JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

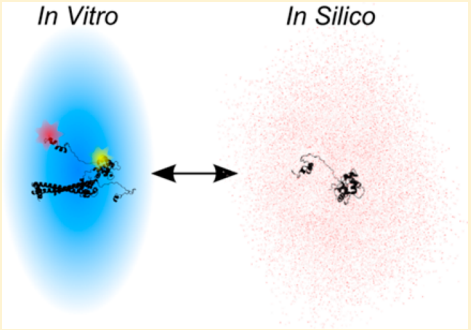
Conformation and Dynamics of the Troponin I C-Terminal Domain: Combining Single-Molecule and Computational Approaches for a Disordered Protein Region

Lauren Ann Metskas[†,§] and Elizabeth Rhoades*[,†,‡,§,∥]

[†]Department of Molecular Biophysics and Biochemistry, [‡]Department of Physics, and [§]Integrated Graduate Program in Physical and Engineering Biology, Yale University, New Haven, Connecticut 06520, United States

Supporting Information

ABSTRACT: In recent years, single-molecule Förster resonance energy transfer (smFRET) has emerged as a critical and flexible tool in structural biology, particularly in the study of highly dynamic regions and molecular assemblies. The usefulness of smFRET can be further extended by combining it with computational approaches, marrying the coarse-grained experimental data with higher-resolution in silico calculations. Here we use smFRET to determine six pairwise distances within the intrinsically disordered C-terminal domain of the troponin I subunit (TnI_C) of the cardiac troponin complex. We used published conflicting structures of TnI_C as starting models for molecular dynamics simulations, which were validated through successful comparison with smFRET measurements before extracting information on conformational dynamics. We find that pairwise distances between residues fluctuate widely in silico, but simulations are generally in good agreement with longer time scale smFRET measurements after averaging across time. Finally, Monte Carlo simulations establish that the lower-energy conformers of TnI_C are indeed varied, but that the highest-sampled clusters resemble the published, conflicting models. In this way, we find that the controversial structures are simply stabilized local minima of this dynamic region, and a population including all three would still be consistent with spectroscopic measurements. Taken together, the combined approaches described here allow us to critically evaluate existing models of TnI_C, giving insight into the conformation and dynamics of TnI_C's disordered state prior to its probable disorder−order transition. Moreover, they provide a framework for combining computational and experimental methods with different time scales for the study of disordered and dynamic protein states.



INTRODUCTION

Due to their importance in protein−protein interactions and numerous human diseases, a growing body of research is focused on intrinsically disordered proteins (IDPs). These proteins are characterized by a broad sampling of conformational space and poorly defined or unstable secondary and tertiary structures. This conformational heterogeneity and fast sampling of relevant states can stymie traditional high-resolution methods. Single-molecule Förster resonance energy transfer (smFRET), however, has proven to be a useful approach for characterizing conformations and dynamics of IDPs. The single-molecule nature of the technique allows for elimination of artifacts from imperfect labeling, photobleaching, aggregation, and dissociation of protein complexes. Furthermore, smFRET measurements can be made in a wide variety of experimental milieu as they are not limited by the size, oligomerization state, or interaction partners of the molecule of interest.[1−4] When combined with simulations capable of extending the analysis beyond the data set of pairwise distances, this approach becomes even more powerful.[4]

The use of coarse-grain experimental data to restrain simulations has been applied to many structural biology methods,[5] with NMR being perhaps the most established.[6] More recently, this approach has been adapted for the combination of smFRET measurements with molecular dynamics (MD) and Monte Carlo simulations to generate low-resolution models of protein−protein and protein−nucleic acid complexes.[7−9] These approaches typically apply smFRET-derived pairwise distances to either restrain sampling during simulations or to screen outputs for structures consistent with experimental measurements. Notably, the most successful applications involve docking proteins with stable conformations and known high-resolution structures.

Inherent in this methodology is the assumption that a good computational output satisfies all experimental constraints simultaneously. While this assumption may be justified in rigid-body docking of folded regions without large conformational changes, it is unlikely to be accurate in cases of highly dynamic proteins. IDPs experience conformational fluctuations on time scales faster than the resolution of most experimental measurements and cannot be treated as rigid bodies. Moreover, the lack of high-resolution structures of these proteins and their

Received: April 29, 2015
Published: September 1, 2015

DOI: 10.1021/jacs.5b04471
J. Am. Chem. Soc. 2015, 137, 11962−11969

conformational sensitivity to environmental conditions can complicate validation of simulated results.

While many studies of IDPs focus on model peptides or fully disordered or denatured proteins, a significant fraction of disordered protein sequences are regions that are adjacent to or bounded by globular domains. Although these intrinsically disordered regions (IDRs) are often truncated to facilitate biophysical characterization, they can be critical in the function of protein complexes. IDRs present unique opportunities for the study of dynamic regions both in silico and in vitro. Their close proximity to folded regions (often well characterized) gives access to more stringent functional and structural controls than is typical for many IDP studies. IDRs also present excellent models for the study of coupled binding and folding due to their requisite proximity to their binding partner.

