JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

Accounting for Fast vs Slow Exchange in Single Molecule FRET Experiments Reveals Hidden Conformational States

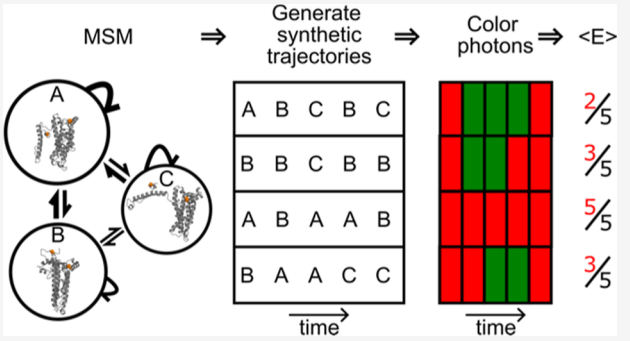
Published as part of Journal of Chemical Theory and Computation special issue "Markov State Modeling of Conformational Dynamics".

Justin J. Miller, Upasana L. Mallimadugula, Maxwell I. Zimmerman, Melissa D. Stuchell-Brereton, Andrea Soranno, and Gregory R. Bowman*

Cite This: J. Chem. Theory Comput. 2024, 20, 10339−10349 Read Online

ACCESS Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Proteins are dynamic systems whose structural preferences determine their function. Unfortunately, building atomically detailed models of protein structural ensembles remains challenging, limiting our understanding of the relationships between sequence, structure, and function. Combining single molecule Förster resonance energy transfer (smFRET) experiments with molecular dynamics simulations could provide experimentally grounded, all-atom models of a protein's structural ensemble. However, agreement between the two techniques is often insufficient to achieve this goal. Here, we explore whether accounting for important experimental details like averaging across structures sampled during a given smFRET measurement is responsible for this apparent discrepancy. We present an approach to account for this time-averaging by leveraging the kinetic information available from Markov state models of a protein's dynamics. This allows us to accurately assess which time scales are averaged during an experiment. We find this approach significantly improves agreement between simulations and experiments in proteins with varying degrees of dynamics, including the well-ordered protein T4 lysozyme, the partially disordered protein apolipoprotein E (ApoE), and a disordered amyloid protein (Aβ40). We find evidence for hidden states that are not apparent in smFRET experiments because of time averaging with other structures, akin to states in fast exchange in nuclear magnetic resonance, and evaluate different force fields. Finally, we show how remaining discrepancies between computations and experiments can be used to guide additional simulations and build structural models for states that were previously unaccounted for. We expect our approach will enable combining simulations and experiments to understand the link between sequence, structure, and function in many settings. Understanding protein dynamics is crucial for understanding protein function, yet few methodologies report on protein motion at an atomic level. Combining single molecule Förster resonance energy transfer (smFRET) experiments with computer simulations could provide atomistic models of protein ensembles which are grounded in experiments, however, there has been limited agreement between the two methods to date. Here, we present an algorithm to recapitulate smFRET experiments from molecular dynamics simulations. This approach significantly improves agreement between simulations and experiments for proteins across the ordered spectrum. Moreover, with this approach we can confidently create atomic models for states observed during smFRET experiments which were otherwise difficult to model due to high amounts of flexibility, disorder, or large deviations from crystal-like states.



INTRODUCTION

A protein's function is determined by the ensemble of structures that it adopts,[1−6] but building atomically detailed models of these ensembles to probe ensemble-function relationships remains challenging.[7] Of course, the high-resolution structures that structural biologists have become adept at solving are of enormous value. Despite this, ongoing challenges with tasks like drug and protein design highlight the limits of the structure−function paradigm.[8] We expect having detailed models of the rest of a protein's structural ensemble would lead to dramatic improvements in our understanding and ability to design such systems.[9−11] Unfortunately, most of the alternative structures a protein can adopt are difficult to

Received: August 15, 2024
Revised: November 15, 2024
Accepted: November 15, 2024
Published: November 26, 2024

detect and/or characterize experimentally because they have too low of a probability (i.e., high energy).[12−14]

smFRET experiments are a powerful tool for studying the distribution of structures that a protein adopts, including high energy states that are invisible to many other techniques.[15−17] In these experiments, a donor and an acceptor fluorophore are attached to two different residues in a protein. The donor fluorophore on a single protein is then excited and one measures how many acceptor and donor photons are emitted.[18,19] The probability of transferring energy from the donor to the acceptor fluorophore, called the FRET efficiency, reports on a variety of valuable structural properties, including the distance between the fluorophores, their relative orientations, and the time scale on which they are rotating. Making many measurements results in a probability distribution of FRET efficiencies. These FRET efficiency distributions report on the distribution of structures the protein adopts and have proved to be a powerful means of revealing the conformational heterogeneity of proteins.[20−23]

Unfortunately, one cannot extract atomically detailed structural models from smFRET data in a manner analogous to fitting structures to electron density from crystallography or cryoEM. smFRET data is inherently sparse, with each experiment reporting on the structure and dynamics of a single pair of dyes. One can perform experiments for multiple dye positions to learn about more of the protein structure. However, each experiment is independent, making it hard to discern any correlations between the behavior of different parts of the protein. While multicolor FRET experiments are being developed,[24] they are quite challenging to perform and still cannot measure many distances in parallel. Another challenge is that there is not a one-to-one mapping between the FRET efficiency and the distance between a pair of residues.

Combining atomically detailed computer simulations with smFRET experiments could yield experimentally grounded models of protein conformational ensembles with the desired resolution.[25,26] Ideally, there would be a method to predict energy transfer distributions from simulations to show that these predictions were in perfect agreement with smFRET experiments. Then one could analyze the simulations, making use of the atomistic structural and dynamical information they provide to generate new hypotheses, and test those hypotheses experimentally.

While there are cases where smFRET experiments and simulations are in good agreement, the agreement between the two approaches is often limited.[25] A variety of approaches have been employed to close this gap. For example, scaling factors have been used to account for the added distance between two alpha carbons modeled in simulation and the distance between two fluorophores as measured in FRET.[27,28] Others have employed reweighting schemes to account for force field biases and shift the relative probabilities of structures from their simulations into closer agreement with experiments.[29−31] There have also been efforts to develop improved methods for predicting the probability of energy transfer from simulations (e.g., by modeling in the dyes) and to improve force fields.[26,28,32−42] However, there is still room for improvement.

Here, we explore the importance of accounting for kinetic effects in smFRET experiments when connecting with simulations. Each FRET efficiency measured in an smFRET experiment is the ratio of acceptor photons to all photons emitted during some time interval. This time interval typically ranges from one to 10 ms depending on the experimental setup (e.g., TIRF vs diffusion confocal, laser power, etc.).[22,23] It has long been recognized in the smFRET community that this means each measured FRET efficiency is, therefore, averaging across whatever conformational dynamics occur during the one to 10 ms time interval. As in nuclear magnetic resonance (NMR) experiments, conformations that are exchanging more quickly than this measurement time will be averaged together, while conformations that are exchanging more slowly will not. Significant effort has gone into dealing with this time-averaging when analyzing experiments.[43−53] For example, it is common to perform global fits to many measurements with different dye positions and solvent conditions,[54−56] fit hidden Markov models to photon traces,[21,46,51] or dissect the correlation between FRET efficiency and other fluorescence observables that report on shorter time scales.[44,45,55] New experimental approaches are also being developed to shorten the time scale over which FRET efficiencies are measured.[57−61] Attempts have also been made to account for time when connecting simulations with smFRET experiments.[41,62−65] However, these approaches are limited by the fact that it is rare to capture all relevant conformational states of a protein in a single trajectory or even a small collection of trajectories. Indeed, given the computational expense of simulations, it is often helpful to use enhanced or adaptive sampling approaches to capture large protein motions. A unified computational approach to analyzing swarms of simulation trajectories while also accounting for time-averaging could dramatically improve agreement between experiments and simulations enabling these two approaches to be used even more effectively to advance our understanding of the ensemble-function relationship.

