

Cite this: *Chem. Sci.*, 2022, 13, 3688

All publication charges for this article have been paid for by the Royal Society of Chemistry

Received 8th November 2021  
Accepted 10th January 2022

DOI: 10.1039/d1sc06184d

rsc.li/chemical-science

# Structure-mechanics statistical learning uncovers mechanical relay in proteins†

Nixon Raj, <sup>a</sup> Timothy H. Click, <sup>a</sup> Haw Yang <sup>b</sup> and Jih-Wei Chu <sup>\*c</sup>

A protein's adaptive response to its substrates is one of the key questions driving molecular physics and physical chemistry. This work employs the recently developed structure-mechanics statistical learning method to establish a mechanical perspective. Specifically, by mapping all-atom molecular dynamics simulations onto the spring parameters of a backbone-side-chain elastic network model, the chemical moiety specific force constants (or mechanical rigidity) are used to assemble the rigidity graph, which is the matrix of inter-residue coupling strength. Using the S1A protease and the PDZ3 signaling domain as examples, chains of spatially contiguous residues are found to exhibit prominent changes in their mechanical rigidity upon substrate binding or dissociation. Such a mechanical-relay picture thus provides a mechanistic underpinning for conformational changes, long-range communication, and inter-domain allostery in both proteins, where the responsive mechanical hotspots are mostly residues having important biological functions or significant mutation sensitivity.

## 1 Introduction

Upon binding a specific molecule such as an inhibitor or a substrate, the interaction network in a protein is expected to self-adjust for the biological function.<sup>1–3</sup> Even without a conformational change, reorganization can proceed as variations in its dynamical fluctuations.<sup>4–6</sup> Although adaptive responses to molecular binding have long been anticipated from structural, biochemical, and spectroscopic measurements,<sup>7–9</sup> this functionally very important property is mostly understood phenomenologically. Quantitative metrics based on physical interactions are lacking due to the difficulties of observing molecular scale processes in a complex system. For example, the respective roles of backbone and side chains in open/close structural transitions that gate an enzyme active site are yet to be resolved.<sup>10–12</sup>

To understand protein reorganization in terms of molecular interactions, a mechanical-coupling dynamics perspective is taken here. According to the equipartition principle in statistical mechanics—the extent of fluctuation inversely reflects the interaction strength—a structure-mechanics statistical learning scheme can be developed to compute the parameters of a coarse-grained (CG) backbone and side-chain elastic network model

(bsENM) from the configurations in an all-atom molecular dynamics (MD) trajectory.<sup>13,14</sup> Using the calculated elastic constants as edge weights between residue nodes gives the rigidity graph of protein dynamics, Fig. 1(a). The mechanical couplings of both backbone and side chains have been shown to exhibit scale-free network properties,<sup>14</sup> indicating that only a fraction of structural contacts exhibit strong interactions. Over the  $\mu\text{s}$  time-scale MD conducted in explicit solvent, the statistically prominent eigenmodes of a protein rigidity graph reveal the specific molecular interaction patterns that persist through stochastic noises.

The unique network features of inter-residue interaction strengths uncovered at a particular state of substrate binding<sup>14</sup> provide a remarkable opportunity for learning the mechanism of protein reorganization by avoiding the difficulties seen in covariance of structural fluctuations<sup>15–17</sup> or sequence changes,<sup>18–20</sup> in which the signals oftentimes do not come from physical interactions. For example, the apparent correlation between distal residues in soft vibrational modes is largely influenced by the topological shape of the native fold.<sup>21–23</sup> Inferring the coupling strengths from the correlated fluctuation of each residue pair<sup>24</sup> is also complicated by its intricate connections in the structure. Collective consideration of the other degrees of freedom, such as by the self-consistent iterations in structure-mechanics statistical learning,<sup>14</sup> is necessary to quantify the sparse mechanical coupling network. In this case, the prominent eigenmodes of rigidity graphs correspond to persistent restraints coming from hydrogen bonding, electrostatic couplings, and hydrophobic contacts.<sup>14</sup> Therefore, comparing the rigidity graphs from the protein dynamics of different molecular binding states is an appealing strategy to study protein reorganization.

<sup>a</sup>Institute of Bioinformatics and Systems Biology, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan, Republic of China

<sup>b</sup>Department of Chemistry, Princeton University, Princeton, NJ 08544, USA

<sup>c</sup>Institute of Bioinformatics and Systems Biology, Department of Biological Science and Technology, Institute of Molecular Medicine and Bioengineering, Center for Intelligent Drug Systems and Smart Bio-devices (IDS<sup>2</sup>B), National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan, Republic of China. E-mail: jwchu@nctu.edu.tw

† Electronic supplementary information (ESI) available: Supporting figures in protein-reorganization. See DOI: 10.1039/d1sc06184d



