

RESEARCH ARTICLE | MARCH 27 2023

Allosteric communication in the gating mechanism for controlled protein degradation by the bacterial ClpP peptidase

Special Collection: [New Views of Allostery](#)

Ashan Dayananda ; T. S. Hayden Dennison; Hewafonsekage Yasan Y. Fonseka ; Mohammad S. Avestan ; Qi Wang; Riina Tehver ; George Stan 



J. Chem. Phys. 158, 125101 (2023)

<https://doi.org/10.1063/5.0139184>



CrossMark

Articles You May Be Interested In

Genomic and proteomic analysis of *Bacillus subtilis* as microplastic bioremediation agents

AIP Conference Proceedings (January 2023)

Molecular analysis of *Brevibacillus brevis* as polyethylene biodegradation agents

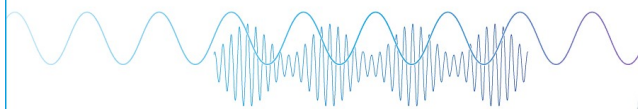
AIP Conference Proceedings (January 2023)

Purification and proteolytic activity of Caseinolytic protease (Clp) from *Lactobacillus plantarum* IIA-1AS

AIP Conference Proceedings (November 2020)

Webinar

Boost Your Signal-to-Noise
Ratio with Lock-in Detection



Sep. 7th – Register now



Zurich
Instruments

Allosteric communication in the gating mechanism for controlled protein degradation by the bacterial ClpP peptidase

Cite as: J. Chem. Phys. 158, 125101 (2023); doi: 10.1063/5.0139184

Submitted: 18 December 2022 • Accepted: 7 March 2023 •

Published Online: 27 March 2023



Ashan Dayananda,¹ T. S. Hayden Dennison,¹ Hewafonsekage Yasan Y. Fonseka,^{1, a)} Mohammad S. Avestan,¹ Qi Wang,^{1, b)} Riina Tehver,^{2, c)} and George Stan^{1, c)}

AFFILIATIONS

¹Department of Chemistry, University of Cincinnati, Cincinnati, Ohio 45221, USA

²Department of Physics and Astronomy, Denison University, Granville, Ohio 43023, USA

Note: This paper is part of the JCP Special Topic on New Views of Allostery.

^{a)}Present address: Biostatistics, Medpace, Inc., 5375 Medpace Way, Cincinnati, OH 45227, United States.

^{b)}Present address: Galixir Technologies, Beijing, China.

^{c)}Authors to whom correspondence should be addressed: tehver@denison.edu and george.stan@uc.edu

ABSTRACT

Proteolysis is essential for the control of metabolic pathways and the cell cycle. Bacterial caseinolytic proteases (Clp) use peptide substrates, such as ClpP, to degrade defective substrate proteins and to regulate cellular levels of stress-responsive proteins. Selective degradation access to the proteolytic chamber of the double-ring ClpP tetradecamer is controlled by a gating mechanism of the two axial pores. The binding of conserved loops of the Clp ATPase component of the protease or small molecules, such as cyclic dipeptide (ADEP), at peripheral ClpP ring sites triggers axial pore opening through dramatic conformational transitions of flexible N-terminal loops between disordered conformations in the “closed” pore state and ordered hairpins in the “open” pore state. We probe the allosteric communication underlying these conformational changes by comparing residue-residue couplings in molecular dynamics simulations of each configuration. Both principal component and normal mode analyses highlight large-scale conformational changes in the N-terminal loop regions and smaller amplitude motions of the peptidase core. Community network analysis reveals a switch between and inter-protomer coupling in the open-closed pore transition. Allosteric pathways that connect the ADEP binding sites to N-terminal loops are rewired in this transition, with shorter network paths in the open pore configuration supporting stronger intra- and inter-loop coupling. Structural perturbations, either through the removal of ADEP molecules or point mutations, alter the allosteric network to weaken coupling.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0139184>

I. INTRODUCTION

Maintaining protein homeostasis at the cellular level is essential in all kingdoms of life.^{1,2} Bacterial caseinolytic proteases (Clp) assist these mechanisms by performing intracellular protein quality control through regulatory protein degradation. Self-compartmentalized Clp nanomachines comprise a central barrel-like peptidase, such as ClpP, and one or two ring-shaped ATPase components, such as ClpA or ClpX, which are axially stacked at the two opposite ends of the peptidase.³ Complex formation,

which is dependent on ATP binding, tightly regulates the degradation process to prevent uncontrolled protein destruction.^{4,7} Isolated ClpP is catalytically inactive as access to its proteolytic chamber is precluded by axial gates locked in a closed configuration, which allow diffusion of short peptides but hinder the entry of longer unfolded polypeptide chains and block internalization of folded proteins.^{8–11} Docking of one or two ATPase partners to ClpP triggers gate opening to unleash the powerful degradation mechanism.^{14–17} Upon recognizing SPs targeted for degradation through short peptide tags attached covalently at one of

polypeptide terminal. The ATPase applies repetitive mechanical forces to effect SP unfolding and translocation through its narrow central channel and to propagate the unfolded polypeptide to the peptidase core. SP degradation takes place processively and yields small peptide fragments of 7–8 amino acids.

Highly conserved N-terminal loops (amino acids 1–19) control access to the degradation chamber through the apertures of the double-ring tetradecameric ClpP. Intermediate “head” regions (residues 20–122 and 149–193) connect N-terminal loops to the inter-ring equatorial interface formed by interlocked “handle” regions (residues 123–148) of each protomer (Fig. 1). In the “closed” configuration, N-terminal loops assemble into a mesh that involves strong inter-loop contacts and that occludes passage to the proteolytic chamber. Removal of the N-terminal loops in ClpP variants abolishes the gating mechanism and enables even the isolated peptidase to indiscriminately destroy unfolded proteins. Functional control of degradation through complex formation with the ATPase is mediated by contacts formed with a ClpP ring at hydrophobic grooves located at peripheries from the axial pore. Although docking of the hexameric ClpA or ClpX ATPases to the heptameric ClpP rings involves a symmetry mismatch, binding of six conserved ATPase loops which contain the IGL or IGF sequence motif in ClpA or ClpX, respectively, ensures robust degradation activity.^{1,29–31} Weaker complexes with only five IGL/F loops are functional, albeit degradation proceeds at a reduced rate. The N-terminal loops are also involved in the complex formation and through interaction with the ATPase pore-2 loops, they form weaker contacts due to their greater conformational flexibility.^{30,32,33} Remarkably, pore opening may be affected without assistance from the ATPase through strongly cooperative binding of seven acyldepsipeptide 1 (ADEP1) molecules to the peripheral ClpP sites (Fig. 1).^{2,34} Kinetic studies indicate that ADEP1 establishes favorable interactions with the ClpP hydrophobic groove through regions (residues 123–148) of each protomer (Fig. 1). In the “closed” configuration, phenylalanine β -methylproline and alanine moieties and the aliphatic tail. This triggers a dramatic conformational transition to a quasi-symmetric configuration of N-terminal β -hairpins pointing outward from the proteolytic chamber that results in nearly doubling the pore diameter.^{26,34,36,37} Notably, the ClpP structure is very similar in the ADEP-bound and ClpX-bound configurations with a global, C α -based root-mean-square deviation of 0.8 Å in one ClpP ring and 0.6 Å in the pore region (defined by the seven N-terminal loops and adjacent helices). ClpP pore opening induced by ADEP binding yields powerful antibacterial action through uncontrolled destruction of bacterial proteins that is pursued for therapeutic applications against pathogenic *Staphylococcus aureus* and *Mycobacterium tuberculosis*.^{35,38–41} Structural plasticity of ClpP probed in

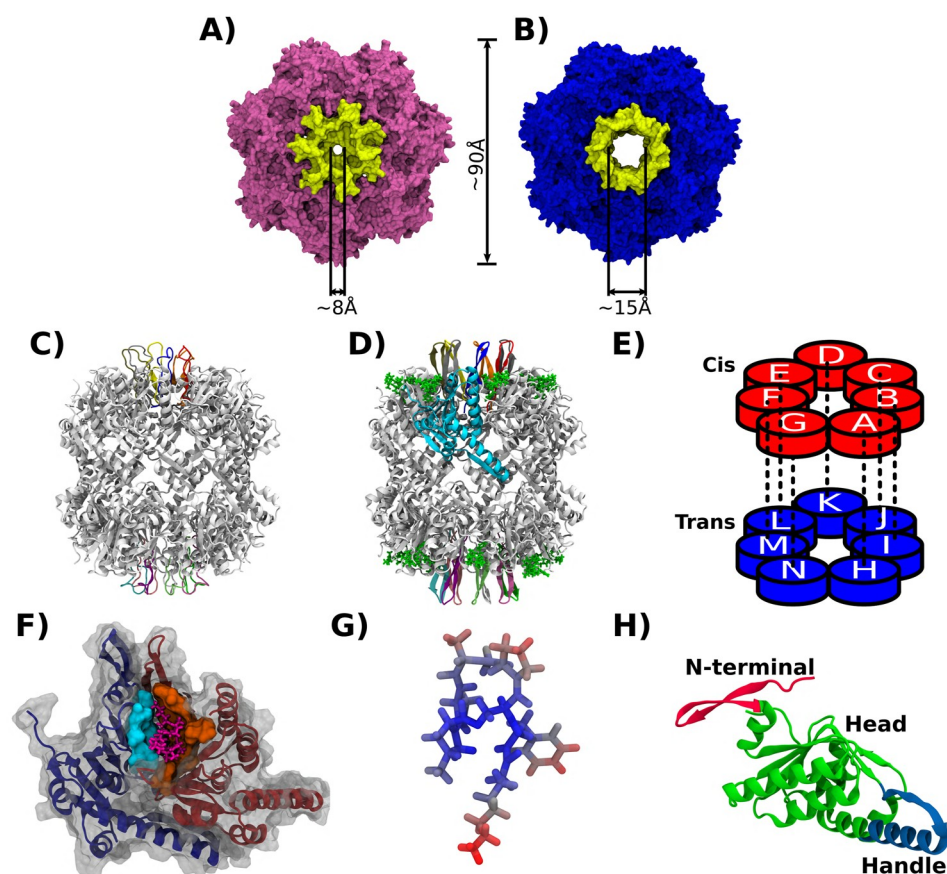


FIG. 1. Structural details of the closed and open pore configurations of ClpP. (a) and (b) The crystal structure of the (a) closed [Protein Data Bank (PDB) ID: 1YG6] and (b) open (PDB ID: 3MT6) pore configuration of ClpP [(a) pink; (b) blue] reveals dramatic gate-controlling conformational transitions of N-terminal loops (yellow) (c) and (d) Side view of the (c) closed and (d) open pore configurations of ClpP (gray). N-terminal loops (color-coded), ADEP1 molecules (green) and one ClpP protomer (cyan) are highlighted. (e) Protomer organization of the *cis* (red) and *trans* (blue) heptameric rings and cross-ring inter-protomer coupling (dashes), mediated by the handle region interface, are indicated. (f) Top view of the ADEP binding site consisting of the hydrophobic pocket (indicated as a molecular surface) formed at the interface between chains A (blue) and B (orange). (g) ADEP1 molecular structure. (h) Domain architecture of ClpP monomer. N-terminal (residues 1–19, red), head (residues 20–122 and 149–193, green), handle (residues 123–148, blue) are highlighted. Molecular images in this work are rendered using Visual Molecular Dynamics.⁶³

crystallographic and computational studies led to the hypothesis that transient exit channels form in the equatorial region to facilitate the release of peptide fragments resulting from the degradation process.^{2,42}

