Supporting Information:

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,[‡] Integrated Graduate Program in Physical and Engineering Biology,[§] Yale University, New Haven, Connecticut 06520

Email: elizabeth.rhoades@sas.upenn.edu

Contents

Materials and Methods (Expanded)

Preparation and Labeling of Proteins

Expression plasmids for human cardiac troponins were the kind gift of C. Redwood. All troponin subunits were expressed in *E. coli* BL21 DE3 cells. Briefly, troponin C (TnC) was purified according to published protocols using hydrophobic exchange chromatography to isolate TnC proteins capable of conformational change in response to shifting calcium concentrations.[1] Proteins bound a phenyl sepharose column equilibrated in a neutral Tris buffer (50 mM Tris-HCl pH 7.5, 1 mM DTT, 50 mM NaCl , 10 mM $CaCl_2$, 1 mM $MgCl_2$); the column was washed with a lower-salt buffer (1 mM NaCl, 0.1 mM $CaCl_2$), and TnC was eluted with the addition of 1 mM EDTA. Troponin T (TnT) was purified according to published protocols using anion and cation exchange chromatography.[1] Cells were lysed into a homogenization buffer containing 8 M urea to solubilize inclusion bodies. Cell lysate was run over a CM sepharose column equilibrated in a low-pH denaturing buffer [50 mM sodium citrate pH 5.5, 6 M urea, 1 mM EDTA, 0.1 mM TCEP] and eluted in a single step elution of 100 mM NaCl. Fractions containing TnT were purified using a DEAE sepharose column equilibrated in a Tris denaturing buffer [50 mM Tris-HCl pH 8.0, 6 M urea, 1 mM EDTA, 0.1 mM TCEP] and eluted in a 0-200 mM NaCl gradient. Finally, troponin I (TnI) cells were chemically lysed, and soluble proteins discarded. Inclusion bodies were dounced into a denaturing buffer [25 mM triethanolamine pH 7.0, 8 M urea, 2 mM EDTA, 1 mM DTT], dialyzed overnight against the same buffer, and sonicated in a water-bath sonicator for 30 seconds before centrifuging at 40,000 rpm on a Ti45 rotor for 30 minutes to remove insoluble material. The inclusion body sample was loaded onto a CM sepharose column and eluted in a 0-0.5 M NaCl gradient. TnI-containing fractions were then purified by size exclusion chromatography using a sepharose S-100 column equilibrated in [25 mM triethanolamine pH 7.0, 6 M urea, 150 mM NaCl, 1 mM DTT]. All proteins were concentrated and flash-frozen for storage.

To assemble troponin complexes (Tn), 16 nmol each TnC and TnT were combined with labeled TnI (typically 16-20 nmol) and incubated on ice for 1 hour in a buffer admixture containing at least 6 M urea. The sample was then dialyzed against a high-salt buffer [25 mM MOPS pH 7.0, 1 M NaCl, 5 mM $CaCl_2$, 1 mM DTT] and the salt concentration slowly reduced to 0.2 mM NaCl / 2 mM $CaCl_2$. Assembled complexes were purified by size exclusion chromatography to ensure a 1:1:1 stoichiometry, with the final buffer conditions being [25 mM MOPS pH 7.0, 0.2 mM KCl, 2 mM $CaCl_2$, 1 mM $MgCl_2$, 1 mM TCEP]. After verifying that "fresh" complexes and stored complexes were indistinguishable in spectroscopic measurements, complexes were flash-frozen for storage.

TnI was labeled prior to assembly of the troponin complexes to preserve native cysteines on TnC. Briefly, native cysteines in TnI were mutated to serine, and cysteine residues were introduced at desired locations. For constructs using the D190 position, a conservative W191F mutation was added to reduce fluorophore quenching; this mutation was employed by another lab and proven to have no effect on function.[2] 25 nmol TnI was brought to 150 µL (final concentration TnI 167 µM) in [50 mM Tris pH 7.4, 6 M urea, 0.15 M KCl, 1 mM EDTA, 1 mM DTT], then dialyzed 3 hours against the same buffer without DTT. The protein was then mixed with 0.04 mg Alexa 488 and 0.2 mg Alexa 594 (both dissolved in DMSO). The protein/dye mixture was slowly stirred at room temperature for 20 minutes, then overnight at 4˚ C. Free dye was removed using two coupled Hi-Trap Desalting Columns incubated in [50 mM Tris pH 7.4, 300 mM KCl, 6 M urea, 1 mM TCEP]. 1 mM DTT was added to labeled Tn to quench any unreacted free dye present before assembling the complex. For TnC 35 / TnI 152, TnC was labeled with Alexa 488 following the above protocol, and TnI was labeled with Alexa 594.

Actin and tropomyosin (α/β mix) were purified from rabbit skeletal muscle acetone powder using published protocols.[3,4] Murine cardiac myosin was purified and cleaved into heavy meromyosin using published protocols.[5,6]

Single-Molecule Förster Resonance Energy Transfer

smFRET measurements were made on a laboratory-built instrument based on an inverted Olympus IX-71 microscope (Olympus), described previously.[7] Laser power (488 nm) was adjusted to 25-30 μW before entry into the microscope. Fluorescence emission was collected through the objective and separated from laser excitation using a Z488RDC long-pass dichroic and a 500 long-pass filter (Chroma). Donor and acceptor photons were separated using an HQ585LP dichroic and further selected using emission filters: an ET525/50M filter for the donor photons, and a 605LP filter for acceptor photons. Photons were detected using avalanche photodiodes (Perkin-Elmer) directly coupled to the microscope through 100 μm-diameter optical fibers (OzOptics).

Samples were placed in eight-well chambered coverglasses (Nunc) passivated by polylysine-conjugated polyethylene glycol to prevent adsorption of Tn to chamber surfaces.[8] Wells contained 6 nM unlabeled (dark) Tn in [20 mM Tris pH 7.4, 150 mM KCl, 1 mM $CaCl_2$, 1 mM TCEP, 0.5 mM $MgCl_2$], to which was added roughly 75 pM of labeled Tn immediately before measurement. Photon traces for the acceptor and donor channels were collected in 1 ms time bins over the course of hour-long measurements. Photon bursts were selected using a threshold of 30 total photons per time bin, with adjacent time bins combined when both surpassed the threshold.

Energy transfer efficiencies (ET_{eff}) were calculated for photon bursts by Equation S1, where I_a and I_d are the fluorescence intensity in the acceptor and donor channels, respectively. The β factor accounts for donor fluorescence bleed-through to the acceptor channel (0.06 for Alexa Fluor 488 on our instrument), and the γ factor accounts for differences in detection efficiency and quantum yield between the two fluorophores (1.3 on our instrument).

$$ET_{eff} = \frac{(I_a - \beta * I_d)}{(I_a - \gamma * I_d)}$$

(Equation S1)

ET_{eff} values for bursts were compiled into histograms, and representative traces chosen for figures. Histograms were fitted as a sum of two or three Gaussian peaks, depending on whether there was a clear monomer peak visible and distinguishable. The peaks present were a zero peak (donor-only labeled protein and background noise), a signal peak, and occasional monomer peak (see details of characterization below). All fitting was performed in MATLAB (MathWorks) using laboratory-written scripts.

Conversion of smFRET Efficiencies to Distances

The ET_{eff} values reflect the distances between the two fluorophores, which are attached to the protein via ~20 Å long flexible linkers. For this reason, all distances in this study were calculated as inter-dye distances rather than inter-residue distances, because the conversion from residue to dye position for simulations was less error-prone than assigning an arbitrary correction to the ET_{eff} values.

smFRET measurements are time-averaged; therefore, it is necessary to convolute the Förster equation with a polymer model to convert ET_{eff} values to distances. Our polymer model assumes less flexibility than the Gaussian chain model, resulting in calculated distances in between those predicted by the Förster equation (most accurate for rigid bodies) and those predicted by the Gaussian chain model (most accurate for completely flexible polymers). A full explanation of this conversion, and discussion of how it compares with other ET_{eff}-to-distance conversions, can be found in a previous publication.[9]

Molecular Dynamics Simulations

All simulations were performed in GROMACS 4.5.4 using the Yale High Performance Computing Facilities.[10] The starting models for simulations were constructed based on published conflicting structures of the C-terminal domain of TnI (TnI_C). The 1J1E crystal structure of human cardiac Tn was used as a base for all simulations, with any discontinuities in the sequence modeled in as extended loops.[11] A fragment of the complex containing TnC 1-88 and TnI 147-210 was chosen as the smallest portion of the complex containing both the region of interest and a folded domain as a control (see

further discussion below). Models were crafted as described in the text and loaded into GROMACS, with hydrogen atoms placed by the software. The force field chosen was Amber99sb, with the Tip4p-EW water model (see Figure S4 for force field and water model tests).[12-14] Models were minimized by iterating through the Conjugate Gradient, L-BFGS, and Steepest Descent algorithms until the maximum force on any atom in a structure was below 1 kJ mol^{-1} nm^{-1} (typically 2-4 iterations through the set). Following the initial minimization, proteins were placed in a dodecahedral water-box with 1.5 nm between the solute and the box. Na and Cl atoms were added to a 0.2 M concentration, with any imbalance necessary to neutralize charge within the box. Periodic boundary conditions were used, with no observed contacts across the periodic boundary.

