

3-13-2024

β -sheets Mediate the Conformational Change and Allosteric Signal Transmission Between the AsLOV2 Termini

Sian Xiao
Southern Methodist University


Mayar Terek Ibrahim
Southern Methodist University

Gennady M. Verkhivker
Chapman University, verkhivk@chapman.edu

Brian D. Zoltowski
Southern Methodist University

Peng Tao
Southern Methodist University

Follow this and additional works at: https://digitalcommons.chapman.edu/scs_articles

 Part of the [Agronomy and Crop Sciences Commons](#), [Amino Acids, Peptides, and Proteins Commons](#), [Computational Chemistry Commons](#), and the [Other Chemistry Commons](#)

Recommended Citation

S. Xiao, M. T. Ibrahim, G. M. Verkhivker, B. D. Zoltowski, P. Tao, *J. Comput. Chem.* 2024, 1. <https://doi.org/10.1002/jcc.27344>

This Article is brought to you for free and open access by the Science and Technology Faculty Articles and Research at Chapman University Digital Commons. It has been accepted for inclusion in Mathematics, Physics, and Computer Science Faculty Articles and Research by an authorized administrator of Chapman University Digital Commons. For more information, please contact laughtin@chapman.edu.

β -sheets Mediate the Conformational Change and Allosteric Signal Transmission Between the AsLOV2 Termini

Comments

This article was originally published in *Journal of Computational Chemistry* in 2024. <https://doi.org/10.1002/jcc.27344>

Creative Commons License





This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Copyright

The authors

RESEARCH ARTICLE

β -sheets mediate the conformational change and allosteric signal transmission between the AsLOV2 termini

Sian Xiao¹  | Mayar Tarek Ibrahim¹ | Gennady M. Verkhivker²  |
Brian D. Zoltowski¹ | Peng Tao¹ 

¹Department of Chemistry, Center for Research Computing, Center for Drug Discovery, Design, and Delivery (CD4), Southern Methodist University, Dallas, Texas, USA

²Department of Biomedical and Pharmaceutical Sciences, Chapman University School of Pharmacy, Irvine, California, USA

Correspondence

Sian Xiao and Peng Tao, Department of Chemistry, Center for Research Computing, Center for Drug Discovery, Design, and Delivery (CD4), Southern Methodist University, Dallas, TX 75205, USA.
Email: sxiao@smu.edu and ptao@smu.edu

Funding information

National Institute of General Medical Sciences of the National Institutes of Health, Grant/Award Number: R15GM122013; National Institute of General Medical Sciences of the National Institutes of Health, Grant/Award Number: 2R15GM109282

Abstract

Avena sativa phototropin 1 light-oxygen-voltage 2 domain (AsLOV2) is a model protein of Per-Arnt-Sim (PAS) superfamily, characterized by conformational changes in response to external environmental stimuli. This conformational change begins with the unfolding of the N-terminal A' α helix in the dark state followed by the unfolding of the C-terminal J α helix. The light state is characterized by the unfolded termini and the subsequent modifications in hydrogen bond patterns. In this photoreceptor, β -sheets are identified as crucial components for mediating allosteric signal transmission between the two termini. Through combined experimental and computational investigations, the H β and I β strands are recognized as the most critical and influential β -sheets in AsLOV2's allosteric mechanism. To elucidate the role of these β -sheets, we introduced 13 distinct mutations (F490L, N492A, L493A, F494L, H495L, L496F, Q497A, R500A, F509L, Q513A, L514A, D515V, and T517V) and conducted comprehensive molecular dynamics simulations. In-depth hydrogen bond analyses emphasized the role of two hydrogen bonds, Asn482-Leu453 and Gln479-Val520, in the observed distinct behaviors of L493A, L496F, Q497A, and D515V mutants. This illustrates the role of β -sheets in the transmission of the allosteric signal upon the photoactivation of the light state.

1 | INTRODUCTION

The Per-Arnt-Sim (PAS) superfamily is widely present across the kingdoms of life, members of which are often found in pathways that regulate responses to environmental changes, including the circadian response pathway.^{1,2} As a subset of the PAS superfamily, the light, oxygen, or voltage sensing (LOV) domains are small sensor modules found in proteins that respond to environmental signals including light, redox potential, and oxygen, regulating various biological processes. *Avena sativa* phototropin 1 light-oxygen-voltage 2 domain (AsLOV2) serves as a model protein to study LOV function and

allostery. It contains a flavin mononucleotide (FMN) cofactor located in the center of the PAS fold. A large α -helical region of the AsLOV2 C-terminal to the PAS fold is termed the J α helix (Figure 1).³ AsLOV2 undergoes an allosteric activation from its resting dark state upon irradiation with blue light. A covalent bond forms between Cys450 side chain in the PAS fold and C(4a) of the FMN, resulting in the formation of cysteinyl-flavin adduct, referred to as the light state.⁴ This light activation also leads to a significant conformational change in the overall protein structure, including the unfolding of the J α helix. The reversion from the light state to the dark state for LOV domains is generally spontaneous and starts when irradiation ceases. This

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Authors. *Journal of Computational Chemistry* published by Wiley Periodicals LLC.

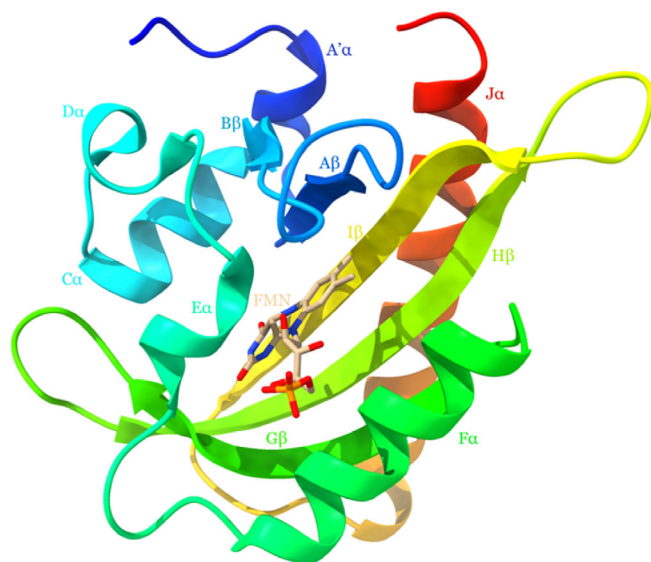


FIGURE 1 AsLOV2 protein structure (PDB ID: 2V1B).

process includes the thiol bond scission and a conformational change back to the dark state, occurring within seconds to hours, depending on the type of LOV domain.⁵ With its characteristic conformational changes upon light activation, AsLOV2 has been extensively studied as a model of designing optogenetic switches.^{6–8}

Past studies have identified the termini as the location of the most characteristic conformational changes in AsLOV2, including the undocking and unfolding of the J α helix at carboxyl terminal, and the unfolding of the A α helix at the amino terminal.⁹ However, the complete signaling propagation pathway remains elusive, with β -sheets being determined to play an important role in this process.^{2,10,11} Zayner et al.¹² carried out an extensive mutational analysis of AsLOV2 domain to study the relation between the sequence of the protein and its function, including 13 distinct mutations on H β sheet and I β sheet (F490L, N492A, L493A, F494L, H495L, L496F, Q497A, R500A, F509L, Q513A, L514A, D515V, and T517V). Photocycle time and conformational change of these mutations are analyzed but no obvious correlation exists between these two properties.¹² However, the ample experimental observations related to multiple mutations provide an opportunity for computational studies focusing on the allosteric mechanism underlying the activation process of the AsLOV2 domain.

The crystal structures are mostly limited to proteins in their equilibrium states. Therefore, little information of the light activation of AsLOV2 could be obtained directly from its crystal structures. Molecular dynamics (MD) simulations can capture the behavior of proteins in full atomic detail and at very fine temporal resolution, serving as important tools for understanding the physical basis of protein structure and function.^{13,14} In this study, extensive MD simulations were conducted on the selected mutations of AsLOV2 to investigate its allosteric mechanisms, focusing on the role of H β and I β sheets. Combining simulation analysis with the Markov state model,^{15–17} we identified some intriguing mutant behaviors potentially due to the changes of the hydrogen bond pattern.

2 | MATERIALS AND METHODS

2.1 | Molecular dynamics simulations

Initial AsLOV2 structures were sourced from the Protein Data Bank (PDB IDs: dark structure—2V1A, light structure—2V1B).¹⁸ In the light state, a covalent bond between C(4a) of FMN and Cys450 of AsLOV2 was introduced, along with the 13 mutations detailed in the previous section.