Here, we report on an IDR present in a heterotrimeric complex: the C-terminal domain of troponin I (TnI_C) within the human cardiac troponin complex of troponin T (TnT), C (TnC), and I (TnI) (Figure 1A). The troponin complex docks

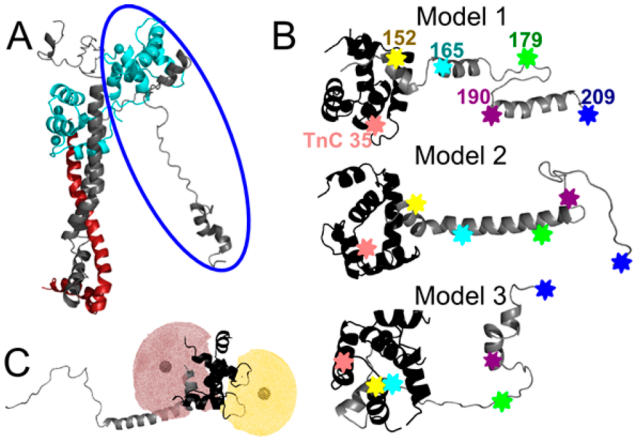


Figure 1. Models of TnI_C. (A) Troponin complex, with TnC (teal), TnI (gray), and the T2 domain of TnT (red). Blue circle shows the fragment used for simulations.[19] (B) Models 1−3 (top to bottom) all contain the N-terminal head of TnC (black, residues 1−88) and TnI_C (gray, residues 147−210). Residues used as labeling positions are marked; measured residue pairs are listed in the text. (C) Example accessible volume calculation of fluorophore positions for the control pair TnC 35/TnI_C 152.[8]

onto actin thin filaments with tropomyosin, functioning as a molecular switch to regulate actomyosin binding in response to changes in calcium concentration. Under the low-calcium conditions found in relaxed muscle, TnI_C is bound to the thin filament where it is likely at least partially folded.[10] Increased calcium leads to a series of conformational changes in troponin that ultimately result in decreased contact between TnI_C and the thin filament,[11] culminating in the complete dislodging of TnI_C upon myosin binding during muscle contraction.[12,13] At this point, TnI_C is likely disordered, though residual helical structure has been reported.[14] Muscle relaxation thus requires coupled binding and folding of TnI_C as it returns to its thin filament-bound state. In the cardiac troponin complex, point mutations throughout TnI_C are associated with familial hypertrophic cardiomyopathy, highlighting the critical function associated with these conformational changes.[15,16]

TnI_C presents a unique case because there are already multiple high-resolution structural models of the region when dissociated from the thin filament (Figure 1B). However, these models conflict with each other, as well as with NMR characterizations indicating high conformational heterogeneity.[17−19] There is currently no consensus on which model is most accurate or even whether TnI_C is truly intrinsically disordered. It appears that frequently a preferred model is simply chosen.[14,20,21]

In this study, we take advantage of TnI_C's conflicting structural models to run short MD simulations of the different proposed conformations of TnI_C. We validate the unrestrained simulations by comparison with smFRET measurements before examining differences and similarities in the dynamic profiles, gaining insight into the effect of the different secondary structure signatures present in the models. Finally, Monte Carlo (MC) simulations are employed to increase conformational sampling and remove starting model bias, generating a library of possible TnI_C conformations. This combination of experimental and computational approaches allows us to cross the numerous time scales of both our own and others' work on TnI_C, drawing the disparate models of this important IDR into one cohesive conformational behavior.

RESULTS

Published Structural Models of TnI_C Are Inconsistent with smFRET. smFRET was used to investigate the disordered state of TnI_C in the calcium-bound troponin complex. We chose five positions on TnI_C (D152, E165, E179, D190, and E209) and one position on TnC (C35) for labeling, for a total data set of seven pairwise distance measurements (TnC 35/TnI_C 152, TnI_C 152/179, 152/190, 152/209, 165/190, 165/209, and 179/209; Figure 1B). smFRET measurements of diffusing troponin complexes were made using ∼75 pM concentrations of dual-labeled complexes in the presence of 6 nM unlabeled complex to improve stability (Figure S1). The primary smFRET histogram peaks were fitted with Gaussian distributions, with the average mean and width (standard deviation) from repeated measurements used for further analysis (Figures 2 and 3, left panels). Additional peaks in

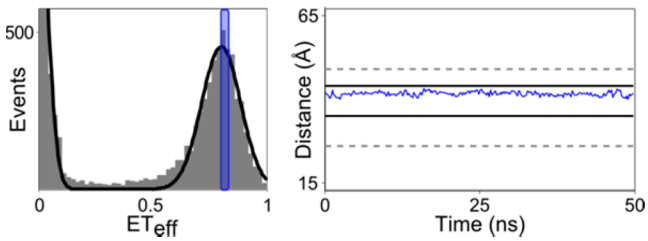


Figure 2. Comparison between smFRET and MD simulations of the TnC 35/TnI_C 152 construct. A representative smFRET histogram on the left is overlaid by the average ET_{eff} calculated from pairwise distances in the MD simulations (blue), with the bar width representing one standard deviation from the mean of the MD pairwise distances. A representative simulated pairwise distance trajectory on the right is overlaid by lines designating one (black, solid) and two (gray, dashed) standard deviations from the mean distance calculated from the ET_{eff} distribution.

the histograms were experimentally assigned as contributions from donor-only signal or complex dissociation and excluded from further analysis (Figures S2, S3).

To compare our measurements to the published NMR and crystal structures, we first built models in silico based on the published partial structures. Structures were converted to the human cardiac isoform through homology modeling,[22,23] filling

