

Simulation of FRET dyes allows quantitative comparison against experimental data

Ines Reinartz, Claude Sinner, Daniel Nettels, Brigitte Stucki-Buchli, Florian Stockmar, Pawel T. Panek, Christoph R. Jacob, Gerd Ulrich Nienhaus, Benjamin Schuler, and Alexander Schug

Citation: *The Journal of Chemical Physics* **148**, 123321 (2018);

View online: <https://doi.org/10.1063/1.5010434>

View Table of Contents: <http://aip.scitation.org/toc/jcp/148/12>

Published by the [American Institute of Physics](http://www.aip.org)



Simulation of FRET dyes allows quantitative comparison against experimental data

Ines Reinartz,^{1,2} Claude Sinner,^{1,2} Daniel Nettels,³ Brigitte Stucki-Buchli,³ Florian Stockmar,⁴ Pawel T. Panek,⁵ Christoph R. Jacob,⁵ Gerd Ulrich Nienhaus,^{4,6,7,8} Benjamin Schuler,^{3,9} and Alexander Schug^{2,10}

¹Department of Physics, Karlsruhe Institute of Technology, Wolfgang-Gaede-Str. 1, 76131 Karlsruhe, Germany

²Steinbuch Centre for Computing, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

³Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

⁴Institute of Applied Physics, Karlsruhe Institute of Technology, Wolfgang-Gaede-Str. 1, 76131 Karlsruhe, Germany

⁵Institute of Physical and Theoretical Chemistry, TU Braunschweig, Gaußstraße 17, 38106 Braunschweig, Germany

⁶HEiKA–Heidelberg Karlsruhe Research Partnership, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

⁷Institute of Nanotechnology and Institute of Toxicology and Genetics, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

⁸Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

⁹Department of Physics, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

¹⁰John von Neumann Institute for Computing, Jülich Supercomputing Centre, Forschungszentrum Jülich, 52425 Jülich, Germany

(Received 25 October 2017; accepted 18 December 2017; published online 9 January 2018)

Fully understanding biomolecular function requires detailed insight into the systems' structural dynamics. Powerful experimental techniques such as single molecule Förster Resonance Energy Transfer (FRET) provide access to such dynamic information yet have to be carefully interpreted. Molecular simulations can complement these experiments but typically face limits in accessing slow time scales and large or unstructured systems. Here, we introduce a coarse-grained simulation technique that tackles these challenges. While requiring only few parameters, we maintain full protein flexibility and include all heavy atoms of proteins, linkers, and dyes. We are able to sufficiently reduce computational demands to simulate large or heterogeneous structural dynamics and ensembles on slow time scales found in, e.g., protein folding. The simulations allow for calculating FRET efficiencies which quantitatively agree with experimentally determined values. By providing atomically resolved trajectories, this work supports the planning and microscopic interpretation of experiments. Overall, these results highlight how simulations and experiments can complement each other leading to new insights into biomolecular dynamics and function. © 2018 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/1.5010434>

I. INTRODUCTION

Life at the molecular level is determined by the interplay of structure and dynamics. While a biomolecular structure can be determined by sophisticated experimental techniques such as NMR or X-ray, detailed insight into its dynamics is more difficult to access. A particularly powerful technique is FRET (Förster Resonance Energy Transfer), where specific sites in biomolecules are labeled by fluorescent dyes. FRET can provide time-resolved information about, e.g., protein-folding dynamics,¹ folding intermediates,^{2,3} unfolded or disordered proteins,^{4–6} or conformational transitions.⁷ Both planning and interpretation of such experiments can be nontrivial, as FRET does not directly provide quantitative distance information within the protein. Instead, it rather measures the transfer efficiency between the dyes which—besides the distance between the dyes—depends on their orientation and

dynamics and hence also on their shapes. A useful complement for such experiments is molecular simulations of dye-labeled biomolecules. Depending on their complexity, additional information up to the full dynamics of all atoms becomes accessible.

To date, many simulation approaches are available. A straightforward approach is predicting the inter-dye distance by estimating the accessible volume of FRET dyes.^{8–12} This does, however, not account for dynamics and the mutual dye orientation, reflected by the orientation factor κ^2 . An approach that has been used extensively for unfolded and intrinsically disordered proteins is the interpretation of FRET data in terms of analytical polymer models,¹³ but such models are not applicable to folded proteins. Molecular dynamics simulations are computationally considerably more demanding and aim at modeling the system dynamics accurately. In this approach, the dyes are typically parametrized for a specific

biomolecular force field.¹⁴ Then, simulations on the ns to μ s time scale are run and compared to experimentally accessible properties, such as fluorescence anisotropy¹⁵ or mean FRET efficiencies.¹⁴ As one might imagine, dye and linker fluctuations play an important role.¹⁶ Similarly, the specific dye shape and mutual orientation are crucial and are not fully captured by simple accessible volume calculations.¹⁷ Molecular dynamics simulations also allow testing assumptions such as approximating the mutual dye-dye orientation factor κ^2 with $2/3$, which is questioned in several studies,^{18,19} in particular, for low linker flexibilities.²⁰ Many simulation protocols have computationally prohibitive demands to reach sufficiently long simulation times for accurate calculations of experimentally accessible properties.²¹ In particular, it is challenging to account for slow or large-scale conformational transitions, structurally diverse ensembles such as unfolded proteins, very large systems, or intrinsically disordered proteins.

Here, we want to introduce a new method for simulating the dynamics of FRET-labeled proteins on a coarse-grained description level²² which maintains full protein flexibility and all heavy atoms of proteins and dyes. This considerably lowers computational costs and allows us to eliminate many challenges regarding the mentioned insufficient sampling of large conformational ensembles. Our model allows us to access and quantitatively analyze these complex dynamic ensembles in atomic detail. The model requires only few and robust parameters to achieve a realistic description of the system. We validate our approach by direct and quantitative comparison against experimental FRET efficiency histograms. As an example, we highlight simulating large and diverse structural ensembles by comparing simulations and measured FRET data of both folded and unfolded proteins. Having established the precision of our model, we then compare our model to descriptions of FRET by the accessible volume approach and analytical polymer models.

II. THEORY

A. Förster resonance energy transfer

Förster Resonance Energy Transfer (FRET)²³ is a non-radiative energy transfer between a donor and an acceptor fluorophore. It has a high sensitivity to distances in the nanometer range and is therefore a powerful “spectroscopic ruler”.²⁴ For example, one can label specific residues of proteins with suitable donor and acceptor fluorophores to distinguish different protein conformations and conformational changes directly through distance changes between the dyes.

FRET experiments measure the FRET efficiency E , which is related to the inter-dye distance R_{DA} by²⁵

$$E = \frac{1}{1 + \left(\frac{R_{DA}}{R_0}\right)^6}. \quad (1)$$

The Förster radius R_0 determines the range of measurable distances and is given by²⁵

$$R_0^6 = \frac{9(\ln 10)\kappa^2 Q_D J(\lambda)}{128\pi^5 n^4 N_A}, \quad (2)$$

where Q_D is the fluorescence quantum yield of the donor in the absence of the acceptor, κ^2 is the dipole orientation factor, n is the refractive index of the medium, N_A is the Avogadro constant, and $J(\lambda)$ is the spectral overlap integral of donor emission and acceptor absorption spectra. The orientation factor κ^2 describes the relative orientation of the transition dipole moments for emission of the donor, μ_D , and absorption of the acceptor, μ_A , and can be calculated via²⁵

$$\kappa^2 = (\sin \theta_D \sin \theta_A \cos \phi - 2 \cos \theta_D \cos \theta_A)^2, \quad (3)$$

where θ_D and θ_A are the angles between μ_D and μ_A and the connecting vector R_{DA} between the two dye centers, respectively. ϕ denotes the angle between the planes defined by μ_A , R_{DA} and μ_D , R_{DA} , respectively.