The details of allosteric communications that underlie the catalytic gate opening mechanism remain elusive even as multiple conformations have been resolved. Such challenges are commonly noted for ring-shaped molecular machines such as GroEL, thermosome and CCT chaperonins, and are attributable to their large conformational plasticity, dynamic rewiring of allosteric networks, and correlated intra- and inter-ring motions. To uncover the allosteric mechanisms in ClpP, we use computational approaches that are able to address such questions in diverse proteins by probing the complex networks of residue-residue interactions underlying the long-distance communication between the effector and the functional site.^{48–51} These approaches use concepts in graph theory combined with residue-residue coupling derived from structural data or conformational dynamics to identify “communities” of strongly coupled amino acids and to map the allosteric pathways connecting them.^{52–62}

In this paper, we describe comparative studies of ClpP conformational dynamics in its open- and closed-pore configurations. To this end, we perform solvated, atomistic, molecular dynamics simulations of these systems and we identify the collective motions using principal component and normal mode analyses. Coupling between regions of the tetradecameric structure revealed by community network analysis indicates a switch between inter- and intra-protomer coupling as a result of the transition from the closed to the open pore configuration. Allosteric paths identified between the ADEP binding sites and the ClpP N-terminal loops highlight stronger intra- and inter-ring coupling between binding sites and N-terminal loops in the open configuration.

II. METHODS

A. Molecular dynamics simulations

The closed and open pore configurations of *Escherichia coli* ClpP were described using the x-ray structures with Protein Data Bank IDs 1YG6⁶⁹ and 3MT6,⁷⁴ respectively. Unresolved regions of the x-ray structures were modeled using the Modeller server. To study the effect of perturbations on each structure, we considered several point mutations which were implemented using PyMOL (Table S1). For each configuration and sequence molecular dynamics (MD) simulations were performed using the Gromacs 2022 package and the CHARMM36 all-atom force field. The CGenFF server was used to generate force field parameters for the ADEP molecule compatible with the CHARMM36 force field. Each protein structure was solvated in a dodecahedral box with dimensions $\sim 122 \times 100 \times 100$ Å with water molecules represented using the CHARMM-modified TIP3P⁶⁹ model. The system was neutralized by adding an appropriate number of Na ions for each setup (Table S1). The solvated system was energy minimized using the steepest descent algorithm for 50 000 steps with the convergence criterion of the maximum force value smaller than 1000 kJ/(mol nm). Next, the systems were equilibrated by performing NVT and NPT dynamics. First, simulations were performed for 500 ps in the NVT ensemble using the V-rescale algorithm with $T = 300$ K, and by restraining the heavy atoms of the protein with a spring constant of

1000 kJ/(mol nm²). In the second equilibration step, the restrained system was simulated for 500 ps in the NPT ensemble, with the constant pressure of 1 atm using the Parrinello–Rahman algorithm.

The time step in all MD simulations was 2 fs. Finally, restraints were removed and for each setup, four unbiased MD simulation trajectories (50 ns each) were performed in the NPT ensemble. For each configuration, 2000 data frames were saved every 100 ps were saved after excluding the first 10 ns of each trajectory. Root-mean-square deviations of simulations in each setup are shown in Fig. S1.

B. Dynamic cross correlation matrix (DCCM)

To quantify the time-dependent directional correlations between residue pairs of the protein, we computed the Dynamic Cross Correlation Matrix (DCCM) of position fluctuations of C atoms of protein residues using the Bio3D package.⁷² DCCM is an $N \times N$ matrix, with N equal to the number of residues, where each element C_{ij} corresponds to the dynamic cross-correlation between residues i and j :

$$C_{ij}(t) = \frac{\langle \Delta \mathbf{r}_i(t') \cdot \Delta \mathbf{r}_j(t') \rangle}{(\langle \|\Delta \mathbf{r}_i(t')\|^2 \rangle \langle \|\Delta \mathbf{r}_j(t')\|^2 \rangle)^{1/2}}. \quad (1)$$

Here, $\mathbf{r}$ and $\mathbf{r}'$ are the spatial positions of the C atoms of residues i and j , as a function of time, $\langle \cdot \rangle$ denotes the ensemble average over all trajectories and all time frames up to time t , and $\Delta \mathbf{r}_i(t') = \mathbf{r}_i(t') - \langle \mathbf{r}_i \rangle$ denotes the instantaneous position fluctuations of residue i from its mean over the simulation time. Correlation values range from -1 to 1 , with positive C_{ij} values corresponding to motions of i and j atoms that are moving in the same direction, whereas the negative values indicate motions in opposite directions. Convergence of the DCCM matrix, shown in Fig. S2, was assessed using $R(t) = (1/N_p) \sum_{(ij)} (C_{ij}(t) - C_{ij}(t - \tau))^2$, where N_p is the number of residue pairs and the time interval $\tau = 5$ ns. We note that evaluation of $R(t)$ using several τ values between 2.5 and 10 ns consistently yield $R(t) \leq 0.001$. Here, C_{ij} is evaluated using data frames up to the total simulation time per trajectory $t \leq 50$ ns.

DCCM maps yield couplings between highly interconnected residue pairs that make it challenging to decipher correlated motion of larger protein regions. To address this limitation, we performed community detection using the strongly coupled residue pairs in DCCM ($|C_{ij}| \geq 0.6$) and the Girvan–Newman algorithm implemented in the `cna()` function in the Bio3D software package. In this approach, the full residue network is split into highly intra-connected communities but weakly inter-connected between communities. In the Girvan–Newman algorithm, the number of communities is selected using the maximum modularity approach. Modularity quantifies how well a network is partitioned into communities. Since this is a probability measure, the values are between 0 and 1. In general, for an ensemble MD derived correlation network, modularity above 0.7 indicates reasonable partitioning in a network.^{72,75}

C. Principal component analysis

Principal Component Analysis (PCA) is a standard technique for reducing high-dimensional datasets comprising atomic positions saved at each time step to glean information about the most significant

conformational dynamics at a coarse-grained level one can employ paths traversing through a given node used to identify the Principal Component Analysis.^{46,78} Here, we probe the functional important residues that contribute to the allosteric network. Analysis of node degeneracy indicates that ~350 paths between the source and sink are sufficient for convergence.⁵⁸

performing Principal Component Analysis (PCA). In PCA, the covariance matrix ($\langle \Delta \mathbf{r} \Delta \mathbf{r}^T \rangle$) comprising positional fluctuations is diagonalized to determine the set of eigenvalues and eigenvectors. Normal mode analysis and structural perturbation method tools *g_covar* and *g_anaeig*, where *g_covar* generates both eigenvalues and eigenvectors by diagonalizing the covariance matrix and *g_anaeig* filters the original trajectory and projects it along a given eigenvector. Prior to PCA calculation we remove translation and rotation degrees of freedom of the entire molecule by superimposing each frame of the MD simulation onto the crystal structure. The comparison between the essential subspaces corresponding to open and closed configurations is performed by calculating the Root Mean Square Inner Product (RMSIP), which computes the overlap between two subsets of eigenvectors, ψ_i and ψ_j , by using^{79,80}

$$\text{RMSIP}_M = \left(\frac{1}{M} \sum_{i,j=1}^M (\psi_i^A \cdot \psi_j^B) \right)^{1/2}. \quad (2)$$

Here, the overlap is computed such that the top M eigenvalues of each configuration account for over 80% of the total variance.

D. Optimal and suboptimal path analysis

To understand the dynamic coupling between the ADEP binding site and the N-terminal loop regions of the ClpP ring, we calculate the optimal and suboptimal paths traversing from each ADEP binding site to the N-terminal loops in one heptameric ring. We use the *cnpath()* function implemented in the Bio3D package.^{58,72,75} In the path analysis, each C_α atom is considered a node and the connection between nodes is weighted by $w_{ij} = -\log(|l_{ij}|/d)$. To remove weakly correlated regions and interactions between residues that are not in contact, the DCCM is filtered using the cutoff value $|l_{ij}| \geq 0.3$ and the dynamic contact map obtained in the MD trajectories. The contact map is generated in two steps. First, we identify persistent contacts in each trajectory. Residue pairs with the C_α distance $d \leq 10$ Å present in at least 75% of the simulation time frames.^{59,81} Next, the dynamic contact map is generated as the consensus map of contacts identified in at least three of the four trajectories. We find that the dynamic contact maps include $\approx 84\%$ of contacts present in the native structure and $\approx 2\%$ of new contacts. Paths are determined by using an efficient bidirectional approach that simultaneously initiates the search from one “source” (ADEP binding site) residue and one “sink” (N-terminal loop) residue and identifies closed paths upon locating common nodes. The path with the shortest length is identified as optimal, whereas slightly longer paths that are closed are identified as suboptimal. Accordingly, analysis of paths traversing between the “source” and the “sink” is useful to glean information about the allosteric regulation of regions that do not show large conformational changes. In our analysis, 300 paths are calculated for each “source” and “sink” residue pair and path length distributions are analyzed to assess the strength of the correlated motions. The extent of the overlap between two path length distributions, $p_i(x)$, $i = 1, 2$, is quantified by using the overlapping coefficient, $OC = \int \min\{p_1(x), p_2(x)\} dx$.^{82,83} Furthermore, the normalized node degeneracy, evaluated as the fraction

of paths traversing through a given node, is used to identify the important residues that contribute to the allosteric network. Analysis of node degeneracy indicates that ~350 paths between the source and sink are sufficient for convergence.⁵⁸

In order to calculate the normal modes of the proteins and analyze the response of the modes to perturbations, we modeled the proteins as elastic networks composed of N nodes where N is the number of amino acid residues in the structure. The nodes are placed at the locations of the C_α atoms of the amino acid residues in the corresponding PDB structures. All nodes that are within a cutoff distance, $R_c = 9$ Å are connected via harmonic potential with the energy function:

$$H = \frac{1}{2} \sum_{i,j: d_{ij} < R_c} \gamma (d_{ij} - d)^2, \quad (3)$$

where γ is the spring constant that defines the energy state, d is the dynamic distance between residues i and j , and d_{ij} is the distance between the residues in the PDB structure.

