

Effect of Pressure on the Conformational Landscape of Human γ D-Crystallin from Replica Exchange Molecular Dynamics Simulations

Arlind Kacirani, Betül Uralcan, Tiago S. Domingues, and Amir Haji-Akbari*



Cite This: <https://doi.org/10.1021/acs.jpcb.4c00178>



Read Online

ACCESS |



Metrics & More

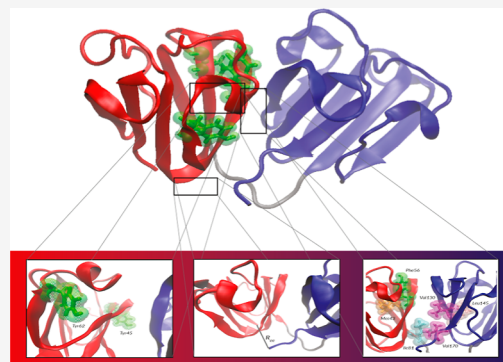


Article Recommendations



Supporting Information

ABSTRACT: Human γ D-crystallin belongs to a crucial family of proteins known as crystallins located in the fiber cells of the human lens. Since crystallins do not undergo any turnover after birth, they need to possess remarkable thermodynamic stability. However, their sporadic misfolding and aggregation, triggered by environmental perturbations or genetic mutations, constitute the molecular basis of cataracts, which is the primary cause of blindness in the globe according to the World Health Organization. Here, we investigate the impact of high pressure on the conformational landscape of wild-type H γ D-crystallin using replica exchange molecular dynamics simulations augmented with principal component analysis. We find pressure to have a modest impact on global measures of protein stability, such as root-mean-square displacement and radius of gyration. Upon projecting our trajectories along the first two principal components from principal component analysis, however, we observe the emergence of distinct free energy basins at high pressures. By screening local order parameters previously shown or hypothesized as markers of H γ D-crystallin stability, we establish correlations between a tyrosine–tyrosine aromatic contact within the N-terminal domain and the protein’s end-to-end distance with projections along the first and second principal components, respectively. Furthermore, we observe the simultaneous contraction of the hydrophobic core and its intrusion by water molecules. This exploration sheds light on the intricate responses of H γ D-crystallin to elevated pressures, offering insights into potential mechanisms underlying its stability and susceptibility to environmental perturbations, crucial for understanding cataract formation.



1. INTRODUCTION

Cataract is the primary cause of blindness worldwide and is the leading factor contributing to vision impairment in the United States.¹ This condition is characterized by the accumulation of heterogeneous and opaque aggregates formed by a family of mammalian lens proteins referred to as crystallins,^{2,3} which are categorized into three groups based on molecular weight and function. α -Crystallins, which possess higher molecular weights, are heat shock proteins that function as chaperones guiding the folding of β - and γ -crystallins that are responsible for maintaining lens transparency and a high refractive index. Due to lack of turnover postbirth, crystallins require high thermodynamic stability to remain folded, and form stable aqueous solutions at concentrations surpassing 400 g/L.⁴ Therefore, the aggregation of crystallins is generally induced through their partial misfolding due to aging or genetic mutations. However, crystallins can also reversibly phase separate into protein-rich domains at low temperature, e.g., in calf lens, leading to a condition known as cold cataracts.⁵ Among crystallins, human γ D-crystallin (H γ D-crys), a monomeric lens protein comprising 173 residues, represents a significant component of the human lens whose aggregation contributes to age-related cataracts.^{6,7} Its secondary structure mainly consists of β -sheets organized into two homologous

domains (Figure 1A). Each domain contains two Greek-key motifs forming a β -sandwich of eight intercalated β -strands.⁸

While crystallin aggregation represents a crucial aspect of cataract formation, significant gaps remain in our understanding of the precise molecular mechanisms underlying this phenomenon. When it comes to H γ D-crys, an aggregation mechanism that involves domain swapping at the hydrophobic interdomain interface followed by partial unfolding of the N-terminal domain (N-td) has been proposed,^{6,9,10} a proposition supported by single molecule atomic force microscopy¹¹ and nuclear magnetic resonance¹² experiments, as well as atomistic classical molecular dynamics (MD) simulations.⁷ The structural changes resulting from physicochemical environmental perturbations—such as acidic pH, ultraviolet light, oxidative damage, heating or freezing,^{2,7,10,13} as well as specific amino-acid substitutions,^{8,12,14,15} reportedly contribute to favoring the formation of such aggregation-prone conformations. There-

Received: January 10, 2024

Revised: April 1, 2024

Accepted: April 8, 2024

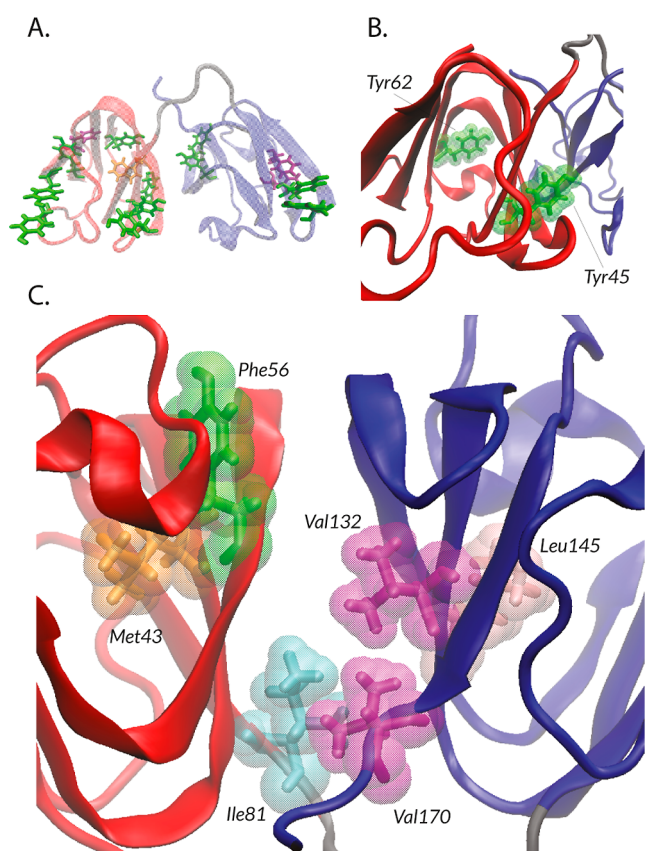


Figure 1. (A) A schematic illustration of HyD-crys with N- and C-terminals depicted in red and blue, respectively, also highlighting the network of relevant aromatic residues. (B) A closer view of the N-terminal domain depicting the van der Waals volumes of tyrosines at positions 45 and 62. (C) Zoomed-in rendering of HyD-crys interdomain interface highlighting the six hydrophobic residues: Met43, Phe56 and Ile81 in N-td, and Val132, Leu145, and Val170 in C-td.

fore, the identification and comprehension of potential initial denaturation sites as early indicators of aggregation are pivotal for enhancing our molecular-level understanding of the mechanism of cataract formation.

Current evidence points to the importance of the aromatic side chain contacts within the N-td (Figure 1A) and the hydrophobic nature of the interdomain interface (Figure 1C) to be key markers in the initial denaturation of HyD-crys.^{6,16–18} For instance, alanine substitutions of aromatic residues in HyD-crys N-td, when expressed in *E. coli*, exhibited reduced thermal stability and faster unfolding compared to the wild-type protein.¹⁸ This underscores the pivotal role of N-td aromatic contacts in maintaining the protein's structural integrity. In fact, the importance of aromatic pair interactions in folding and stability of a diverse range of proteins is supported by a strong body of experimental evidence,^{19–21} especially in the case of proteins rich in β -sheet content.^{22–24} Figure 1A illustrates the structure of HyD-crys, showcasing its N- and C-terminal domains in red and blue, respectively. This visualization highlights the residue pairs engaged in aromatic–aromatic interactions, a critical aspect known to uphold the protein's folding stability, as previously reported. More specifically, tyrosines emerge as significant contributors to protein structural stability compared to other residues,²⁵ due to their ability to form aromatic π – π contacts, strong hydrogen

bonds through the phenol group, and other interactions through their multiple side-chain binding modes, as confirmed both experimentally and computationally.^{26,27}

Additionally, HyD-crys, akin to other proteins prone to aggregation, harbors a hydrophobic core within its structure, primarily composed of six amino acid residues.^{6,7,16,17} Among these residues, Met43, Phe56, and Ile81 belong to the N-terminal domain, while Val132, Leu145, and Val170 belong to the C-terminal domain. Mutations of these residues, particularly those within the N-td, have been experimentally demonstrated to significantly destabilize the protein structure and enhance its unfolding kinetics by facilitating access to new metastable states.^{10,17} To preserve the hydrophobicity of the interdomain interface, the wild-type HyD-crys employs two flanking pairs, namely Gln54–Gln143 and Arg79–Met147, which act as molecular tethers, encircling the hydrophobic core formed by the N- and C-terminal domains.¹⁶ Moreover, its N- and C-terminals shield its hydrophobic core, with the C-terminal end forming a loop over the protein's interdomain interface, resembling a structural anchor between the two protein lobes.