Most experimental studies assume fast rotational diffusion of the dyes with respect to the excited state lifetime, resulting in an averaging over all possible orientations. This is referred to as the “isotropic averaging regime,” in which case a constant value of $\kappa^2 = 2/3$ can be used.²⁵ The assumption of $\kappa^2 = 2/3$ has been questioned in several studies^{12,18–20} but can be tested experimentally, e.g., by measuring time-dependent fluorescence anisotropy decays.¹⁴

B. Rotational correlation time

A sample of randomly oriented dyes illuminated by polarized light depolarizes over time as a consequence of rotational diffusion. Observing the time dependent fluorescence anisotropy $r(t)$ leads to a measure of the flexibility and rotational speed of the dyes, the rotational correlation time τ_{rot} . The time-resolved fluorescence anisotropy can be measured experimentally and calculated from simulations using the normalized transition dipole moment vector $\hat{\mu}$ via²⁶

$$r(t) = r_0 \langle P_2 [\hat{\mu}(s) \cdot \hat{\mu}(s+t)] \rangle, \quad (4)$$

where r_0 is the fundamental anisotropy and P_2 is the second-order Legendre polynomial. $r(t)$ allows a direct comparison of experimental observations with simulations. Here, we assume that the absorption and emission dipole moments are collinear for our dyes, so the fundamental anisotropy is given by $r_0 = 0.4$, close to the experimental value.²⁷

In the analysis of experimental data, often a double-exponential decay is required to account for the contributions of rotational motion of both the fluorophore and the protein.^{14,26,28} In the systems we investigate here, the proteins are large in comparison to the dyes and the measured dye rotational correlation times are one order of magnitude smaller than the expected rotational correlation time of the proteins. In the simulations, we focus on the rotational motion of the dye in the inertial system of the protein, where we find it sufficient to use only one exponential and fit with

$$r(t) = r_0 \exp\left[-\frac{t}{\tau_{\text{rot}}}\right]. \quad (5)$$

C. Structure based models

For our simulations, we use native Structure Based Models (SBMs), also referred to as Gō-type models.²⁹ They

are based on energy landscape theory and the principle of minimal frustration.^{30–33} SBMs are used successfully to study a wide range of phenomena³⁴ ranging from, e.g., protein folding,^{35,36} misfolding,^{37,38} structure prediction,³⁹ and conformational dynamics^{40,41} to large biomolecules such as the ribosome⁴² or RNA.⁴³ SBM simulations show good agreement with experimental results, e.g., they are used to reproduce transition state ensembles and “en-route” intermediates,⁴⁴ and folding rates comparable to experimental measurements.³⁵ By design, they are limited to unfrustrated systems with negligible non-native interactions.^{29–33} With their high computational efficiency, SBMs facilitate simulations of several folding and unfolding transitions on regular desktop computers while offering full flexibility for all parts of the protein. The SBM potential employs the native structure of the protein as its ground state but also allows one to model systems with multiple stable conformations by using potentials with multiple minima.^{7,40,41,45}

We use an all-atom SBM²² implemented in eSBMTools,⁴⁶ which takes into account all heavy atoms of the protein and uses a simplified potential of the form⁴⁷

$$V = \sum_{\text{bonds}} K_b(r - r_0)^2 + \sum_{\text{angles}} K_a(\theta - \theta_0)^2 + \sum_{\text{impr. d.}} K_i(\chi - \chi_0)^2 + \sum_{\text{prop. d.}} K_d \left[[1 - \cos(\phi - \phi_0)] + \frac{1}{2} [1 - \cos(3(\phi - \phi_0))] \right] + \sum_{\text{contacts}} K_c C_G(r_{ij}, r_0^{ij}) + \sum_{\text{non-native contacts}} K_{nc} \left(\frac{\tilde{\sigma}}{r_{ij}} \right)^{12}. \quad (6)$$

The potential includes harmonic potentials for bonds, angles, and improper dihedral angles and a potential for proper dihedral angles which allows the occupation of isomeric conformations. The minima are set to r_0 , θ_0 , χ_0 , and ϕ_0 , the values for bond lengths, angles, improper dihedral angles, and proper dihedral angles in the native state, respectively. The contact potential $C_G(r_{ij}, r_0^{ij})$ stabilizes the native structure by introducing attractive potentials for atom pairs forming contacts in the native state, whereas an overall repulsive term is included for the excluded volume of all atoms. r_0^{ij} and r_{ij} are the native and the actual distances of the atom pair (i, j) and $\tilde{\sigma}$ represents the excluded volume for Pauli repulsion with the value $\tilde{\sigma} = 2.5$ Å. The energetic weights for bonds, angles, improper dihedral angles, and non-native contacts are set to the values $K_b = 20\,000$ ϵ/nm^2 , $K_a = 40$ ϵ/deg , $K_i = 40$ ϵ/deg , and $K_{nc} = 0.01$ ϵ , where deg refers to degree and ϵ is the reduced energy unit.⁴⁸ The energetic weights for the contact potential K_c and for the proper dihedral angle potential K_d are determined as described in Ref. 48. The Gaussian contact potential is given by^{47,49}

$$C_G(r_{ij}, r_0^{ij}) = \left(1 + \left(\frac{\tilde{\sigma}}{r_{ij}} \right)^{12} \right) \left(1 - \exp \left[-\frac{(r_{ij} - r_0^{ij})^2}{2\sigma^2} \right] \right) - 1, \quad (7)$$

with $\sigma^2 = (r_0^{ij})^2 / (50 \ln 2)$ for each native contact pair (i, j) determined by the shadow map algorithm.⁴⁷ All atoms are handled equally with identical parameters for the excluded volume and a unit mass of $m = 1.0$.

As SBMs do not have an inherent temperature scale comparable to physical temperature, the temperatures are typically

measured in reduced units. Accordingly, the folding temperature in the SBMs is about 1.0, corresponding to around $T = 120$ in GROMACS units.⁴⁸ Here, we report the temperature in GROMACS units.

III. METHODS

A. Dye structure

We want to find a minimal and robust set of parameters describing the dyes which is sufficient to replicate and predict experimental data.

SBM simulations require an initial structure of the system. For many dyes, only chemical structures are available. As we want a systematic and consistent workflow, we perform quantum-chemical calculations to get three-dimensional structures of the dyes. Details and the structures of the Alexa Fluor dyes with available chemical structures are given in the [supplementary material](#) (see SI 1). Subsequently we add the respective linkers and maleimide which are used to bind the dyes to the protein. In this study, we use two dye pairs—the Alexa Fluor 488 dye with C_5 -linker (AF488) and Alexa Fluor 594 dye with C_5 -linker (AF594), and the Alexa Fluor 546 dye with C_5 -linker (AF546) and Alexa Fluor 647 dye with C_2 -linker (AF647). In general, every dye and linker with an available chemical structure can be utilized accordingly.

B. Dye parameters

For the initial dye parameters, we choose the same parametrization as for the protein in the SBM. We determine the angles and dihedral angles automatically from the bond information given in the chemical structure. A detailed description can be found in the [supplementary material](#) (see SI 2). The only interaction between both dyes and also between dyes and protein is the excluded volume repulsion term.