Here, we present an approach for accounting for time-averaging when predicting FRET efficiencies from simulations and assess its performance on three well-studied systems that exemplify different extents of dynamics. An important methodological contribution is the use of adaptive sampling to accelerate exploration of conformational space and the use of Markov state models (MSMs) to describe conformational dynamics. In particular, using MSMs to generate synthetic trajectories captures even more variations on dynamics than the original simulations since these synthetic trajectories can include sequences of transitions that were never observed in a single molecular dynamics simulation. To demonstrate our approach, we start with Apolipoprotein E4 (ApoE4), as it contains both ordered and disordered regions and addresses the applicability of our approach to each.[54,66] We also apply our approach to T4 Lysozyme, a well-ordered system that has recently been extensively characterized using 33 distinct smFRET labeling positions.[55] Finally, we apply our approach to amyloid-β40 (Aβ40), a 40 amino acid highly disordered protein.[28] One recent study produced simulations of Aβ40 with a variety of force fields, giving us the chance to test how well different force fields perform when combined with our approach for accounting for time-averaging when predicting the experimentally observed energy transfer distribution.[67]

RESULTS

Accounting for Time-Averaging Ddramatically Improves Agreement between Simulations and Experiments for a Partially Disordered Protein. We developed an approach for predicting FRET efficiencies from simulations in a manner that accounts for time-averaging by drawing on

https://doi.org/10.1021/acs.jctc.4c01068
J. Chem. Theory Comput. 2024, 20, 10339−10349

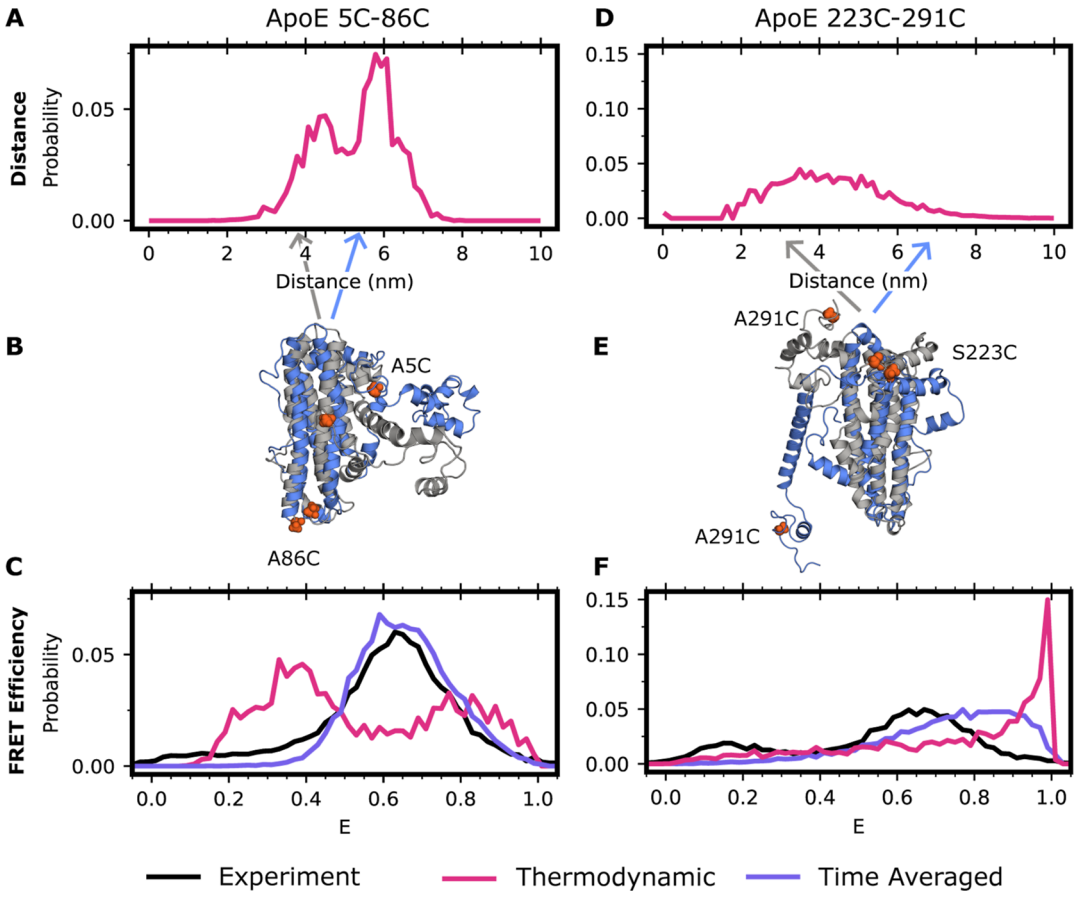


Figure 1. Accounting for time averaging significantly alters the apparent structural distribution from our model and increases agreement with experiments. (A) Interdye distances for apolipoprotein E labeled with Alexafluor 488 and Alexafluor 594 at positions 5 and 86 or (D) 223 and 291. In red is the result from taking a purely thermodynamic perspective and accounting for the distance added or subtracted by dye positioning (B,E) Exemplar structures of ApoE at two distinct dye-distance positions. Arrows indicate the portion of the distance distribution the structure occupies. (C) FRET efficiencies obtained for positions 5 and 86 or (F) 223 and 291. In black is the experimental distribution, in red is the thermodynamic distribution of ApoE, and in purple is the time-averaged distribution resulting from the red trace.

MSMs[68−71] built from molecular dynamics simulations. An MSM is a network model that describes a molecule's conformational space in terms of the structural states it adopts and the probabilities of hopping between every pair of states in a fixed time interval. These models integrate information from many independent simulations to capture length and time scales that are far beyond the reach of any individual simulation. Importantly, we can use an MSM to generate a synthetic trajectory using a kinetic Monte Carlo scheme in which one chooses a random starting state and then iteratively adds new random states based on the transition probabilities from the current state to all other states. To mimic a smFRET experiment, we use one MSM to describe the protein's conformational dynamics and separate MSMs for each of the dyes. First, we select a random experimental photon time trace and use our protein MSM to generate a synthetic trajectory of the same length. Then we identify conformations in our synthetic trajectory that correspond to the times photons were detected in the experiment. We assume these are the conformations that emit photons and then choose whether to label each photon as coming from the donor or acceptor dye as follows. First, we generate a set of plausible dye conformations by mapping representative structures from each state in our dye MSM onto the protein structure and removing any that form steric clashes with the protein and recreating a MSM based on the remaining states. Next, we use our dye MSMs to simulate the dynamics of the dye (on a fixed protein structure) leading up to emission of a photon. At each step of these dye simulations, we use a Monte Carlo move to decide if the donor emits a photon, transfers energy leading to emission of an acceptor photon, or stays excited similar to previous efforts.[40,41] If the dye remains excited, both dyes are allowed to hop to another state in the MSM. We repeat this process for the remainder of the synthetic trajectory to simulate a photon burst, returning the average FRET efficiency, or the number of acceptor photons divided by the total photons, for that burst. Finally, we repeat this process over multiple trajectories until an adequate number of photon bursts have been sampled. Since we model dyes as a postprocessing step instead of including the dyes in the simulations, it is easy to scale this approach to predict the observed FRET for many dye positions. Furthermore, simulating the dye dynamics allows us to minimize the number of adjustable parameters, as we do not need to select constant values like a Förster radius that are required by other approaches.

To test our approach, we applied it to the partially disordered protein ApoE4. Our recent work presented smFRET measurements for five different pairs of dye positions on this protein. Some of these dye positions report on

https://doi.org/10.1021/acs.jctc.4c01068
J. Chem. Theory Comput. 2024, 20, 10339−10349

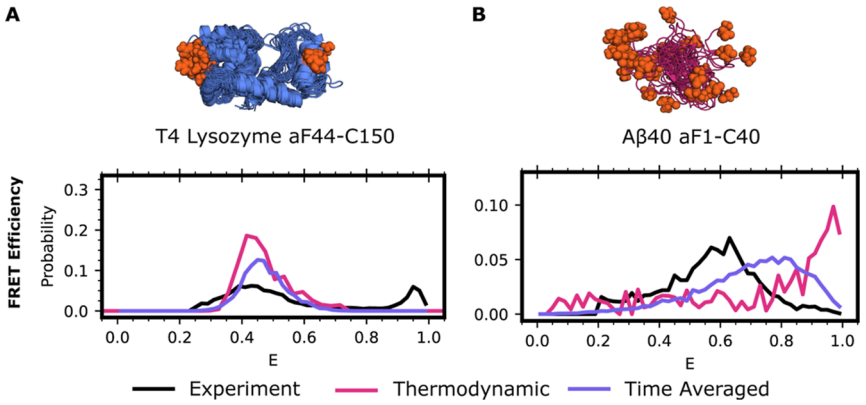


Figure 2. smFRET time averaging impacts proteins across the ordered spectrum. (A) FRET efficiencies for T4 Lysozyme labeled at 44 (*para*-acetylphenylalanine) and 150 (cysteine) with Alexa 488 and Alexa 647 or (B) Aβ40 labeled at positions 1 (*para*-acetylphenylalanine) and 40 (cysteine) with Alexa 488 and Alexa 647. In black is the experimental distribution, red the thermodynamic FRET efficiency, and purple accounting for protein dynamics via time-averaging. Protein structures are the 15 most probable states in the MSM with labeling positions indicated in orange spheres. Experimental donor only counts ($E < 0.25$) have been removed for ease of comparison.

