COMPUTATIONAL
AND STRUCTURAL
BIOTECHNOLOGY
J O U R N A L

Mechanical couplings of protein backbone and side chains exhibit scale-free network properties and specific hotspots for function

Nixon Raj [a], Timothy Click [a], Haw Yang [e], Jhih-Wei Chu [a,b,c,d,]*

[a] Institute of Bioinformatics and Systems Biology, National Yang Ming Chiao Tung University, 75 Bo-Ai Street, Hsinchu 30010, Taiwan, ROC
[b] Department of Biological Science and Technology, National Yang Ming Chiao Tung University, 75 Bo-Ai Street, Hsinchu 30010, Taiwan, ROC
[c] Institute of Molecular Medicine and Bioengineering, National Yang Ming Chiao Tung University, 75 Bo-Ai Street, Hsinchu 30010, Taiwan, ROC
[d] Center for Intelligent Drug Systems and Smart Bio-devices (IDS^{2}B), National Yang Ming Chiao Tung University, 75 Bo-Ai Street, Hsinchu 30010, Taiwan, ROC
[e] Department of Chemistry, Princeton University, Princeton, NJ 08544, USA

A R T I C L E   I N F O

A B S T R A C T

A backbone-side-chain elastic network model (bsENM) is devised in this contribution to decipher the network of molecular interactions during protein dynamics. The chemical details in 5 μs all-atom molecular dynamics (MD) simulation are mapped onto the bsENM spring constants by self-consistent iterations. The elastic parameters obtained by this structure-mechanics statistical learning are then used to construct inter-residue rigidity graphs for the chemical components in protein amino acids. A key discovery is that the mechanical coupling strengths of both backbone and side chains exhibit heavy-tailed distributions and scale-free network properties. In both rat trypsin and PDZ3 proteins, the statistically prominent modes of rigidity graphs uncover the sequence-specific coupling patterns and mechanical hotspots. Based on the contributions to graphical modes, our residue rigidity scores in backbone and side chains are found to be very useful metrics for the biological significance. Most functional sites have high residue rigidity scores in side chains while the biologically important glycines are generally next to mechanical hotspots. Furthermore, prominent modes in the rigidity graphs involving side chains oftentimes coincide with the co-evolution patterns due to evolutionary restraints. The bsENM specifically devised to resolve the protein chemical character thus provides useful means for extracting functional information from all-atom MD.

1. Introduction

Proteins exhibit remarkable properties such as thermal stability, specific molecular binding, and catalytic activities. These functionally important features are sensitive to mutation and can trace their origin to both the polypeptide backbone that frames the structure and the side chains that define the chemical specificity [1–3]. Deciphering the contributions of these two components to functional properties, though, remains a fundamental challenge [4–6]. The folded topology was recognized to host a variety of sequences in structural families [7,8] and has inspired artificial protein engineering and design [9–11]. Constructing potential energy function with the structural network, such as the elastic network model (ENM) [12–14] and the Gō model [15–17], is very

useful in studying functional motions [12–14], protein folding [15–17], and allosteric wiring [18–20]. In addition to the Hamiltonian-based methods, graphical analysis [21–23] is frequently applied to analyze the protein structural network. If the distance between a residue pair is within a cutoff, their edge in the adjacency matrix **A** is typically set to one, and the diagonal degree matrix **D** records the residue contact numbers. This topology-based approach corresponds to using a universal spring constant in ENM. The Laplacian matrix ($\mathbf{L} = \mathbf{D} - \mathbf{A}$) [24,25] was found to offer good approximation for low-frequency motions [13,14], and the structural network is often used to study the collective vibrations that are not very sensitive to the sequence specificity due to side chains. [26–29].

Yet, beyond the backbone-framed protein structure, how to delineate the networks of molecular interactions, what are their differences in comparison to the structural network, and what are the manifestations of the side-chain sequence and dynamical motions? In particular, if the interaction network of side chains

* Corresponding author at: Institute of Bioinformatics and Systems Biology, National Yang Ming Chiao Tung University, 75 Bo-Ai Street, Hsinchu 30010, Taiwan, ROC.
E-mail address: jwchu@nctu.edu.tw (J.-W. Chu).

could be studied, its properties are expected to differ from that of the backbone. To address these key issues of the sequence-structure-function relationship, a graph-theoretic methodology is devised here to compute the mechanical interactions mediated by side chains and backbone from all-atom molecular dynamics (MD) simulations in an explicit solvent.

The guiding principle is that given the unique structural position and chemical environment of a protein residue, its couplings with surrounding atoms would assume specific strength during dynamical motions. A mechanical-coupling dynamics perspective is thus taken to unravel the network behaviors of physical interactions. In particular, the elastic parameters in a model of backbone and side-chain nodes connected by harmonic springs, i.e., the backbone-side-chain elastic network model, bsENM, proposed in this work and illustrated in Fig. 1A, is used to represent the effective interaction strengths. Our design of bsENM is to explicitly represent the protein chemical components for resolving the backbone and side-chain contributions in the mechanical coupling network. In particular, the scope of bsENM spring constants is expanded for an unexplored context: to statistically learn mechanical coupling strengths of backbone and side chains from all-atom MD simulations. The effective elasticities calculated by self-consistent iterations [30–32] are then used to construct inter-residue graphs. As will be shown later, this bsENM-graph approach can be used to map the protein dynamics into distinct rigidity groupings, namely the backbone-backbone (BB), the backbone-side-chain (BS), and the side-chain-side-chain (SS). Investigating them using spectral analysis allows us to uncover their unique mechanical topologies for comparing with experimental observables such as residue conservation and co-evolution in multiple sequence alignment (MSA), mutation sensitivity, residue flexibility profile, and signals reflecting residue micro-environments. Under this structure-mechanics statistical learning framework, in contrast to the aforementioned topology-only approach, even interaction pairs of similar distance separation can have very different coupling strengths. Using the elastic properties statistically learned from an all-atom MD trajectory as the edge weights in $\mathbf{A}$ and $\mathbf{D}$ thus offers a new perspective—the rigidity graphs of protein dynamics. To reveal the impact of chemical details, the specifically designed scheme is dividing the bsENM harmonic potentials into (a) skeleton springs as those linking the nearest and second nearest residues and (b) non-skeleton springs as the rest. With $\mathbf{A} = \mathcal{A}(\text{skeleton springs}) + \mathbb{A}(\text{non-skeleton springs})$ and $\mathbf{D} = \mathcal{D} + \mathbb{D}$, we first establish that the low-frequency modes of the skeleton Laplacian ($\mathcal{L} = \mathcal{D} - \mathcal{A}$) are exceedingly insensitive to the variation in strength during protein dynamics. Analysis of mechanical couplings is hence focused on non-skeleton springs. A key finding is that the non-skeleton signless Laplacian ($\mathbb{K} = \mathbb{D} + \mathbb{A}$) reveals the specific patterns very clearly. The statistically prominent features of $\mathbb{K}$ can thus be extracted
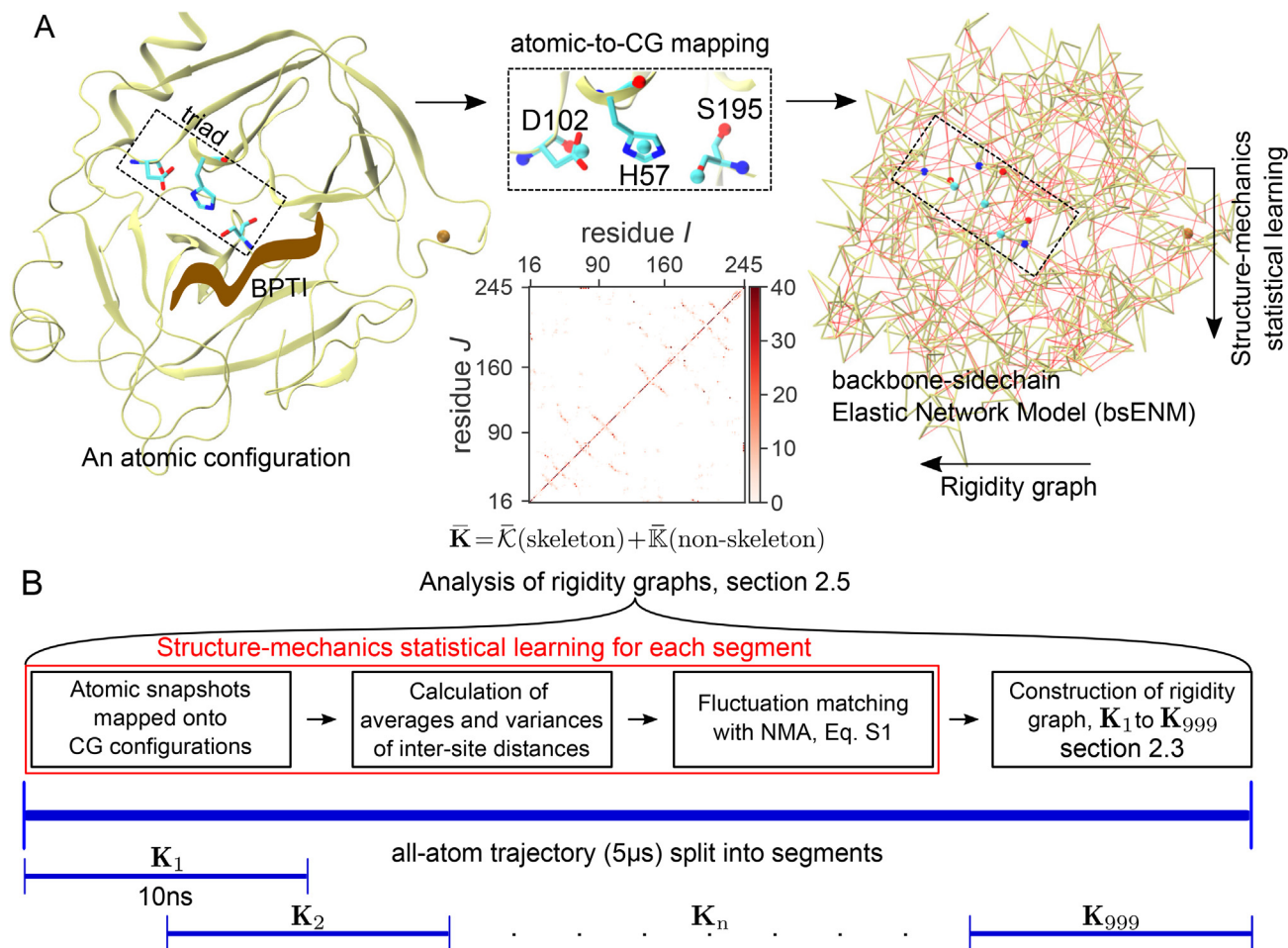
Fig. 1. Rigidity graphs of the protein mechanical coupling network statistically learned from all-atom MD, see Materials and Methods for details. (A) A sampled atomic structure is mapped onto the coordinates of backbone and side-chain sites in the coarse-grained (CG) representation of bsENM. Left: a ribbon representation for the atomic structure of RT bound with BPTI. The catalytic triad is highlighted and a zoomed-in view showcases the atomic-to-CG mapping in Table S1. Right: the mapped CG configuration. (B) Dividing the 5 μs all-atom MD trajectory into consecutive 10 ns windows, and the work flow for computing the bsENM parameters of each segment by structure-mechanics statistical learning.

from the all-atom MD trajectory to reveal the mechanical topologies of different interactions. To illustrate this approach, we choose as case studies a serine protease family member rat trypsin (RT) [33,34] and PDZ3 (the third PDZ signaling domain) [35–37] because they both are β-strand rich and have comprehensive data for site-specific mutagenesis and functional activity experiments. In addition, the wealth of sequence information allows the calculation of inter-residue co-evolution from statistical coupling analysis (SCA) [38–40] or direct coupling analysis (DCA) [41,42] from MSA and provides complementary insights. The BB (backbone-backbone), SS (side chain-side chain), and BS (backbone-side chain) rigidity graphs can thus be compared with these observables for linking physical interactions with biological functions and evolutionary pressure.

