Cell Stress and Chaperones (2025) 30:100086
https://doi.org/10.1016/j.cstres.2025.100086

Pathogenic mechanism of the K141E mutation in HSPB8: Insights from smFRET and simulations

Daniele Montepietra[1,2,†,#] · Sveinn Bjarnason[3,†] · Kristinn R. Óskarsson[3] · Ciro Cecconi[1,2] · Serena Carra[4] · Pétur O. Heidarsson[3,5,*] · Giorgia Brancolini[2,*]

Received: 8 April 2025 / Revised: 21 May 2025 / Accepted: 22 May 2025

Abstract

Pathogenic mutations can have a large impact on the conformational ensemble of intrinsically disordered proteins, but revealing those effects and their physiological relevance can be challenging. We used large-scale all-atom explicit-solvent molecular dynamics simulations and single-molecule Förster resonance energy transfer (smFRET) experiments to investigate the conformational dynamics of the chaperone protein HSPB8 and its K141E mutant that is linked to motor neuropathies. Our findings revealed that the HSPB8-K141E mutant exhibits increased conformational flexibility compared to the wild-type protein, particularly at high physiological ionic strengths, leading to a more extended conformational ensemble. Bayesian maximum entropy reweighting was applied to improve agreement between simulated and experimental smFRET data, further emphasizing the mutation's influence on protein dynamics. While both WT and K141E showed similar primary smFRET peaks after reweighting, the mutant displayed a higher occurrence of a secondary peak at lower FRET, indicative of an unfolded state. Additionally, differences in salt bridge networks between the variants highlighted the role of ionic interactions in modulating protein structure and suggest a possible connection between rapid dynamics and conformational stability. These results suggest that the pathogenicity of the K141E mutation may be, at least in part, due to the enhanced conformational variability that negatively influences the protein function. The study underscores the significance of ionic strength in the structural dynamics of intrinsically disordered proteins like HSPB8, providing insights into the functional implications of these changes and how stability changes can manifest across different timescales.

Keywords Heat shock protein B8 · Protein mutations · Intrinsically disordered proteins · Enhanced sampling simulations · Neural networks · Förster resonance energy transfer

Abbreviations: smFRET, single-molecule Förster resonance energy transfer; HSPB8, Heat Shock Protein B8; ACD, α-crystallin domain; IDRs, intrinsically disordered regions; IDPs, Intrinsically Disordered Proteins; NMR, Nuclear Magnetic Resonance; CD, Circular Dichroism; MD, Molecular dynamics simulations; NTD, N-terminal; CTD, C-terminal; KDE, Kernel Density Estimation; TREMD, temperature replica exchange molecular dynamics simulations; sHsps, small heat shock proteins; distal HMN, Distal Hereditary Motor Neuropathy; CMT2L, Charcot-Marie-Tooth Neuropathy Type 2L; EDTA, ethylenediaminetetraacetic acid; DTT, dithiothreitol; GSH, glutathione; RP-HPLC, reversed-phase high-performance liquid chromatography; TFA, trifluoroacetic acid; GdmCl, guanidinium chloride; RT, room temperature; TCSPC, time-correlated single photon counting; IOS, ionic strengths; PME, Particle Mesh Ewald; RMSF, Root Mean Square Fluctuations; SAW, self-avoiding walk

* Pétur O. Heidarsson
poh@bio.ku.dk, pheidarsson@hi.is

* Giorgia Brancolini
giorgia.brancolini@nano.cnr.it

[1] Department of Physics, Computer Science and Mathematics, University of Modena and Reggio Emilia, Modena, Italy

[2] Nanoscience Institute – CNR-NANO, Center S3, Modena, Italy

[3] Department of Biochemistry, University of Iceland, Reykjavík, Iceland

[4] Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy

[5] Department of Biology, University of Copenhagen, Copenhagen N, Denmark

[†] Present address: Department of Biotechnology, University of Verona, Strada Le Grazie 15, 37134 Verona, Italy.
[#] These authors contributed equally.

Published online: 29 May 2025

Introduction

Heat shock protein B8 (HSPB8) belongs to the family of small HSPs (HSPBs), which are classified as molecular chaperones and participate in the maintenance of cellular proteostasis.[1] HSPB8 forms a stoichiometric complex with the cochaperone BAG3 and the master chaperone HSP70. The complex is expressed in different cell types, including skeletal muscle cells and neurons, where it promotes the autophagic removal of mutated and misfolded proteins,[2–4] whose aggregation is associated with neuromuscular and neurodegenerative diseases.[2,5] In contrast to Adenosine triphosphate (ATP)-dependent chaperones such as HSP70 that hydrolyzes ATP to assist protein refolding and are classified as "foldase," HSPB8 and the other HSPB family members are ATP-independent chaperones that lack refolding activity and are, as such, classified as "holdases."[1] Recent studies using Optical Tweezers have shown that HSPB8 prevents protein aggregation without affecting native protein folding, acting as a "holder" that halts the progression of misfolded conformations.[6]

Structurally, HSPB8, as well as the other HSPB family members, comprises a conserved α-crystallin domain (ACD), along with intrinsically disordered N-terminal (NTD) and C-terminal (CTD) regions and exist as an ensemble of oligomers of different sizes, ranging from dimers to 4-mer, up to 24-mer.[1] The α-crystallin domain is required for the chaperone activity of the HSPBs, while the disordered N-terminal and C-terminal regions are involved in dimer and oligomer formation.[1] While the functional role of HSPB8 in preventing protein aggregation and promoting autophagy-mediated clearance of misfolded and mutated substrates[7] is gradually becoming clearer,[2,4,8–10] detailed atomistic insights into its structure remain elusive, as no experimental 3D structure of HSPB8 has yet been determined. The structural characterization of its intrinsically disordered NTD and CTD regions poses significant challenges, as most experimental techniques are optimized for structured proteins. Consequently, computational simulations have become an invaluable complement to experimental methods, offering insights into the conformational ensembles of intrinsically disordered proteins (IDPs) and regions (IDRs). Concerning HSPB8 and small heat shock proteins in general, their highly flexible IDRs shape the protein's free energy landscape, resulting in multiple shallow minima and enabling the existence of a broad ensemble of conformations under physiological conditions. Molecular simulations, coupled with dimensionality reduction techniques, have provided insights into the interplay between these domains, particularly the functional contributions of the disordered regions.[11]

As mentioned above, ATP-dependent and ATP-independent chaperones play an important role in the maintenance of protein homeostasis, whose loss is associated with aging and neurodegeneration.[12] An increasing number of mutations in genes coding for chaperones is associated with neurodegenerative disorders, especially neuromuscular and muscular disorders. These include mutations in genes coding for members of the HSP40/DNAJ family and the HSPB8 family, including HSPB8 itself and its stoichiometric partner BAG3.[13–15] Understanding how these mutations alter the stability and function of these chaperones is instrumental for the design of future therapeutic approaches. We recently characterized the K141E variant within the α-crystallin domain of HSPB8, which is associated with distal hereditary motor neuropathy, Charcot–Marie–Tooth neuropathy type 2L[16] and distal myopathy[17] (Figure 1). Despite maintaining the native folding propensity, the K141E substitution negatively impacts the protein's antiaggregation activity *in vitro*, as well as in cellular and animal model characterized by protein aggregation,[8,18] The reduced chaperone-activity of HSPB8 K141E compared to the WT protein can be at least in part be ascribed to a decreased interaction with the partner BAG3, an impaired HSPB8 dimerization, and a reduced binding affinity to non-native proteins.[4,18]

To gain further insight into how the K141E mutation induces structural changes under varying ionic strengths, we combined experimental techniques, such as single-molecule Förster resonance energy transfer (smFRET) and fluorescence lifetime measurements, with atomistic molecular dynamics (MD) simulations to compute smFRET efficiencies. Site-specific fluorophore labeling was achieved by replacing selected cysteine residues with donor and acceptor dyes, enabling precise measurement of intramolecular distances. Using smFRET and lifetime analysis, we measured distances and dynamics between selected amino acid residues in both HSPB8 and its K141E variant, examining how these distances change across a range of ionic strengths, from low to high. We then integrated single-molecule spectroscopy with large-scale all-atom explicit-solvent MD simulations to gain insight into the intramolecular distance distributions at nanometer scales and dynamics down to nanosecond timescales. Using explicit solvent and ions, we performed unbiased molecular large-scale simulations with IDP-specific force field parameters and ion models of HSPB8 and the mutation K141E. Our analyses revealed that the protein chain undergoes different degrees of compaction or expansion in response to ionic strength variations, due to favorable ion interactions with the protein backbone and specific interactions with various residue side-chains.

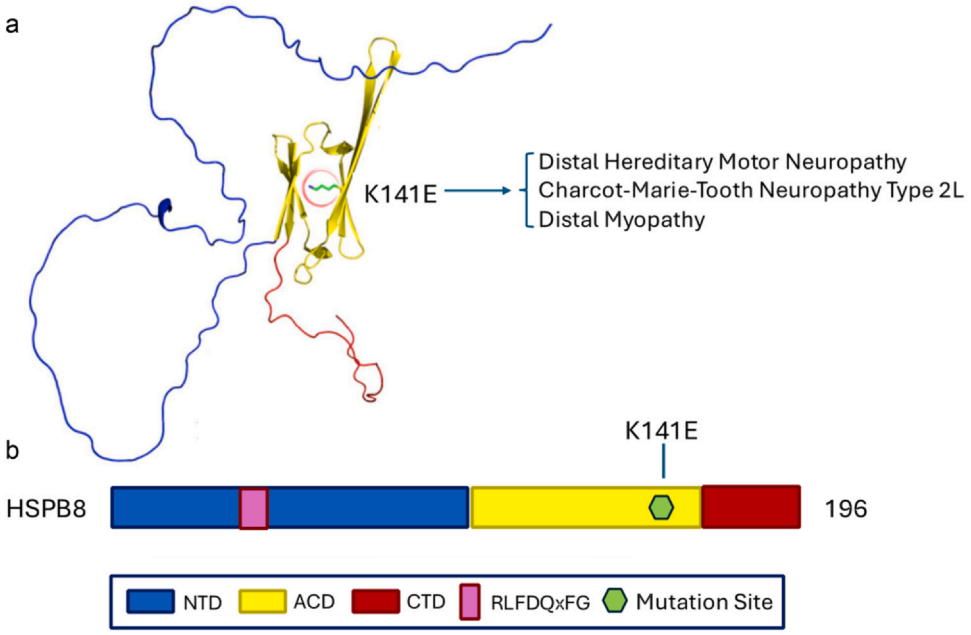

Fig. 1 A graphical overview of human small heat shock protein B8 (HSPB8) and its K141E mutant: (a) 3D structure prediction of HSPB8 obtained with Alphafold3, highlighting the K141E single-point mutation in a licorice representation, and (b) a schematic diagram of HSPB8 domains. The diagram identifies the variable N-terminal (NTD) in blue and C-terminal (CTD) domain in red, as well as the conserved alpha-crystallin domain (ACD) in yellow. The purple box shows the conserved RLFDQxFG sequence. The green hexagon indicates the K141E mutation site.

We applied the new FRETpredict algorithm,[19] using a rotamer library approach, to estimate the accessible volume of FRET probes (Figure S1) and predict the mean dye distance for each simulated protein structure. FRETpredict places donor and acceptor dyes independently at specified sites, computes distances and relative orientations for all rotamer combinations, and evaluates nonbonded interaction energies within 1.0 nm. The statistical weights of each rotamer pair are combined to calculate the average FRET efficiency. Integrating FRETpredict with smFRET data provides a robust framework for correlating experimental and computational findings. By calculating FRET efficiencies and applying Bayesian maximum entropy (BME) reweighting, we confirmed that the K141E mutation impacts protein dynamics.

Our findings revealed different structural variability between HSPB8 WT and the K141E mutation, which promoted increased conformational variability, particularly under physiological ionic strength. Elevated ionic strength promotes more extended conformations, with the K141E mutant exhibiting increased conformational variability with respect to WT. This heightened structural flexibility may contribute to the observed destabilization from smFRET measurements, and functional impairments and pathogenicity associated with the K141E mutation, highlighting a potential connection between conformational dynamics and disease progression. The insights gained here can readily extend to other IDPs, including other chaperones associated with neuromuscular and muscular diseases, potentially advancing the understanding of the pathogenic effects of mutations.