Following solvation, a short 100 ps simulation was run in lieu of a second minimization. This MD run was too short to allow any significant movement of the protein; instead, its primary purpose was to allow water molecules to settle. Simulations were NPT ensembles (number of particles, pressure, and temperature held constant); starting conditions were 1 bar pressure and 298 K temperature, consistent with laboratory conditions. Berendsen pressure coupling and the velocity rescaling thermostat were used for pressure and temperature, respectively, with 1 ps and 0.1 ps time constants for coupling pressure and temperature, respectively. Treatment of electrostatics was in keeping with the Amber99sb force field parameterization: nonbonded interactions were cut off at 0.9 nm, and the particle mesh Ewald method was used for calculating long-range interactions. In these short simulations, a 0.1 fs step size was used.

The output of the 100 ps simulation was then used as the frame zero for all simulations of the same model. Three simulations of each model were run using the above parameters, with different random seeds, and a step size of 1 fs. An additional two simulations of each model were run with a step size of 2 fs after confirming that this increase in step size did not affect simulation output. Trajectory

position files were updated every 20 ps throughout the simulation. For the purposes of maintaining a manageable dataset, pairwise distances used for AV calculations (below) were calculated at 200 ps intervals.

The mass-weighted covariance matrices[15] were calculated for TnI residues 150-210 using the g_covar function in Gromacs, integrating over the 25-50 ns time interval for each replicate and using an averaged structure as the reference. The backbone residues were used for least-squares fitting to the reference structure, and the covariance calculated for alpha carbon positions. Covariance was summed across the xx, yy and zz covariances, then converted to cross-correlation by normalizing over the movement of the respective atoms (Equation S2):

$$C_{(i,j)} = \frac{c(i,j)}{\sqrt{2}\sqrt{|c(i,i)|}\sqrt{|c(j,j)|}} \qquad \text{(Equation S2)}$$

where $C_{(i,j)}$ is the element of the cross-correlation matrix for residue i in relation to residue j, and c(i,j) is the element of the covariance matrix for residues i in relation to residue j. The cross-correlation matrices for the five replicates of each model were averaged together, and a threshold of 0.2 applied for negative and positive cross-correlation values. While these matrices are naturally symmetrical across the diagonal, we split the matrices into positive and negative cross-correlations, with only positive cross-correlation displayed on the top left half, and only negative cross-correlation displayed on the bottom right half.

To search for common elements in all simulations irrespective of starting model, we calculated a 'commonality' matrix. Each element of a cross-correlation matrix was assigned a value of 1 if $|C_{(i,j)}| > 0.2$, then the matrices were summed over all fifteen simulations (5 simulations for each of the 3 models) for total scores of 0-15 for each element.

In an effort to understand the potential mechanism behind the prominent anti-correlated motion from the cross-correlation analysis, we computed a charge-charge interaction matrix. Briefly, net charge was calculated over a six-residue sliding window. Relative interactions between charge groups were calculated as a multiplication of the two net charges, flipping the resultant sign so that negative values indicate repulsive terms. The matrix was then divided across the diagonal mirror plane, with all positive calculated interactions on the top left and all negative calculated interactions on the bottom right. This analysis was intended as a rough estimate of what correlated chain motions might be explainable by sequence alone regardless of structure, and is not intended to stand alone as a robust analysis of electrostatics.

To aid comparison with CAMPARI simulations, we calculated contact maps for both starting models and time-averaged simulations. All initial contact map calculations were performed using the g_mdmat command, which calculates the mean distance between residues over the analysis period (25-50 ns) for each simulation. Distance maps were then converted to binary contact maps (1 for distances at or below 3.5 Å, 0 for distances above 3.5 Å) in Matlab, then averaged across the five replicates for the model. Contacts representing i,i+1,i+2 contacts were removed prior to plotting. This analysis mirrors the CAMPARI contact map calculation as closely as possible.

All analyses use trajectory data taken every 20 ps unless otherwise indicated.

Accessible Volume (AV) Calculations

AV calculations were performed using the FPS software published by the Seidel lab.[16] The donor fluorophore (Alexa Fluor 488) was modeled with the program's parameters of length 20, width 4.5, and dye radius 3.5 Å; the acceptor fluorophore (Alexa Fluor 594) was modeled using length 20, width 4.5, and a three-radii AV calculation with radii 6.7, 4.5, and 1.5 Å.

For each simulation, the representative trajectory used in Figure 3 was printed to PDB file every 200 ps. Each printed frame was analyzed with the AV modeling. The difference between the pairwise

residue distance and the AV mean dye position was then applied as the "linker contribution" to the other trajectories of the same model. To confirm that this application was sound, we performed AV modeling on six frames from each trajectory and compared the mean linker contributions (data not shown). In all cases, the linker contributions were less than the standard deviation of the MD positions.

Because MD frame-by-frame outputs can be modeled as instantaneous in comparison with the timescale of smFRET measurements, we convert them to ET_{eff} using the Förster equation (Equation S3):

$$ET_{eff} = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6} \quad \text{(Equation S3)}$$

where r is the inter-dye distance, and R_0 is the Förster radius of the dye pair (54 Å).[17]

Monte Carlo Simulations

All Monte Carlo (MC) simulations were run using the CAMPARI package with OPLS parameterization and the Absinth implicit solvation model.[18,19] Extended conformers of TnI_C 161-210 were run in 15 mM NaCl spherical water droplets with a 75 Å radius and atom-based soft-wall boundary conditions (moveset parameters available upon request). Replica exchange simulations were run for 46 million steps, with a one million step equilibration period. Replicates at 280, 285, 289, 294, 298, 310, 320, 330, 340, 355, 370, 385 and 400 K were run, with 10 swaps between neighboring replicas attempted every 50,000 steps. Unless other temperatures are indicated, all analyses were performed on the 298 K replica.

Helical propensities were calculated using the BBSEG, DSSP, and FYC keyword analyses. Contact map analysis and clustering were performed in CAMPARI using 3.5 Å cut-offs and eliminating i, i+1 and i+2 contacts; additional clustering of contact maps was examined using Matlab clustering functions. Contact maps use distances calculated by choosing the smallest pairwise distance between any two

atoms of two different residues. Hydrodynamic radius, asphericity, and pairwise distance between residues were all calculated using standard CAMPARI functions and with additional Matlab analysis where necessary.

Supplemental Experiments: smFRET Controls

Fluorescence Correlation Spectroscopy

Because Tn is a three-subunit complex, we sought to confirm that the complex was stable at single molecule concentrations. To do this, we used fluorescence correlation spectroscopy (FCS) to monitor the diffusion time of Tn over two hours at varying concentrations; a decrease in diffusion time would indicate dissociation of the complex.

FCS measurements were made on a laboratory-built instrument based on an inverted Olympus IX-71 microscope (Olympus).[7] Laser power was adjusted to 4-6 µW before entering the microscope. Fluorescence emission was collected through the objective, and separated from the laser excitation using a Z488RDC long-pass dichroic and an HQ600/200M band-pass filter (Chroma) before being focused onto the aperture of a 50-µM-diameter optical fiber (OzOptics) directly coupled to an avalanche photodiode (Perkin-Elmer). Autocorrelation curves were generated by a digital correlator (Correlator.com). For each measurement, 30 fluorescence intensity traces of 15 seconds each were recorded and averaged together to obtain statistical variations. The average curve was then fitted with a diffusion equation for a single component (Equation S4) weighted by the inverse-square of the standard deviation.

$$G(\tau) = \frac{1}{N}\frac{1}{1+\frac{\tau}{\tau_D}}\sqrt{\frac{1}{1+\frac{s^2\tau}{\tau_D}}} \qquad \text{(Equation S4)}$$

In the above equation, G(τ) is the autocorrelation as a function of time τ, N is the average number of fluorescent particles in the focal volume at any given time, τ_D is the average diffusion time of the fluorescent particles, and s is the ratio of the radial to axial dimension of the focal volume. In these measurements, Tn was labeled at the TnC C35 position with Alexa Fluor 488; TnC is the only subunit of the Tn complex that is stably folded as a monomer and could be expected to remain in solution were the complex to disassemble, which would lower the diffusion time. τ_D was measured every 30 minutes for two hours at room temperature in the eight-chambered passivated wells described above for smFRET measurements.