Simulation systems were prepared with the CHARMM¹⁹ program package (version c41b1), using force field parameters for the FMN from a previous work.²⁰ Protonation states of titratable residues were set at neutral pH, and protein systems were solvated in TIP3P²¹ water boxes, each with a 10 Å buffer zone. Sodium and chloride ions were added to each system to neutralize and to maintain an ionic strength of approximately 0.15 M. Structures underwent initial minimization using the steep descent method for 200 steps, followed by adopted basis Newton–Raphson minimization for 1000 steps. Subsequently, 24 picosecond (ps) simulations raised the temperature from 0 to 300 K. Systems were subjected to 10 ns isothermal–isobaric (NPT) equilibrations at 300 K (equilibrium simulation), followed by 1 μ s canonical (NVT) simulations (production simulation) at the same temperature. Snapshots of product simulations were saved every 100 ps. Three independent simulations were conducted for the wild type (WT) AsLOV2 in dark and light states, and for the light states of the 13 selected mutants.

The SHAKE²² constraint was applied in all simulations to maintain the length of bonds involving hydrogen atoms. During the simulation, the water molecules were kept rigid, with fixed bond lengths and angles. The long-range electrostatic interactions were accounted for using the particle mesh Ewald summation method.²³ The nonbonding interactions were cut off at 12 Å. The simulation was conducted using OpenMM²⁴ (version 7.6.0).

2.2 | Simulation analysis

Root-mean-square fluctuation (RMSF) evaluates the flexibility of individual residues. RMSF_{*i*} quantifies the fluctuation of residue *i* around its average position \bar{r}_i over *T* simulation frames. RMSF_{*i*} is calculated as:

$$\text{RMSF}_i = \sqrt{\frac{1}{T} \sum_{t=1}^T (r_i(t) - \bar{r}_i)^2} \quad (1)$$

where $r_i(t)$ is the coordinate for residue *i* in frame *t*.

Hydrogen bonds are identified using the Baker–Hubbard method, which sets cutoffs for distances between hydrogen atoms and acceptors (*r*) and angles between donors, hydrogens, and acceptors (θ).²⁵ The method considers NH and OH as donors, and O and N as acceptors. The hydrogen bonds analysis was carried out using MDTraj²⁶ (version 1.9.7). The following criterion and the occurrence fractions during simulations were employed for validation of the hydrogen bonds:

$$\theta > 120^\circ \text{ and } r < 2.5 \text{ \AA} \quad (2)$$

The dictionary of protein secondary structure (DSSP)²⁷ algorithm assigns standard secondary structure annotations to amino acid residues of a protein based on three-dimensional structure. It classifies residues into α -helices, β -sheets, turns, and bridges, using hydrogen-bonding patterns. The DSSP calculation is conducted by MDTraj using implementation of DSSP-2.2.0, giving the secondary structure assignments of each residue in each frame.

Dynamic cross correlation (DCC) analysis evaluates MD simulations trajectories to identify correlative motions between residues or residue groups within a molecule, crucial for understanding molecular function and interaction. It involves calculating correlation coefficients between the movements of pairs of atoms over the course of a simulation, revealing how the motion of one part of a molecule is coupled to another. The DCC between the p th and q th atoms is defined by the following equation:

$$\text{DCC}(p, q) = \frac{\langle \Delta r_p(t) \cdot \Delta r_q(t) \rangle_t}{\sqrt{\langle \|\Delta r_p(t)\|^2 \rangle_t} \sqrt{\langle \|\Delta r_q(t)\|^2 \rangle_t}} \quad (3)$$

where $r_p(t)$ denotes the vector of the p th atom's coordinates as a function of time t , $\langle \cdots \rangle_t$ means the time ensemble average, and $\Delta r_p(t) = r_p(t) - \langle r_p(t) \rangle_t$. The similar notation applies for the q th atom.

Essential dynamics of proteins distinguish between two configurational subspaces: the “essential” subspace, characterized by a few degrees of freedom and anharmonic motion that encompasses major positional fluctuations relevant to protein function, and the “physically constrained” subspace, where motion follows a narrow Gaussian distribution, indicating only local fluctuations.²⁸ Essential dynamics analysis is performed using ProDy package^{29,30} (version 2.4.0).

2.3 | Markov state model

Examining MD simulations provide valuable insights into the conformational landscape and dynamic movements of biomolecules. However, the complexity and high dimensionality of data from these simulations can obscure the crucial dynamic signatures.³¹ To address this challenge and improve conformational landscape analysis, time-lagged independent component analysis (t-ICA) is employed for dimensionality reduction, projecting extensive MD simulation data onto a more intuitive and interpretable low-dimensional space.

The t-ICA method identifies the slowest degrees of freedom, preserving kinetic information in the MD trajectories. It finds coordinates of maximal autocorrelation at a specific lag time, making it an excellent choice for dimensionality reduction in MD data transformation for Markov state model construction.^{32–34}

After obtaining the low-dimensional representation from t-ICA analysis, the k -means clustering method is applied to the projected low-dimensional space. Clusters identified in this analysis are referred to as microstates, with each simulation frame assigned to a microstate. The next step tests the kinetical relevance of these microstates

and identifies an appropriate lag time τ . Relaxation timescales, or implied timescales, are examined against various lag time to generate a model that satisfies the Markov assumption. Upon determining an appropriate lag time, microstates that can transit within the relaxation timescale are clustered together as one macrostate. Perron-Cluster cluster analysis was used to identify macrostates, that is, kinetically separate equilibrium states.^{35,36} Further theoretical details are available in a previous study.³⁷ t-ICA and Markov state model building were conducted using the PyEMMA package³⁸ (version 2.5.12).

3 | RESULTS

3.1 | AsLOV2 wild type

The RMSF of alpha Carbons (C_α) were calculated for the WT AsLOV2 dark and light states. The RMSF values reported in Figure 2A represent averages from three independent simulations. RMSF based on each independent simulation of each state is provided in Figure S1. Enhanced fluctuations in the light state, observed in both A' α and J α helices compared with the dark state during simulations, indicate increased flexibility in these regions upon light activation and formation of the FMN-Cys450 covalent bond. Additionally, RMSF elevation is observed in the loop connecting the H β sheet to the I β sheet, the terminal residues of the I β sheet, and the loop between the I β sheet and J α helix. Mean RMSF values for individual residues in the A' α and J α helices, and H β and I β sheets, were calculated for further comparison (Figure 2B). The A' α helix at N-terminal shows the highest mean RMSF values among the selected secondary structures in both states. Furthermore, the A' α helix has a higher mean RMSF value in the light state than the dark state, indicating heightened mobility upon light exposure. The J α helix follows this trend but to a smaller degree possibly due to its larger size. The H β and I β sheets have lower mean RMSF values, with a slight elevation in the light state, denoting notable flexibility changes with light activation. This supports the conclusion that light activation induces a state of increased flexibility in specific regions of the protein, potentially facilitating its functional conformational changes.

The undocking and unfolding of both A' α and J α helices represent signature conformational changes when transitioning from the dark state to the light state of AsLOV2. The preservation of secondary structures in WT AsLOV2 during simulations of both states is accordingly investigated. The ratio of residues recognized as secondary structure and the total number of residues is calculated for each structure, based on the simulation data for each state (Figure 3). For example, if a helix contains ten residues and eight are recognized with helical character using DSSP analysis in a given simulation frame, the preservation ratio for this helix in that frame is 0.8. This is calculated and averaged for all frames. A significant decrease of the preservation ratio of the A' α helix suggests its unfolding during the simulation. Ratio analysis at the residue level for A' α helix is presented in Figure S2 and Table S1. In contrast, the J α helix displays a stable helical ratio, with no significant changes between the two states. This suggests that the unfolding timescale of J α helix exceeds the scope of

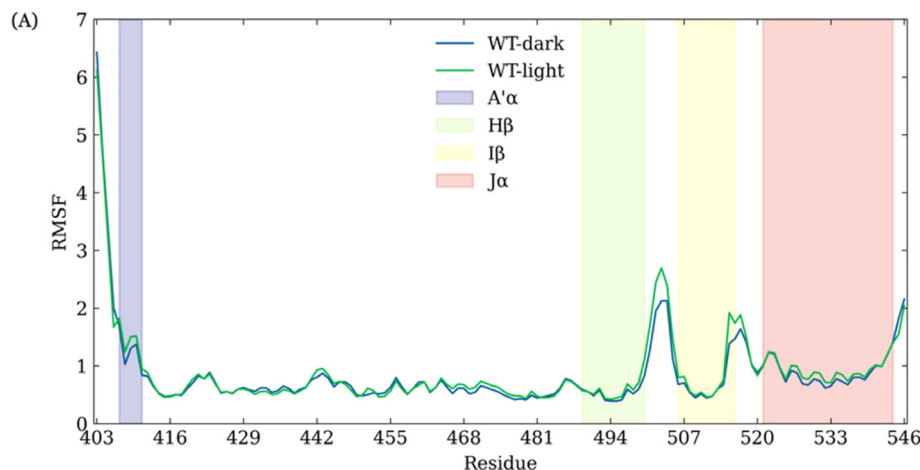


FIGURE 2 Root-mean-square fluctuations (RMSFs) analysis of wild type (WT) of AsLOV2 dark and light states. (A) RMSF of WT AsLOV2 dark state (blue) and light state (green). The plotted RMSF is the averaged value over three independent simulations of each configuration. Four shadowed regions indicate the A'α helix (purple), Hβ sheet (green), Iβ sheet (yellow), and Jα helix (red), respectively. (B) Mean RMSF values of residues in four secondary structures for WT AsLOV2 dark state (blue) and light state (orange), respectively.