2. Materials and methods

The elastic parameters in bsENM (k_{ij}'s) are calculated from an all-atom MD trajectory by matching the fluctuations of inter-site distances. Effectively, this structure-mechanics statistical learning integrates out the other degrees of freedom by self-consistent iteration with normal mode analysis (NMA) [30–32]. Moreover, mechanical coupling strengths represented by the spring constants are used to construct BB, BS, and SS rigidity graphs. This computational framework is detailed in the following using RT as the example.

2.1. All-atom MD simulation

The X-ray structure of BPTI bound RT (PDB ID: 3TGI) is used to construct its all-atom model [33] whereas for PDZ3, the apo X-ray structure (PDB ID: 1BFE) [37] is used. All systems are solvated in orthorhombic dodecahedron TIP3P water boxes and neutralized with NaCl ions at 0.15 M. The CHARMM36 all-atom force field [43] is used to compute the potential energy and the GROMACS software [44] is used for MD runs. The production run for both the RT and PDZ3 systems is at 300 K and 1 atm for 5 μs, during which a snapshot is saved every 1 ps for analysis. The other details are reported in SI.

2.2. Structure-mechanics statistical learning of bsENM parameters

With the raw data of protein dynamics generated with full atomic details, the CG sites in bsENM serve to read out the statistics of inter-site elasticity. The goal is to capture the significant mechanical couplings that can survive thermal noise, and the CG sites are thus located where specific molecular interactions are typically observed in the trajectories. In particular, the backbone is represented by two coarse-grained (CG) sites at the amide nitrogen and carbonyl oxygen positions as they are the loci of hydrogen bonding, Fig. 1A. For side chains, a CG site is placed at the position of a representative atom at which specific interactions are formed. For example, the alanine side-chain site is at C_β and that of lysine is at N_ζ. For hydrophobic side-chains, the center of mass of heavy atoms is used. Table S1 lists the details of this atomic-to-CG mapping, which is used to convert each frame in the MD trajectory to a bsENM configuration.

Given a set of bsENM configurations, the spring length l_{ij}^0 between sites i and j is their averaged distance. To parametrize k_{ij}, the variance of distance fluctuation, $\langle \delta l_{ij}^2 \rangle_{AA}$, in the all-atom MD data is the targeted value. At each iteration step, NMA of the bsENM gives the predicted distance fluctuation, and k_{ij} is adjusted to match the targeted value [31,32]. Since the bsENM springs are connected in the structure, their coupled fluctuations are handled by self-consistent iterations in this fluctuation matching. Other

details are reported in SI. To construct the starting model (initial guess) for statistical learning, a cutoff distance l_c is used to include a harmonic potential in the bsENM for every inter-site pair with $l_{ij}^0 < l_c$. The cutoff is thus an adjustable parameter and is determined by scanning the value and comparing the resulting residue RMSF (root-of-mean-squared-fluctuation) and the low-frequency vibrational modes of the statistically learned bsENM with those calculated from all-atom MD. As discussed in Fig. S1, consistent behaviors are observed over a wide range of cutoff values since the bsENM springs are trained to match the all-atom MD target data. The default l_c is set to 7.8 Å as shown in Fig. S1. After convergence of the first round, connectivity trimming followed by another round of fluctuation matching is conducted to prevent having excessive springs in the network; other details are reported in SI. An approximation of bsENM is using a universal value for all springs within the cutoff. This 0th-order construction solely relying on the structural network is denoted bsENM$_0$. To highlight the effects of chemical details within the native topology, the bsENM statistically learned from all-atom MD is compared with the bsENM$_0$ of the same equilibrium structure and l_c.

2.3. Construction of inter-residue rigidity graphs from bsENM

In our graphical representation of inter-residue interaction networks, the edge weights between residue nodes I and J in the adjacency matrix are $k_{IJ} = \sum_{i \in I, j \in J} k_{ij}$, the sum over the bsENM spring constants linking their CG sites. For bsENM$_0$, k_{ij} is set to 1 for the springs within l_c. In this graphical theory, the degree matrix components record the total coupling strength of each residue $k_{II} = \sum_{I \neq J} k_{IJ}$, or, in the case of bsENM$_0$, the residue contact number of CG sites within l_c. Since bsENM springs can be categorized by the types (backbone or side chain) of their CG sites, the graphs of different rigidity groups can be constructed accordingly. For example, $\mathbb{K}^{BB}$, $\mathbb{K}^{BS}$, and $\mathbb{K}^{SS}$ are the signless Laplacian of the non-skeleton springs in backbone-backbone, backbone-side-chain, and side-chain-side-chain groups, respectively. For $\mathbb{K}^{SS}$, disulfide bond springs are exceedingly strong and are skipped to focus on non-covalent mechanical couplings.

2.4. Comparison of rigidity graphs

A protein of N residues can thus have different $N \times N$ rigidity graphs that are symmetric and positive-semidefinite or positive-definite. The eigenvectors, which also form an orthonormal basis set, are ordered according to their eigenvalues in a descending order with mode index α. The eigenvectors of a rigidity graph are the specific patterns of mechanical couplings between protein residues. To quantify whether different rigidity graphs have similar behaviors, the following mode-based procedure is developed.

The similarity of graph $\mathbf{L}$ with respect to a reference graph $\mathbf{R}$ along its mode α is defined as $r_\alpha = \max_\beta \left| \boldsymbol{v}_\beta^{\mathbf{L}} \cdot \boldsymbol{v}_\alpha^{\mathbf{R}} \right|$, i.e., by finding the eigenvector β_{max} in $\mathbf{L}$ that the dot product with mode α in $\mathbf{R}$ has the largest magnitude. Here, $\boldsymbol{v}_\beta^{\mathbf{L}}$ and $\boldsymbol{v}_\alpha^{\mathbf{R}}$ are the eigenvectors of the $\mathbf{L}$ and $\mathbf{R}$ rigidity graphs, respectively. For example, if $r_\alpha = 1$ for $\mathbf{L}$, the bsENM Laplacian, with respect to $\mathbf{L}_0$, the bsENM$_0$ Laplacian, then $\mathbf{L}$ has an identical counterpart as mode α in $\mathbf{L}_0$. The mode of the compared graph that delivers the phase, β_{max}, may not be the same as α in the reference graph, since the respective mode rankings may differ.

2.5. Statistical analysis of protein rigidity graphs

With the 5 μs production runs of RT and PDZ3, a single bsENM using the entire trajectory only provides an equilibrium harmonic

approximation for structural fluctuations, omitting the potentially interesting and informative fluctuations on the interaction-network level. This limitation is overcome by dividing the trajectory into consecutive 10-ns windows to compute a series of rigidity graphs (see Fig. 1B for an illustration of windowed trajectory segmentation for statistical analysis). The 10-ns window appears to be a reasonably robust window size in terms of extracting mechanical coupling parameters. This is illustrated in Fig. S2, which shows that the distribution of the inter-site interaction strength, k_{ij}'s, appears unchanged in 5-ns, 10-ns, or 20-ns window sizes. Therefore, the coupling network (e.g., the rigidity graph) extracted from a 10-ns window trajectory is used as an element in further statistical analyses.

2.5.1. Mean-modes of fluctuating rigidity graphs

The rigidity graph of a protein fluctuates and evolves over time as the protein experiences thermal fluctuations or interacts with the surrounding molecules. To capture and quantify these fluctuations on the network level using the rigidity graph, we begin by defining "mean-modes" which are to be understood as the average eigenmodes of an otherwise fluctuating rigidity graph. They are computed as follows: From the bsENM of each trajectory window indexed by n, the non-skeleton springs are used to calculate the off-diagonal k_{IJ} and the diagonal k_{II} terms in the inter-residue rigidity graph $\mathbb{K}_n$. Averaging the $\mathbb{K}_n$ graphs over the temporal segments gives the mean rigidity graph $\bar{\mathbb{K}}$. From the mean rigidity graph $\bar{\mathbb{K}}$, one finds eigenvectors, $\boldsymbol{v}_{\alpha'}$'s, each of which is a mean-mode with a corresponding eigenvalue $\lambda_{\alpha'}$—the coupling strength for the α' mean-mode. Note that hereafter a superscripted index such as α' is used for variables and functions derived from the mean rigidity graph, $\bar{\mathbb{K}}$.

2.5.2. Content of a mean-mode in an analysis time window

Each mean-mode is a unit vector for the N residues of the rigidity graph, and can be understood as an N-vector mechanical coupling pattern. In this light, in addition to the coupling strength, another important property is how much the pattern of a particular mean-mode is retained in the mechanical coupling network of each of the analysis window. This property is determined by first calculating the mean-mode content in each trajectory window following the description in Section 2.4 as $r_{n\alpha'} = \max_\beta |\boldsymbol{v}_{n\beta} \cdot \boldsymbol{v}_{\alpha'}|$, where $\boldsymbol{v}_{n\beta}$ is an eigenvector of $\mathbb{K}_n$ indexed by β. A high mean-mode content of $r_{n\alpha'} \sim 1$ indicates that the pattern of $\boldsymbol{v}_{\alpha'}$ stays the same in the trajectory window. Averaging the mean-mode content over all trajectory windows then gives the averaged mean-mode content $r_{\alpha'}$. It measures the extent to which the eigenvector-pattern is retained throughout the entire all-atom MD trajectory.