This analysis demonstrated that Tn diffusion times for concentrations 5 nM and above were stable for 60 minutes at room temperature (Figure S1). At 90 minutes, diffusion times displayed a greater standard deviation and decrease of the mean, consistent with roughly 25% monomeric TnC. 1 nM and 0.5 nM concentrations experienced greater losses in diffusion time, with the 0.5 nM signal already lower than the 1 nM signal at 90 minutes. We therefore chose the 5 nM concentration as appropriate for smFRET, adding an additional 1 nM to correct for any error.

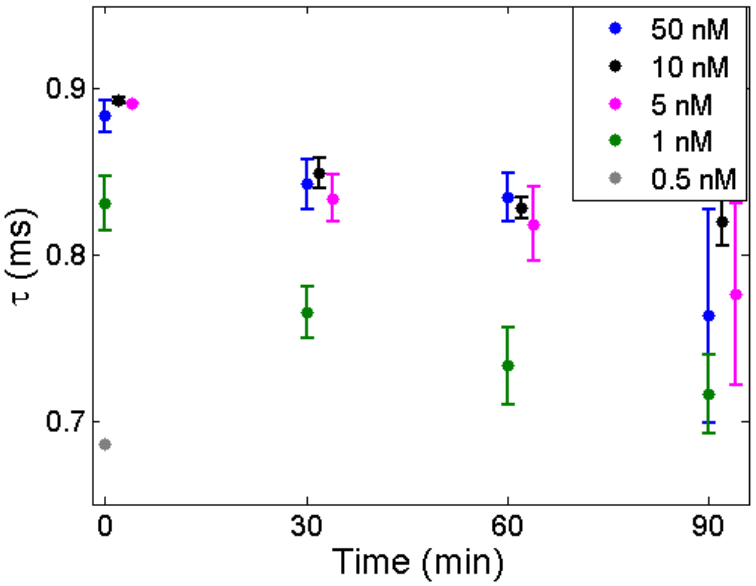

Figure S1. Diffusion time of the Tn complex over 90 minutes at varying concentrations. Concentrations 5 nM and above are stable over 60 minutes, indicated by the tightly clustered diffusion times compared with the lower 1 nM concentration. Timepoints are spread along the abscissa for clarity.

Fluorescence Anisotropy

All labeling positions were checked for rotational freedom of the fluorophore using fluorescence anisotropy. Steady state anisotropy was measured on a QuantaMaster C-61 fluorescence spectrometer (PTI). 100 second intensity scans were taken over a wavelength range to verify lack of wavelength dependence in detection, and the intensities were averaged. Alexa Fluor 488 was excited at 490 nm and emission collected over 510-530 nm; Alexa Fluor 594 was excited at 590 nm and emission collected over 610-630 nm. Concentrations ranging from 0.1-10 µM protein were used for analysis. Anisotropy (r) was calculated as

$$r = \frac{I_{VV} - G * I_{VH}}{I_{VV} + 2G * I_{VH}}$$

(Equation S5)

where $G = I_{HV}/I_{HH}$. Some fluorophore positions were measured on FRET-labeled Tn, resulting in the Alexa 594 measurement being an average of anisotropy for two positions. In these cases, anisotropy was calculated for two separate constructs to confirm that acceptable anisotropies were not the result of averaging between a high and low number. For example, position 190 was measured in both the 152/190 and 165/190 constructs; TnI 152 is documented as having the highest anisotropy of the set, so TnI 190 must have a lower anisotropy than this maximum because it's two-position average is below the TnI 152 anisotropy value.

Labeling Position	Fluorophore	Anisotropy (r)
TnC C35	Alexa 488	0.167
TnI 152	Alexa 488	0.194
	Alexa 594	0.243
TnI 165*	Alexa 594	0.166; 0.162
TnI 179	Alexa 488	0.132
TnI 190*	Alexa 594	0.172; 0.166
TnI 209*	Alexa 594	0.162; 0.135

*Labeling positions marked with an asterisk are a mixture of two positions. In these cases, all were measured with two separate partners. Position 165 was measured in combination with positions 190 and 209, position 190 was measured with positions 152 and 165, and position 209 was measured with positions 165 and 179.

Table S1. Anisotropy measurements for all labeling positions show high orientational freedom, indicating that the $\kappa^2 = 2/3$ assumption holds true in these conditions.

Fluorescence Lifetime Measurements

We performed fluorescence lifetime measurements to confirm that the Förster radius was not affected by quenching of the donor fluorophore. Measurements were performed on a TCSPC TD-Fluor Horiba Fluorolog 3 Time Domain Fluorimeter using a 459 nm nano-LED for excitation and 517 nm emission monochrometer. Slit widths were adjusted for an alpha measurement of 0.5-2% photon collection, and measurements were performed until 10,000 photons were collected in the highest bin. Traces were fitted with single-component fits, adding a second component only if it contained more than 5% relative abundance (R.A.). In this case, the T2 component was fixed for the Alexa488 lifetime calculated with a dye standard in that fitting session.

Position	T1 Component	R.A. (%)	T2 Component	R.A. (%)	Chi^2
TnC 35	2.3 ns	18.7 %	4.1 ns (fixed)	81.3 %	1.18
TnI 152	4.1 ns				1.44
TnI 165	3.9 ns				0.98
TnI 179	2.4 ns	14.7 %	4.1 ns (fixed)	85.3 %	1.27
TnI 190	4.0 ns				0.95

Table S2. Fluorescence lifetime calculations on Alexa 488 at selected labeling positions. Both single- and double-component fits were calculated, and the single-component fit reported except in cases with a more than 5% relative abundance of a second lifetime component.

To confirm that the small quenched components in the TnC35 and TnI 179 positions would not affect the measured transfer efficiency, we performed a simple mathematical modeling experiment. A representative smFRET dataset was taken, and each single event was given the R.A. percent chance of experiencing the T1 component lifetime (and the ET recalculated according to the change in the Förster radius). The dataset was then compiled into a histogram and fitted with our standard smFRET functions. We observed that even the strongest quenching from our labeling positions resulted in a histogram fit that was within one standard deviation of the dataset, indicating that the quenching would not affect our results.

Alternating Laser Excitation to Confirm Labeling

In some labeled complexes, we observed a minor contribution from a high-ET peak. This peak behaved differently from the primary Tn complex peak in several key ways: event rate in this peak was steady over time while Tn events decrease due to sticking and dissociation, molecules in the high-ET peak diffused more quickly than molecules in the primary peak, and events in the high-ET peak contained fewer photons. First, we used alternating laser excitation (ALEX) to confirm that the high-ET peak was indeed double-labeled protein rather than a contaminant.[20] Briefly, this was performed on the laboratory-built instrument used for smFRET (see above methods). The lasers were controlled by acousto-optic modulators (Isomet) that interleaved 100 µs laser pulses of 488 nm and 561 nm with 5 µs of dark time in between each pulse, allowing for the capture of both traditional smFRET and a "stoichiometry" ratio characterizing labeling (Equation S6):

$$S = \frac{I_a + I_d}{I_d + I_a + I_a^a}$$

(Equation S6)

where I_a and I_d are the intensity of the acceptor and donor channels during 488 nm excitation and I_a^a is the intensity of the acceptor channel during 561 nm excitation.

We observed that, while the smFRET histograms for several constructs are trimodal, the stoichiometry histograms are only bimodal. The high-stoichiometry peak corresponds with the zero-transfer efficiency peak, consistent with donor-only labeling. However, both non-zero smFRET peaks share the same stoichiometry, indicating that they share the same labeling characteristics (Figure S2).