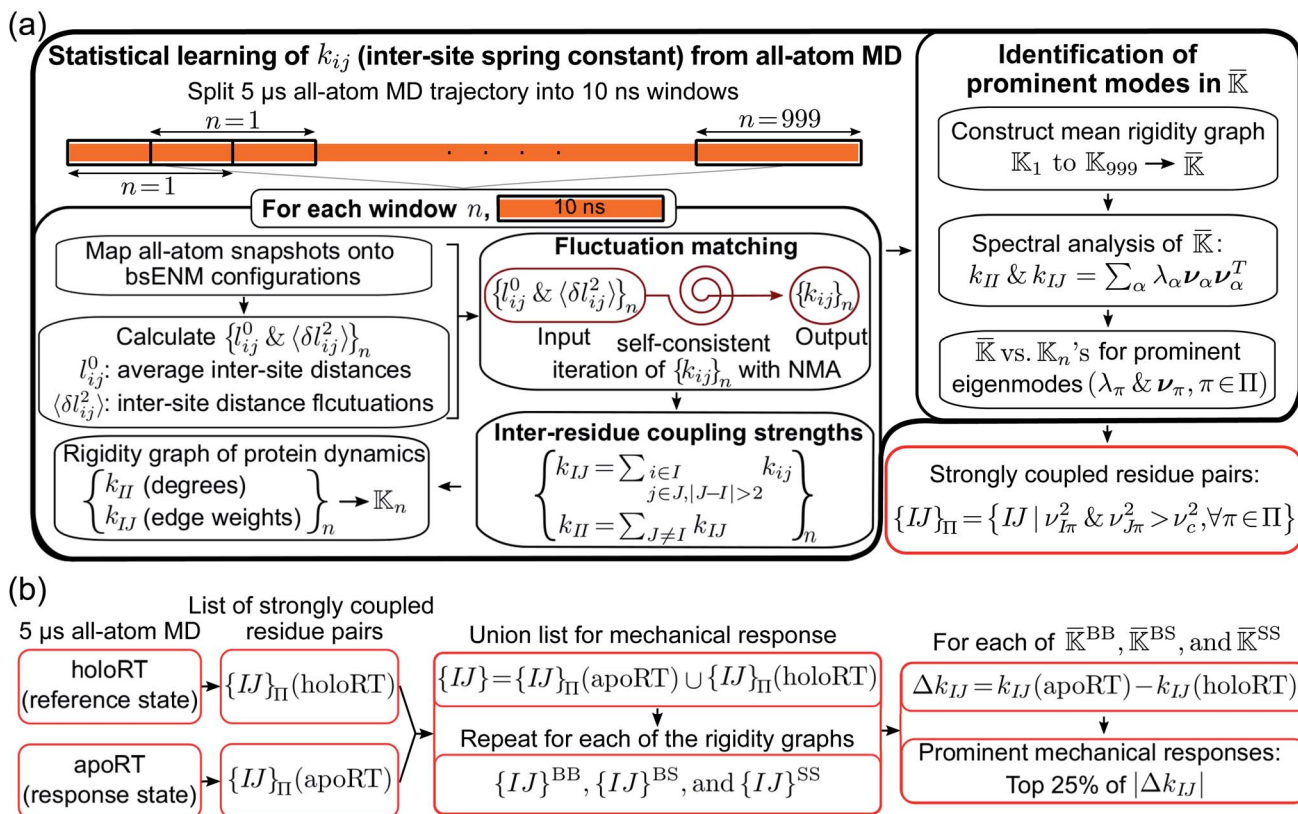


Fig. 1 Flowcharts for statistical learning of protein reorganization upon substrate binding or dissociation. (a) Computation of inter-site spring constants ( $k_{ij}$ 's) in the backbone-side-chain elastic network model (bsENM) from all-atom MD, construction of rigidity graphs with  $k_{ij}$ 's, identification of their statistically prominent eigenmodes ( $\Pi = \{\pi\}$ ), and generation for the list of strongly coupled residue pairs,  $\{IJ\}_{\Pi}$ , from  $\pi \in \Pi$  (red box). (b) Construction of a union list  $\{IJ\}$  from the  $\{IJ\}_{\Pi}$  at different substrate binding states. Given the union lists of  $\mathbb{K}^{\text{BB}}, \mathbb{K}^{\text{BS}}, \text{ and } \mathbb{K}^{\text{SS}}$ , protein reorganization is the differences in their matrix components. The top 25 percentile of  $|\Delta k_{IJ}|$  in a union list are the prominent mechanical responses of the rigidity graph upon molecular binding. BPTI unbinding of RT is used here for illustration.

This approach is taken here to study the reorganization of a serine protease family member rat trypsin (RT) upon unbinding the BPTI inhibitor and of the third PDZ signaling domain (PDZ3) after binding the CRIPT peptide substrate. In RT, substrate variation or mutation at sites away from the catalytic triad still impact its activity; yet, how such a long-range communication is achieved is not clear.<sup>25–29</sup> Whereas for PDZ3, the linkage between its intra-domain<sup>5,6</sup> and inter-domain allostery<sup>30</sup> has not been established.

## 2 Materials and methods

The computational framework integrating explicit-solvent all-atom MD simulations, structure-mechanics statistical learning, and graph theory is summarized in Fig. 1. After the calculation of rigidity graph from protein dynamics and identification of the set of statistically prominent eigenmodes,  $\Pi$ , the list of strongly coupled residues pairs,  $\{IJ\}_{\Pi}$ , is determined as shown in the lower-right corner of Fig. 1(a). Next, the scheme devised here for comparing the protein mechanical coupling networks at different substrate binding states is summarized in Fig. 1(b). The different components of this approach are discussed in more detail in the following.

### 2.1 All-atom MD at different substrate binding states

The X-ray structure of RT bound with BPTI (holoRT, PDB ID: 3TGI)<sup>31</sup> is used to construct the all-atom model of apoRT based on the procedure<sup>14</sup> of model development for holoRT. For PDZ3, the 1BE9 X-ray structure<sup>32</sup> is used to construct the model of holoPDZ3 following the steps of setting up apoPDZ3 (ref. 14) with 1BEF. 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<sup>33</sup> is used to compute the potential energy and the GROMACS software<sup>34</sup> is used for MD runs. After equilibration, the production run is at 300 K and 1 atm for 5  $\mu$ s for all simulations, during which a snapshot is saved every 1 ps for analysis. In probing variation in the protein mechanical coupling network, holoRT and apoPDZ3 are taken as the reference states and their trajectories were collected as reported in ref. 14. Here, the same simulation protocol is used for the all-atom MD simulations of the response states, apoRT and holoPDZ3.

### 2.2 Calculation of rigidity graphs from protein dynamics

The 5  $\mu$ s all-atom MD production run is obtained for both the apo and holo states of each protein system. Each trajectory is



split into overlapping 10 ns windows as shown in Fig. 1(a), and the atomistic configurations in each are used to compute the spring parameters (equilibrium length  $l_{ij}^0$  and spring constant  $k_{ij}$  between each site pair in the potential energy function) of the coarse-grained (CG) backbone-side chain elastic network model (bsENM) as defined in ref. 14. This split of a long trajectory into window segments is to overcome the harmonic approximation of bsENM. The structure-mechanics statistical learning method involves self-consistent iterations with normal mode analysis (NMA) to compute the length fluctuation of each spring,  $\langle \delta l_{ij}^2 \rangle_{\text{NMA}}$ , and match the corresponding fluctuation observed in all-atom MD simulation,  $\langle \delta l_{ij}^2 \rangle_{\text{AA}}$ . In each trajectory window, the equilibrium length  $l_{ij}^0$  is the averaged distance between CG sites  $i$  and  $j$  in the all-atom MD segment. For an isolated spring, the variance of length fluctuation at thermal equilibrium is inversely proportional to the magnitude of spring constant according to the equipartition theorem. Given the system of coupled springs in bsENM, the elastic parameters are thus updated as  $k_{ij}^{(n+1)} = k_{ij}^{(n)} - \alpha(1/\langle \delta l_{ij}^2 \rangle_{\text{NMA}}^{(n)} - 1/\langle \delta l_{ij}^2 \rangle_{\text{AA}})$  with  $(n)$  the iteration number and  $\alpha$  the learning rate.<sup>35</sup> The other components of the structure-mechanics statistical learning, including the determination of cutoff length for including a spring in bsENM and the quantitative calibration, are reported in ref. 14.

The rigidity graphs of protein dynamics are constructed from the statistically learned inter-site force constants,  $k_{ij}$ 's. Residues indexed by  $I$  and  $J$  are nodes and the inter-residue coupling strengths  $k_{IJ}$ 's are edge weights. The inter-residue coupling strength is calculated as  $k_{IJ} = \sum_{i \in I, i \in J} k_{ij}$  for  $|J - I| > 2$  and zero

otherwise to focus on the strengths of non-covalent interactions and the springs of disulfide bonds are also excluded. With the degree as  $k_{II} = \sum_{j \neq I} k_{ij}$ , a rigidity graph ( $\mathbb{K}_n$ ) is constructed for each  $n$  of the trajectory windows. Since the bsENM is composed of backbone and side-chain sites, the rigidity graph can be categorized as backbone-backbone ( $\mathbb{K}_n^{\text{BB}}$ ), backbone-side chain ( $\mathbb{K}_n^{\text{BS}}$ ) and side chain-side chain ( $\mathbb{K}_n^{\text{SS}}$ ).

### 2.3 Statistical analysis of rigidity graphs

For statistical analysis of the fluctuating rigidity graphs along the all-atom MD trajectory, we consider the mean rigidity graph by averaging over those of each trajectory window, *i.e.*,  $\bar{\mathbb{K}} = \bar{\mathbb{K}}^{\text{BB}} + \bar{\mathbb{K}}^{\text{BS}} + \bar{\mathbb{K}}^{\text{SS}}$ . For each type (BB, BS, or SS) of the mean rigidity graph  $\bar{\mathbb{K}}$ , spectral decomposition is conducted and  $\bar{\mathbb{K}} = \sum \lambda_\alpha \mathbf{v}_\alpha \mathbf{v}_\alpha^T$ . Here, the superscript is dropped for simplicity, and this analysis is conducted for each of  $\bar{\mathbb{K}}^{\text{BB}}$ ,  $\bar{\mathbb{K}}^{\text{BS}}$ , and  $\bar{\mathbb{K}}^{\text{SS}}$ . In each case, the mean modes have the eigenvectors ( $\mathbf{v}_\alpha$ 's) and eigenvalues ( $\lambda_\alpha$ 's). Each mean mode  $\alpha$  is then compared with all the modes of every window to compute its mean-mode content in the window as  $r_{n\alpha} = \max_\beta |\mathbf{v}_{n\beta} \cdot \mathbf{v}_\alpha|$ . Averaging the mean-mode contents of all gives the averaged mean-mode content  $\langle r_\alpha \rangle$ .