2.5.3. Prominent modes of a rigidity graph

With the coupling strength $\lambda_{\alpha'}$ and the average content $r_{\alpha'}$ defined for a given mechanical-coupling pattern (designated by the mean-mode α'), we next ask which mean-mode features are most prominent as the protein undergoes dynamical structural fluctuations. For the present work, we define a "prominent mode" as a mean-mode that exhibits both (a) strong strength in the mechanical coupling and (b) high averaged content during the protein dynamics. More quantitatively, for (a), mean-modes exhibiting strong couplings are defined as those showing higher than the upper fence of an empirically defined quantity, $Q_3 + 1.5 \times$ (inter-quartile region), within the $\{\lambda_{\alpha'}\}$ distribution (see Fig. S3). For (b), mean-modes having high contents during the protein dynamics are based on the cumulative density function (CDF) of $r_{\alpha'}$. The cutoff for high-content designation is assigned empirically depending on the type of mechanical coupling. For $\bar{\mathbb{K}}^{BB}$ and $\bar{\mathbb{K}}^{BS}$, only those at top 25% of $r_{\alpha'}$ values are considered

the candidates of prominent modes, while the empirical percentile cutoff is top 32% for the $\bar{\mathbb{K}}^{SS}$ modes (see Fig. S3 for summary plots). As shown in Figs. S4–S6, the prominent modes of $\mathbb{K}$ rigidity graphs exhibit pointed patterns in residue weights ($v_{I\alpha'}^2$), and mechanical hotspots are filtered out as the residues having significant population.

2.6. Residue rigidity scores for backbone and side chain

The set of prominent modes in a $\bar{\mathbb{K}}$ rigidity-graph, Π, contains the strong and high-content mechanical coupling patterns during protein dynamics and are potentially important for functional activities. Therefore, key residues in the prominent modes likely have significant biological importance, thereby providing yet further refined chemical specificity. With this insight, the participation of each protein residue in the prominent modes is used to compute a quantitative metric in the mechanical coupling network—the residue rigidity score. In a rigidity graph, we start by finding the characteristic mode $\mathcal{I}'$ for each residue indexed by I. The protein residue is first associated with a rigidity-graph mode $\pi' \in \Pi$ that its weight is the highest, and the $\mathcal{I}'$ of I is set to π' if the weight is significant ($v_{I\pi'}^2 \geqslant v_c^2$).

$$\mathcal{I}' = \begin{cases} \pi' = \arg\max\limits_{\beta' \in \Pi} v_{I\beta'}^2 \text{ if } v_{I\pi'}^2 \geqslant v_c^2 \\ \alpha' = \arg\max\limits_{\beta' \notin \Pi} v_{I\beta'}^2 \text{ if } v_{I\pi'}^2 < v_c^2. \end{cases} \quad (1)$$

In this equation, if residue I does not play a significant role in any of the prominent modes ($v_{I\pi'}^2 < v_c^2$), $\mathcal{I}'$ is one of the rest eigenvectors ($\beta' \notin \Pi$) that it has the highest weight. As shown in Figs. S4–S6, the empirical $v_c^2 = 0.1$ is used to identify the significantly populated residues in a prominent mode. Next, the residue rigidity score in the graph is defined as $\kappa_I = \langle r_{\mathcal{I}'} \rangle \lambda_{\mathcal{I}'}$, i.e., the mechanical strength weighted by the averaged mean-mode content. If the $\mathcal{I}'$ of residue I is in Π, its residue rigidity score is high, while if $\mathcal{I}'$ has a weak strength and/or low averaged content during protein dynamics, κ_I is low. The residue rigidity score in backbone is defined as $\kappa_I^B = \max(\kappa_I^{BB}, \kappa_I^{BS})$, the maximum score the residue delivers though its backbone, and κ_I^{BS} is compared only if residue I contributes backbone in the mode. Similarly, the residue rigidity score in side-chain is calculated as $\kappa_I^S = \max(\kappa_I^{SS}, \kappa_I^{BS})$, and κ_I^{BS} is only considered if residue I participates by its side-chain.

3. Results and discussion

With the all-atom MD simulation, structure-mechanics statistical learning, and rigidity graph analysis for both the RT and the PDZ3 proteins, the main text primarily uses RT for introducing the rich and quantitative information made available by our new approach. The two systems illustrate the common general features of sparse mechanical coupling network, scale-free network coupling strengths, hotspots in both backbone and side-chain interaction networks, and the residue rigidity scores during protein-dynamics as a new metric for biochemical functions. In what follows, the mechanical coupling network is analyzed in detail.

3.1. The protein mechanical coupling network is sparse

Our structure-mechanics statistical learning with bsENM provides a way to map the chemical details during protein dynamics onto k_{ij} values. The bsENM calculated from a trajectory segment thus contains a list of springs each with a length l_{ij}^0 and a positive-definite elastic constant k_{ij}. The diverse mechanical coupling strengths can be seen in the k_{ij} distribution from a RT trajec-

tory window, Fig. 2A. The skeleton springs between residues neighbors and disulfide bonds are exceedingly strong and will be discarded later in the analysis of rigidity graphs. In Fig. 2B, the number of inter-site pairs in the protein structure separated by the l_{ij}^0 value in each bin is shown with its fraction of springs that converge to $k_{ij} > 0$ after the self-consistent iterations. With increasing l_{ij}^0, it can be seen that the fraction of positive-definite springs drops further, and the protein mechanical coupling network is thus progressively sparser than the structural contact network. Sparsity of the mechanical coupling network signifies that the protein fold can afford sequence variations to accommodate different functions, and the network properties of mechanical couplings during protein dynamics can potentially serve to capture the functionally important interactions.

Although the inter-residue Laplacian matrix ($\mathbf{L} = \mathbf{D} - \mathbf{A}$) of topological contacts is useful in mimicking collective modes, capturing the specific mechanical coupling patterns may call for a different representation. This aspect is illustrated by using the $\mathbf{L}_0$ of bsENM$_0$ from topological contacts (no chemical details) as the reference. For the lowest-frequency modes of $\mathbf{L}$ from all-atom MD, the similarity r_α being close to 1 with respect to those of $\mathbf{L}_0$ (Fig. S7A)
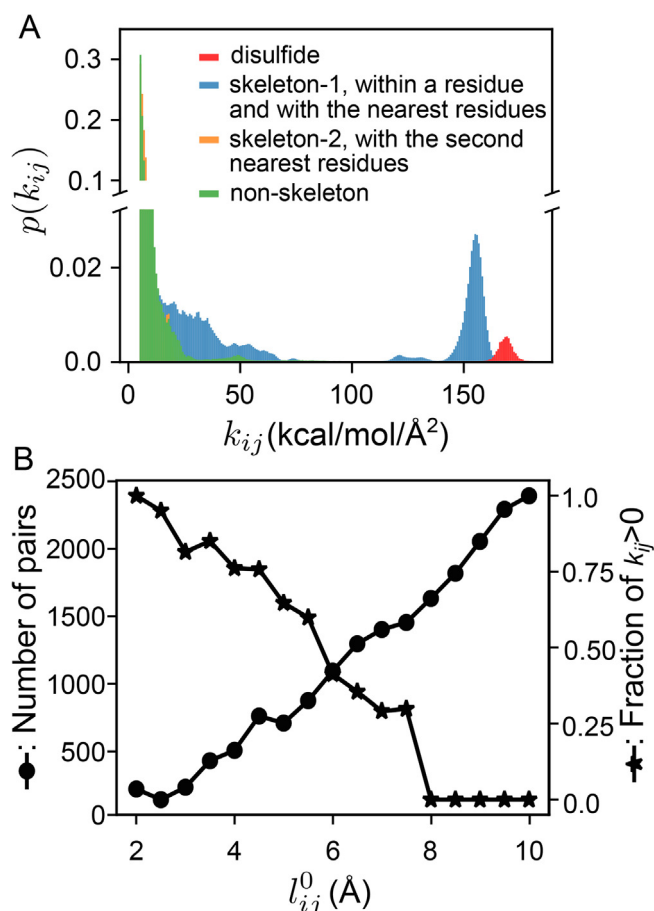

Fig. 2. The elastic spring parameters of a bsENM statistically learned from a trajectory window of RT. (A) Normalized histogram of coupling strengths. Between CG sites i and j, k_{ij} is the coupling strength. Disulfide springs that connect the Sγ atoms of a cysteine pair are very strong. Skeleton-1 includes the springs within a residue and between the nearest residues. The very high strengths in skeleton-1 are peptide bonds, whereas the other springs in this category are on the left. Skeleton-2 is the springs between the second nearest residues. Non-skeleton springs are the rest with disulfide bonds also excluded. (B) In each 0.5 Å bin of the spring length l_{ij}^0, the number of pairs in the protein structure (left) versus the number of springs converging to a non-zero k_{ij} in the fluctuation matching of a trajectory window (right).

shows that the collective vibrations in $\mathbf{L}$ are indeed insensitive to the chemical details. If considering only the skeleton springs, the $\mathcal{L}$ modes are essentially identical to those of $\mathcal{L}_0$ (Fig. S7A), whereas the modes of $\mathbb{L}$ from non-skeleton springs exhibit lower r_α values with respect to those of $\mathbb{L}_0$. Robustness of the low-frequency modes in $\mathbf{L}$ thus mostly comes from the skeleton springs.

To better reveal the molecular specificities in inter-residue elasticities (k_{IJ}'s), we seek to the $\mathbf{K} = \mathbf{D} + \mathbf{A}$ rigidity graph. The similarity r_α of the $\mathbf{K}$ modes with respect to those of $\mathbf{K}_0$ shows that the lowest-frequency modes of $\mathbf{K}$ still exhibit specific behaviors as the r_α values are low, Fig. S7. Even for skeleton springs, $\mathcal{K}$ exhibit significant differences comparing to $\mathcal{K}_0$. Furthermore, inspection of the $\mathbb{K}$ eigenmodes for non-skeleton springs (Fig. S8) shows that they reveal clear signals for the patterns of non-covalent k_{IJ}'s. Therefore, we employ the $\mathbb{K}$ rigidity graph and its BB, BS, and SS parts in the following to uncover the interplay of backbone and side chains in the protein mechanical coupling network.