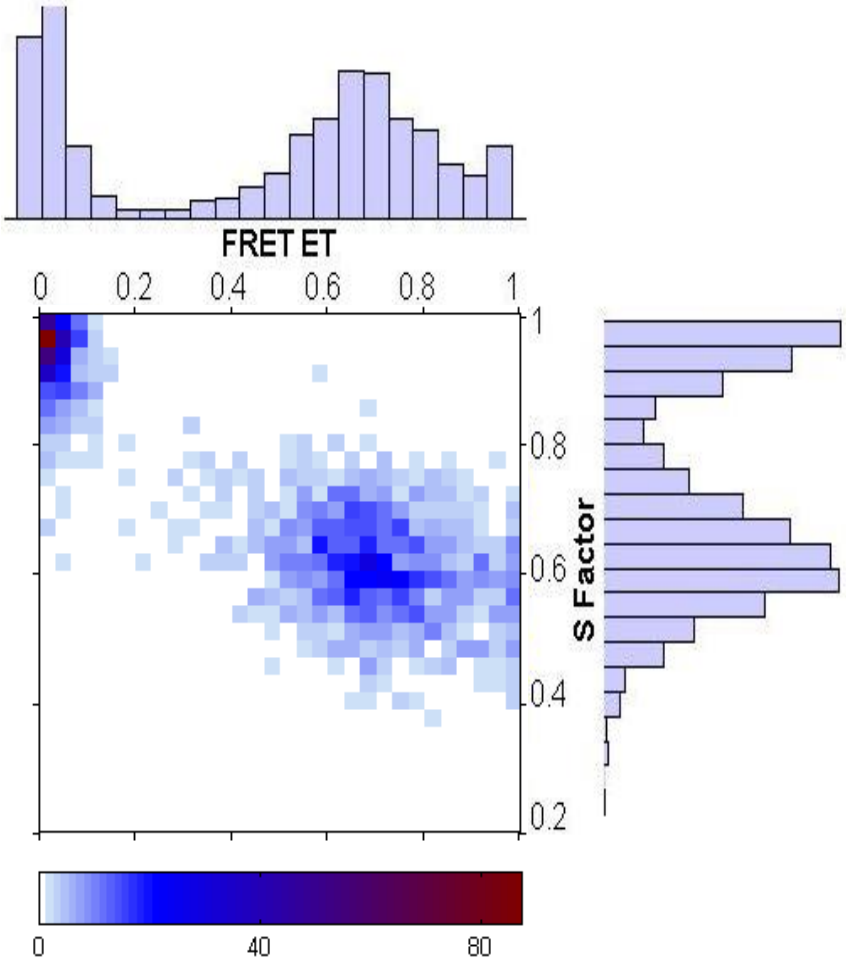

Figure S2. A representative scatter plot of transfer efficiency (ET_{eff}) and stoichiometry (S) for the 152/209 construct measured by ALEX, with histograms for each along the axes. Both non-zero smFRET peaks share the same stoichiometry, indicating similarity in labeling and ruling out labeling artifacts as the source of the high-ET peak.

Identification of the TnI Monomer Peak

To confirm that the high-ET peak seen in some histograms is signal from dissociated TnI as it lost from solution, we performed smFRET measurements on TnI monomer. Labeled TnI was maintained in the smFRET buffer with 8 M urea; to begin measurements, 75 pM of labeled protein was rapidly pipetted into non-denaturing buffer and measured as the signal decreased (TnI in non-denaturing conditions is quickly lost to sticking and precipitation, with an 80% loss over 30 minutes). We observed that the monomer collapsed quickly from a slower-diffusing state to a faster-diffusing high-ET_{eff} state during these measurements; the latter state is consistent with the both the high ET_{eff} values and faster diffusion times observed in the high-ET_{eff} peak of Tn histograms (Figure S3).

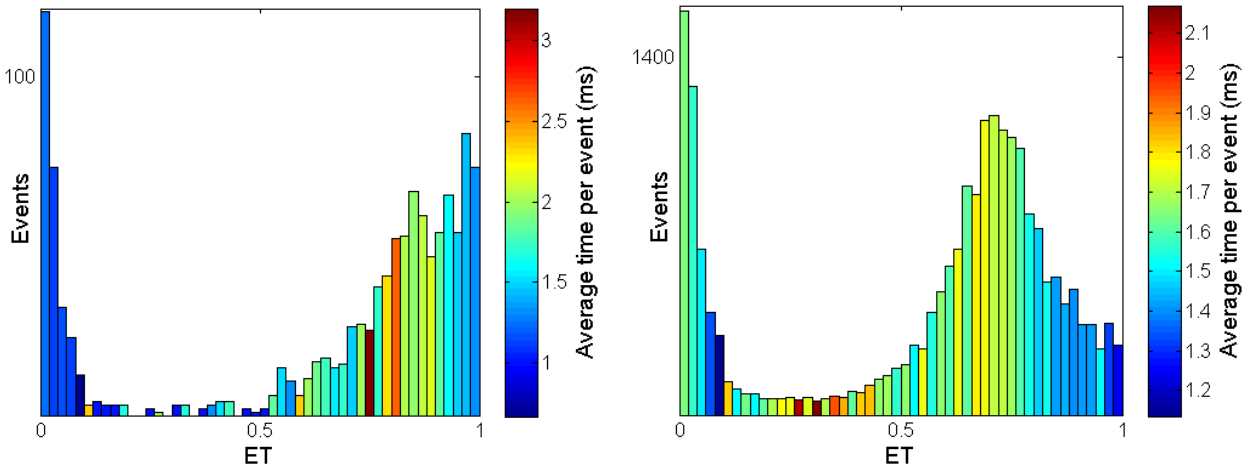

Figure S3. Representative smFRET histograms for TnI 152/190 monomer (left) and Tn complex (right), color-coded by the duration of time in the focal volume per event. The TnI monomer collapses and falls out of solution over time, with higher ET_{eff} values and faster diffusion time than the Tn complex. The high-ET_{eff} shoulder of the Tn histogram is explained by contribution from TnI monomer.

Analysis of Histogram Width and Thresholding

We observed a range of peak widths in our smFRET histograms, from 0.068 to 0.138, with an average value of 0.109. The width of these peaks is a convolution of many contributing factors: shot noise, conformational dynamics, static conformational variability over a small distance range, subtle variations in fluorophore lifetime, and many other factors. Therefore, we do not ascribe meaning to the

peak width beyond its effect on a reliable location of the true mean of the distribution (the definition of the σ parameter in a Gaussian fit).

Histogram widths can be affected by thresholding, or the number of photons required for defining a photon burst for smFRET analysis. The threshold of 30 photons was chosen as the lowest threshold that consistently limits noise from spurious photon counts on our hardware and biological system. (It is important to note that thresholding beyond this level decreases data quality by decreasing the number of events in the histogram and selectively eliminating high-ET events due to differences in detection efficiency and fluorophore brightness.) We performed additional analyses of a subset of smFRET data to determine the effect of a lower threshold (20 photons), or more stringent thresholds (40-50 photons) (Table S3). Because thresholding also affects shot noise contributions, we have included its effect on those estimates as well.

While increasing thresholds result in modestly decreased Gaussian peak widths, this accompanies a decrease in shot noise, indicating that the relative peak widths between constructs is not affected. Furthermore, the effects are quite modest, confirming that the choice of a different threshold would not impact interpretation of our data.

	T=20			T=30			T=40			T=50		
Construct	x	σ	Shot Noise	x	σ	Shot Noise	x	σ	Shot Noise	x	σ	Shot Noise
152/179	0.822 ± 0.003	0.093 ± 0.002	0.086	0.819 ± 0.002	0.084 ± 0.004	0.070	0.820 ± 0.001	0.080 ± 0.003	0.061	0.818 ± 0.005	0.075 ± 0.007	0.054
152/190	0.720 ± 0.007	0.124 ± 0.004	0.100	0.711 ± 0.006	0.118 ± 0.005	0.083	0.704 ± 0.008	0.112 ± 0.004	0.072	0.700 ± 0.009	0.107 ± 0.003	0.065
152/209	0.668 ± 0.005	0.151 ± 0.009	0.105	0.655 ± 0.009	0.131 ± 0.018	0.087	0.652 ± 0.003	0.128 ± 0.011	0.075	0.645 ± 0.004	0.122 ± 0.010	0.068
165/190	0.836 ± 0.011	0.139 ± 0.007	0.083	0.818 ± 0.003	0.116 ± 0.006	0.070	0.813 ± 0.006	0.111 ± 0.006	0.062	0.806 ± 0.006	0.106 ± 0.007	0.056
165/209	0.756 ± 0.003	0.149 ± 0.009	0.096	0.747 ± 0.001	0.138 ± 0.008	0.079	0.738 ± 0.007	0.131 ± 0.011	0.070	0.733 ± 0.004	0.128 ± 0.008	0.063
179/209	0.847 ± 0.001	0.079 ± 0.002	0.080	0.844 ± 0.001	0.068 ± 0.002	0.066	0.842 ± 0.002	0.064 ± 0.002	0.058	0.839 ± 0.002	0.059 ± 0.003	0.052

Table S3. Mean and standard deviation of the mean for Gaussian peak center and width calculations across a subset of three measurements for each construct. Each dataset was analyzed with varying photon thresholds and compared with estimated shot noise contribution at that specific threshold and transfer efficiency. Numbers are reported to three significant figures in order to display the standard deviation of the mean value. smFRET measurements used in analysis are only considered accurate to two significant figures.