DOI: 10.1021/jacs.5b04471
J. Am. Chem. Soc. 2015, 137, 11962−11969

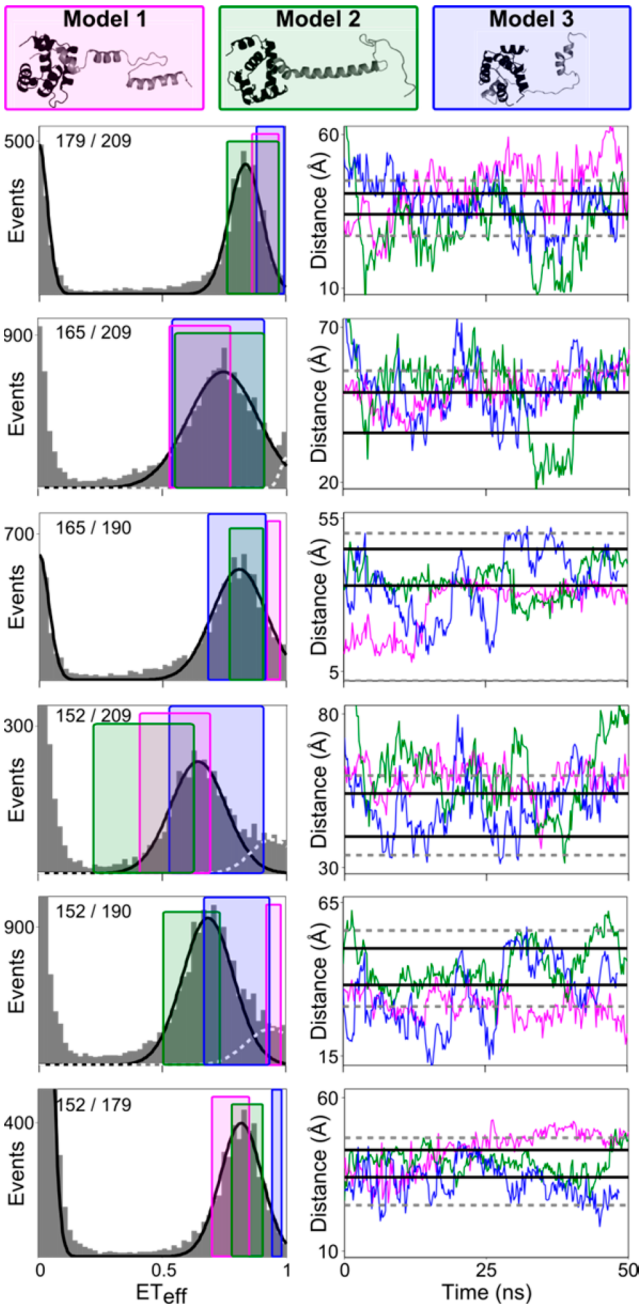


Figure 3. Comparison between smFRET pairwise distances and MD simulations of TnI_C. smFRET histograms (left panels; constructs are indicated in upper left of these plots) are overlaid by the average pairwise distances calculated from MD simulations, with colors indicating the starting model used for the group of simulations and the overlay width representing one standard deviation. White dashed lines fit the monomeric TnI peak, which was excluded from analysis (see the SI for details). The simulated pairwise distance trajectories (right panels) are overlaid by dotted lines representing one (black solid) and two (gray dashed) standard deviations from the smFRET mean. The starting models used for simulations are shown at the top of the figure.

in any missing residues with minimized antiparallel beta strand. Fluorophore positions were estimated using accessible volume (AV) models (Figure 1C).[8,24] The calculated interdye distances of the models were then compared to those measured by smFRET (Table 1; see the Supporting Information (SI) for details). We found very good agreement between the crystal structure and the smFRET construct TnC 35/TnI_C 152, which probed the folded/interaction domain of TnC and TnI_C. However, in stark contrast, no model of the dynamic C-terminal domain of TnI was consistent with more than half of the smFRET measurements (Table 1).

Table 1. Predicted Interdye Distances for Models Compared with Interdye Distances Measured by smFRET

residue pair	model 1	model 2	model 3	smFRET
TnC 35/TnI 152	39 Å	39 Å	39 Å	40 (5) Å
TnI 152−179	41 Å	42 Å	18 Å	39 (5) Å
TnI 152−209	67 Å	79 Å	45 Å	46 (6) Å
TnI 179−209	29 Å	64 Å	25 Å	38 (4) Å
TnI 165−209	50 Å	83 Å	53 Å	42 (7) Å
TnI 152−190	42 Å	46 Å	22 Å	44 (6) Å
TnI 165−190	22 Å	46 Å	28 Å	39 (5) Å

Short MD Simulations Can Bridge the Time Scale Gap between Experimental Techniques for Dynamic Proteins. The observation that measured smFRET distances are inconsistent with the models derived from published structures motivated a need to bridge the substantial gap in time scales between atomic-resolution structural characterization and spectroscopy. The structural models are snapshots of a highly dynamic region, while smFRET measurements are binned over 1 ms—sufficient time to average conformational dynamics. If a protein experiences conformational fluctuation on the time scale of short MD simulations, these provide a mechanism for deriving dynamic ensembles from the structural models. These ensembles allow a more meaningful comparison with smFRET measurements.

Our goal with the MD simulations was to probe dynamics while maintaining the unique signatures of the different structural models. Because TnI_C experiences rapid dynamic fluctuations in silico, simulations were run for only 50 ns to avoid interconversion between models. While overall sampling was intentionally restricted by this time limit, local sampling was increased by running five independent replicas of each model. We chose the AMBER99sb force field with a TIP4P-Ew water model, commonly used for IDPs.[2,14,25−28] We validated that this combination maintains the structure of folded troponin regions without artificially compacting its disordered regions (Figure S4; see the SI for details of other force fields and water models considered). Pairwise distances in MD simulations were converted to interdye distances using AV calculations.[8] All starting models demonstrated large sampling of radius of gyration and maintained helical propensities in areas containing helices at the start of the simulation.

For the TnC 35/TnI_C 152 construct, which probes a folded region, the simulated pairwise distances are in excellent agreement with the smFRET distances (Figure 2).[19] Simulations of this folded region displayed a fairly static conformation, as seen by the relatively small magnitude of the fluctuations in the simulation trajectory (Figure 2, right panel). To compare the simulations with the smFRET experiments, we used the peak widths (the Gaussian standard deviation of the population) as the largest estimate of error in fixing the mean ET_{eff} (a discussion of error calculation is included in the SI). Lines corresponding to one (1σ) and two (2σ) standard deviations from the Gaussian fit of the ET_{eff} distribution were plotted along with the simulated pairwise distance trajectory. The trajectory remains within the 1σ lines (Figure 2, right panel).

