the N terminus- $\beta_3$  and  $\alpha$ -Loop<sub>2</sub> breaks earlier than  $\beta_2$ - $\beta_3$  unlike 2ci2. 1mit and 1tin has a much longer N-terminus than 2ci2 and the loop at this terminus has a very different orientation compared to the other proteins. The bond breaking order is similar to 3ci2 except that all the bonds in  $\beta_2$ - $\beta_3$  breaks in 1mit before  $\beta_1$ - $\beta_2$ .

### Results on Barnase thermal unfolding

In Fig 6-A we illustrate the TM-score vs the correlations between the thermal unfolding pathways of the 49 members of the Barnase family. Both the contour plots exhibit similar qualitative features. Proteins 1yvs and 1bnr (red lines) and proteins 2kf3, 2kf4, 2kf5 and 2kf6 (lighter band near 30) have significant differences in unfolding pathways as well as lower TM-score. The divergent behavior of these proteins, particularly for 1yvs and 1bnr (green dotted), also show up in the floppy curves in Fig 6-B. The native structures of these two proteins differ



**Fig. 5.** Thermal unfolding pathways for CI2 family. (A) TM-score measuring the structural similarity between two proteins is obtained from the optimized residue-to-residue alignment<sup>63</sup>. TM-score between each pair of 36 proteins of CI2 family is plotted in blue. Also shown in green are the correlations of the bond breaking sequences between each pair of member proteins. The trivial diagonal elements (TM-score / correlations of the unfolding pathways of a protein with itself) are not plotted. The first 23 proteins have a  $\alpha$ -helix at the N-terminus, next 10 proteins have  $\beta$ -strand at the N-terminus and the last 3 proteins have neither. (B) The behavior of the fraction of floppy modes  $f$  as a function of the coordination number  $\langle z \rangle$  for the family members for the three types of proteins. The floppy curves exhibiting the greatest degree of variation are represented by dotted lines. We have marked these proteins by red lines in (A). (C) The native structures of the proteins with distinct floppy mode curves (dotted) in (B).

significantly from the other members (see Fig 6-C and Fig 6-D). They have 3  $\alpha$ -helices making the hydrophobic core 2 weaker. Additionally we find that the hydrophobic core 1 is partially open in 1yvs with the  $\alpha_1$  and  $\alpha_2$  far from the rest of protein.