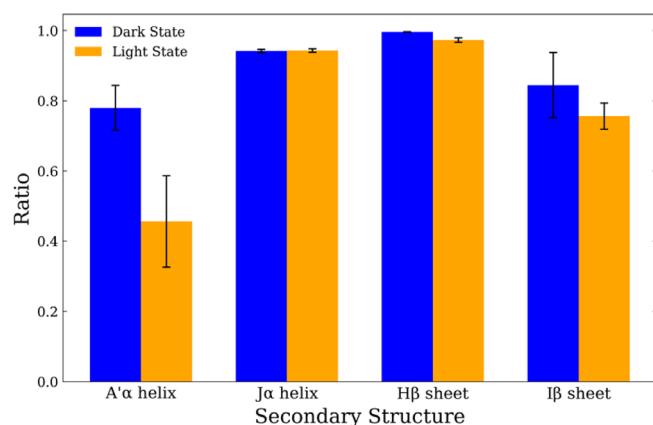
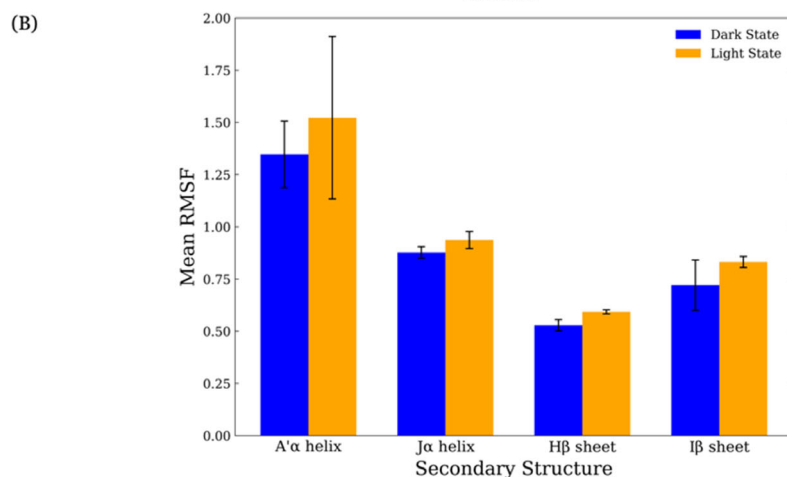


FIGURE 3 Average preservations of secondary structures for WT AsLOV2 dark state (blue) and light state (orange) in four key secondary structures (A'α helix, Jα helix, Hβ sheet, and Iβ sheet).

conventional MD simulations. The Hβ and Iβ sheets exhibit a slight and moderate decrease in their sheet ratios, respectively. Given the typically stable nature of β-sheets, these observed decreases could be indicative of their involvement in the protein's response to light activation, potentially playing a role in signal transduction. Combining with the RMSF analysis, these observations shed more light onto the protein conformation and dynamics change in response to light activation.

The dynamical correlation within the dark and light states of WT AsLOV2 is represented by the dynamic cross-correlation matrices (DCCMs) (Figure 4). For clarity, a discrete color scale with a segmentation threshold of 0.4 is used in the plots presented in Figure 4 to accentuate the correlation contrasts between residues. The discrete color scale distinctly marks highly correlated or anticorrelated regions. DCCM plots using a continuous color scale are available in Figure S3. In Figure 4A, deep red hues along the diagonal indicate strong positive correlation among neighboring residues, and similarly, robust correlations are noted among anti-parallel β-sheets. With light activation, the overall correlation structure is retained. Notable shifts in the Jα helix's correlation patterns, particularly with the antiparallel β-sheet regions and the Fα helix emerge (Figure 4B), suggesting a potential undocking tendency, where the Jα helix exhibits a distinct movement away from the core structure in response to the light activation. This tendency supports the light-induced undocking behavior of the Jα helix, even though the undocking of the Jα helix was not observed in the current study.

3.2 | AsLOV2 light states mutants

Simulations of the selected AsLOV2 mutants in the light states were performed to compare their impacts on the dynamics of the AsLOV2

domain. The comparative analysis focuses on the four key secondary structures of AsLOV2: A' α and J α helices, H β and I β sheets. Average RMSF values for these four secondary structures across all light states

are plotted in Figure 5. The RMSF plot for each mutant is shown in Figure S4. The comparison of average RMSF values over the whole AsLOV2 domain for each light state is shown in Figure S5. The RMSF

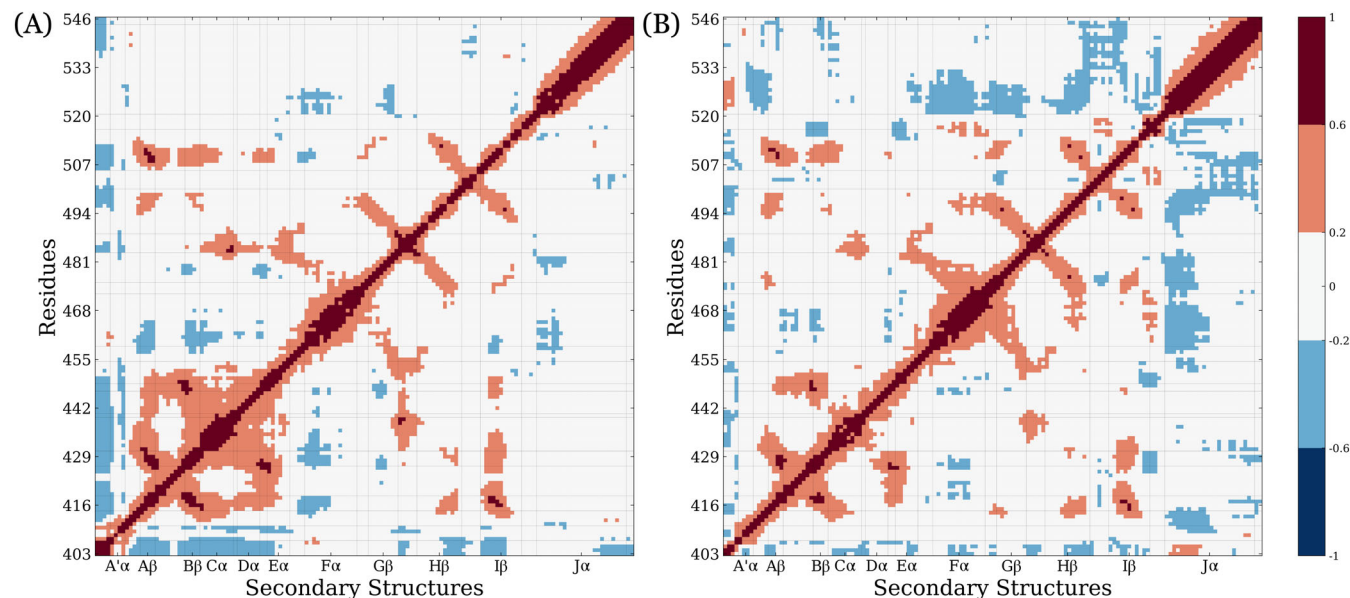


FIGURE 4 Discrete average dynamic cross-correlation matrix for WT AsLOV2 in (A) dark state and (B) light state over three parallel simulations. For clarity, a discretized color scale with a segmentation threshold of 0.4 is used.