DOI: 10.1021/jacs.5b04471
J. Am. Chem. Soc. 2015, 137, 11962−11969

To address the difference in time scale between measurements and simulations, we averaged each pairwise distance over the final 25 ns of the MD simulation, then averaged across the five replicates. These simulated distance averages were converted to predicted ET_{eff} values (see SI for details) and overlaid onto the smFRET histograms to compare them with the measured ET_{eff} values (Figure 2, left panel, blue overlay). The standard deviation of simulated distances, shown by the width of the colored overlay, was low for this folded region; simulations only sampled conformational space consistent with the median bin of the smFRET histogram.

In marked contrast to the stable pairwise distance of the folded region, the pairwise distances within TnI_C display extensive sampling of conformational space, passing outside even the 2σ bounds to sample both more compact and more extended conformations (Figure 3, right panels). This behavior is true even for areas with helical structure. Of particular interest is the fact that the sampling is different for each pairwise distance, making it unlikely that any group of pairwise distances will be within even our most liberal estimate of the measured ET_{eff} value (the 2σ bounds) at any given time. To quantify this, we performed an analysis of all simulations from 25 to 50 ns (5 simulations, 25 ns per simulation, 200 ps per frame for a total of 1250 frames), calculating the percentage of frames in which all smFRET measurements would be met within one or two standard deviations. With this analysis, we found that no MD frame ever samples a conformation in which every smFRET measurement is consistent within one standard deviation (Table 2); the "ideal" structure as defined by smFRET is in fact never sampled. When we relaxed the criteria to two standard deviations, the percentage of frames meeting all constraints increased; this increase was specific to the model, ranging from approximately 6−27% of frames (Table 2). However, the TnI_C conformations in MD runs were so dynamic that even with these modest criteria, the majority of frames still had at least one pairwise distance out of bounds.

Table 2. Fraction of Simulation Frames Which Simultaneously Satisfy all smFRET Constraints within One or Two Standard Deviations Based on Gaussian Fits to the Data

bounds	model 1	model 2	model 3
1 σ	0%	0%	0%
2 σ	5.6%	27.0%	17.2%

While instantaneous snapshots of the dynamic TnI_C trajectories were inconsistent with smFRET measurements, and trajectories frequently explored areas outside our experimentally determined boundaries, averaging pairwise distances over time and across replicates brought simulations and experiments into far greater agreement (Figure 3, left panel). All ET_{eff} values calculated from time-averaged pairwise distances approached smFRET measurements for all constructs, regardless of the starting model. In fact, most averaged simulated pairwise distances overlap with one standard deviation of the mean in the smFRET measurements (Figure 3, left panels). These results emphasize the critical importance of considering time scales when comparing experiments and/or simulations, particularly for disordered proteins and regions.

We can use these results to gain greater insight into the structure of TnI_C by considering regions where there is particularly good or particularly poor agreement between the time-averaged simulations and the smFRET measurements. It is important to note that the three starting models mostly agree with smFRET measurements following conformational averaging, despite preserving different secondary structure propensities. This observation argues that all may be equally valid as snapshots of a highly dynamic region. However, there are three instances of poor performance: model 1 (pink) fares poorly in smFRET comparisons focusing on the central region of TnI_C in which it is more compact than the other models (Figure 3, constructs 152/190 and 165/190), and model 3 (blue) performs poorly in the 152/179 comparison where the absence of a helix allows overcompaction (Figure 3). Considering the rest of the models and where they perform best, we can conclude that the average conformation of TnI_C is likely relatively extended and contains helical propensity in its N-terminal segment.

Negatively Correlated Motions Present in Areas with Helical Propensity in MD Starting Models. We computed cross-correlation matrices for each model (Figure 4A),

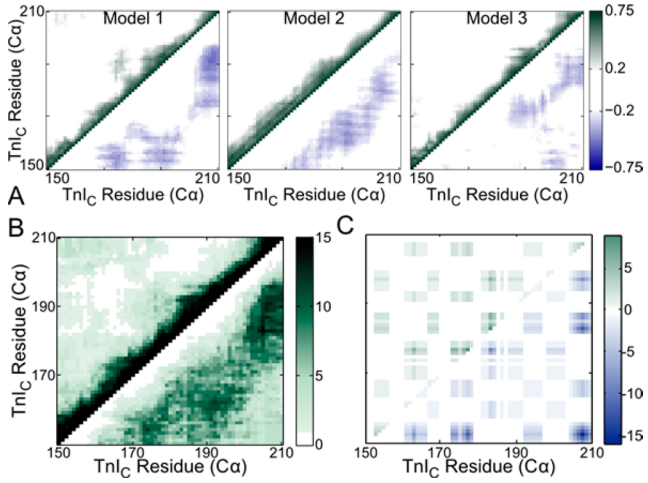


Figure 4. Correlated motions within TnI_C. (A) Cross-correlation was calculated over the final 25 ns of the simulation, then averaged between model replicates (models 1−3); maps are divided along the diagonal, with positive correlation on the upper left and negative correlation on the lower right. (B) The commonality matrix scores the number of separate simulations in which there is correlation at that location on the map; for example, the value of 14 in the bottom right half for positions 190/206 indicates that 14 simulations showed a correlation < −0.2 at this position. (C) Net charge was calculated on a six-residue sliding window and compared across the region. Relative attraction/repulsion is divided across the diagonal mirror plane, with all attractive calculations on the upper left and all repulsive calculations on the lower right.

identifying areas of the protein that move together (positive correlation) or in opposition (negative correlation). Similar to the pairwise distance analysis described above, the trajectories were integrated over the final 25 ns to generate a cross-correlation matrix, and the matrices were then averaged across the five replicates. While each model has a different cross-correlation pattern, there are clear similarities between the correlated motion patterns for the different models. All models show relatively little positively correlated motion but have substantial negative correlation evident throughout the region, consistent with a flexible, self-avoiding, extended chain.