The prominent modes of a particular rigidity graph, are then identified by two criteria.<sup>14</sup> The first is to find the mechanically strong modes with a high eigenvalue (statistical outliers of the  $\lambda_\alpha$  distribution). The second is to identify the highly persistent

eigenvectors by fitting the distribution of  $\langle r_\alpha \rangle$  and getting the values greater than a cumulative density function (CDF) cutoff. The modes that satisfies both these conditions are the prominent modes of ultra-high mechanical coupling strengths that persist through thermal noise in the all-atom MD trajectory. From each member in the set of prominent modes  $\Pi$ , the strongly coupled residue pairs are identified as both residues taking a weight higher than a cutoff in the eigenvector,  $\{IJ\}_\Pi = \{IJ | \nu_{I\pi}^2 \ \& \ \nu_{J\pi}^2 > \nu_c^2, \ \forall \ \pi \in \Pi\}$  and putting all such residues together forms the list of strongly coupled residues pair in the rigidity graph, *i.e.*,  $\{IJ\}_\Pi$  in Fig. 1(a). For consistent comparison of rigidity graphs, the numerical details reported in ref. 14 ( $\nu_c^2 = 0.1$ ) are employed here for the alternative substrate binding state.

### 2.4 Identification of mechanically responsive residues upon substrate binding

Since the protein mechanical coupling network depends on the state of inhibitor or substrate binding, the list of strongly coupled residues as in the prominent modes of rigidity graph,  $\{IJ\}_\Pi$ , would also vary. As shown in Fig. 1(b), the identification of mechanically responsive residues upon substrate binding starts by finding the union list  $\{IJ\}$  from the prominent rigidity graph modes at the reference state and the response state. For BPTI unbinding of RT, the reference state is holoRT and the response state is apoRT. From the union list of strongly coupled residues  $\{IJ\}$ , variation in the mechanical coupling network is measured as  $\Delta k_{IJ}$  and the prominent mechanical responses are residue pairs within top 25% of the  $|\Delta k_{IJ}|$  values, Fig. 1(b).

### 2.5 Multiple sequence alignment for PDZ3

The third PDZ domain of PSD-95 (PDZ3) has a CT-extension  $\alpha$ -helix,  $\alpha_3$ , packed against the  $\beta$ -sandwich which acts as a part of the linker between the third PDZ domain and the SH3 domain.<sup>36</sup> To identify the PDZ3 sequences possessing the  $\alpha_3$  extension, the subrange sequence from H372 to A402 of 1BE9 is used to BLAST against the non-redundant database excluding the models and uncultured/environmental samples. A seed length of 3 is used to increase the specificity against the sequence database which resulted in 1457 sequences. The hypothetical, clone variants and other redundant sequences are the removed, resulting in the final 45 sequences with diverse taxonomic lineage. The multiple sequence alignment of these strings is conducted by T-Coffee.<sup>37</sup>

## 3 Results and discussion

Beginning with the case BPTI unbinding in RT, reorganization in the mechanical coupling network is captured based on the rigidity graphs of different chemical components (BB, BS, and SS) that all have the same dimension as the total residue number. For a particular rigidity graph, the prominent mechanical responses—top 25 percentile of the  $|\Delta k_{IJ}|$  values in the union list  $\{IJ\}$  (holoRT and apoRT) of strongly coupled residues pairs—are softening ( $\Delta k_{IJ} < 0$ ) or stiffening ( $\Delta k_{IJ} > 0$ ), and the rest are neutral. Putting together the responses of





$\Delta k_{ij}^{BB}$ ,  $\Delta k_{ij}^{BS}$ , and  $\Delta k_{ij}^{SS}$  uncovers the specific routes of mechanical relay. Whether such intriguing patterns would appear in a different functionality of signaling is then analyzed by comparing the rigidity graphs learned from the protein dynamics of holoPDZ3 and apoPDZ3.

### 3.1 Occlusion of RT active site upon BPTI unbinding

Over the 5  $\mu$ s trajectories of holoRT and apoRT, the active site residues involving the triad and the oxyanion hole (Fig. 2(a)) remain close to the conformation in the reference X-ray structure as their  $C_\alpha$  RMSD (root-of-mean-squared-deviation) values are small ( $\sim 1$  Å), Fig. 2(b). Similarly, the inhibitor unbinding does not affect the  $\beta$ -barrel structures in RT and they have very small RMSD values throughout the production runs. The conformational responses, though, mostly occur at surface loops and their  $C_\alpha$  RMSD values are significantly higher, Fig. 2(b). As the space occupied by BPTI becomes available in apoRT, specific residues in these regions shift to interact with

active site residues. Furthermore, even the  $l_5$  loop, *i.e.*, the  $Ca^{2+}$ -binding loop, distal to the active site exhibits significant conformational changes in apoRT. To understand such behaviors in terms of physical interactions, the protein reorganization is analyzed based on the prominent changes in the mechanical coupling network.

### 3.2 RT reorganization exhibits routes of mechanical relay

The prominent mechanical responses in the RT rigidity graphs are shown in Fig. 3(a) with a color bar to quantitatively indicate the stiffening/softening levels of inter-residue couplings. Marking the significant  $\Delta k_{ij}$  values on the corresponding residue indices reveals physical contiguity in the prominent mechanical responses. In Fig. 3(a), a line is put between two marks if a common residue is shared or if they have residues of nearby sequences (mostly  $\leq 2$ , no more than 4 residues). Consecutive connection of such lines leads to the routes in Fig. 3(a) that have nearly vertical and/or horizontal sections. On the RT structure, residues participating in the prominent mechanical responses are shown in Fig. 3(b).

Here, the reference state is holoRT and apoRT is the response state, Fig. 1(b). From the all-atom MD trajectory of holoRT, the strong couplings with BPTI were identified to be  $\Delta k_{K15^{BPTI}-S190}^{BS}$ ,  $\Delta k_{K15^{BPTI}-G193}^{BB}$ ,  $\Delta k_{A16^{BPTI}-S195}^{SS}$ , and  $\Delta k_{R17^{BPTI}-E151}^{SS}$ .<sup>14</sup> Since they become absent in apoRT, the off mark “ $\times$ ” is labeled on the diagonal of Fig. 3(a) at the corresponding residues of RT. In capturing the prominent mechanical responses in RT upon the inhibitor unbinding (Fig. 1(b)), the significant  $\Delta k_{ij}$  changes are obtained by comparing the residue pairs in the protease. In Fig. 3(a), the off mark is also labelled over the intra-RT couplings that are lost upon BPTI unbinding. On the other hand, the on mark “+” indicates the strong interactions formed in apoRT but not in holoRT. In RT, S195, D102, and H57 are the catalytic triad, and G193 is the oxyanion hole. The secondary structures and functional sites of RT are annotated in Fig. 4(a).

Considering the reorganization branching from S195 with  $\Delta k_{A16^{BPTI}-S195}^{SS}$  off in apoRT, the neighboring  $\Delta k_{G196-V213}^{BB}$  stiffens, Fig. 3(a). W215 nearby V213 at the C-terminal end of  $\beta_{12}$  then moves toward the active site (Fig. 3(b)) and detaches V227 at S1 site, *i.e.*,  $\Delta k_{W215-V227}^{SS}$  off. This positional shift of the W215 side chain at the end of  $\beta_{12}$  and beginning of  $l_{11}$  that softens the nearby couplings at the S1 site is coupled with R96 in  $l_6$ , which also migrates to the space originally occupied by BPTI and turns on  $\Delta k_{R96-W215}^{SS}$ , Fig. 3(a). With R96 coupling two triad residues ( $\Delta k_{R96-D102}^{SS}$  and  $\Delta k_{R96-H57}^{SS}$ ) through this conformational change in apoRT,  $\Delta k_{H57-S195}^{SS}$  is switched off in the response state. Therefore, losing the interactions with BPTI leads to a chain of adjustment in the interaction network that eventually occludes the catalytic triad in apoRT by W215 and R96. This route of coupling strength variation over contiguous residues—a system of mechanical relay—is colored orange in Fig. 3 and 4. In this case, W215, R96, and H57 act as on-off switches by interacting with alternative partners at different states of inhibitor binding. In the S1A protease family, allosteric active site occlusion has been observed for the W215 loop in thrombin<sup>43–45</sup> while for the R96 containing loop in coagulation