MD and MC Simulation Controls

MD Force Field and Water Model Tests

The GROMACS package includes implementation of OPLS, CHARMM, Amber, and Gromos, as well as recommended water models for each. While we have found Amber99sb/tip4p-EW to perform best with IDPs,[9] we set to test several force fields and water models. Force fields were evaluated by three criteria: radius of gyration of TnI_C, freedom of conformational sampling, and maintenance of folded regions.

Simulations of disordered proteins suffer from several known issues, most notably artifactual compaction of the disordered region. These proteins typically have high solvent exposure, high charge

density and low hydrophobicity, all of which differ substantially from the folded proteins for which force fields were originally optimized. Amber99sb has a well-documented history of good performance with both disordered proteins and folded proteins;[9,21-26] however, because TnI_C is both an IDR and has controversial structure, we chose to compare the Amber99sb force field and Tip4p-EW water model with other major force field and water model combinations rather than assume that good practices for IDP simulations would be best suited for this region. We tested force field and solvent models on a TnI 160-210 fragment, monitoring the radius of gyration over time. All simulations were run using a 1 fs step size, with other parameters as described in the SI Methods section above. Simulations were started from Model 3 (see main text), allowing simulations to start from an extended conformation. Radius of gyration was calculated at each frame, and the mean and standard deviation between the set of five replicates was calculated.

We tested four combinations of force field and water model: Amber99sb/tip4p-EW, Gromos53a6/spc, OPLS-AA/L/tip4p, and CHARMM27/tip4p. Water models were chosen based on recommendations in the Gromacs program. Of the four force field/water model combinations, Gromos53a6 and OPLS-AA/L simulations resulted in decreased mean radius of gyration, which is inconsistent with experiments (Figure S4). CHARMM27 displayed a lower mean radius of gyration than Amber99sb, but both force fields allowed large conformational sampling and simulations ended in good agreement with each other (Figure S4).

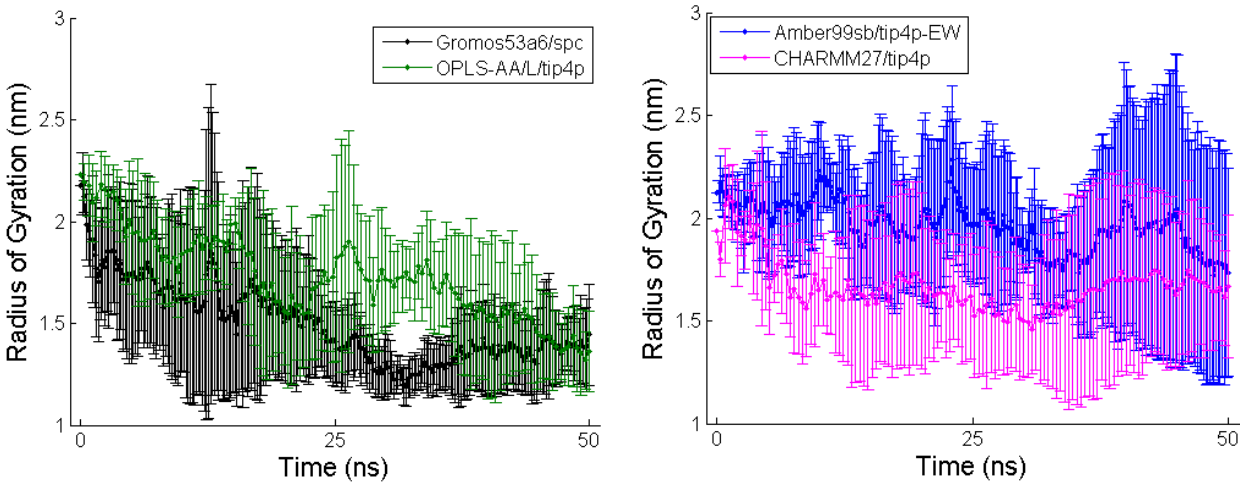

Figure S4. Force fields and water models tested for simulations of TnI_C. The Gromos53a6/spc and OPLS-AA/L/tip4p pairs results in lower radii of gyration and decreased conformational sampling at later time points (left panel) compared with Amber99sb/tip4p-EW and CHARMM27/tip4p (right panel).

Freedom of conformational sampling was also addressed through the radius of gyration over time. The NMR experiments of the Sykes lab document extensive dynamics in this region, consistent with an IDR.[27] Based on general behavior of disordered proteins, we would anticipate that this would result in a frame-by-frame comparison of multiple simulations having a large standard deviation. The force fields that resulted in artificial compaction of the region, Gromos53a6 and OPLS, also resulted in low standard deviations and poor conformational sampling, inconsistent with the literature NMR results. However, both CHARMM27 and Amber99sb had larger standard deviations at later time points.

As expected, the folded regions of the Tn complex (from the 1J1E crystal structure) were less sensitive to force field, with little effect on structure over time with either Amber99sb/tip4p-EW or Gromos53a6/spc (data not shown). Limited breaking and melting of the long helices in the coiled coil region occurred when the TnC/TnI/TnT interface region was outside proximity of a binding partner; however, when both trimers within the 1J1E unit cell were simulated, no breaking of helices occurred. We confirmed that the TnC 1-88 fraction bound to TnI_C remains folded in the Amber99sb force field, and that there was no observable benefit to simulating the full folded complex compared with the TnC 1-88/TnI 147-210 fragment chosen (data not shown). We also tested temperature (310 K vs. 298 K) and

PME cut-off (1.2 nm vs. 0.9 nm) on the 1J1E crystal structure, which again showed no benefit to increasing either (data not shown).

Given the choice between Amber99sb and CHARMM27, we chose Amber99sb based on experience simulating other IDPs. The Amber99sb/tip4p-EW combination resulted in the highest mean radius of gyration as well as the largest standard deviation at later time points.

Correlation of Pairwise Movements Compared with Cross-Correlations

As a means of bridging the experimentally-validated pairwise distances with the cross-correlation matrices calculated from the simulations, we compared the changes in simulated pairwise distances over time with the cross-correlation element representing the relevant pair of residues. For example, searching for a relationship between the 152/179 and 152/209 pairwise distances over simulation time would be searching for a relationship between movement of residue 179 and residue 209. Because these relationships are coarse and specific to the residues used as reference points, the agreement between them and the more refined cross-correlation calculations is imperfect; however, we were able to determine that the relationships were consistent (Figure S5).