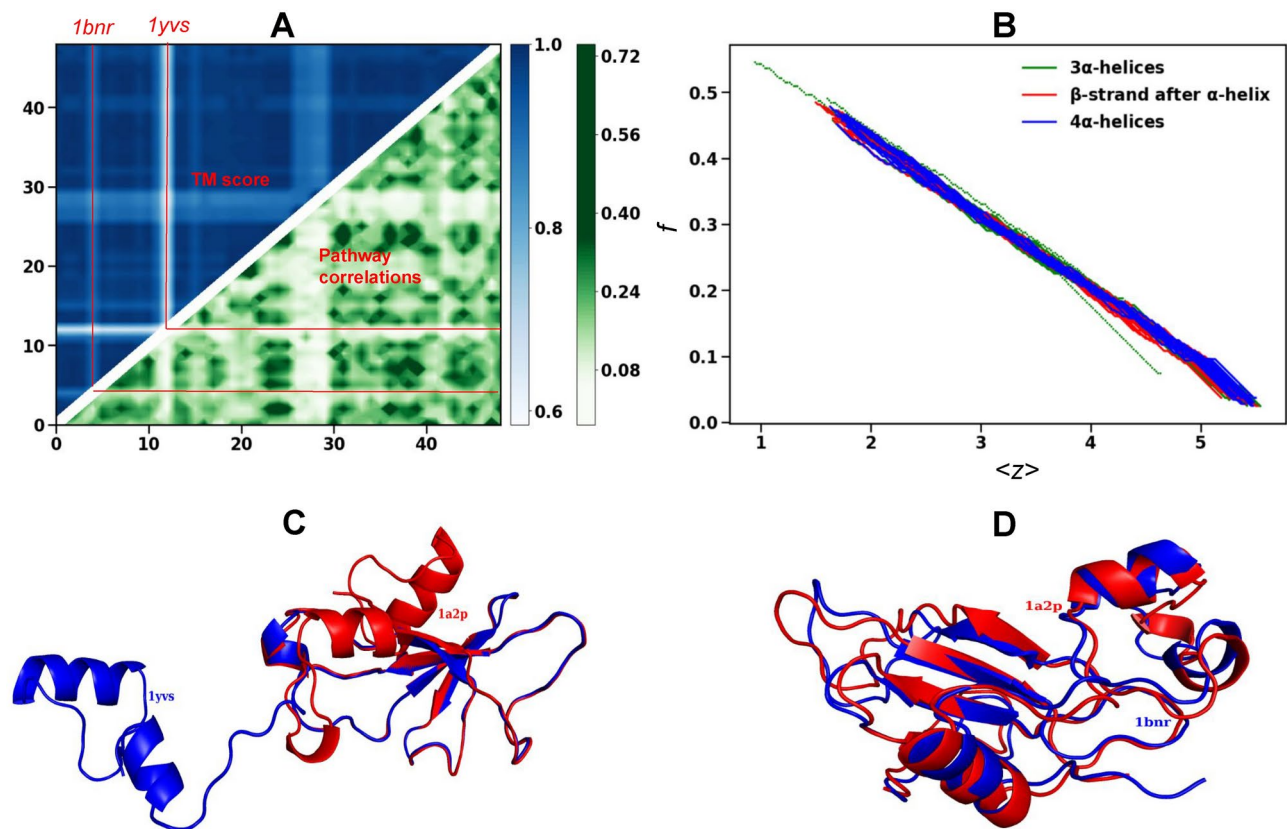
### Results on CI2 force-induced unfolding

For force-induced unfolding of CI2 family proteins by pulling at the ends, we observe a minimal resemblance between the TM-score and the correlation of unfolding pathways in Fig 7-A. In fact the plot shows high correlations between the unfolding pathways for almost all the proteins. However, the floppy mode curves in Fig 7-B show significant differences for proteins 1ypb, 1sbn, 1acb, 3ci2 and 1vbw. It is important to note that most of these proteins show different unfolding pathways even for thermal unfolding.