We calculated a "commonality matrix" to determine which inter-residue relationships were preserved across the models

DOI: 10.1021/jacs.5b04471
J. Am. Chem. Soc. 2015, 137, 11962−11969

(Figure 4B). In this analysis, each simulation replicate of each model was analyzed independently, and each matrix element scored from 0 to 15 to indicate the number of simulations showing positive or negative correlation for that residue pair. According to this analysis, the strongest relationships are the correlations between the C-terminal end of TnI_C and the 185−195 region, 190−196/160−167, and 173−177/150−158 (Figure 4B). All three are negative correlations, and all involve segments that are helical in at least one starting model.

These relationships between regions with greater helical propensity (residues 169−176, 190−196) and other nonhelical areas of the chain could prove significant in the frequent order/disorder transitions of TnI_C during cardiac contraction cycles. We extended our analysis of the areas of high correlation, examining charge distribution to determine whether charge−charge repulsion could play a role in the pockets of negative correlation described above. This analysis uses a six-residue sliding window of net charge to generate a matrix of relative attraction or repulsion across different parts of the protein chain based on sequence alone (Figure 4C; see the SI for details).[29] We found high charge−charge repulsion in two of the areas with strong negative correlation (the C-terminal/185−195 correlation and the 173−177/150−158 correlation), providing some explanation for the source of this movement bias. Interestingly, the third area of anticorrelation, 190−196/160−167, occurs in an area of sequence space that can be expected to have very low charge−charge repulsion, indicating that another mechanism must be responsible for this motion. Because the 160−167 area is both immediately adjacent to the calcium-dependent switch peptide helix and has helical propensity in two starting models, it is interesting to speculate that this area of correlated motion could provide a mechanism for communication between TnC and TnI_C upon calcium binding.

Unbiased Monte Carlo Simulations Corroborate Models Used in MD Simulations. Monte Carlo (MC) simulations allow for greatly enhanced sampling, enabling us to generate the large libraries of conformers necessary for such a dynamic, flexible protein region. We performed MC simulations to determine conformational ensembles of TnI_C using CAMPARI software.[30] In contrast to the MD simulations, the MC simulations do not rely on a specific starting structure of the protein, which reduces the user input and enables visualization of any additional highly populated conformations that were not present in the original published models. Replica exchange simulations of TnI_C 161−210 were performed in duplicate, using OPLS-AA/L parametrization with the Absinth implicit solvation model.[31]

The MC simulations show an extended conformation for TnI_C with high asphericity and R_g (Figure 5), and with no persistent long-range contacts within 3.5 Å. This marked preference for extension is consistent with the anticorrelation of the MD simulations (Figure 4) and both the MD and smFRET pairwise distances. Although TnI_C is highly dynamic, this preference for extension limits its conformational sampling. We quantified this using conformational heterogeneity analysis,[32] which involves a comparison with a Flory random coil simulation set having the same sequence as the protein of interest. The results are scaled from 0 to 1, with higher numbers indicating higher homogeneity. TnI_C has a conformational heterogeneity of 0.60, which is similar to that of polyproline and even some folded peptides.[32] This result

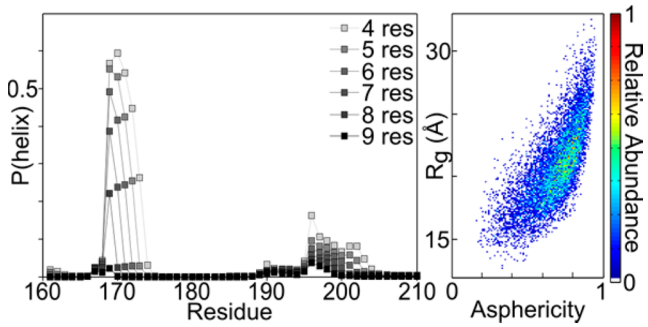


Figure 5. Monte Carlo simulation results for TnI_C. Left: MC simulations are consistent with the MD-based prediction of a short N-terminal helix similar in length and location to that of model 1, with over 50% of conformers having a four-residue helix in the 169−173 region. Helicity in the C-terminal region of the domain is diffuse and lower probability. Right: MC conformers populate an extended conformation with high asphericity and R_g.

indicates that while the extended state is dynamic with little secondary structure, its conformation is not random.

As predicted from the comparison of MD simulations and smFRET, the MC simulations show helical propensity in the N-terminal region. Over 50% of conformers have a helix at residue L169, and the majority of these helices span seven residues in length (Figures 5, S9, S10). This is consistent with the MD simulations, in which published structures frequently have helices in this region and coil immediately preceding it (Figure 6; a full analysis of the effect of the helix on global and local TnI_C conformation can be found in the SI). We also observed a

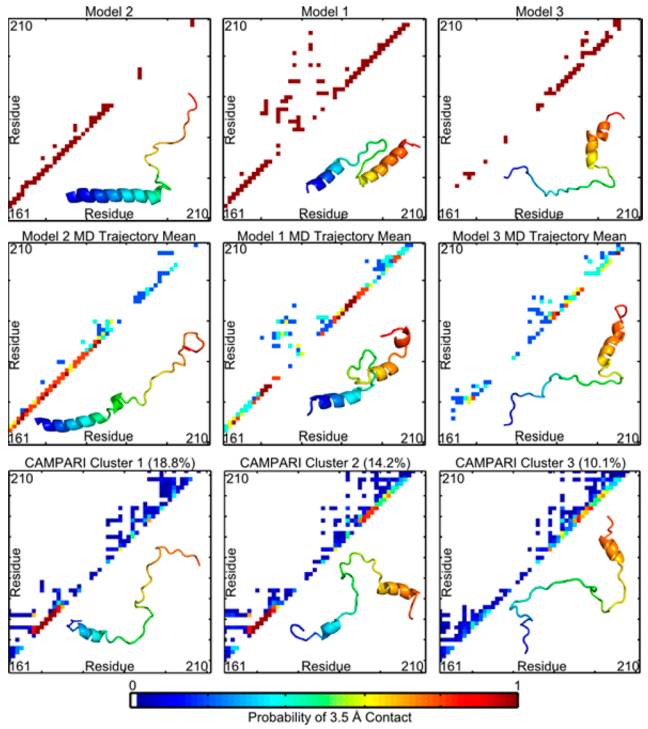


Figure 6. 3.5 Å contact maps for starting models (top panel), MD trajectories averaged over the 25−50 ns window (center panel), and three most-populated MC clusters (lower panel). Insets show the starting model of TnI_C (top panel), the TnI_C structure at 25 ns in the MD trajectory (center panel), or a representative snapshot from the MC cluster (lower panel).