Results and discussion

K141E pathogenic mutation increases conformational sensitivity to ionic strength

The K141E mutation lies in the very conserved alpha-crystallin domain of the protein (Figure 1). To explore the physiological implications of the K141E mutation, which involves a reversal of the residue charge state, it is essential to perform simulations and experiments across a range of ionic strengths, from an "ideally" low 0 mM to 165 mM, mimicking the physiological concentration.

The simulated data were collected and analyzed to study the effects of the K141E mutation in the HSPB8 protein under different ionic conditions. Five replicas of 1 μs all-atom MD simulations were performed for both the WT and the K141E mutant protein. For each variant, simulations were repeated at two ionic strengths, namely 0 mM and 165 mM.

The purpose of these simulations was to investigate the impact of the K141E mutation on the structure and

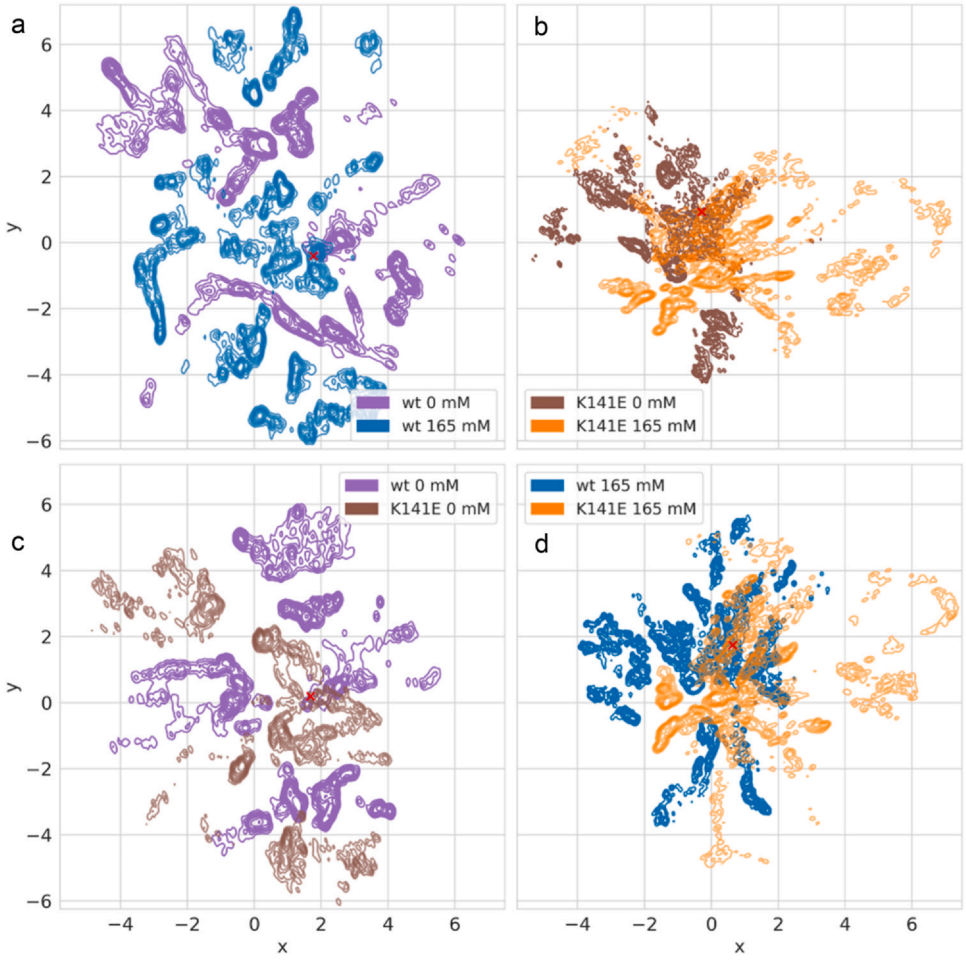

Fig. 2 Encodermap: 2D kernel density estimate (KDE) plots illustrating the molecular dynamics (MD) trajectories of HSPB8 variants under varying ionic strengths. Analysis was performed on *concatenated* trajectories, meaning that for each system, the five independent replica simulations (each 1 μs long) were combined into a single 5 μs trajectory. (a) The trajectories of the wild type (wt) at 0 mM (violet) and 165 mM (blue) ionic strength. (b) The K141E variant at the same ionic strengths, represented in brown (0 mM) and orange (165 mM). (c) The wt (violet) and K141E (light brown) at 0 mM, while (d) contrasts the two variants at 165 mM, with wt in blue and K141E in orange. The x-axis and y-axis represent coordinates assigned to the structures in the trajectories by the EncoderMap algorithm. Darker colors indicate regions of the plot that were more frequently visited during the trajectories, highlighting the dynamic behavior of the proteins under different ionic conditions. Red crosses correspond to the common MD starting structures.

dynamics of HSPB8, particularly under different ionic strengths. EncoderMap[20] was utilized to analyze and visualize the large dataset generated. In Figures 2 and S7, the data were processed using the EncoderMap 2D kernel density estimate (KDE), which reveals the structural dynamics of both the wild-type (WT) protein and the K141E variant of HSPB8 at ionic strengths of 0 mM and 165 mM. These plots, created from concatenated MD trajectories, utilize dimensionality reduction with EncoderMap,[20] a neural network-based algorithm for decreasing dimensionality. This approach provides a clear visualization of how each protein variant explores the conformational space under varying environmental conditions.

Thicker contours in the KDE plots correspond to the conformations most frequently sampled during the simulations. Qualitatively, we can see notable differences in how the WT and K141E variants cover the conformational space at different ionic strengths. For example, in Figure 2(a), the WT protein at 0 mM exhibits fewer, more concentrated conformational "islands" (as enclosed by rectangles), indicating a restricted range of conformations. In contrast, at 165 mM, the protein explores a broader and more diverse set of conformations. In Figure 2(b), the K141E variant shows distinct behavior, with wider or more scattered conformational islands at both ionic strengths, suggesting a more dynamic or flexible conformational behavior.

In Figure 2(c) and (d), the visual differences in the plots highlight how the mutation (K141E) and changes in ionic strength affect the structural dynamics of HSPB8, underscoring the mutation's potential impact on protein function under physiological and stress conditions.

In Figure 2(d), the comparison between WT and K141E at 165 mM ionic strength shows that the conformational rectangle area for K141E is larger than that for WT, while the covered area is smaller (as detailed in Table S1). This leads to a lower covered area density for K141E. The rectangle is defined by the minimum and maximum x and y coordinates of the nonempty bins in the 2D conformational space.

The rectangular area provides an estimate of the total conformational space accessible to the protein, while the covered area reflects the extent of that space that is actually sampled by the protein. If a protein has low conformational variability, it is likely to stay within a confined region of its conformational space, resulting in a well-sampled intermediate region between its most frequently visited structures. This is indicated by a high coverage area. Conversely, high conformational variability implies sampling of a wider array of structures, with less focus on any one intermediate region, thus resulting in a lower covered area.

These distinctions are visually noticeable: areas of low coverage tend to feature elongated, thin "conformational islands" on the plot, while areas of high coverage appear as dense, dark islands. The combination of rectangle and covered areas produces the "covered area density," which measures how extensively the 2D conformational space is sampled.

Overall, K141E consistently shows lower covered areas than WT at both 0 mM and 165 mM. Notably, the K141E rectangle area is 10% smaller than WT at 0 mM but 24% larger at 165 mM, underscoring its increased conformational flexibility at higher, more physiologically relevant ionic strength.

Considering the common area values and overall overlap in the EncoderMap plots (Figure S7), it appears that both WT and K141E undergo greater structural changes with the shift from 0 mM to 165 mM than when comparing the two variants at the same ionic strength. The common area values more than double when comparing the same variant across different ionic strengths, as opposed to comparing the two variants at identical ionic strengths, underscoring the influence of ionic conditions on structural dynamics. The EncoderMap plots (Figures 2 and 3(d)) structural simulation data (Figure 3) reveal no significant differences in conformational variability between the K141E mutant and WT HSPB8 under 0 mM conditions. Overall, the results suggest that WT and mutant populate similar structures, but the mutant has more frequent interconversions between structural states.

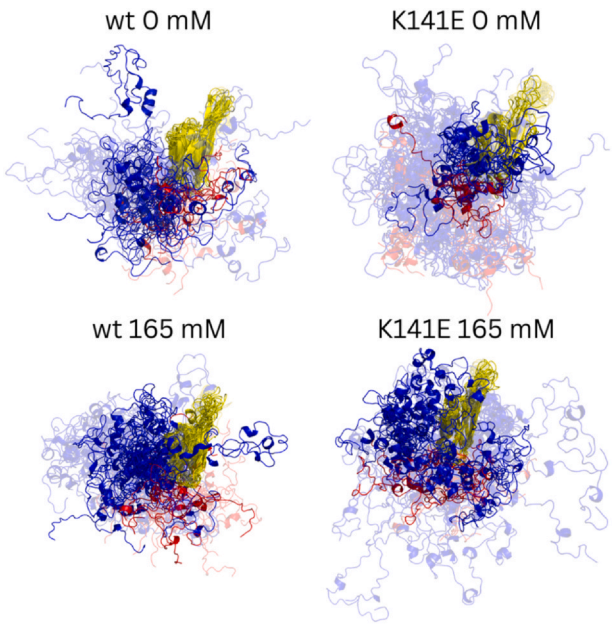

Fig. 3 Representative HSPB8 structures obtained from clustering the conformational islands in the EncoderMap for wild-type low ionic strenght (0 mM), K141E (0 mM), and high ionic strength wild type (165 mM) and K141E (165 mM), where the α-crystallin domains have been superimposed. The structures illustrate the main conformational states sampled at each condition, with the N-terminal in blue, the α-crystallin domain in yellow, and the C-terminal in red. Darker structures indicate regions corresponding to minima in the EncoderMap, highlighting the most stable conformations observed.

Further root-mean-square fluctuations analysis has revealed that the increased conformational variability of HSPB8 K141E at 0 mM, compared to the WT (Figure S8). The most pronounced difference in mobility is observed in the NTD. This increased mobility is further reflected in the larger conformational states of K141E illustrated in Figure 3.