It is necessary to note that conformational perturbations that lead to cataract are generally subtle, as several cataract-associated single and double amino-acid substitutions within the wild type HyD-crys have been previously reported to retain a native-like conformation showing no evidence of large-scale misfolding.^{8,12} It, therefore, appears that crystallin aggregation is triggered by subtle conformational changes that may be difficult to capture experimentally. In contrast, environmental perturbations such as changes in temperature, or chemical denaturants, including urea and guanidine hydrochloride (GdnHCl), induce complete unfolding of the protein, henceforth significantly distorting its equilibrium conformational landscape.^{16,17}

To bridge the gap between extensive conformational changes triggered by “harsh” stress factors and subtle conformational instabilities in aggregation-prone mutants, we turn to high hydrostatic pressure as a milder physical perturbation of the protein structure. Crucially, nature favors life at extremely high pressure, as evidenced by genetic adaptations enabling deep-sea proteins to function at extreme pressures.²⁸ Notably, even minor amino acid substitutions at critical functional sites can adapt the pressure sensitivity of a protein by favoring chemical conformations that are less prone to changes in volume.²⁹ However, such adaptations are pressure-specific, limiting the depth at which deep-sea organisms can thrive while preserving in vitro sensitivity to pressure.³⁰ Pressure affects protein folding through stabilizing folding intermediates with smaller specific volumes.^{31–33} Additionally, pressure serves as a valuable tool for inferring mechanistic information about protein folding and refolding by perturbing a protein's thermodynamic equilibrium.^{34–36} At the molecular level, the response of proteins to high pressure predominantly involves changes to intramolecular hydrophobic interactions. However, it has been experimentally observed that increasing pressure can elicit both increases and decreases in protein stability.³⁷ These intricacies have resulted in a resurgent interest in understanding the impact of high pressure on protein folding in general and have further motivated our investigation of how pressure impacts the stability indicators associated with the N-terminal domain and the interdomain interface of HyD-crystallin. To the best of our knowledge, this

is the first computational investigation of how pressure impacts the conformational landscape of a large globular protein.

However, studying these subtle conformational changes within HyD-crystallin poses a challenge for conventional experimental techniques due to their limited spatiotemporal resolution in capturing small variations in the protein structure. Meanwhile, traditional computational methods such as MD might become inefficient in sampling conformational landscapes characterized by basins separated by substantial free energy barriers.^{38,39} To address these limitations, a wide variety of advanced sampling techniques have been developed to facilitate traversing energy barriers in MD simulations.^{40–42} Notably, replica exchange MD (REMD) have emerged as a powerful tool for exploring rough potential energy landscapes with moderate bottlenecks, making it ideal for studying conformational landscapes of proteins.⁴³ Here, we utilize REMD simulations to investigate the impact of pressure on the conformational landscape of wild-type HyD-crystallin, which we probe using principal component analysis (PCA). Our investigation quantifies the impact of pressure on both conventional global metrics of protein stability and localized structural features confirmed or hypothesized to play crucial roles in stabilizing the aromatic contacts within the N-terminal domain and modulating the hydration status of the hydrophobic interdomain interface. Intriguingly, our findings reveal correlations between the two primary PCA components and the equilibrium distributions of the local order parameters associated with N-td and the interdomain interface. These observed correlations shed light on the relationship between conformational dynamics and key stability markers affected by pressure in HyD-crystallin.

2. METHODS

2.1. MD Simulations. All MD simulations are performed using the open-source GROMACS⁴⁴ software, with Newton's equations of motion integrated using the leapfrog algorithm with a time step of 2 fs. A modified Amber force-field is used to model the protein (Amber ff99SB-ILDN),⁴⁵ while water molecules are represented using the TIP4P model.⁴⁶ The compatibility of this force field with the TIP4P water model has been validated by other simulation studies on proteins.^{47–49} (We also confirm the stability of HyD-crys upon utilizing this force-field, as depicted in Figure S1.) All bond and angle constraints are maintained using the linear constraint solver (LINCS) algorithm.⁵⁰ All production runs are conducted within the isothermal isobaric (NPT) ensemble with temperature and pressure controlled using the Nosé-Hoover thermostat⁵¹ and the Parinello-Rahman barostat,⁵² with time constants of 0.4 and 2 ps, respectively. All long-range electrostatic interactions are treated using the smooth particle mesh Ewald⁵³ method with a grid spacing of 0.1 nm, while all short-range interactions are truncated at 1 nm.

2.2. System Setup and Equilibration. The initial configuration of the wild-type human γ D-crystallin is obtained from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB code 1HK0).⁸ We use the pdb2gmx tool of GROMACS to assign protonation states of all residues assuming canonical pK_a values and pH of 7. Subsequently, the electro-neutral protein is solvated with 15,133 water molecules, where the protein is kept 2 nm away from its closest periodic image. This corresponds to simulating an isolated protein, i.e., in the infinite dilution limit. The resulting system undergoes energy minimization employing

5000 steps of the steepest descent method in order to remove any unphysical atomic overlaps. Following energy minimization, the system undergoes a two-step equilibration process. First, a 100 ps position-restrained equilibration within the canonical (NVT) ensemble is conducted employing the Brendsen thermostat⁵⁴ in order to properly equilibrate the solvent molecules surrounding the protein. Subsequently, a 100 ps NPT simulation is performed using the Parinello-Rahman barostat to stabilize the density and pressure. After equilibration, a pressure ramp is applied, starting from the NPT-equilibrated structures. Subsequently, 100 ns production runs are initiated at each target pressure, namely 1 bar, and 5, 7, 10, and 15 kbar, respectively.

2.3. Dimensionality Reduction. We employ principal component analysis⁵⁵ (PCA) to analyze all our MD trajectories. Functioning as a linear dimensionality reduction technique, PCA identifies key collective variables that are pivotal in explaining the variability in data. Starting from a trajectory, $x(t)$, which could be a collection of time frames, or samples taken from a probability distribution, one first computes the covariance matrix C , whose components are given by

$$C_{ij} = \langle [x_i(t) - \overline{x_i(t)}][x_j(t) - \overline{x_j(t)}] \rangle_t$$

Note that C is a $3N \times 3N$ symmetric matrix, which can be diagonalized as

$$R^T C R = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{3N})$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{3N} \geq 0$. Here, λ_i 's are the eigenvalues of C and R is an orthogonal matrix containing the eigenvectors of C as its columns. In other words, R specifies a unitary transformation that can be used to map $x(t)$ along its normal modes, yielding the projection vector γ given by

$$\gamma(t) = R^T[x(t) - \overline{x(t)}]$$

The eigenvectors corresponding to C 's leading eigenvalues are capable of capturing substantial deviations from the mean, $\overline{x(t)}$. Therefore, the corresponding projections can serve as suitable collective variables for embedding a high-dimensional trajectory. PCA has become a standard technique for constructing free energy landscapes of proteins.⁵⁶ In our analysis, $x(t)$'s fed into PCA consist of the instantaneous Cartesian coordinates of α -carbons (Ca) of all residues. After conducting PCA, the fraction of variance explained (FVE), evaluating the proportion of variance retained by each eigenvalue is computed as

$$\text{FVE}_i = \frac{\lambda_i}{\sum_j \lambda_j} \quad (1)$$

The cumulative FVE for a set of eigenvalues can be determined by adding up their respective individual FVEs. To investigate the conformational landscape at different pressures, we project $x(t) - \overline{x}$ onto the two primary eigenvectors of C . These individual projections are then used for constructing free energy profiles, as discussed in the text.

2.4. REMD Simulations. REMD constitutes an advanced sampling technique rooted in the idea of parallel tempering and was initially proposed by Sugita and Okamoto in 1999.⁴¹ This method enables the effective exploration of rugged potential energy landscapes by concurrently conducting multiple simulations at varying temperatures and intermittently exchanging configurations between adjacent simulation

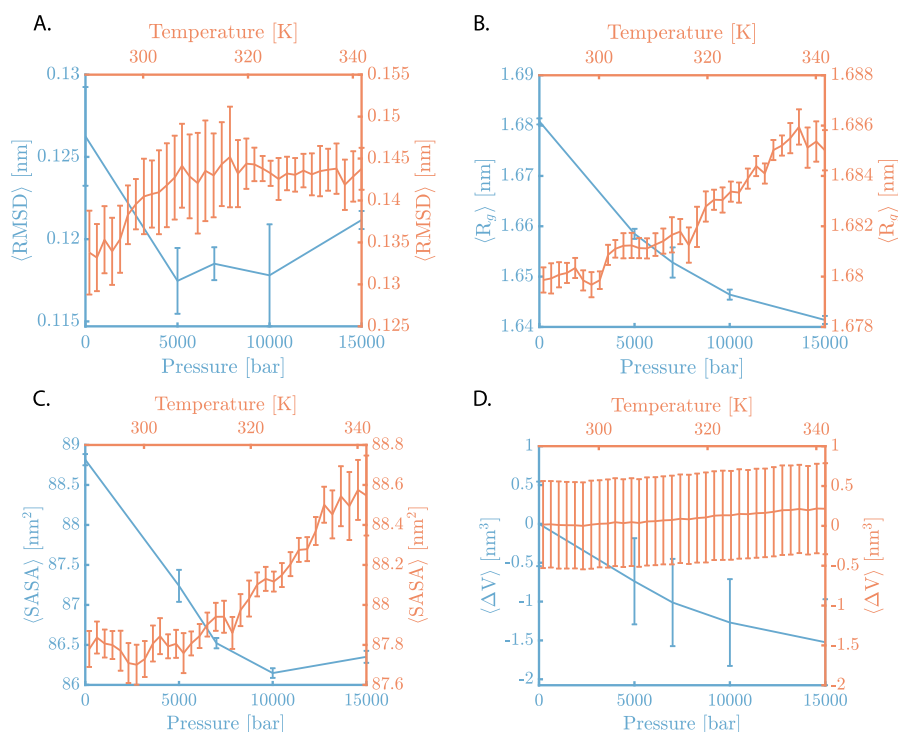


Figure 2. Dependence of global measures of protein stability to temperature and pressure: (A) root mean-square deviation, (B) radius of gyration, (C) solvent accessible surface area, and (D) change in protein volume.