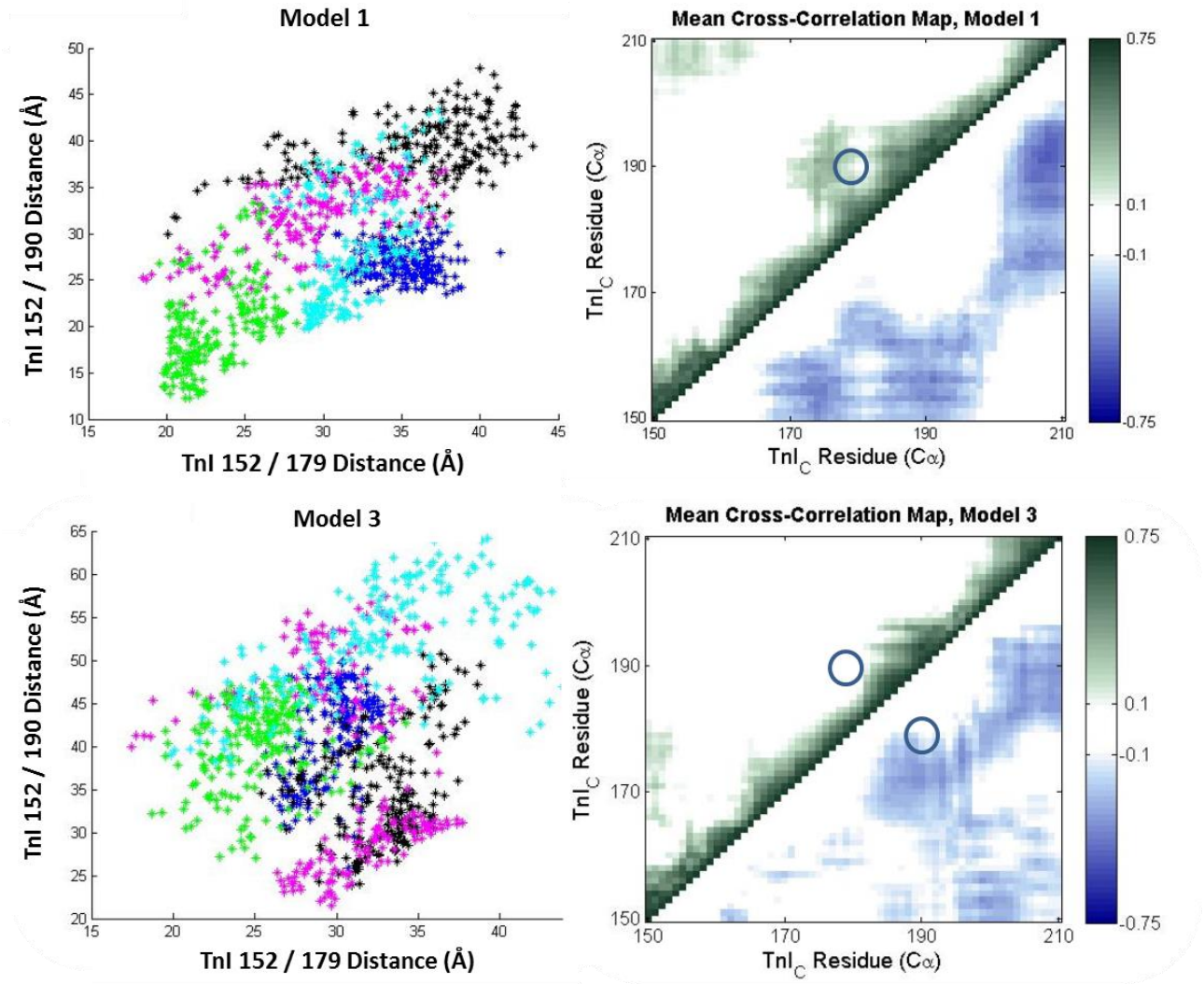

Figure S5. Representative comparisons between cross-correlation matrices (right panels) and rough pairwise distance correlations over time (left panels). Pairwise distances were averaged over 120 ps, then plotted over time for each simulation to generate the rough pairwise distance plots. While sampling is not complete enough for the latter to be a robust analysis, still the positive correlation (upper panels) is clear in both plots. The negative or uncorrelated motions are harder to distinguish due to sampling (lower panels). Circles on the cross-correlation map are centered on the element indicated by the pairwise sampling plots; correlation cut-offs have been relaxed to 0.1 to indicate similarities more clearly.

Simulation of a Wholly Incorrect Starting Model (Negative Control)

As an additional starting model, we created a random coil with the TnI$_C$ 161-210 sequence, then used it to replace those residues in the 1J1E fragment used in the other simulations. The random coil was generated in Python using random torsion angles taken from the allowed space of a Ramachandran plot, then manually adjusted in PyMol as needed to join to the crystal structure fragment. Five

independent MD simulations were run using our standard method, with the exception of our short equilibration run being reduced to 1 ps to increase independence in the trajectories. No linker correction was used due to the frequent reorientation of the chain.

We compared the ET_{eff} calculated from these simulations to our smFRET data, as done with our other experiments. In contrast with other models, the random coil simulation showed overly collapsed conformations at all labeling positions (Figure S6). While longer-distance labeling pairs were able to sample in the mean-plus-1σ area of the histogram, the four shorter distances were unable to sample within 1 σ of the smFRET mean position and showed model bias toward shorter pairwise distances in all tested pairs. This negative control confirms that the success of our comparison of MD simulations and smFRET measurements is not an artifact of fluorophore positioning or labeling site choices.

It is possible that there are alternative TnI_C starting conformations that would be consistent with our smFRET measurements, either independent of time-averaging or assisted by it. However, due to the strong influence of starting models on short MD simulations, we did not attempt to pursue additional simulations started from models that were not experimentally determined.

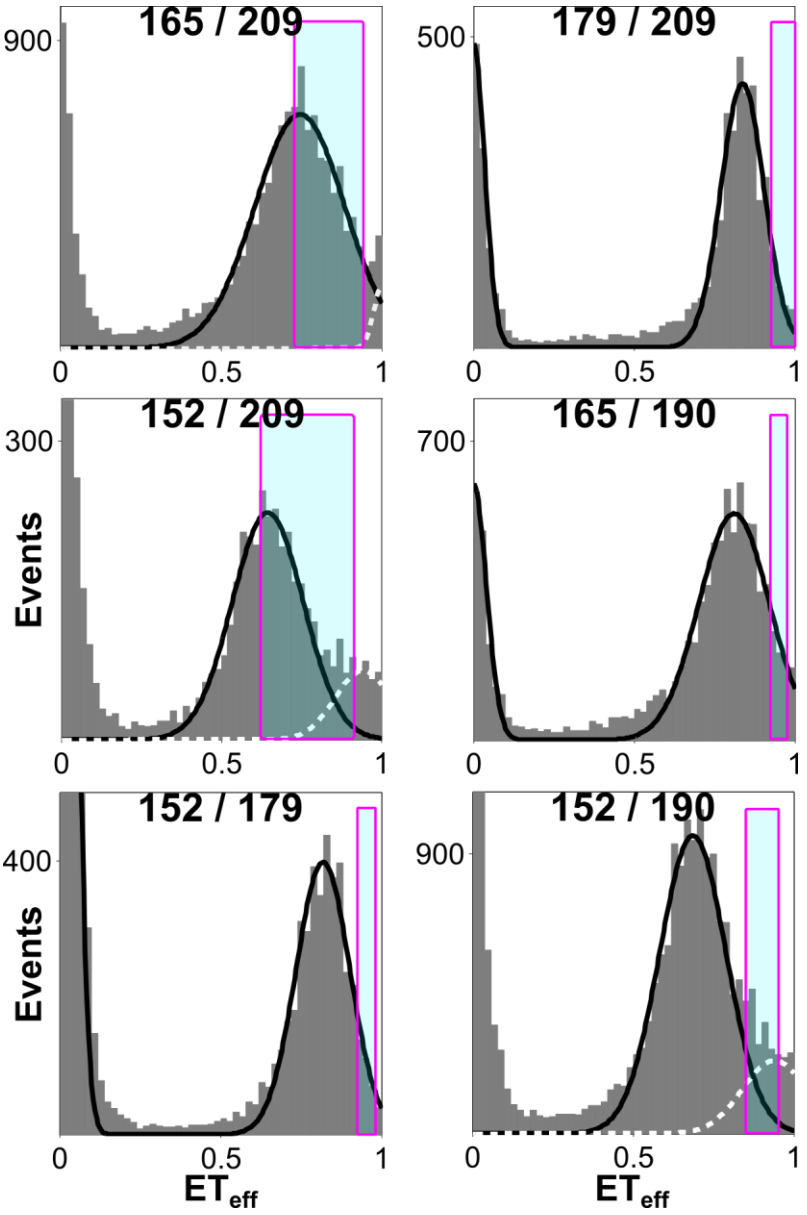

Figure S6. smFRET histograms (gray) overlaid with the ET_{eff} predicted from MD simulations of a random coil model for TnI_C 161-210 (magenta with cyan shading). The width of the overlay reflects the +/- 1 σ bounds for 5 independent simulations. White dashed lines in the smFRET histograms are contributions from TnI monomer and were excluded from analysis.

Troponin Controls

Co-Sedimentation Assays and Densitometry

We confirmed the identity of purified TnC using mass spectrometry, and verified the other components by their co-purifying with TnC during size exclusion and assembling with 1:1:1

stoichiometry based on SDS-PAGE densitometry analysis. We further confirmed that actin filaments pull down the troponin complex with tropomyosin, again with the correct stoichiometry (7:1:1 actin:tropomyosin:troponin). The S1 domain of murine cardiac myosin binds these thin filaments well with ATP dependence, as anticipated (data not shown).

Calcium Regulation of Myosin ATPase

To confirm troponin function following purification of the complex, we performed ATPase assays to monitor calcium-based regulation of myosin accessibility to actin. To assemble thin filaments, 5 µM phalloidin-stabilized actin and 1 µM tropomyosin were incubated on ice for 30 minutes, then 1 µM Tn was added and incubated another 30 minutes. 0.445 µM rabbit skeletal heavy muromyosin (Cytoskeleton.com) was added to tubes, followed by 2 mM ATP at time=0. Reactions were incubated at 37˚ C for 15 minutes, then assessed for free phosphate by a molybdate-based colorimetric assay. These measurements were performed in [10 mM imidazole pH 7.5, 25 mM KCl, 2 mM $MgCl_2$, 1 mM EGTA, 1 mM DTT], with the calcium concentrated adjusted to pCa 3.2 or 8.2 using an on-line calculator.[28] Regulated thin filaments resulted in a strong calcium dependence for myosin ATPase activity, with a 60% decrease in activity when reducing calcium from pCa 3 (the 'open' state) to pCa 8 (the 'blocked' state, Figure S7). The nonzero pCa 8.2 activity reflects background absorbance as well as ATPase activity due to inactive HMM forcing regulatory units into the 'open' conformation.