3.2. Protein mechanical coupling networks have scale-free behaviors

The non-skeleton spring constants are used to construct the off-diagonal inter-residue coupling strength k_{IJ} in $\mathbb{K}$, and the diagonal k_{II} is the total coupling strength of residue I. In the graphical theory terms of $\mathbb{K}$, k_{IJ} is the edge weight between the I and J nodes, and k_{II} is the degree of node I. For a network, the degrees exhibiting power-law scaling in its high-value tale is an indicator of the scale-free property [45–47]. Such heavy-tailed profiles are due to most nodes having low values, but a small fraction of hotspots exhibits high couplings [48–50]. As a counter example, residue contact number m_{II} in the $\mathbb{K}_0$ graph of the bsENM$_0$ is not scale-free given the packing density in a native fold, Fig. 3A. The mechanical coupling network of the bsENM springs statistically learned from all-atom MD, on the other hand, behaves fundamentally different. The probability density $p(k_{II})$ exhibits a long tail and fits quite well with the classical Lomax distribution [51] for heavy-tailed profiles, Fig. 3B. As such, the mechanical coupling strengths during protein dynamics exhibit power-law scaling despite the data having more complicated patterns, and the exponent γ is in the typically encountered range ($2 < \gamma < 3$) of real-world scale-free networks [45–50]. The protein structural network, however, lacks such scale-free property. The data in Fig. 3 include the bsENM and bsENM$_0$ of every 10-ns trajectory window in the 5-μs production run of RT.

The scale-free behavior indicates that structural contacts within similar distances have highly diverse coupling strengths. Our finding of this network property of protein dynamics is consistent with the observation of mutational tolerance [52,53] that only a certain percentage of mutations would impact the phenotype. Protein rigidity graphs, similarly, have just a fraction of inter-residue edges carrying significant weights, and can potentially serve as the molecular-scale mechanistic basis for mutation sensitivity. Next, the functional connection is analyzed by first focusing on the specific properties of backbone and side-chain mechanical coupling networks.

A key advantage in our design of bsENM is that the CG sites are either the backbone or side-chain type. The backbone-backbone, backbone-side-chain, and side-chain-side-chain rigidity graphs can thus be constructed to uncover their separate behaviors in the mechanical coupling network, and $\mathbb{K} = \mathbb{K}^{BB} + \mathbb{K}^{BS} + \mathbb{K}^{SS}$. This decomposition of the mechanical coupling network shows that the rigidity graphs of different chemical components all exhibit heavy tails and scale-free behaviors, Fig. 4. Containing the overall weaker non-polar interactions, Fig. S9, the exponent of power-law scaling for the k_{II} values in $\mathbb{K}^{SS}$ is steeper ($\gamma = 4.55$), Fig. 4. The high-value outliers in $\mathbb{K}^{SS}$, though, exhibit complicated behav-
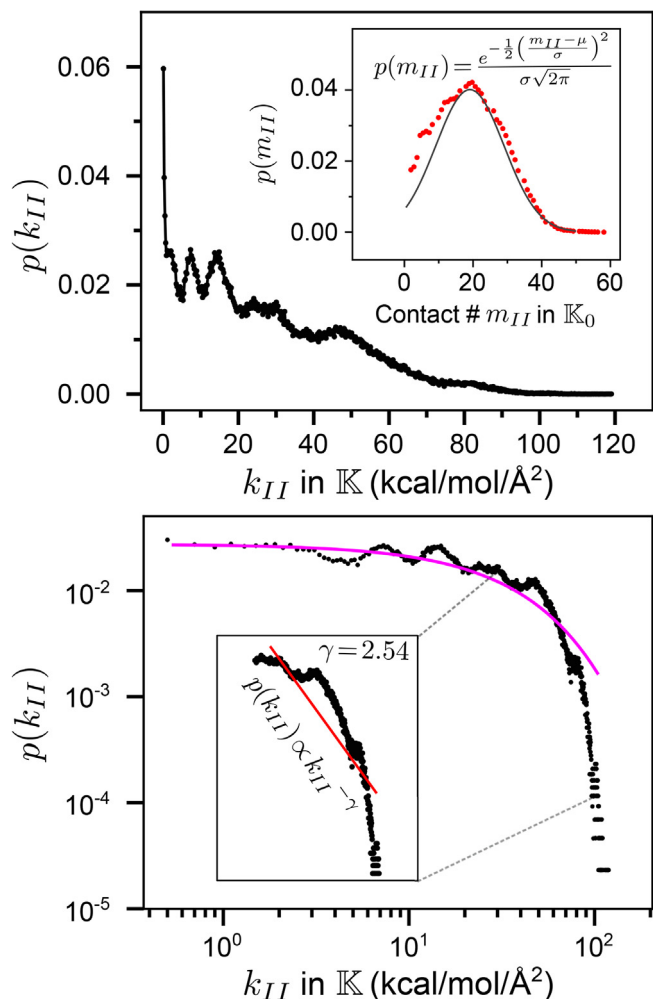
Fig. 3. Mechanical couplings during protein dynamics exhibit a heavy tailed distribution and scale-free network behavior. The diagonal components of $\mathbb{K}$ (bsENM) and $\mathbb{K}_0$ (bsENM$_0$) in every trajectory window of the 5 μs production run of RT are included. Top panel: $p(k_{II})$, the probability density distribution of k_{II}, the residue coupling strength due to non-skeleton springs in bsENM. Insert: $p(m_{II})$, where the $\mathbb{K}_0$ diagonal m_{II} is the residue contact number. It follows a Gaussian distribution given the packing density in the protein structure. Bottom panel: $p(k_{II})$ in the log-log scale. The orange line is the best-fit Lomax distribution. The red line in the insert is a power-law fit with the scaling exponent γ.

iors that deviate from the simple power-law equation. In an alternative representation by spectral analysis, the rigidity graph eigenvalues (λ_α's) are the coupling strengths of different modes, and they also exhibit heavy-tailed distributions and power-law scaling. For the λ_α distribution of $\mathbb{K}^{SS}$, the 2.39 exponent is similar to that of $\mathbb{K}^{BB}$ and $\mathbb{K}^{BS}$, Fig. 4, indicating synergistic combination of inter-residue couplings in the collective modes.

For the protein mechanical coupling networks of a fixed number of nodes (residues) to have scale-free behaviors, their edge weights exhibit high-strength tails as shown in Fig. S9. To analyze the dependence of network properties on chemical differences, the off-diagonal terms of $\mathbb{K}^{BB}$ (k_{IJ}'s) are grouped according to the secondary structure (sheet, helix, and loop) while those of $\mathbb{K}^{SS}$ are divided into polar and nonpolar groups. The $\mathbb{K}^{BS}$ couplings are all polar since backbone is involved. If either residue I or J is not in a helix or sheet, the pair is counted as in loop. The residue composition of RT secondary structures is listed in Fig. 5A. The power-law scalings of these categories are indeed different, and each case has evident higher and/or lower-valued outliers deviating from the simple formula. The α helix k_{IJ}'s are the highest populated in the
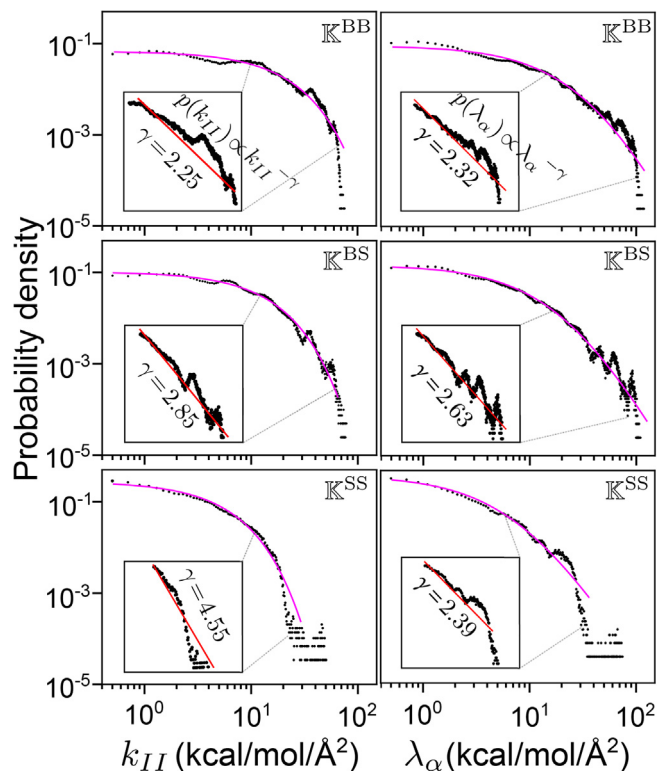
Fig. 4. The $p(k_{II})$ (left panels) and $p(\lambda_\alpha)$ (right panels) of $\mathbb{K}^{BB}$, $\mathbb{K}^{BS}$, and $\mathbb{K}^{SS}$ rigidity graphs on a log-log scale. The rigidity graphs of all trajectory windows in the 5 μs production run of RT are used. The orange line is the best-fit Lomax distribution for the heavy-tailed profiles. The red line in the insert is a power-law fit with the scaling exponent γ.

5-10 kcal/mol/Å^2 range, but they do not have any instance of very strong strengths (>15 kcal/mol/Å^2) and are the least-tailed group, Fig. S9. On the other hand, β sheets have much higher chances of exhibiting exceptional strengths during the dynamical motions and have the lowest γ. Heavy-tailed distributions of strengths are also seen in the BS, BB-loop, and SS-polar couplings. For the nonpolar side chains in RT, its k_{IJ} tail is shorter and the power-law scaling exponent is higher.

In the 5 μs dynamics of PDZ3, scale-free behaviors of the BB, BS, and SS rigidity graphs are similarly observed, Fig. S10, and the BB-helix and SS-nonpolar couplings are also less heavy-tailed, Fig. S11. The BB-helix k_{IJ}'s, though, have a better fit with the Lomax distribution than those in RT. Interestingly, the polar $\mathbb{K}^{SS}$ components exhibit a much extended tail in PDZ3, and strengths even higher than those of β sheet $\mathbb{K}^{BB}$ show up, Fig. S11, illustrating protein specific behaviors in the mechanical coupling networks.

The results of both RT and PDZ3 show that the mechanical couplings of backbone and side chains have different network properties. To further illustrate this point, specific patterns in the $\mathbb{K}$ rigidity graphs of non-skeleton k_{IJ}'s are characterized by spectral decomposition. This analysis also provides the data in identifying the mechanical hotspots based on the contributions of residue backbone and side chains.