The normal mode calculation yields a set of $3N$ -dimensional eigenvectors $\mathbf{q}_i$ and corresponding eigenvalues ω_i^2 for each mode M . We also analyze the responses of these modes to perturbations that mimic to point mutations of specific amino acids. This approach has been termed the Structural Perturbation Method (SPM).^{86,87} In practice, we calculate the response to a mutation at the site i :

$$\delta \omega_M = \frac{1}{2} \sum_{j: d_{ij} < R_c} \delta \gamma (d_{ij} - d)^2, \quad (4)$$

where $\delta \gamma$ is the perturbed spring constant. The sum is over all other nodes in the network. The greater the response $\delta \omega$, the more dynamically significant a specific residue is to a given mode. We highlight the nodes that have $\delta \omega$ values that are three-fold above the average value for a mode as significant.

The overlap function quantifies how a given normal mode compares with the conformational change along a transition pathway. We compute the overlap function by projecting the eigenvector $\mathbf{q}$ of a given mode M onto the displacement vector between the open and closed configurations according to the formula,⁸⁸

$$I_M^{\text{open} \rightarrow \text{closed}} = \frac{\sum_{i=1}^{N_p} \mathbf{q}_{iM} \Delta \mathbf{r}_i^{\text{open} \rightarrow \text{closed}}}{\left[\sum_{i=1}^{N_p} \mathbf{q}_{iM}^2 \sum_{i=1}^{N_p} (\Delta \mathbf{r}_i^{\text{open} \rightarrow \text{closed}})^2 \right]^{1/2}}, \quad (5)$$

where the sum is over the N nodes, $\Delta \mathbf{r}_i^{\text{open} \rightarrow \text{closed}} = \mathbf{r}_i^{\text{open}} - \mathbf{r}_i^{\text{closed}}$, and $\mathbf{r}_i^{\text{open(closed)}}$ are position vectors of the i th node in the open (closed) structures, respectively. A value of one for the overlap corresponds to the direction given by the eigenvector being identical with that of $\Delta \mathbf{r}$. The relative amplitude of node i in mode M is obtained using $\delta \mathbf{q}_i = \frac{\Delta \mathbf{r}_i \cdot \mathbf{q}_i}{\sqrt{a_{i,M}^2 + a_{y,M}^2 + a_{z,M}^2}}$ where $a_{i,M}$, a_x , a_y , a_z , are the components of the normalized eigenvector $\mathbf{q}$.

III. RESULTS

A. Collective motions underlying the gate-opening conformational transition involve primarily N-terminal loop regions of ClpP

To glean the collective motions of ClpP protomers underlying the gate-opening transition, we use Normal Mode Analysis (NMA) and Principal Component Analysis (PCA) (see Sec. II). These approaches have proved highly effective in studies of conformational changes and dynamics of multisubunit biomolecules, as the ring-shaped chaperonins, ClpXP, or immunoglobulin. We use these analysis methods in conjunction to obtain detailed information about the collective motions as inter-residue couplings in NMA are restricted to an elastic network model of the native protein structure, whereas in PCA they reflect dynamic, i.e., native as well as non-native contacts. Nonlinear contributions to these couplings quantified by considering the mutual information between residue coordinates have also been used to probe allosteric networks.^{68,60,61,94–97} We note that a recent study of the tetracycline repressor dimer identified similar trends in the inter-residue coupling when comparing linear and nonlinear contributions.⁹⁷ A distinct advantage of network-based approaches is the ability to probe the propagation of allosteric signals at timescales accessible to MD simulations, yet yielding results consistent with much longer biological timescales.^{98,99}

In PCA, diagonalization of the covariance matrix of atomic fluctuations yields the set of independent modes of motions of protein that characterizes its essential space.⁷⁷ Eigenvectors of the matrix characterize the orthogonal directions of maximal variance, whereas eigenvalues determine the amplitude of conformational deviations. We focus on the PC modes that correspond to the largest eigenvalues (Fig. S3), which provide the major contribution to the variance of fluctuations. In the open state of ClpP, we find that eigenvalues corresponding to the top 3 PC modes contribute $\approx 40\%$ of the cumulative variance of the fluctuations, and eigenvalues of the top 20 modes are required to explain $\approx 67\%$ of the total variance. In the closed state, the top 3 eigenvalues contributed $\approx 52\%$ of the total variance, whereas the top 20 eigenvalues contribute $\approx 74\%$ of the total variance. In both cases, examination of the motions corresponding to the top 2 PC modes indicates a combination of large amplitude swing-type and torsional motions of the N-terminal loops that enable pore opening and closing [Figs. 2(a) and 2(b) and Movies SM1–4 in the supplementary material].

Handling domains in the inter-ring interface undergo slight contracting motions. More specifically, in the open state, PC1 corresponds to swing-like motions that underlie pore opening and closing and PC2 corresponds to torsional motions of N-terminal loops. In the closed configuration, PC1 corresponds to torsional motions and PC2 corresponds to a combination of swing-like and torsional motions.

To discern the contribution of harmonic vibrations to these motions we perform normal mode analysis of the open state conformation (see Sec. II). We focus on the subset of normal modes whose eigenvectors have the largest overlap (see Sec. II) with the conformational changes corresponding to the transition between open and closed conformations. As shown in Fig. 2(c), five modes have overlap $0.15 \leq \leq 0.2$, whereas all other modes have smaller contributions (≤ 0.1). The absence of a single, dominant, mode indicates the lack of strong coordination of motions of the several

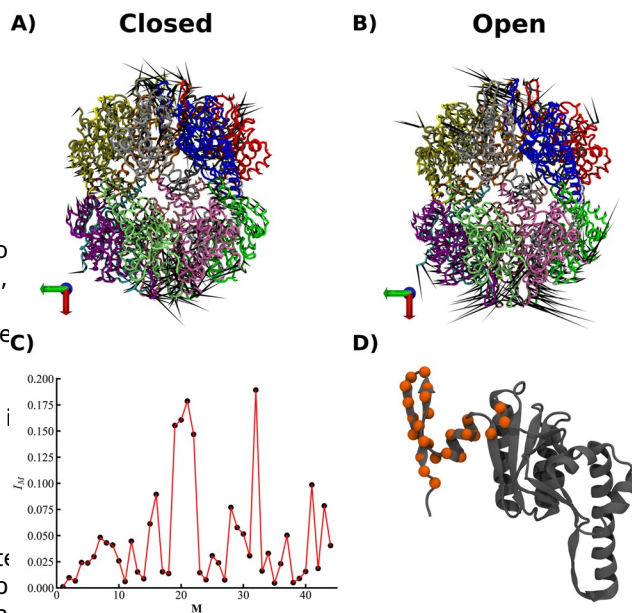


FIG. 2. Principal Component and Normal Mode Analysis for ClpP conformations (a) and (b) Motions corresponding to the first principal component (a) closed and (b) open pore configurations (see Movies SM1–6 in the supplementary material). (c) Overlap of normal modes of ClpP with amino acid displacements in the open $\rightarrow$ close transition. (d) The hot-spot residues (orange) extracted from the structural perturbation results. The list of hot-spot residues for the top two modes is shown in Table S2.

subunits of each ring to enable the transition. We also note that the lowest frequency modes, which describe global motions of the ClpP tetradecamer, have small contributions to the open $\rightarrow$ closed pore transition, whereas the five higher frequency modes, which describe more local motions, are more suitable to describe the transition. Consistently, the amplitudes of residue motions in the top five modes indicate large values only in the loop regions and negligible amplitudes outside of these regions (Fig. S4). These results reveal that the dominant conformational changes in this transition are primarily associated with motions of the N-terminal loops. On the basis of the comparison between PCA and NMA results, we surmise that dynamic fluctuations provide a stronger contribution in the handle and head regions than in the N-terminal loop regions. Hot-spot residues that are critical for allosteric communication are identified by employing the Structural Perturbation Method (see Sec. II). Here, the perturbations imposed by point mutations capture the change in the local network energy and hot-spot residues correspond to nodes whose response exceeds by three-fold the average value. Table S2 shows the hot-spot residues derived from the two modes with the largest overlap (28 and 28 (Fig. S5 and Movies SM5–6) and Fig. 2(d) illustrates the location of these hot-spot residues projected onto a single ClpP monomer. We note that NMA and SPM analysis identify hot-spot residues clustered primarily in the N-terminal loop region, in accord with the dominant structural flexibility of this region in the harmonic approximation. In addition to the hot-spot residue cluster within and near the

N-terminal loops, SPM analysis highlights two residues (His191, intra-subunit coupling within six subunits that involves strong coordination between the N-terminal loop, handle, and head regions Arg192) in the C-terminal region.

Both PCA and NMA results are consistent with observations of structural studies, which highlighted that the peptidase core, ClpP(20–193), has virtually identical conformations in the open and closed configurations. On this basis, it was proposed that ADEP1 binding causes a significant conformational change in the N-terminal loops and only small-amplitude motions of the equatorial belt formed between two rings. To quantify these contributions to the motions and to pinpoint the regions with the largest contribution to the open $\rightarrow$ closed transition, we further probe the PCs associated with the motions of the N-terminal loops and of the peptidase core, ClpP(20–193), respectively, in closed and open configurations. To this end, we perform separate PCA of each of the two ClpP regions in each of the two configurations and we determine the RMSIP of the two subsets of eigenvectors, which provides a measure of the similarity of motions described by the PC modes (see Sec. II).

Comparison of collective motions of N-terminal loops was computed by considering the eigenvectors corresponding to the largest 11 eigenvalues, which collectively account for 80% of the variance, thus representing the essential subspace. Quantitatively, the overlap between the essential subspaces corresponding to N-terminal regions indicates weak similarity, with $\text{RMSIP} \approx 0.37$. By contrast, for the peptidase core ClpP(20–193), where the top 110 eigenvectors must be included to describe the essential subspace, we find a strong overlap between the PC modes, with $\text{RMSIP} \approx 0.73$. This indicates that ADEP binding has only a weak effect on the dynamics of the peptidase core.