Single-molecule FRET insights into HSPB8 structural dynamics

To test the computational predictions on HSPB8 at physiological ionic strengths, we turned to smFRET, which is a powerful technique to study the structure and dynamics of IDRs.[21] We took advantage of three native cysteine residues in the polypeptide sequence for labeling. We made three variants for both WT and mutant, each with one of the three cysteines (Cys10, Cys99, and Cys195) individually removed, and chemically modified the two remaining cysteines with the fluorophores Cy3B and CF660R. This allowed us to then

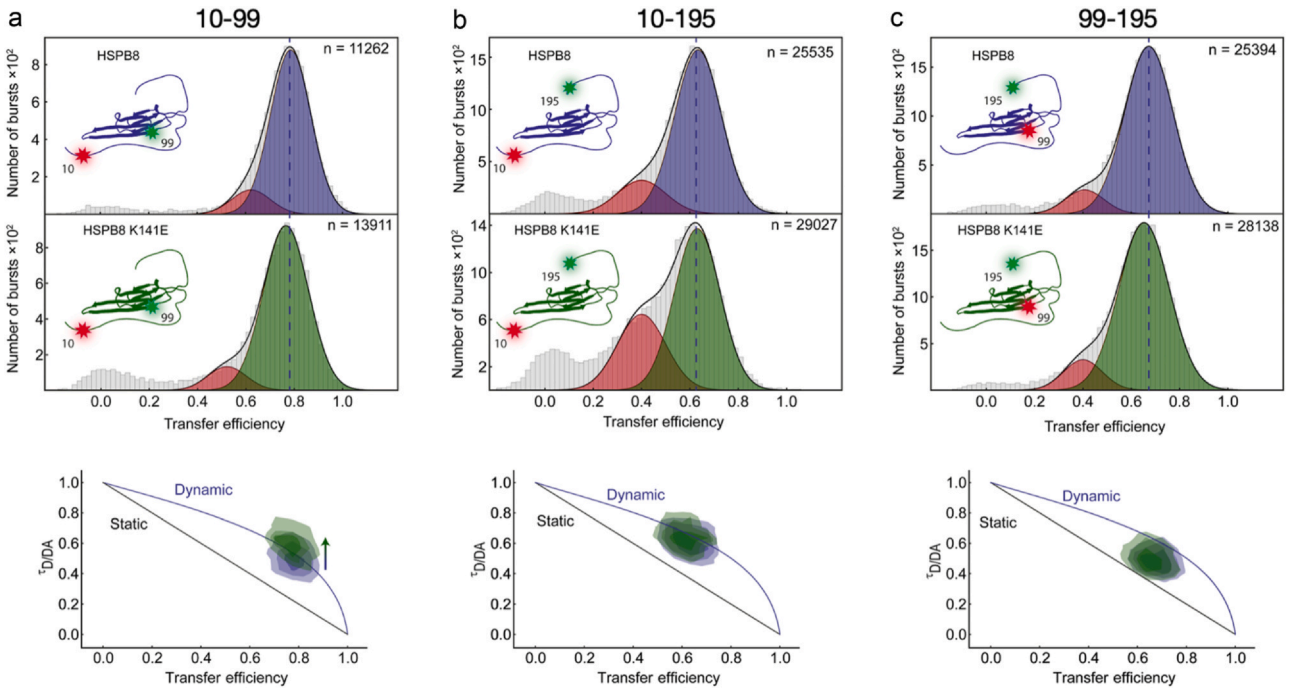

Fig. 4 Single-molecule FRET (smFRET) experimental data obtained at 165 mM. Förster resonance energy transfer efficiency distributions for HSPB8 labeled at (a) 10–99, (b) 10–195, and (c) 99–195. The wild type is shown in blue and the K141E mutant in green, with *E* values for each peak shown in the corresponding panel. The lower panels show the corresponding relative donor fluorescence lifetimes. The distributions illustrate the conformational dynamics and variations in Förster resonance energy transfer efficiency between the wild type and K141E mutant under the same experimental conditions. The number of molecules (*n*) is indicated for each histogram. Abbreviation used: HSPB8, heat shock protein B8.

probe three distinct intramolecular distances (between positions 10–99, 10–195, and 99–195) by measuring the mean transfer efficiency, *E*, for each variant (Figure 4(a)). The smFRET data, recorded on thousands of individual molecules for both the WT and K141E variant, highlight the correlation between FRET efficiencies and conformational changes upon single-point mutation. For both the WT and mutant, a major population is observed with an additional well-defined minor population at lower FRET. However, the relative sizes of the major and minor populations are different between the WT and mutant. The distinctive behavior of the HSPB8 K141E variant with respect to WT is especially evident in the smFRET data for the 10–195 pair, which essentially reflects the conformation of the entire protein. While the mean *E*-values for the WT and K141E are similar (0.38 and 0.40, respectively), the secondary peak frequency is significantly higher for K141E than for the WT. Conversely, we observe that the secondary peak *E*-values for the 99–195 dye pair (representing the ACD + CTD regions) are nearly identical for both variants (0.42 for WT and 0.40 for K141E), with similar frequencies, although K141E has a slightly higher ratio between major and minor populations (Figure 4). For the Cys 10–99 pair (representing the NTD region), the secondary peak frequencies at 165 mM are similar between WT and K141E, but the *E*-value of the peak decreases from ∼0.6 in the WT to ∼0.5 in the K141E mutant. This lower *E*-value for K141E suggests a larger distance between the fluorophores, indicating that the NTD of the mutated protein generally adopts more extended conformations than the WT, potentially expanding its 3D conformational space.

To probe rapid fluctuations in interdye distance and detect potential changes between wild type and variant, we analyzed relative fluorescence lifetimes of donor (Figures 4 and S9). Sub-μs dynamics can be detected in a plot of transfer efficiency (E) as a function of relative donor lifetimes ($\tau_{D/DA}$) as a deviation from the static line toward the dynamic line which describes the maximum of such dynamics in the case of a gaussian chain[22] As expected, a strong deviation is observed for the variants probing the full-length protein (10–195) and the disordered NTD (10–99), whereas the lifetimes of the variant probing mostly the crystallin domain are much closer to the static line. The K141E mutant shows very similar behavior but has notably increased dynamics in the NTD, which can be visualized as a shift upwards, away from the static line. This indicates that the mutation causes increased distance dynamics in the NTD on a sub-μs timescale. This timescale can be compared with the timescale of the simulations, which

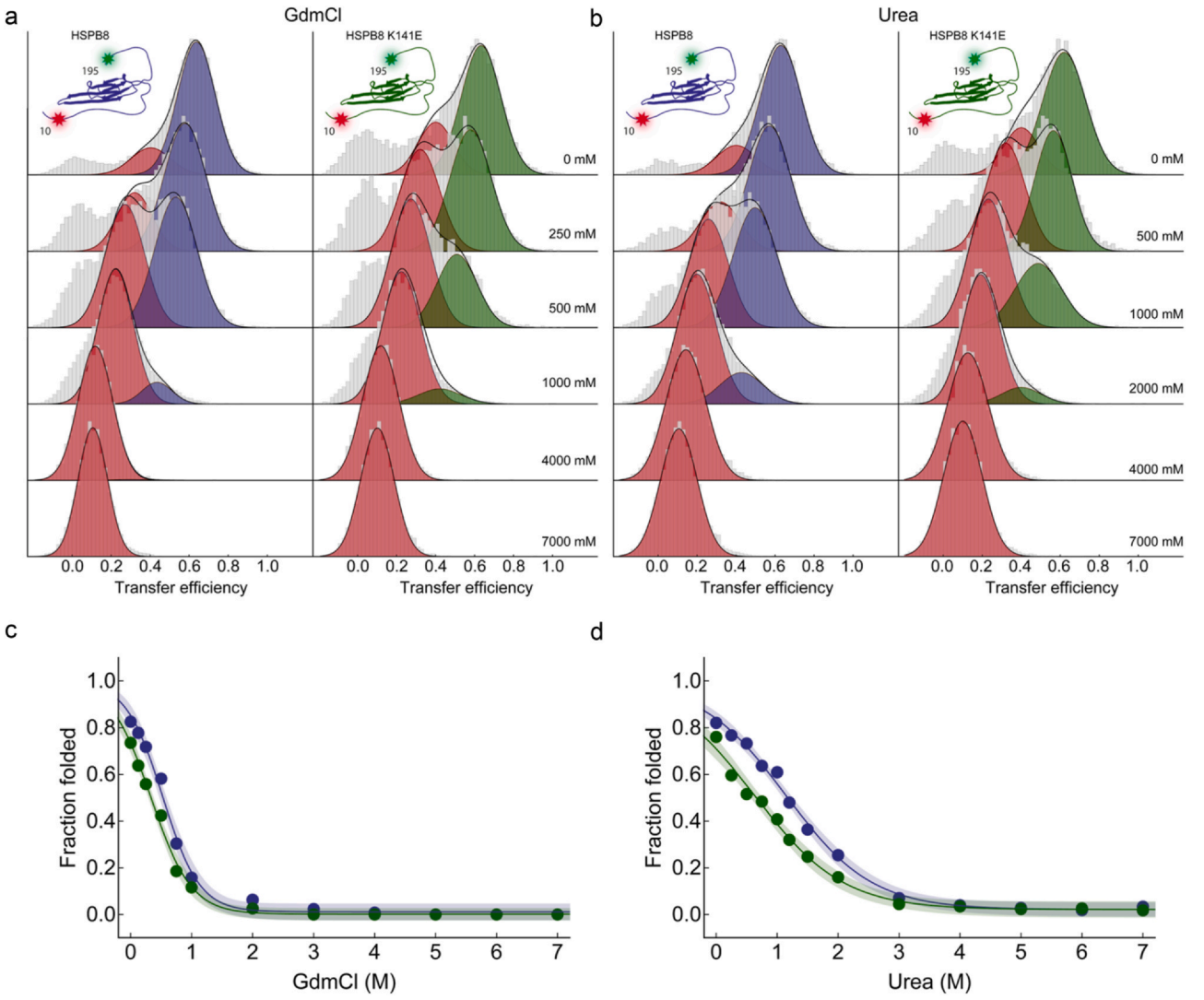

Fig. 5 Denaturation curves for HSPB8 wild type (WT) and K141E mutant labeled at Cys 10–195. (a) and (b) Transfer efficiency histograms at different (a) GdmCl or (b) urea concentrations for HSPB8 wt (blue/red) and K141E (green/red). (c) and (d) Fraction of the folded population as a function of (c) GdmCl or (d) urea. The solid lines are fits to a two-state unfolding model (see Materials and methods). Shaded regions represent 95% confidence intervals of the fits. Abbreviations used: GdmCl, guanidinium chloride; HSPB8, heat shock protein B8.

shows good agreement and a greater tendency for the mutant to populate more extended states within the simulation timescale of nanoseconds.

We wondered whether the minor peak in the Förster resonance energy transfer (FRET) efficiency histograms corresponded to the unfolded state of HSPB8. To probe conformational changes in more detail, we performed chemical denaturation experiments using both urea and guanidinium chloride (GdmCl) on the variant probing the full-length protein. In Figure 5, we have plotted denaturation curves for HSPB8 wild-type (WT) and K141E mutant labeled at Cys 10-195, while in Figure 6, we have plotted denaturation curves for HSPB8 WT and K141E mutant labeled at Cys 99-195. In urea, there is a more pronounced difference in the unfolding response with respect to GdmCl. We used the areas of the peaks in the transfer efficiency histograms to calculate the fraction of folded protein as a function of denaturant. By fitting the denaturation data to a two-state unfolding model, we could extract the free energy of unfolding, ΔG_u^{H2O}. The free energy difference due to the mutation is almost two-fold, reduced from ~1.0 kcal/mol to 0.5 kcal/mol (Table S2).

Importantly, the minor peak shows a smooth expansion as a function of denaturant for all labeling variants, whereas the major population resists denaturation at low denaturant concentrations for those labeling variants that probe the crystallin domain (Figure S10).

Finally, we measured the effect of denaturants on only the NTD (Figure 7). The main population shifts to