“windows” based on their respective Boltzmann weights. In this work, we conduct two rounds of REMD simulations, each employing 35 replicas within the temperature range, $290 \text{ K} \leq T \leq 342 \text{ K}$. The temperature differences between adjacent replicas are carefully selected to maintain an exchange acceptance rate of approximately 20–25%. The initial round of simulations, spanning 90 ns, commences with configurations periodically collected along the 100 ns conventional MD trajectories outlined in Section 2.2. However, due to the rugged nature of the potential energy landscape at elevated pressures, effectively exploring the free energy landscape, particularly the emerging metastable basins, poses a challenge even with extended trajectories. To address this challenge, we randomly selected 35 configurations from the two PCA basins identified at 5 kbar with equal probabilities. These configurations serve as initial points for a subsequent 60 ns round of simulations. For ambient pressure, configurations are randomly chosen from those periodically collected at ambient temperature during the initial round of REMD. All analyses and results presented in this paper rely on trajectories obtained during this second round of REMD. We ensure the adequacy of sampling and, therefore, the robustness and reliability of our findings by verifying that extending the REMD trajectories does not alter the qualitative shapes or positions of PCA basins.

2.5. Structural Analysis. In order to assess how HyD-crys stability is impacted by environmental perturbations, a collection of global and local order parameters (OPs) are computed. The global OPs computed here are all standard measures of protein stability. The first is the root-mean-square deviation (RMSD), expressed as

$$\text{RMSD}(t) = \sqrt{\frac{1}{n_a} \sum_{i=1}^{n_a} |\mathbf{r}_i(t) - \mathbf{r}_{i,\text{ref}}|^2} \quad (2)$$

The summation is performed across all n_a C α atoms. $\mathbf{r}_i(t)$ represents the instantaneous position of the i th C α at time t , while $\mathbf{r}_{i,\text{ref}}$ denotes its corresponding position within the protein’s least-squares fitted reference structure, i.e., the RCSB structure.

Another important metric for assessing protein stability is the radius of gyration, R_g , defined as

$$R_g(t) = \sqrt{\frac{\sum_{i=1}^{n_h} m_i |\mathbf{r}_i(t) - \mathbf{r}_{\text{cm}}(t)|^2}{\sum_{i=1}^{n_h} m_i}} \quad (3)$$

Here, m_i and $\mathbf{r}_i(t)$ correspond to the mass and instantaneous position of atom i , respectively. The summation is over all n_h heavy (nonhydrogen) atoms within the protein. This approach is chosen to minimize the influence of highly fluctuating hydrogen nuclei on the computed R_g .

In addition to RMSD and radius of gyration, we also compute the solvent-accessible surface area (SASA) using the double cubic lattice method.⁵⁷ This also enables us to compute the average volume of the protein, more precisely van der Waals volume approximation of Connolly molecular volume,⁵⁸ using the Gauss-Ostrogradskii theorem. The change in protein’s volume ΔV (with respect to ambient conditions) is then computed using

$$\langle \Delta V \rangle = \langle V|_{T_i, P_i} \rangle_t - \langle V|_{T_a, P_a} \rangle_t \quad (4)$$

Here, the subscripts i and a refer to target and ambient conditions, respectively.

The local OPs considered in this work are all specific to HyD-crys and correspond to distances between specific residues, as will be discussed in Section 3. Unless otherwise specified, the distance between two residues is defined as the minimum distance between their heavy atoms. More precisely,

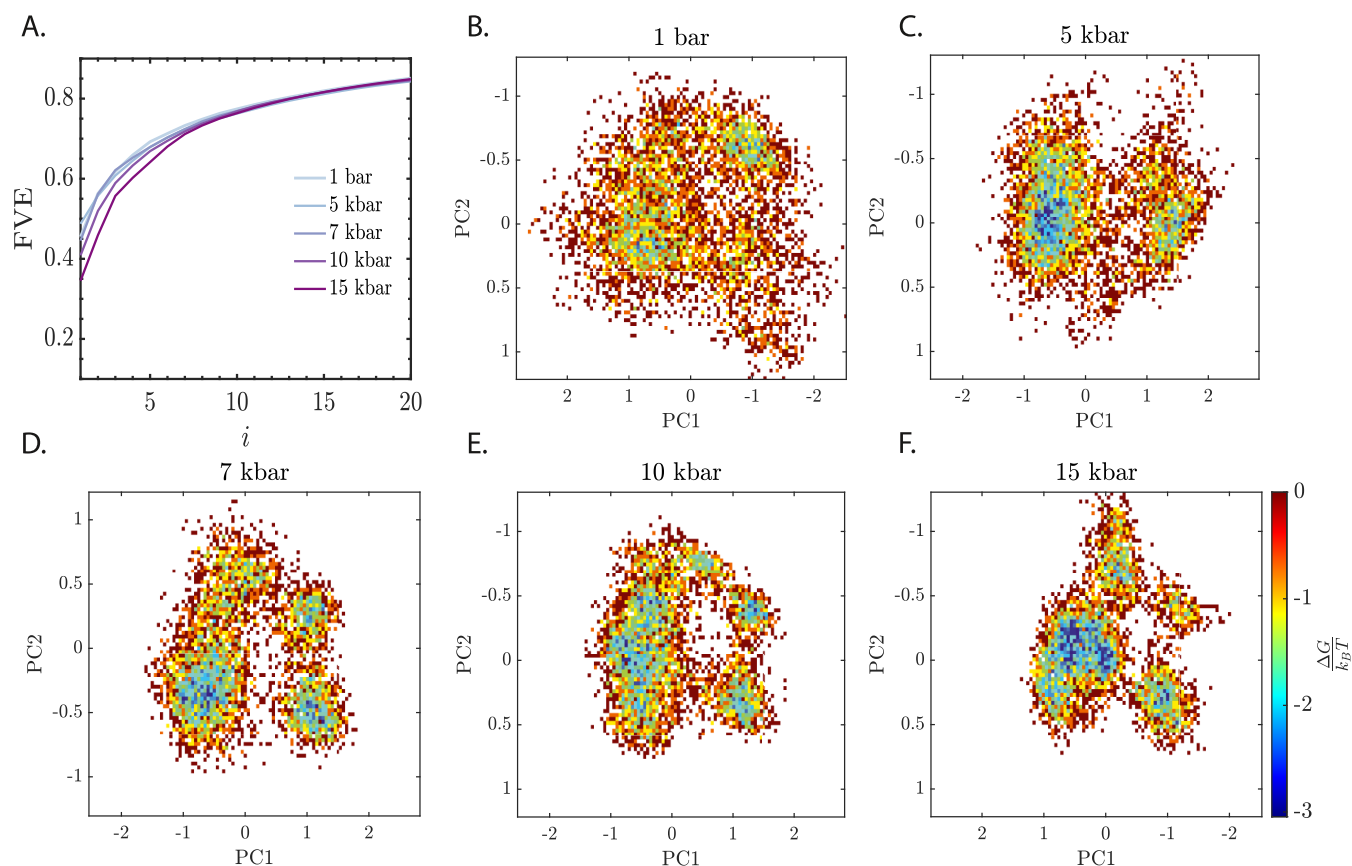


Figure 3. (A) Cumulative FVE for the first 20 eigenvalues of all target pressures. (B–F) Free energy profiles within the space of projections along the first two principal components at (B) 1 bar, (C) 5 kbar, (D) 7 kbar, (E) 10 kbar, and (F) 15 kbar.

if residues i and j comprise n_i and n_j heavy atoms, the instantaneous distance between them is defined as

$$d_{ij}(t) = \min_{l \leq n_i, m \leq n_j} |\mathbf{r}_{i,l}(t) - \mathbf{r}_{j,m}(t)| \quad (5)$$

where $\mathbf{r}_{i,l}$ and $\mathbf{r}_{j,m}$ correspond to the position of the l -th and m -th heavy atom within the i -th and j -th residue, respectively.

3. RESULTS AND DISCUSSION

3.1. Global Order Parameters. In order to assess the impact of pressure and temperature on the native structure of HyD-crys, we first calculate the global OPs described in Section 2.5 along the 60 ns REMD trajectories, with their averages and error bars depicted in Figure 2. Since elevating pressure results in an increase in the density of the water molecules surrounding the protein, both $\langle \text{RMSD} \rangle$ (Figure 2A) and $\langle R_g \rangle$ (Figure 2B) tend to decrease with pressure, indicating enhanced compactness and stiffness of the protein upon compression. In contrast, increasing temperature results in an increase in both of these quantities (Figure 2A,B). A similar trend is observed for $\langle \text{SASA} \rangle$, which increases slightly with temperature (Figure 2C) but exhibits a reduction of approximately 3 nm² with pressure (Figure 2C). This observation suggests that exposing the protein to high pressure has a more pronounced impact on its overall hydration structure compared to elevating temperature. This further reinforces the contrasting effects of temperature and pressure on these global order parameters. Our reported $\langle R_g \rangle$ and $\langle \text{SASA} \rangle$ values are consistent with those reported in earlier studies of HyD-crys.⁵⁹

Figure 2D depicts the change in protein volume upon an increase in pressure and temperature. Once more, the influence of high pressure on $\langle \Delta V \rangle$ is directionally different from that induced by high temperature. Additionally, akin to SASA, $\langle \Delta V \rangle$ undergoes more substantial changes with pressure than temperature. The impact of pressure on $\langle \Delta V \rangle$ aligns with Le Chatelier's principle, which predicts that an increase in pressure will prompt a shift in thermodynamic equilibrium toward configurations with smaller volumes.