dynamics within the largely folded N-terminal domain while others report on the partially disordered and highly flexible C-terminal domain.[54] Therefore, comparisons between simulations of this protein and experiments speak to the utility of our approach for both well-folded and disordered structures. We previously showed that ApoE4 predominately adopts three conformational states: a closed state, an open state, and an extended state. Identifying these three states experimentally required an enormous number of measurements, including multiple dye positions each at varying levels of denaturant. Even with this wealth of data, building atomically detailed structural models of the different states required extensive molecular dynamics simulations, totaling over 3 ms of aggregate simulation, which we showed were in reasonable agreement with experiments using a simplified version of the approach presented here. That approach did not model dye dynamics and, therefore, required us to choose constant values for parameters like the fluorescence anisotropy used in the Förster radius.

Assessing different ways of predicting FRET efficiencies from simulations and examining the actual distance distributions in those simulations highlights the importance of accounting for time-averaging (Figure 1). For example, Figure 1A shows the modeled interdye distance distribution between residues 5 and 86 in the folded, N-terminal domain. This distribution has two peaks, which roughly correspond to the closed and open states of ApoE4 (Figure 1B). If one assumes that smFRET measurements are instantaneous (i.e., there is no time-averaging and the observed signal is entirely determined by the thermodynamics of the ensemble), then the distribution of FRET efficiencies that one predicts retains these two peaks. We call this the thermodynamic distribution for brevity. However, accounting for time-averaging causes these two peaks to collapse into a single peak because the different populations are in fast exchange (Figure 1C). Importantly, the FRET efficiency we predict by accounting for time-averaging is in good agreement with the experimental data. Without accounting for time-averaging (i.e., using a purely thermodynamic approach), we would have come to the erroneous conclusion that our simulations were in poor agreement with experiments. By accounting for time-averaging, we instead find good agreement with experiments and can use the simulation data to help identify the different populations that give rise to the experimentally observed smFRET data.

Repeating the analysis above for the other dye positions supports the importance of accounting for time-averaging. For example, the modeled interdye distance distribution between residues 223 and 291 in the disordered C-terminal domain is broad and symmetrical (Figure 1D). The distribution of FRET efficiencies one would predict from a purely thermodynamic perspective (i.e., without accounting for time-averaging) is skewed to large FRET efficiencies, in poor agreement with the experimentally observed distribution of FRET efficiencies. Accounting for time-averaging improves the agreement between simulations and experiments (Figure 1F). Importantly, the MSM accounts for motions occurring over multiple time scales enabling us to automatically average together states which are interconverting rapidly in a single energy basin while simultaneously capturing the differences between states that are interconverting slowly and thus broadening the histogram. Similar results are found for other dye positions (Figure S2). Treating the dyes as a point cloud rather than modeling their dynamics also gives similar results (Figure S3), though this approach requires the choice of a constant Förster radius that can be a source of error if a poor choice is made or if the dyes are not isotropically rotating.

Time-Averaging Improves Experiment-Simulation Agreement for the Entire Spectrum from Ordered to Disordered Proteins. Given that ApoE has a mix of ordered and disordered regions, we reasoned that our time-averaging approach should be equally applicable to fully ordered and disordered systems. To test this hypothesis further, we used our approach on the highly ordered protein, T4 lysozyme, and the intrinsically disordered protein (IDP), Aβ40. Both T4 lysozyme and Aβ40 benefit from a plethora of prior structural studies, including experimental smFRET characterization.[28,55] For Aβ40, we make use of an existing 30 μs long simulation in the amber99sb force field, which was found to match NMR order parameters reasonably well.[67] For lysozyme, we performed 5 independent 5 μs long all-atom molecular dynamics simulations in explicit tip3p solvent and the amber03 force field as described in the methods section. For both Aβ40 and lysozyme, we clustered our data sets, made MSMs, modeled on the appropriate dye pairs to match the

https://doi.org/10.1021/acs.jctc.4c01068
J. Chem. Theory Comput. 2024, 20, 10339−10349

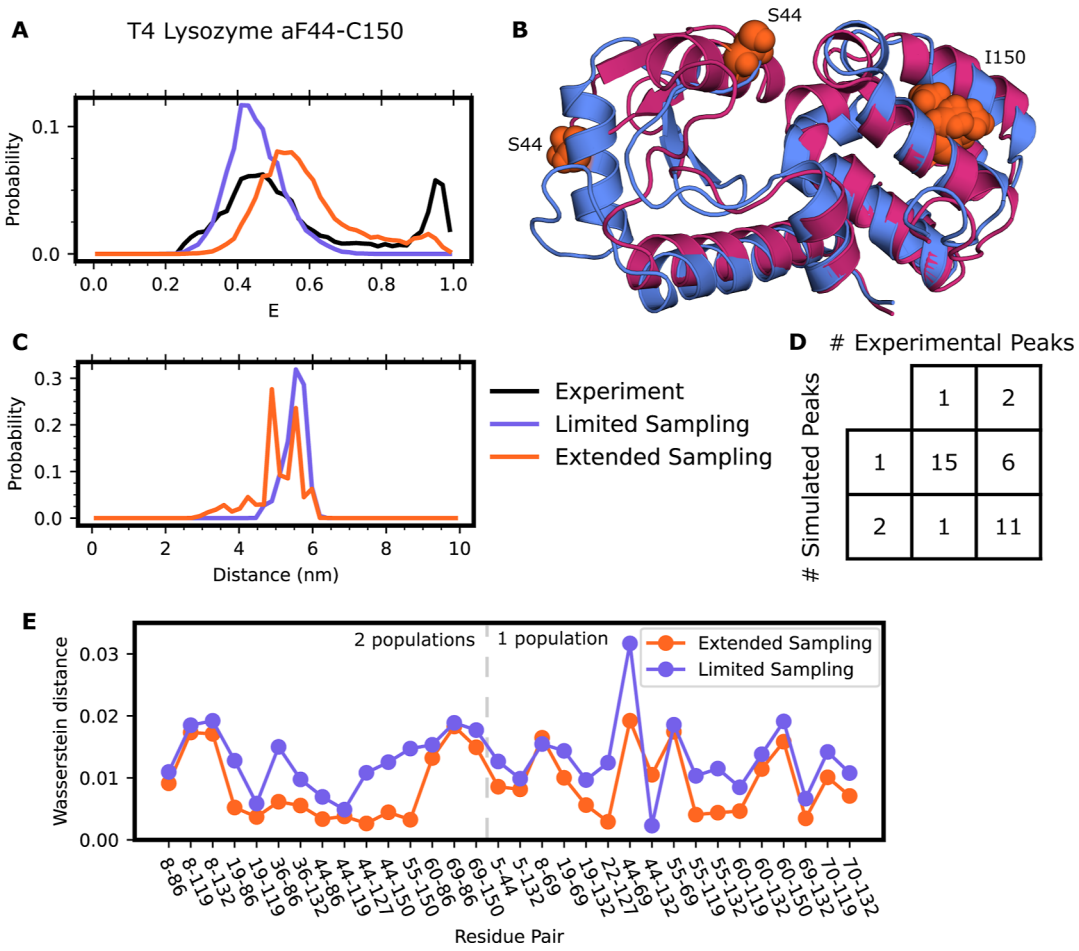


Figure 3. Discrepancies between smFRET and time averaging results enable discovery of a novel lysozyme fold. (A) FRET efficiencies and (C) observed interdye distances for T4 Lysozyme labeled at position 44 and 150 with Alexa 488 and Alexa 647. Experimental traces are in black, calculations resulting from an MSM that only included crystal-like states in purple, and calculations resulting from an MSM including the alternate state in orange. Experimental donor only counts ($E < 0.25$) have been removed for ease of comparison. (B) Example conformations of the crystal-like state of lysozyme (blue), or the alternate state (red). Residues 44 and 150 labeled for clarity. (D) Qualitative comparison of smFRET distributions from experiment and simulation results including the alternate pose of lysozyme. (E) Wasserstein distance between experimental and limited sampling data set (purple) or extended sampling data set (orange) for all labeled pairs.

experimental setup, and investigated the predicted FRET efficiencies using our time-averaging approach.