SBMs do not have an inherent temperature scale. To adjust the dye parameters independently of the protein in a consistent way, we set the dye temperature T_{dye} separately from the protein temperature T . In the simulation, we treat dyes and protein as different groups and couple them to separate temperature baths. This has the advantage of uncoupling the behavior of dyes and protein, so we can use the same dye temperature regardless of investigating, e.g., a folded or an unfolded protein, realized by low and high protein temperatures, respectively. We tested the effect of the dye temperature on the protein dynamics and found only a slight increase of the fluctuations of the respective residues the dyes are attached to, whereas the dynamics of the adjacent residues remain unaltered. Due to numerical instabilities,⁵⁰ we cannot sufficiently increase the temperature to account for the high dye mobility. Hence, we additionally change the mass of the dye atoms to 0.2, rather than the unit atom mass of 1.0 the SBM assigns to the protein.

C. Merging dyes and protein

The protein structures used for the simulations are taken from the Protein Data Bank (PDB).⁵¹ The three used proteins are the chymotrypsin inhibitor 2 (CI-2, PDB: 2CI2⁵²) as a test system, the tenth type III module of fibronectin (¹⁰FNIII,

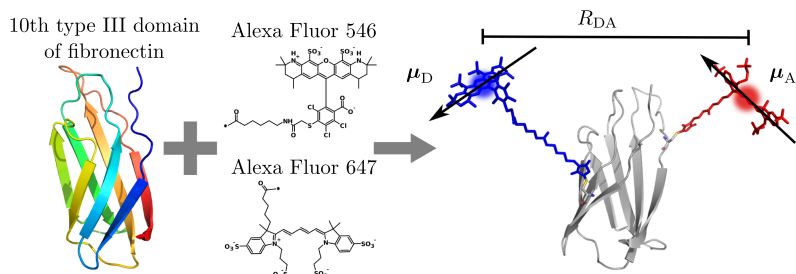


FIG. 1. Structures of 10 FNIII (left) and AF546 and AF647 dyes⁵⁵ (middle). The dyes are attached with the respective linkers via a maleimide bond to residues 11 and 86 of 10 FNIII. The merged structure is shown with the assumed transition dipole moments (black arrows) for donor (blue, μ_D) and acceptor (red, μ_A) dyes as well as the distance between the dyes' centers (R_{DA}) (right).

PDB: 1TTG⁵³), and the cold-shock protein from the hyperthermophilic bacterium *Thermotoga maritima* (CspTm, PDB: 1G6P⁵⁴).

The Alexa Fluor dyes are typically attached to cysteine residues of proteins via a maleimide group. In our simulations, the respective residue in the protein is replaced by a cysteine and the dye-maleimide structure is attached to the sulfur atom of the cysteine. This is done by attaching the dye preferably orthogonally to the protein surface while avoiding steric clashes between atoms. As an example, 10 FNIII with AF546 and AF647 attached to residues 11 and 86, respectively, is shown in Fig. 1. Details of the procedure can be found in the [supplementary material](#) (see SI 2).

D. Simulation protocol

We perform our simulations using GROMACS⁵⁶ with the SBM potential [see Eq. (6)] and the Langevin dynamics simulation protocol.⁵⁶ As we use CI-2 as a simple test system, we choose a protein temperature of $T = 50$ below the folding temperature for the simulation of the folded state.

For the proteins CspTm and 10 FNIII, we want to achieve a quantitative comparison against experimental measurements. To identify the SBM temperature for the folded states corresponding to the experimental setup, we perform an initial all-atom simulation in the AMBER99 force field:⁵⁷ We choose the SBM protein temperature by comparing the native-basin fluctuations of the proteins simulated with the AMBER99 force field at the physiological temperatures used in the experiments [$T = 296/295$ K (10 FNIII/CspTm)] with SBM simulations of varying temperatures. We compare the root mean square fluctuations of the C_α -atoms and choose the SBM temperature which results in the least deviation from the AMBER99 simulation, according to Ref. 58. For the results, see Table I. For the unfolded state, we choose a temperature well above the folding temperature so that the protein is unfolded throughout the simulation. We check the influence of choosing different

temperatures above the folding temperature, which does not alter the results. Future work will focus on the challenges of simulating conformational transitions (cf. Ref. 59).

To determine appropriate dye temperatures, we perform simulations with different dye temperatures and determine the respective rotational correlation times by using Eqs. (4) and (5). The temperature directly affects the dye mobility and the rotational correlation time τ_{rot} . Higher temperatures lead to smaller τ_{rot} . To obtain a high dye mobility, we choose the temperature for the faster rotating dye at $T_{dye} = 250$ and for the slower dye such that the ratio of both rotational correlation times from experiments is matched (see Table S1 in the [supplementary material](#)). For AF647, we do not have an experimental value of τ_{rot} and use the same temperature as for AF546. Varying the dye temperature slightly does not alter the final results.

We then perform simulations with 10^9 steps and $dt = 0.0005$ for CspTm and with 2.5×10^9 steps and $dt = 0.0002$ for CI-2 and 10 FNIII. The temperature coupling constant is set to $\tau_T = 0.1$. From the simulations, we extract the inter-dye distance $R_{DA}(t)$ and the transition dipole moments of donor $\mu_D(t)$ and acceptor $\mu_A(t)$ (see Fig. 1, for details see SI 3 of the [supplementary material](#)).

SBMs do not have a time scale directly comparable to the physical time scale. To evaluate simulated trajectories and obtain FRET efficiency histograms, we introduce a time scale based on a comparison of the experimental and theoretical rotational correlation times τ_{rot} of the dyes. We calculate the fluorescence anisotropy decay $r(t)$ from simulations according to Eq. (4) and fit it with Eq. (5). From the comparison, we get a conversion factor to adjust the trajectory time scale.

This time scale is necessary as we want to calculate FRET efficiency histograms directly comparable to experiments. Photon statistics are gained from Monte Carlo photon simulations (as described in Refs. 8, 60, and 61) and fully account for shot-noise. The simulation parameters are given in Table S1 in the [supplementary material](#). We generate

TABLE I. Simulation parameters for different protein/dye combinations. The protein temperatures for simulation of the folded T_F and the unfolded T_U states, the dye pair, the dye-labeled residues, and the dye temperature(s) are given.

| Protein | T_F | T_U | Dye pair | | T_{dye} |
|---------------|-------|-------|----------------|------------------------------------|-----------|
| | | | Donor/acceptor | Labeled residues Donor/acceptor | |
| CI-2 | 50 | 170 | AF546/AF647 | 20/78 | 250 |
| CI-2 | 50 | 170 | AF488/AF594 | 20/78 | 190/250 |
| CspTm | 90 | 150 | AF488/AF594 | 2/68, 68/2, 11/68, 68/11, 23/68 | 190/250 |
| 10 FNIII | 60 | 200 | AF546/AF647 | 11/86 | 250 |

donor and acceptor photons for the whole simulated trajectory, corresponding to around 100 μs on the physical dye time scale, until a specified number of photons (the burst size) is collected. Each of these bursts is used to calculate a single FRET efficiency value. Details are given in the [supplementary material](#) (see SI 3).

We perform simulations with three different proteins (CspTm, CI-2, and $^{10}\text{FNIII}$) and two different dye pairs (AF488 and AF594, AF546 and AF647). Because of the dependence of the Förster radius on the refractive index of the medium, R_0 changes slightly for higher denaturant concentrations in the experiments. Hence, for evaluation of the unfolded state, we use the changed Förster radii (see Table S1 in the [supplementary material](#)).