DOI: 10.1021/jacs.5b04471
J. Am. Chem. Soc. 2015, 137, 11962−11969

prominent bend in the structure occurring around the 175−185 region, but no beta-turn propensities (Figure 6, insets).

MC outputs were clustered by 3.5 Å contact maps to determine whether certain conformations were more frequently represented in the simulation. Strikingly, even though no starting structure was used to generate the MC ensembles, the three most populated clusters resemble the three TnI_C models derived from crystallography and NMR structures (Figure 6). Clusters 1 and 3 of the MC are highly similar to models 2 and 3, respectively, and a simple untwisting of the central loop region brings cluster 2 and model 1 into greater resemblance. While the published models show longer helices than the MC outputs, this is explained by the strong temperature dependence of the N-terminal helix (Figure S9) and likely increased stability from crystal packing. Furthermore, the gap between the published models (Figure 6, top panel) and MC clusters (Figure 6, bottom panel) is narrowed if the published structures are permitted to relax and move. Contact maps of the 25−50 ns simulation window allow modest melting of helices and increased longer-range contacts (Figure 6, center panel), though the short time limits of the simulations does not permit shifting of helix positions. In these contact maps, the conformations populated by TnI_C can be seen to be similar across atomic-resolution experiments, MD simulations, and MC simulations. The presence of the three clusters in the MC work argues that all three published models likely represent stabilized conformations of this dynamic region.

DISCUSSION

In this paper, we demonstrate a combined experimental and computational approach to examining the conformation and dynamics of TnI_C. Each of the three techniques employed both complements and extends the others. smFRET experimentally determines pairwise distances for residue pairs, which can be used to validate MD simulations that provide a glimpse of correlated motions in the disordered domain, while MC simulations extend the sampling to define a library of potential conformers. Our results provide both a critical step toward understanding the functional mechanism of an important IDR and a framework for effectively combining smFRET measurements with simulations of dynamic proteins.

Using this combination of experiments and simulations, we were able to gain insight into the conformation and dynamics of TnI_C in the context of the calcium-bound state of the troponin complex. Our results provide evidence that this IDR is flexible, with a tendency to populate several distinct non-random conformations. The TnI_C domain of the troponin complex plays a critical role in its regulation of actin-myosin binding through a putative fly casting mechanism.[12,33] Our study of this system illuminates the possibility that an ensemble of structures may be responsible for the biological function of this protein complex, rather than a single conformation. Most broadly, in contrast to well-studied globular proteins, this work highlights the importance of considering dynamic ensembles as the relevant functional forms of IDPs and IDRs.

Specifically, our results indicate that the conformation of TnI_C is likely extended, with moderate helical propensity in the 169−175 region and weaker helical propensity closer to the C-terminus. While flexible and dynamic, the conformation is somewhat homogeneous in its extension: TnI_C does not have access to all conformational space. The asphericity of the domain is reflected in the preponderance of anticorrelation seen in the cross-correlation matrices, and is at least partially explained by charge repulsion (Figure 4). This observation is consistent with previous studies relating general charge distribution to IDP conformation.[34]

TnI_C has a specific area of interest surrounding residue 164. Adult mammalian cardiac TnI isoforms experience calcium desensitization at pH 6.2, which is relevant for cardiac acidosis and ischemia/reperfusion damage. It has been established that a single point mutation mimicking the neonatal TnI isoform, A164H, confers resistance to these challenges in vivo.[35] While the mechanism is still unclear as to how single point mutations in IDPs can have far-reaching functional effects, the finding of correlated motion between a residue cluster including A164 and the C-terminal helix area may provide some clue as to the mechanism behind this phenotypic shift.

Our results also provide a partial explanation of the controversy in the literature regarding TnI_C. While flexible, TnI_C has substantial conformational preferences, which could argue for a possibility of stabilizing one conformation over others. Furthermore, it is evident that the different models could arise from conformational sampling without requiring extensive domain movements. The MC simulations sample all three literature-derived MD starting models with similar frequency (and greater frequency than any other conformation), which can be seen clearly in the contact maps and structural snapshots (Figure 6). As the MC, MD, and experimental structure determination are all fully orthogonal techniques, this agreement is all the more striking, and proposes the hypothesis that all three models may be thermodynamically accessible as low-energy conformations. Finally, the overall extended nature of the domain could result in coarse experimental measurements being consistent either with conformational sampling about any of the three predominant conformations tested or with conformational sampling across them. This finding could be useful in other cases of conflicting structures: dissimilar structures may be more consistent with each other than first appearances would indicate, particularly in the case of dynamic proteins that could be predisposed by experimental conditions to favor one ensemble of conformers over another.

Finally, our combined experimental and computational approach is broadly applicable to studies of other disordered protein states. As was evident in Figure 2, the application of smFRET bounds to a simulation of a folded region would not negatively affect a simulation; "breathing" motions of the molecule are well within the 1σ smFRET bounds. However, our MD results suggest that the application of smFRET-based restraints to a dynamic protein may be misleading, even when used with the large windows for error that such measurements typically require. Specifically, simulations in good agreement with experimental data may not in fact sample structures meeting all these criteria simultaneously, as found for the model case investigated here (Figure 3, Table 2). In such cases, an "averaged" structure based on FRET restraints may not be informative, or at least such a structure would require vastly different interpretation in comparison to one generated for a stably folded protein. Furthermore, because simulated proteins may explore conformational space far outside of experimentally based expectations, the analysis of correlated motion may be impacted by the use of restraints. Instead, it may well be more informative to validate unrestrained simulations with experimental data, applying limited guidance or screening after the initial run, if necessary, to bring the results from both methods into agreement: in short, a population-based treatment, in

DOI: 10.1021/jacs.5b04471
J. Am. Chem. Soc. 2015, 137, 11962−11969

keeping with the nature of the single-molecule experiments themselves.