As expected, we found that accounting for time-averaging is important for both systems (Figure 2). We first calculate the lysozyme FRET efficiency for one of the experimental FRET probe distances, residues 44−150, using Alexa 488 maleimide and Alexa 647 hydroxylamine dyes. We find strong agreement between our time-averaging approach and the experimental results for lysozyme (Figure 2A). We note that both time-averaging and thermodynamic FRET are in reasonable agreement with experimental data for this probe position, though both miss a population at high FRET efficiency. We next calculate the FRET efficiency for Aβ40 using Alexa 488 hydroxylamine and Alexa 647 maleimide attached to positions 1 and 40. We find that the results significantly improve upon the distribution obtained without time-averaging. However, the distribution is shifted overall toward higher FRET efficiencies, suggesting either insufficient sampling or force fields issues (Figure 2B). Overall, these findings demonstrate that accounting for time-averaging is helpful when there are conformations in fast exchange and is equivalent to other approaches when such exchange is absent.

Directing Sampling Based on Discrepancies between Predicted and Observed FRET Reveals a Novel Conformation of Lysozyme. Given the strong agreement between our predicted energy transfer distributions and experiments for folded and partially disordered systems, we reasoned that remaining discrepancies may point to under sampled regions of conformational space in simulations. Indeed, the prior study on lysozyme concluded that the minor population could not be explained by any structure of lysozyme existing in the protein data bank.[55] If this is true, then we should be able to improve the agreement between simulations and experiments by driving simulations to sample structures with FRET values that are not observed often enough compared to experiments.

To explore this possibility, we sought to provide a structural model for a minor population that was previously observed in an extensive smFRET study of lysozyme. That study presented smFRET measurements for 33 pairs of dye positions. For 17 of these dye pairs, the authors observe a minor state in the FRET efficiency that they could not explain based on any of the numerous published crystal structures of this protein. When probing residues 44−150, this minor population has high FRET efficiency, a metric which would require the dyes to

https://doi.org/10.1021/acs.jctc.4c01068
J. Chem. Theory Comput. 2024, 20, 10339−10349

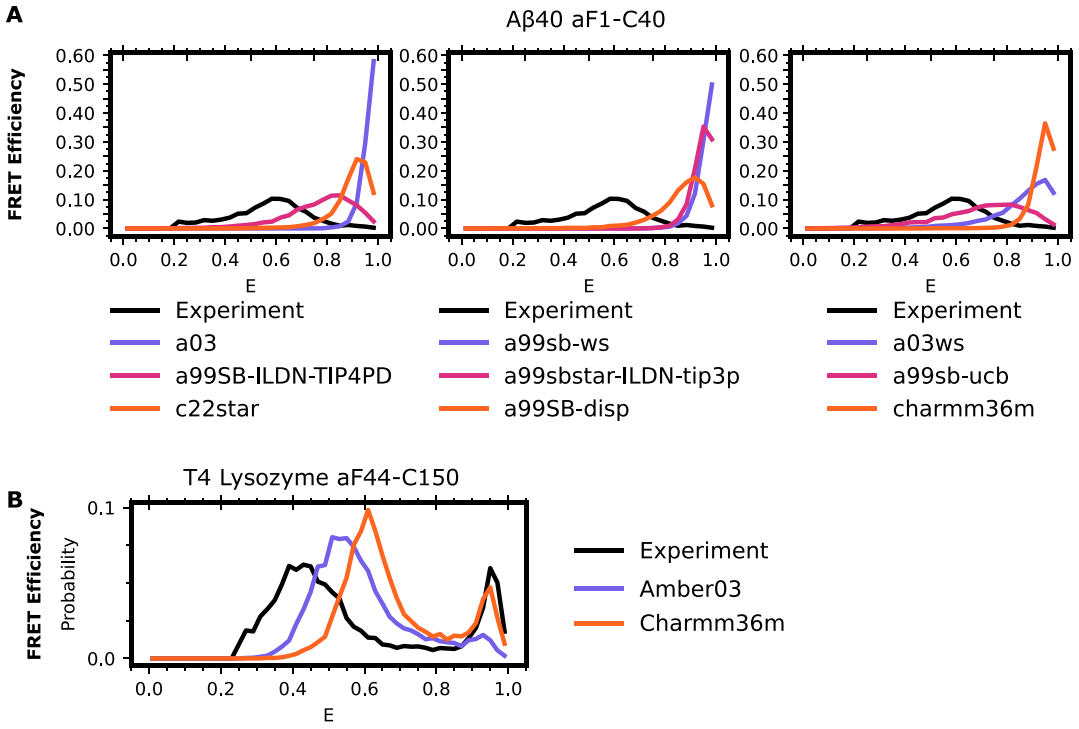


Figure 4. Force field choice has a significant impact on the level of agreement between simulations and experiments on IDPs even when accounting for time averaging. (A) FRET efficiencies for Aβ40 labeled at positions 1 and 40 with Alexa 488 and Alexa 647 in different force fields. In each trace, the black line represents the experimental smFRET result, individual force field and water combinations indicated in the legend below the column. (B) FRET efficiency distributions for T4-lysozyme residues 44 and 150 in amber03 with TIP3P water (purple) or charmm36m with TIP3P water (orange) or the experimental result (black). Experimental donor only counts ($E < 0.25$) have been removed for ease of comparison.

come closer together than is conceivable based on a clamshell motion of the two lobes of lysozyme. Furthermore, we would not expect our simulations (aggregate simulation time of 25 μs) to reach this minor state since the experiments suggest that it is accessed with a rate of ~4 ms^{-1}. Indeed, we find that our simulations stay near the starting structure and that our predicted energy transfer distributions agreed well with the major population seen experimentally but missed the minor population seen for constructs like $lysozyme_{44-150}$ (Figure 3A).

To provide a structural explanation for the minor population seen experimentally, we employed a combination of metadynamics and MSMs. First, we used metadynamics simulations to find structures that are consistent with the high FRET of the minor population. In metadynamics, one adds an external biasing force to drive dynamics along a preselected collective variable. In this case, we pushed the system along the distance between residues 44 and 150 to see if we could find structures where they come close together, as this would result in a high FRET efficiency. These simulations revealed that the β domain can undergo minor unfolding which enables a swiveling motion to bring residue 44 much closer to residue 150 of the α domain (Figure 3B). To test if this alternative structural state is metastable, we selected four conformations where residues 44 and 150 are near one another and ran >500 ns conventional molecular dynamics simulations of each of them. All the simulations stayed near the starting point, confirming that the alternative structure state we discovered in metadynamics is a metastable free energy state. To determine the relative probability of this alternative state and those observed in our original simulations, we sought to build an MSM that captured transitions between the crystal-like states and the new alternative state. We ran goal-oriented adaptive sampling simulations using the fluctuation amplification of specific traits (FAST) algorithm[72] to promote transitions from the crystallographic state to the alternative state and vice versa. FAST works by iteratively running a batch of simulations, building an MSM, and choosing states from the MSM as starting points for new simulations in a manner that balances between exploring further around states with a specified geometric property (called exploitation) and broad exploration of conformational space. In this case, we started one batch of FAST simulations from the crystallographic structure and set the exploitation term to favor states with a low RMSD to the novel fold we discovered and a second with the targets reversed. Both sets of simulations captured transitions between the two folds, providing a basis for building an MSM that captures the relative probabilities of both folds of lysozyme.

After building a new MSM that incorporates this data, we find that the computationally calculated energy transfer distribution now includes a minor state in agreement with experiments with multiple dye positions (Figure 3). Specifically, including the novel fold of lysozyme greatly improves agreement between our time-averaged results and the experimental traces for labeling pair 44−150 (Figure 3A,B). As a further test of our model, we then calculated energy transfer distributions for the remaining 32 FRET probe positions. Agreement between our model and these experiments would be strong support for our model, given that none of these experiments influenced our simulation strategy. In support of the alternative fold we predicted, we see the addition of minor peaks to 11 of the 17 FRET probe positions that were sensitive to this minor population, and only one additional peak in probe positions not reporting on the minor population (Figures 3D and S4). Inclusion of the alternate

https://doi.org/10.1021/acs.jctc.4c01068
J. Chem. Theory Comput. 2024, 20, 10339−10349

state greatly reduces the Wasserstein distance between the experimental and predicted histograms for all states except 2:8−69 and 44−132.