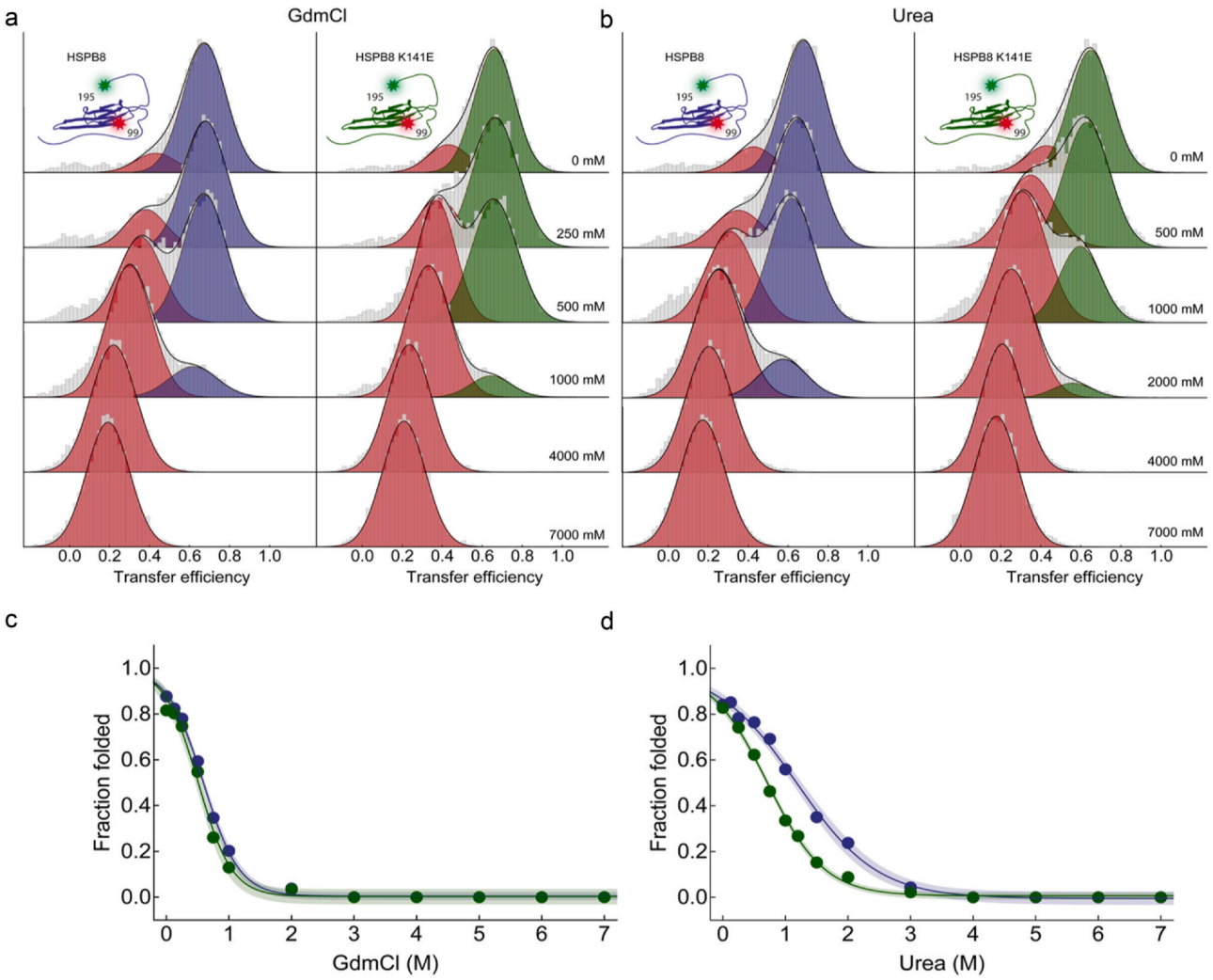

Fig. 6 Denaturation curves for HSPB8 wild type (WT) and K141E mutant labeled at Cys 99–195. (a) and (b) Transfer efficiency histograms at different (a) GdmCl or (b) urea concentrations for HSPB8 WT (blue/red) and K141E (green/red), (c) and (d) Fraction of the folded population as a function of (c) GdmCl or (d) urea. The solid lines are fits to a two-state unfolding model (see Materials and methods). Shaded regions represent 95% confidence intervals of the fits. Abbreviations used: GdmCl, guanidinium chloride; HSPB8, heat shock protein B8.

lower FRET efficiency at low denaturant concentrations, indicating expansion of the disordered region, and it quickly merges with the minor unfolded state population. We thus converted the mean *E*-values to root-mean-square distance, which allows us to estimate an apparent radius of gyration (R_g) using a self-avoiding walk (SAW) polymer model (see Materials and methods) (Figure 7).

Interestingly, this region resists changes in R_g at low urea concentrations, indicating that even though this region is dynamic (as observed from the fluorescence lifetimes in Figure 4) it is held in a compact state presumably by interactions with other parts of the protein.

Given the overall charge of the disordered regions and the reversal of charge state due to the mutation, we examined the influence of ionic strength on protein structure by conducting FRET experiments at different concentrations of KCl (Figure 8). These experiments focused on the effect of ionic strength on the ratio of the secondary to the primary peak in FRET efficiency distributions, based on the concept that mutations affecting the protein's charge might also impact its overall dimensions and conformation.[23] Interestingly, the ratio of the secondary peak for the WT remains relatively stable as KCl concentration increases (Figure 8(a)). In contrast, the K141E mutant shows a significant shift in this ratio with varying KCl concentrations, with the most pronounced change occurring at physiological ionic strength, around 100–200 mM (Figure 8(c)). smFRET data across different KCl concentrations reveal a decrease in *E*-values for both the main and secondary peaks for both variants. Specifically, the WT main peak *E*-value decreases from 0.68 to

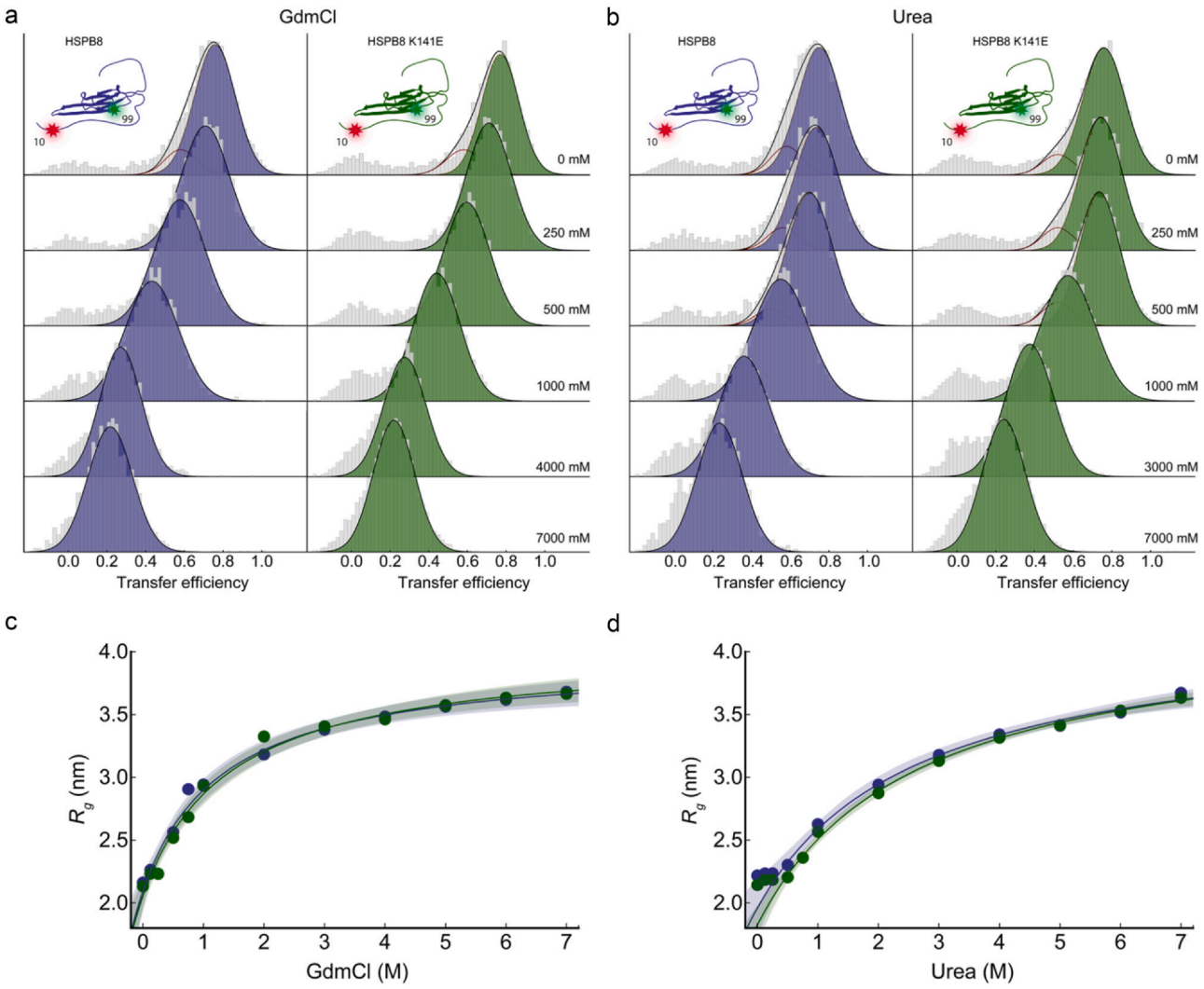

Fig. 7 Denaturation of HSPB8 wild type (WT) and K141E mutant labeled at Cys 10–99. (a) and (b) Transfer efficiency histograms at different (a) GdmCl or (b) urea concentrations for HSPB8 WT (blue/red) and K141E (green/red), (c) and (d) R_g as a function of (c) GdmCl or (d) urea. The solid lines are fits to a weak denaturant binding model (see Materials and methods). Shaded regions represent 95% confidence intervals of the fits. In (d), only data points after 0.5 M urea are used for the fit. Abbreviations used: GdmCl, guanidinium chloride; HSPB8, heat shock protein B8.

0.60, while for K141E, it decreases from 0.65 to 0.59. Similarly, the secondary peak *E*-values also drop, from 0.38 to 0.33 for WT, and from 0.41 to 0.34 for K141E. Moreover, the height of the secondary peak, corresponding to the unfolded state, increases with increasing ionic strength, particularly for K141E. This suggests that higher ionic strength destabilizes the folded state more significantly in K141E compared to the WT.

This finding aligns with simulation results that highlight the role of salt bridge network disruption in the K141E mutant.[5] In Figure S11, the salt bridge networks for both WT and K141E variant were analyzed at 0 mM and 165 mM ionic strength. In Figure S11, protein residues are depicted in a circular arrangement, connected by lines where salt bridges are present during the simulation. The line thickness represents the frequency of each salt bridge formation, with thicker lines indicating more frequent interactions.

This structural extension observed at higher ionic strength is likely driven by salt bridge rearrangements (Figure S11). Increased ion concentrations alter the electrostatic interactions among HSPB8′s charged residues, disrupting intraprotein salt bridges and facilitating the formation of salt bridges with surrounding ions. This phenomenon is not unique to HSPB8, as it has been observed in other IDPs.[24] The balance between intraprotein and protein–ion electrostatic interactions may explain the distinct conformational behaviors of the WT and K141E variants under varying ionic conditions.[25,26]

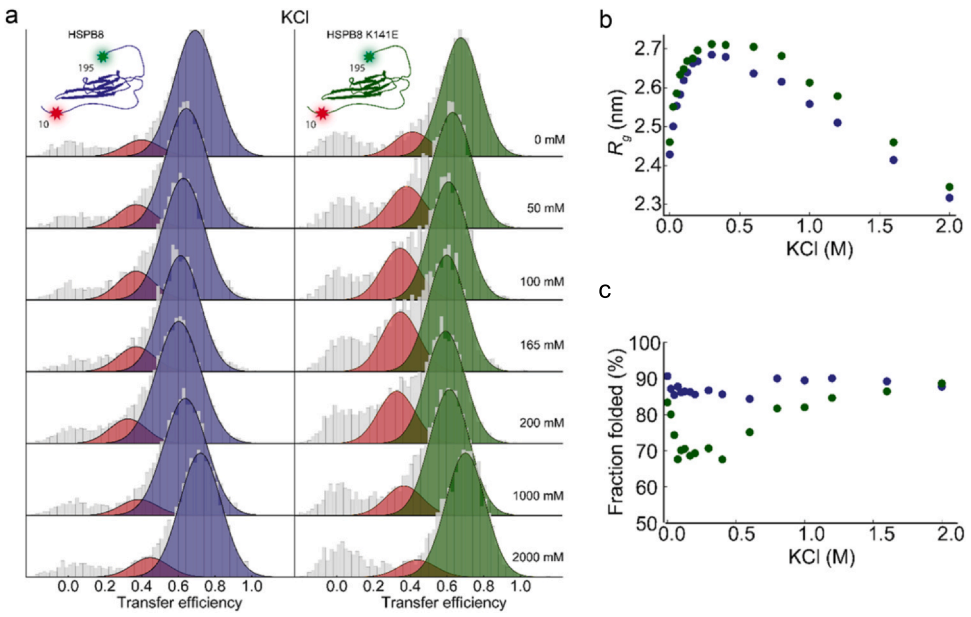

Fig. 8 KCl titration curves for HSPB8 WT and K141E mutant labeled at Cys 10–195. (a) Transfer efficiency histograms at different concentrations of KCl, for HSPB8 WT (left, blue) and the K141E mutant (right, green). (b) Apparent R_g as a function of KCl concentration for WT (blue) and K141E variant (green). (c) Fraction of folded protein as a function of KCl concentration for WT (blue) and K141E variant (green). Abbreviations used: GdmCl, guanidinium chloride; HSPB8, heat shock protein B8.