MATERIALS AND METHODS

Preparation and Labeling of Proteins. All troponin subunits were expressed in *E. coli* using modified published protocols, with native cysteines mutated to serines in TnI. Cysteines were introduced at desired positions (TnI D152C, D165C, E179C, D190C, E209C), and Alexa 488 (donor) and Alexa 594 (acceptor) fluorophores attached by cysteine-maleimide chemistry. Tn complexes were assembled in vitro after labeling, then purified by size exclusion to ensure correct stoichiometry. Further details can be found in the SI.

smFRET Measurements. smFRET measurements were made on a laboratory-built instrument, described previously (details in the SI).[4] Samples were measured in [20 mM Tris pH 7.4, 150 mM KCl, 1 mM CaCl2, 0.5 mM MgCl2, 1 mM TCEP] at 75 pM labeled Tn complex, with an additional 6 nM of unlabeled Tn complex. We confirmed that Tn is stable under these conditions for the duration of our measurements (Figure S1). Photon traces were collected in 1 ms time bins over hour-long measurements, with photon bursts selected using a threshold of 30 photons per time bin, and adjacent time bins combined when both surpassed the threshold. Zero-peaks and contribution from dissociated TnI were identified and excluded from further analysis (see SI).

ET_{eff} values were calculated according to eq S1, then compiled into histograms and fitted with Gaussian distributions using laboratory-written MATLAB scripts. Because ET_{eff} values reflect the distances between two fluorophores, smFRET distances in this study are reported as interdye distances rather than interresidue distances. When converting ET_{eff} values to distance, the Förster equation was convoluted with a polymer model to account for time-averaging (see the SI).[2]

Models. Models of TnI_C were constructed in PyMol based on the published conflicting structures of TnI_C. Briefly, all structures contain TnC 1−88 and TnI 147−160 of the 1J1E crystal structure,[19] with TnI 161−210 attached following construction. Model 1 was fully built using the online I-TASSER platform constrained to the 1VDJ structure for fitting.[22,23] Model 2 simply extends the 1J1E crystal structure using minimized antiparallel beta strand for the missing residues. Model 3 uses homology modeling based loosely on another crystal structure of the skeletal isoform,[22,23,36] again with missing residues added as antiparallel beta strand. All models were minimized prior to the start of simulations (see the SI).

MD Simulations. MD simulations were performed in GROMACS 4.5.4 with AMBER99sb force field and TIP4P-EW water model (see the SI for a comparison of several water models and force fields, and quality controls).[25−28] Simulations used a fragment of the troponin complex, TnC 1−88 with TnI 147−210; this fragment contains both the region of interest and a folded domain to serve as a control (see the SI for discussion of MD simulation controls). Proteins were minimized (see the SI), built into periodic dodecahedral water boxes with 1.5 nm between solute and box, and NaCl added to 0.2 M concentration. A short 10 ps simulation was run at 0.1 fs step size to generate the starting conformation.

Simulations were performed as NPT (concentration, pressure and temperature maintained). Each group of simulations started with the same conformation, but with different seeds to generate different starting velocities. Five 50 ns simulations were performed for each starting model, three with 1 fs step size and two with 2 fs step size. Following completion of the trajectory runs, analyses were performed with standard GROMACS functions using the last 25 ns of each simulation. Contact maps used the minimum distance between any atom in one residue and any atom in another residue. Further details on analysis can be found in the SI.

Finally, MD pairwise distances were converted to intradye distances for comparison with smFRET measurements. AV calculations were performed using the FPS software published by the Seidel lab (details in the SI),[8] with one MD simulation from each model group calculated frame-by-frame to determine an average "linker contribution" for other trajectories of the same model (controls and details in the SI). MD frames were converted to ET_{eff} using the Förster equation because they were considered instantaneous in comparison to the time scale of smFRET measurements (eq S3).

Monte Carlo Simulations. Monte Carlo simulations were performed using the CAMPARI distribution in collaboration with the Pappu lab.[30,31] NVT ensemble simulations of TnI 161−210 were run in spherical water droplets with 100 Å radius and 15 mM NaCl. Parameterization was done with the OPLS-AA/L force field and ABSINTH implicit solvation. Replica exchange trajectories were run in duplicate for 46 000 000 steps, with the first 1 000 000 discarded as equilibration. All analyses were performed on the 298 K replicates. Contact maps were generated using all-atom conditions and clustered using standard CAMPARI commands. Further details can be found in the SI.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.5b04471.

Description of the methods, smFRET controls, molecular dynamics controls and force field tests, additional CAMPARI analyses, and troponin binding and function controls (PDF)

AUTHOR INFORMATION

Corresponding Author

*elizabeth.rhoades@sas.upenn.edu

Present Address

‖E.R.: Department of Chemistry, University of Pennsylvania, Philadelphia, PA 19104, United States.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank R. Pappu and A. Mittal for assistance with CAMPARI simulations and analysis; we also thank them and R. Das for helpful discussions. We thank J. Mittal and G. Ozer for providing frames of their MD simulations for testing purposes. The authors also thank C. Redwood for wild-type human cardiac TnC, TnI, and TnT plasmids; M. Mooseker and E. De La Cruz for guidance in thin filament protein work; and A. Nath and A. Miranker for helpful discussions, insights into MD simulations, and feedback on the manuscript. This work benefited from the facilities and staff of the Yale University Faculty of Arts and Sciences High Performance Computing Cluster. We acknowledge the support of the Raymond and Beverly Sackler Institute. This work was supported by NIH NS079955 (to E.R.) and American Heart Association 13PRE16570013 (to L.A.M.).