The slight changes observed in these global measures of protein stability upon increasing the temperature and pressure align well with the known high stability of HyD-crys, which is highly resistant to unfolding. These global order parameters typically increase with temperature, yet the extent of this increase remains notably minimal within the considered temperature range. Conversely, in response to pressure, they uniformly exhibit a slight decrease, indicating that pressure-induced conformational changes are likely local in nature, while the overall fold remains mostly unchanged. This is consistent with earlier investigations of pressure-induced unfolding, which generally culminate in the formation of more compact conformations.³⁶

3.2. Free Energy Landscapes. Our examination of global order parameters confirms that the impact of pressure on the global structure of HyD-crys remains relatively mild, even under pressures as large as 15 kbar. This implies that increasing pressure only induces subtle changes in the protein's structure, evading detection through global order parameters. To identify and characterize such changes, we first scrutinized our REMD trajectories using PCA. However, the efficacy of PCA hinges

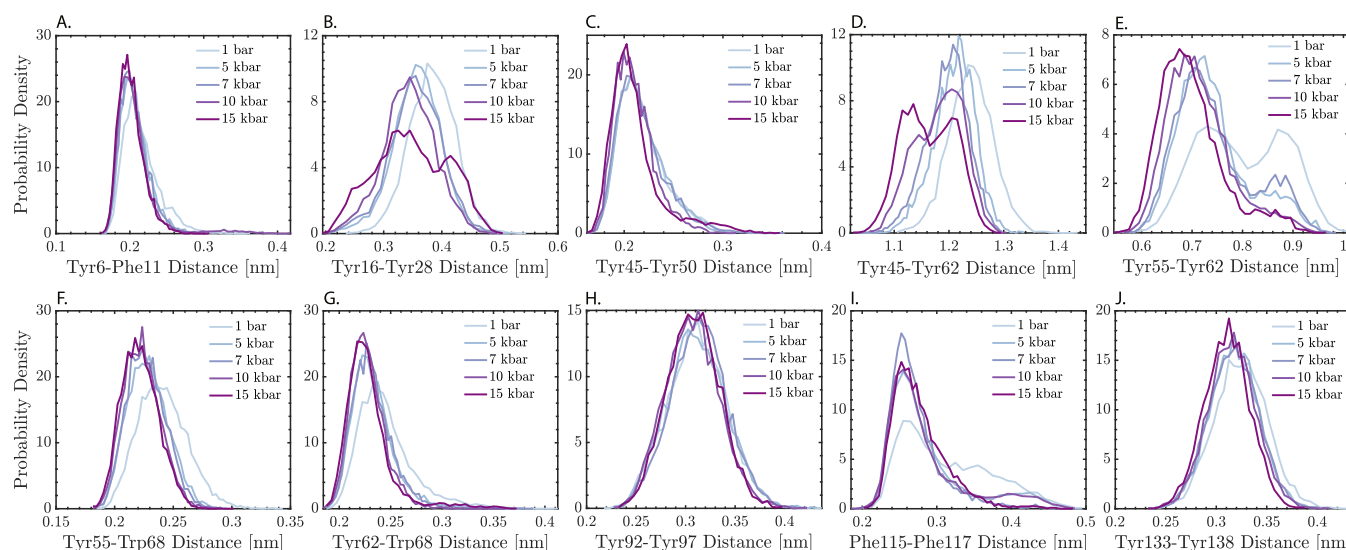


Figure 4. Histograms of minimum distances of the ten aromatic pairs previously identified as potential markers of H₇D-crys stability as a function of pressure. Among these, (A–G) and (H–J) correspond to contacts within the N-td and C-td, respectively.

upon the protein's local dynamics possessing sufficient simplicity to be amenable to the linearization inherent in PCA. Under such conditions, crucial conformational changes within the protein can be succinctly captured through a limited number of principal components. We believe that the absence of dramatic changes in H₇D-crys conformation, and the overall stability of global OPs justify the use of PCA here.

Based on the FVEs computed across all target pressures (Figure 3A), the first two leading eigenvalues capture over 50% of variability in the data. Consequently, we project each REMD trajectory onto the corresponding eigenvectors associated with these leading eigenvalues, resulting in a time series projection, $(\gamma_1(t), \gamma_2(t))$, with γ_i corresponding to the projection along the i th principal component. Subsequently, $\Delta G(\gamma_1, \gamma_2)$ is computed as follows. First, the (γ_1, γ_2) space is discretized into distinct bins, and the number of observations within each bin (N_{ij}) is determined. The free energy associated with each bin, ΔG_{ij} , is then estimated as⁶⁰

$$\frac{\Delta G_{ij}}{k_B T} = -\ln N_{ij} + C \quad (6)$$

where k_B denotes the Boltzmann constant and C is a normalization constant.

It is clearly evident that an increase in pressure results in the emergence of new basins within the free energy landscape (Figure 3B–F). Moreover, such basins are separated by modest free energy barriers of approximately $2-3k_B T$. This is in contrast to ambient pressure wherein PCA projections scatter within a single free energy basin, comprising shallow wells separated by barriers of approximately $0.5k_B T$. The structuring of the free energy landscape at elevated pressures suggests preferential stabilization of specific conformational states. To discern and characterize such conformational states and to associate them with different basins within the free energy landscape, it is imperative to identify local physically interpretable order parameters that correlate with PCA basins. In search of such local OPs, we focus on structural features that have been experimentally demonstrated to correlate with denaturation under environmental perturbations. Specifically, we investigate aromatic contacts within the N-terminal domain

as well as markers associated with the stability and hydration of the hydrophobic core.

3.3. Local Order Parameters. **3.3.1. N-Terminal Domain Stability.** As detailed in Section 1, aromatic contacts within the N-td have been demonstrated to be critical for H₇D-crys stability (Figure 1A).^{6,16–18,61} We, therefore, probe the impact of pressure on the minimum distances among specific aromatic pairs known to establish contacts within both N-td and C-td. Specifically, we examine six tyrosine–tyrosine pairs (Y16–Y28, Y45–Y50, Y45–Y62, Y55–Y62, Y92–Y97, and Y133–138), two tyrosine–tryptophan pairs (Y55–W68, Y62–W68), one tyrosine–phenylalanine pair (Y6–F11), and one phenylalanine–phenylalanine pair (F115–F117). Among these aromatic pairs, seven are associated with the reportedly less stable N-td, while the remaining three pairs facilitate interactions within the C-td.

Figure 4 illustrates the evolution of the histograms of these ten minimum distances upon increasing pressure. In general, most of these distances tend to decrease as pressure increases, with only a few remaining almost unchanged, such as Y6–F11 (Figure 4A), Y45–Y50 (Figure 4C), and Y92–Y97 (Figure 4H). For a few pairs, however, the qualitative shapes of the histograms also change, in addition to leftward shifts in their means. Specifically, both Tyr45–Tyr62 (Figure 4D) and Tyr55–Tyr62 (Figure 4E) pairs reveal the emergence of new peaks at 1.12 nm (Figure 4D) and 0.65 nm (Figure 4E), respectively. Furthermore, the singular peak of the Tyr16–Tyr28 distance bifurcates into two distinct peaks under high-pressure conditions (Figure 4B).

Based on these observations, it is reasonable to surmise the existence of a potential correlation between (some of) these distances and the basins identified through PCA. To facilitate the visual exploration of such possibility, we use MATLAB's linkage implementation⁶² of the agglomerate hierarchical cluster linkage algorithm to systematically group $(\gamma_1(t), \gamma_2(t))$'s, i.e., the PCA projections, into clusters, color-coding them based on their association with different principal component projections. This is conducted by computing Euclidean distances between all different PCA projections and grouping those with the closest distances into clusters, which are then recursively merged based on the largest distance