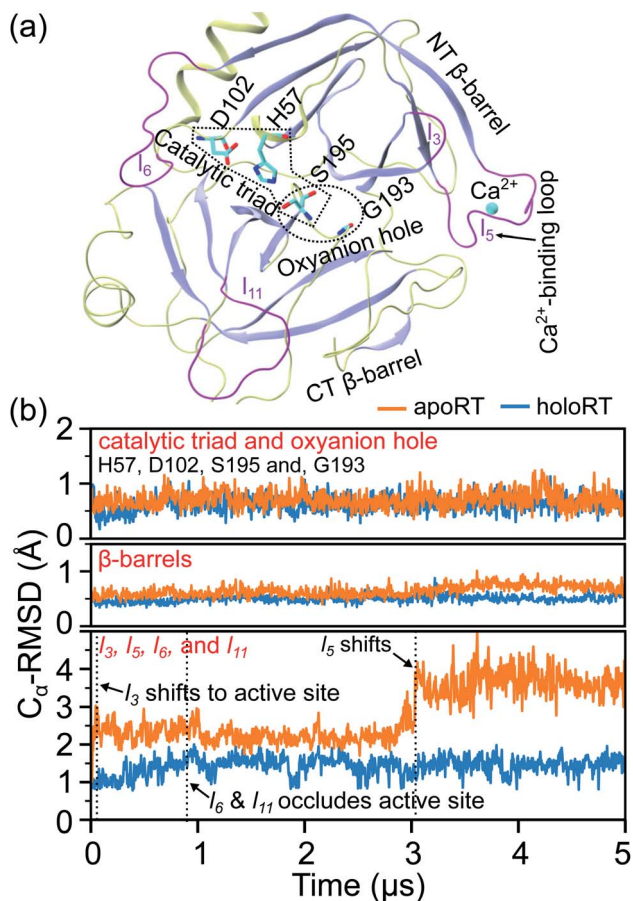
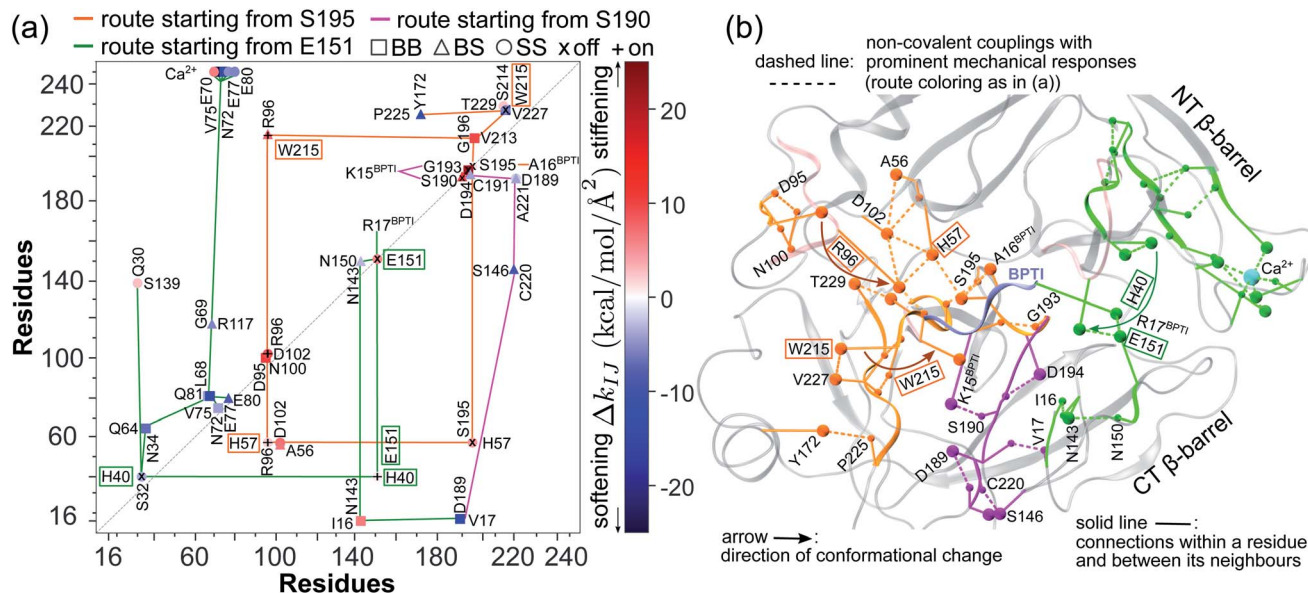


Fig. 2 The  $C_\alpha$  RMSD (root-of-mean-squared-deviation) of RT from the coordinates in the reference X-ray structure in holoRT and apoRT simulations. (a) A ribbon representation of the RT structure with the  $\beta$ -barrels colored in iceblue and the NT- and CT-barrels labelled. The catalytic triad and oxyanion hole at the  $\beta$ -barrel interface are shown in licorice. The  $Ca^{2+}$  ion is shown in a ball representation. (b) The  $C_\alpha$  RMSD of catalytic triad and oxyanion hole residues (top), the  $\beta$ -barrel residues (middle), and the surface loops exhibiting active site occlusion upon BPTI unbinding and the  $Ca^{2+}$ -binding loop (bottom).





**Fig. 3** Reorganization of the mechanical coupling network in RT upon BPTI unbinding. (a) The prominent mechanical responses of inter-residue couplings—top 25 percentile of the  $|\Delta k_{IJ}|$  values in the union list of the strongly coupled pairs in each rigidity graph,  $\mathbb{K}^{\text{BB}}$  (square),  $\mathbb{K}^{\text{BS}}$  (triangle), and  $\mathbb{K}^{\text{SS}}$  (circle). The levels of softening ( $\Delta k_{IJ} < 0$ , blue) and stiffening ( $\Delta k_{IJ} > 0$ , red) responses are represented by the color bar. The prominent couplings with BPTI in holoRT are labelled on the diagonal. The key RT residue interacting with BPTI that starts a specific route of mechanical relay is used to annotate the chains of interaction network reorganization. If  $k_{IJ} = 0$  in the response state, the pair is labelled "off" with  $\times$ . If  $k_{IJ} = 0$  in the reference state, the pair is labelled "on" with  $+$ . (b) The residues of mechanical relay systems in (a) on the RT structure.

factor IXa and in trypsin-like hepatocyte growth factor activator.<sup>46,47</sup> The previously unknown mechanism as identified here for RT is the route of mechanical relay over spatially contiguous residues.

Another scenario is E151 that strongly interacts with R17<sup>BPTI</sup> in holoRT. As  $\Delta k_{\text{R17}^{\text{BPTI}}-\text{E151}}^{\text{SS}}$  becomes off in apoRT, the imidazole side chain of H40 at the start of  $\beta_3$  shifts to interact with E151 and  $\Delta k_{\text{H40}-\text{E151}}^{\text{SS}}$  is turned on. This conformational change also shuts off  $\Delta k_{\text{S32}-\text{H40}}^{\text{SS}}$  in apoRT, and the E151 and H40 mechanical switches start a chain of softening responses that reach the Ca<sup>2+</sup>-binding loop around E70. This distal loop has been shown to relate to the long-term stability of family members and exhibit stimulatory effects on enzyme activities.<sup>42</sup> For example, in urokinase-type plasminogen activator, communication between H40 and E70 loops has been implied from the inhibitor binding and enzyme activity experiments.<sup>41</sup> Our finding in RT uncovers the underlying mechanism as the domino-like variation in mechanical couplings. This route seeding from E151 also involves a reported autolysis site R117 (ref. 38–40) and is colored green in Fig. 3 and 4(a).