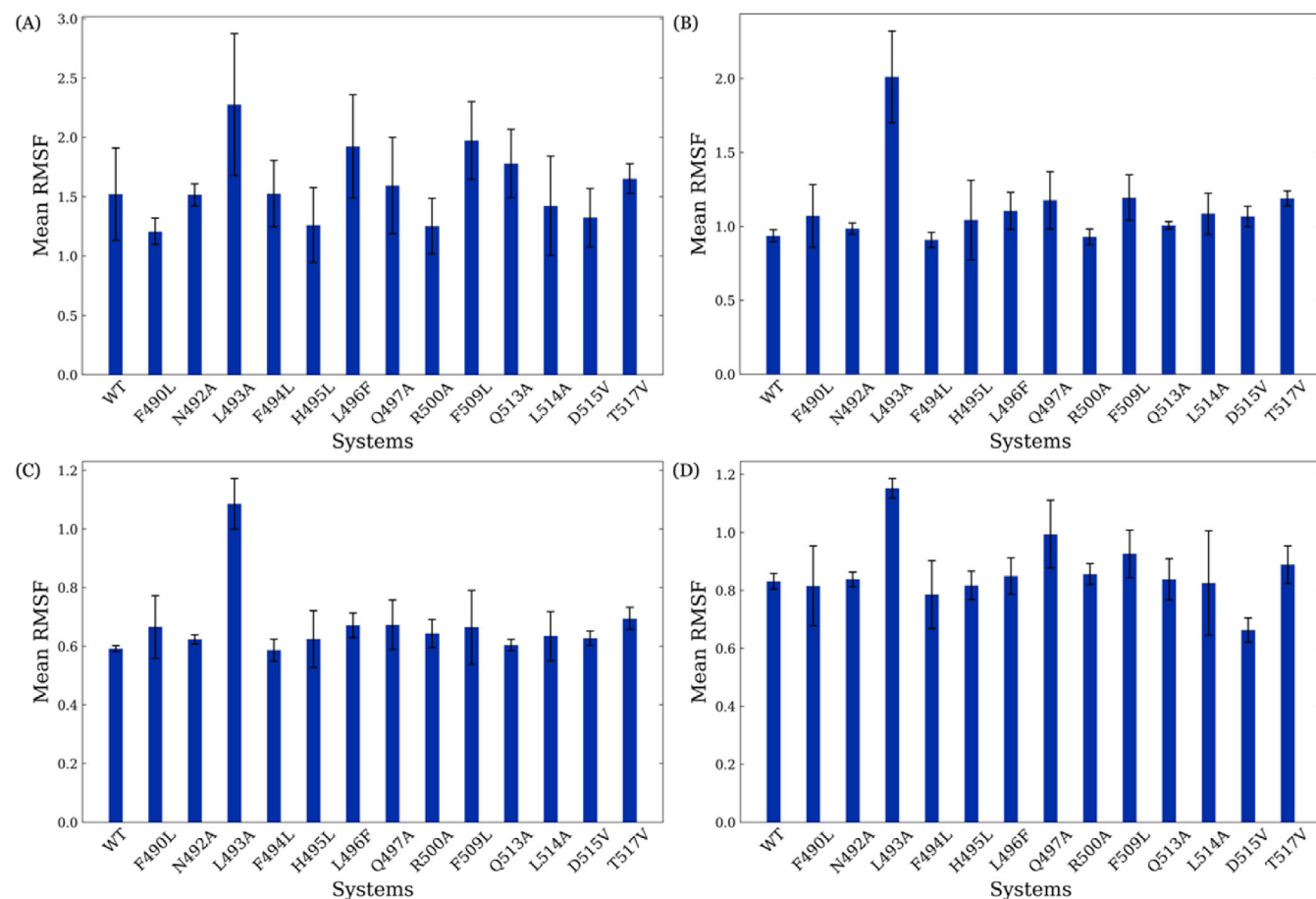


FIGURE 5 Average RMSF values over residues of WT and mutants in light states for (A) A' α helix, (B) J α helix, (C) H β sheet, and (D) I β sheet.

values of A α helix vary significantly among all light states under investigation (Figure 5A). Large error bars suggest significant variability among the simulations. This could also be attributed to the short length of the A α helix and its inherent flexibility given its location at the flexible N-terminal region. In contrast, the H β sheet, I β sheet, and J α helix of most mutants exhibit RMSF values close to the WT value, indicating minimal impact from the corresponding mutants on protein flexibility. L493A emerges as an outlier with elevated RMSF values across the entire protein (Figure S5), suggesting a significant increase in flexibility due to the mutation. Additionally, the D515V mutation on the I β sheet uniquely lowers the mean RMSF of the I β sheet, suggesting localized structural stabilization. These specific changes in dynamics necessitate additional investigation to elucidate the mutants' impact on the protein's behavior and interactions.

Preservation of the four secondary structures in the light states of AsLOV2 WT and mutants is calculated and plotted for comparison (Figure 6). Consistency is observed in the H β sheet and J α helix across all mutants, with the proportion of these structures close to that of the WT. The A α helix, on the other hand, exhibits increased helical content in the light state among most mutants compared with the WT, yet still lower than the WT dark state values, suggesting a light-dependent partial unfolding or destabilization. While the A α helix

ratios show a general increase, the elongated error bars signal considerable variability, complicating definitive conclusions about the mutants' influence. A general but not precisely negative correlation between helix ratio and RMSF underscores a correlated relationship between flexibility and unfolding behavior. Unlike a previous study focusing on mutants located on A α helix itself, it is difficult to ascertain definitive conclusions about H β sheet and I β sheet mutants' influence on the unfolding behavior of A α helix. The I β sheet displays interesting responses to Q497A and D515V. Compared with other mutants, Q497A exhibits a higher I β sheet RMSF (Figure 5D) and a correspondingly lower I β sheet ratio (Figure 6D), suggesting an increase in flexibility. Conversely, D515V shows a lower I β sheet RMSF and a higher I β sheet ratio, implying enhanced stability. Therefore, Q497A and D515V present as prime candidates for further investigation to deepen our understanding of their impact on the I β sheet's dynamics.

Dynamic cross-correlation analysis reveals consistent negative correlation between the J α helix and the adjacent antiparallel β -sheet region in various mutants. A subset of mutants demonstrates DCCM patterns remarkably akin to the WT with mutant R500A as an example (Figure 7A). Intriguingly, in several mutants including L496F (Figure 7B), a novel trend emerges wherein the F α helix exhibits