3.3. Backbone and side chains exhibit specific mechanical hotspots

Graphical analysis of the bsENM quantitatively reveals the backbone and side-chain contributions in the mechanical coupling network. The high-strength tails in the eigenvalue distributions imply that their mechanical coupling patterns are more resistant to thermal noises during protein dynamics. The averaged mean-mode contents during protein dynamics, $r_{\alpha'}$, for the eigenmodes

Secondary structure	Range	Secondary structure	Range
β_1	Y20 − T21	I_I	I16 − G19
β_2	Q30 − N34	I_2	C22 − Y29
β_3	H40 − N48	I_3	S37 − Y39
β_4	W51 − S54	I_4	A55 − I63
β_5	Q64 − L68	I_5	G69 − E80
β_6	Q81 − K90	I_6	H91 − I103
β_7	M104 − L108	I_7	S109 − T134
β_8	Q135 − G140	I_8	W141 − L155
β_9	Q156 − P161	I_9	Y172 − N179
β_{10}	M180 − V183	I_{10}	G184 − G197
β_{11}	P198 − C201	I_{11}	G216 − P225
β_{12}	E204 − W215	α_1	Q165 − S171
β_{13}	G226 − K230	α_2	V231 − A243

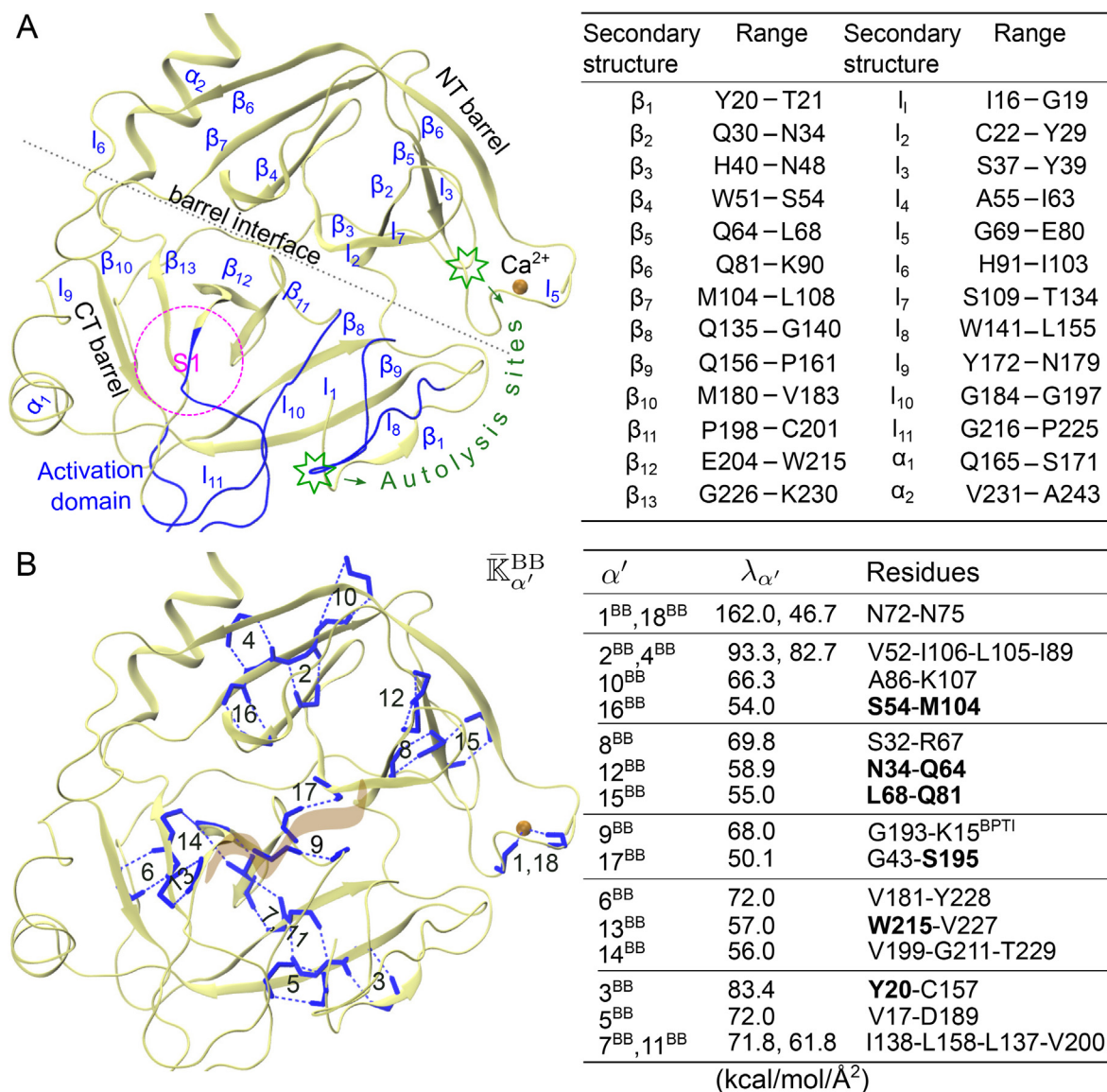| α'	$\lambda_{\alpha'}$	Residues
1^{BB}, 18^{BB}	162.0, 46.7	N72-N75
2^{BB}, 4^{BB}	93.3, 82.7	V52-I106-L105-I89
10^{BB}	66.3	A86-K107
16^{BB}	54.0	**S54-M104**
8^{BB}	69.8	S32-R67
12^{BB}	58.9	**N34-Q64**
15^{BB}	55.0	**L68-Q81**
9^{BB}	68.0	G193-K15^{BPTI}
17^{BB}	50.1	G43-**S195**
6^{BB}	72.0	V181-Y228
13^{BB}	57.0	**W215**-V227
14^{BB}	56.0	V199-G211-T229
3^{BB}	83.4	**Y20**-C157
5^{BB}	72.0	V17-D189
7^{BB}, 11^{BB}	71.8, 61.8	I138-L158-L137-V200

$$(\text{kcal/mol/Å}^2)$$

Fig. 5. In RT, the prominent backbone-only mechanical couplings during protein dynamics. (A) A ribbon representation the structure and residue composition of secondary structures. (B) Prominent mode residues in $\mathbb{K}^{BB}$ are in licorice. The BPTI inhibitor is labeled as brown ribbon. The index, eigenvalue, and residues of prominent modes are listed.

of $\mathbb{K}^{BB}$, $\mathbb{K}^{SS}$, or $\mathbb{K}^{BS}$ are calculated following the description in 2.5. The eigenvectors that show a statistically prominent $\lambda_{\alpha'}$ and $r_{\alpha'}$ over the RT trajectory are then identified in Fig. S3, and their pointed patterns inform the participating residues, Figs. S4–S6. For the 5 μs trajectory of RT bound with BPTI, the rigidity graph includes the inhibitor residues and the Ca^{2+} ion in RT is also treated as an additional residue. It is thus straightforward to adapt the bsENM-graph framework for studying complex protein systems. We focus on the modes of RT residues, and those of BPTI only are not presented.

The trypsin fold of RT containing NT and CT barrels (Fig. 5A) is a useful structural template for therapeutic design [54,55]. The high-strength $\mathbb{K}^{BB}$ eigenvectors indeed have high $r_{\alpha'}$ values and Fig. 5B shows the mechanical wiring of the 18 prominent modes. The $\mathbb{K}^{BB}$ eigenvector components $v_{I\alpha'}^{BB}$ of prominent modes often pick up residue pairs with very strong hydrogen bonds, such as the oxyanion hole G193 coupling to BPTI in $\mathbb{K}_9^{BB}$, yet more collective patterns (α'=2, 7, and 14) are also observed. Most of the prominent $\mathbb{K}^{BB}$ modes disperse in separate β strand-rich regions, and a notice-

able pattern is the strongest coupling locating at the cluster center with few nearby modes containing residues at secondary structure peripheries (edge residues of a β strand or α helix) as boldfaced in Fig. 5B. Out of the 34 hotspot residues in prominent $\mathbb{K}^{BB}$ modes, 20 are in secondary structures, 9 are at peripheries, and 5 are in loops, Fig. 6B. The catalytic triad S195, H57, and D102 are within 2-3 residues to β11, β4, and β7, respectively, and are considered at their peripheries. The backbone coupling of S195 with G43 as in $\mathbb{K}_{17}^{BB}$ links the NT-barrel and CT-barrel.

Spectral analysis of $\mathbb{K}^{SS}$ illustrates a different network topology, Fig. 6A and Fig. S6. In the SS prominent modes, dual-residue patterns often appear at the interface between NT and CT barrels due to very strong hydrogen bonds or salt bridges, such as $\mathbb{K}_4^{SS}$ between H57 and D102. Polar and nonpolar side chains, though, do not mix in the same prominent modes, demonstrating mechanical coupling separation due to chemical differences. Mostly locating in β strands, the eigenvectors populated by hydrophobic residues tend to involve more mechanically linked partners and can still emerge as prominent modes even the individual k_{IJ}^{SS} values

A $\overline{\mathbb{K}}_{\alpha'}^{SS}$

α'	$\lambda_{\alpha'}$	Residues
$1,11^{SS}$	164.9,15.8	E70-E77-E80-Ca^{2+}
$2,3^{SS}$	52.8,37.0	E70-E77-E80
13^{SS}	14.7	S32-**H40**
18^{SS}	11.8	L33-L46-V66
$10^{SS},17^{SS}$	16.2,12.5	V31-**L68**-V52-I106
20^{SS}	11.6	V53-L105-I238
$5^{SS},16^{SS}$	21.8,12.7	W237-**H91**-**Y234**
4^{SS}	26.9	**H57-D102**
12^{SS}	15.7	**H57**-S214-T229
$15^{SS},21^{SS}$	13.3,11.1	**D102-A56-H57-S195**
23^{SS}	8.8	**A55**-C58
14^{SS}	13.5	**Q30**-S139
22^{SS}	10.2	Y29-**P198**-L209
19^{SS}	11.7	E151-R17^{BPTI}
8^{SS}	18.8	**W141-L155**
$6^{SS},9^{SS}$	20.0,18.4	I16-D194-V213-Y228
7^{SS}	19.4	**W215**-V227

(kcal/mol/Å²)

B $\overline{\mathbb{K}}_{\alpha'}^{BS}$

α'	$\lambda_{\alpha'}$	Residues
2^{BS}	46.9	E77-E80
8^{BS}	31.3	G69-R117
9^{BS}	31.2	**W51**-I238
7^{BS}	31.8	S45-**S54**-G196-V213
11^{BS}	27.2	H71-**W141-L155**
5^{BS}	37.4	S190-K15^{BPTI}
$1^{BS},6^{BS}$	76.6,34.2	I16-D194-C191
3^{BS}	43.6	N143-N150
10^{BS}	30.5	S146-C220-A221-D189
4^{BS}	40.1	**Y172-P225**
12^{BS}	23.7	C128-Q210

(kcal/mol/Å²)

Number of RT's prominent mode residues in		BB	SS	BS
	β/α interior	20	19	4
	β/α periphery	9	14	6
	loop	5	8	17

Fig. 6. In RT, the prominent mechanical couplings during protein dynamics involving a side chain. The prominent modes in (A) $\mathbb{K}^{SS}$ and (B) $\mathbb{K}^{BS}$ rigidity graphs. The BPTI inhibitor is in brown ribbon. The RT residues in hydrogen bonds and salt bridges are in licorice and those in hydrophobic couplings are in ball-and-stick. The index, eigenvalue, and residues of prominent modes are listed. The prominent mode residues are hotspots in the mechanical coupling network, and the numbers within a secondary structure, β/α interior, at a secondary structure periphery, β/α periphery, or in a loop are reported.

may be lower. Comparing to $\overline{\mathbb{K}}^{BB}$, the $\overline{\mathbb{K}}^{SS}$ prominent modes have higher percentages of periphery and loop residues, Fig. 6B. As for $\overline{\mathbb{K}}^{BS}$, the prominent modes are scattered hydrogen bonds primarily in loops, some at secondary structure peripheries, but very few within a secondary structure. In RT, the very strong $\overline{\mathbb{K}}^{BS}$ couplings primarily occur in the CT barrel that contains the activation domain, Fig. 1A and Fig. 6B. Backbone and side-chain mechanical couplings thus exhibit specific patterns in the structure.