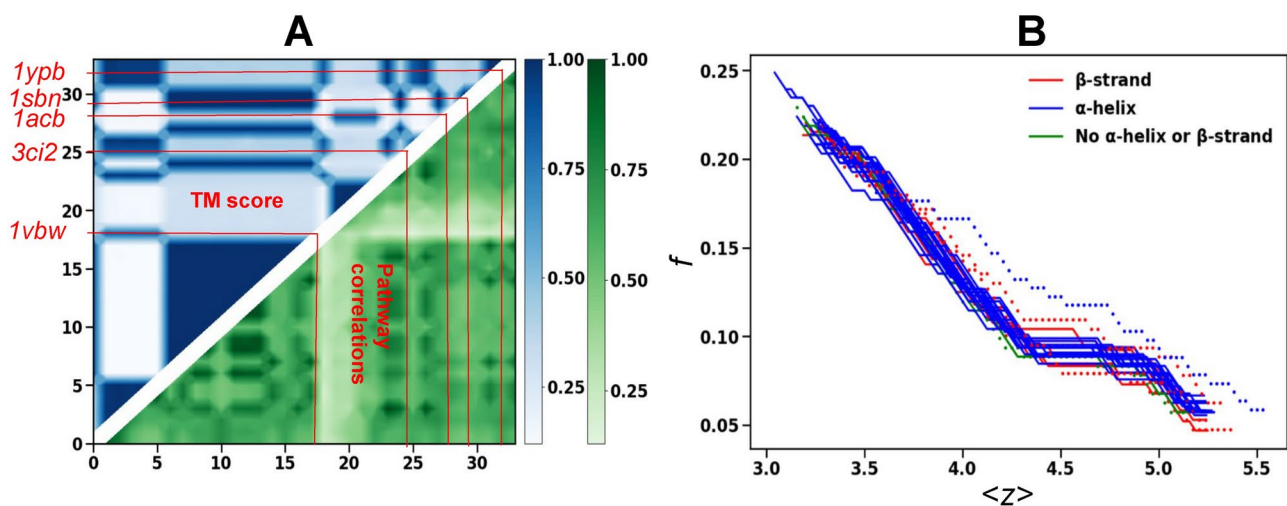
### Conclusion

We investigated the changes in the local and global structure of small single domain proteins during thermal and force-induced unfolding. Coarse-grained simulations based on the Gaussian Network Model<sup>43</sup> were used to obtain the unfolding pathways. The local structural features were captured using the shear part of the strain which peaks at the flexible regions of the protein. The high shear regions occur at the ends and at the hinge site of the proteins. We found strong correlations between the high shear and the order of bond breaking which suggests that the bond breaking starts at the high shear regions and spreads to regions of low shear until the entire protein becomes floppy.

We simulated the unfolding pathways for the members of CI2 and Barnase families. For thermal unfolding of both CI2 and families, the unfolding pathways have qualitative agreement with the TM-score implying that the structural dissimilarity lead to very different unfolding pathways. However for force-induced unfolding for CI2



**Fig. 6.** Thermal unfolding pathways for Barnase family. (A) TM-score between 49 proteins of Barnase family superimposed with the correlation between the unfolding pathways. The first 14 proteins have three  $\alpha$ -helices, next 19 proteins have  $\beta$ -strand after  $\alpha$ -helix and the last 16 proteins have four  $\alpha$ -helices. 1yvs and 1bnr have least correlated pathways with low TM-scores. (B) The floppy modes  $f$  as a function of  $\langle z \rangle$  show largest variation for 1yvs and 1bnr (green dotted). (D) The structures of 1yvs and 1bnr superimposed with Barnase (1a2p).



**Fig. 7.** Force induced pathways for CI2. The TM-score between pair of proteins in CI2 family superimposed with the correlation between the unfolding pathways. The order of proteins are same as Fig 1-A. The most divergent floppy mode curves as a function of  $\langle z \rangle$  are represented by dotted lines in (B) and by red lines in (A).

this agreement is very weak. The force-induced unfolding pathways are very similar for almost all the proteins in the family even for low TM-scores.

Additionally, we explored how the global structural parameter, the number of floppy modes, behaves during unfolding. The number of floppy modes increases with unfolding and we found that the proteins which have similar unfolding pathways have converging behavior of these floppy mode curves, while the proteins with very different unfolding pathways have divergent floppy mode curves. This implies that the number of floppy modes serves as a tool to characterize variations in protein unfolding pathways. Analysis of the proteins' 3D structures reveals that divergent floppy mode behavior stems more from the differences in their tertiary structures than their secondary structures.

Overall we find two tools: 1. the shear that contains the local structural information about the protein and 2. the number of floppy modes that contains the global structural information about the protein. The interplay of these two will be useful to study the structural transitions in biomolecules.

## Data availability

All data generated or analyzed during this study are included in this published article.