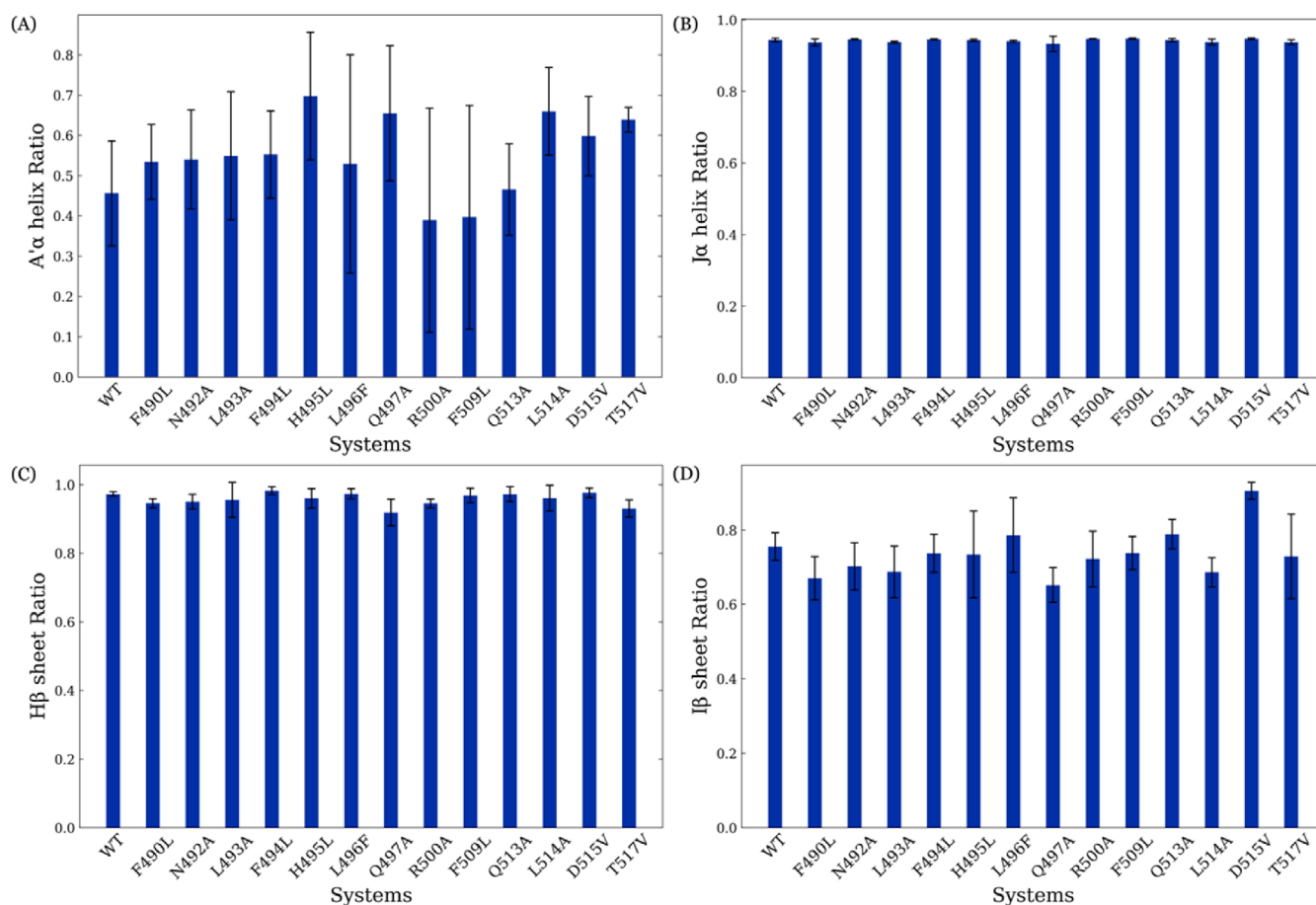


FIGURE 6 Average preservations of secondary structures for WT and mutants in light states for (A) A α helix, (B) J α helix, (C) H β sheet, and (D) I β sheet.

increased negative correlations with spatially distant secondary structures: $A\beta$ and $B\beta$ sheets, as well as the $C\alpha$ and $D\alpha$ helices, despite being separated by the $E\alpha$ helix and FMN. These mutants exhibit an uptick in negative correlations between intersecondary structures, such as the loops connecting $F\alpha$ helix and $G\beta$ sheet and the loop connecting $G\beta$ sheet and $H\beta$ sheet. In F509L mutant, the C-terminal displays stronger negative correlations with the distant cluster comprising the $A\beta$ and $B\beta$ sheets as well as the $C\alpha$ and $D\alpha$ helices (Figure 7C). An extreme case is presented by L493A, in which the cluster of $A\beta$ and $B\beta$ sheets, and the $C\alpha$ and $D\alpha$ helices, shows pronounced positive intragroup correlations. The antiparallel β -sheets also display strong positive correlations with other β -sheets. Meanwhile, the $J\alpha$ helix and the C-terminal reveal more negative correlations with these groups but increased positive correlations with the $A'\alpha$ helix and $F\alpha$ helix located in the same periphery (Figure 7D).

In each light-state system of WT and mutants, the top 10 essential modes are extracted, with mean square-fluctuations of each residue calculated and plotted (Figure 8). The most significant contributions to these modes arise from regions including the N-terminal with $A'\alpha$ helix, loops between $H\beta$ sheet and $I\beta$ sheet, between $I\beta$ sheet and $J\alpha$ helix, and the C-terminal with $J\alpha$ helix. Most mutants exhibit higher flexibility comparing to WT, especially at the loop between $H\beta$ sheet and $I\beta$ sheet, consistent with the RMSF results. L493A, H495L, and F509L show some extra essential movements at $C\alpha$ helix (residues 431–439, spatially close to $A'\alpha$ helix) and the loop between $G\beta$ sheet and $H\beta$ sheet (residues 483–489, spatially close to $C\alpha$ helix). Several mutants, including F490L, L493A, L496F, Q497A, and F509L show higher flexibility at the loop between $I\beta$ sheet and $J\alpha$ helix. Especially, L493A extends the movement flexibility from this loop to the $J\alpha$ helix.

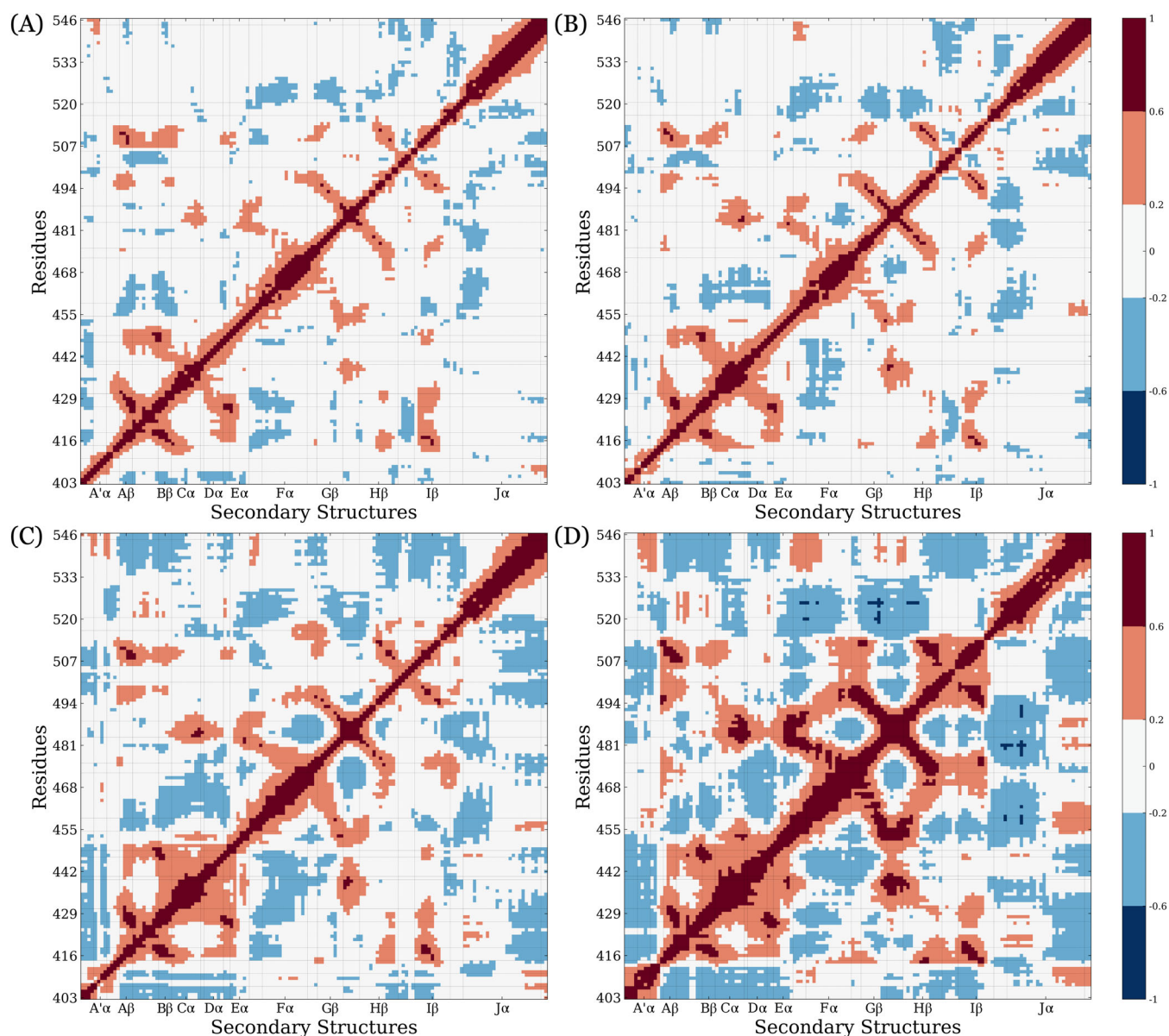


FIGURE 7 Discrete average dynamic cross-correlation matrix in the light state for representative mutants: (A) R500A, (B) L496F, (C) F509L, and (D) L493A.

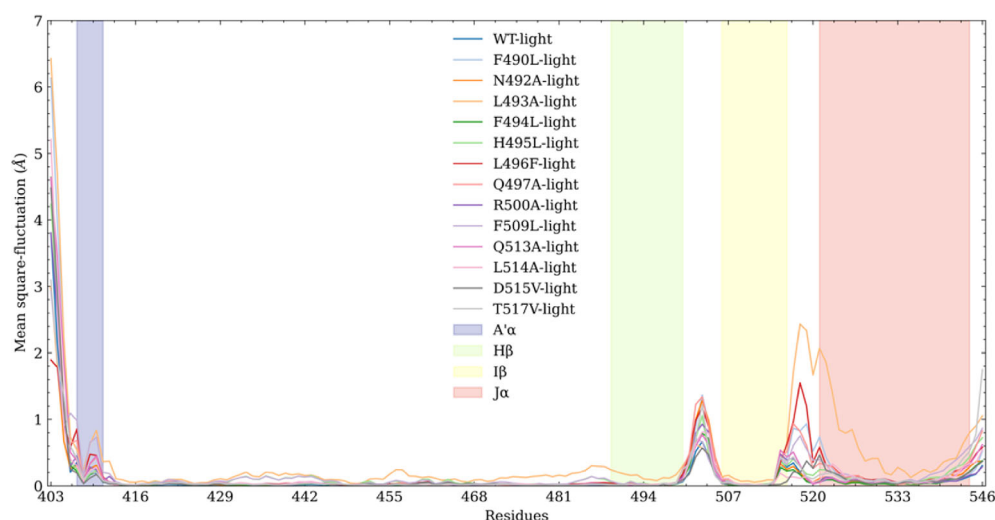


FIGURE 8 Mean square-fluctuations for top 10 essential modes for each system in light state.