IV. SIMULATION RESULTS

A. Test simulations with CI-2

As an example, we simulate the test system CI-2. The resulting distributions for CI-2 with two different dye pairs, attached at residues 20 and 78, respectively, are shown in Fig. 2. The inter-dye distances R_{DA} [see Figs. 2(a) and 2(b)] and the

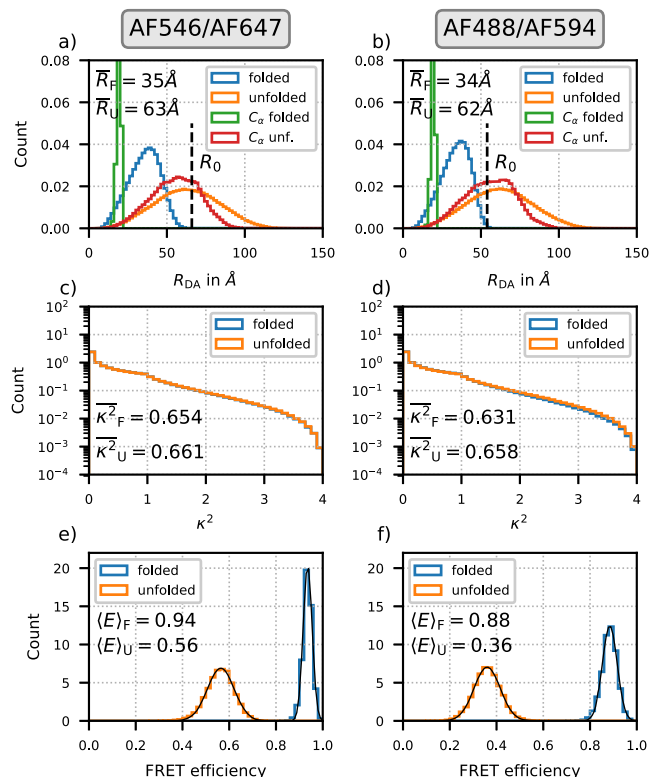


FIG. 2. Simulation results for CI-2 for two different dye pairs—AF546 and AF647 (left) and AF488 and AF594 (right)—attached to residues 20 and 78, respectively. The results are shown for simulations of the folded state (blue) and the unfolded state (orange). [(a) and (b)] Distributions of the inter-dye distances R_{DA} and the average distances for folded (\bar{R}_F) and unfolded states (\bar{R}_U) and respective Förster radii R_0 (dashed black lines). Further, the respective distributions of C_{α} -distances between residues 20 and 78 for folded (green) and unfolded (red) states are shown. [(c) and (d)] Distributions of the orientation factors κ^2 and the average orientation factors for folded ($\overline{\kappa^2}_F$) and unfolded ($\overline{\kappa^2}_U$) states. [(e) and (f)] Distributions of the FRET efficiencies with a Gaussian fit (black lines) and the positions of the peaks for folded ($\langle E \rangle_F$) and unfolded ($\langle E \rangle_U$) states.

orientation factors κ^2 [see Figs. 2(c) and 2(d)] for folded and unfolded conformations along with the respective mean values are depicted. We have also carefully investigated the effect of sampling (see SI 4 of the [supplementary material](#)).

Clearly, the unfolded state has a much broader distance distribution, whereas the folded state is narrower. Comparison of the distances of the C_{α} -atoms with the inter-dye distances shows that the distribution of R_{DA} in the folded state is dominated by dye dynamics. For both dye pairs, the distance distributions are similar. The ideal Förster radius to get best-separated states would be between the average distances of folded and unfolded states. Considering the Förster radii of the two dye pairs used here, the distance distributions show that the pair AF488/AF594 is better suitable for studying CI-2 experimentally.

The distributions of the orientation factors in the folded and unfolded states are similar, and the mean values $\overline{\kappa^2}_U$ are in good agreement with the approximation of $\overline{\kappa^2} = 2/3$. As expected, in the case of unfolded proteins, the dyes can rotate freely and their orientations are almost unrestricted. However, we observe slightly larger deviations of 2%-5% for the mean values in the folded state $\overline{\kappa^2}_F$, probably due to steric restrictions imposed by the protein.

The resulting FRET efficiency histograms along with the mean efficiency values calculated by Gaussian fits are illustrated in Figs. 2(e) and 2(f). As an alternative, log-normal

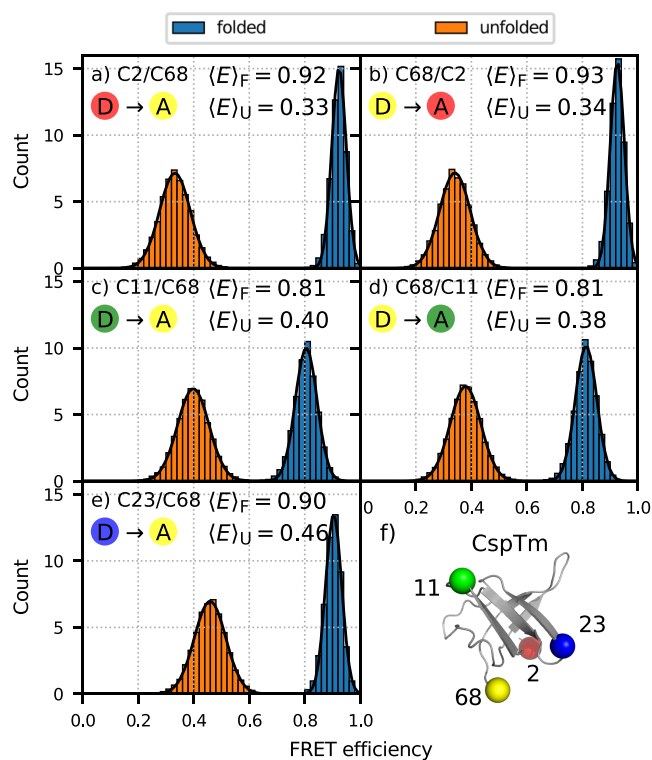


FIG. 3. FRET efficiency distributions for CspTm with AF488 (donor) and AF594 (acceptor) at different labeling sites. The labeling positions of donor (D) and acceptor (A) are color coded in (a)–(e). Colors refer to the positions marked in (f), the residues 2 (red), 11 (green), 23 (blue), and 68 (yellow). The residue numbering within the protein is given in the [supplementary material](#) (see SI 6). The distributions are shown for the folded (blue) and unfolded (orange) states and fitted with Gaussians. The positions of the peaks are given for the folded ($\langle E \rangle_F$) and unfolded ($\langle E \rangle_U$) states.

distributions or a straightforward median can be used, but this does only have negligible effects on the quantitative results. Different hypothetical Förster radii can be tested with this model (see SI 5 of the [supplementary material](#)).

B. CspTm with different dye positions

Another advantage of the efficient simulation approach we propose is that different labeling positions can readily be implemented and investigated. As an example, we perform simulations for CspTm with AF488 and AF594 at different labeling positions. The resulting FRET efficiency histograms are shown in Fig. 3. The labeling sites are the residue pairs C2/C68, C68/C2, C11/C68, C68/C11, and C23/C68 for donor/acceptor,^{62,63} respectively (for the numbering scheme, see SI 6 of the [supplementary material](#)). As expected, the peaks for the unfolded state shift to higher efficiencies with

shorter sequence separation. Also, permutations of the dyes with respect to the attachment points have, unsurprisingly, no relevance.

Both results (see Figs. 2 and 3) show the strength of the simulation method as it establishes a way to directly relate distance distributions to FRET efficiency distributions and allows one to vary and test different parameters such as the Förster radius, dye pair, linker length, or labeling sites. This facilitates improvement in planning, interpreting, and validating experimental results.