In contrast, the "roll-over" effect shows only slight differences between the WT and the K141E mutant, as both demonstrate a similar pattern in the variation of the R_g at each ionic strength (Figure 8(b)). This observation aligns with the R_g values obtained from atomistic simulations (Figure S3).

FRET efficiency and BME reweighting

Large-scale all-atom explicit-solvent molecular simulations of both HSPB8 and the K141E mutant were validated by comparing the simulated data to experimental results (a schematic overview of the integrative workflow is depicted in Figure S12). According to a study by Aznauryan *et al.*[27] the end-to-end distance reconfiguration times for IDPs and unfolded proteins typically range between 50 and 150 nanoseconds. In our MD simulations, we assessed the convergence of sampling by analyzing the autocorrelation times (Figure S4), which indicate whether the simulation has sufficiently explored the protein's conformational space. Since the autocorrelation times are less than one-tenth of the total simulation duration, this suggests that the simulations have likely achieved convergence, meaning they have captured the key conformational states necessary for analysis.

To improve agreement between the simulation and experimental results, BME reweighting was applied to the simulated data, reported in Figure 9. This method refines the simulated FRET efficiencies, a critical metric for measuring intramolecular distances and comparing different protein conformations.

When analyzing MD simulation data, FRET efficiencies can be reweighted using the BME method to better align with experimental FRET measurements.[28] This reweighting process assigns varying weights to simulated conformations, ensuring that the average FRET efficiency of the weighted sample matches the experimental values. Reweighting improves the agreement between simulations and experimental data, even if some difference arises because enhanced sampling MD simulations typically capture shorter timescales compared to those accessible in experiments.

Starting with our microsecond-scale all-atom explicit solvent MD simulations, we applied reweighting to examine and compare additional structural properties of the proteins, such as salt bridges (Figure S6), EncoderMap plots (Figure 2), and R_g (Figure S3). This method allows for a more detailed understanding of the conformational landscapes of HSPB8 and the K141E variant.

Figure 9 shows the FRET efficiency values predicted by FRETpredict using rotamer libraries at the physiological ionic strength of 165 mM for HSPB8 WT (top panel) and K141E mutant (bottom panel) labeled at Cys 10–99, 10–195, and 99–195. In Figure S13, we also report a comparison with the values obtained at 0 mM. Calculated values of FRET efficiencies compared with experimental values indicate convergence. Notably, upon BME reweighting, the alignment between simulated and experimental FRET efficiencies for all the pairs of

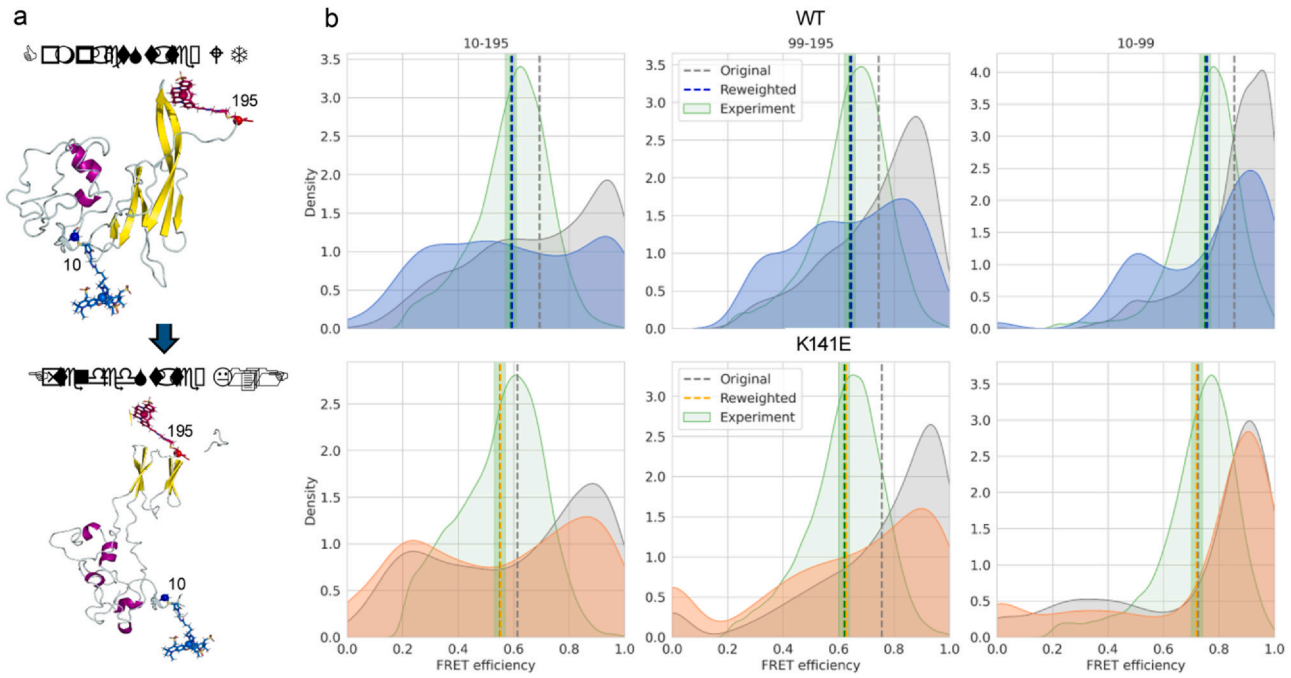

Fig. 9 (a) HSPB8 wild-type (WT) protein resides predominantly in a stable, compact state, while the K141E variant exhibits an increased population of extended conformations, as indicated by MD simulations and smFRET experiments. These extended conformations are associated with greater instability, promoting unfolding and aggregation, which may contribute to functional inactivation. (b) Comparison of FRET efficiency distributions at 165 mM salt concentration, for HSPB8 WT (top panel) and K141E mutant (bottom panel) labeled at Cys 10–99, 10–195, and 99–195, respectively. The figure shows the experimental FRET distribution alongside the simulated data before and after Bayesian maximum entropy (BME) reweighting (blue dashed line). The reweighting procedure (red dashed line) adjusts the simulated FRET efficiencies to better match experimental values (light green). The plots demonstrate the differences between the original simulated distributions, reweighted distributions, and the experimental data for both the WT and K141E variant. Abbreviation used: HSPB8, heat shock protein B8.

labeled residues of HSPB8 WT and K141E mutant is refined, providing better agreement.

Conclusion

This study elucidates the impact of ionic strength on the conformational dynamics of HSPB8 WT and the disease-linked K141E mutation and demonstrates that variations in ionic strength significantly influence the conformational ensemble of the protein, due to the structural characteristics of its disordered regions.

Utilizing 2D KDE plots and EncoderMap analysis, we observed distinct differences in the conformational exploration of the WT and K141E variant under varying ionic conditions. While the WT exhibited concentrated conformational islands at low ionic strength, indicating limited structural diversity, the K141E mutant showed broader and more scattered islands, suggesting a heightened conformational flexibility, particularly at physiological salt concentrations. Notably, the K141E variant demonstrated a marked shift in structural sampling between the two ionic strengths, in contrast to the WT, which displayed more stability across conditions.

smFRET experiments further supported these findings, revealing a correlation between FRET efficiencies and the conformational dynamics of both variants. Fluorescence lifetime analysis indicated increased interdye distance fluctuations on the sub-μs timescale. The elevated secondary peak frequency in the K141E variant at physiological ionic strength suggests a destabilization of the folded state, which may be linked to its propensity to populate extended conformations. It will be important to resolve whether these structural and dynamical changes correlate with the functional implications due to the mutation. Additionally, the presence of distinct differences in computed salt bridge networks between WT and K141E further underscores the role of ionic interactions in modulating conformational landscapes.

Our findings indicate that the K141E mutation in WT HSPB8 enhances conformational variability. This aligns with experimental data from flies, cellular studies, and *in vitro* assays, which demonstrate reduced chaperone and antiaggregation activity associated with this mutation.[6,8,11,18] Importantly, this study highlights parallels with other disease-linked chaperones that harbor IDRs and their associated mutations, emphasizing the

broader impact of such alterations on protein function and pathogenicity. By shedding light on the structural dynamics of chaperones like HSPB8, our work provides a framework for understanding how mutations influence chaperone conformational states and their functional consequences in cellular physiology, and it provides a possible link to mutations in folded regions having global and long-range effects on IDRs, which may result in global destabilization.

Materials and methods

smFRET experiments

Mutagenesis, expression, and purification

A pGEX-4T-GST vector containing the gene for human HSPB8 was used to generate a vector library of three dual-Cys variants of both HSPB8 and HSPB8-K141E. Mutants were made with the QuikChange Lightning kit from Agilent using primers from Integrated DNA Technologies and verified by Sanger sequencing (Azenta). All variants were expressed in BL21 competent *Escherichia coli* cells (NEB) and grown in LB media (0.1 mg/mL ampicillin). Expression was induced at OD_{600} 0.5–0.7 with 0.4 mM Isopropyl β-d-1-thiogalactopyranoside (IPTG) and cells grown for 4 h after induction at 30 °C. Cells were harvested by centrifugation at 4500g and 4 °C and kept at −20 °C until further use. Cell pellets were lysed by resuspending the pellet in a lysis-PBS (phosphate buffered saline) buffer (11.8 mM phosphate, 137 mM NaCl, 2.7 mM KCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 1 μg/mL DNase, 1 mg/mL lysozyme, 1 cOmplete inhibitor tablet and 1 mM PMSF, pH 7.4). The pellet was resuspended by gentle shaking at 4 °C for 2 h prior to sonication. Sonication was performed on ice with 5 s bursts and 15 s cooling time for a total of 5 min burst time.

The cell lysate was centrifuged at 40,000g for 45 min at 4 °C to remove cell debris, and the supernatant was collected. The supernatant was passed through a 0.45 μm syringe filter (Whatman) prior to being loaded onto a 5 mL GSTrap column (Cytiva) equilibrated with a PBS buffer containing 1 mM dithiothreitol (DTT) and 1 mM EDTA (wash buffer). After loading the sample, the column was washed with 10 column volumes (CV) of wash buffer, and the sample was eluted with a PBS buffer containing 1 mM DTT, 1 mM EDTA, and 20 mM reduced glutathione. The GST-tag of the HSPB8 construct was removed by adding 200 units of Human Rhinovirus 3C protease (HRV 3C) to the eluted sample while dialyzing overnight at 4 °C against PBS buffer containing 1 DTT and 1 mM EDTA. The GST-tag was removed from the solution by heating the sample to 72 °C for 10 min, resulting in the precipitation of the tag, followed by a centrifugation step at room temperature (RT) at 5000 g, keeping the supernatant.

The sample was then concentrated using Amicon Ultracentrifugal filters (Merck), reduced with 100 mM DTT and purified by reversed-phase high-performance liquid chromatography (RP-HPLC) using a ZORBAX 300SB-C3 column (Agilent) with flow rate of 2.5 mL/min starting at 95% solvent A (99.9% H_20, 0.1% trifluoroacetic acid (Sigma) and 5% solvent B (99.9% acetonitrile, 0.1% trifluoroacetic acid) and going to 100% solvent B over 95 min. Protein purity was confirmed by SDS-PAGE, and samples were lyophilized and stored at −20 °C.

Protein labeling

Prior to maleimide–thiol conjugation, lyophilized protein samples were resuspended in a degassed 100 mM potassium phosphate labeling buffer at pH 7.0 and labeled overnight at 4 °C using Cy3B maleimide (donor) (Cytiva) (0.85:1 dye to protein ratio) dissolved in anhydrous Dimethyl sulfoxide. The fluorophores were covalently attached to cysteine residues introduced by site-directed mutagenesis to enable site-specific labeling. The reaction was quenched using 100 mM DTT, and RP-HPLC was then used to remove unreacted dye and separate unlabeled and double donor-labeled proteins, using the same protocol as in the purification process. The protein aliquots were lyophilized and resuspended in the labeling buffer and labeled overnight at 4 °C using CF660R maleimide (acceptor) (Sigma) dissolved in anhydrous Dimethyl sulfoxide. The reaction was quenched using 100 mM DTT, and RP-HPLC was then used to remove unreacted dye and separate donor-donor doubly labeled and acceptor-acceptor doubly labeled proteins. Aliquots of labeled proteins were lyophilized, resuspended in 8 M GdmCl, frozen in liquid N_2, and stored at −80 °C.