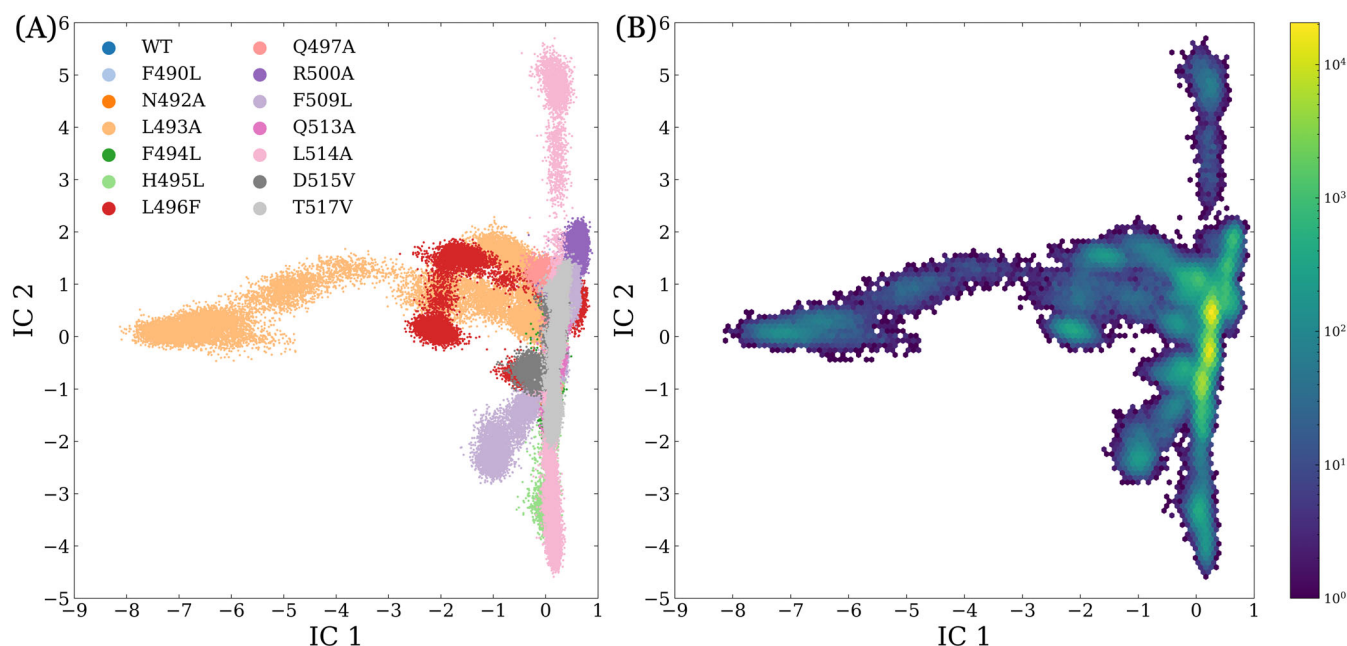


FIGURE 9 Projection of light states simulation onto two-dimensional space based on t-ICA. (A) Distribution of each individual system. (B) Overall population of combined light states simulations.

3.3 | Dimensionality reduction analysis and Markov state model

Simulations of all WT and mutant light states were combined and analyzed using t-ICA, with each state projected onto a two-dimensional (2D) space defined by the two vectors: IC 1 and IC 2. The mutants, especially L493A and L514, display significantly different distributions (Figure 9A). The overall population from the combined simulations is shown in Figure 9B. While most mutant simulations mainly resemble the WT distribution (depicted as a high-density light-yellow region) with only slight deviation, mutants such as L493A, H495L, L496F, F509L, and L514A show pronounced deviations, exploring a much larger conformational space.

Markov state model (MSM) was used to reveal dynamic patterns by kinetically clustering the conformations. Conformations are divided into 50 microstates via *k*-means clustering analysis within the t-ICA 2D space. Top 15 implied timescales, calculated with lag times ranging from 0.1 to 30 ns (Figure S6), showed convergence after about 20 ns, which was chosen as the lag time for MSM construction. The time-scale differentiation resulted in the definition of seven distinct macrostates, with PCCA sequentially used for kinetic clustering.

All seven macrostates and their populations are depicted in Figure 10A and Figure 10B, respectively. Macrostates 2 and 1 exhibit the most dominant occupancies at 78.72% and 11.92%, respectively. The WT light state occupies both macrostates 1 and 2, mirroring the patterns shown in Figure 9B, where these states are the most

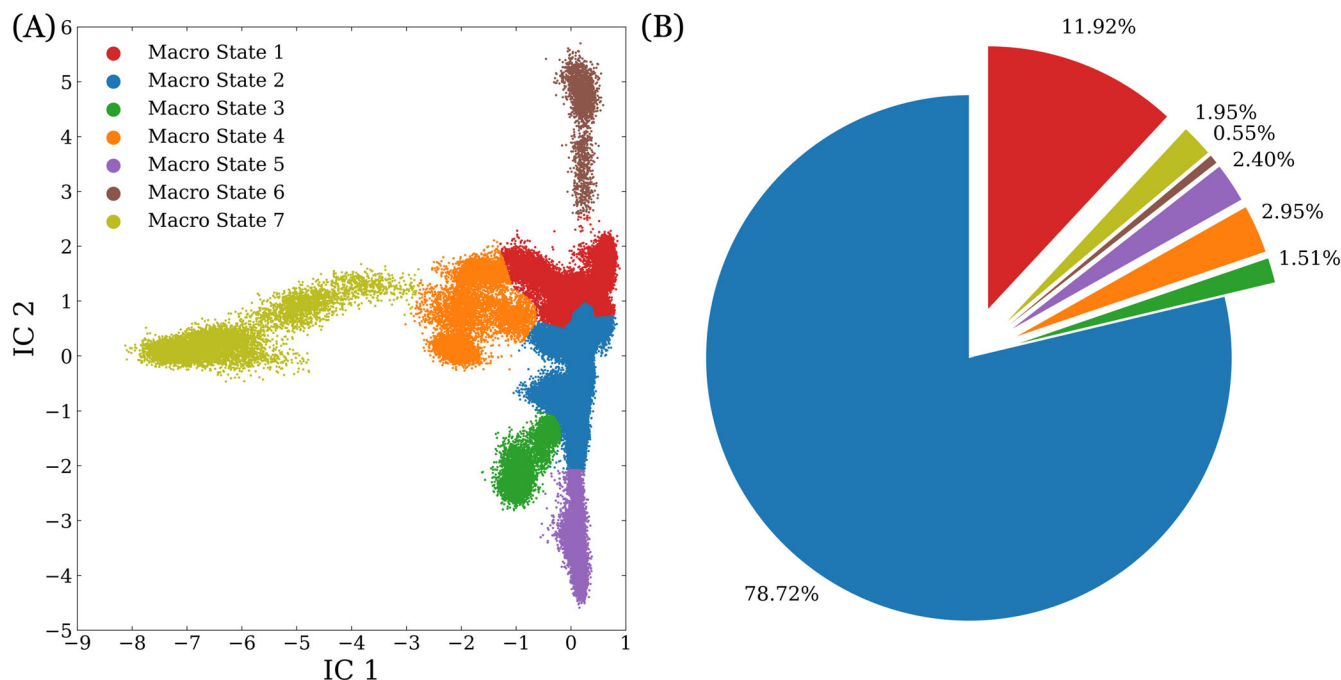
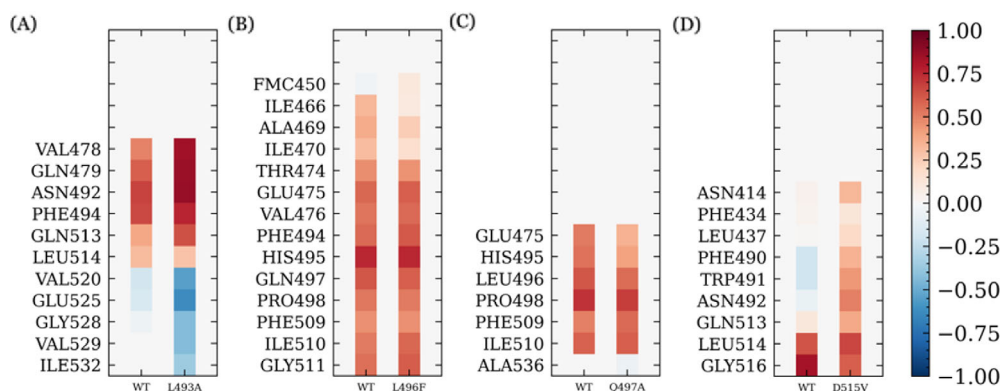


FIGURE 10 Markov state model of AsLOV2 light states. (A) The macrostates represented on the t-ICA 2D space. (B) Relative population composition of the macrostates.