Results for IDPs are More Sensitive to Force Field Choice than Globular Proteins. We hypothesized that accounting for time-averaging in our Aβ40 prediction was insufficient to give strong agreement with experiment because the force field preferred overly compact states of Aβ40. Historically, force fields which govern the underlying physics of the simulation, have yielded large differences between simulations and the corresponding experimental data for IDPs.[67,73,74] While amber99sb notably performs better with IDPs than others, force fields that were parametrized for folded proteins tend to lead to over compaction of IDPs, largely due to an imbalance between protein−protein and protein−water interactions. This systematic compaction of Aβ40 would skew the observed FRET values toward higher efficiencies, exactly as we have observed (Figure 2B).

To test whether discrepancies in experimental agreement are due to force field errors, we predicted Aβ40 smFRET using simulations conducted with a suite of nine different force fields/water models. Seven of these data sets were taken from a previous study that examined how well 30 μs simulations with each force field recapitulated NMR measurements. The seven force field/water combinations are amber99SB*-ILDN with TIP3P, C22* with TIP3P-CHARMM, C36m with TIP3P-CHARMM, a03ws with TIP4P, a99SB with TIP4P-Ew with Head-Gordon vdW and dihedral modifications (a99sb-ucb), a99SB-ILDN with TIP4P-D, and a99SB-disp with a modified TIP4P water. In addition to these data sets, we also ran our own simulations using amber03[75] and a99sb-ws, both with TIP3P water. For each of our simulations, we ran 250 ns long simulations in triplicate starting from the 10 most distinct Aβ40 structures captured in the previous simulations.

We find that force field choice substantially affects the quality of smFRET prediction for IDPs and that there is still room for improvement. For each of the above force field−water combinations, we generated MSMs and calculated the expected smFRET. All force field−water combinations result in Aβ40 distributions that are more collapsed than the experimental distribution (Figure 4A). We note that in our data sets, we explicitly started simulations from expanded states of Aβ40. However, these states exhibit a rapid compaction event which is not reversed during the simulation, consistent with previous findings that most force fields are biased toward more compact IDP structures than are experimentally observed.[42,73,76] Of the force field−water combinations, a99SB-ILDN-TIP4PD and a99sb-ucb showed the strongest agreement with experiment.

While there is large variation in force field performance for IDPs, we find less variation between force fields for lysozyme. We performed another 5 independent, 5 μs long replicate simulations of lysozyme in tip3p water using charmm36m as a force field. As with our amber03 simulations, we fail to uncover the third, minor population, of lysozyme (Figure S5). Using the previously discovered minor state of lysozyme as both a starting and target structure, we again performed goal-oriented sampling to promote transitions between the crystal-like poses of lysozyme and the minor state in the charmm36m force field. Both sets of simulations again capture transitions between the two states, enabling us to construct an MSM. Though there are slight differences in energy transfer efficiency between our data sets constructed using amber03 and charmm36m, both produce good agreement between the experimental energy transfer distributions and those predicted from our MSMs (Figure 4B).

DISCUSSION

Here, we have explored how conformational averaging during smFRET measurements impacts the observed distribution of FRET efficiencies. Our results show that accounting for time-averaging across protein and dye dynamics improves the agreement between simulations and experiments for three proteins across the ordered spectrum (Figures 1 and 2). These results agree with an existing experimental understanding that single molecule energy transfer distributions report an average of all protein motion during the measurement window. Our work adds to these prior experimental efforts by both identifying which states are being averaged together, while also providing an atomistic view of the protein conformations. While prior computational efforts have often focused on accounting for protein dynamics, improving protein and dye force field accuracy, or accounting for dye dynamics in FRET predictions, often these efforts either require additional simulations for every labeled dye position, or are unable to account for the effect of dye dynamics without a priori knowledge of the dynamical nature of the dyes. Our approach is unique in that it leverages MSMs to account for both protein and dye dynamics without the need for additional, computationally expensive, simulations. This removes additional modeling choices, such as choosing a Förster radius or time scale to average dye motions over, from the calculation.

Historically it has been difficult to determine why simulations and experiments have failed to agree. While simulations can have systematic errors due to parametrization or incomplete conformational sampling, experimental limitations and artifacts may also lead to disagreements. Here, we highlight examples where predicted and experimentally obtained energy transfer efficiency measurements appeared to disagree until we properly accounted for details of the experiment like time-averaging. Accounting for these experimental details in our modeling approach did not provide a structural rationale for a previously observed minor population of T4 lysozyme. However, we were able to use this persistent discrepancy to guide additional simulations to find structures that are consistent with the energy transfer in the minor populations. This approach allowed us to propose a structural model for the previously unexplained minor population that is consistent with most of the experimental measurements (Figure 3). We also find that simulations of Aβ40 with force field and water combinations parametrized for folded proteins result in an overly compact ensemble compared to the experimentally determined ensemble (Figure 4). As expected, based on previous publications, force fields designed to improve performance on disordered proteins performed better in our tests. While there were still differences between amber03 and charmm36m force fields for our lysozyme simulations, the choice of force field was less impactful for lysozyme than Aβ40.

We expect our approach will enable combining simulations and experiments to understand the link between sequence, structure, and function in many settings. While smFRET experiments are extremely valuable, one cannot readily derive atomistic models of conformational distributions from this data alone. The approach we have outlined here enables robust calculation of energy transfer distributions from protein

https://doi.org/10.1021/acs.jctc.4c01068
J. Chem. Theory Comput. 2024, 20, 10339−10349

ensembles, providing a direct link between energy transfer distributions and atomic models. While our approach led to strong agreement between simulations and experiments, there are some exceptions where our data sets diverge. One explanation could be that we do not consider alternative mechanisms of donor energy emission—such as quenching via nearby residues such as tryptophan and tyrosine. Another explanation could be that the mutagenesis required for dye attachment during experiments, as well as the attachment of the dye itself, disrupt the conformational landscape of the protein in question. Indeed, in many of our apo-simulation models we observe states that are incompatible with dye labeling, such as when amino acid to-be-labeled becomes buried or interacts closely with other regions of the protein. Nonetheless, our method is implemented postsimulation, modeling additional dye positions is rapid and requires minimal additional computational cost. Accordingly, once one has a satisfactory simulation data set, it is facile to use these tools to design novel probe pairs which report on identified motions of interest.

MATERIALS AND METHODS

Molecular Dynamics Simulations. All simulations were performed in explicit solvent at 300 K. Simulations of Apolipoprotein E4 were generated using OpenMM8.0[77] and amber03[75] with TIP3P water[78] and a time step of 4 fs. The data set was generated using a diverse composition of starting structures of ApoE initially based on PDB ID 2L7B{Citation} and totals 3.61 ms. Clustering was performed based on the pairwise distances between 15 selected residue pairs and the MSM was created using row normalization and a 2 ns lagtime.

Simulations of Aβ40 in the following force fields were obtained from prior work:[67] a99SB*-ILDN with TIP3P, C22* with TIP3P-CHARMM, C36m with TIP3P-CHARMM, a03ws with TIP4P/2005 interactions, a99SB with TIP4P-Ew with the Head-Gordon vdW and dihedral modifications (a99SB-UCB), a99SB-ILDN with TIP4P-D, and a99SB-disp with a modified TIP4P-D water. For each combination, a total of ∼30 μs data were collected using Anton hardware. Simulations of Aβ40 in amber03 + TIP3P and amber99sb-ws + TIP3P were generated in this work using GROMACS and 10 diverse starting structures from the above Aβ40 data set and running each simulation for 250 ns in triplicate using unique initial velocities for each (aggregate 7.5 μs). Clustering for both simulation data sets was performed using the distance between every fifth residue as a feature, and the MSM was created using row normalization and a 5 ns (Anton data sets) or a 0.2 ns lag time (GROMACS).

Simulations of T4 lysozyme were performed in amber03 with TIP3P water. Initial unbiased simulations were started from PDB structure 5LZM[79] and 5 replicates were performed for 5 μs each using differing initial velocities. Metadynamics simulations were performed using PLUMED and a biasing potential of 0.3 between residues 44 and 150 for a total of 250 ns. Unbiased simulations were started from 4 alternate states uncovered by metadynamics with 5 replicas using differing initial velocities, each for a total length of 1 μs. FAST adaptive sampling was performed from both alternate and crystal-like states to the opposing state to capture the transition pathways in forward and reverse using RMSD as a progress metric. MSMs were built based on cluster centers from initial unbiased simulations, or the entire data set excluding the metadynamics runs (total 94.8 μs). Clustering was performed using backbone RMSD to a radius of 2.5 Å and the MSM was built using row normalization and a lag time of 2 ns.