Overall, NMA and PCA data suggest that the conformational transition of ClpP protomers during the gate opening and closing is best described by an ensemble of motions. Taken together, both PC and normal mode data reveal that large conformation changes only exert at the N-terminal loop regions while the core remains intact.

B. Distinct coupling between ClpP regions in closed and open pore configurations

To investigate the correlated motions of regions in the ClpP tetradecamer, we employ the community network analysis, which uses the residue–residue coupling quantified by the directional cross-correlation map (DCCM, see Sec. II). DCCM maps are highly interconnected at the residue level, which makes it complicated to extract information for large systems. Therefore, we probe community network clusters by converting the atomic cross-correlations to a coarse-grained type community network using the Girvan–Newman clustering method to investigate both intra- and inter-protomer coupling. We select the cutoff of the cross-correlation between residues i and j , $|C_{ij}|$, such that the maximum modularity analysis requires the additional evaluation of suboptimal paths, (see Sec. II and Fig. S6) corresponds to a larger number of community clusters than the number of ClpP protomers in each of the three ClpP configurations. We find that this requirement is satisfied by $|C_{ij}| \geq 0.6$, which yields $\approx 30 - 40$ community clusters for ClpP configurations (Tables S3–S5). As shown in Fig. 3, the community network analysis reveals distinct patterns of domain coupling in ADEP-bound and unbound conformations of ClpP. Interestingly, the ADEP-bound open conformation is characterized by extensive

By contrast, the limited coupling is observed across the equatorial interface between protomers of the *cis* (A–G) and *trans* (H–N) rings (according to the ClpP protomer organization shown in Fig. 1), involving the handle regions of two protomer pairs (C–J and D–K). Removal of ADEP from the open pore conformation results in dramatic changes in intra-protomer coupling, which nearly abolishes the correlation between the handle region and the loop and inter-ring coupling between handle regions of protomer pairs is enhanced through coordinated motions of handle regions of five protomer pairs. These changes yield coupling patterns similar to the closed conformation, in which intra-protomer coupling is weaker than inter-ring coupling between handle regions of protomer pairs is the important role of the handle domain in tetradecamer formation and stacking of the two rings (*cis* and *trans*).

The Optimal and suboptimal path analysis reveal stronger coupling between allosteric and active sites of ADEP bound ClpP

The absence of large-scale rigid-body domain motions of ClpP protomers and the proximity of the ADEP binding sites to the N-terminal loops suggest that allosteric signals can be transmitted by relatively short, intra-protomer, paths. The ring structure of ClpP, however, also allows effective allosteric communication to take place between ADEP binding site of one protomer and N-terminal loops of the other intra-ring protomer. To probe these allosteric mechanisms in a quantitative manner, we use approaches that combine residue–residue positional correlations and concepts in graph theory (see Sec. II). To this end, we map the allosteric paths that connect one ADEP binding site, termed “source,” and the seven N-terminal loops within one ring, or “sink.” In the correlation network, each node represents one protein residue and the connecting edges have associated lengths that reflect the cross-correlation value (see Sec. II). The path length between nodes located in the source and sink is then identified as the sum of the lengths of all individual edges that connect these nodes. We emphasize that, given the construction of an allosteric map in the correlation space, the relative importance of the allosteric paths depends on the strength of the coupling between constituent residue pairs rather than their proximity in the physical space. In this framework, the shortest paths between the source and the sink reveal the strongest allosteric couplings within a protein. Interestingly, the minimal length, or “optimal,” path was not found to be very sensitive to changes in the protein configuration, therefore, analysis of allosteric paths may yield misleading conclusions about the signaling pathways. Instead, a comprehensive analysis requires the additional evaluation of suboptimal paths, which have slightly longer lengths than the optimal path. Using this framework, we computed 300 paths for each residue and this end, we select as source one ADEP binding site [Fig. 4(a)], comprising residues Val44, Leu48, Phe49, Glu51, Ala52, Phe82 from protomer 1 and Arg22, Leu23, Val28, Phe30, Tyr60, Tyr62, Ile90, Met92, Leu114, Leu189 from the adjacent protomer i , and as sink two representative residues, Thr10 and Gly13, located near the turn of each

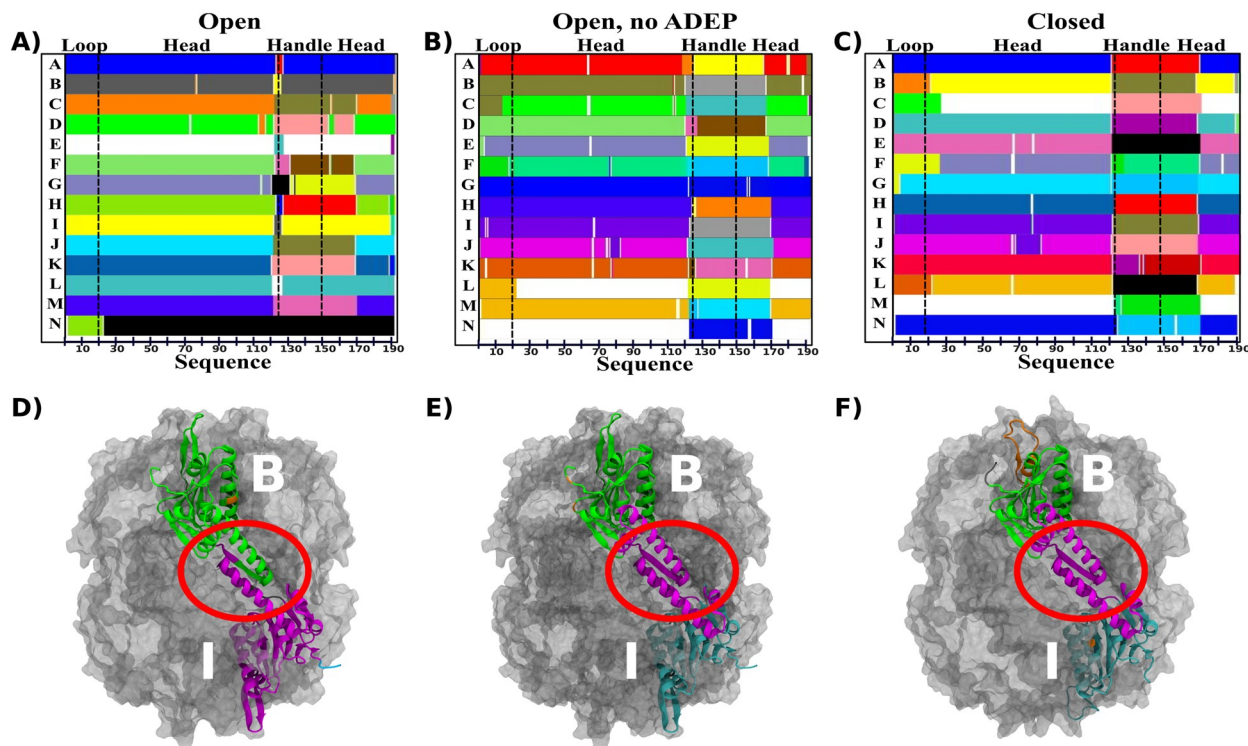


FIG. 3. Community network analysis for ClpP tetradecamer configurations. Community maps for (a) open pore (b) open, no ADEP bound (c) closed pore configurations. The N-terminal loop, head, and handle regions are indicated for the *cis* (protomers A–G) and *trans* (H–N) rings of ClpP. (d)–(f) Structural details of inter-protomer coupling across the ClpP rings. (d) Strong intra-protomer and weak inter-protomer coupling in the open pore configuration (e) and (f) Strong inter-protomer coupling mediated by the equatorial interface, including the handle region (highlighted by red ellipse), in (e) open, no ADEP and (f) closed pore configuration. Communities identified in (a)–(c) are highlighted.

N-terminal loop in the same (*cis*) ring. The selection of representative loop residues suffices for the purposes of mapping paths. The allosteric signal propagates within the loop exclusively through intra-loop residues. By considering paths connecting each ADEP binding site to each loop, we examined a total of 470 400 pathways traversing one ClpP ring (see Sec. I). In all three ClpP configurations, the optimal pathways are intra-protomeric and connect Arg22 to loop residues 13–17 (Table S3). Thus, optimal pathways are largely stable among the three ClpP configurations, which is in accord with observations made on small proteins noted above.

In order to characterize the strength of allosteric communication from a broad perspective, we examine the set of suboptimal paths in each pore configuration as well as the subsets of paths connecting each ADEP binding site to a specific loop [Fig. 4(a)]. As shown in Fig. 4(b), the histograms of suboptimal path lengths in the three ClpP configurations highlight the stronger coupling between binding sites and loops affected by ADEP binding, with the shortest paths corresponding to the open state. In contrast, in the closed state, the path length distribution is shifted toward longer paths, with lengths up to ≈ 8 , which indicates a weaker coupling between the ADEP binding site and the N-terminal loop. Perturbation of the open configuration through the removal of the ADEP molecules results in a slight shift of the suboptimal path length distribution toward paths of intermediate lengths between those found in the open and closed configurations.