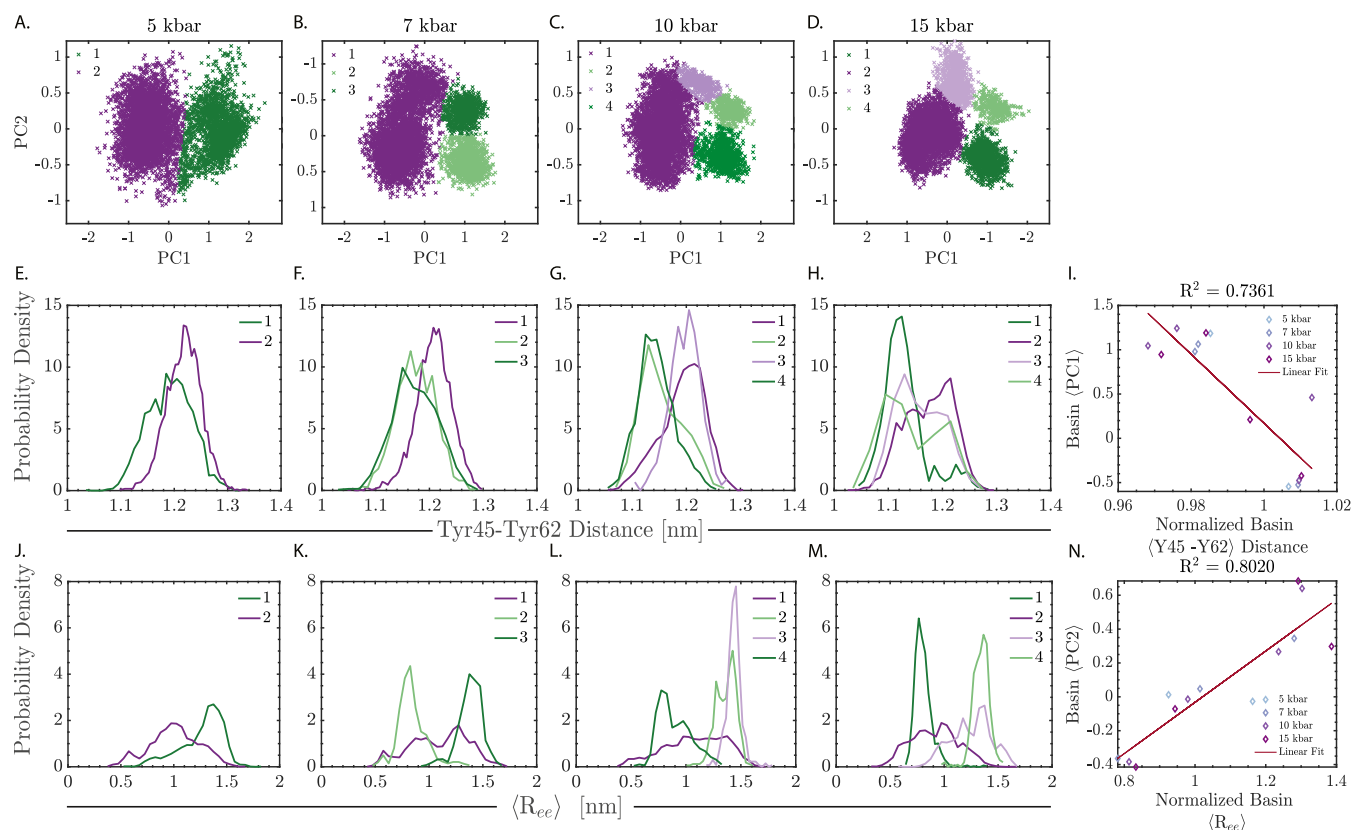


Figure 5. (A–D) Distinct free energy basins within the PC1–PC2 space colored based on their association with different principal component projections: green and purple for PC1 and dark and light for PC2. (E–H, J–M) Conditional histograms of (E–H) Y45–Y62 and (J–M) end-to-end minimum distances for the basins distinctly colored in (A–D). (I,N) Linear correlations between the normalized mean (I) Y45–Y62 and (N) end-to-end distances with (I) mean PC1 and (N) mean PC2 per basin, respectively.

between projections in any two distinct clusters. Clusters linked to distinct values of PC1 are denoted in purple and green, while those associated with distinct PC2 values are represented by darker and lighter shades of each color (Figure 5A–D). Subsequently, conditional histograms of the minimum distances considered in Figure 4 are constructed for the basins identified through the hierarchical cluster linkage algorithm to examine the existence of any correlations between either of those distances and the first two PCA projections.

Among all of the pairs considered in Figure 4, the Tyr45–Tyr62 pair (Figure 1B) exhibits the strongest correlation with the first principal component projection (PC1), with the other pairs exhibiting no meaningful correlation (Figure S2). As depicted in Figure 4D, the Y45–Y62 minimum distance histogram exhibits a peak at 1.25 nm at ambient pressure. Upon increasing pressure, this peak undergoes a leftward shift. Moreover, a second peak emerges at approximately 1.15 nm, initially appearing faintly at 5 kbar (Figure 5E), but gaining prominence at 7 kbar (Figure 5F) and beyond. Examining the conditional histograms of the Tyr45–Tyr62 distance (depicted in Figure 5E–H) elucidates a discernible correlation with PC1. Specifically, basins associated with smaller PC1 values consistently exhibit reduced Y45–Y62 distances despite some overlap evident in the respective histograms.

In order to make our analysis more quantitative, we computed the mean values of PC1 and the Y45–Y62 distance for each PCA basin. Moreover considering the concurrent decrease in the mean Y45–Y62 distance with pressure, we normalize each mean basin distance by the total mean distance

at the corresponding pressure. As illustrated in Figure 5I, a meaningful correlation emerges between the normalized basin mean Y45–Y62 distances and PC1 values, displaying a correlation coefficient of $R^2 = 0.7361$. This correlation underscores the association between the leading principal component projections and the structural alterations signified by a change in the Y45–Y62 distance.

The situation is more nuanced for other top contenders, most notably the Tyr55–Tyr62 pair (Figure S2J–N). At ambient pressure, the minimum distance histogram exhibits two peaks at 0.73 and 0.89 nm (Figure 4E). However, increasing pressure leads to a distinct leftward shift of the first peak, accompanied by a gradual reduction in the prominence of the second peak, which starts diminishing at 5 kbar and nearly disappears by 10 kbar. Despite the evident pressure-induced changes in the shape of the histogram, we fail to discern any significant correlation between the absolute or normalized mean distances within the PCA basins and either of the mean PCA projections (Figure S2N). A similar scenario unfolds for other important contenders, such as Y16–Y28 (Figure S2E–I) and F115–F117 (Figure S2O–S), where changes in histogram shapes occur with increasing pressure, yet meaningful correlations with PCA projections are not apparent (Figure S2I,S).

Collectively, these observations suggest that pressure stabilizes the native contacts among the aromatic residues, most importantly the tyrosine pairs, thereby intensifying the crowding effect within the β -sheet motifs of the N-td domain. This phenomenon aligns with observations from studies on

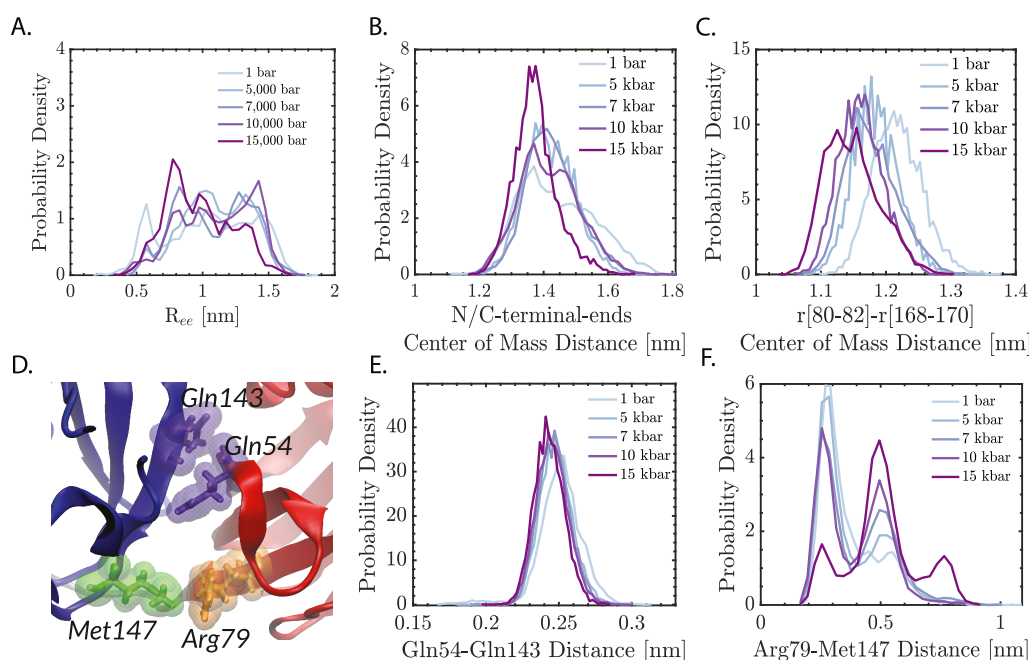


Figure 6. Histograms of local order parameters associated with the interdomain interface of HyD-crys, namely distance between (A) N- and C-terminal residues, (B,C) centers of masses of structureless regions within each terminal and (C) residues 80–82 from the N-terminal and residues 168–170 from the C-terminal. (D) Rendering of HyD-crys highlighting the flanking pairs Gln54–Gln143 and Arg79–Met147, (E,F) Histograms of minimum distances between (E) Gln54–Gln143 and (F) Arg79–Met147 as a function of pressure.

proteins of comparable sizes, where the crowding of β sheets has been identified as a hallmark of heightened native structural stability.^{22–24,63} Moreover, our observations suggest a correlation between the strength of aromatic contacts within the N-terminal domain and the first principal component from our PCA analysis.

3.3.2. Interdomain Interface Stability. Another pivotal determinant of HyD-crys stability is the integrity of the interdomain interface, which, if compromised due to mutations or environmental variations, can adversely affect the stability of the hydrophobic core within the protein.^{6,7,16,17} To investigate the pressure-induced effects on the stability of the interdomain interface, we examined several associated local order parameters.