C. Comparison of simulation and experiment

As a validation of our model, we compare the resulting FRET efficiency histograms against experimental data. FRET efficiency histograms from experiments and simulations for four different systems are shown in Fig. 4. Figure 4(a) shows

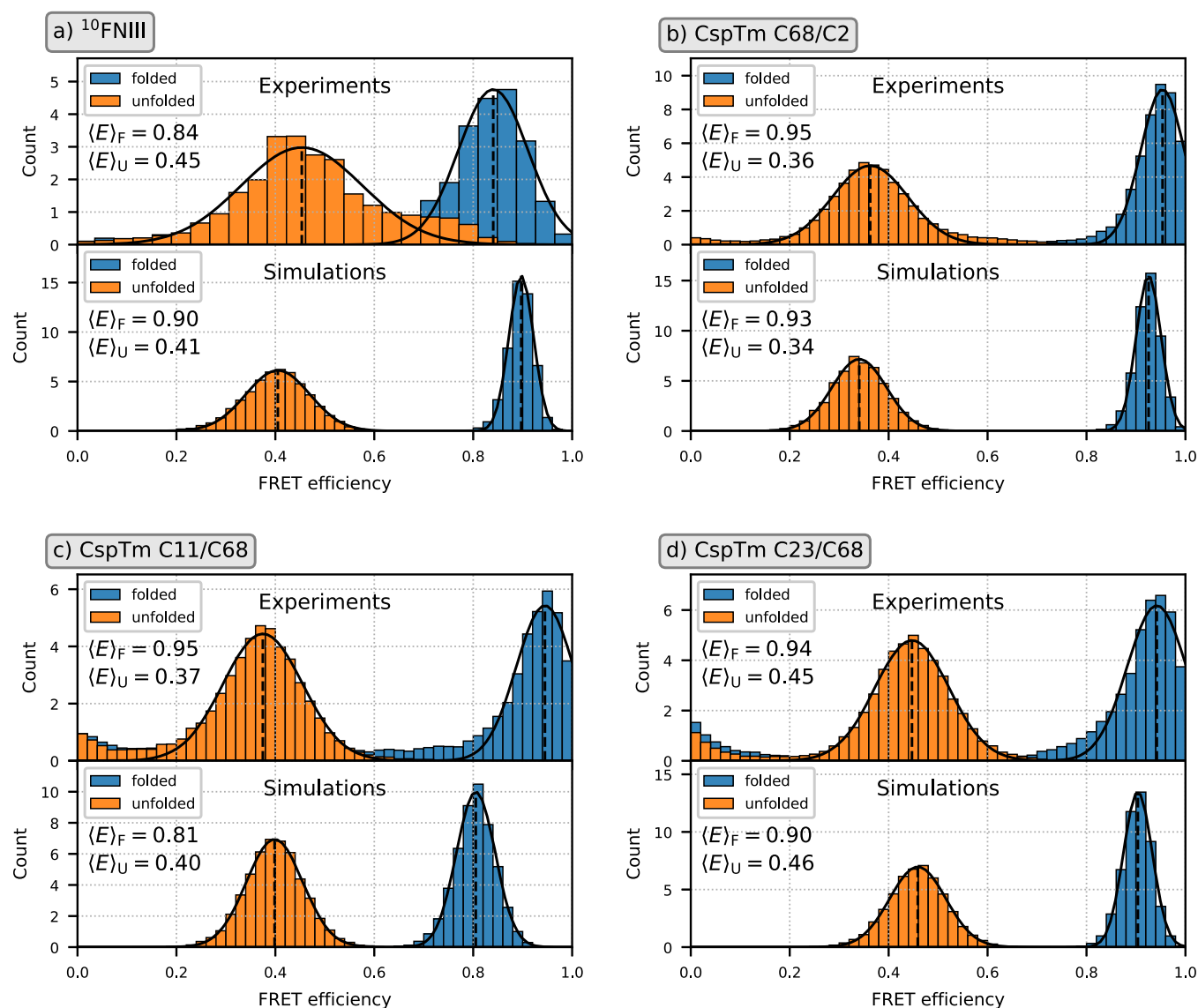


FIG. 4. Comparison of FRET efficiency histograms from simulations and experiments. (a) Results for ¹⁰FNIII with AF546 and AF647 at residues 11 and 86. [(b)–(d)] Results for CspTm with AF488 and AF594 at (b) residues 68 and 2, (c) residues 11 and 68, and (d) residues 23 and 68, respectively.⁶³ The distributions are shown for the folded (blue) and unfolded (orange) states. The experimental FRET efficiency values below 0.0 and above 1.0 are not shown. The distributions are fitted with Gaussians, and the peak positions for folded ($\langle E \rangle_F$) and unfolded ($\langle E \rangle_U$) states together with the respective standard deviations σ_F and σ_U can be found in Table II.

the results for $^{10}\text{FNIII}$ with AF546 and AF647, and Figs. 4(b)–4(d) show the results for CspTm with AF488 and AF594 at different labeling positions.

The experimental data for the respective protein in 0.0 M GdmCl (folded) and in 4.63 M and 7.0 M GdmCl (unfolded) for $^{10}\text{FNIII}$ and CspTm⁶³ are shown, respectively. The experimental data are already corrected for background, different quantum yields of donors and acceptors, different detection efficiencies, cross talk, and direct acceptor excitation. From these errors, only the different quantum yields are reflected in our simulation protocol and are already corrected for in the Monte Carlo photon simulations.

The simulated data agree well with the experimental results except in the case of CspTm C11/C68 [see Fig. 4(c)], where the transfer efficiency of the folded state in the simulation is shifted to lower values. In this case, the dyes are attached on opposite sides of the protein, which would explain a lower FRET efficiency compared to the other two labeling schemes. The deviation between simulations and experiments may be caused by residual attractive interactions between the fluorophores and protein surface, an aspect that could be tested by detailed time-resolved fluorescence anisotropy measurements of the different variants.

The mean and standard deviations of the Gaussian fits in Fig. 4 are summarized in Table II. The widths of the FRET efficiency distributions from the simulations are dominated by shot-noise, caused by the limited number of photons collected for each burst.

In experiments, dye photophysics, e.g., donor and acceptor quenching or blinking, additional deviations through the necessity to correct for background, cross talk, and detection efficiencies, lack of site-specific labeling or other chemical heterogeneity, and other experimental artifacts can lead to additional broadening of the FRET efficiency distribution.^{25,64} Incomplete labeling or photobleaching can lead to a donor-only peak near zero FRET efficiency which is also not present in the simulations.

For CspTm, the width of the Gaussian from the simulations is only slightly lower than the ones from the experiment, which indicates that the width of the experimental histograms is already close to the shot-noise limit.

TABLE II. Parameters of the Gaussian fits of the FRET efficiency histograms from simulations (Sim) and experiments (Expt.). The peak positions of the Gaussian fit for the folded $\langle E \rangle_{\text{F}}$ and unfolded $\langle E \rangle_{\text{U}}$ states are given with the respective standard deviations (σ_{F} , σ_{U}). (a) $^{10}\text{FNIII}$ with AF546 and AF647 at residues 11 and 86. [(b)–(d)] CspTm with AF488 and AF594 at (b) residues 68 and 2, (c) residues 11 and 68, and (d) residues 23 and 68.

| System | | Folded | | Unfolded | |
|-------------------------|-------|--------------------------------|---------------------|--------------------------------|---------------------|
| | | $\langle E \rangle_{\text{F}}$ | σ_{F} | $\langle E \rangle_{\text{U}}$ | σ_{U} |
| (a) $^{10}\text{FNIII}$ | Expt. | 0.84 | 0.07 | 0.45 | 0.12 |
| | Sim | 0.90 | 0.03 | 0.41 | 0.06 |
| (b) CspTm C68/C2 | Expt. | 0.95 | 0.04 | 0.36 | 0.08 |
| | Sim | 0.93 | 0.03 | 0.34 | 0.06 |
| (c) CspTm C11/C68 | Expt. | 0.95 | 0.06 | 0.37 | 0.08 |
| | Sim | 0.81 | 0.04 | 0.40 | 0.06 |
| (d) CspTm C23/C68 | Expt. | 0.94 | 0.06 | 0.45 | 0.07 |
| | Sim | 0.90 | 0.03 | 0.46 | 0.06 |

D. Comparison to simple models for data analysis

To compare the simulation results with different simple models for data analysis, we plot the inter-dye distance R_{DA} for the folded state and R_{DA} in relation to the C_{α} -distance (the distance between the C_{α} -atoms of the respective residues the dyes are attached to) in the unfolded state in Fig. 5. This is shown for three different protein-dye systems.