Next, we probed the intra-ring propagation of the allosteric signal by examining the path length distributions corresponding to paths connecting each binding site to the N-terminal loop of each protomer [Figs. 4(c)–4(f)]. In accord with the above observations, path lengths corresponding to the open state are consistently shorter than those for the closed state. In both open and closed pore configurations, the shortest paths originating from the ADEP binding site of protomer *i* correspond to paths to the N-terminal loop of the same protomer [Figs. 4(c)–4(f)]. In the open state, we find relatively short paths, of length ≈ 2 , to the nearest neighbor loops in both clockwise (CW) and counterclockwise (CCW) directions. This strong intra-ring coupling supports the ability of the hexameric ATPase to trigger ClpP gate-opening even with the substoichiometric occupation of distal binding sites. Decreasing coupling strength is found in the N-terminal loops of the second and third nearest neighbor protomers, however, with slightly shorter paths in the CW direction [Figs. S7(A)–S7(C)]. In the closed state, we note the larger overlap than in the open state between path length distributions corresponding to neighboring loops and that of the same protomer loop, which indicates a slower intra-loop decay of the weaker allosteric signal in this state. In addition, the CW and CCW distributions corresponding to N-terminal loops

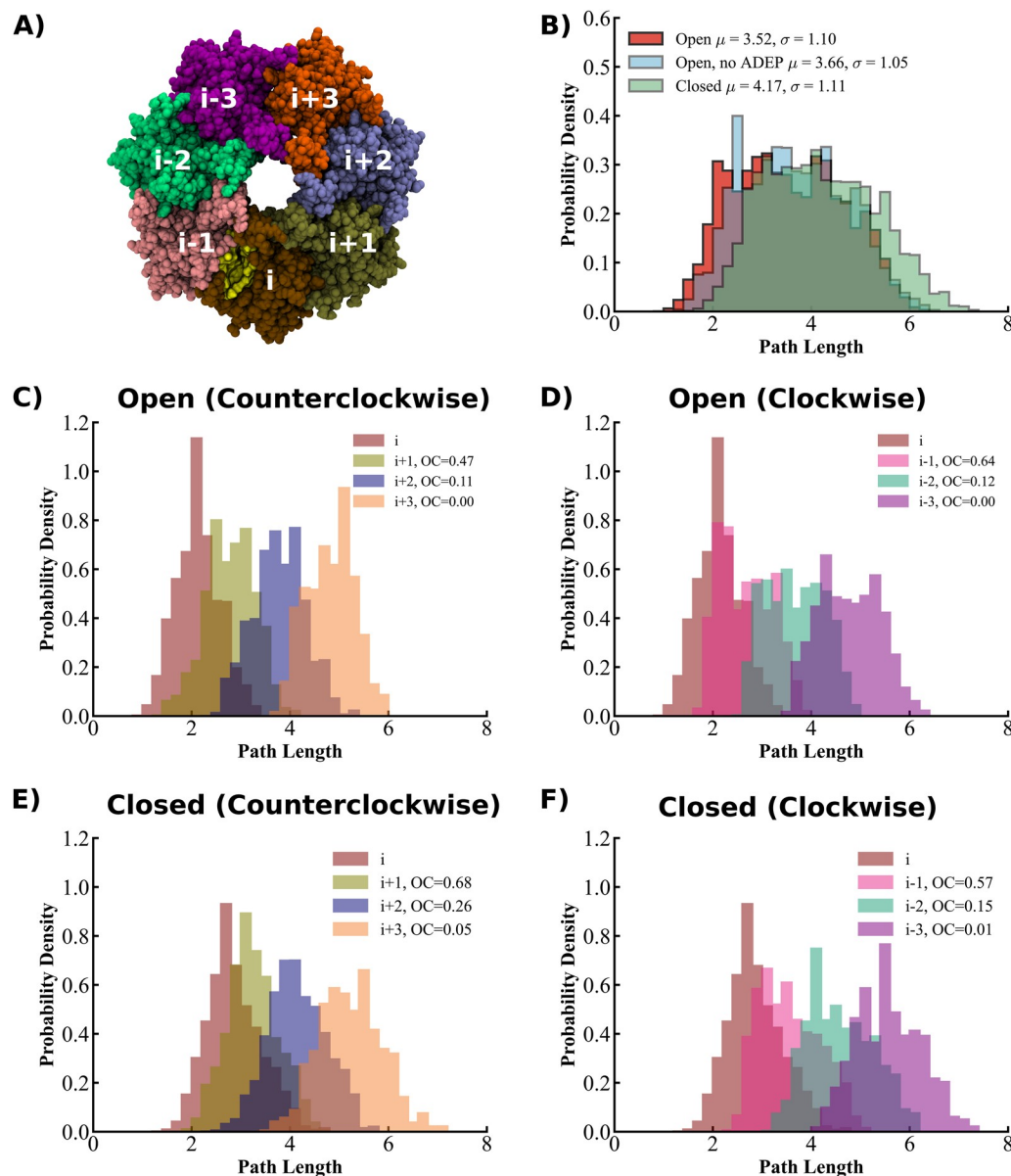


FIG. 4. Path length distributions in allosteric signaling in ClpP configurations. (a) Intra-ring protomer ordering relative to a given ADEP binding site (yellow). Allosteric paths are mapped between the binding site and the N-terminal loops in the same protomer i , and in protomers in counterclockwise ($+j$, $j = 1, 3$) and in clockwise ($-j$, $j = 1, 3$) directions. (b) Probability density distributions of path lengths are shown for the complete set of paths between each binding site and all intra-ring loops in the open (red); open, no ADEP (blue); and closed (green) configurations of ClpP. The legend indicates the mean and standard deviations of distributions. (c) and (d) Path length distributions of paths mapped in the open ClpP configuration between binding site i and loops in protomers in (c) counterclockwise and (d) clockwise directions are compared with the path length distribution of paths to the loop in protomer i . (e) and (f) Same as in (c) and (d) for the closed ClpP configuration. The legends indicate the overlapping coefficient, OC, between each distribution and the same protomer distribution.

of equidistant nearest neighbors overlap nearly completely [Figs. communication To obtain the microscopic understanding of the changes induced by perturbations, we probe the detailed paths in each configuration. To this end, we examine the three-dimensional configurations of ClpP, as well as the marked effect of perturbations on the paths associated with suboptimal paths (Fig. 5). The structural maps of the path lengths indicate a significant dynamic rewiring of allosteric pathways in the open configuration indicate strong

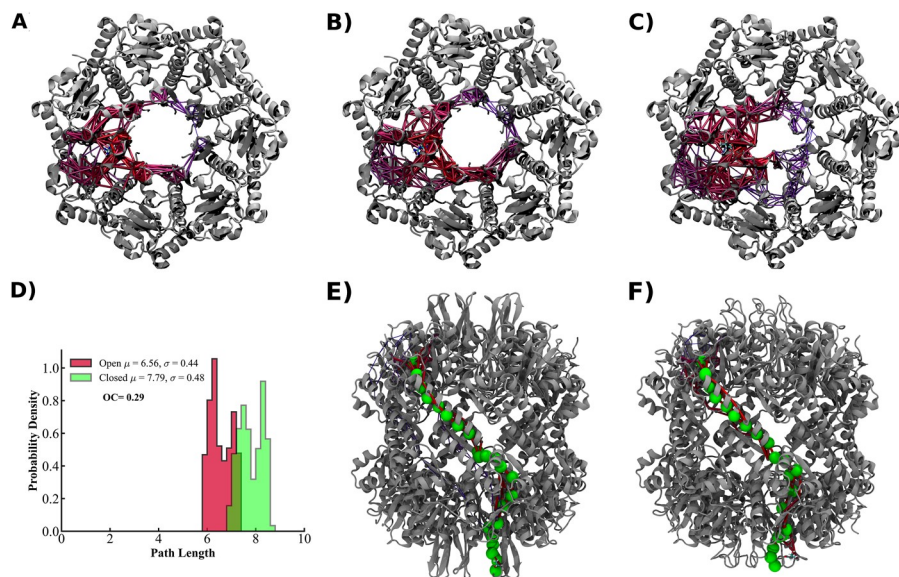


FIG. 5. Signaling pathways for ClpP configurations derived from the dynamic network analysis. (a)–(c) Suboptimal paths (red) between one ADEP binding site and all N-terminal loops in the same ClpP ring (gray) in (a) open; (b) open, no ADEP; and (c) closed pore configurations. (d) Probability density distributions of suboptimal inter-ring paths in open (red) and closed (green) pore configurations that connect the ADEP binding site of a *cis* ring protomer and the N-terminal loop in its *trans* ring protomer partner. The mean and standard deviations of distributions and the overlapping coefficient between them are indicated. (e) and (f) Structural details of the optimal inter-ring paths in (e) open and (f) closed pore configurations.

TABLE I. Normalized node degeneracy derived from suboptimal path analysis. Residues are grouped according to the ClpP protomer they belong to (shown as a subscript). Protomers are numbered in the counterclockwise direction, in the top view of the *cis* ring. In the path analysis, the ADEP binding site “source” includes residues from protomers i and $i-1$. Highlighted residues represent critical nodes (degeneracy ≥ 0.10) in all three ClpP configurations examined.

Residue	Ser21 ₋₁	Leu24 ₋₁	Ala45 ₋₁	Gln46 ₋₁	Leu50 ₋₁
Open	0.17	0.26	0.07	0.14	0.19
Open, no ADEP	0.17	0.26	0.08	0.14	0.15
Closed	0.14	0.17	0.11	0.14	0.17

Residue	Ile19	Tyr20	Ser21	Leu24	Ile29	Ser21 ₊₁
Open	0.22	0.11	0.19	0.11	0.10	0.06
Open, no ADEP	0.11	0.19	0.26	0.12	0.08	0.13
Closed	0.16	0.23	0.22	0.05	0.09	0.06

signaling propagated from the ADEP binding site to the nearest three N-terminal loops in the *cis* ring [Fig. 5(a)]. Removal of the ADEP molecules yields weaker coupling (indicated by thinner lines connecting the nodes) and increasing number of paths connecting the ADEP binding site to the more distant loops [Fig. 5(b)]. In the closed configurations, paths that connect multiple loops are increasingly found [Fig. 5(c)].

Interestingly, as shown in Fig. 5(d), the analysis of inter-ring pathways connecting the binding site of one protomer (B) in the *cis* ring and the N-terminal loops of its partner (I) in the *trans* ring, using $l_{ij} \geq 0.1$ reveals shorter paths and therefore tighter inter-ring coupling, in the open configuration compared with the closed configuration. In the open configuration, allosteric communication between rings is primarily mediated by paths with lengths ≤ 6.5 , which are not available in the closed configuration, and the overlap between the two distributions is small, with the overlapping coefficient $OC \approx 0.29$ [Fig. 5(d)]. Although, as noted in Sec. II B, the inter-ring handle interface is abolished in the open configuration, shorter paths become available through alternate routes, as indicated

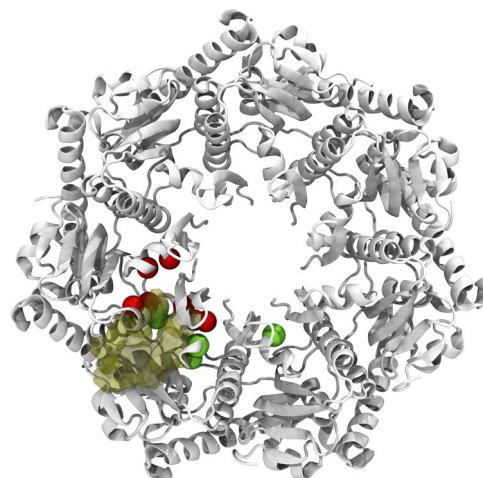


FIG. 6. Hot-spot residues for ClpP configurations derived from the dynamic network analysis. Residues with similar (distinct) node degeneracy in open and closed pore configurations are indicated with red (green) dots.