The initial set of such order parameters focuses on quantifying the proximity between the N- and C-terminals. In the HyD-crys structure, the C-terminal extends over the interdomain interface, serving as both an anchoring point between protein lobes and a barrier to the protein's hydrophobic core. In particular, we consider the following order parameters: (i) the minimum distance between the first residue of the N-terminal and the last residue of the C-terminal (R_{ee}), representing the initial N-td and the terminal C-td residues, respectively, (ii) the distance between the centers of mass of the structureless loops of the N- and C-terminals, and (iii) the distance between the centers of mass of residues 80–81 and 168–170, corresponding to the closest β -sheets of the interdomain interface. The histograms of these order parameters as a function of pressure are depicted in Figure 6A–C. Unlike the consistent decrease observed in aromatic pair distances with increasing pressure, the influence of pressure on these order parameters is more intricate and, therefore, less straightforward to interpret. The center of mass distances between the N- and C-terminals, and the two closest β -sheets, generally exhibit a decline with escalating pressure

(Figure 6B,C). The end-to-end distance (R_{ee}), however, demonstrates rich and intricate behavior (Figure 6A). Despite its mean being virtually insensitive to pressure, the positions and magnitudes of peaks within its histogram exhibit a complex dependence on pressure.

The second category of order parameters are associated with the Gln54–Gln143 and Arg79–Met147 pairs, the two flanking pairs highlighted in Figure 6D and discussed in Section 1 that function as gateways to the hydrophobic core.¹⁶ The histograms of minimum distances of both pairs are depicted in Figure 6E,F. As depicted in Figure 6E, pressure has a minimal impact on the Gln54–Gln143 pair, leaving their distance virtually unchanged. In contrast, the histogram corresponding to the Arg79–Met147 pair distance (Figure 6F) exhibits a shift toward larger values at higher pressures. Once more, this underscores that the effect of pressure on the interdomain interface is complex and nontrivial, which is in stark contrast to the behavior observed for aromatic contacts.

Upon closer examination of conditional histograms of these order parameters for basins identified from PCA, we find a distinct separation between the conditional histograms of the end-to-end distance histograms (Figure 6A) associated with basins of distinct PC2 values. Naturally, such a distinction is absent at 5 kbar, where both basins share the same PC2 values (Figure 5J) and manifests itself at 7 kbar (Figure 5K) and above (Figure 5L,M) where basins with distinct PC2 values emerge. Similar to the approach discussed in Section 3.3.1, we compute the mean end-to-end distance per basin and normalize it with the corresponding mean at each pressure. As illustrated in Figure 5N, we observe a strong correlation (with a correlation coefficient of $R^2 = 0.8020$) among these normalized distances and the corresponding per-basin mean PC2 values. A weaker, but nonetheless meaningful, correlation ($R^2 = 0.7153$) is observed between mean center-of-mass distance of the loop regions of N- and C-terminals (Figure

S3I). This is not surprising considering the expected correlation between the center-of-mass distance and the end-to-end distance. No strong correlation was observed for the other three order parameters considered in Figure 6 (see Figure S3).

This observation confirms that the two collective variables derived from PCA are linked to physically interpretable local order parameters. More specifically, the Y45–Y62 aromatic pair contact within the N-terminal domain correlates with PC1, while the end-to-end distance correlates to PC2. Consequently, the basins identified via PCA do not represent modeling artifacts; rather, they possess distinct and clear physical interpretations.

Motivated by the correlation between the second principal component and the end-to-end distance, which is a proxy for the stability of the interdomain interface, we probe the effect of pressure on the size and solvent exposure of the hydrophobic core. To quantify the size of the hydrophobic core, we employ the volume of the convex hull (Figure 7A) constructed from the centers of masses of six crucial hydrophobic residues pivotal for maintaining the integrity of the interdomain interface (Figure 1C). As expected, the volume of the hydrophobic core decreases upon increasing pressure, as indicated by the histograms of convex hull volumes presented

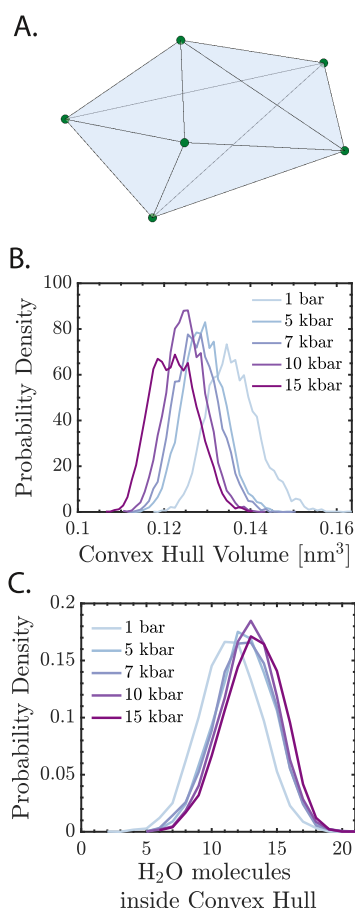


Figure 7. (A) Schematic representation of the convex hull approximation of the hydrophobic core using the centers of masses of the six hydrophobic residues, Met43, Phe56, Ile81, Val132, Leu145, and Val170 (shown in Figure 1C). (B,C) Histograms of (B) the volume of and (C) the number of water molecules within the convex hull as a function of pressure.

in Figure 7B. Interestingly though, the number of water molecules enclosed within the convex hull rises with pressure (Figure 7C), implying that pressure alters water's density and structure around the protein in a manner that counteracts the reduction in the accessible volume of the hydrophobic core. Evaluating the number of hydrogen bond donor–acceptor pairs surrounding the six hydrophobic residues reveals a minor increment in the number of protein–solvent hydrogen bonds, while the number of intramolecular hydrogen bonds within the hydrophobic core remains mostly unchanged (Figure S4). This observation once again validates the intrusion of water into the hydrophobic core under high-pressure conditions. The expansion of the hydrophobic core is likely triggered by the increase in the Arg79–Met147 pair distance, as depicted in Figure 6F.

4. CONCLUSIONS AND FUTURE OUTLOOK

In this study, we utilized REMD simulations augmented with PCA to explore the conformational landscape of human γ D-crystallin under elevated pressures. As anticipated, our findings reveal that high pressure only has a modest impact on global measures of protein stability, such as RMSD and radius of gyration. However, through the projection of REMD trajectories using the first two principal components, an intriguing picture emerges: the singular basin observed at ambient pressure divides into multiple distinct basins with increasing pressure. We probe multiple local order parameters previously demonstrated or hypothesized to be markers of H γ D-crys stability, namely, various aromatic contacts within the N-td, and a handful of characteristic distances associated with the interdomain interface. Our analyses reveal a correlation between the minimum distance of the aromatic pair, Tyr45–Tyr62, with PC1, as well as a relationship between the N-/C-td end-to-end distance and PC2. Furthermore, we observe the contraction of the hydrophobic core, accompanied by its intrusion by water molecules, possibly facilitated by the increased separation between Arg79 and Met147. These findings collectively provide deeper insights into the structural alterations occurring within H γ D-crystallin under elevated pressure conditions, shedding light on specific molecular changes influencing its stability.

It is generally understood that pressure affects protein folding through favoring intermediates that possess smaller specific volumes, which can result in both stabilizing and destabilizing the folded state. When it comes to H γ D-crystallin, it is unclear whether elevated pressures lead to its stabilization or disruption. On one hand, the consistent reduction in distances between aromatic pairs within the N-td suggests a potential overall stabilizing effect. Conversely, the infiltration of water molecules into the hydrophobic core implies potential destabilization of the interdomain interface. Consequently, the overall effect of pressure on the stability of H γ D-crystallin remains undetermined, as the opposing influences on different structural elements necessitate a comprehensive evaluation to ascertain the net impact.

To comprehensively address this issue, it is imperative to determine the change in folding free energy, typically achieved through methods such as thermodynamic integration that rely on the two-state model for protein folding.^{36,64} However, this task poses significant challenges due to substantial free energy barriers separating the folded and unfolded states of large proteins such as H γ D-crys. Consequently, standard enhanced sampling techniques, such as REMD, struggle to efficiently

capture unfolding–folding transitions, despite their ability to sample rugged free energy landscapes. Remedying this issue might involve utilizing bias-based or path sampling techniques. However, such methods often require the specification of predefined reaction coordinates, which are generally unknown *a priori* for complex processes such as protein folding. Moreover, the accuracy of the two-state folding model remains unverified, contributing an additional layer of uncertainty to this conundrum. This uncertainty, coupled with the substantial energy barriers, renders the computation of folding free energy a daunting task that is beyond the scope of this work.

We also emphasize that the pressure-induced conformational changes observed here can be characterized using a host of other dimensionality reduction techniques. Our choice of PCA is motivated by the subtle nature of the observed changes to HyD-crystallin's conformation. Through PCA, we establish meaningful correlations between crucial physically interpretable local order parameters and projections along the two principal components. However, it is important to note that alternative nonlinear methods, such as diffusion maps,^{65,66} isometric feature maps⁶⁷ (Isomaps), and *t*-distributed stochastic neighbor embedding⁶⁸ (*t*-SNE), might offer increased efficacy in capturing relevant low-dimensional embeddings of the free energy landscapes, potentially yielding a qualitatively different picture. Nonetheless, a significant challenge with applying such methods lies in the often less straightforward interpretability of the collective variables that they produce. Assessing the sensitivity of our findings to the employed dimensionality reduction technique can be the focus of future explorations.

To identify physically interpretable collective variables that align with the leading principal component projections, we focus on local order parameters previously established or suggested as meaningful indicators of HyD-crystallin stability in prior experimental studies. While this targeted strategy enhances our ability to pinpoint relevant local order parameters, it also poses a potential limitation. By relying primarily on previously identified order parameters, we may inadvertently restrict our exploration, potentially overlooking alternative local order parameters that might contribute to understanding the system but have not been detected in earlier investigations.