For the folded state, we compare the mean and standard deviation from the distance distribution of the simulation (black crosses and error bars) to the respective values of an accessible volume calculation^{11,12} (blue). The accessible volume approach calculates all sterically possible dye positions within the linkage length and considers them as equally probable. Then, it calculates the mean distance of the dyes R_{DA} .

Both methods agree very well, with the simulations having a lower standard deviation. This originates from the dynamics in the simulations, which entropically disfavors dye states close to the protein surface, whereas the accessible volume method assumes all states to be equally probable.

For the unfolded state, the accessible volume approach faces the challenge of properly treating an unfolded ensemble of diverse conformations with dye distributions. By contrast, our model can directly simulate the whole ensemble with dyes.

Proteins in the unfolded state can be described as polymer chains. At high temperatures in the unfolded state, the SBM results in an excluded volume polymer chain.⁶⁵ To test this, we consider the equation

$$\langle r^2 \rangle^{1/2} = C \times N^{\nu} \quad (8)$$

with the spatial distance r between two chain elements, the sequence separation between two elements N , the length scaling exponent ν , and a constant C . We fit the mean C_{α} -distance $\langle (R_{C_{\alpha}}^2)_{ij} \rangle^{1/2}$ between residues i and j , averaged over the trajectory, as a function of the sequence separation $N = |i - j|$ with Eq. (8). We obtain values for the fit parameters ν which correspond well with the expected value for an excluded volume chain of $\nu = 3/5$ ⁶⁵ (see SI 7 of the [supplementary material](#)). Now we determine an effective segment length $N_{\text{eff}} = N + L$ with a “length” L of the dye pair which was already used to interpret experimental FRET measurements.^{66,67} We take the dyes’ mean square separation $\langle R_{\text{DA}}^2 \rangle^{1/2}$ and assign an N_{eff} from the respective fit. An example fit and details of the calculation are given in the [supplementary material](#) (see Fig. S12). It turns out that for the AF488/AF594 dye pair, the length is about $L = 11.3 \pm 1.3$ residues, which is in the same range as found experimentally in Ref. 67. For the AF546/AF647 dye pair, it is 14.15 ± 0.65 residues. With this effective length, a correction factor $m = ((N + L)/N)^{\nu}$ for the relation between $\langle R_{\text{DA}}^2 \rangle^{1/2}$ and the distance of the respective C_{α} -atoms $\langle R_{C_{\alpha}}^2 \rangle^{1/2}$ can be calculated. It is given by $m = 1.10$ for CspTm (with AF488/AF594), $m = 1.14$ for CI-2 (with AF546/AF647), and $m = 1.11$ for $^{10}\text{FNIII}$ (with AF546/AF647) (for details see SI 7 of the [supplementary material](#)).

For the unfolded state, the histogram for the inter-dye distance and the corresponding C_{α} -distance is shown in Fig. 5. It can be seen clearly that the values are not identical (which would be indicated by the dotted lines). By correcting this for the effective chain length, we expect a different relation

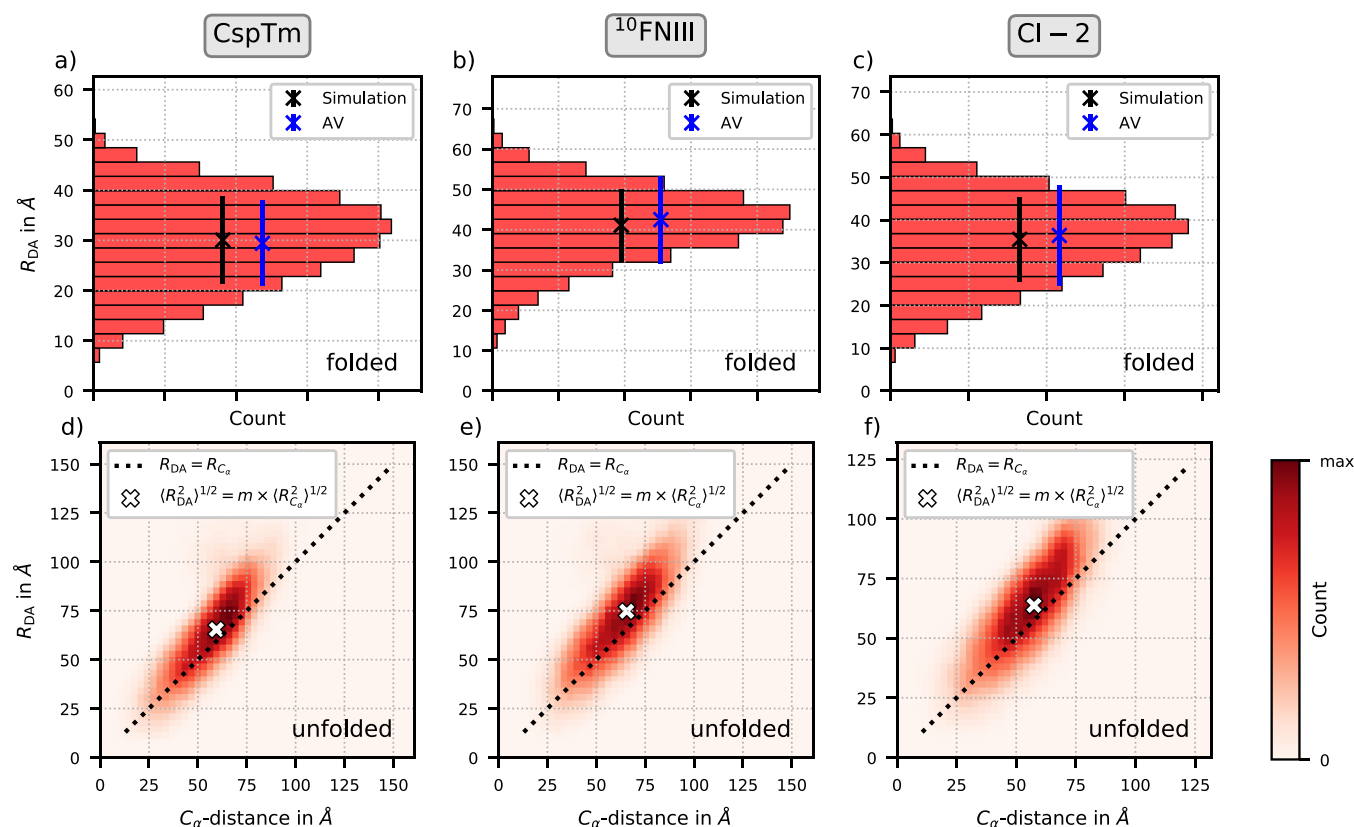


FIG. 5. Distributions of inter-dye distances R_{DA} and distances of the C_α -atoms of the respective residues for CspTm (with AF488 and AF594 at residues 2 and 68, left), ¹⁰FNIII (with AF546 and AF647 at residues 11 and 86, middle), and CI-2 (with AF546 and AF647 at residues 20 and 78, right). For the folded state [(a)–(c)], distance distributions (red), mean distances (crosses), and standard deviation (error bars) from simulations (black) and from accessible volume calculations (blue) are depicted. For the unfolded state [(d)–(f)], the histograms of the inter-dye distances and the corresponding C_α -distances are shown. The expected dependency for equality of both values (dotted lines) and for the expectation corrected by an effective segment length of the chain (white crosses) is shown. The histogram count is scaled according to the maximum of each histogram separately.

between R_{DA} and the C_α -distance, as indicated by the white crosses, providing a good approximation of the dependence observed in the simulations.