### 3.3 Hotspots in mechanical relay capture functional sites

Functional regulation in many S1A proteases has been found to proceed by movements of surface loops and the participating segments are a key signature for member classification.<sup>42</sup> The previously unknown behaviors of RT are uncovered here to have a unique profile, including  $l_6$  of R96,  $l_{11}$  starting at W215,  $l_3$  ends at H40, and the Ca<sup>2+</sup>-binding loop, Fig. 3. Along a route of mechanical relay, an important notice is that having a conformational change is not necessary. For example, as  $\Delta k_{\text{K15}^{\text{BPTI}}-\text{S190}}^{\text{BS}}$

and  $\Delta k_{\text{K15}^{\text{BPTI}}-\text{G193}}^{\text{BB}}$  become off in apoRT, the pathway colored purple in Fig. 3 and 4 involves soft responses without having any on-off switch. S146, another reported autolysis site in thrombin and bovine trypsin,<sup>38–40</sup> is on this route.

An intriguing finding in the case of RT reorganization upon BPTI unbinding is that the mechanically responsive residues during the different protein dynamics are mostly the functional sites identified experimentally, Fig. 4(a). For example, the salt bridge between I16 and D194 retains similar strengths in both holoRT and apoRT trajectories, but the two biologically important residues exhibit significant variation in their strengths of the couplings with other partners as seen on the purple route in Fig. 3 and 4(a). Occurring through specific chains of physically contiguous amino acids, the significant coupling strength variation in the interaction network reorganization of all-atom MD simulations—the hotspots of mechanical relay—is thus a useful metric for the functional importance of residues.

### 3.4 PDZ3 exhibits mechanical relay upon binding CRIPT

To study whether similar behaviors could be observed in an alternative functionality of signaling rather than in an enzyme, the self-adjustment in PDZ3 upon binding the CRIPT peptide is analyzed by comparing the rigidity graphs holoPDZ3 (response state) and apoPDZ3 (reference state) following the schemes outlined in Fig. 1. With the binding groove in between  $\beta_2$ ,  $l_2$ , and  $\alpha_2$ , Fig. 4(b) and 5(a), the  $\beta$ -sandwich core has been shown to conduct intra-domain allostery without structural changes.<sup>5,6</sup> Substrate binding in PDZ3 also impacts C-terminal (CT)-



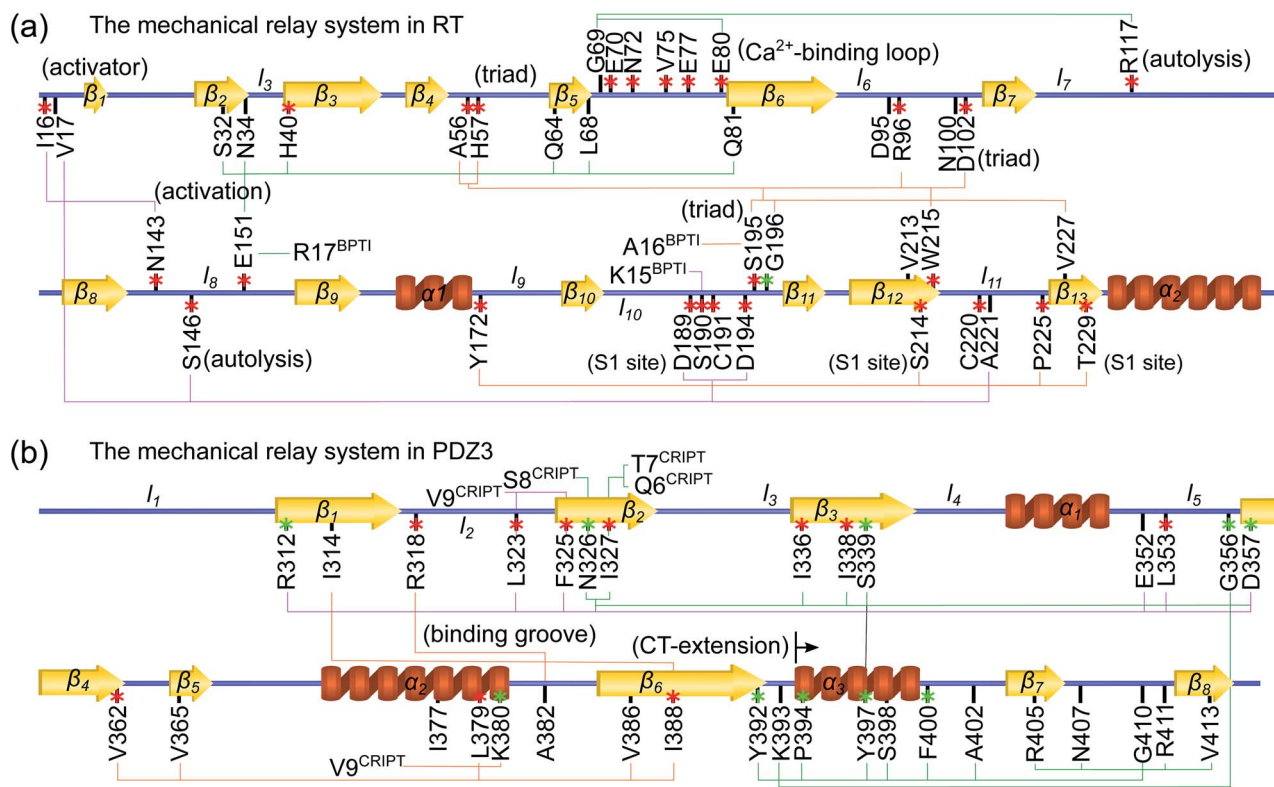


Fig. 4 Functional and evolutionary significance of protein intramolecular mechanical relays. Mechanically responsive residues in molecule-binding induced reorganization are mostly functional sites identified in experiments (red star), or have high sequence conservation (green star). The mechanical relay systems identified in Fig. 3 for RT and in Fig. 6 for PDZ3 are listed here on the primary sequence with secondary structures annotated. (a) In RT, H57, D102, S195, and G193 are the catalytic triad and oxyanion hole; nearby D189 and S214 are the S1 site for specificity control in substrate binding;<sup>29</sup> R117 and S146 are protease autolysis sites;<sup>38–40</sup> H40,<sup>41</sup> E70,<sup>42</sup> and W215 (ref. 43–45) at regulatory loop edges exhibit long-range effects on activities; R96 in a surface loop conducts active site occlusion in a homolog;<sup>46,47</sup> with I16 and D194 activators forming a salt bridge, the activation domain involves N143 in  $l_8$ , D189–C191 in  $l_{10}$ , and C220 in  $l_{11}$ ,<sup>41</sup> and functional mutation sensitivity has been demonstrated for the other labelled residues.<sup>48–50</sup> (b) In PDZ3, the identified functional sites showing mutation sensitivity are mostly in the  $\beta$ -sandwich core.<sup>51,52</sup> Including CT-extension in MSA as motivated by their inter-domain couplings in the rigidity graphs reveals the highly conserved residues shown here.

extension conformation as illustrated by NMR and SAXS measurements,<sup>53</sup> *i.e.*, exhibiting inter-domain allostery.

Over the 5  $\mu$ s production runs of the all-atom MD simulation of apoPDZ3 and holoPDZ3 in explicit solvent, the  $\beta$ -sandwich stays close to the reference X-ray structures<sup>32</sup> as illustrated by the small  $C_\alpha$  RMSD values in Fig. 5(b). The  $\alpha_3$  helix in CT-extension is more flexible, but remains attached to the  $\beta$ -sandwich in both apoPDZ3 and holoPDZ3. On the other hand, the  $\beta_7$ – $\beta_8$  hairpin in CT-extension remains attached to the  $\beta$ -sandwich throughout the apoPDZ3 reference state, but in holoPDZ3, it detaches and loosens (Fig. S1†), showing pronounced RMSD values in Fig. 5(b). This is a first atomic-scale *in silico* observation that the substrate binding in PDZ3 affects the conformation of distal CT-extension as speculated in NMR and SAXS studies.