The above results of backbone and side chains having distinct mechanical coupling networks provide molecular basis for their separate adjustability as empirically adopted in protein engineering and design [9–11]. Whether backbone or side chains are more important in shaping the folding funnel is also an unresolved debate [3–6]. Rather than lumping each residue as a single unit [12–17], our strategy is explicit representation of backbone and side chains. With their rigidity graphs computed from protein dynamics, this framework provides a refined way for delineating the free-energy landscape around the structure. Next, whether the mechanical couplings would exhibit extended patterns for understanding protein allostery is addressed.

3.4. Emergence of extensive mechanical couplings

The RT rigidity graphs reveal that the prominent couplings between RT and BPTI lie in the hydrogen bonding modes $\overline{\mathbb{K}}_9^{BS}$, $\overline{\mathbb{K}}_5^{BS}$, and $\overline{\mathbb{K}}_{19}^{SS}$ (Fig. 5 and Fig. 6). They are next to several prominent modes within the trypsin fold, including $\overline{\mathbb{K}}_4^{SS}$, $\overline{\mathbb{K}}_{12}^{SS}$, $\overline{\mathbb{K}}_{17}^{BB}$ of the triad, and the C42-C58 disulfide bond at the S1' site. With such spatial arrangement, the prominent modes $\overline{\mathbb{K}}_{15}^{SS}$ and $\overline{\mathbb{K}}_{21}^{SS}$ indeed come out as long-range mechanical couplings containing BPTI A16, S1' site C42, the catalytic triad, oxyanion hole, A56, and S1 site S214 and T229, Fig. 6A. From the activation domain to active site, Fig. 5B, extensive prominent modes also emerge as $\overline{\mathbb{K}}_1^{BS}$, $\overline{\mathbb{K}}_5^{BS}$, and $\overline{\mathbb{K}}_{10}^{BS}$, Fig. 6B. A mystery of the serine protease family is that substrate variation or mutation at sites away from the triad still impact the k_{cat}/K_M of cleavage [56–60]. Our result is a first demonstration that under thermal noise, specific molecular interactions can integrate into significant mechanical signals across distal sites.

During the 5 μs all-atom trajectory of PDZ3, the mechanical hotspots are also captured as the significantly populated residues in the prominent modes, Figs. S12–S15. Similarly, the backbone

and side-chain mechanical coupling networks of PDZ3 exhibit different patterns, Figs. S16–S17. Most BB prominent modes are in the β-sandwich with certain extensive patterns like $\bar{\mathbb{K}}_9^{BB}$. The prominent BS modes of PDZ3, on the other hand, are more scattered and contain the extensive $\bar{\mathbb{K}}_1^{BS}$, $\bar{\mathbb{K}}_3^{BS}$, and $\bar{\mathbb{K}}_7^{BS}$ that link the β-sandwich and CT-extension. Most of the prominent SS modes in PDZ3 are hydrophobic and rather extensive, while the $\bar{\mathbb{K}}_1^{SS}$ salt bridge at the β-sandwich facing CT-extension is exceeding strong, Fig. S17. Similar to RT, the prominent mode residues of PDZ3 $\bar{\mathbb{K}}^{SS}$ have a significantly higher percentage at secondary structure peripheries, such as the two residues of $\bar{\mathbb{K}}_1^{SS}$, than those of $\bar{\mathbb{K}}^{BB}$ do. The mechanical hotspots of PDZ3 $\bar{\mathbb{K}}^{BS}$ also very frequently occur at secondary structure peripheries (Fig. S17) rather than in loops as in the case of RT (Fig. 6B). While consistent overall patterns in mechanical coupling networks are observed, the two protein systems exhibit specific features in their prominent modes of rigidity graphs.

At the core of allosteric communication in proteins is the physical interactions that are persistent under thermal noise for connecting distal sites [4–6]. In attempting to capture such functionally important long-range couplings, many approaches are based on positional covariance [18–20], structural contacts [21–23], or sequence co-evolution [38–42]. However, a fundamental difficulty is that the observed signals do not necessarily correspond to molecular interactions. For example, in using low-frequency vibrational modes to study intra-protein communication, positional fluctuations of unconnected, distal residues can be highly correlated due to the structural topology [12–14]. From

protein dynamics, our computational framework of identifying the prominent modes of $\mathbb{K}$ rigidity graphs thus provides a way to capture the molecularly specific patterns that can survive the stochastic fluctuations. In both the RT and PDZ3 protein systems, extensive mechanical couplings composed of physical interactions are identified.

3.5. Residue rigidity scores in backbone and side chains during protein dynamics as metrics for biological functions

Being the strong mechanical couplings persistent through protein dynamics, the prominent modes of $\bar{\mathbb{K}}^{BB}$, $\bar{\mathbb{K}}^{BS}$, and $\bar{\mathbb{K}}^{SS}$ likely have important implications in biological functions. They also represent the routes through which protein backbone and side chains are wired in the mechanical coupling network. Based on this mechanistic insight, we propose to quantify the biological importance of residues by deducing the residue rigidity scores in $\bar{\mathbb{K}}^{BB}$, $\bar{\mathbb{K}}^{BS}$, and $\bar{\mathbb{K}}^{SS}$. Since their eigenvectors exhibit pointed patterns, each residue is specifically populated in few modes. The residue rigidity score for I in a particular rigidity graph thus comes from the characteristic eigenvector $\mathcal{I}'$ that the residue is most representative (see Eq. 1 discussed in 2.6). The residue rigidity score κ_I is then the mechanical strength weighted by the averaged content of mode $\mathcal{I}'$ during protein dynamics, $\langle r_{\mathcal{I}'} \rangle \lambda_{\mathcal{I}'}$. By putting together the results of $\bar{\mathbb{K}}^{BB}$, $\bar{\mathbb{K}}^{BS}$, and $\bar{\mathbb{K}}^{SS}$, the residue rigidity score in backbone, κ_I^B, is the largest mechanical contribution from the backbone of residue I, and κ_I^S is that from its side chain. Therefore, if residue I plays a significant role in a prominent mode in $\bar{\mathbb{K}}^{BB}$, $\bar{\mathbb{K}}^{BS}$, and/or $\bar{\mathbb{K}}^{SS}$, it would
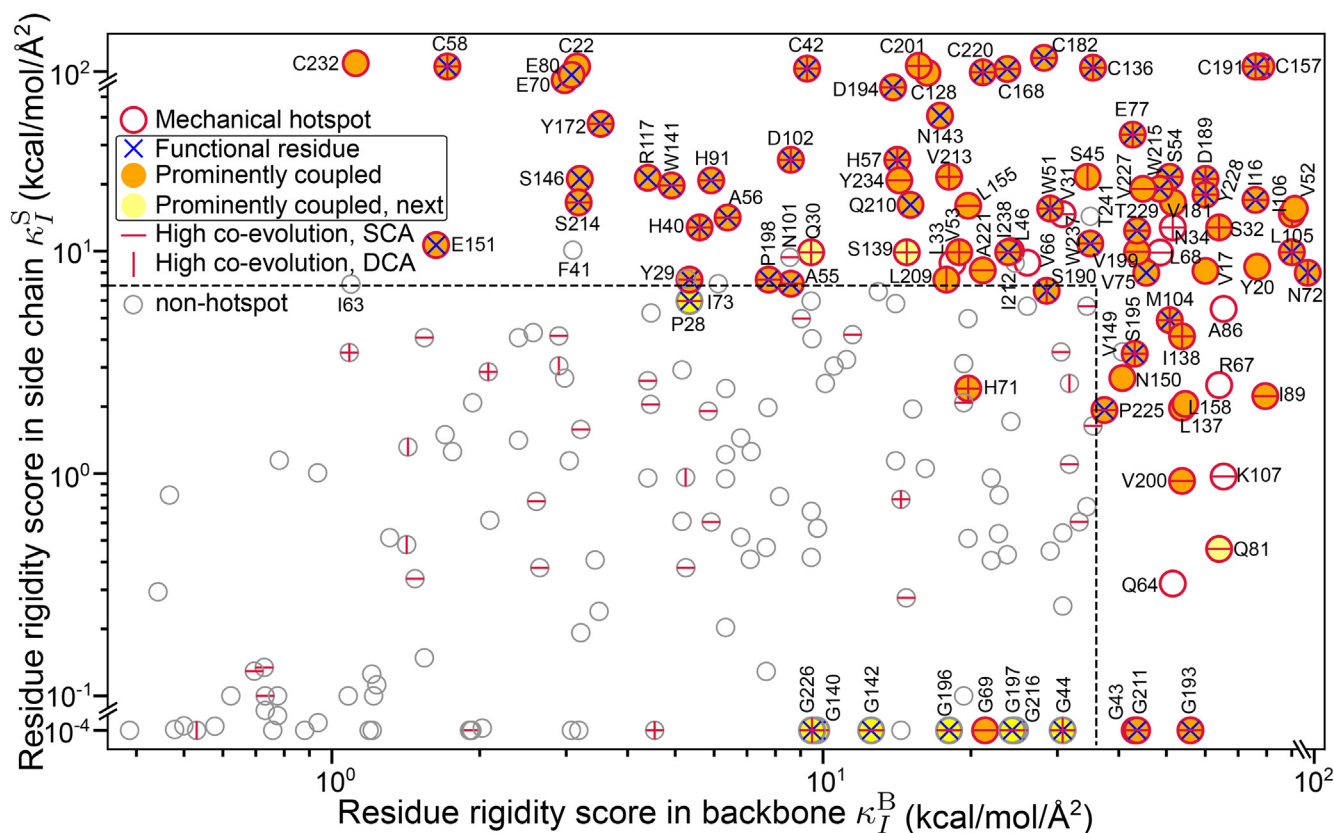

Fig. 7. For RT, the residue rigidity score during protein dynamics in backbone (κ_I^B, x-axis) and in side chain (κ_I^S, y-axis) for residues indexed by I. Since glycine residues do not have a side-chain score, they are provided with a minimal κ_I^S and hence locate at the bottom. Red circles denote mechanical hotspots: the significantly weighted residues in the prominent modes of rigidity graphs. Dashed lines are the hotspot boundary due to $\bar{\mathbb{K}}^{BB}$ and $\bar{\mathbb{K}}^{SS}$ prominent modes, and mechanical hotspots are mostly enclosed in the band except G69, H71, and S190 that contribute backbone to the prominent $\bar{\mathbb{K}}^{BS}$ modes. Labelled ✕ as "functional residue" are those listed in Table S2 that have experimentally verified function or ultra high conservation in MSA. Filled orange as "prominently coupled" are the mechanical hotspots of a single rigidity graph mode that contain least one functional residue. Filled yellow as "prominently coupled, next" are the hotspots next to a functional residue in sequence or non-hotspot functional residues next to a hotspot. A residue is labeled — or | for high co-evolution in SCA or DCA, respectively. Grey circles are non-hotspot residues.

have high κ_I^B and/or κ_I^S. On the other hand, if a residue only appears in modes having low strength and/or averaged content, it would have low residue rigidity scores. The residue rigidity scores of RT backbone and side chains in the all-atom MD simulation are shown in Fig. 7; glycine residues listed in the bottom are provided with a minimal κ_I^S as they do have side chain and hence the corresponding score.