In summary, we achieve our aim of capturing both folded and unfolded states in the same simulation method, paving the way to simulate complex large-scale conformational transitions between multiple states in the future.

V. DISCUSSION

In this work, we introduce a new simulation method for dye-labeled biomolecules with only a few robust parameters. The approach suggested here is most useful in the absence of pronounced non-native interactions or structure formation not included in the structure-based model. We are able to achieve a realistic description of the system as evidenced by direct comparison and quantitative agreement of FRET efficiency histograms with experimental data. This establishes a way to directly complement experimental FRET efficiency distributions with atomically resolved structural ensembles from simulations, which provides novel and detailed insights into biomolecular processes. Even complex scenarios involving large structural ensembles, such as unfolded proteins, intrinsically disordered proteins, conformational transitions, and folding intermediates, can be sufficiently sampled with modest computational resources.

The model can improve planning experimental measurements as one can vary and test different parameters such as the Förster radius, linker length, and labeling sites *in silico*. On the basis of these simulations, one can then choose the dye combinations which would, e.g., best distinguish the conformational states of interest. Further, we found that the approximation of $\overline{\kappa^2} = 2/3$ is mostly appropriate for the tested typical systems. Similarly, we could probe estimations of dye distance distributions based on simple models.

Having established the simulation techniques, many new scenarios are now accessible. These include simulating strongly restricted or slow dye motions and very short or inflexible linkers to test their effect on κ^2 . The low computational costs make investigating conformational transitions between multiple stable conformations and other large scale motions accessible. In addition, it can be used to simulate more complex FRET measurements, such as three-color FRET,^{68–70} where the interpretation of the data is even more challenging.

SUPPLEMENTARY MATERIAL

See [supplementary material](#) for additional information about the used methods and the three-dimensional structures of the Alexa Fluor dyes.

ACKNOWLEDGMENTS

I.R., C.S., and A.S. received funding from the Impuls- und Vernetzungsfond of the Helmholtz association. We would like to thank Pascal Friederich for fruitful discussions. G.U.N. acknowledges funding by the Karlsruhe Institute of Technology in the context of the Helmholtz Program Science and Technology of Nanosystems (STN). B.S. was supported by the Swiss National Science Foundation.