Received: 19 April 2024; Accepted: 4 October 2024

Published online: 13 October 2024

## References

- Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem* **48**, 545–600 (1997).
- Lu, H., Israilewitz, B., Krammer, A., Vogel, V. & Schulten, K. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys J* **75**, 662–671 (1998).
- Habibi, M., Rottler, J. & Plotkin, S. S. As simple as possible, but not simpler: exploring the fidelity of coarse-grained protein models for simulated force spectroscopy. *PLoS Comput Biol* **12**, e1005211 (2016).
- Kouza, M., Hu, C.-K., Li, M. S. & Kolinski, A. A structure-based model fails to probe the mechanical unfolding pathways of the titin i27 domain. *J. Chem. Phys.* **139** (2013).
- Sotomayor, M. & Schulten, K. Single-molecule experiments in vitro and in silico. *Science* **316**, 1144–1148 (2007).
- Brockwell, D. J. et al. Pulling geometry defines the mechanical resistance of a  $\beta$ -sheet protein. *Nat Struct Mol Biol* **10**, 731–737 (2003).
- Carrion-Vazquez, M. et al. The mechanical stability of ubiquitin is linkage dependent. *NNat Struct Mol Biol* **10**, 738–743 (2003).
- Daura, X., Jaun, B., Seebach, D., Van Gunsteren, W. F. & Mark, A. E. Reversible peptide folding in solution by molecular dynamics simulation. *J Mol Biol* **280**, 925–932 (1998).
- Shea, J.-E. & Brooks, C. L. III. From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu Rev Phys Chem* **52**, 499–535 (2001).
- Zhuravlev, P. I., Reddy, G., Straub, J. E. & Thirumalai, D. Propensity to form amyloid fibrils is encoded as excitations in the free energy landscape of monomeric proteins. *J Mol Biol* **426**, 2653–2666 (2014).
- Greenfield, N. J. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nature protocols* **1**, 2527–2535 (2006).
- Eftink, M. R. The use of fluorescence methods to monitor unfolding transitions in proteins. *Biophys J* **66**, 482–501 (1994).
- Schuler, B. & Hofmann, H. Single-molecule spectroscopy of protein folding dynamics-expanding scope and timescales. *Curr Opin Struct Biol* **23**, 36–47 (2013).
- Schuler, B. & Eaton, W. A. Protein folding studied by single-molecule fret. *Curr Opin Struct Biol* **18**, 16–26 (2008).
- Mojumdar, S. S. et al. Multiple intermediates in the folding of superoxide dismutase 1 revealed by single molecule force spectroscopy. *Biophys J* **110**, 497a (2016).
- Habibi, M., Rottler, J. & Plotkin, S. S. The unfolding mechanism of monomeric mutant sod1 by simulated force spectroscopy. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1865**, 1631–1642 (2017).
- Lipfert, J. & Doniach, S. Small-angle x-ray scattering from rna, proteins, and protein complexes. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 307–327 (2007).
- Dyson, H. J. & Wright, P. E. Unfolded proteins and protein folding studied by nmr. *Chem Rev* **104**, 3607–3622 (2004).
- Sułkowska, J. I., Kloczkowski, A., Sen, T. Z., Cieplak, M. & Jernigan, R. L. Predicting the order in which contacts are broken during single molecule protein stretching experiments. *Proteins: Structure, Function, and Bioinformatics* **71**, 45–60 (2008).
- Toofanny, R. D. & Daggett, V. Understanding protein unfolding from molecular simulations. *WIREs Computational Molecular Science* **2**, 405–423. <https://doi.org/10.1002/wcms.1088>.
- Alm, E. & Baker, D. Matching theory and experiment in protein folding. *Curr Opin Struct Biol* **9**, 189–196 (1999).
- Alm, E. & Baker, D. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proceedings of the National Academy of Sciences* **96**, 11305–11310 (1999).
- Perl, D. et al. Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nat Struct Mol Biol* **5**, 229–235 (1998).
- Rader, A., Hespeneheide, B. M., Kuhn, L. A. & Thorpe, M. F. Protein unfolding: rigidity lost. *Proceedings of the National Academy of Sciences* **99**, 3540–3545 (2002).
- Jacobs, D. J., Rader, A., Kuhn, L. A. & Thorpe, M. Protein flexibility predictions using graph theory. *Proteins: Structure, Function, and Bioinformatics* **44**, 150–165. <https://doi.org/10.1002/prot.1081>.
- Habibi, M., Plotkin, S. S. & Rottler, J. Soft vibrational modes predict breaking events during force-induced protein unfolding. *Biophys J* **114**, 562–569 (2018).
- Mitchell, M. R., Tlusty, T. & Leibler, S. Strain analysis of protein structures and low dimensionality of mechanical allosteric couplings. *Proceedings of the National Academy of Sciences* **113**, E5847–E5855 (2016).
- Dutta, S., Eckmann, J.-P., Libchaber, A. & Tlusty, T. Green function of correlated genes in a minimal mechanical model of protein evolution. *Proceedings of the National Academy of Sciences* **115**, E4559–E4568 (2018).
- Sartori, P. & Leibler, S. Evolutionary conservation of mechanical strain distributions in functional transitions of protein structures. *bioRxiv* 2023–02 (2023).
- Widmer-Cooper, A., Perry, H., Harrowell, P. & Reichman, D. R. Irreversible reorganization in a supercooled liquid originates from localized soft modes. *Nature Physics* **4**, 711–715 (2008).
- Tanguy, A., Mantsi, B. & Tsamados, M. Vibrational modes as a predictor for plasticity in a model glass. *Europhys Lett* **90**, 16004 (2010).