in Figs. 5(e) and 5(f). The shorter path lengths corresponding to the open configuration can be rationalized in terms of the connectivity of the intra-protomer communities [Figs. 3, 5(e), and 5(f)]. The differential gap between community networks of the *cis* and *trans* protomers results in a length of ≈ 5.9 for the optimal path in the open configuration and ≈ 6.9 for the closed pore configuration, as the structural details of the paths are similar (Table S7). We surmise that the handle interface, rather than facilitating inter-ring allosteric communication, acts as a constraint in the closed configuration,

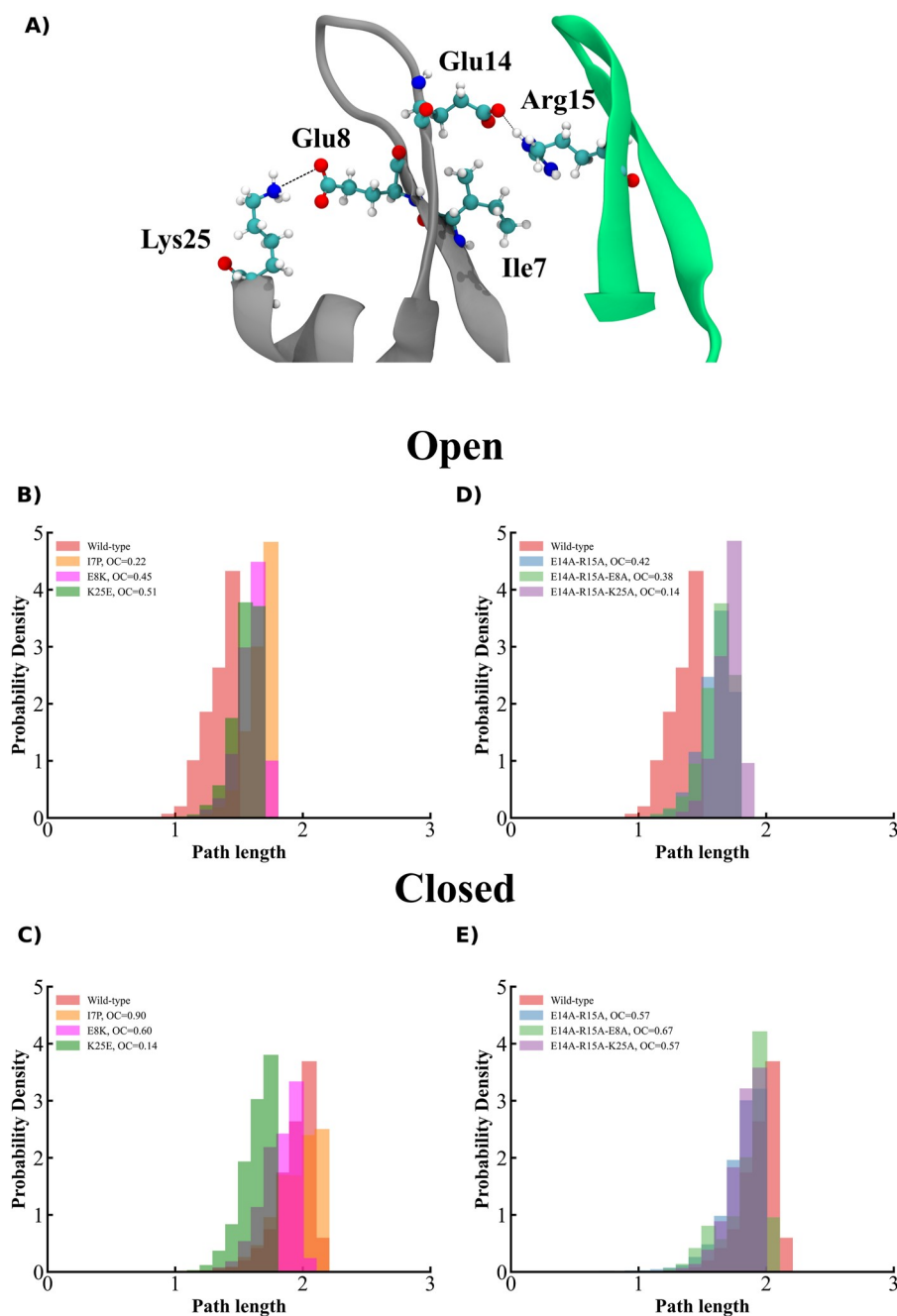


FIG. 7. Effect of mutations on allosteric signaling. (a) Single point-mutations considered, I7P, E8K, and K25E, affect the stability of individual loops (gray) and double and triple mutations, E14A-R15A, E14A-R15A-E8A, and E14A-R15A-K25A, affect both intra-loop and inter-loop (gray and green) stabilization. Salt bridges formed in the open pore configuration are highlighted. (b) and (c) Comparative probability distribution of the shortest 5000 paths in wild-type and single mutants in (b) open and (c) closed pore configurations. (d) and (e) The same as in (b) and (c) for double and triple mutants. Overlapping coefficients between each mutant and the wild-type distribution are indicated in the legend.

resulting in two intra-protomer path length penalties. In contrast, in the open configuration, the inter-protomer constraint is removed and a single path length penalty is applied.

Next, to identify residues that are critical for allosteric signaling, we compute the normalized intra-ring node degeneracy, the fraction of paths that include each node (see Sec. II C and Table I). Table I summarizes the residues in the *cis* ring that have the node degeneracy ≥ 0.10 in at least one configuration, as revealed by path analysis. The structural location of residues that act as critical nodes is highlighted in (Fig. 6). Notably, residues Ile19, Tyr20, Ser21, Ser21₋₁, and Leu24₋₁, at the base of N-terminal loops proximal to the ADEP binding site and Gln46₋₁ and Leu50₋₁, which are in the proximity of ADEP binding site, have node degeneracy ≥ 0.10 in all three setups and therefore are likely to have a critical contribution to allosteric communication. (In our notation, residue labels include as subscript the protomer location in the *cis* ring, with the “source” binding site occupying protomers $i - 1$ and i . Protomers are numbered in the counterclockwise direction in the top view of the ring.) One set of residues, Ile19, Ser21₋₁, Leu24₋₁, and Leu50₋₁, has a slightly reduced degeneracy in the closed pore configuration and upon removal of ADEP molecules in the open configuration, whereas the other residues have increased degeneracy. Interestingly, we also note that residue Gln46 shows no change in node degeneracy values in the three setups, indicating a weak sensitivity to structural perturbations. Structural studies revealed that the presence of Ile19 is crucial for the stability of the N-terminal loops and substrate translocation in the ADEP bound open state. We note that Ile19 is present in all three setups with a slightly higher value in the open setup.

D. N-terminal mutations differentially alter allosteric communication in open and closed pore configurations

We further explore how perturbations alter the coupling between the allosteric and active sites of ClpP by engineering mutations at N-terminal sites indicated to be functionally important (see Sec. II and Table S1).¹⁰⁰ We focus, on the one hand, on single point-mutations such as I7PE8K, and K25E, that alter the stability of individual N-terminal loops, and, on the other hand, on double, E14A-R15A, and triple, E14A-R15A-E8A and E14A-R15A-K25A, mutations that affect intra- as well as inter-loop stabilization (Fig. 7). To this end, we compare the path length distributions between each ADEP binding site and N-terminal loops of distant protomers. According to these results, in this configuration, neighboring intra-ring protomers are strongly coupled, consistent with the observed ability of the ATPase to trigger gate opening even as it activates only six binding sites. Interestingly, inter-ring coupling is also strengthened in the open configuration even as the hand-off interface present in the closed configuration between protomer partners is removed. Stronger coupling in the open configuration is affected through efficient crossing of a single gap between residue communities that reduces the penalties of crossing gaps between multiple communities in the closed configuration.

Allosteric communication in the ClpP peptidase is likely to be further modulated by two external factors, namely its interactions with the ATPase partner, such as the single-ring ClpX or the double-ring ClpA, and with the substrate protein being degraded. The effect of the ATPase interaction reflects the variability of IGL/IGF loop binding to the seven ClpP binding sites during the catalytic

IV. DISCUSSION

Our computational studies probe the allosteric mechanisms of the ClpP peptidase in response to effects that activate its gate-opening conformational transition. Intriguingly, this transition involves limited structural rearrangement outside of the gate-controlling N-terminal loop region. How is the allosteric signal propagated in the absence of large-scale rigid-body motions of ClpP subunits? To address this question, we undertook comparative studies of the open and closed pore configurations of ClpP by performing equilibrium MD simulations of each of these states. Our results, quantified through principal component and normal mode analysis, highlight the similarity of motions of the peptidase core in the two states even as the loop motions are significantly different. In our analysis, both structural perturbation, derived from Normal Mode calculations, and node degeneracies, computed using the positional cross-correlations, highlighted a set of hot-spot residues that are critical for the gate opening transition of ClpP. We find that the hot-spots derived from the harmonic approximation used in NMA reveal regions that are highly flexible and dynamic through their proximity to the N-terminal loop regions or the C-terminal regions. Node degeneracy values derived from dynamic cross-correlations account for hot-spot residues that are distributed in both N-terminal loops and protease core. Here we note that residues Ile19, Tyr20, and Leu24 were common to both SPM and node degeneracy calculations indicating these residues as critical for the allosteric regulation. Our results are in agreement with structural studies, which have shown that large non-polar side chains of Ile are critical for the integrity of the N-terminal loops.

Our detailed analysis of intra- and inter-ring allosteric pathways reveals stronger communication in the open configuration between each ADEP binding site and N-terminal loops of distant protomers. According to these results, in this configuration, neighboring intra-ring protomers are strongly coupled, consistent with the observed ability of the ATPase to trigger gate opening even as it activates only six binding sites. Interestingly, inter-ring coupling is also strengthened in the open configuration even as the hand-off interface present in the closed configuration between protomer partners is removed. Stronger coupling in the open configuration is affected through efficient crossing of a single gap between residue communities that reduces the penalties of crossing gaps between multiple communities in the closed configuration.