- ¹B. Schuler and H. Hofmann, *Curr. Opin. Struct. Biol.* **23**, 36 (2013).
- ²R. Rieger, A. Kobitski, H. Sielaff, and G. U. Nienhaus, *ChemPhysChem* **12**, 627 (2011).
- ³R. Rieger and G. U. Nienhaus, *Chem. Phys.* **396**, 3 (2012).
- ⁴S. Muller-Spath, A. Soranno, V. Hirschfeld, H. Hofmann, S. Ruegger, L. Reymond, D. Nettels, and B. Schuler, *Proc. Natl. Acad. Sci. U. S. A.* **107**, 14609 (2010).
- ⁵D. Nettels, S. Muller-Spath, F. Kuster, H. Hofmann, D. Haenni, S. Ruegger, L. Reymond, A. Hoffmann, J. Kubelka, B. Heinz, K. Gast, R. B. Best, and B. Schuler, *Proc. Natl. Acad. Sci. U. S. A.* **106**, 20740 (2009).
- ⁶E. Sherman and G. Haran, *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11539 (2006).
- ⁷Y. Gambin, A. Schug, E. A. Lemke, J. J. Lavinder, A. C. M. Ferreón, T. J. Magliery, J. N. Onuchic, and A. A. Deniz, *Proc. Natl. Acad. Sci. U. S. A.* **106**, 10153 (2009).
- ⁸Z. Wang and D. E. Makarov, *J. Phys. Chem. B* **107**, 5617 (2003).
- ⁹A. Muschiolok, J. Andrecka, A. Jawhari, F. Brückner, P. Cramer, and J. Michaelis, *Nat. Methods* **5**, 965 (2008).
- ¹⁰A. Muschiolok and J. Michaelis, *J. Phys. Chem. B* **115**, 11927 (2011).
- ¹¹S. Sindbert, S. Kalinin, H. Nguyen, A. Kienzler, L. Clima, W. Bannwarth, B. Appel, S. Müller, and C. A. M. Seidel, *J. Am. Chem. Soc.* **133**, 2463 (2011).
- ¹²S. Kalinin, T. Peulen, S. Sindbert, P. J. Rothwell, S. Berger, T. Restle, R. S. Goody, H. Gohlke, and C. A. M. Seidel, *Nat. Methods* **9**, 1218 (2012).
- ¹³B. Schuler, A. Soranno, H. Hofmann, and D. Nettels, *Annu. Rev. Biophys.* **45**, 207 (2016).
- ¹⁴R. B. Best, H. Hofmann, D. Nettels, and B. Schuler, *Biophys. J.* **108**, 2721 (2015).
- ¹⁵G. F. Schröder, U. Alexiev, and H. Grubmüller, *Biophys. J.* **89**, 3757 (2005).
- ¹⁶R. B. Best, K. A. Merchant, I. V. Gopich, B. Schuler, A. Bax, and W. A. Eaton, *Proc. Natl. Acad. Sci. U. S. A.* **104**, 18964 (2007).
- ¹⁷M. J. Shoura, R. U. Ranatunga, S. A. Harris, S. O. Nielsen, and S. D. Levene, *Biophys. J.* **107**, 700 (2014).
- ¹⁸M. M. Reif and C. Oostenbrink, *J. Comput. Chem.* **35**, 2319 (2014).
- ¹⁹E. V. Kuzmenkina, C. D. Heyes, and G. U. Nienhaus, *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15471 (2005).
- ²⁰D. B. VanBeek, M. C. Zwier, J. M. Shorb, and B. P. Krueger, *Biophys. J.* **92**, 4168 (2007).
- ²¹B. Corry and D. Jayatilaka, *Biophys. J.* **95**, 2711 (2008).
- ²²P. C. Whitford, J. K. Noel, S. Gosavi, A. Schug, K. Y. Sanbonmatsu, and J. N. Onuchic, *Proteins: Struct., Funct., Bioinf.* **75**, 430 (2009).
- ²³T. Förster, *Ann. Phys.* **437**, 55 (1948).
- ²⁴L. Stryer, *Annu. Rev. Biochem.* **47**, 819 (1978).
- ²⁵E. Sisamakias, A. Valeri, S. Kalinin, P. J. Rothwell, and C. A. Seidel, in *Single Molecule Tools, Part B: Super-Resolution, Particle Tracking, Multi-parameter, and Force Based Methods*, Methods in Enzymology, edited by N. G. Walter (Academic Press, 2010), Vol. 475, pp. 455–514.
- ²⁶G. Lipari and A. Szabo, *Biophys. J.* **30**, 489 (1980).
- ²⁷F. Hillger, D. Hänni, D. Nettels, S. Geister, M. Grandin, M. Textor, and B. Schuler, *Angew. Chem., Int. Ed.* **47**, 6184 (2008).
- ²⁸D. Nettels, A. Hoffmann, and B. Schuler, *J. Phys. Chem. B* **112**, 6137 (2008).
- ²⁹J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, *Proteins: Struct., Funct., Genet.* **21**, 167 (1995).
- ³⁰J. D. Bryngelson and P. G. Wolynes, *Proc. Natl. Acad. Sci. U. S. A.* **84**, 7524 (1987).
- ³¹P. E. Leopold, M. Montal, and J. N. Onuchic, *Proc. Natl. Acad. Sci. U. S. A.* **89**, 8721 (1992).
- ³²H. Frauenfelder, S. Sligar, and P. Wolynes, *Science* **254**, 1598 (1991).
- ³³J. N. Onuchic and P. G. Wolynes, *Curr. Opin. Struct. Biol.* **14**, 70 (2004).
- ³⁴C. Sinner, B. Lutz, S. John, I. Reinartz, A. Verma, and A. Schug, *Isr. J. Chem.* **54**, 1165 (2014).
- ³⁵L. L. Chavez, J. N. Onuchic, and C. Clementi, *J. Am. Chem. Soc.* **126**, 8426 (2004).
- ³⁶C. Sinner, B. Lutz, A. Verma, and A. Schug, *J. Chem. Phys.* **143**, 243154 (2015).
- ³⁷M. B. Borgia, A. Borgia, R. B. Best, A. Steward, D. Nettels, B. Wunderlich, B. Schuler, and J. Clarke, *Nature* **474**, 662 (2011).
- ³⁸P. Tian and R. B. Best, *PLoS Comput. Biol.* **12**, e1004933 (2016).
- ³⁹A. Schug, M. Weigt, J. N. Onuchic, T. Hwa, and H. Szurmant, *Proc. Natl. Acad. Sci. U. S. A.* **106**, 22124 (2009).
- ⁴⁰P. C. Whitford, O. Miyashita, Y. Levy, and J. N. Onuchic, *J. Mol. Biol.* **366**, 1661 (2007).
- ⁴¹A. Schug, P. C. Whitford, Y. Levy, and J. N. Onuchic, *Proc. Natl. Acad. Sci. U. S. A.* **104**, 17674 (2007).
- ⁴²P. C. Whitford and K. Y. Sanbonmatsu, *J. Chem. Phys.* **139**, 121919 (2013).
- ⁴³P. C. Whitford, A. Schug, J. Saunders, S. P. Hennelly, J. N. Onuchic, and K. Y. Sanbonmatsu, *Biophys. J.* **96**, L7 (2009).
- ⁴⁴C. Clementi, H. Nymeyer, and J. N. Onuchic, *J. Mol. Biol.* **298**, 937 (2000).
- ⁴⁵J. K. Noel, A. Schug, A. Verma, W. Wenzel, A. E. Garcia, and J. N. Onuchic, *J. Phys. Chem. B* **116**, 6880 (2012).
- ⁴⁶B. Lutz, C. Sinner, G. Heuermann, A. Verma, and A. Schug, *Bioinformatics* **29**, 2795 (2013).
- ⁴⁷J. K. Noel, P. C. Whitford, and J. N. Onuchic, *J. Phys. Chem. B* **116**, 8692 (2012).
- ⁴⁸J. K. Noel, P. C. Whitford, K. Y. Sanbonmatsu, and J. N. Onuchic, *Nucleic Acids Res.* **38**, W657 (2010).
- ⁴⁹H. Lammert, A. Schug, and J. N. Onuchic, *Proteins: Struct., Funct., Bioinf.* **77**, 881 (2009).
- ⁵⁰M. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and GRO-MACS Development Team, GROMACS User Manual version 5.0, URL: <http://www.gromacs.org> (2014), pp. 78–79, 84.
- ⁵¹H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, *Nucleic Acids Res.* **28**, 235 (2000).
- ⁵²C. A. McPhalen, I. Svendsen, I. Jonassen, and M. N. James, *Proc. Natl. Acad. Sci. U. S. A.* **82**, 7242 (1985).
- ⁵³A. L. Main, T. S. Harvey, M. Baron, J. Boyd, and I. D. Campbell, *Cell* **71**, 671 (1992).
- ⁵⁴W. Kremer, B. Schuler, S. Harrieder, M. Geyer, W. Gronwald, C. Welker, R. Jaenicke, and H. R. Kalbitzer, *Eur. J. Biochem.* **268**, 2527 (2001).
- ⁵⁵See URL: <http://www.thermofisher.com> for information about the Alexa Fluor dyes.
- ⁵⁶D. van der Spoel, E. Lindahl, B. Hess, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. A. Feenstra, R. van Drunen, and H. J. C. Berendsen, Gromacs User Manual version 4.5.6, URL: <http://www.gromacs.org> (2010).
- ⁵⁷J. Wang, P. Cieplak, and P. A. Kollman, *J. Comput. Chem.* **21**, 1049 (2000).
- ⁵⁸B. Lutz, M. Faber, A. Verma, S. Klumpp, and A. Schug, *Nucleic Acids Res.* **42**, 2687 (2014).
- ⁵⁹J. Jackson, K. Nguyen, and P. Whitford, *Int. J. Mol. Sci.* **16**, 6868 (2015).
- ⁶⁰M. Hoeffling, N. Lima, D. Haenni, C. A. M. Seidel, B. Schuler, and H. Grubmüller, *PLoS One* **6**, e19791 (2011).
- ⁶¹M. Hoeffling and H. Grubmüller, *Comput. Phys. Commun.* **184**, 841 (2013).
- ⁶²A. Hoffmann, A. Kane, D. Nettels, D. E. Hertzog, P. Baumgartel, J. Lengefeld, G. Reichardt, D. A. Horsley, R. Seckler, O. Bakajin, and B. Schuler, *Proc. Natl. Acad. Sci. U. S. A.* **104**, 105 (2007).
- ⁶³A. Soranno, B. Buchli, D. Nettels, R. R. Cheng, S. Muller-Spath, S. H. Pfeil, A. Hoffmann, E. A. Lipman, D. E. Makarov, and B. Schuler, *Proc. Natl. Acad. Sci. U. S. A.* **109**, 17800 (2012).
- ⁶⁴B. Schuler, “Application of single molecule Förster resonance energy transfer to protein folding,” in *Protein Folding Protocols*, edited by Y. Bai and R. Nussinov (Humana Press, Totowa, NJ, 2006), pp. 115–138.
- ⁶⁵P. J. Flory, *J. Chem. Phys.* **17**, 303 (1949).
- ⁶⁶H. Hofmann, A. Soranno, A. Borgia, K. Gast, D. Nettels, and B. Schuler, *Proc. Natl. Acad. Sci. U. S. A.* **109**, 16155 (2012).
- ⁶⁷M. Aznauryan, L. Delgado, A. Soranno, D. Nettels, J.-r. Huang, A. M. Labhardt, S. Grzesiek, and B. Schuler, *Proc. Natl. Acad. Sci. U. S. A.* **113**, E5389 (2016).
- ⁶⁸S. Hohng, C. Joo, and T. Ha, *Biophys. J.* **87**, 1328 (2004).
- ⁶⁹S. Lee, J. Lee, and S. Hohng, *PLoS One* **5**, e12270 (2010).
- ⁷⁰S. Milles, C. Koehler, Y. Gambin, A. A. Deniz, and E. A. Lemke, *Mol. Biosyst.* **8**, 2531 (2012).