32. Manning, M. L. & Liu, A. J. Vibrational modes identify soft spots in a sheared disordered packing. *Phys Rev Lett.* **107**, 108302 (2011).
33. Smessaert, A. & Rottler, J. Structural relaxation in glassy polymers predicted by soft modes: a quantitative analysis. *Soft Matter* **10**, 8533–8541 (2014).
34. Smessaert, A. & Rottler, J. Correlation between rearrangements and soft modes in polymer glasses during deformation and recovery. *Physical Review E* **92**, 052308 (2015).
35. Sastry, S., Debenedetti, P. G. & Stillinger, F. H. Signatures of distinct dynamical regimes in the energy landscape of a glass-forming liquid. *Nature* **393**, 554–557 (1998).
36. Wang, J., Plotkin, S. S. & Wolynes, P. G. Configurational diffusion on a locally connected correlated energy landscape; application to finite, random heteropolymers. *Journal de Physique I*(7), 395–421 (1997).
37. Schweizer, K. S. & Saltzman, E. J. Entropic barriers, activated hopping, and the glass transition in colloidal suspensions. *J Chem Phys.* **119**, 1181–1196 (2003).
38. Bhattacharyya, S. M., Bagchi, B. & Wolynes, P. G. Facilitation, complexity growth, mode coupling, and activated dynamics in supercooled liquids. *Proceedings of the National Academy of Sciences* **105**, 16077–16082 (2008).
39. Spaepen, F. A microscopic mechanism for steady state inhomogeneous flow in metallic glasses. *Acta metallurgica* **25**, 407–415 (1977).
40. Langer, J. Microstructural shear localization in plastic deformation of amorphous solids. *Physical Review E* **64**, 011504 (2001).
41. Dasgupta, R., Hentschel, H. G. E. & Procaccia, I. Microscopic mechanism of shear bands in amorphous solids. *Phys Rev Lett.* **109**, 255502 (2012).
42. Wisitsorasak, A. & Wolynes, P. G. Dynamical theory of shear bands in structural glasses. *Proceedings of the National Academy of Sciences* **114**, 1287–1292 (2017).
43. Su, J. G., Li, C. H., Hao, R., Zu Chen, W. & Wang, C. X. Protein unfolding behavior studied by elastic network model. *Biophys. J.* **94**, 4586–4596 (2008).
44. Haliloglu, T., Bahar, I. & Erman, B. Gaussian dynamics of folded proteins. *Phys Rev Lett.* **79**, 3090 (1997).
45. Bahar, I., Atilgan, A. R. & Erman, B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding and Design* **2**, 173–181 (1997).
46. Eyal, E. & Bahar, I. Toward a molecular understanding of the anisotropic response of proteins to external forces: insights from elastic network models. *Biophys J.* **94**, 3424–3435 (2008).
47. Chennubhotla, C., Rader, A., Yang, L.-W. & Bahar, I. Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies. *Phys Biol.* **2**, S173 (2005).
48. Tama, F. & Sanejouand, Y.-H. Conformational change of proteins arising from normal mode calculations. *Protein engineering* **14**, 1–6 (2001).
49. Yang, L., Song, G. & Jernigan, R. L. How well can we understand large-scale protein motions using normal modes of elastic network models?. *Biophys J.* **93**, 920–929 (2007).
50. Delarue, M. & Sanejouand, Y.-H. Simplified normal mode analysis of conformational transitions in dna-dependent polymerases: the elastic network model. *J Mol Biol.* **320**, 1011–1024 (2002).