Allosteric communication in the ClpP peptidase is likely to be further modulated by two external factors, namely its interactions with the ATPase partner, such as the single-ring ClpX or the double-ring ClpA, and with the substrate protein being degraded. The effect of the ATPase interaction reflects the variability of IGL/IGF loop binding to the seven ClpP binding sites during the catalytic

cycle. Given the asymmetric binding of the six ATPase loops to the seven binding sites of ClpP, it is plausible that the signaling induced by ADEP binding represents the upper bound to the coupling strength between binding sites and the ClpP N-terminal loop. Additionally, the non-concerted conformational transitions of the ATPase hexamer further weakens the allosteric coupling and the ring symmetry. Nevertheless, asymmetric intra-ring allostery may support ClpP's active internalization of the polypeptide chain in the degradation process through non-concerted conformational changes of the pore loops. In support of the active action studies using a ClpAP complex with one or more IGL loops of ClpA covalently crosslinked to the ClpP binding sites allow degradation to proceed at a slightly reduced rate compared with the noncovalent ClpAP complex.¹¹

Allosteric communication within the ATPase itself has a high complexity and can therefore give rise to multiple responses in ClpP. An illustration of the complex ATPase allostery is provided by the double-ring ClpB disaggregase, which is not a cellular partner of ClpP, but can be engineered to form a complex with ClpP. Community network analysis of apo, nucleotide and/or substrate-bound configurations of ClpB revealed distinct intra-ring communication within the nucleotide-binding domain (NBD) 1 and 2 rings. Whereas, in the NBD1 ring, strong coupling is found between large and small subdomains of neighboring subunits, in the NBD2 ring intra-protomer coupling is dominant.

The interaction between the ClpP peptidase and its substrate protein partner also has the potential to effect changes in the allosteric signaling. As noted in a recent study, ClpP forcefully grips the titin substrates with forces that exceed those of the native ATPase.¹⁰³ Such strong signaling may dramatically alter the allosteric paths and break the symmetry of communications between the ADEP binding site and N-terminal loops.

SUPPLEMENTARY MATERIAL

See the [supplementary material](#) for Table S1 showing the summary of the MD setups for ClpP wild-type and mutant setups, Table S2 for hot-spot residues derived from the structural perturbation method, Tables S3–S5 for highly correlated residue communities and Tables S6 and S7 for the list of optimal paths in each ClpP configuration. Fig. S1 for RMSD time series in each ClpP configuration, Fig. S2 for DCCM convergence computed over multiple trajectories, Fig. S3 for the largest 20 eigenvalues of the PC modes for each ClpP configuration, Fig. S4 for normalized amplitudes of amino-acid motions associated with the top five normal modes, Fig. S5 for motions associated with top normal modes of the ClpP tetradecamer, Fig. S6 for network modularity, Fig. S7 for path length distributions of equidistant protomer loops, Movies SM1–4 for motions corresponding to principal components in the open (SM1–2) and closed pore configurations (SM3–4) and Movies SM5–6 for motions associated with the top two normal modes.

ACKNOWLEDGMENTS

The authors gratefully acknowledge stimulating discussions with Sue Wickner and Mike Maurizi. This work has been supported by the National Science Foundation Grants Nos. MCB-1516918

and MCB-2136816 to G.S. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by NSF Grant No. ACI-1548562. XSEDE Bridges resources at the Pittsburgh Supercomputer Center were used through allocation TG-MCB170020 to G.S. Partial support was received through research cyberinfrastructure resources and services provided by the Advanced Research Computing center at the University of Cincinnati.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Ashwin Dayananda: Conceptualization (equal); Data curation (lead); Formal analysis (lead); Investigation (equal); Methodology (equal); Validation (equal); Visualization (equal); Writing – original draft (lead); Writing – review & editing (lead). **T. S. Hayden Dennison:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Validation (equal); Visualization (equal). **Hewafonsekage Yasar Fonseka:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Validation (equal); Visualization (equal). **Mohammad S. Avesta:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Validation (equal); Visualization (equal). **Qi Wang:** Conceptualization (equal); Investigation (supporting); Methodology (equal); Validation (supporting); Visualization (supporting). **Rajna Tehver:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **George Stan:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

1. S. Wickner, M. R. Maurizi, and S. Gottesman, *Science* **286**, 1888 (1999).
2. M. Gersch, K. Famulla, M. Dahmen, C. Göbl, I. Malik, K. Richter, V. S. Korotkov, P. Sass, H. Rübsamen-Schaeff, and T. Madl, *Nat. Commun.* **6**, 6320 (2015).
3. J. Ortega, H. S. Lee, M. R. Maurizi, and A. C. Steven, *J. Struct. Biol.* **146**, 217 (2004).
4. J. Wang, J. A. Hartling, and J. M. Flanagan, *Cell* **91**, 447 (1997).
5. X. Fei, T. A. Bell, S. Jenni, B. M. Stinson, T. A. Baker, S. C. Harrison, and R. T. Sauer, *eLife* **9**, e52774 (2020).
6. K. Katayama-Fujimura, S. Gottesman, and M. R. Maurizi, *J. Biol. Chem.* **262**, 4477 (1987).