In contributing to the mechanical coupling network, Fig. 7 illustrates that the protein amino acids have diverse residue rigidity scores in backbone and side chains. Mechanical hotspots are the significantly populated residues in the prominent mode of rigidity graphs as defined in 2.6, and the dashed lines in Fig. 7 are the boundaries due to $\mathbb{K}^{BB}$ and $\mathbb{K}^{SS}$. The general importance of the trypsin fold in RT has led to a variety of functional characterization including mutagenesis at different sites. Combining the residues with experimentally verified function and ultra high conservation in MSA provides the functional residues of RT as summarized in Table S2 and marked on Fig. 7. It can be seen that the mechanical hotspots cover most of the functional residues as well as their prominently coupled associates in the rigidity graphs. Although glycine residues do not have a side chain are often flexible, few of them still emerge as mechanical hotspots through backbone such as the oxyanion hole G193. Even though several functional glycines are not in a prominent mode of the rigidity graphs, their signature in the networks is being sequence neighbors of mechanical hotspots, Fig. 7. While the rigidity scores are based on non-skeleton springs, sequence neighbors are prominently coupled through skeleton connections. Therefore, the "prominently coupled, next" category that the aforementioned glycine residues reside also includes the mechanical hotspots that are next to functional residues. Moreover, many mechanical hotspots exhibit strong signals in SCA and/or DCA [39,40]; Table S3 summarizes the residues having high co-evolution as marked in Fig. 7. For example, the residues in the aforementioned $\mathbb{K}_{15}^{SS}$ and $\mathbb{K}_{21}^{SS}$ prominent modes that form a spatially extensive set of mechanical couplings are all in a SCA sector, which also includes the mechanical hotspots in $\mathbb{K}_1^{BS}$ (Fig. 6 and Table S3).

From the 5 μs all-atom MD data of PDZ3, the mechanical hotspots covering most of its functional residues [38] is also observed in its κ_I^B-κ_I^S plot, Fig. S18. The mechanical hotspot F325 important for substrate recognition has been shown to co-evolve with another hotspot H372 at a distal site with A347 and L353 on the communication pathway, but the underlying physical interactions are unclear [61]. As a molecular mechanism for this SCA-based prediction, Fig. S17 shows that F325 is in a cluster of prominent hydrophobic modes ($\mathbb{K}_{2-5,7}^{SS}$) that together with $\mathbb{K}_6^{BS}$ mechanically link the residue with H372. One of the mechanical hotspots in these modes is I341, which has been proposed as an alternative route in a thermal-diffusion MD study [62]. Our rigidity graph analysis based on all-atom MD simulations thus offers a unified mechanistic picture for the various data on intra-PDZ3 communication.

Although most of the characterized residues are in the β-sandwich [38] of PDZ3, several mechanical hotspots are found at the interface contacting CT-extension. For example, D357 coupling to Y392 at the interface in $\mathbb{K}_1^{BS}$ is the most conserved residues in the β-sandwich [61]. It would thus be valuable to specifically examine the functional roles of such residues in inter-domain communication [63,64]. Overall, the residue rigidity scores in backbone and side chains are very useful metrics for the functional importance of RT and PDZ3 sites. This establishment opens a new door for using molecular simulation to study the mechanistic basis of biological activities and evolutionary restraints.

4. Conclusions

Given a protein fold, residue contact numbers center around a value due to the packing density. But then, what is the manifestation of sequence specificities in the structure? This question is addressed here by developing a bsENM with structure-mechanics statistical learning to compute the elastic parameters from 5 μs all-atom MD simulation in explicit solvent. To analyze the network behaviors of the complicated molecular interactions in structural fluctuations, the newly devised graph-theoretic framework introduces the concept of protein rigidity graphs. A key discovery is that the chemical details during protein dynamics render scale-free network properties in the mechanical coupling strengths of both backbone and side chains. In the nano-scale network of a single protein, exhibition of small-world-like features has not been shown to the best of our knowledge. The significantly populated residues in the statistically prominent modes of rigidity graphs are thus recognized as mechanical hotspots.

Furthermore, our bsENM-graph approach enables the direct comparison of backbone and side-chain mechanical couplings to accentuate their differences. Such outcomes point to an important notion that protein residues have diverse combinations of backbone and side-chain contributions to the mechanical coupling network as seen in the κ_I^S-κ_I^B plot of RT (Fig. 7) and PDZ3 (Fig. S18). Encouragingly, functional residues of the two protein systems are largely mechanical hotspots themselves or next to one in sequence as for glycine. While most functional residues have high residue rigidity scores for their side chains, some also have prominent backbone couplings as in the top-right corner of Fig. 7 and Fig. S18. Only a specific set of sites having top residue rigidity scores in both the side chains and backbone indicates sophisticatedly tuned interaction network and has implications in shaping the folding funnel and in rendering proper conformational flexibilities for function. Another finding is that a significant portion of side-chain related mechanical hotspots locate at secondary peripheries (Fig. 6 and Fig. S17), i.e., the edges of foldons [65], and are potentially important factors in adopting the foldon inspired models [65,66] for protein folding. Although the all-atom MD data depend on empirical force fields, specific conditions, and duration, the structure-mechanics statistical learning and graphical analysis schemes as well as the concepts therein can be readily applied to different cases. The mechanical hotspots and prominent rigidity graph modes identified in molecular simulation also bear similarities with the co-evolution in MSA, which also suffers from statistical noises. For RT, the physically contiguous residues in $\mathbb{K}_{15,21}^{SS}$ and $\mathbb{K}_{1,10}^{BS}$ also exhibit prominent co-evolution signals in MSA as a single sector, Fig. 6 and Table S3. For PDZ3, the multiple prominent modes involving F325 ($\mathbb{K}_{2-5}^{SS}$) primarily involve sector residues, Fig. S18. This work thus suggests a molecular mechanism for the coupled sequence variation due to evolutionary restraints, that it be associated with having prominent patterns in the protein mechanical coupling network.

CRediT authorship contribution statement

Nixon Raj: Conceptualization, Methodology, Software, Writing - original draft. **Timothy Click:** Methodology, Software. **Haw Yang:** Conceptualization, Writing - review & editing. **Jhih-Wei Chu:** Conceptualization, Methodology, Software, Writing - original draft, Writing - review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.csbj.2021.09.004.

References

[1] Dill KA. Dominant forces in protein folding. Biochemistry 1990;29 (31):7133–55. https://doi.org/10.1021/bi00483a001.

[2] Onuchic JN, Wolynes PG. Theory of protein folding. Curr Opin Struct Biol 2004;14(1):70–5. https://doi.org/10.1016/j.sbi.2004.01.009.

[3] Rose GD, Fleming PJ, Banavar JR, Maritan A. A backbone-based theory of protein folding. Proc Natl Acad Sci USA 2006;103(45):16623–33. https://doi.org/10.1073/pnas.0606843103.

[4] Mittermaier AK, Kay LE. Observing biological dynamics at atomic resolution using NMR. Trends Biochem Sci 2009;34(12):601–11. https://doi.org/10.1016/j.tibs.2009.07.004.

[5] Nussinov R, Tsai C-J. Allostery in disease and in drug discovery. Cell 2013;153 (2):293–305. https://doi.org/10.1016/j.cell.2013.03.034.

[6] Lee AL. Contrasting roles of dynamics in protein allostery: NMR and structural studies of CheY and the third PDZ domain from PSD-95. Biophys Rev 2015;7 (2):217–26. https://doi.org/10.1007/s12551-015-0169-3.

[7] Dawson NL, Lewis TE, Das S, Lees JG, Lee D, Ashford P, Orengo CA, Sillitoe I. CATH: an expanded resource to predict protein function through structure and sequence. Nucl Acids Res 2017;45(D1):D289–95. https://doi.org/10.1093/nar/gkw1098.

[8] Chandonia J-M, Fox NK, Brenner SE. SCOPe: classification of large macromolecular structures in the structural classification of proteins-extended database. Nucl Acids Res 2019;47(D1):D475–81. https://doi.org/10.1093/nar/gky1134.

[9] Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, St Clair JL, et al. Computational design of an enzyme catalyst for a stereoselective bimolecular diels-alder reaction. Science 2010;329(5989):309–13. https://doi.org/10.1126/science.1190239.

[10] Kazlauskas R. Engineering more stable proteins. Chem Soc Rev 2018;47 (24):9026–45. https://doi.org/10.1039/C8CS00014J.

[11] Nisthal A, Wang CY, Ary ML, Mayo SL. Protein stability engineering insights revealed by domain-wide comprehensive mutagenesis. Proc Natl Acad Sci USA 2019;116(33):16367–77. https://doi.org/10.1073/pnas.1903888116.

[12] Tirion. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys Rev Lett 1996;77(9):1905–8. doi:10.1103/PhysRevLett.77.1905. .

[13] Haliloğlu T, Bahar I, Erman B. Gaussian dynamics of folded proteins. Phys Rev Lett 1997;79(16):3090–3. https://doi.org/10.1103/PhysRevLett.79.3090.

[14] Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 2001;80(1):505–15. https://doi.org/10.1016/S0006-3495(01)76033-X.

[15] Taketomi H, Ueda Y, Gō N. Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. Int J Pept Protein Res 1975;7(6):445–59. https://doi.org/10.1111/j.1399-3011.1975.tb02465.x.