We conclude with a few broader implications of investigating the conformational landscape of globular proteins such as HyD-crystallin under elevated pressures. Examining pressure effects is relevant to the field of extreme biology, offering insights into the physiology of organisms adapted to high-pressure environments, such as the deep ocean or subsea floor.^{28–30} The adaptability of biological macromolecules to elevated pressures serves as a foundation for diverse biotechnological applications utilizing high pressure.⁶⁹

Furthermore, the use of pressure, often combined with cosolvents, has emerged as a valuable tool in investigating the stability, energetics, packing, and intermolecular interactions of protein aggregates, such as amyloid fibrils.⁷⁰ These fibrils result from the misfolding of proteins such as insulin, islet amyloid polypeptide, β 2 microglobulin, amyloid β , and α -synuclein in various tissues.^{71–73} Specifically, high hydrostatic pressure facilitates distinguishing various stages of amyloid formation, enabling the acquisition of crucial structural and thermodynamic data at different transformation stages. This approach provides valuable insights into the mechanisms underlying amyloid formation and offers opportunities for developing

therapeutic interventions targeting these protein misfolding disorders.

Another intriguing aspect of pressure effects lies in the striking resemblance between the structural changes in pure water under high pressure and those induced by the introduction of solutes. This observation suggests a correlation between such similarities and subsequent dynamic processes. For instance, it is well established that the rate of homogeneous ice nucleation in both aqueous solutions and pressurized pure water depends solely on the water activity.⁷⁴ Given the paramount importance of water structure and dynamics in protein folding, it is plausible to posit a similar connection between water activity and the thermodynamics as well as kinetics of protein folding. Consequently, investigating pressure effects becomes appealing as pressure could serve as a mild perturbation mimicking solute effects. This is important as, under physiologically relevant conditions, nearly all globular proteins fold within aqueous solutions comprised of a large number of different solute types. Therefore, exploring pressure effects would offer an attractive alternative to model diverse solutes if the existence of such a hypothesized correlation can be established. This is an exciting possibility that can be the topic of future experimental and computational explorations.

On a related note, the impact of high pressure on HyD-crys is most vividly evident in the negative $\langle \Delta V \rangle$, an effect that can also be induced by other perturbations such as crowding agents and cosolutes. For instance, negative $\langle \Delta V \rangle$ can arise from the disruption of salt bridges and ionic bonds along with associated changes in hydration due to the presence of ionic salts. Considering these diverse influences, it becomes crucial to investigate the effect of cosolutes, including simple monovalent ions (such as K^+) and small organic molecules, on both the conformational landscape of the monomer and its capacity to assemble into high molecular weight oligomers. Such research will contribute significantly to understanding the intricate interplay between various perturbations and their impact on the behavior and structural characteristics of HyD-crys.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcb.4c00178>.

Further details about the suitability of the employed force-field and the procedure utilized for characterizing the hydrogen bond network, as well as four additional figures on stability of global OPs, and screening of local OPs for correlation with principal component projections and the analysis of the hydrogen bond network (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Amir Haji-Akbari — Department of Chemical and Environmental Engineering, Yale University, New Haven, Connecticut 06520, United States; orcid.org/0000-0002-2228-6957; Email: amir.hajiakbaribalou@yale.edu

Authors

Arlind Kacirani — Department of Chemical and Environmental Engineering and Integrated Graduate Program in Physical and Engineering Biology, Yale University, New Haven, Connecticut 06520, United States

Betül Uralcan — Department of Chemical and Environmental Engineering, Yale University, New Haven, Connecticut 06520, United States; Department of Chemical Engineering, Boğaziçi University, Istanbul 34342, Turkey; orcid.org/0000-0002-0669-8441

Tiago S. Domingues — Department of Chemical and Environmental Engineering and Graduate Program in Applied Mathematics, Yale University, New Haven, Connecticut 06520, United States

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jpcb.4c00178>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank J. Howard for useful discussions. A.H.-A. gratefully acknowledges the support of the National Science Foundation CAREER Award (CBET-1751971). These calculations were performed at the Yale Center for Research Computing (YCRC). This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant no. ACI-1548562.

REFERENCES

- (1) Flaxman, S. R.; Bourne, R. R.; Resnikoff, S.; Ackland, P.; Braithwaite, T.; Cicinelli, M. V.; Das, A.; Jonas, J. B.; Keeffe, J.; Kempen, J.; et al. Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. *Lancet Global Health* **2017**, *5*, e1221–e1234.
- (2) Benedek, G. B. Cataract as a protein condensation disease: the Proctor Lecture. *Invest. Ophthalmol. Visual Sci.* **1997**, *38*, 1911–1921.
- (3) Andley, U. P. Crystallins in the eye: Function and pathology. *Prog. Retin. Eye Res.* **2007**, *26*, 78–98.
- (4) Bloemendal, H.; De Jong, W.; Jaenicke, R.; Lubsen, N. H.; Slingsby, C.; Tardieu, A. Ageing and vision: Structure, stability and function of lens crystallins. *Prog. Biophys. Mol. Biol.* **2004**, *86*, 407–485.
- (5) Tanaka, T.; Benedek, G. B. Observation of protein diffusivity in intact human and bovine lenses with application to cataract. *Invest. Ophthalmol. Visual Sci.* **1975**, *14*, 449–456.
- (6) Das, P.; King, J. A.; Zhou, R. β -strand interactions at the domain interface critical for the stability of human lens γ D-Crystallin. *Protein Sci.* **2010**, *19*, 131–140.
- (7) Das, P.; King, J. A.; Zhou, R. Aggregation of γ -crystallins associated with human cataracts via domain swapping at the C-terminal β -strands. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10514–10519.
- (8) Basak, A.; Bateman, O.; Slingsby, C.; Pande, A.; Asherie, N.; Ogun, O.; Benedek, G. B.; Pande, J. High-resolution X-ray Crystal Structures of Human γ D Crystallin (1.25 Å) and the R58H Mutant (1.15 Å) Associated with Aculeiform Cataract. *J. Mol. Biol.* **2003**, *328*, 1137–1147.
- (9) Kosinski-Collins, M. S.; King, J. In vitro unfolding, refolding, and polymerization of human γ D Crystallin, a protein involved in cataract formation. *Protein Sci.* **2003**, *12*, 480–490.
- (10) Serebryany, E.; King, J. A. The $\beta\gamma$ -crystallins: native state stability and pathways to aggregation. *Prog. Biophys. Mol. Biol.* **2014**, *115*, 32–41.
- (11) Garcia-Manyes, S.; Giganti, D.; Badilla, C. L.; Lezamiz, A.; Perales-Calvo, J.; Beedle, A. E.; Fernández, J. M. Single-molecule Force Spectroscopy Predicts a Misfolded, Domain-swapped Conformation in human γ D-Crystallin Protein. *J. Biol. Chem.* **2016**, *291*, 4226–4235.
- (12) Boatz, J. C.; Whitley, M. J.; Li, M.; Gronenborn, A. M.; Van Der Wel, P. C. Cataract-associated P23T γ D-Crystallin retains a native-like fold in amorphous-looking aggregates formed at physiological pH. *Nat. Commun.* **2017**, *8*, 15137.
- (13) Moran, S. D.; Zhang, T. O.; Zanni, M. T. An alternative structural isoform in amyloid-like aggregates formed from thermally denatured human γ D-Crystallin. *Protein Sci.* **2014**, *23*, 321–331.
- (14) Khan, A. R.; James, S.; Quinn, M. K.; Altan, I.; Charbonneau, P.; McManus, J. J. Temperature-Dependent Interactions Explain Normal and Inverted Solubility in a γ D-Crystallin Mutant. *Biophys. J.* **2019**, *117*, 930–937.
- (15) Serebryany, E.; King, J. A. Wild-type Human γ D-Crystallin Promotes Aggregation of Its Oxidation-mimicking, Misfolding-prone W42Q Mutant. *J. Biol. Chem.* **2015**, *290*, 11491–11503.
- (16) Flaugh, S. L.; Kosinski-Collins, M. S.; King, J. Interdomain side-chain interactions in human γ D Crystallin influencing folding and stability. *Protein Sci.* **2005**, *14*, 2030–2043.
- (17) Flaugh, S. L.; Kosinski-Collins, M. S.; King, J. Contributions of hydrophobic domain interface interactions to the folding and stability of human D-Crystallin. *Protein Sci.* **2005**, *14*, 569–581.
- (18) Kong, F.; King, J. Contributions of aromatic pairs to the folding and stability of long-lived human γ D-Crystallin. *Protein Sci.* **2011**, *20*, 513–528.
- (19) Serrano, L.; Bycroft, M.; Fersht, A. R. Aromatic-aromatic interactions and protein stability. *J. Mol. Biol.* **1991**, *218*, 465–475.
- (20) Simkovsky, R.; King, J. An elongated spine of buried core residues necessary for in vivo folding of the parallel β -helix of P22 tailspike adhesin. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 3575–3580.
- (21) Roskamp, K. W.; Paulson, C. N.; Brubaker, W. D.; Martin, R. W. Function and Aggregation in Structural Eye Lens Crystallins. *Acc. Chem. Res.* **2020**, *53*, 863–874.
- (22) Macdonald, J. T.; Purkiss, A. G.; Smith, M. A.; Evans, P.; Goodfellow, J. M.; Slingsby, C. Unfolding crystallins: The destabilizing role of a β -hairpin cysteine in β B2-Crystallin by simulation and experiment. *Protein Sci.* **2005**, *14*, 1282–1292.
- (23) Waters, M. L. Aromatic interactions in peptides: Impact on structure and function. *Protein Sci.* **2004**, *76*, 435–445.
- (24) Rea, A. M.; Simpson, E. R.; Meldrum, J. K.; Williams, H. E.; Searle, M. S. Aromatic Residues Engineered into the β -Turn Nucleation Site of Ubiquitin Lead to a Complex Folding Landscape, Non-Native Side-Chain Interactions, and Kinetic Traps. *Biochemistry* **2008**, *47*, 12910–12922.
- (25) Koide, S.; Sidhu, S. S. The Importance of Being Tyrosine: Lessons in Molecular Recognition from Minimalist Synthetic Binding Proteins. *ACS Chem. Biol.* **2009**, *4*, 325–334.
- (26) Martin, E. W.; Holehouse, A. S.; Peran, I.; Farag, M.; Incicco, J.; Bremer, A.; Grace, C. R.; Soranno, A.; Pappu, R. V.; Mittag, T. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **2020**, *367*, 694–699.
- (27) Bremer, A.; Farag, M.; Borchers, W. M.; Peran, I.; Martin, E. W.; Pappu, R. V.; Mittag, T. Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains. *Nat. Chem.* **2022**, *14*, 196–207.
- (28) Yancey, P. H.; Gerringer, M. E.; Drzen, J. C.; Rowden, A. A.; Jamieson, A. Marine fish may be biochemically constrained from inhabiting the deepest ocean depths. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 4461–4465.
- (29) Morita, T. Structure-based analysis of high pressure adaptation of α -actin. *J. Biol. Chem.* **2003**, *278*, 28060–28066.
- (30) Somero, G. N. Biochemical ecology of deep-sea animals. *Experientia* **1992**, *48*, 537–543.
- (31) Gross, M.; Jaenicke, R. Proteins under pressure. *Eur. J. Biochem.* **1994**, *221*, 617–630.
- (32) Mozhaev, V. V.; Heremans, K.; Frank, J.; Masson, P.; Balny, C. High Pressure Effects on Protein Structure and Function. *Proteins* **1996**, *24*, 81–91.
- (33) Silva, J. L.; Foguel, D.; Royer, C. A. Pressure provides new insights into protein folding, dynamics and structure. *Trends Biochem. Sci.* **2001**, *26*, 612–618.
- (34) Hummer, G.; Garde, S.; García, A. E.; Paulaitis, M. E.; Pratt, L. R. The pressure dependence of hydrophobic interactions is consistent