The underlying mechanism of this inter-domain, long-range effect of CRIPT binding is unraveled here by comparing the rigidity graphs of different protein dynamics, *i.e.*, holoPDZ3 *versus* apoPDZ3. In the former, T7<sup>CRIPT</sup> and V9<sup>CRIPT</sup> are found to make most of the stronger interactions with PDZ3 as labeled with the on mark along the Fig. 6(a) diagonal, consistent with their

determining roles in substrate selectivity.<sup>54,55</sup> The backbone sites of I327 on  $\beta_2$  turn on  $\Delta k_{T7^{\text{CRIPT}}-I327}^{\text{BB}}$  and  $\Delta k_{Q6^{\text{CRIPT}}-I327}^{\text{BB}}$  to anchor the substrate, while its neighbor N326 mediates  $\Delta k_{T7^{\text{CRIPT}}-N326}^{\text{BB}}$  and  $\Delta k_{S8^{\text{CRIPT}}-N326}^{\text{SS}}$ , Fig. 4(b) and 6. Stiffening responses ( $\Delta k_{I327-I336}^{\text{SS}}$  and  $\Delta k_{N326-I339}^{\text{BB}}$ ) then proceed on the other side of  $\beta_2$  facing  $\beta_3$ . Next, the mechanical relay further passes to  $\beta_4$  ( $\Delta k_{I338-D357}^{\text{BB}}$  softening) and to the  $\beta_6$  end facing CT-extension  $\alpha_3$  ( $\Delta k_{G356-K393}^{\text{BB}}$  stiffening). Following this route across the  $\beta$ -sandwich, the coupling of Y392 side chain at the C-terminal edge of  $\beta_6$  with G410 in between  $\beta_7$  and  $\beta_8$ ,  $\Delta k_{Y392-G410}^{\text{BS}}$ , is off, leading to detachment of the  $\beta_7$ – $\beta_8$  hairpin. This route of mechanical relay seeding from the T7<sup>CRIPT</sup>–I357 coupling in the substrate binding groove and reaching CT-extension is colored green in Fig. 4(b) and 6. The inter-domain allostery of CRIPT binding can thus be understood as the chain of physically contiguous residues exhibiting significant coupling strength variation.

Another route starts on the  $\beta_2$  side of the binding groove. Turning on the prominent hydrophobic interaction  $\Delta k_{V9^{\text{CRIPT}}-F325}^{\text{SS}}$  in holoPDZ3 softens the  $\beta_2$ – $l_5$  coupling of  $\Delta k_{F325-L353}^{\text{SS}}$  in apoPDZ3 and a chain of softening responses ( $\Delta k_{E352-R312}^{\text{BS}}$  and  $\Delta k_{R312-D357}^{\text{SS}}$ ) colored purple in Fig. 4(b) and 6 follows. On the  $\alpha_2$  side of the binding





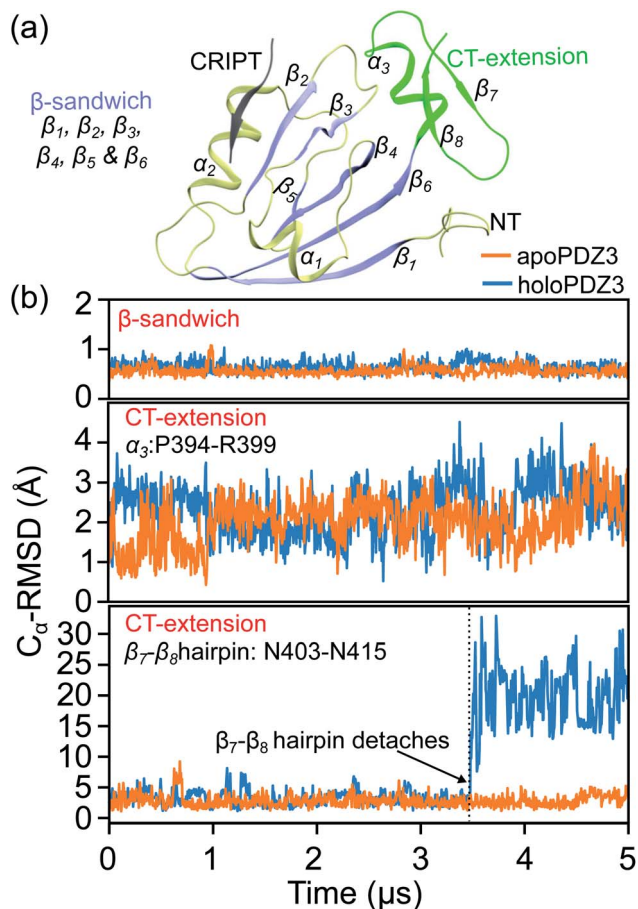


Fig. 5 The  $C_\alpha$  RMSD (root-of-mean-squared-deviation) of PDZ3 from the coordinates in the reference X-ray structure in holoPDZ3 and apoPDZ3 simulations. (a) A ribbon representation of the PDZ3 structure with the  $\beta$ -sandwich colored in iceblue. The CT extension is colored green and the CRIPT peptide is colored gray. (b) The  $C_\alpha$  RMSD of  $\beta$ -sandwich (top), the  $\alpha_3$ -helix in CT-extension (middle), and  $\beta_7$ - $\beta_8$  hairpin in CT-extension (bottom).

groove, on the other hand, turning on  $\Delta k_{V9^{CRIP1-L379}}^{SS}$  leads to the mechanical relay route colored orange consisting mostly of stiffening responses that reaches the distal I388 in  $\beta_6$ . Intra-domain communication in PDZ3 can thus proceed through multiple pathways. The mechanical coupling variation also provides comprehensive mechanistic basis for the NMR observed rigidification in structural flexibility after substrate binding.<sup>6</sup>

### 3.5 Mutation sensitive sites affecting the substrate binding of PDZ3 participate in mechanical relay

Although most of the prominent mechanical responses upon binding CRIPT are stiffening, certain inter-residue couplings soften with pronounced levels, indicating heterogeneity and complexity in the interaction network reorganization. By performing high-throughput mutagenesis over the  $\beta$ -sandwich residues, several sites not at the substrate-binding groove were shown to still have significant impact on CRIPT binding.<sup>51</sup> In comparing the rigidity graphs of holoPDZ3 and apoPDZ3 all-atom MD simulations, most of the mechanically responsive

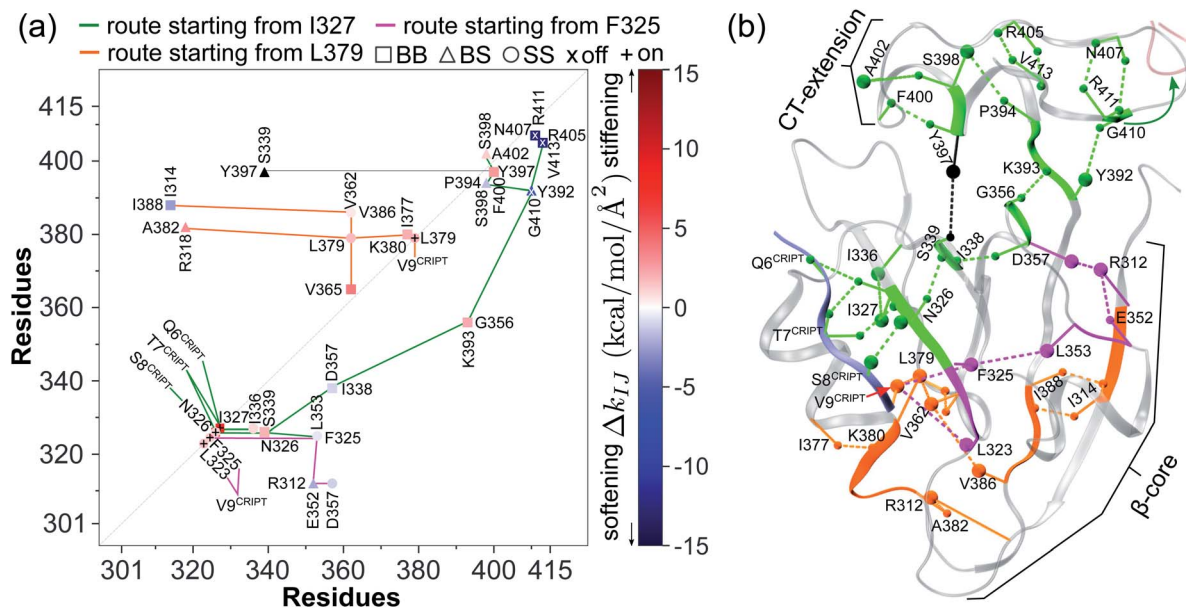
residues away from the binding groove, *i.e.*, on the mechanical relay routes shown in Fig. 4(b) and 6, predict the mutation sensitive sites identified experimentally, including I336 and I338 in  $\beta_3$ , L353 in  $I_5$ , V362 in  $\beta_4$ , and I388 in  $\beta_6$ . Our results indicate that the experimentally measurable functional property of PDZ3 residues as their extents of imposing long-range effects on substrate binding can be linked to the specific variation in their interaction strengths during the protein dynamics at different substrate binding states. The property of mechanical relay thus provides the mechanistic basis for propagation of substrate binding signals as altered coupling strengths. Certain residues such as I338 exhibit responses in both its backbone and side-chain couplings, showcasing the cooperation of chemical components. In another experimental work, R318 not captured by the high-throughput mutagenesis<sup>51</sup> was found to affect CRIPT binding,<sup>52</sup> and our rigidity graph analysis identifies it to also exhibit prominent mechanical response.