Single-molecule spectroscopy

Single-molecule fluorescence experiments were conducted at RT using a MicroTime 200 (PicoQuant) connected to an Olympus IX73 inverted microscope. The donor dye was excited using a 520 nm diode laser (LDH-D-C-520, PicoQuant) using pulsed interleaved excitation[29] with a 640 nm diode laser (LDH-D-C-640, PicoQuant) to alternate excitation of donor and acceptor dyes with a repetition rate of 40 MHz. The laser intensities were adjusted to 40 μW at

520 nm and 20 μW at 640 nm, measured with a handheld optical power and energy meter (PM100D, Thorlabs). Excitation and emission light were focused and collected using 60× water objective (UPLSAPO60XW, Olympus). Emitted fluorescence was focused through a 100 μm pinhole before being separated first by polarization and then by donor (582/64 BrightLine HC, Semrock) and acceptor (690/70 H Bandpass, AHF) emission wavelengths, into four detection channels. Detection of photons took place using single-photon avalanche diodes (SPCM-AQRG-TR, Excelitas Technologies). Arrival time of detected photons was recorded with a MultiHarp 150P time-correlated single-photon counting module (PicoQuant). Experiments were performed in μ-Slide sample chambers (Ibidi) at RT in a TE buffer (10 mM Tris, 0.1 mM EDTA, pH 7.4) with varying KCl concentrations. For photoprotection 143 mM 2-mercaptoethanol (Sigma) was added, along with 0.01% (v/v) Tween-20 (AppliChem) to reduce surface adhesion. In experiments using denaturants, the exact concentration of denaturant was determined from measurement of the solution refractive index.[30]

Transfer efficiency histograms were collected from 50 to 100 pM of freely diffusing double-labeled proteins. Data were analyzed using the Mathematica scripting package "Fretica" (https://schuler.bioc.uzh.ch/programs/) developed by Daniel Nettels and Ben Schuler. Fluorescence bursts were first identified by combining all detected photons with less than 100 μs interphoton times. Transfer efficiencies within each fluorescence burst were calculated according to

$$E = n'_A/(n'_A + n'_D)$$

where n'_A and n'_D are the number of acceptor and donor photons, respectively. The number of photons were corrected for background, direct acceptor excitation, channel crosstalk, differences in dye quantum yields and photon detection efficiencies.[31] The resulting bursts were then filtered to remove bursts where the acceptor bleaches during the transit of the molecule through the confocal volume,[32] which would otherwise cause a bias toward lower FRET. Fluorescence bursts with photon counts exceeding three times the mean signal, binned at 1 s, corresponding to oligomeric states or aggregates, were removed before data analysis. The labeling stoichiometry ratio (S) was determined according to

$$S = \frac{n_D^D + n_A^D}{n_D^D + n_A^D + n_A^A}$$

where $n_{D/A}^D$ is the number of detected donor or acceptor photons after donor excitation and n_A^A is the number of detected acceptor photons after acceptor excitation. To construct the final transfer efficiency histograms, bursts that had $S = 0.3 - 0.7$ were kept, filtering out bursts that originate from molecules that are without an active acceptor.

Mean FRET efficiencies $\langle E \rangle$ were extracted from histograms by fitting to an appropriate number of Gaussian distribution functions, corresponding to one or two populations.

Fluorescence anisotropy values were determined for fluorescently labeled variants using polarization-sensitive detection in the single-molecule instrument,[33] and values for all variants were between 0.00 and 0.16. Similar values (± 0.03) for anisotropy were obtained at 0 M and 7 M GdmCl, indicating sufficiently rapid orientational averaging of the fluorophores to justify the approximation $\kappa^2 \approx 2/3$ used in Förster theory.[34]

Fluorescence lifetime analysis

Fluorescence lifetimes were estimated as the mean detection time of donor photons $\langle t_D \rangle$ after their excitation pulse. In the case of a fixed distance between donor and acceptor, the ratio between the mean donor lifetime in the presence of acceptor $\langle \tau_{DA} \rangle / \tau_D$ is equal $1 - E$, whereas in a system that rapidly samples a broad distance distribution, this ratio can significantly deviate from $1 - E$. A rapidly fluctuating distance can be described by a Gaussian chain probability density function $P_{Gauss}(r)$[35] of the interdye distance r, given as

$$P_{Gauss}(r) = 4\pi r^2 \left(\frac{3}{2\pi \langle r^2 \rangle} \right)^{\frac{3}{2}} e^{\left(\frac{-3r^2}{2\langle r^2 \rangle} \right)}$$

where $\langle r^2 \rangle$ is the mean squared end-to-end distance of the region probed. The distribution of distances affects the average fluorescence lifetime $\langle \tau_{DA} \rangle$ according to

$$\frac{\langle \tau_{DA} \rangle}{\tau_D} = 1 - \langle E \rangle + \frac{\sigma^2}{1 - \langle E \rangle}$$

The variance σ^2 given by

$$\sigma^2 = \langle E^2 \rangle - \langle E \rangle^2 = \int_0^\infty dr \, [E(r) - \langle E \rangle]^2 P(r)$$

Thus, fluorescence lifetimes were plotted against their corresponding transfer efficiencies in two-dimensional scatter plots, where $\tau_{DA}/\tau_D = \langle t_D \rangle / \tau_D$ was calculated for each burst for an intrinsic donor lifetime τ_D and plotted along with the static distance line and the Gaussian chain probability density function (diagonal line and a curved line, respectively).

Analysis of smFRET denaturation experiments

Prior to measurements, samples were incubated at RT in TEK buffer (10 mM Tris, 165 mM KCl, 0.1 mM

EDTA, pH 7.4) containing varying concentrations of either urea or GdmCl for at least 30 min, allowing samples to reach equilibrium. For samples that showed apparent two-state unfolding, the ratios between the areas under the curve of the two different populations were used to calculate the fraction of folded protein. Data sets were normalized by using the absolute ratios between the two peaks. The resulting sigmodal curve was fitted with[36]

$$f = \frac{(a_1 + b_1[denaturant]) + (a_2 + b_2[denaturant])\left(\frac{\Delta G_u - m_{eq}[denaturant]}{RT}\right)}{1 + \left(\frac{\Delta G_u - m_{eq}[denaturant]}{RT}\right)}$$

where f is the fraction folded, a_1 (set to 1) and a_2 are the signals of the folded and the unfolded states, respectively; b1 and b2 are the slopes of the pretransition and post-transition lines both set to 0, respectively; ΔG_u is the free energy of unfolding; and m_{eq} is the equilibrium m-value.

Samples where unfolding did not result in state change (smooth, uniform unfolding), mean transfer efficiencies $\langle E \rangle$ were converted to root-mean-square end-to-end distances $R = \sqrt{\langle r^2 \rangle}$, numerically solving the transcendental equation:

$$\langle E \rangle = \int_0^\infty dr E(r) P_{SAW-\nu}(r).$$

Here, $P_{SAW-\nu}(r)$ denotes the distance probability density function of the SAW-ν,[35] given by

$$P_{SAW-\nu}(r) = A\frac{4\pi}{R}\left(\frac{r}{R}\right)^{2+(\gamma-1)/\nu} \exp\left(-\alpha\left(\frac{r}{R}\right)^{1/(1-\nu)}\right),$$

where ν and $\gamma(\approx 1.1615)$ are critical exponents. The constants A and α are determined by requiring $P_{SAW-\nu}(r)$ to be normalized and to satisfy $\langle r^2 \rangle = R^2$, respectively. The dependency on ν in $P_{SAW-\nu}(r)$ is removed by assuming that a scaling law $R = bN^\nu$ holds and by substituting $\nu = \left(\frac{R}{b}\right)/\ln(N)$ into the expression for $P_{SAW-\nu}(r)$, where $b \approx 0.55$ nm for proteins and N denotes the number of residues between the fluorescent groups. The radius of gyration R_g can then be approximated as

$$R_g \approx R\sqrt{\frac{\gamma(\gamma+1)}{2(\gamma+2\nu)(\gamma+2\nu+1)}}.$$

In all denaturant experiments, the Förster radius (R_0) of the Cy3B/CF6660R dye pair (6 nm) was corrected for changes in refractive index ($R_0(c_D)$) according to[26]

$$R_0^6(c_D) = R_0^6\left(\frac{n_0}{n(c_D)}\right)^4$$

where n_0 denotes the refractive index without denaturant, and $n(c_D)$ denotes the refractive index of the sample at denaturant concentration c_D. The resulting radius of gyration ($R_{g(c_D)}$) values were then fitted to a weak denaturant binding model with the form[37,38]

$$R_g(c_D) = R_{g,0} + \left(\frac{\Delta R_g \quad K_a \quad c_D}{1 + K_a \quad c_D}\right)$$

where c_D is the denaturant concentration, K_a is the association constant of denaturants to the polypeptide chain, using a simple binding model that assumes identical independent binding sites, ΔR_g is the total change of R_g upon saturation of denaturant, and $R_{g,0}$ is the R_g value at zero denaturant.

MD simulations

Five replicas of 1 μs all-atom explicit solvent MD simulations of wt and K141E HSPB8 at 0 mM and 165 mM NaCl concentration were performed in explicit water solvent with Amber99sb-disp + TIP4PD as force field and water model,[39,40] using GROMACS 2022.3 as simulation software[41] (see Table 1). To ensure independence, each of the five replicas for every system was initiated using a different random seed during the velocity assignment step at the beginning of the MD simulations.

Although the experimental ionic strength is due to KCl, NaCl was used in the simulations to reproduce

Table 1
Summary of the simulated systems, including the relative ionic strengths (IOS), number of replicas per system, simulation length per replica, and total simulation time per system.

MD	Replica 1	Replica 2	Replica 3	Replica 4	Replica 5	Total time
IOS 0 mM						
WT	1 μs	1 μs	1 μs	1 μs	1 μs	5 μs
K141E	1 μs	1 μs	1 μs	1 μs	1 μs	5 μs
IOS 165 mM						
WT	1 μs	1 μs	1 μs	1 μs	1 μs	5 μs
K141E	1 μs	1 μs	1 μs	1 μs	1 μs	5 μs

ionic strength, a common and widely accepted practice in MD studies due to the similar ionic properties of Na^+ and K^+ in this context. The starting structures used for both ionic strengths correspond to the I-TASSER homology models.[11] These models were selected as the initial structure for MD simulations because the AlphaFold3 prediction, although generally considered reliable and comparable to other homology modeling tools, was found to be inconsistent with the low intrinsic disorder predicted for HSPB8 by the IUPRED3 web server. Specifically, AlphaFold3 overestimated the presence of IDRs, whereas the I-TASSER models showed a much better agreement with the IUPRED3 prediction. Notably, the structure of the ordered ACD was similarly well predicted by both I-TASSER and AlphaFold3, but the overall consistency with experimentally informed disorder profiles favored the use of I-TASSER-derived models. All MD simulations were carried out using periodic boundary conditions in cubic boxes centered on the protein center of mass, with side lengths equal to the protein diameter plus 1.1 nm, to avoid protein self-interactions during the simulation. For the 0 mM simulations, counterions were added after solvation to obtain neutral systems, considering a neutral protonation state for Histidine residues. 6 Na+ ions were added for the wt and 8 for the K141E mutant. 64 or 66 Na+ and 57 Cl-ions were added to the simulation box for simulations at experimental ionic strengths to obtain 165 mM concentration for the wt and K141E variants, respectively. Initially, the systems were minimized using the steepest descent algorithm to remove van der Waals contacts of high potential energy, with the maximum force threshold value set at 1000 kJ/mol/nm. Minimization was followed by a 100 ps relaxation of the solvent around the position-restrained protein and a 100 ps NPT equilibration with isotropic Berendsen pressure coupling at 1 bar.[42] The temperature was kept at 298.15 K using a velocity rescaling thermostat.[43] The 1 μs full MD simulations of the systems were performed using the leap-frog algorithm with a 2 fs time step, the Verlet cutoff scheme for Van der Waals interactions, and the Particle Mesh Ewald method for the treatment of electrostatic interactions with 1.0 nm cutoff was adopted. The temperature coupling method used was velocity rescale, with a time coupling constant of 0.1 ps. Covalent bonds involving hydrogen atoms were constrained using the LINCS algorithm. Positions and coordinates were saved every 20 ps. The total number of frames was 10,001 for each replica simulation. Standard structural analysis on each MD trajectories as root-mean-square deviation (Figure S2) was performed. The radius of gyration (Rg) (Figure S3) was performed using GROMACS. Trajectories and structures were visualized using VMD[44] and PyMol.[45] KDE plots were realized from custom Python scripts using the seaborn Python package.[46] Throughout the work, we employed Python version 3.8.10.