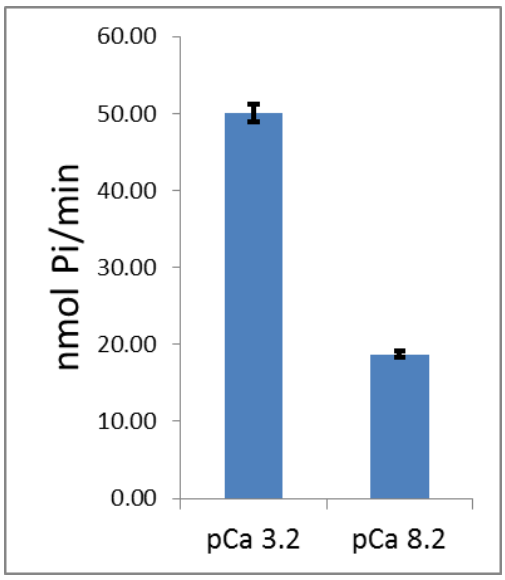

Figure S7. ATPase assays on regulated thin filaments confirm troponin function. Lowering calcium concentrations from pCa 3.2 to pCa 8.2 results in decreased myosin access to actin, reducing the actin-activated ATPase activity of the myosin.

smFRET Reproducibility and Error Reporting

smFRET measurements were highly reproducible, with standard deviation of the mean peak position (Table S4) less than error from sources such as shot noise, linker contribution or fluorophore orientation. Therefore, this error was treated as negligible. Because it was the most generous error metric, we used the standard deviation according to the Gaussian fit of the smFRET peak for analysis purposes and comparison with simulations.

Construct	Mean of Gaussian Fit (Peak Center)	Standard Deviation of Gaussian Fit (Peak Width)	Expected Shot Noise Contribution
TnI 152/179	**0.82 ± 0.005**	**0.083 ± 0.007**	**0.07**
TnI 152/190	**0.71 ± 0.007**	**0.12 ± 0.003**	**0.08**
TnI 152/209	**0.66 ± 0.02**	**0.13 ± 0.01**	**0.09**
TnI 165/190	**0.82 ± 0.007**	**0.11 ± 0.006**	**0.07**

TnI 165/209	**0.75 ± 0.004**	**0.14 ± 0.005**	**0.08**
TnI 179/209	**0.84 ± 0.003**	**0.068 ± 0.004**	**0.07**

Table S4. Means and standard deviations for peak parameters show high reproducibility in smFRET measurements. Expected shot noise contribution is lower than the Gaussian fit standard deviation for all constructs.

The upper limit for shot noise was estimated using the formula

$$\sigma = \sqrt{\frac{ET * (1 - ET)}{T}}$$

(Equation S6)

where T is the threshold for the number of photons defining a smFRET event (30 photons in these experiments). This analysis shows that some of the more narrow distributions fall near the shot noise estimation, indicating that the contribution of other factors is likely low. Other, wider distributions (165/209, 152/209) have widths higher than this estimate, possibly indicating dynamics or conformational heterogeneity. Due to the pitfalls associated with deconvoluting histogram widths, we have not pursued this further.

Additional Experiments

Comparison of smFRET Distances to Models

Before running simulations, we initially compared our smFRET inter-dye distances to models derived from published structures (Models 1-3, with described changes). For this comparison, we assume that published structural data is in a near-minimum energy state, and therefore do not minimize regions resolved in the crystal or NMR structure. Model 1 (1VDJ, converted to cardiac isoform) was not subjected to minimization following I-TASSER modeling; Model 2 (1J1E) was only minimized after residue 190 (residues 196-210 are minimized coil). Model 3 was based fully on homology modeling, and therefore was minimized using the GROMACS protocol described in the Extended Methods section. While we compared all pairwise distances in all models, of course the distances lying within resolved

areas of the structures are the most informative; these are bolded in the following table (Table S5). We also considered distances for an ideal random coil based solely on residue separation.

All modeled distances were converted using AV modeling of the frozen structure, and smFRET distances were considered using 1 standard deviation of the Gaussian peak fit. It is important to note that there is a shot noise contribution to the smFRET histogram that can increase the width of the peak; therefore smFRET uncertainty may be overestimated. Conversely, the random coil calculation lacks a structure, and thus cannot be directly modeled for AV calculations; therefore, linker contributions are an unknown source of error for this comparison but are likely 0-5 Å based on the range from other calculations. Due to these limitations, this snapshot comparison is for the purpose of assessing consistency between models and smFRET calculations only.

Residue Pair	Model 1	Model 2	Model 3	Random Coil	smFRET
TnI 152-179	**41 Å**	**42 Å**	18 Å	34 Å	39 ± 5 Å
TnI 152-209	**67 Å**	79 Å	45 Å	53 Å	46 ± 6 Å
TnI 179-209	**29 Å**	64 Å	**25 Å**	36 Å	38 ± 4 Å
TnI 165-209	**50 Å**	83 Å	53 Å	46 Å	42 ± 7 Å
TnI 152-190	**42 Å**	**46 Å**	22 Å	42 Å	44 ± 6 Å
TnI 165-190	**22 Å**	**46 Å**	28 Å	33 Å	39 ± 5 Å

Table S5. Comparisons of smFRET data to pairwise distances in structural models. Bolded distances in models are those based solely on the original structure (without addition of unresolved residues).

In this comparison, we found that no published structure was fully consistent with all smFRET data. In the case of Model 2, resolved areas approached smFRET data, but the long helix prevented other areas of the model from approaching their experimental values without excessive user input as to placement of the disordered 195-210 extension.

Additional Monte Carlo Analyses

In addition to the analyses reported in the main text, we performed several other standard analyses of the Monte Carlo simulations.

An internal scaling plot of sequence separation vs. inter-residue distance shows that the TnI_C 161-210 sequence shows polymer scaling behavior intermediate between the excluded volume (repulsive Leonnard-Jones potentials) and ideal gas (entropically-driven) polymer models, again appropriate for an IDR (Figure S8). The polymer scaling plot is smooth, without distance-dependent deviations in the shape of the curve.

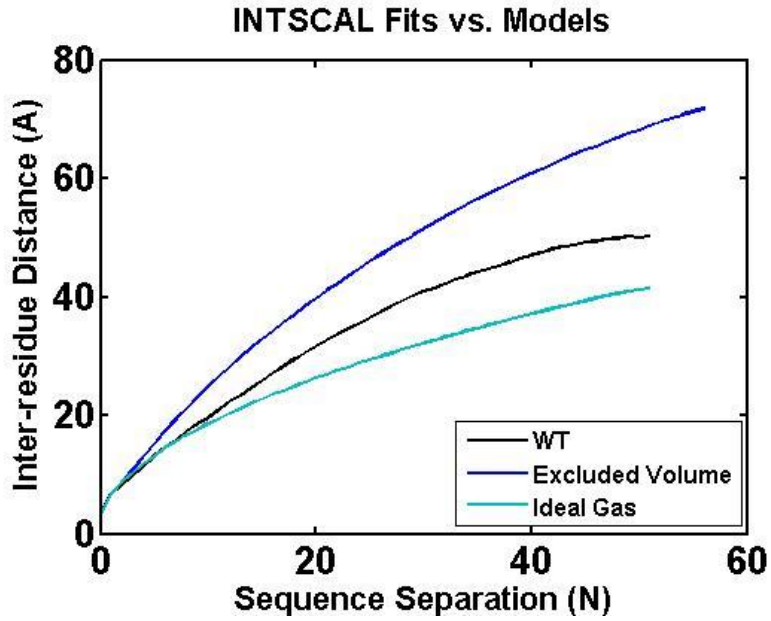

Figure S8. Internal scaling of TnI_C, plotted against the excluded volume and ideal gas polymer models.

More of interest is the behavior of the predominant N-terminal helix. We examined the presence of this helix across the temperatures used in our replica exchange simulations, and found a strong temperature dependence in the range of reasonable experimental measurements (Figure S9). This behavior could in part explain some of the conflicting reports of structure in TnI_C, in which a range of temperatures can be found. AGADIR calculations performed on this region also show a salt

dependence for stabilization of the helix (salt destabilizes the helix),[29] which is another experimental parameter that could vary significantly and impact helix stability in different treatments of the protein for different experiments.