Simulations of dyes were performed in amber03ws using the modified amber dye parameters[28,39] with TIP3P water and a time step of 2 fs. A single run was performed for each dye for 500 ns. Simulation frames were saved every 20 fs. A 5000 state MSM with a lag time of 2 ps was built for each dye using RMSD of heavy atoms as a clustering metric.

Simulation of smFRET. Dye color determination was achieved by building a MSM for each dye of interest (see simulations, above). Briefly, the dye MSM is modeled onto each state in the protein MSM, removing any positions from the dye MSM that clash with the protein and creating a new, renormalized MSM with clashing states removed. Next, a random dye position is chosen for both the donor and acceptor dye. Probabilities of radiative decay, energy transfer, nonradiative decay, or remaining excited are calculated and a random outcome is chosen accordingly, similar to prior work.[40,41] If the dye remains excited, dye positions are allowed to update along with transfer probabilities until the donor dye is no longer excited. In the case of point clouds, a conformational ensemble of dyes[26] was modeled onto the protein and all steric clashes were discarded. Photon colors were determined by choosing a random distance between the donor and acceptor dye emission centers and the Förster relationship (eq 1).

$$E = R_0^6/(R_0^6 + r^6) \tag{1}$$

To determine which protein states to average, we recolor an experimental photon trace from Apolipoprotein E4.[54] We choose a random state from our protein MSM and build a synthetic trajectory to match the length of the experimental photon burst. We apply a time correction factor of 10,000 to slow the simulation time scale to match the experimental time scale. Each time an experimental photon is observed, we select the corresponding state in our trajectory and evaluate the photon identity as above. We determine the overall energy transfer efficiency as the ratio of acceptor photons to the total observed photons and repeat this process for all bursts (∼14,000), yielding the displayed distributions. The code for these calculations is available on github (https://github.com/bowman-lab/enspara).

Experimental smFRET Data. Data for Apolipoprotein E4 was obtained from Stuchell-Brereton et al.[54] Data for T4 lysozyme was obtained from Sanabria et al.[55] Data for Aβ40 was obtained from Meng et al.[28]

Analysis/Software. Simulations generated during this manuscript were performed in GROMACS2020[80] or OpenMM8.0[77] as noted. Adaptive sampling was performed using FAST,[72] and metadynamics[81] simulations were performed using PLUMED[82] and GROMACS2020. Structure imaging was performed in PyMOL. Trajectory analysis was performed using MDtraj.[83] Clustering and MSMs were created using ENSPARA.[84] All graphs were generated using Matplotlib.[85]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.4c01068.

Additional methodological details on simulation setup, parameters, and MSM building. Additional methodo-

https://doi.org/10.1021/acs.jctc.4c01068
J. Chem. Theory Comput. 2024, 20, 10339−10349

logical details on modeling of FRET processes computationally. Implied time scales for dye and protein MSMs. FRET distributions for three additional dye pairs for ApoE. Impact of modeling FRET dyes as point clouds versus accounting for dye dynamics for protein ApoE. FRET distributions for all 33 labeling positions of lysozyme pre and postextended sampling. FRET distribution for lysozyme 44−150 with simulations performed in charmm36m and amber03 force fields (PDF)

AUTHOR INFORMATION

Corresponding Author

Gregory R. Bowman − *Departments of Biochemistry & Biophysics and Bioengineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104, United States;* orcid.org/0000-0002-2083-4892; Email: grbowman@seas.upenn.edu

Authors

Justin J. Miller − *Departments of Biochemistry & Biophysics and Bioengineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104, United States;* orcid.org/0000-0001-9400-8916

Upasana L. Mallimadugula − *Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110, United States*

Maxwell I. Zimmerman − *Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110, United States;* orcid.org/0000-0003-0721-0652

Melissa D. Stuchell-Brereton − *Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110, United States*

Andrea Soranno − *Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110, United States;* orcid.org/0000-0001-8394-7993

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.4c01068

Author Contributions

JJM: conceptualization, data curation, methodology, software, writing-original and final draft. ULM: conceptualization, data curation, writing-review and editing. MIZ: conceptualization, writing-review and editing. MDSB: conceptualization, writing-review and editing. AS: conceptualization, writing-review and editing. GRB: conceptualization, writing-initial and final drafts, supervision, funding acquisition.

Notes

The authors declare no competing financial interest.

Preprint Servers: This manuscript has been deposited on bioRxiv with license CC-BY.

ACKNOWLEDGMENTS

We thank the citizen scientists who contributed computational time and technical expertise for the simulation of ApoE via Folding@home. This work was supported by the National Institutes of Health through U19AG069701 (project 1: A.S., core B: G.R.B.), RF1AG067194 (G.R.B. and A.S.), NIGMS R35GM152085 (G.R.B.), and NSF MCB 2218156 (G.R.B.). J.J.M. was funded by the NIH training grant T32AG05851804.

REFERENCES

(1) Knoverek, C. R.; et al. *Opening of a Cryptic Pocket in β-Lactamase Increases Penicillinase Activity*; Proceedings of the National Academy of Sciences, 2021; Vol. *118*, pe2106473118.

(2) Cruz, M. A.; Frederick, T. E.; Mallimadugula, U. L.; Singh, S.; Vithani, N.; Zimmerman, M. I.; Porter, J. R.; Moeder, K. E.; Amarasinghe, G. K.; Bowman, G. R. A cryptic pocket in Ebola VP35 allosterically controls RNA binding. *Nat. Commun.* **2022**, *13*, 2269.

(3) Xie, T.; Saleh, T.; Rossi, P.; Kalodimos, C. G. Conformational states dynamically populated by a kinase determine its function. *Science* **2020**, *370*, No. eabc2754.

(4) Fraser, J. S.; Clarkson, M. W.; Degnan, S. C.; Erion, R.; Kern, D.; Alber, T. Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **2009**, *462*, 669−673.

(5) Knoverek, C. R.; Amarasinghe, G. K.; Bowman, G. R. Advanced methods for accessing protein shape-shifting present new therapeutic opportunities. *Trends Biochem. Sci.* **2019**, *44*, 351−364.

(6) Chen, S.; Wiewiora, R. P.; Meng, F.; Babault, N.; Ma, A.; Yu, W.; Qian, K.; Hu, H.; Zou, H.; Wang, J.; et al. The dynamic conformational landscape of the protein methyltransferase SETD8. *Elife* **2019**, *8*, No. e45403.

(7) Miller, M. D.; Phillips, G. N. Moving beyond static snapshots: Protein dynamics and the Protein Data Bank. *J. Biol. Chem.* **2021**, *296*, 100749.

(8) Fraser, J. S.; Murcko, M. A. Structure is beauty, but not always truth. *Cell* **2024**, *187*, 517−520.

(9) Smith, L. G.; Novak, B.; Osato, M.; Mobley, D. L.; Bowman, G. R. PopShift: A Thermodynamically Sound Approach to Estimate Binding Free Energies by Accounting for Ligand-Induced Population Shifts from a Ligand-Free Markov State Model. *J. Chem. Theory Comput.* **2024**, *20*, 1036−1050.

(10) Jones, D.; Kim, H.; Zhang, X.; Zemla, A.; Stevenson, G.; Bennett, W. F. D.; Kirshner, D.; Wong, S. E.; Lightstone, F. C.; Allen, J. E. Improved Protein−Ligand Binding Affinity Prediction with Structure-Based Deep Fusion Inference. *J. Chem. Inf. Model.* **2021**, *61*, 1583−1592.

(11) Nam, K.; Wolf-Watz, M. Protein dynamics: The future is bright and complicated. *Struct. Dynam.* **2023**, *10*, 014301.

(12) Boehr, D. D.; McElheny, D.; Dyson, H. J.; Wright, P. E. The Dynamic Energy Landscape of Dihydrofolate Reductase Catalysis. *Science* **2006**, *313*, 1638−1642.

(13) Sekhar, A.; Kay, L. E. An NMR View of Protein Dynamics in Health and Disease. *Annu. Rev. Biophys.* **2019**, *48*, 297−319.