The root-mean-square deviation autocorrelation curves, Figure S4, for each MD simulation were computed with the Python package statsmodels.[47] Autocorrelation times were extrapolated from the intersection between the autocorrelation curve and the shaded area representing 95% confidence in Figure S4 and plotted as vertical dashed lines. The resulting autocorrelation times vary between 37 ns and 84 ns.

FRETpredict calculations

We used the recently developed FRETpredict[19] to calculate the FRET efficiency distributions from the fluorophore-free HSPB8 trajectories. This Python-based software uses a rotamer library approach to describe the FRET probes covalently bound to the protein. The software is available for download from GitHub.[48]

We used the pre-calculated rotamer libraries of Cy3b and CF660R from Klose *et al.*,[49] aligning them on the backbone of the labeled Cys residue pairs (10–195, 99–195, 10–99) on HSPB8 and the mutant. The Forster radius was fixed at the experimental value of 6.0 nm (according to ensemble spectroscopic measurements), by setting in FRETpredict fixed_R0 = True and r0 = 6.0.

To address the inherent inversion symmetry during the synthesis of protein-fluorophore complexes, specifically the inability to control whether Cy3b and CF660R bind to Cys10 and Cys195 or *vice versa*, FRET efficiencies were calculated for both possible binding configurations. Final FRET values and distributions were derived from the combined data. An example of the Python code used to compute FRETpredict efficiencies is reported in Figure S5.

The experimentally observed behavior of freely rotating dyes where the rotational correlation time is significantly faster than both the FRET rate constant and the diffusion rate constant,[50] are expressed by the dynamic relationship $k_R >> k_{FT} >> k_D$ where k_{FT} is the FRET rate constant, k_R is the rotational correlation time of the fluorophore, and k_D is the diffusion rate constant. This "dynamic relationship" ensures that the rotational motion of the fluorophores occurs much faster than the FRET energy transfer process, which itself is faster than molecular diffusion. This means that, during the timescale of FRET, the dyes are able to freely and rapidly rotate, effectively sampling all possible orientations. As a result, the orientation factor averages out to its dynamic isotropic value, typically 2/3. This dynamic averaging justifies the use of Förster theory, allowing FRET efficiency to depend primarily on

the distance between donor and acceptor, rather than their specific orientations.[6]

Thus, we have considered the following formula for per-frame $\langle E \rangle_{Dynamic}$: $\langle E \rangle_{Dynamic} = \left\langle \frac{R_0^6}{R_0^6 + R_i^6} \right\rangle$, with $R_i = |R_{D(i)} - R_{A(i)}|$ where the Donor and Acceptor dye positions given by the vectors R_D and R_A.

Protein conformations that exhibit steric clashes were excluded in all averaging regimes, while the remaining conformations were treated with equal weighting.

With FRETpredict, it is possible to calculate per-frame weights based on dye–protein interactions through the reweight function.

The reweight function assigns a continuous weight to each frame in the MD trajectory, based on how favorable the dye–protein interactions are within that frame. These weights reflect the likelihood that a given dye conformation avoids steric clashes or unfavorable contacts with the protein. Frames where the dye is in close proximity to the protein surface, potentially resulting in steric hindrance, receive lower weights, while conformations where the dye can freely sample space receive higher weights. This is implemented using a Boltzmann-like scheme, where interaction energies are translated into probabilities. All frames remain part of the analysis, but their contributions to the final FRET efficiency or distance distributions are scaled accordingly. In systems like HSPB8, where the protein adopts highly extended conformations and the dyes do not come into close contact with the protein, most frames are similarly favorable, and thus reweighting has a negligible effect. For this reason, the weighted and unweighted FRET distributions are nearly identical, and we decided not to apply the reweight function in this case.

To compare the FRET efficiency distributions from FRETpredict with the experimental results, we excluded values of FRET efficiency below 0.2 (indicating donor-only signal within one standard deviation σ of the Gaussian peak fitting the experimental data and molecules lacking or having inactive acceptor molecules) and above 1.0 (indicating proteins with two acceptor molecules), as they correspond to experimental artifacts. The experimental mean FRET efficiency values were calculated by averaging the resulting distributions.

Reweighting calculations with Bayesian/MaxEnt approach (BME)

Simulated ensembles often do not accurately reflect true biological ensembles due to differences in experimental and simulated conditions, the complex nature of conformational space in IDP, and limitations in the energy functions used in MD simulations. To enhance the fidelity of simulated ensembles, it is beneficial to refine them based on experimental data.[51]

The BME approach uses experimental results to adjust the simulations *a posteriori,* ensuring that (1) the new calculated averages closely match experimental values, accounting for uncertainty, and (2) it maximizes the relative Shannon entropy compared to the original simulation ensemble.[35]

In more detail, if the simulation model is a reasonably good approximation and only a small shift in the distribution of configurations is needed, the observed conformations x_i are reweighted each by a corresponding weight w_i to give a reweighted efficiency $\langle \epsilon \rangle_{rw}(\{w_i\}) = \frac{\sum_i w_i \epsilon_i(x_i)}{\sum_i w_i}$. In this expression, $\epsilon(x_i)$ is the FRET efficiency computed for a single conformation x_i. This can be optimized to match the experiment by minimizing the χ^2 function.

$\chi^2(\{w_i\}) = \sum_{k=1}^{N_{obs}} \frac{(\langle \epsilon \rangle_{rw}(k,\{w_i\}) - \langle \epsilon \rangle_{\exp}(k))^2}{\delta \epsilon^2(k)}$, with the mean efficiency and experimental uncertainty of observation k given by $\langle \epsilon \rangle_{\exp}(k)$ and $\delta\epsilon^2(k)$, respectively.

To ensure that the new set of weights deviates minimally from the original uniform distribution, an additional term is incorporated into the optimization function[52]

$$G(\{w_i\}) = \chi^2(\{w_i\}) - \theta S(\{w_i\})$$

where the Shannon entropy of the weight set measures the deviation from uniform weights:

$$S(\{w_i\}) = -\sum_i w_i \ln\,(w_i)$$

The factor θ controls the relative importance of the penalty (i.e., regularization) term. It is chosen to be as large as possible, thereby maintaining the weights as uniform as possible without causing a significant χ^2 increase.[35] A practical approach to balance this trade-off between θ and χ^2 is to identify the "elbow" of the curve: this involves testing different θ values, starting with a high number, until a further decrease in θ does lead to a significant reduction in the associated χ^2.[53]

To compute BME weights, we used the Python library BME, available for download on the github.[54] The BME Python library refines molecular simulation ensembles by incorporating experimental data. Given a simulated conformational ensemble, which do not perfectly match experimental measurements. BME adjusts the statistical weights of each frame in the ensemble using a Bayesian/maximum entropy approach, maximizing the relative Shannon entropy, so the reweighted ensemble better fits the experimental data while remaining as close as possible to the original distribution. The output is a new set of weights that can be used to compute ensemble-averaged properties in better agreement with experiments.

In our work, the uncertainty associated with each experimental mean FRET efficiency value ($\delta\epsilon^2(k)$) for HSPB8 variants was 0.02.

For each protein variant and ionic strength, the average FRET efficiencies for the pairs (10–195, 99–195, 10–99) were used as experimental restraints for BME reweighting. The mean FRET efficiency for the MD distributions was determined using linear averaging.

The newly computed optimized weights were applied to reweight the computational distributions $E_{Dynamic}$, including the radius of gyration distributions, the EncoderMap plots, and the salt bridge frequencies. Detailed analysis of each aspect are provided in the following sections.

Radius of gyration

In addition to obtaining FRET efficiency distributions, we would often like to know the radius of gyration: $R_g = \sqrt{\frac{1}{2}\langle r_{ij}^2(m)\rangle_{i,j,m}}$, here defined in terms of the average distances $r_{ij}(m)$ over all atom pairs i, j and all conformations m. From the SAW-ν model, we have that $R_g = RMS\sqrt{\frac{\gamma(\gamma+1)}{2(\gamma+2\nu)(\gamma+2\nu+1)}}$, with $RMS = \sqrt{\langle(\langle R_{DA}\rangle_E)^2\rangle}$, $\gamma = 1.1615$, and $b = 0.55nm$ [25]. The SAW-ν model describes the size of a flexible polymer, like an IDP, accounting for excluded volume effects. It relates the radius of gyration to the root-mean-square distance between atom pairs. RMS is the root mean square of the average inter-residue distances over all conformations, γ is a geometric constant, ν is the Flory scaling exponent, and b is the effective segment length, used when estimating RMS distances.

Here, $\langle R_{DA}\rangle_E$ is the FRET-averaged distance between the dyes. It is determined from time-averaged fluorescence intensity measurements on the single-molecule or ensemble level $\langle R_{DA}\rangle_E = R_0\,(E^{-1}-1)^{-1/6}$, where R_0 is 6.0 nm for our chromophore pair.[55] The E values for the experimental $\langle R_{DA}\rangle_E$ distributions were taken from the experimental FRET efficiency data, while the computational $\langle R_{DA}\rangle_E$ values were obtained from the FRETpredict $E_{Dynamic}$ distributions, with the correct BME weights applied.

EncoderMap

EncoderMap[56] is a dimensionality reduction algorithm that combines multidimensional scaling with a variational autoencoder, extracting and optimizing essential features from a large trajectory dataset and representing them with minimal loss of information. The result is a 2D map that provides global information about the conformational ensembles visited during a trajectory.[5] The point density on the map is related to the free energy of the corresponding conformations. The EncoderMap code was downloaded from github.[57]

In this work, the EncoderMap training set comprised the five 1 μs replicas for wt and K141E MD trajectories at 0 mM and 165 mM NaCl concentrations, 10,001 frames each, setting the initial frame of wt at 0 mM as the reference structure for the pairwise distance calculations. The pairwise distances were calculated for every 500 input frames. The Sketchmap Cartesian sigmoid parameters were selected according to the Sketch-map literature[58] and are reported in Table S1. The total number of training steps was set to 50,000, with the first 45,000 conducted without the Cα cost, and the final 5000 steps incorporating the Cα cost. The architecture of the EncoderMap neural network followed the same parametrization used in a recent paper.[5] The EncoderMap 2D plot in Figure 2 has been realized with the *kdeplot* function of the seaborn Python package,[46] considering the per-frame weights obtained with BME. In Figure 2, the x-axis and y-axis correspond to abstract coordinates in a reduced-dimensional space learned from the input structural data. These variables, X and Y, do not represent specific physical quantities but instead reflect major collective differences in conformations across the simulation trajectories. Similar structures are placed close together, while dissimilar ones are positioned further apart. The area covered by each trajectory plotted in the 2D map is determined by counting the number of nonempty bins in a 250×250 bin 2D histogram generated using the histogram2d function of the NumPy Python package. The density of the covered area is calculated as the ratio of the covered area to the area of the smallest rectangle that encompasses the trajectory, where the rectangle's dimensions correspond to the extreme x and y values of the nonempty bins.

Salt bridges

A salt bridge refers to the electrostatic interaction between two nearby residues with opposite charges. In this study, a salt bridge was considered to be present between two oppositely charged residues (such as Asp, Glu, Arg, and Lys) if the distance between their oxygen and nitrogen atoms was less than 4.5 Å. The VMD Salt Bridges tool was used to compute these interactions, and the results are illustrated in Figure S6 using the Inkscape software.[59] The frequency of salt bridges between two protein residues is calculated as the percentage of the trajectory frames in which the salt bridge is observed, adjusted by the corresponding BME weight for each frame.