REFERENCES

(1) Ferreon, A. C.; Gambin, Y.; Lemke, E. A.; Deniz, A. A. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 5645.

(2) Nath, A.; Sammalkorpi, M.; DeWitt, D. C.; Trexler, A. J.; Elbaum-Garfinkle, S.; O'Hern, C. S.; Rhoades, E. *Biophys. J.* **2012**, *103*, 1940.

(3) Soranno, A.; Koenig, I.; Borgia, M. B.; Hofmann, H.; Zosel, F.; Nettels, D.; Schuler, B. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 4874.

(4) Trexler, A. J.; Rhoades, E. *Biochemistry* **2009**, *48*, 2304.

(5) Lasker, K.; Forster, F.; Bohn, S.; Walzthoeni, T.; Villa, E.; Unverdorben, P.; Beck, F.; Aebersold, R.; Sali, A.; Baumeister, W. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 1380.

(6) Jensen, M. R.; Zweckstetter, M.; Huang, J. R.; Blackledge, M. *Chem. Rev.* **2014**, *114*, 6632.

DOI: 10.1021/jacs.5b04471
J. Am. Chem. Soc. 2015, 137, 11962−11969

(7) Brunger, A. T.; Strop, P.; Vrljic, M.; Chu, S.; Weninger, K. R. *J. Struct. Biol.* **2011**, *173*, 497.
(8) Kalinin, S.; Peulen, T.; Sindbert, S.; Rothwell, P. J.; Berger, S.; Restle, T.; Goody, R. S.; Gohlke, H.; Seidel, C. A. *Nat. Methods* **2012**, *9*, 1218.
(9) Sabir, T.; Schroder, G. F.; Toulmin, A.; McGlynn, P.; Magennis, S. W. *J. Am. Chem. Soc.* **2011**, *133*, 1188.
(10) Yang, S.; Barbu-Tudoran, L.; Orzechowski, M.; Craig, R.; Trinick, J.; White, H.; Lehman, W. *Biophys. J.* **2014**, *106*, 855.
(11) Galinska, A.; Hatch, V.; Craig, R.; Murphy, A. M.; Van Eyk, J. E.; Wang, C. L.; Lehman, W.; Foster, D. B. *Circ. Res.* **2010**, *106*, 705.
(12) Hoffman, R. M.; Blumenschein, T. M.; Sykes, B. D. *J. Mol. Biol.* **2006**, *361*, 625.
(13) Xing, J.; Jayasundar, J. J.; Ouyang, Y.; Dong, W. J. *J. Biol. Chem.* **2009**, *284*, 16432.
(14) Julien, O.; Mercier, P.; Allen, C. N.; Fisette, O.; Ramos, C. H.; Lague, P.; Blumenschein, T. M.; Sykes, B. D. *Proteins: Struct., Funct., Genet.* **2011**, *79*, 1240.
(15) Chung, M. W.; Tsoutsman, T.; Semsarian, C. *Cell Res.* **2003**, *13*, 9.
(16) Hilfiker-Kleiner, D.; Knoll, R. *Circulation* **2008**, *117*, 1775.
(17) Murakami, K.; Yumoto, F.; Ohki, S. Y.; Yasunaga, T.; Tanokura, M.; Wakabayashi, T. *J. Mol. Biol.* **2005**, *352*, 178.
(18) Blumenschein, T. M.; Stone, D. B.; Fletterick, R. J.; Mendelson, R. A.; Sykes, B. D. *Biophys. J.* **2006**, *90*, 2436.
(19) Takeda, S.; Yamashita, A.; Maeda, K.; Maeda, Y. *Nature* **2003**, *424*, 35.
(20) Zhou, Z.; Li, K. L.; Rieck, D.; Ouyang, Y.; Chandra, M.; Dong, W. J. *J. Biol. Chem.* **2012**, *287*, 7661.
(21) Manning, E. P.; Tardiff, J. C.; Schwartz, S. D. *Biochemistry* **2011**, *50*, 7405.
(22) Roy, A.; Kucukural, A.; Zhang, Y. *Nat. Protoc.* **2010**, *5*, 725.
(23) Zhang, Y. *BMC Bioinf.* **2008**, *9*, 40.
(24) Sindbert, S.; Kalinin, S.; Nguyen, H.; Kienzler, A.; Clima, L.; Bannwarth, W.; Appel, B.; Muller, S.; Seidel, C. A. *J. Am. Chem. Soc.* **2011**, *133*, 2463.
(25) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435.
(26) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J. Chem. Phys.* **2004**, *120*, 9665.
(27) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 712.
(28) Sorin, E. J.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2472.
(29) Das, R. K.; Ruff, K. M.; Pappu, R. V. *Curr. Opin. Struct. Biol.* **2015**, *32*, 102.
(30) Vitalis, A.; Pappu, R. V. *Annu. Rep. Comput. Chem.* **2009**, *5*, 49.
(31) Vitalis, A.; Pappu, R. V. *J. Comput. Chem.* **2009**, *30*, 673.
(32) Lyle, N.; Das, R. K.; Pappu, R. V. *J. Chem. Phys.* **2013**, *139*, 121907.
(33) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 8868.
(34) Das, R. K.; Pappu, R. V. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 13392.
(35) Day, S. M.; Westfall, M. V.; Fomicheva, E. V.; Hoyer, K.; Yasuda, S.; La Cross, N. C.; D'Alecy, L. G.; Ingwall, J. S.; Metzger, J. M. *Nat. Med.* **2006**, *12*, 181.
(36) Vinogradova, M. V.; Stone, D. B.; Malanina, G. G.; Karatzaferi, C.; Cooke, R.; Mendelson, R. A.; Fletterick, R. J. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 5038.

DOI: 10.1021/jacs.5b04471
J. Am. Chem. Soc. 2015, 137, 11962−11969