FIGURE 11 The comparison of correlation between the mutated residue and its adjacent residues in the mutant and in WT for (A) L493A, (B) L496F, (C) Q497A, and (D) D515V.



populated. Beyond macrostate 2, mutants such as L493A, H495L, L496F, F509L, and L514A distinctly occupy macrostates 3 through 7 (Figure S7). This variation in occupancy reveals unique dynamic behaviors, highlighting the diverse impact of these mutations on the conformational landscape.

4 | DISCUSSION

Integrating findings from experimental analyses with computational data, we selected four mutants for an in-depth investigation from a pool of thirteen. This selection includes three mutants located on the H β sheet (L493A, L496F, and Q497A) and one on the I β sheet (D515V), based on comprehensive simulation dynamics, conformation exploration, and experimental results. Our study aims to uncover the underlying reasons for the distinct behaviors exhibited

by these specific mutants. First, we extract the correlation values from DCCM for residues surrounding the mutant sites. The criteria for defining surrounding residues are that in the first frame of WT or mutant production runs, any residue with at least one atom within 3 Å to any atom in the mutant site (Figure 11). The Baker-Hubbard method is used to identify hydrogen bonds in the selected region appearing at least 50% of the total frames across three simulations.

To further investigate and compare the dynamical behavior of these four mutants relative to the WT light state, we analyzed the hydrogen bonds (Table S2 and Figure 12). We found that all four mutants lack the hydrogen bond between Thr407 and Ala452 present in the WT. Thr407 was identified to be essential for the unfolding mechanism of A' α helix.¹⁵ The absence of this hydrogen bond is suggested to be the main cause of the high fluctuations observed in the N-terminal.

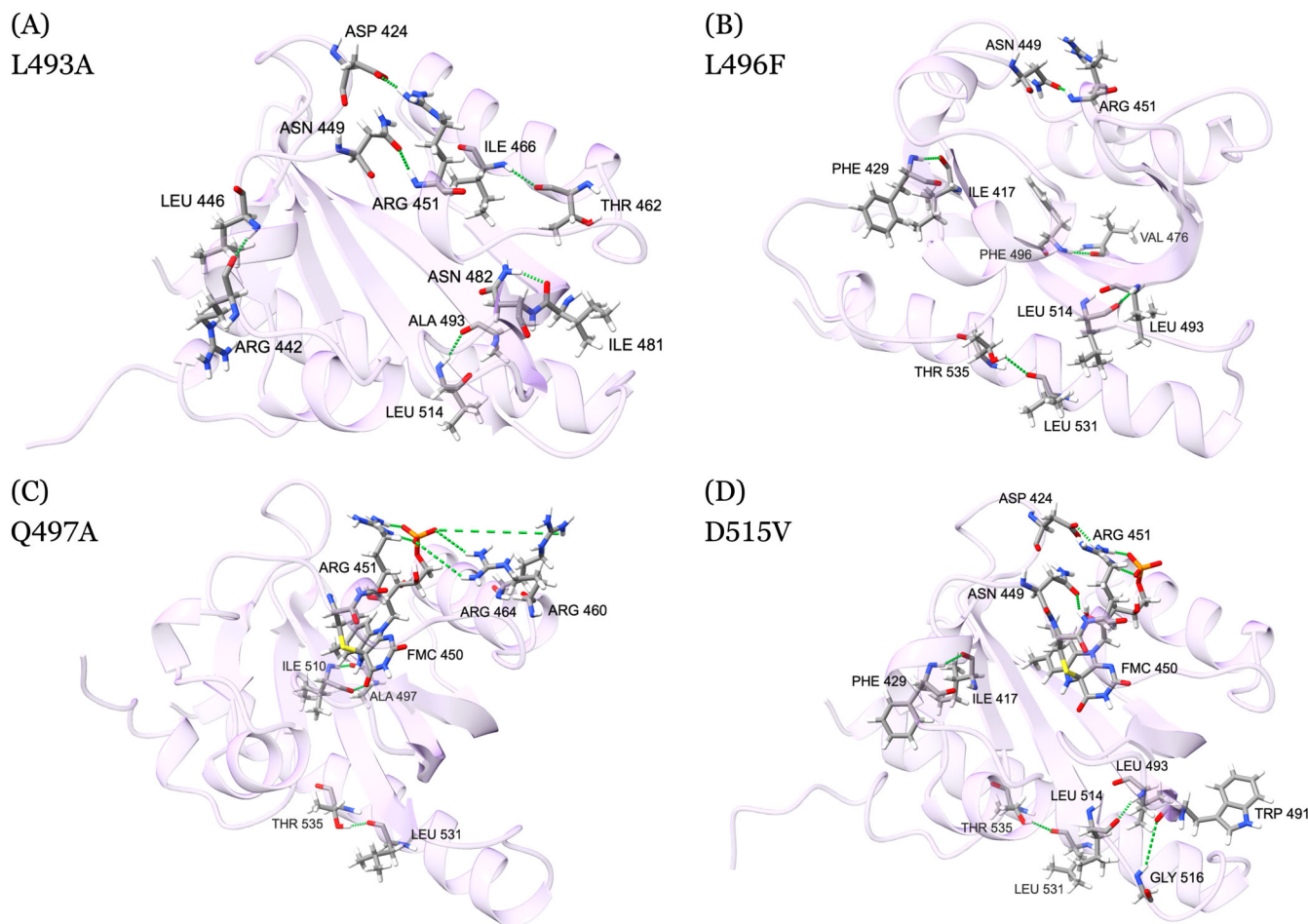


FIGURE 12 The unique hydrogen bonds patterns formed with the most disruptive mutants compared with WT for (A) L493A, (B) L496F, (C) Q497A, and (D) D515V. Each conformation is from the first frame of the first parallel simulation, respectively.

Compared with WT, we observe several missing hydrogen bonds between flexible secondary structures in the L493A mutant. Among these, two missing hydrogen bonds between $I\beta$ sheet and $A'\alpha$ helix, one between Ser431-O and GLys413-O and one between Ser433-OG and Glu412-O, could lead to the increased flexibility of the $A'\alpha$ helix. Another two missing hydrogen bonds between $I\beta$ sheet and loop connecting $I\beta$ sheet and $J\alpha$ helix, one between Gln479-NE2 and Val520-O and one between Gln479-OE1 and Val520-N, may lead to the increased flexibility of the $J\alpha$ helix. There is another missing hydrogen bond between $G\beta$ sheet and $E\alpha$ helix between Ser486-OG and Glu439-OE2. On the other hand, the L493A mutant has more hydrogen bonds within secondary structures or positive correlated regions than the WT. As shown in the local correlation patterns, residue Ala493 shows a more negative correlation with $J\alpha$ helix but a more positive correlation with $G\beta$ sheet compared with Leu493 (Figure 11A). As a result, L493A exhibits a higher flexibility with high fluctuations and the amplified correlation patterns (Figure 12A).

For the L496F mutant, Gln479-NE2 with Val520-O and Gln479-OE1 with Val520-N are missing compared with the WT, leading to a more flexible $J\alpha$ helix. The hydrogen bond between Asn482-ND2 on the $G\beta$ sheet and Leu453-O on the $E\alpha$ helix is also absent. This leads to a more flexible $E\alpha$ helix and $F\alpha$ helix, and it is observed

that the correlation between Phe496 to $F\alpha$ helix is less positive in the mutant (Figure 11B). These absent hydrogen bonds might result from the steric hindrance of the larger Phenylalanine side chain (Figure 12B), which could contribute to the different distribution of L493A and L496F mutants from the WT and other mutants in the t-ICA 2D space (Figure 9A).