Associated content

The data supporting this article have been included as part of the ESI.

Funding and support GB acknowledges the funding by DFM.AD004.324. PRIN2020-2020LW7XWH and supported by the EU within the NextGenerationEU-MUR PNRR Extended Partnership initiative on Emerging Infectious Diseases (Project No. PE00000007, INF-ACT). P.O.H. acknowledges funding from the European Research Council (ERC StG 101040601-PIONEER).

CRediT authorship contribution statement **Ciro Cecconi:** Writing – review & editing, Supervision, Conceptualization. **Kristinn R. Óskarsson:** Writing – original draft, Investigation, Formal analysis. **Sveinn Bjarnason:** Writing – original draft, Methodology, Investigation. **Daniele Montepietra:** Writing – original draft, Methodology, Investigation. **Giorgia Brancolini:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Conceptualization. **Pétur O. Heidarsson:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Serena Carra:** Writing – review & editing, Conceptualization.

Data availability statement We have shared the link to the code.

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

Acknowledgments The authors would like to express their gratitude to Prof Kresten Lindorff-Larsen for his invaluable suggestions and for his support in the development of the FRETpredict code. G.B. and D.M. acknowledge the CINECA award under the ISCRA initiative, for the availability of high-performance computing resources and support. S.C. and C.C. acknowledge FAR2016.

Appendix A. Supplementary Data

Supplementary data associated with this article can be found online at doi:10.1016/j.cstres.2025.100086.

References

1. Ecroyd H, Bartelt-Kirbach B, Ben-Zvi A, et al. The beauty and complexity of the small heat shock proteins: a report on the proceedings of the fourth workshop on small heat shock proteins. *Cell Stress Chaperones.* 2023;28:621–629.
2. Carra S, Seguin SJ, Lambert H, Landry J. HspB8 chaperone activity toward poly(Q)-containing proteins depends on its association with Bag3, a stimulator of macroautophagy. *J Biol Chem.* 2008;283:1437–1444.
3. Ulbricht A, Eppler FJ, Tapia VE, et al. Cellular mechanotransduction relies on tension-induced and chaperone-assisted autophagy. *Curr Biol.* 2013;23:430–435.(o).
4. Carra S, Sivilotti M, Chavez Zobel AT, et al. HspB8, a small heat shock protein mutated in human neuromuscular disorders, has in vivo chaperone activity in cultured cells. *Hum Mol Genet.* 2005;14:1659–1669.
5. Klaips CL, Jayaraj GG, Hartl FU. Pathways of cellular proteostasis in aging and disease. *J Cell Biol.* 2018;217:51–63.
6. Choudhary D, Mediani L, Avellaneda MJ, et al. Chaperone mechanisms in protein folding and aggregation. *J Am Chem Soc.* 2023;145:15188–15196.
7. Haslbeck M, Weinkauf S, Buchner J. Small heat shock proteins: simplicity meets complexity. *J Biol Chem.* 2019;294:2121–2132.
8. Carra S, Boncoraglio A, Kanon B, et al. Identification of the Drosophila ortholog of HSPB8: implication of HSPB8 loss of function in protein folding diseases. *J Biol Chem.* 2010;285:37811–37822.
9. Boczek EE, Fürsch J, Niedermeier ML, et al. HspB8 prevents aberrant phase transitions of FUS by chaperoning its folded RNA-binding domain. *Elife.* 2021;10:e69377.
10. Zhou X, Wang M, Sun M, Yao N. HSPB8 attenuates lipopolysaccharide-mediated acute lung injury in A549 cells by activating mitophagy. *Mol Med Rep.* 2024;30:171.
11. Montepietra D, Cecconi C, Brancolini G. Combining enhanced sampling and deep learning dimensionality reduction for the study of the heat shock protein B8 and its pathological mutant K141E. *RSC Adv.* 2022;12:31996–32011.
12. Hipp MS, Kasturi P, Hartl FU. The proteostasis network and its decline in ageing. *Nat Rev Mol Cell Biol.* 2019;20:421–435.
13. Irobi J, Van Impe K, Seeman P, et al. Hot-spot residue in small heat-shock protein 22 causes distal motor neuropathy. *Nat Genet.* 2004;36:597–601.
14. Sarparanta J, Jonson PH, Kawan S, Udd B. Neuromuscular diseases due to chaperone mutations: a review and some new results. *Int J Mol Sci.* 2020;21:1409.
15. Selcen D, Muntoni F, Burton BK, et al. Mutation in BAG3 causes severe dominant childhood muscular dystrophy. *Ann Neurol.* 2009;65:83–89.
16. Irobi J, Almeida-Souza L, Asselbergh B, et al. Mutant HSPB8 causes motor neuron-specific neurite degeneration. *Hum Mol Genet.* 2010;19:3254–3265.
17. Ghaoui R, Palmio J, Brewer J, et al. Mutations in HSPB8 causing a new phenotype of distal myopathy and motor neuropathy. *Neurology.* 2016;86:391–398.
18. Shemetov AA, Gusev NB. Biochemical characterization of small heat shock protein HspB8 (Hsp22)-Bag3 interaction. *Arch Biochem Biophys.* 2011;513:1–9.
19. Montepietra D, Tesei G, Martins JM, et al. FRETpredict: a Python package for FRET efficiency predictions using rotamer libraries. *Commun Biol.* 2024;7:298.
20. Lemke T, Peter C. EncoderMap: dimensionality reduction and generation of molecule conformations. *J Chem Theory Comput.* 2019;15:1209–1215.
21. Schuler B, Soranno A, Hofmann H, Nettels D. Single-molecule FRET spectroscopy and the polymer physics of unfolded and intrinsically disordered proteins. *Annu Rev Biophys.* 2016;45:207–231.
22. Gopich IV, Szabo A. Theory of the energy transfer efficiency and fluorescence lifetime distribution in single-molecule FRET. *Proc Natl Acad Sci USA.* 2012;109:7747–7752.
23. Pesce F, et al. Design of intrinsically disordered protein variants with diverse structural properties. *Sci Adv.* 2024;10.eadm9926.
24. Basu S, Biswas P. Salt-bridge dynamics in intrinsically disordered proteins: a trade-off between electrostatic interactions and structural flexibility. *Biochim Biophys Acta Proteins Proteom.* 2018;1866:624–641.

25. Bjarnason S, McIvor JAP, Prestel A, et al. DNA binding redistributes activation domain ensemble and accessibility in pioneer factor Sox2. *Nat Commun.* 2024;15:1445.
26. Vancraenenbroeck R, Harel YS, Zheng W, Hofmann H. Polymer effects modulate binding affinities in disordered proteins. *Proc Natl Acad Sci USA.* 2019;116:19506–19512.
27. Aznauryan M, Delgado L, Soranno A, et al. Comprehensive structural and dynamical view of an unfolded protein from the combination of single-molecule FRET, NMR, and SAXS. *Proc Natl Acad Sci USA.* 2016;113:E5389–E5398.
28. Köfinger J, Stelzl LS, Reuter K, et al. Efficient ensemble refinement by reweighting. *J Chem Theory Comput.* 2019;15:3390–3401.
29. Müller BK, Zaychikov E, Bräuchle C, Lamb DC. Pulsed interleaved excitation. *Biophys J.* 2005;89:3508–3522.
30. Pace CN. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* 1986;131:266–280.
31. Holmstrom ED, Holla A, Zheng W, et al. Accurate transfer efficiencies, distance distributions, and ensembles of unfolded and intrinsically disordered proteins from single-molecule FRET. *Methods Enzymol.* 2018;611:287–325.
32. Hillger F, Nettels D, Dorsch S, Schuler B. Detection and analysis of protein aggregation with confocal single molecule fluorescence spectroscopy. *J Fluoresc.* 2007;17:759–765.
33. Sisamakis E, Valeri A, Kalinin S, et al. Accurate single-molecule FRET studies using multiparameter fluorescence detection. *Methods Enzymol.* 2010;475:455–514.
34. Förster T. Zwischenmolekulare Energiewanderung Und Fluoreszenz. *Ann Phys.* 1948;437:55–75.1–2.
35. Holmstrom ED, Holla A, Wenwei Z, et al. Transfer efficiencies, distance distributions, and ensembles of unfolded and intrinsically disordered proteins from single-molecule FRET. *Methods Enzymol.* 2018;611:287–325.
36. Maxwell KL, Best RB, Zitzewitz JA, et al. Protein folding: defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci.* 2005;14:602–616.
37. Hoffmann A, Kane A, Nettels D, et al. Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc Natl Acad Sci USA.* 2007;104:105–110.
38. Schellman JA. Fifty years of solvent denaturation. *Biophys Chem.* 2002;96:91–101.
39. Robustelli P, Piana S, Shaw DE. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc Natl Acad Sci USA.* 2018;115:E4758–E4766.
40. Piana S, Donchev AG, Robustelli P, Shaw DE. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J Phys Chem B.* 2015;119:5113–5123.
41. Bauer P, Hess B, Lindahl E. GROMACS. 2022.3 Source Code; Zenodo, 2022. https://doi.org/10.5281/zenodo.7037338.
42. Berendsen HJC, Postma JPM, van Gunsteren WF, et al. Molecular dynamics with coupling to an external bath. *J Chem Phys.* 1984;81:3684–3690.
43. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys.* 2007;126:014101.
44. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graphics.* 1996;14:33–38.
45. Schrödinger, LLC, & DeLano, WL. PyMOL – The PyMOL Molecular Graphics System, Version 2.4. Retrieved 2020, from https://pymol.org.
46. Waskom ML. Seaborn: statistical data visualization. *J Open Source Softw.* 2021;6:3021.
47. Seabold, S., & Perktold, J. Statsmodels: Econometric and statistical modeling with Python. In van der Walt, S. & Millman, J. (Eds.), Proceedings of the 9th Python in Science Conference SciPy. 2010;92–96. doi:10.25080/Majora-92bf1922-011.
48. KULL-Centre. (2024). FRETpredict [Computer software]. https://github.com/KULL-Centre/FRETpredict.
49. Klose D, Holla A, Gmeiner C, et al. Resolving distance variations by single-molecule FRET and EPR spectroscopy using rotamer libraries. *Biophys J.* 2021;120:4842–4858.
50. Woźniak A, Schröder GF, Grubmüller H, et al. Single-molecule FRET measures bends and kinks in DNA. *Proc Natl Acad Sci USA.* 2008;105:18337–18342.
51. Crehuet R, Buigues PJ, Salvatella X, Lindorff-Larsen K. Bayesian-maximum-entropy reweighting of IDP ensembles based on NMR chemical shifts. *Entropy.* 2019;21:898.
52. Hummer G, Köfinger J. Bayesian ensemble refinement by replica simulations and reweighting. *J Chem Phys.* 2015;143:243150.
53. Bottaro S, Bengtsen T, Lindorff-Larsen K. Integrating molecular simulation and experimental data: a Bayesian/maximum entropy reweighting approach. In: Gáspári Z, ed. *Structural Bioinformatics: Methods in Molecular Biology.* ed. Humana Press; 2020:219–240.
54. KULL-Centre. (n.d.). BME: Integrating Molecular Simulation and Experimental Data. GitHub. Retrieved June 6, 2025, from https://github.com/KULL-Centre/BME.
55. Sindbert S, Pardo-Avila F, Hohng S, Schuler B. Elucidating the properties of fluorescent dyes in single-molecule FRET experiments. *J Am Chem Soc.* 2011;133:2463–2480.
56. Lemke T, Berg A, Jain A, Peter C. EncoderMap(II): visualizing important molecular motions with improved generation of protein conformations. *J Chem Inf Model.* 2019;59:4550–4560.
57. Peter, A. G. (n.d.). encodermap, software. GitHub. https://github.com/AG-Peter/encodermap.
58. Tribello GA, Ceriotti M, Parrinello M. Using sketch-map coordinates to analyze and bias molecular dynamics simulations. *Proc Natl Acad Sci USA.* 2012;109:5196–5201.
59. Inkscape 1.4 Release NotesRelease Date: October 13, 2024, Official URL: https://inkscape.org/release/inkscape-1.4/.