- ⁷M. R. Maurizi, W. P. Clark, Y. Katayama, S. Rudikoff, J. Pumphrey, B. Bowers, and S. Gottesman, *J. Biol. Chem.* **265**, 12536 (1990).
- ⁸K. M. Woo, W. J. Chung, D. B. Ha, A. L. Goldberg, and C. H. Chung, *J. Biol. Chem.* **264**, 2088 (1989).
- ⁹M. W. Thompson, S. K. Singh, and M. R. Maurizi, *J. Biol. Chem.* **269**, 18209 (1994).
- ¹⁰S. Gottesman, M. R. Maurizi, and S. Wickner, *Cell* **91**, 435 (1997).
- ¹¹K. R. Schmitz, D. W. Carney, J. K. Sello, and R. T. Sauer, *Proc. Natl. Acad. Sci. U. S. A.* **111**, E4587 (2014).
- ¹²M. E. Lee, T. A. Baker, and R. T. Sauer, *J. Mol. Biol.* **399**, 707 (2010).
- ¹³M. W. Thompson and M. R. Maurizi, *J. Biol. Chem.* **269**, 18201 (1994).
- ¹⁴J. A. Alexopoulos, A. Guarné, and J. Ortega, *J. Struct. Biol.* **179**, 202 (2012).
- ¹⁵A. Martin, T. A. Baker, and R. T. Sauer, *Mol. Cell* **27**, 41 (2007).
- ¹⁶G. Effantin, M. R. Maurizi, and A. C. Steven, *J. Biol. Chem.* **285**, 14834 (2010).
- ¹⁷A. Martin, T. A. Baker, and R. T. Sauer, *Nat. Struct. Mol. Biol.* **15**, 1147 (2008).
- ¹⁸C. Lee, M. P. Schwartz, S. Prakash, M. Iwakura, and A. Matouschek, *Mol. Cell* **6**, 627 (2001).
- ¹⁹M.-E. Aubin-Tam, A. O. Olivares, R. T. Sauer, T. A. Baker, and M. J. Lang, *Cell* **145**, 257 (2011).
- ²⁰R. A. Maillard, G. Chistol, M. Sen, M. Righini, J. Tan, C. M. Kaiser, C. Hodges, A. Martin, and C. Bustamante, *Cell* **145**, 459 (2011).
- ²¹A. Kravats, M. Jayasinghe, and G. Stan, *Proc. Natl. Acad. Sci. U. S. A.* **108**, 2234 (2011).
- ²²A. O. Olivares, A. R. Nager, O. Iosefson, R. T. Sauer, and T. A. Baker, *Nat. Struct. Mol. Biol.* **21**, 871 (2014).
- ²³A. Javidialesaadi, S. M. Flournoy, and G. Stan, *J. Phys. Chem. B* **123**, 2623 (2019).
- ²⁴H. Y. Y. Fonseka, A. Javidi, L. F. L. Oliveira, C. Micheletti, and G. Stan, *J. Phys. Chem. B* **125**, 7335 (2021).
- ²⁵R. A. Varikoti, H. Y. Y. Fonseka, M. S. Kelly, A. Javidi, M. Damre, S. Mullen, J. Nugent, C. M. Gonzales, G. Stan, and R. I. Dima, *Nanomaterials* **12**, 1849 (2022).
- ²⁶L. D. Jennings, J. Bohon, M. R. Chance, and S. Licht, *Biochemistry* **47**, 1031 (2008).
- ²⁷S. G. Kang, M. R. Maurizi, M. Thompson, T. Mueser, and B. Ahvazi, *J. Struct. Biol.* **148**, 338 (2004).
- ²⁸A. Gribun, M. S. Kimber, R. Ching, R. Sprangers, K. M. Fiebig, and W. A. Houry, *J. Biol. Chem.* **280**, 16185 (2005).
- ²⁹M. C. Bewley, V. Graziano, K. Griffin, and J. M. Flanagan, *J. Struct. Biol.* **153**, 113 (2006).
- ³⁰Y.-I. Kim, I. Levchenko, K. Fraczowska, R. V. Woodruff, R. T. Sauer, and T. A. Baker, *Nat. Struct. Mol. Biol.* **8**, 230 (2001).
- ³¹D. Y. Kim and K. K. Kim, *J. Biol. Chem.* **278**, 50664 (2003).
- ³²Z. Maglica, K. Kolygo, and E. Weber-Ban, *Structure* **17**, 508 (2009).
- ³³A. J. Amor, K. R. Schmitz, J. K. Sello, T. A. Baker, and R. T. Sauer, *ACS Chem. Biol.* **11**, 1552 (2016).
- ³⁴D. H. S. Li, Y. S. Chung, M. Gloyd, E. Joseph, R. Ghirlando, G. D. Wright, Y.-C. Cheng, M. R. Maurizi, A. Guarné, and J. Ortega, *Chem. Biol.* **17**, 959 (2010).
- ³⁵H. Brötz-Oesterhelt, D. Beyer, H.-P. Kroll, R. Endermann, C. Ladel, W. Schroeder, B. Hinzen, S. Raddatz, H. Paulsen, K. Henninger, E. Bandow, H.-G. Sahl, and H. Labischinski, *Nat. Med.* **11**, 1082 (2005).
- ³⁶M. C. Bewley, V. Graziano, K. Griffin, and J. M. Flanagan, *J. Struct. Biol.* **165**, 118 (2009).
- ³⁷S. Vahidi, Z. A. Ripstein, M. Bonomi, T. Yuwen, M. F. Mabanglo, J. B. Juravsky, K. Rizzolo, A. Velyvis, W. A. Houry, M. Vendruscolo, J. L. Rubinstein, and L. E. Kay, *Proc. Natl. Acad. Sci. U. S. A.* **115**, E6447 (2018).
- ³⁸J. Kirstein, A. Hoffmann, H. Lilie, R. Schmidt, H. Rübsem-Waigmann, H. Brötz-Oesterhelt, A. Mogk, and K. Turgay, *EMBO Mol. Med.* **1**, 37 (2009).
- ³⁹P. Sass, M. Josten, K. Famulla, G. Schiffer, H.-G. Sahl, L. Hamoen, and H. Brötz-Oesterhelt, *Proc. Natl. Acad. Sci. U. S. A.* **108**, 17474 (2011).
- ⁴⁰R. M. Raju, A. L. Goldberg, and E. J. Rubin, *Nat. Rev. Drug Discovery* **11**, 726 (2012).
- ⁴¹B. P. Conlon, E. S. Nakayasu, L. E. Fleck, M. D. LaFleur, V. M. Isabella, K. Coleman, S. N. Leonard, R. D. Smith, J. N. Adkins, and K. Lewis, *Nature* **503**, 365 (2013).
- ⁴²J. Zhang, F. Ye, L. Lan, H. Jiang, C. Luo, and C.-G. Yang, *J. Biol. Chem.* **286**, 37590 (2011).
- ⁴³O. Keskin, I. Bahar, D. Flatow, D. G. Covell, and R. L. Jernigan, *Biochemistry* **41**, 491 (2002).
- ⁴⁴W. Zheng, B. R. Brooks, and D. Thirumalai, *Biophys. J.* **93**, 2289 (2007).
- ⁴⁵M. Jayasinghe, P. Shrestha, X. Wu, R. Tehver, and G. Stan, *Biophys. J.* **103**, 1285 (2012).
- ⁴⁶R. Gruber and A. Horovitz, *Chem. Rev.* **116**, 6588 (2016).
- ⁴⁷R. Gruber, M. Levitt, and A. Horovitz, *Proc. Natl. Acad. Sci. U. S. A.* **114**, 5189 (2017).
- ⁴⁸A. R. Atilgan and C. Atilgan, *Briefings Funct. Genomics* **11**, 479 (2012).
- ⁴⁹R. Nussinov, C.-J. Tsai, and B. Ma, *Annu. Rev. Biophys.* **42**, 169 (2013).
- ⁵⁰V. A. Feher, J. D. Durrant, A. T. Van Wart, and R. E. Amaro, *Curr. Opin. Struct. Biol.* **25**, 98 (2014).
- ⁵¹R. Guo and H.-X. Zhou, *Chem. Rev.* **116**, 6503 (2016).
- ⁵²A. R. Atilgan, P. Akan, and C. Baysal, *Biophys. J.* **86**, 85 (2004).
- ⁵³C. Chennubhotla and I. Bahar, *Mol. Syst. Biol.* **2**, e172 (2006).
- ⁵⁴A. Sethi, J. Eargle, A. A. Black, and Z. Luthey-Schulten, *Proc. Natl. Acad. Sci. U. S. A.* **106**, 6620 (2009).
- ⁵⁵I. Rivalta, M. M. Sultan, N. S. Lee, G. A. Manley, J. P. Loria, and V. S. Batista, *Proc. Natl. Acad. Sci. U. S. A.* **109**, E1428 (2012).
- ⁵⁶Y. Miao, S. E. Nichols, P. M. Gasper, V. T. Metzger, and J. A. McCammon, *Proc. Natl. Acad. Sci. U. S. A.* **110**, 10982 (2013).
- ⁵⁷P. M. Gasper, B. Fuglestad, E. A. Komives, P. R. L. Markwick, and J. A. McCammon, *Proc. Natl. Acad. Sci. U. S. A.* **109**, 21216 (2012).
- ⁵⁸G. Scarabelli and B. J. Grant, *Biophys. J.* **107**, 2204 (2014).
- ⁵⁹A. T. Van Wart, J. Durrant, L. Votapka, and R. E. Amaro, *J. Chem. Theory Comput.* **10**, 511 (2014).
- ⁶⁰G. Palermo, C. G. Ricci, A. Fernando, R. Basak, M. Jinek, I. Rivalta, V. S. Batista, and J. A. McCammon, *J. Am. Chem. Soc.* **139**, 16028 (2017).
- ⁶¹G. Stetz and G. M. Verkhivker, *PLoS Comput. Biol.* **13**, e1005299 (2017).
- ⁶²G. M. Verkhivker and L. Di Paola, *J. Phys. Chem. B* **125**, 850 (2021).
- ⁶³W. Humphrey, A. Dalke, and K. Schulten, *J. Mol. Graph.* **14**, 33 (1996).
- ⁶⁴A. Fiser and A. Sali, *Bioinformatics* **19**, 2500 (2003).
- ⁶⁵Schrödinger, LLC, The PyMOL molecular graphics system, version 1.8, 2015.
- ⁶⁶D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, *J. Comput. Chem.* **26**, 1701 (2005).
- ⁶⁷R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, and A. D. MacKerell, *J. Chem. Theory Comput.* **8**, 3257 (2012).
- ⁶⁸K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, et al., *J. Comput. Chem.* **31**, 671 (2010).
- ⁶⁹E. Neria, S. Fischer, and M. Karplus, *J. Chem. Phys.* **105**, 1902 (1996).
- ⁷⁰G. Bussi, D. Donadio, and M. Parrinello, *J. Chem. Phys.* **126**, 014101 (2007).
- ⁷¹M. Parrinello and A. Rahman, *J. Appl. Phys.* **52**, 7182 (1981).
- ⁷²B. J. Grant, A. P. C. Rodrigues, K. M. ElSawy, J. A. McCammon, and L. S. D. Caves, *Bioinformatics* **22**, 2695 (2006).
- ⁷³H. Kamberaj and A. Van der Vaart, *Biophys. J.* **96**, 1307 (2009).
- ⁷⁴M. Girvan and M. E. J. Newman, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 7821 (2002).
- ⁷⁵B. J. Grant, L. Skjærven, and X. Q. Yao, *Protein Sci.* **30**, 20 (2021).
- ⁷⁶S. Hayward and L. Groot, *Molecular Modeling of Proteins* (Springer, 2008), 89–106.
- ⁷⁷A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen, *Proteins* **17**, 412 (1993).
- ⁷⁸M. A. Balsera, W. Wriggers, Y. Oono, and K. Schulten, *J. Phys. Chem.* **100**, 2567 (1996).
- ⁷⁹A. Amadei, B. L. De Groot, M.-A. Ceruso, M. Paci, A. Di Nola, and H. J. C. Berendsen, *Proteins* **35**, 283 (1999).
- ⁸⁰R. Cossio-Pérez, J. Palma, and G. Pierdominici-Sottile, *J. Chem. Inf. Model.* **57**, 226 (2017).
- ⁸¹X.-Q. Yao, R. U. Malik, N. W. Griggs, L. Skjærven, J. R. Traynor, S. Sivaramakrishnan, and B. J. Grant, *J. Biol. Chem.* **291**, 4742 (2016).
- ⁸²L. E. Bradley, "Overlapping coefficient," in *Encyclopedia of Statistical Sciences* (John Wiley, New York, 1985), Chap. 6, pp. 546–547.

- ⁸³ M. S. Weitzman, "Measures of overlap of income distributions of white and negro families in the United States," Technical Paper 2212, U.S. Bureau of the Census, 1970).
- ⁸⁴ M. M. Tirion, *Phys. Rev. Lett.* **77**, 1905 (1996).
- ⁸⁵ I. Bahar and A. Rader, *Curr. Opin. Struct. Biol.* **15**, 586 (2005).
- ⁸⁶ W. Zheng, B. R. Brooks, and D. Thirumalai, *Proc. Natl. Acad. Sci. U. S. A.* **103**, 7664 (2006).
- ⁸⁷ R. Tehver, J. Chen, and D. Thirumalai, *J. Mol. Biol.* **387**, 390 (2009).
- ⁸⁸ O. Marques and Y.-H. Sanejouand, *Proteins* **23**, 557 (1995).
- ⁸⁹ L. Skjaerven, A. Martinez, and N. Reuter, *Proteins* **79**, 232 (2011).
- ⁹⁰ M. Damre, A. Dayananda, R. A. Varikoti, G. Stan, and R. I. Dima, *Biophys. J.* **120**, 3437 (2021).
- ⁹¹ M. T. Posgai, S. Tonddast-Navaei, M. Jayasinghe, G. M. Ibrahim, G. Stan, and A. B. Herr, *Proc. Natl. Acad. Sci. U. S. A.* **115**, E8882 (2018).
- ⁹² L. González-Paz, C. Lossada, M. L. Hurtado-León, F. V. Fernández-Materán, J. L. Paz, S. Parvizi, R. E. Cardenas Castillo, F. Romero, and Y. J. Alvarado, *AC Omega* **8**(8), 7302–7318 (2023).
- ⁹³ O. F. Lange and H. Grubmüller, *Proteins* **62**, 1053 (2006).
- ⁹⁴ L. G. Ahuja, S. S. Taylor, and A. P. Kornev, *IUBMB Life* **71**, 685 (2019).
- ⁹⁵ M. C. R. Melo, R. C. Bernardi, C. De La Fuente-Nunez, and Z. Luthey-Schulten, *J. Chem. Phys.* **153**, 134104 (2020).
- ⁹⁶ L. Astl, G. Stetz, and G. M. Verkhivker, *J. Chem. Inf. Model.* **60**, 3616 (2020).
- ⁹⁷ Y. Yuan, J. Deng, and Q. Cui, *J. Am. Chem. Soc.* **144**, 10870 (2022).
- ⁹⁸ M. Iljina, H. Mazal, A. Dayananda, Z. Zhang, G. Stan, I. Riven, and G. Haran, *bioRxiv:2023.02.02.526786* (2023).
- ⁹⁹ V. Bhandari, K. S. Wong, J. L. Zhou, M. F. Mabanglo, R. A. Batey, and W. A. Houry, *ACS Chem. Biol.* **13**, 1413 (2018).
- ¹⁰⁰ J. Alexopoulos, B. Ahsan, L. Homchaudhuri, N. Husain, Y.-Q. Cheng, and J. Ortega, *Mol. Microbiol.* **90**, 167 (2013).
- ¹⁰¹ S. Kim, K. L. Zuromski, T. A. Bell, R. T. Sauer, and T. A. Baker, *eLife* **9**, e61451 (2020).
- ¹⁰² J. Weibezahn, C. Schlieker, B. Bukau, and A. Mogk, *J. Biol. Chem.* **278**, 32608 (2003).
- ¹⁰³ S. D. Walker and A. O. Olivares, *Biophys. J.* **121**, 3907 (2022).