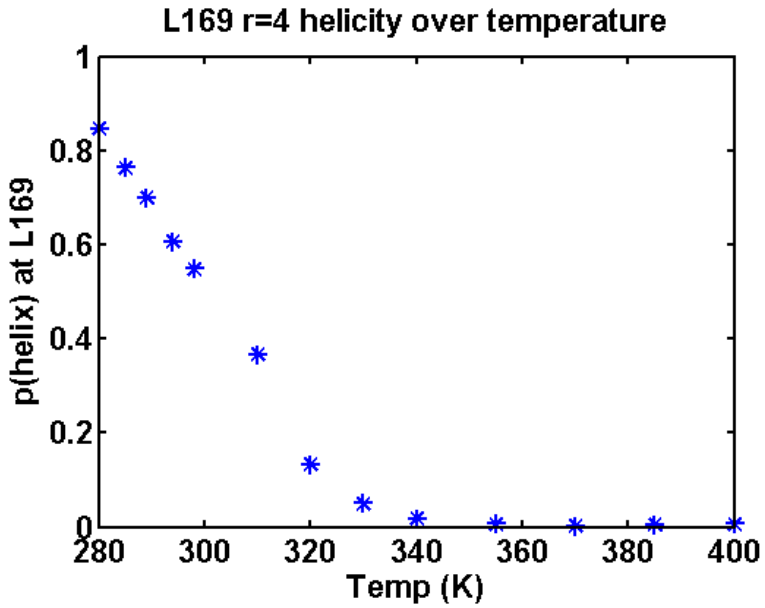

Figure S9. Percent of simulations with at least a 4-residue helix at the N-terminal helix position according to simulation temperature. Room temperature (298 K) is roughly at the center of the temperature dependence curve, making TnI_C highly sensitive to reasonable temperature differences between experimental methods.

While the presence of the helix is of strong interest due to its close proximity to regions with correlated motion, its effect on the structure of TnI_C itself appears to be low, at least *in silico* and in isolation. The presence of a helix at the N-terminal site does not have a statistically significant effect on the probability of a helix at the C-terminal site. However, the helix does have modest effects on the global conformation of TnI_C. The helix has a small but statistically significant effect on the radius of gyration, which is 0.64 Å more compact when a four-residue helix is present compared with cases where it is not ($p<0.0001$ by unpaired T-test). We investigated this further by examining pairwise distances within TnI_C to uncover which areas of the chain were most affected (Figure S10). Conformers with a helix of any length at the N-terminal location experience a roughly 1 Å expansion of the C-terminal

region of the protein, away from the site of the helix. There is a stronger compaction in the N-terminal region, with effects shortening the distance between residues around 170 and residues around 190 by up to 4 Å, but it is difficult to deconvolve the effect of the helix itself and possible allosteric movements in response to its presence.

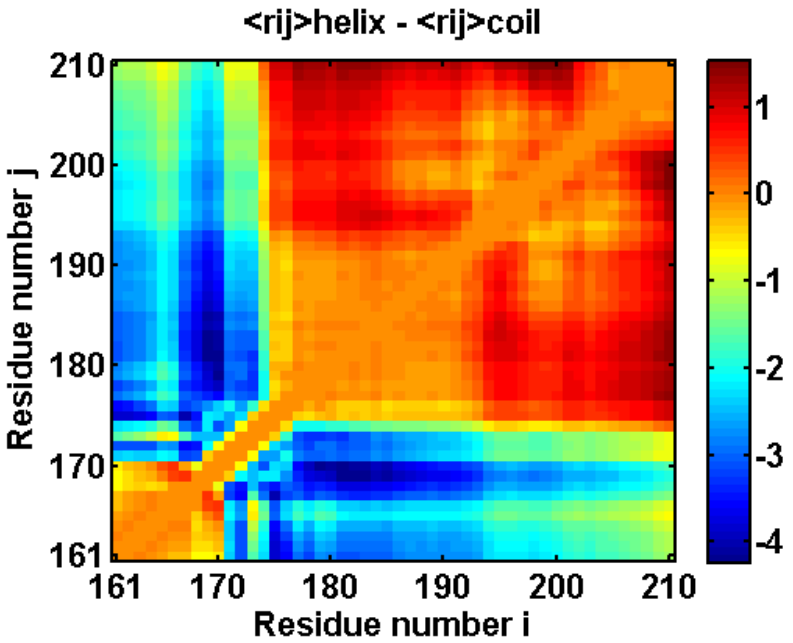

Figure S10. Heat map showing the difference between interresidue distances when the 169-173 region is helical or random coil. In this map, positive numbers indicate an expansion (increased distance when the helix is present), and negative numbers indicate compaction (decreased distance when the helix is present).

The absence of the TnI switch peptide, which immediately precedes the TnI_C 161-210 region and is helical in the high-Ca^{2+} state, may stabilize the helix and serve as a way of communicating the calcium state of the complex across TnI_C. While our results were inconclusive as to the effect of the helix, this remains a topic for further research.

References

(1) Kruger, M.; Pfitzer, G.; Stehle, R. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2003**, *786*, 287.
(2) Robinson, J. M.; Cheung, H. C.; Dong, W. *Biophys J* **2008**, *95*, 4772.
(3) Spudich, J. A.; Watt, S. *J Biol Chem* **1971**, *246*, 4866.

(4) Simile, L. B. *Methods Enzymol* **1982**, *85*, 234.
(5) Ramamurthy, B.; Cao, W.; De la Cruz, E. M.; Mooseker, M. S. *Cytoskeleton (Hoboken)* **2012**, *69*, 59.
(6) Margossian, S. S.; Lowey, S. *Methods Enzymol* **1982**, *85 Pt B*, 55.
(7) Trexler, A. J.; Rhoades, E. *Biochemistry* **2009**, *48*, 2304.
(8) Middleton, E. R.; Rhoades, E. *Biophys J* **2010**, *99*, 2279.
(9) Nath, A.; Sammalkorpi, M.; DeWitt, D. C.; Trexler, A. J.; Elbaum-Garfinkle, S.; O'Hern, C. S.; Rhoades, E. *Biophys J* **2012**, *103*, 1940.
(10) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *Journal of Chemical Theory and Computation* **2008**, *4*, 435.
(11) Takeda, S.; Yamashita, A.; Maeda, K.; Maeda, Y. *Nature* **2003**, *424*, 35.
(12) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712.
(13) Sorin, E. J.; Pande, V. S. *Biophys J* **2005**, *88*, 2472.
(14) 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.
(15) Ichiye, T.; Karplus, M. *Proteins* **1991**, *11*, 205.
(16) 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.
(17) Schuler, B.; Lipman, E. A.; Eaton, W. A. *Nature* **2002**, *419*, 743.
(18) Vitalis, A.; Pappu, R. V. *Annu Rep Comput Chem* **2009**, *5*, 49.
(19) Vitalis, A.; Pappu, R. V. *J Comput Chem* **2009**, *30*, 673.
(20) Kapanidis, A. N.; Lee, N. K.; Laurence, T. A.; Doose, S.; Margeat, E.; Weiss, S. *Proc Natl Acad Sci U S A* **2004**, *101*, 8936.
(21) Showalter, S. A.; Johnson, E.; Rance, M.; Bruschweiler, R. *J Am Chem Soc* **2007**, *129*, 14146.
(22) Sgourakis, N. G.; Merced-Serrano, M.; Boutsidis, C.; Drineas, P.; Du, Z.; Wang, C.; Garcia, A. E. *J Mol Biol* **2011**, *405*, 570.
(23) Cino, E. A.; Choy, W. Y.; Karttunen, M. *Journal of Chemical Theory and Computation* **2012**, *8*, 2725.
(24) Lange, O. F.; van der Spoel, D.; de Groot, B. L. *Biophys J* **2010**, *99*, 647.
(25) Trbovic, N.; Kim, B.; Friesner, R. A.; Palmer, A. G., 3rd *Proteins* **2008**, *71*, 684.
(26) Wickstrom, L.; Okur, A.; Simmerling, C. *Biophys J* **2009**, *97*, 853.
(27) Julien, O.; Mercier, P.; Allen, C. N.; Fisette, O.; Ramos, C. H.; Lague, P.; Blumenschein, T. M.; Sykes, B. D. *Proteins* **2011**, *79*, 1240.
(28) Patton, T., Epel *Cell Calcium* **2010**, *99*, 1.
(29) Munoz, V.; Serrano, L. *Nat Struct Biol* **1994**, *1*, 399.

The calculator cited as reference #21 is located at maxchelator.stanford.edu; the program used was the Ca-Mg-ATP-EGTA Calculator v1.0 using constants from NIST database #46 v8.