### 3.6 Mechanical relay of the inter-domain allostery in PDZ3 exhibit strong signals in multiple sequence alignment

For PDZ3, inter-domain allostery is functionally more relevant but much less understood. Our analysis of rigidity graphs at different substrate binding states shows that the two processes are actually linked and hence provides a unified view. The intra-domain routes of mechanical relay all reach the edges of  $\beta_1$  (R312),  $\beta_4$  (G356 and D357), and  $\beta_6$  (Y392 and K393) that face the CT-extension as revealed in their coupling strength responses, and this cluster links to the  $\Delta k_{Y392-G410}^{BS}$  knob that drives the inter-domain allostery for the detachment of the  $\beta_7$ - $\beta_8$  hairpin from the  $\beta$ -sandwich, Fig. 4(b) and 6.

Another prominent coupling of CT-extension with  $\beta$ -sandwich is between Y397 in  $\alpha_3$  and S339 in  $\beta_3$ , and the latter is on the green route of mechanical relay. Unlike the turning off of  $\Delta k_{Y392-G410}^{BS}$  upon substrate binding,  $\Delta k_{S339-Y397}^{BS}$  is neutral and the coupling persists with high strengths in both holoPDZ3 and apoPDZ3. Y397 phosphorylation has been shown to affect the inter-domain allostery with SH3,<sup>30</sup> and our result points to the disruption of this coupling as a likely mechanism. Although  $\alpha_3$  remains linked to  $\beta_3$  with  $\Delta k_{S339-Y397}^{BS}$  neutral, the  $\beta_7$ - $\beta_8$  detachment upon  $\Delta k_{Y392-G410}^{BS}$  turned off leads to significant reorganization in the mechanical couplings in  $\alpha_3$ , particularly at the F400 site facing the  $\beta$ -sandwich. For the residues conducting inter-domain allostery as predicted by the rigidity graph analysis, their functional relevance is further analyzed by designing a multiple sequence alignment (MSA) limited to  $\alpha_3$ -containing PDZ3 analogs as described in Materials and methods. D357, the mostly conserved residue in the  $\beta$ -sandwich-only MSA,<sup>56</sup> is found to be preserved among the  $\alpha_3$ -bearing sequences of PDZ3, and the other key residues for the inter-domain allostery observed in all-atom MD simulations, R312, the Y392 knob, Y397, and F400 are all very highly conserved when CT-extension is considered in MSA, Fig. S2.† The mechanical relay identified here for PDZ3 inter-domain allostery thus exhibits strong signals in the MSA concerning the sequences that contain the CT-extension  $\alpha_3$ .





**Fig. 6** Reorganization of the mechanical coupling network in PDZ3 upon binding the CRIPT peptide. (a) The prominent mechanical responses of inter-residue couplings—top 25 percentile of the  $|\Delta k_{IJ}|$  values in the union list of the strongly coupled pairs in each rigidity graph,  $\mathbb{K}^{\text{BB}}$  (square),  $\mathbb{K}^{\text{BS}}$  (triangle), and  $\mathbb{K}^{\text{SS}}$  (circle). The levels of softening ( $\Delta k_{IJ} < 0$ , blue) and stiffening ( $\Delta k_{IJ} > 0$ , red) responses are represented by the color bar. The prominent couplings with CRIPT in holoPDZ3 are labelled on the diagonal. The key PDZ3 residue interacting with CRIPT that starts a specific route of mechanical relay is used to annotate the chains of interaction network reorganization. If  $k_{IJ} = 0$  in the response state, the pair is labelled off with  $\times$ . If  $k_{IJ} = 0$  in the reference state, the pair is labelled on with  $+$ . (b) The residues of mechanical relay systems in (a) on the PDZ3 structure. Definitions of solid/dash lines and arrows are as in Fig. 3(a). The S339–Y397 coupling prominent in both apoPDZ3 and holoPDZ3 is colored black.

## 4 Conclusions

In both RT and PDZ3, comparing the rigidity graphs of protein dynamics at different substrate binding states unravels specific sets of spatially contiguous residues as routes of mechanical relay. In such a connection over distal sites, the inter-linked couplings include backbone as well as side chains with an intricate coordination of their polar and hydrophobic interactions. For residues exhibiting prominent responses in the reorganization of the protein interaction network, the close correspondence with the experimentally identified functional sites and the highly conserved spots in our simulation-motivated MSA indicates that the mechanical relay system is under significant evolutionary pressure. From the explicit-solvent all-atom MD simulations conducted at different molecular binding states, comparing their residue rigidity graphs hence provides a useful approach to understanding the protein functional properties in terms of specific molecular interactions. In many important cases, the mechanical coupling network of a substrate like DNA also exhibits complicated behaviors.<sup>13</sup> Capturing the coupled reorganization over the rigidity graphs of the enzyme–substrate pair is likely key to understand the sequence-specific functional properties<sup>57</sup> such as binding affinities, catalytic mechanism, and kinetics. In these complex scenarios, the framework of our structure-mechanics statistical learning and the data structure of graphical analysis are readily applicable.

## Data availability

The Python code for construction of rigidity graphs, identification of prominent modes, and quantification of mechanical

responses due to substrate association/dissociation can be found at <https://github.com/nixnmtm/MechanicalRelay>.

## Author contributions

NR: conceptualization, methodology, software, writing – original draft preparation; TC: methodology and software; HY: conceptualization, writing – review & editing; JW: conceptualization, methodology, software, writing – original draft preparation, writing – review & editing, supervision.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

This work was supported by the Princeton University (to HY), the Ministry of Science and Technology of Taiwan (110-2113-M-A49-024), and the Ministry of Education of Taiwan through the IDS<sup>2</sup>B center and the “Smart Platform of Dynamic Systems Biology for Therapeutic Development” project in The Featured Areas Research Center Program. The National Center for High-Performance Computing of Taiwan supported part of the computational resources.

## References

- J. P. Changeux, *Science*, 2005, **308**, 1424–1428.
- R. G. Smock and L. M. Gierasch, *Science*, 2009, **324**, 198–203.