[16] Gō N. Theoretical studies of protein folding. Annu Rev Biophys Bioeng 1983;12:183–210. https://doi.org/10.1146/annurev.bb.12.060183.001151.

[17] Takada S. Gō model revisited. Biophys Physicobiol 2019;16:248–55. https://doi.org/10.2142/biophysico.16.0_248.

[18] Zheng WJ, Brooks BR, Thirumalai D. Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations. Proc Natl Acad Sci USA 2006;103(20):7664–9. https://doi.org/10.1073/pnas.0510426103.

[19] Li H, Chang Y-Y, Lee JY, Bahar I, Yang L-W. DynOmics: dynamics of structural proteome and beyond. Nucl Acids Res 2017;45(W1):W374–80. https://doi.org/10.1093/nar/gkx385.

[20] Thirumalai D, Hyeon C, Zhuravlev PI, Lorimer GH. Symmetry, rigidity, and allosteric signaling: from monomeric proteins to molecular machines. Chem Rev 2019;119(12):6788–821. https://doi.org/10.1021/acs.chemrev.8b00760.

[21] Brinda KV, Vishveshwara S. A network representation of protein structures: implications for protein stability. Biophys J 2005;89(6):4159–70. https://doi.org/10.1529/biophysj.105.064485.

[22] Di Paola L, De Ruvo M, Paci P, Santoni D, Giuliani A. Protein contact networks: an emerging paradigm in chemistry. Chem Rev 2013;113(3):1598–613. https://doi.org/10.1021/cr3002356.

[23] Kayikci M, Venkatakrishnan AJ, Scott-Brown J, Ravarani CNJ, Flock T, Babu MM. Visualization and analysis of non-covalent contacts using the protein contacts atlas. Nat Struct Mol Biol 2018;25(2):185–94. https://doi.org/10.1038/s41594-017-0019-z.

[24] Kannan N, Vishveshwara S. Identification of side-chain clusters in protein structures by a graph spectral method. J Mol Biol 1999;292(2):441–64. https://doi.org/10.1006/jmbi.1999.3058.

[25] Pavlopoulos GA, Secrier M, Moschopoulos CN, Soldatos TG, Kossida S, Aerts J, et al. Using graph theory to analyze biological networks. BioData Min. 4 (UNSP 10). doi:10.1186/1756-0381-4-10. .

[26] Ma J. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. Structure 2005;13(3):373–80. https://doi.org/10.1016/j.str.2005.02.002.

[27] Bahar I, Lezon TR, Bakan A, Shrivastava IH. Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. Chem Rev 2010;110(3):1463–97. https://doi.org/10.1021/cr900095e.

[28] Vashisth H, Brooks ICL. Conformational sampling of maltose-transporter components in cartesian collective variables is governed by the low-frequency normal modes. J Phys Chem Lett 2012;3(22):3379–84. https://doi.org/10.1021/jz301650q.

[29] López-Blanco JR, Chacón P. New generation of elastic network models. Curr Opin Struct Biol 2016;37:46–53. https://doi.org/10.1016/j.sbi.2015.11.013.

[30] Chu J-W, Voth GA. Coarse-grained modeling of the actin filament derived from atomistic-scale simulations. Biophys J 2006;90(5):1572–82. https://doi.org/10.1529/biophysj.105.073924.

[31] Click T, Raj N, Chu J-W. Calculation of enzyme fluctuograms from all-atom molecular dynamics simulation. In: Voth GA, editor. Computational approaches for studying enzyme mechanism Part B, vol. 578 of Meth. Enzymol. Academic Press; 2016. Ch. 14. p. 327–42. doi:10.1016/bs.mie.2016.05.024. .

[32] Chen Y-T, Yang H, Chu J-W. Structure-mechanics statistical learning unravels the linkage between local rigidity and global flexibility in nucleic acids. Chem Sci 2020;11(19):4969–79. https://doi.org/10.1039/D0SC00480D.

[33] Pasternak A, Ringe D, Hedstrom L. Comparison of anionic and cationic trypsinogens: The anionic activation domain is more flexible in solution and differs in its mode of BPTI binding in the crystal structure. Protein Sci 1999;8 (1):253–8. https://doi.org/10.1110/ps.8.1.253.

[34] Page MJ, Di Cera E. Serine peptidases: classification, structure and function. Cell Mol Life Sci 2008;65(7–8):1220–36. https://doi.org/10.1007/s00018-008-7565-9.

[35] Lee H-J, Zheng JJ. PDZ domains and their binding partners: structure, specificity, and modification. Cell Commun Signal 8(8). doi:10.1186/1478-811X-8-8. .

[36] Hung AY, Sheng M. PDZ domains: structural modules for protein complex assembly. J Biol Chem 2002;277(8):5699–702. https://doi.org/10.1074/jbc.R100065200.

[37] Doyle DA, Lee A, Lewis J, Kim E, Sheng M, MacKinnon R. Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. Cell 1996;85(7):1067–76. https://doi.org/10.1016/s0092-8674(00)81307-0.

[38] McLaughlin J, Richard N, Poelwijk FJ, Raman A, Gosal WS, Ranganathan R. The spatial architecture of protein function and adaptation. Nature 2012;491 (7422):138–U163. doi:10.1038/nature11500. .

[39] Halabi N, Rivoire O, Leibler S, Ranganathan R. Protein sectors: evolutionary units of three-dimensional structure. Cell 2009;138(4):774–86. https://doi.org/10.1016/j.cell.2009.07.038.

[40] Rivoire O, Reynolds KA, Ranganathan R. Evolution-based functional decomposition of proteins. PLoS Comput Biol 2016;12(6): . https://doi.org/10.1371/journal.pcbi.1004817e1004817.

[41] Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, Zecchina R, Onuchic JN, Hwa T, Weigt M. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proc Natl Acad Sci USA 2011;108(49):E1293–301. https://doi.org/10.1073/pnas.1111471108.

[42] Cocco S, Monasson R, Weigt M. From principal component to direct coupling analysis of coevolution in proteins: low-eigenvalue modes are needed for structure prediction. PLoS Comput Biol 9(8). doi:10.1371/journal.pcbi.1003176. .

[43] Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, et al. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi1 and chi2 dihedral angles. J Chem Theory Comput 2012;8(9):3257–73. https://doi.org/10.1021/ct300400x.

[44] Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. GROMACS: High performance molecular simulations through multi-level parallelism from

laptops to supercomputers. SoftwareX 2015;1–2(9):19–25. https://doi.org/10.1016/j.softx.2015.06.001.

[45] Albert R, Jeong H, Barabási A-L. Diameter of the World-Wide Web. Nature 1999;401(6749):130–1. https://doi.org/10.1038/43601.

[46] Strogatz SH. Exploring complex networks. Nature 2001;410(6825):268–76. https://doi.org/10.1038/35065725.

[47] Jeong H, Mason SP, Barabási AL, Oltvai ZN. Lethality and centrality in protein networks. Nature 2001;411(6833):41–2. https://doi.org/10.1038/35075138.

[48] Albert R, Barabási A-L. Statistical mechanics of complex networks. Rev Mod Phys 2002;74(1):47. https://doi.org/10.1103/RevModPhys.74.47.

[49] Caldarelli G, Capocci A, De Los Rios P, Muñoz MA. Scale-free networks from varying vertex intrinsic fitness. Phys Rev Lett 2002;89(25): . https://doi.org/10.1103/PhysRevLett.89.258702258702.

[50] Cimini G, Squartini T, Saracco F, Garlaschelli D, Gabrielli A, Caldarelli G. The statistical physics of real-world networks. Nat Rev Phys 2019;1(1):58–71. https://doi.org/10.1038/s42254-018-0002-6.

[51] Lomax KS. Business failures: another example of the analysis of failure data. J Am Stat Assoc 1954;49:847–52. https://doi.org/10.2307/2281544.

[52] Thyagarajan B, Bloom JD. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. eLife 3. .

[53] Fuller ZL, Berg JJ, Mostafavi H, Sella G, Przeworski M. Measuring intolerance to mutation in human genetics. Nat Genet 2019;51(5):772–6.

[54] Craik CS, Page MJ, Madison EL. Proteases as therapeutics. Biochem J 2011;435:1–16. https://doi.org/10.1042/BJ20100965.

[55] Batt AR, St Germain CP, Gokey T, Guliaev AB, Baird Jr T. Engineering trypsin for inhibitor resistance. Protein Sci 2015;24(9):1463–74. https://doi.org/10.1002/pro.2732.

[56] Sprang SR, Fletterick RJ, Gráf L, Rutter WJ, Craik CS. Studies of specificity and catalysis in trypsin by structural analysis of site-directed mutants. Crit Rev Biotechnol 1988;8(3):225–36. https://doi.org/10.3109/07388558809147559.

[57] Evnin LB, Vásquez JR, Craik CS. Substrate specificity of trypsin investigated by using a genetic selection. Proc Natl Acad Sci USA 1990;87(17):6659–63. https://doi.org/10.1073/pnas.87.17.6659.

[58] Vindigni A, Di Cera E. Role of P225 and the C136–C201 disulfide bond in tissue plasminogen activator. Protein Sci 1998;7(8):1728–37. https://doi.org/10.1002/pro.5560070807.

[59] Krem MM, Prasad S, Di Cera E. Ser214 is crucial for substrate binding to serine proteases. J Biol Chem 2002;277(43):40260–4. https://doi.org/10.1074/jbc.M206173200.

[60] Hedstrom L. Serine protease mechanism and specificity. Chem Rev 2002;102 (12):4501–23. https://doi.org/10.1021/cr000033x.

[61] Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. Science 1999;286(5438):295–9. https://doi.org/10.1126/science.286.5438.295.

[62] Ota N, Agard DA. Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion. J Mol Biol 2005;351(2):345–54.

[63] Zhang J, Petit CM, King DS, Lee AL. Phosphorylation of a PDZ domain extension modulates binding affinity and interdomain interactions in postsynaptic density-95 (PSD-95) protein, a membrane-associated guanylate kinase (MAGUK). J Biol Chem 2011;286(48):41776–85. https://doi.org/10.1074/jbc.M111.272583.

[64] Zhang J, Lewis SM, Kuhlman B, Lee AL. Supertertiary structure of the MAGUK core from PSD-95. Structure 2013;21(3):402–13. https://doi.org/10.1016/j.str.2012.12.014.

[65] Englander SW, Mayne L, Krishna MMG. Protein folding and misfolding: mechanism and principles. Quart Rev Biophys 2007;40(4):287–326. https://doi.org/10.1017/S0033583508004654.

[66] Rollins GC, Dill KA. General mechanism of two-state protein folding kinetics. J Am Chem Soc 2014;136:11420–7. https://doi.org/10.1021/ja5049434.