(14) Agarwal, P. K.; Doucet, N.; Chennubhotla, C.; Ramanathan, A.; Narayanan, C. Conformational Sub-states and Populations in Enzyme Catalysis. *Methods Enzymol.* **2016**, *578*, 273−297.

(15) Roy, R.; Hohng, S.; Ha, T. A practical guide to single-molecule FRET. *Nat. Methods* **2008**, *5*, 507−516.

(16) Ha, T.; Enderle, T.; Ogletree, D. F.; Chemla, D. S.; Selvin, P. R.; Weiss, S. Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6264−6268.

(17) 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.

(18) Förster, T. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys.* **1948**, *437*, 55−75.

(19) Stryer, L. Fluorescence Energy Transfer as a Spectroscopic Ruler. *Annu. Rev. Biochem.* **1978**, *47*, 819−846.

(20) Agam, G.; Gebhardt, C.; Popara, M.; Mächtel, R.; Folz, J.; Ambrose, B.; Chamachi, N.; Chung, S. Y.; Craggs, T. D.; de Boer, M.; et al. Reliability and accuracy of single-molecule FRET studies for characterization of structural dynamics and distances in proteins. *Nat. Methods* **2023**, *20*, 523−535.

(21) Mazal, H.; Haran, G. Single-molecule FRET methods to study the dynamics of proteins at work. *Curr. Opin. Biomed. Eng.* **2019**, *12*, 8−17.

(22) Lerner, E.; Barth, A.; Hendrix, J.; Ambrose, B.; Birkedal, V.; Blanchard, S. C.; Börner, R.; Sung Chung, H.; Cordes, T.; Craggs, T. D.; et al. FRET-based dynamic structural biology: Challenges, perspectives and an appeal for open-science practices. *Elife* **2021**, *10*, No. e60416.
(23) Lerner, E.; Cordes, T.; Ingargiola, A.; Alhadid, Y.; Chung, S.; Michalet, X.; Weiss, S. Toward dynamic structural biology: Two decades of single-molecule Förster resonance energy transfer. *Science* **2018**, *359*, No. eaan1133.
(24) Lee, J.; Lee, S.; Ragunathan, K.; Joo, C.; Ha, T.; Hohng, S. Single-molecule Four-color FRET. *Angew. Chem., Int. Ed. Engl.* **2010**, *49*, 9922−9925.
(25) Alston, J. J.; Soranno, A.; Holehouse, A. S. Integrating single-molecule spectroscopy and simulations for the study of intrinsically disordered proteins. *Methods* **2021**, *193*, 116−135.
(26) Montepietra, D.; Tesei, G.; Martins, J. M.; Kunze, M. B. A.; Best, R. B.; Lindorff-Larsen, K. FRETpredict: a Python package for FRET efficiency predictions using rotamer libraries. *Commun. Biol.* **2024**, *7*, 298.
(27) Zheng, W.; Zerze, G. H.; Borgia, A.; Mittal, J.; Schuler, B.; Best, R. B. Inferring properties of disordered chains from FRET transfer efficiencies. *J. Chem. Phys.* **2018**, *148*, 367a.
(28) Meng, F.; Bellaiche, M. M.; Kim, J. Y.; Zerze, G. H.; Best, R. B.; Chung, H. S. Highly Disordered Amyloid-β Monomer Probed by Single-Molecule FRET and MD Simulation. *Biophys. J.* **2018**, *114*, 870−884.
(29) Peran, I.; Holehouse, A. S.; Carrico, I. S.; Pappu, R. V.; Bilsel, O.; Raleigh, D. P. Unfolded states under folding conditions accommodate sequence-specific conformational preferences with random coil-like dimensions. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 12301−12310.
(30) Bottaro, S.; Bengtsen, T.; Lindorff-Larsen, K. Integrating Molecular Simulation and Experimental Data: A Bayesian/Maximum Entropy Reweighting Approach. In *Structural Bioinformatics: Methods and Protocols*; Gáspári, Z., Ed.; Springer US: New York, NY, 2020; pp 219−240.
(31) Köfinger, J.; Różycki, B.; Hummer, G. Inferring Structural Ensembles of Flexible and Dynamic Macromolecules Using Bayesian, Maximum Entropy, and Minimal-Ensemble Refinement Methods. In *Biomolecular Simulations: Methods and Protocols*; Bonomi, M., Camilloni, C., Eds.; Springer: New York, NY, 2019; pp 341−352.
(32) Woźniak, A. K.; Schröder, G. F.; Grubmüller, H.; Seidel, C. A. M.; Oesterhelt, F. Single-molecule FRET measures bends and kinks in DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 18337−18342.
(33) Lehmann, K.; Felekyan, S.; Kühnemuth, R.; Dimura, M.; Tóth, K.; Seidel, C. A. M.; Langowski, J. Dynamics of the nucleosomal histone H3 N-terminal tail revealed by high precision single-molecule FRET. *Nucleic Acids Res.* **2020**, *48*, 1551−1571.
(34) Kalinin, S.; Peulen, T.; Sindbert, S.; Rothwell, P. J.; Berger, S.; Restle, T.; Goody, R. S.; Gohlke, H.; Seidel, C. A. M. A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat. Methods* **2012**, *9*, 1218−1225.
(35) Walczewska-Szewc, K.; Corry, B. Accounting for dye diffusion and orientation when relating FRET measurements to distances: three simple computational methods. *Phys. Chem. Chem. Phys.* **2014**, *16*, 12317−12326.
(36) Sindbert, S.; Kalinin, S.; Nguyen, H.; Kienzler, A.; Clima, L.; Bannwarth, W.; Appel, B.; Müller, S.; Seidel, C. A. M. Accurate Distance Determination of Nucleic Acids via Förster Resonance Energy Transfer: Implications of Dye Linker Length and Rigidity. *J. Am. Chem. Soc.* **2011**, *133*, 2463−2480.
(37) Merchant, K. A.; Best, R. B.; Louis, J. M.; Gopich, I. V.; Eaton, W. A. Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1528−1533.
(38) Dimura, M.; Peulen, T. O.; Hanke, C. A.; Prakash, A.; Gohlke, H.; Seidel, C. A. Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems. *Curr. Opin. Struct. Biol.* **2016**, *40*, 163−185.
(39) Graen, T.; Hoefling, M.; Grubmüller, H. AMBER-DYES: Characterization of Charge Fluctuations and Force Field Parameterization of Fluorescent Dyes for Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 5505−5512.
(40) Speelman, A. L.; Muñoz-Losa, A.; Hinkle, K. L.; VanBeek, D. B.; Mennucci, B.; Krueger, B. P. Using Molecular Dynamics and Quantum Mechanics Calculations To Model Fluorescence Observables. *J. Phys. Chem. A* **2011**, *115*, 3997−4008.
(41) Hoefling, M.; Lima, N.; Haenni, D.; Seidel, C. A. M.; Schuler, B.; Grubmüller, H. Structural Heterogeneity and Quantitative FRET Efficiency Distributions of Polyprolines through a Hybrid Atomistic Simulation and Monte Carlo Approach. *PLoS One* **2011**, *6*, No. e19791.
(42) Rauscher, S.; Gapsys, V.; Gajda, M. J.; Zweckstetter, M.; de Groot, B. L.; Grubmüller, H. Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field: A Comparison to Experiment. *J. Chem. Theory Comput.* **2015**, *11*, 5513−5524.
(43) Tomov, T. E.; Tsukanov, R.; Masoud, R.; Liber, M.; Plavner, N.; Nir, E. Disentangling Subpopulations in Single-Molecule FRET and ALEX Experiments with Photon Distribution Analysis. *Biophys. J.* **2012**, *102*, 1163−1173.
(44) Barth, A.; Opanasyuk, O.; Peulen, T. O.; Felekyan, S.; Kalinin, S.; Sanabria, H.; Seidel, C. A. M. Unraveling multi-state molecular dynamics in single-molecule FRET experiments. I. Theory of FRET-lines. *J. Chem. Phys.* **2022**, *156*, 141501.
(45) Opanasyuk, O.; Barth, A.; Peulen, T. O.; Felekyan, S.; Kalinin, S.; Sanabria, H.; Seidel, C. A. M. Unraveling multi-state molecular dynamics in single-molecule FRET experiments. II. Quantitative analysis of multi-state kinetic networks. *J. Chem. Phys.* **2022**, *157*, 031501.
(46) Pirchi, M.; Tsukanov, R.; Khamis, R.; Tomov, T. E.; Berger, Y.; Khara, D. C.; Volkov, H.; Haran, G.; Nir, E. Photon-by-Photon Hidden Markov Model Analysis for Microsecond Single-Molecule FRET Kinetics. *J. Phys. Chem. B* **2016**, *120*, 13065−13075.
(47) Gopich, I. V.; Szabo, A. Decoding the Pattern of Photon Colors in Single-Molecule FRET. *J. Phys. Chem. B* **2009**, *113*, 10965−10973.
(48) Chung, H. S.; Gopich, I. V. Fast single-molecule FRET spectroscopy: theory and experiment. *Phys. Chem. Chem. Phys.* **2014**, *16*, 18644−18657.
(49) Torella, J. P.; Holden, S. J.; Santoso, Y.; Hohlbein, J.; Kapanidis, A. N. Identifying Molecular Dynamics in Single-Molecule FRET Experiments with Burst Variance Analysis. *Biophys. J.* **2011**, *100*, 1568−1577.
(50) Chung, H. S.; Cellmer, T.; Louis, J. M.; Eaton, W. A. Measuring ultrafast protein folding rates from photon-by-photon analysis of single molecule fluorescence trajectories. *Chem. Phys.* **2013**, *422*, 229−237.
(51) Harris, P. D.; Narducci, A.; Gebhardt, C.; Cordes, T.; Weiss, S.; Lerner, E. Multi-parameter photon-by-photon hidden Markov modeling. *Nat. Commun.* **2022**, *13*, 1000.
(52) Laurence, T. A.; Kong, X.; Jäger, M.; Weiss, S. Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17348−17353.
(53) Nettels, D.; Hoffmann, A.; Schuler, B. Unfolded Protein and Peptide Dynamics Investigated with Single-Molecule FRET and Correlation Spectroscopy from Picoseconds to Seconds. *J. Phys. Chem. B* **2008**, *112*, 6137−6146.
(54) Stuchell-Brereton, M. D.; Zimmerman, M. I.; Miller, J. J.; Mallimadugula, U. L.; Incicco, J. J.; Roy, D.; Smith, L. G.; Cubuk, J.; Baban, B.; DeKoster, G. T.; et al. Apolipoprotein E4 has Extensive Conformational Heterogeneity in Lipid-Free and Lipid-Bound Forms. *Proc. Natl. Acad. Sci. U.S.A.* **2023**, *120*, No. e2215371120.
(55) Sanabria, H.; Rodnin, D.; Hemmen, K.; Peulen, T. O.; Felekyan, S.; Fleissner, M. R.; Dimura, M.; Koberling, F.; Kühnemuth, R.; Hubbell, W.; et al. Resolving dynamics and function of transient states in single enzyme molecules. *Nat. Commun.* **2020**, *11*, 1231.
(56) Müller-Späth, S.; Soranno, A.; Hirschfeld, V.; Hofmann, H.; Rüegger, S.; Reymond, L.; Nettels, D.; Schuler, B. Charge Interactions