51. Srivastava, A. & Granek, R. Cooperativity in thermal and force-induced protein unfolding: integration of crack propagation and network elasticity models. *Phys Rev Lett.* **110**, 138101 (2013).
52. Guest, W. C., Cashman, N. R. & Plotkin, S. S. A theory for the anisotropic and inhomogeneous dielectric properties of proteins. *Phys Chem Chem Phys.* **13**, 6286–6295 (2011).
53. Eyal, E., Yang, L.-W. & Bahar, I. Anisotropic network model: systematic evaluation and a new web interface. *Bioinfo* **22**, 2619–2627 (2006).
54. Atilgan, A. R. et al. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J.* **80**, 505–515 (2001).
55. Zhang, H., Jiang, T., Shan, G., Xu, S. & Song, Y. Gaussian network model can be enhanced by combining solvent accessibility in proteins. *Sci. Rep.* **7**, 7486. <https://doi.org/10.1038/s41598-017-07677-9>.
56. Dietz, H. & Rief, M. Elastic bond network model for protein unfolding mechanics. *Phys. Rev. Lett.* **100**, 098101. <https://doi.org/10.1103/PhysRevLett.100.098101>.
57. Tirion, M. M. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys Rev Lett.* **77**, 1905 (1996).
58. Srivastava, A. & Granek, R. Protein unfolding from free-energy calculations: Integration of the gaussian network model with bond binding energies. *Phys. Rev. E.* **91**, 022708. <https://doi.org/10.1103/PhysRevE.91.022708>.
59. Micheletti, C., Carloni, P. & Maritan, A. Accurate and efficient description of protein vibrational dynamics: comparing molecular dynamics and gaussian models. *Proteins: Structure, Function, and Bioinformatics* **55**, 635–645 (2004).
60. Doruker, P., Atilgan, A. R. & Bahar, I. Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: Application to  $\alpha$ -amylase inhibitor. *Proteins: Structure, Function, and Bioinformatics* **40**, 512–524 (2000).
61. Atilgan, C. & Atilgan, A. R. Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein. *PLoS Comput Biol.* **5**, e1000544 (2009).
62. Gullett, P. M., Horstemeyer, M. F., Baskes, M. I. & Fang, H. A deformation gradient tensor and strain tensors for atomistic simulations. *Model. Simulat. Mat. Sci. Eng.* **16**, 015001. <https://doi.org/10.1088/0965-0393/16/1/015001>.
63. Zhang, Y. & Skolnick, J. Tm-align: a protein structure alignment algorithm based on the tm-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).
64. Baker, D. A surprising simplicity to protein folding. *Nature* **405**, 39–42 (2000).
65. Mu noz, V. & Eaton, W. A. A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proceed. Nat. Acad. Sci.* **96**, 11311–11316 (1999).

## Author contributions

SD designed the research. RK carried out all simulations. SD and RK analyzed the data and wrote the article.

## Funding

The author acknowledges financial support by Science and Engineering Research Board (SERB), Government of India (File SRG/2022/000598) and (File MTR/2022/000281). The author also acknowledges financial support from the Birla Institute of Technology and Science, Pilani (File PLN/AD/2022-23/3). The author also thanks DST-FIST for the computational resources provided to the Department of Physics, BITS Pilani. Open access funding provided by Birla Institute of Technology and Science.

## Declarations

## Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to S.D.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024