- 3 N. R. Latorraca, A. J. Venkatakrishnan and R. O. Dror, *Chem. Rev.*, 2016, **117**, 139–155.
- 4 A. Cooper and D. T. Dryden, *Eur. Biophys. J.*, 1984, **11**, 103–109.
- 5 C. M. Petit, J. Zhang, P. J. Sapienza, E. J. Fuentes and A. L. Lee, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 18249–18254.
- 6 A. L. Lee, *Biophys. Rev.*, 2015, **7**, 217–226.
- 7 S. Cyphers, E. F. Ruff, J. M. Behr, J. D. Chodera and N. M. Levinson, *Nat. Chem. Biol.*, 2017, **13**, 402–408.
- 8 S. P. Meisburger, W. C. Thomas, M. B. Watkins and N. Ando, *Chem. Rev.*, 2017, **117**, 7615–7672.
- 9 K. Banerjee-Ghosh, S. Ghosh, H. Mazal, I. Riven, G. Haran and R. Naaman, *J. Am. Chem. Soc.*, 2020, **142**(48), 20456–20462.
- 10 J. A. Hanson, K. Duderstadt, L. P. Watkins, S. Bhattacharyya, J. Brokaw, J. W. Chu and H. Yang, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 18055–18060.
- 11 J. B. Brokaw and J.-W. Chu, *Biophys. J.*, 2010, **99**, 3420–3429.
- 12 T. E. Morrell, I. U. Rafalska-Metcalf, H. Yang and J.-W. Chu, *J. Am. Chem. Soc.*, 2018, **140**, 14747–14752.
- 13 Y.-T. Chen, H. Yang and J.-W. Chu, *Chem. Sci.*, 2020, **11**, 4969–4979.
- 14 N. Raj, T. Click, H. Yang and J.-W. Chu, *Comput. Struct. Biotechnol. J.*, 2021, **19**, 5309–5320.
- 15 I. Bahar, T. R. Lezon, A. Bakan and I. H. Shrivastava, *Chem. Rev.*, 2010, **110**, 1463–1497.
- 16 W. J. Zheng, B. R. Brooks and D. Thirumalai, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 7664–7669.
- 17 D. Thirumalai, C. Hyeon, P. I. Zhuravlev and G. H. Lorimer, *Chem. Rev.*, 2019, **119**, 6788–6821.
- 18 F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. S. Marks, C. Sander, R. Zecchina, J. N. Onuchic, T. Hwa and M. Weigt, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, E1293–E1301.
- 19 N. Halabi, O. Rivoire, S. Leibler and R. Ranganathan, *Cell*, 2009, **138**, 774–786.
- 20 O. Rivoire, K. A. Reynolds and R. Ranganathan, *PLoS Comput. Biol.*, 2016, **12**, e1004817.
- 21 M. M. Tirion, *Phys. Rev. Lett.*, 1996, **77**(9), 1905–1908.
- 22 T. Haliloğlu, I. Bahar and B. Erman, *Phys. Rev. Lett.*, 1997, **79**, 3090–3093.
- 23 A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin and I. Bahar, *Biophys. J.*, 2001, **80**, 505–515.
- 24 G. Stetz and G. M. Verkhivker, *J. Chem. Inf. Model.*, 2016, **56**, 1490–1517.
- 25 S. R. Sprang, R. J. Fletterick, L. Gráf, W. J. Rutter and C. S. Craik, *Crit. Rev. Biotechnol.*, 1988, **8**, 225–236.
- 26 L. B. Evin, J. R. Vásquez and C. S. Craik, *Proc. Natl. Acad. Sci. U. S. A.*, 1990, **87**, 6659–6663.
- 27 A. Vindigni and E. Di Cera, *Protein Sci.*, 1998, **7**, 1728–1737.
- 28 M. M. Krem, S. Prasad and E. Di Cera, *J. Biol. Chem.*, 2002, **277**, 40260–40264.
- 29 L. Hedstrom, *Chem. Rev.*, 2002, **102**, 4501–4523.
- 30 J. Zhang, C. M. Petit, D. S. King and A. L. Lee, *J. Biol. Chem.*, 2011, **286**, 41776–41785.
- 31 A. Pasternak, D. Ringe and L. Hedstrom, *Protein Sci.*, 1999, **8**, 253–258.
- 32 D. A. Doyle, A. Lee, J. Lewis, E. Kim, M. Sheng and R. MacKinnon, *Cell*, 1996, **85**, 1067–1076.
- 33 R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, J. MacKerell and D. Alexander, *J. Chem. Theory Comput.*, 2012, **8**, 3257–3273.
- 34 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1–2**, 19–25.
- 35 J.-W. Chu and G. A. Voth, *Biophys. J.*, 2006, **90**, 1572–1582.
- 36 A. Camara-Artigas, J. Murciano-Calles and J. C. Martínez, *Acta Crystallogr., Sect. D: Struct. Biol.*, 2019, **75**, 381–391.
- 37 C. Notredame, D. G. Higgins and J. Heringa, *J. Mol. Biol.*, 2000, **302**, 205–217.
- 38 A. Bódi, G. Kaslik, I. Venekei and L. Gráf, *Eur. J. Biochem.*, 2001, **268**, 6238–6246.
- 39 X. F. Li, X. Nie and J. G. Tang, *Biochem. Biophys. Res. Commun.*, 1998, **250**, 235–239.
- 40 E. Várallyay, G. Pál, A. Patthy, L. Szilágyi and L. Gráf, *Biochem. Biophys. Res. Commun.*, 1998, **243**, 56–60.
- 41 T. Kromann-Hansen, E. L. Lange, H. P. Sørensen, G. Hassanzadeh-Ghassabeh, M. Huang, J. K. Jensen, S. Muyldermans, P. J. Declerck, E. A. Komives and P. A. Andreasen, *Sci. Rep.*, 2017, **7**, 1–11.
- 42 P. Goettig, H. Brandstetter and V. Magdolen, *Biochimie*, 2019, **166**, 52–76.
- 43 A. O. Pineda, Z.-W. Chen, A. Bah, L. C. Garvey, F. S. Mathews and E. Di Cera, *J. Biol. Chem.*, 2006, **281**, 32922–32928.
- 44 P. S. Gandhi, Z. Chen, F. S. Mathews and E. Di Cera, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 1832–1837.
- 45 W. Niu, Z. Chen, P. S. Gandhi, A. D. Vogt, N. Pozzi, L. A. Pelc, F. Zapata and E. Di Cera, *Biochemistry*, 2011, **50**, 6301–6307.
- 46 K. Sichler, E. Kopetzki, R. Huber, W. Bode, K.-P. Hopfner and H. Brandstetter, *J. Biol. Chem.*, 2003, **278**, 4121–4126.
- 47 C. Eigenbrot, R. Ganesan and D. Kirchhofer, *FEBS J.*, 2010, **277**, 2215–2222.
- 48 E. Várallyay, Z. Lengyel, L. Gráf and L. Szilágyi, *Biochem. Biophys. Res. Commun.*, 1997, **230**, 592–596.
- 49 L. Hedstrom, J. J. Perona and W. J. Rutter, *Biochemistry*, 1994, **33**, 8757–8763.
- 50 F. C. Peterson, N. C. Gordon and P. G. Gettins, *Biochemistry*, 2001, **40**, 6275–6283.
- 51 J. McLaughlin, N. Richard, F. J. Poelwijk, A. Raman, W. S. Gosal and R. Ranganathan, *Nature*, 2012, **491**, 138–U163.
- 52 C. N. Chi, A. Engström, S. Gianni, M. Larsson and P. Jemth, *J. Biol. Chem.*, 2006, **281**, 36811–36818.
- 53 J. Zhang, S. M. Lewis, B. Kuhlman and A. L. Lee, *Structure*, 2013, **21**, 402–413.
- 54 S. Gianni, S. R. Haq, L. C. Montemiglio, M. C. Jürgens, Å. Engström, C. N. Chi, M. Brunori and P. Jemth, *J. Biol. Chem.*, 2011, **286**, 27167–27175.
- 55 K. Luck, S. Charbonnier and G. Trave, *FEBS Lett.*, 2012, **586**, 2648–2661.
- 56 S. W. Lockless and R. Ranganathan, *Science*, 1999, **286**, 295–299.
- 57 R. Venkatramani and R. Radhakrishnan, *Phys. Rev. Lett.*, 2008, **100**, 088102.