with the observed pressure denaturation of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1552–1555.

(35) Hillson, N.; Onuchic, J. N.; Garcia, A. E. Pressure-induced protein-folding/unfolding kinetics. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14848–14853.

(36) Uralcan, B.; Debenedetti, P. G. Computational Investigation of the Effect of Pressure on Protein Stability. *J. Phys. Chem. Lett.* **2019**, *10*, 1894–1899.

(37) Chen, C. R.; Makhatadze, G. I. Molecular determinant of the effects of hydrostatic pressure on protein folding stability. *Nat. Commun.* **2017**, *8*, 14561.

(38) Zhuravlev, P. I.; Materese, C. K.; Papoian, G. A. Deconstructing the native state: energy landscapes, function, and dynamics of globular proteins. *J. Phys. Chem. B* **2009**, *113*, 8800–8812.

(39) Falahati, H.; Haji-Akbari, A. Thermodynamically driven assemblies and liquid-liquid phase separations in biology. *Soft Matter* **2019**, *15*, 1135–1154.

(40) Dellago, C.; Bolhuis, P. G.; Chandler, D. Efficient transition path sampling: Application to Lennard-Jones cluster rearrangements. *J. Chem. Phys.* **1998**, *108*, 9236–9245.

(41) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(42) Hussain, S.; Haji-Akbari, A. Studying rare events using forward-flux sampling: Recent breakthroughs and future outlook. *J. Chem. Phys.* **2020**, *152*, 060901.

(43) Shea, J. E.; Levine, Z. A. Studying the early stages of protein aggregation using replica exchange molecular dynamics simulations. *Methods Mol. Biol.* **2016**, *1345*, 225–250.

(44) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.

(45) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950–1958.

(46) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(47) Jorgensen, W. L.; Tirado-Rives, J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6665–6670.

(48) Kùhrová, P.; De Simone, A.; Otyepka, M.; Best, R. B. Force-Field Dependence of Chignolin Folding and Misfolding: Comparison with Experiment and Redesign. *Biophys. J.* **2012**, *102*, 1897–1906.

(49) Beauchamp, K. A.; Lin, Y. S.; Das, R.; Pande, V. S. Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements. *J. Chem. Theory Comput.* **2012**, *8*, 1409–1414.

(50) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.

(51) Nosé, S.; Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983**, *50*, 1055–1076.

(52) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(53) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(54) Berendsen, H. J.; Postma, J. P.; Van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(55) Greenacre, M.; Groenen, P. J.; Hastie, T.; d'Enza, A. I.; Markos, A.; Tuzhilina, E. Principal component analysis. *Nat. Rev. Methods Primers* **2022**, *2*, 100.

(56) Maisuradze, G. G.; Liwo, A.; Scheraga, H. A. Principal component analysis for protein folding dynamics. *J. Mol. Biol.* **2009**, *385*, 312–329.

(57) Eisenhaber, F.; Lijnzaad, P.; Argos, P.; Sander, C.; Scharf, M. The double cubic lattice method: Efficient approaches to numerical

integration of surface area and volume and to dot surface contouring of molecular assemblies. *J. Comput. Chem.* **1995**, *16*, 273–284.

(58) Connolly, M. L. Computation of molecular volume. *J. Am. Chem. Soc.* **1985**, *107*, 1118–1124.

(59) Brudar, S.; Gujt, J.; Spohr, E.; Hribar-Lee, B. Studying the mechanism of phase separation in aqueous solutions of globular proteins via molecular dynamics computer simulations. *Phys. Chem. Chem. Phys.* **2021**, *23*, 415–424.

(60) Dill, K. A.; Bromberg, S.; Yue, K.; Chan, H. S.; Ftebig, K. M.; Yee, D. P.; Thomas, P. D. Principles of protein folding—a perspective from simple exact models. *Protein Sci.* **1995**, *4*, 561–602.

(61) Hsueh, S. S.; Wang, S. S.; Chen, S. H.; Wang, C. L.; Wu, W. J.; Lin, T. H. Insights to Human γ D-Crystallin Unfolding by NMR Spectroscopy and Molecular Dynamics Simulations. *Int. J. Mol. Sci.* **2022**, *23*, 1591.

(62) The MathWorks Inc. Linkage, Agglomerative hierarchical cluster tree, Matlab Version 2023 Statistical and Machine Learning Toolbox. Online, 2023 <https://www.mathworks.com/help/stats/linkage.html>.

(63) Stagg, L.; Zhang, S. Q.; Cheung, M. S.; Wittung-Stafshede, P. Molecular crowding enhances native structure and stability of α/β protein flavodoxin. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18976–18981.

(64) Hawley, S. A. Reversible Pressure-Temperature Denaturation of Chymotrypsinogen. *Biochemistry* **1971**, *10*, 2436–2442.

(65) Ferguson, A. L.; Panagiotopoulos, A. Z.; Debenedetti, P. G.; Kevrekidis, I. G. Systematic determination of order parameters for chain dynamics using diffusion maps. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 13597–13602.

(66) Kim, S. B.; Dsilva, C. J.; Kevrekidis, I. G.; Debenedetti, P. G. Systematic characterization of protein folding pathways using diffusion maps: Application to Trp-cage miniprotein. *J. Chem. Phys.* **2015**, *142*, 85101.

(67) Tenenbaum, J. B.; Silva, V. d.; Langford, J. C. A global geometric framework for nonlinear dimensionality reduction. *Science* **2000**, *290*, 2319–2323.

(68) van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.

(69) Knorr, D.; Heinz, V.; Buckow, R. High pressure application for food biopolymers. *Biochim. Biophys. Acta* **2006**, *1764*, 619–631.

(70) Ostermeier, L.; de Oliveira, G. A.; Dzwolak, W.; Silva, J. L.; Winter, R. Exploring the polymorphism, conformational dynamics and function of amyloidogenic peptides and proteins by temperature and pressure modulation. *Biophys. Chem.* **2021**, *268*, 106506.

(71) Baba, M.; Nakajo, S.; Tu, P. H.; Tomita, T.; Nakaya, K.; Lee, V. M.; Trojanowski, J. Q.; Iwatsubo, T. Aggregation of alpha-synuclein in Lewy bodies of sporadic Parkinson's disease and dementia with Lewy bodies. *Am. J. Pathol.* **1998**, *152*, 879.

(72) Hardy, J.; Selkoe, D. J. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* **2002**, *297*, 353–356.

(73) Akter, R.; Cao, P.; Noor, H.; Ridgway, Z.; Tu, L. H.; Wang, H.; Wong, A. G.; Zhang, X.; Abedini, A.; Schmidt, A. M.; et al. Islet Amyloid Polypeptide: Structure, Function, and Pathophysiology. *J. Diabetes Res.* **2016**, *2016*, 1–18.

(74) Koop, T.; Luo, B.; Tsias, A.; Peter, T. Water activity as the determinant for homogeneous ice nucleation in aqueous solutions. *Nature* **2000**, *406*, 611–614.