Can Dominate the Dimensions of Intrinsically Disordered Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 14609−14614.
(57) Acuna, G. P.; Möller, F. M.; Holzmeister, P.; Beater, S.; Lalkens, B.; Tinnefeld, P. Fluorescence Enhancement at Docking Sites of DNA-Directed Self-Assembled Nanoantennas. *Science* **2012**, *338*, 506−510.
(58) Farooq, S.; Hohlbein, J. Camera-based single-molecule FRET detection with improved time resolution. *Phys. Chem. Chem. Phys.* **2015**, *17*, 27862−27872.
(59) Bohlen, J.; Cuartero-González, Á.; Pibiri, E.; Ruhlandt, D.; Fernández-Domínguez, A. I.; Tinnefeld, P.; Acuna, G. P. Plasmon-assisted Förster resonance energy transfer at the single-molecule level in the moderate quenching regime. *Nanoscale* **2019**, *11*, 7674−7681.
(60) Nüesch, M. F.; Ivanović, M. T.; Claude, J. B.; Nettels, D.; Best, R. B.; Wenger, J.; Schuler, B. Single-molecule Detection of Ultrafast Biomolecular Dynamics with Nanophotonics. *J. Am. Chem. Soc.* **2022**, *144*, 52−56.
(61) Grabenhorst, L.; Sturzenegger, F.; Hasler, M.; Schuler, B.; Tinnefeld, P. Single-Molecule FRET at 10 MHz Count Rates. *J. Am. Chem. Soc.* **2024**, *146*, 3539−3544.
(62) Girodat, D.; Pati, A. K.; Terry, D. S.; Blanchard, S. C.; Sanbonmatsu, K. Y. Quantitative comparison between sub-millisecond time resolution single-molecule FRET measurements and 10-second molecular simulations of a biosensor protein. *PLoS Comput. Biol.* **2020**, *16*, No. e1008293.
(63) Matsunaga, Y.; Sugita, Y. Linking time-series of single-molecule experiments with molecular dynamics simulations by machine learning. *Elife* **2018**, *7*, No. e32668.
(64) Wolf, S.; Sohmen, B.; Hellenkamp, B.; Thurn, J.; Stock, G.; Hugel, T. Hierarchical dynamics in allostery following ATP hydrolysis monitored by single molecule FRET measurements and MD simulations. *Chem. Sci.* **2021**, *12*, 3350−3359.
(65) Hoefling, M.; Grubmüller, H. In silico FRET from simulated dye dynamics. *Comput. Phys. Commun.* **2013**, *184*, 841−852.
(66) Frieden, C.; Wang, H.; Ho, C. M. W. *A Mechanism for Lipid Binding to apoE and the Role of Intrinsically Disordered Regions Coupled to Domain−Domain Interactions*; Proceedings of the National Academy of Sciences, 2017; Vol. *114*, pp 6292−6297.
(67) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, E4758−E4766.
(68) Husic, B. E.; Pande, V. S. Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.* **2018**, *140*, 2386−2396.
(69) Konovalov, K. A.; Unarta, I. C.; Cao, S.; Goonetilleke, E. C.; Huang, X. Markov State Models to Study the Functional Dynamics of Proteins in the Wake of Machine Learning. *JACS Au* **2021**, *1*, 1330−1341.
(70) Pande, V. S.; Beauchamp, K.; Bowman, G. R. Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* **2010**, *52*, 99−105.
(71) Suárez, E.; Wiewiora, R. P.; Wehmeyer, C.; Noé, F.; Chodera, J. D.; Zuckerman, D. M. What Markov State Models Can and Cannot Do: Correlation versus Path-Based Observables in Protein-Folding Models. *J. Chem. Theory Comput.* **2021**, *17*, 3119−3133.
(72) Zimmerman, M. I.; Bowman, G. R. FAST Conformational Searches by Balancing Exploration/Exploitation Trade-Offs. *J. Chem. Theory Comput.* **2015**, *11*, 5747−5757.
(73) Best, R. B.; Zheng, W.; Mittal, J. Balanced Protein−Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theory Comput.* **2014**, *10*, 5113−5124.
(74) Lincoff, J.; Haghighatlari, M.; Krzeminski, M.; Teixeira, J. M. C.; Gomes, G. N. W.; Gradinaru, C. C.; Forman-Kay, J. D.; Head-Gordon, T. Extended experimental inferential structure determination method in determining the structural ensembles of disordered protein states. *Commun. Chem.* **2020**, *3*, 74.
(75) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(76) Best, R. B.; Mittal, J. Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse. *J. Phys. Chem. B* **2010**, *114*, 14916−14923.
(77) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017**, *13*, No. e1005659.
(78) 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.
(79) Bell, J. A.; Wilson, K. P.; Zhang, X.; Faber, H. R.; Nicholson, H.; Matthews, B. W. Comparison of the crystal structure of bacteriophage T4 lysozyme at low, medium, and high ionic strengths. *Proteins: Struct., Funct., Bioinf.* **1991**, *10*, 10−21.
(80) 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.
(81) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.
(82) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.
(83) McGibbon, R. T.; Beauchamp, K.; Harrigan, M.; Klein, C.; Swails, J.; Hernández, C.; Schwantes, C.; Wang, L. P.; Lane, T.; Pande, V. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528−1532.
(84) Porter, J. R.; Zimmerman, M. I.; Bowman, G. R. Enspara: Modeling molecular ensembles with scalable data structures and parallel computing. *J. Chem. Phys.* **2019**, *150*, 044108.
(85) *Matplotlib. A 2D Graphics Environment*; IEEE Journals & Magazine; IEEE Xplore. https://ieeexplore.ieee.org/document/4160265.