For the Q497A mutant, the two hydrogen bonds between Gln497 and Ile510 remain. In addition, the Cys450-FMN adduct has more hydrogen bonds associated with Arg451, Arg460, and Arg464 on $F\alpha$ helix (Figure 12C). These hydrogen bonds may help the Q497A mutant maintain the local correlation, protein flexibility, and dynamics similar to the WT (Figure 11C). Similar changes in the hydrogen bonds patterns of the Cys450-FMN core were observed in D515V mutant. This can be attributed that Asn482 form an extra hydrogen bond in both mutants with Leu453, compared with other mutants (Figure 12D). This hydrogen bond is disrupting the hydrogen bonds of Asn482 with Cys450-FMN core in both Q497A and D515V, explaining the drastic changes in the hydrogen bond patterns observed in the FMN core.

For the D515V mutant, we observe an increased number of hydrogen bonds, particularly two hydrogen bonds near residue Val515: one between Leu493-N and Leu514-O and one between

Gly516-N and Trp491-O. As comparison, the WT exhibits one hydrogen bond between Leu514-N and Ala493-O. In this mutant, residues Leu493 and Leu514, located on two adjacent antiparallel β -sheets, are likely to have stronger correlation with each other due to these hydrogen bonds. Similarly, the hydrogen bond between Gly516 and Trp491 also helps to maintain or strengthen the positive correlation comparing to the WT especially given the fact that residue Gly516 is at the end of the $l\beta$ sheet and behaves more similarly to a loop. The correlations between the mutated residue Val515 and residues 490–492 are more positive than in the WT (Figure 11D). As a result, the $l\beta$ sheet, especially the connecting region to the $J\alpha$ helix, becomes more constrained, leading to reduced flexibility and increased higher $l\beta$ sheet ratio. It is possible when Aspartic Acid changes to Valine at position 515, the negative side chain becomes neutral and the electrostatic attractions from surroundings (opposite to residues 490–492) mainly becomes steric effect and contribute to the attractions of residues 490–492.

5 | CONCLUSION

Extensive molecular dynamics simulations were conducted for the wild type and 13 different mutations introduced into the H β and $l\beta$ sheets of AsLOV2 to investigate the role of β -sheets in the protein's allosteric mechanism. RMSF, DCCM, and MSM analyses indicated that four mutations accounted for most of the changes. The L493A mutant exhibits the highest fluctuations across various regions of the protein. The L493A and L496F mutants are among those exploring the most conformational space, likely due to the missing hydrogen bond between Val520 and Gln479. The Q497A and D515V mutants alter the hydrogen bond patterns with the FMN molecule significantly more than other mutations, likely due to the missing hydrogen bond between Asn482 and Leu453.

AUTHOR CONTRIBUTIONS

Sian Xiao: Conceptualization; methodology; software; formal analysis; investigation; data curation; validation; visualization; writing—original draft; writing—review and editing. **Mayar Tarek Ibrahim:** Conceptualization; methodology; formal analysis; investigation; validation; writing—original draft; writing—review and editing. **Gennady M. Verkhivker:** Writing—review and editing. **Brian D. Zoltowski:** Writing—review and editing. **Peng Tao:** Supervision; project administration; funding acquisition; resources; writing—review and editing.

ACKNOWLEDGMENTS

Computational time was provided by the Southern Methodist University's Center for Research Computing.

FUNDING INFORMATION

Research reported in this article was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award no. R15GM122013 (PT) and 2R15GM109282 (BDZ).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The processed results presented or mentioned in this study can be found in Zenodo repository at <https://doi.org/10.5281/zenodo.10247350>.

ORCID

Sian Xiao  <https://orcid.org/0000-0002-3451-5227>

Gennady M. Verkhivker  <https://orcid.org/0000-0002-4507-4471>

Peng Tao  <https://orcid.org/0000-0002-2488-0239>

REFERENCES

- [1] B. E. McIntosh, J. B. Hogenesch, C. A. Bradfield, *Annu. Rev. Physiol.* **2010**, 72, 625.
- [2] A. Möglich, R. A. Ayers, K. Moffat, *Structure* **2009**, 17, 1282.
- [3] O. I. Lungu, R. A. Hallett, E. J. Choi, M. J. Aiken, K. M. Hahn, B. Kuhlman, *Chem. Biol.* **2012**, 19, 507.
- [4] S. Crosson, K. Moffat, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, 98, 2995.
- [5] B. D. Zoltowski, B. Vaccaro, B. R. Crane, *Nat. Chem. Biol.* **2009**, 5, 827.
- [6] A. Pudasaini, K. K. El-Arab, B. D. Zoltowski, *Front. Mol. Biosci.* **2015**, 2, 18.
- [7] D. Strickland, K. Moffat, T. R. Sosnick, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, 105, 10709.
- [8] D. Strickland, X. Yao, G. Gawlak, M. K. Rosen, K. H. Gardner, T. R. Sosnick, *Nat. Methods* **2010**, 7, 623.
- [9] J. P. Zayner, C. Antoniou, T. R. Sosnick, *J. Mol. Biol.* **2012**, 419, 61.
- [10] A. I. Nash, W.-H. Ko, S. M. Harper, K. H. Gardner, *Biochemistry* **2008**, 47, 13842.
- [11] A. Yamamoto, T. Iwata, Y. Sato, D. Matsuoka, S. Tokutomi, H. Kandori, *Biophys. J.* **2009**, 96, 2771.
- [12] J. P. Zayner, C. Antoniou, A. R. French, R. J. Hause, T. R. Sosnick, *Biophys. J.* **2013**, 105, 1027.
- [13] S. A. Hollingsworth, R. O. Dror, *Neuron* **2018**, 99, 1129.
- [14] M. Karplus, J. A. McCammon, *Nat. Struct. Biol.* **2002**, 9, 646.
- [15] M. T. Ibrahim, F. Trozzi, P. Tao, *Comput. Struct. Biotechnol. J.* **2022**, 20, 50.
- [16] S. Xiao, G. M. Verkhivker, P. Tao, *Trends Biochem. Sci.* **2023**, 48, 375.
- [17] G. Verkhivker, M. Alshahrani, G. Gupta, S. Xiao, P. Tao, *Int. J. Mol. Sci.* **2023**, 24, 7747.
- [18] A. S. Halavaty, K. Moffat, *Biochemistry* **2007**, 46, 14001.
- [19] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kucsera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, *J. Comput. Chem.* **2009**, 30, 1545.
- [20] P. L. Freddolino, K. H. Gardner, K. Schulten, *Photochem. Photobiol. Sci.* **2013**, 12, 1158.
- [21] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, 79, 926.
- [22] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, 23, 327.
- [23] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, 98, 10089.
- [24] P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern,

- R. P. Wiewiora, B. R. Brooks, V. S. Pande, *PLoS Comput. Biol.* **2017**, 13, e1005659.
- [25] E. N. Baker, R. E. Hubbard, *Prog. Biophys. Mol. Biol.* **1984**, 44, 97.
- [26] R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L.-P. Wang, T. J. Lane, V. S. Pande, *Biophys. J.* **2015**, 109, 1528.
- [27] W. Kabsch, C. Sander, *Biopolymers* **1983**, 22, 2577.
- [28] A. Amadei, A. B. M. Linssen, H. J. C. Berendsen, *Proteins* **1993**, 17, 412.
- [29] A. Bakan, L. M. Meireles, I. Bahar, *Bioinformatics* **2011**, 27, 1575.
- [30] S. Zhang, J. M. Krieger, Y. Zhang, C. Kaya, B. Kaynak, K. Mikulska-Ruminska, P. Doruker, H. Li, I. Bahar, *Bioinformatics* **2021**, 37, 3657.
- [31] S. Xiao, Z. Song, H. Tian, P. Tao, *J. Comput. Biophys. Chem.* **2023**, 22, 489.
- [32] G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, F. Noé, *J. Chem. Phys.* **2013**, 139, 015102.
- [33] F. Noé, C. Clementi, *J. Chem. Theory Comput.* **2015**, 11, 5002.
- [34] L. Molgedey, H. G. Schuster, *Phys. Rev. Lett.* **1994**, 72, 3634.
- [35] S. Röblitz, M. Weber, *Adv. Data Anal. Classif.* **2013**, 7, 147.
- [36] B. Trendelkamp-Schroer, H. Wu, F. Paul, F. Noé, *J. Chem. Phys.* **2015**, 143, 174101.
- [37] S. Xiao, M. Alshahrani, G. Gupta, P. Tao, G. Verkhivker, *J. Chem. Inf. Model.* **2023**, 63, 5272.
- [38] M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz, F. Noé, *J. Chem. Theory Comput.* **2015**, 11, 5525.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: S. Xiao, M. T. Ibrahim, G. M. Verkhivker, B. D. Zoltowski, P. Tao, *J. Comput. Chem.* **2024**, 1, <https://doi.org/10.1002/jcc.27344>