

Resolving distance variations by single-molecule FRET and EPR spectroscopy using rotamer libraries

Daniel Klose,^{1,*} Andrea Holla,² Christoph Gmeiner,¹ Daniel Nettels,² Irina Ritsch,¹ Nadja Bross,⁵ Maxim Yulikov,¹ Frédéric H.-T. Allain,³ Benjamin Schuler,^{2,4,*} and Gunnar Jeschke¹

¹Department of Chemistry and Applied Biosciences, ETH Zurich, Zurich, Switzerland; ²Department of Biochemistry, University of Zurich, Zurich, Switzerland; ³Institute of Biochemistry, ETH Zurich, Zurich, Switzerland; ⁴Department of Physics and ⁵Department of Chemistry, University of Zurich, Zurich, Switzerland

ABSTRACT Förster resonance energy transfer (FRET) and electron paramagnetic resonance (EPR) spectroscopy are complementary techniques for quantifying distances in the nanometer range. Both approaches are commonly employed for probing the conformations and conformational changes of biological macromolecules based on site-directed fluorescent or paramagnetic labeling. FRET can be applied in solution at ambient temperature and thus provides direct access to dynamics, especially if used at the single-molecule level, whereas EPR requires immobilization or work at cryogenic temperatures but provides data that can be more reliably used to extract distance distributions. However, a combined analysis of the complementary data from the two techniques has been complicated by the lack of a common modeling framework. Here, we demonstrate a systematic analysis approach based on rotamer libraries for both FRET and EPR labels to predict distance distributions between two labels from a structural model. Dynamics of the fluorophores within these distance distributions are taken into account by diffusional averaging, which improves the agreement with experiment. Benchmarking this methodology with a series of surface-exposed pairs of sites in a structured protein domain reveals that the lowest resolved distance differences can be as small as ~ 0.25 nm for both techniques, with quantitative agreement between experimental and simulated transfer efficiencies within a range of ± 0.045 . Rotamer library analysis thus establishes a coherent way of treating experimental data from EPR and FRET and provides a basis for integrative structural modeling, including studies of conformational distributions and dynamics of biological macromolecules using both techniques.

SIGNIFICANCE Combining data from different experimental techniques is often essential for taking advantage of the complementary information they can provide. An area in which this approach has been particularly fruitful is the integrative structural modeling of biological macromolecules, their conformational changes, and their assemblies. Förster resonance energy transfer (FRET) and electron paramagnetic resonance (EPR) spectroscopy are two powerful biophysical techniques that have long been used for this purpose, but integrating them has been complicated by the lack of a suitable analysis framework. Here, we establish such a framework based on rotamer libraries of the labels used in FRET and EPR, and we apply it to a challenging experimental benchmark.

INTRODUCTION

Revealing the detailed functional mechanisms of proteins, nucleic acids, and the biomolecular complexes that they form requires not only information on static structures but also on their conformational distributions and transitions, as well as structural dynamics. Prominent examples include

the nuclear pore complex (1,2), protein-RNA complexes involved in splicing regulation (3,4), and G-protein-coupled receptor complexes (5,6). The intrinsic flexibility or structural heterogeneity of these systems is commonly described by ensembles of structures, in which each member contributes with a certain probability. This ensemble forms the basis for a description of the mechanism of action (5,7).

Arriving at such a detailed and complex description is, however, challenging and may require a combination of different techniques. For the structural characterization of a single well-structured state, classical structure determination methods such as X-ray crystallography, NMR spectroscopy, and single-particle cryo-electron microscopy are

Submitted February 26, 2021, and accepted for publication September 13, 2021.

*Correspondence: daniel.klose@phys.chem.ethz.ch or schuler@bioc.uzh.ch

Editor: Samrat Mukhopadhyay.

<https://doi.org/10.1016/j.bpj.2021.09.021>

© 2021 Biophysical Society.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).



well established for determining biomolecular structures with atomic resolution (3,8,9). These structures provide the starting point for characterizing transitions to other functional states or conformational heterogeneity and dynamics of mobile elements, such as flexible linkers in multidomain proteins or disordered, functionally relevant regions. In these cases, long-range information on distances and dynamics is particularly important. Long-range distance information can be accessed, for instance, by small-angle scattering, either of X-rays or neutrons (10). These methods offer the advantage—like NMR—that the system is investigated in solution and thus under near-physiological conditions. Although inferring structure from scattering curves is a severely ill-posed problem, small-angle scattering is a valuable tool for providing low-resolution shape information, which has been successfully combined in integrative structure modeling approaches with data from other techniques such as X-ray crystallography, NMR, EPR, FRET, or cryo-electron microscopy (3,4,10–19).

Another important class of techniques allows for measuring site-selective distances in the low-nanometer range and for characterizing dynamics: electron paramagnetic resonance (EPR) and fluorescence spectroscopy via Förster resonance energy transfer (FRET). Both techniques rely on the site-directed covalent attachment of labels, namely a pair of paramagnetic spin labels or a pair of fluorophores that make the investigated systems EPR or FRET active, respectively. Although FRET is performed in solution at ambient temperature, at which information on dynamics is directly accessible, pulse EPR spectroscopy typically requires low temperatures in the range of 10–50 K, depending on the type of spin label, to reduce longitudinal and transverse relaxation of the electron spins (20–25).

The distance information accessible from freeze-trapping of different states and the application of pulse EPR spectroscopy is in the range of 1.5–16 nm (20,26,27). Importantly, besides average distances, distance distributions can be obtained that provide information on the width of the structural ensemble of the molecular system (28). The distance distributions between spin labels can be determined by pulse dipolar spectroscopy (PDS), a term that summarizes different methods such as Double Electron-Electron Resonance (DEER) (29,30), RIDME (31), DQC (32), and SIFTER (33). All these techniques measure the distribution of electron dipolar spin-spin coupling, which has an r^{-3} distance dependence. The transformation of the time-domain data from PDS experiments to distance distributions is moderately ill posed. Data analysis procedures such as Tikhonov regularization with the L-curve criterion are well established (20,26) and enable the detection of distance distributions that is direct in the sense that the forward calculation of time-domain data from distance distributions is calibration-free because it only depends on fundamental constants. Orientation selection with respect to the dipolar coupling can usually be neglected for solvent-exposed nitro-

xide spin labels because of their wide conformational distributions. Alternatively, when the orientation between two labels is correlated, the angular information on the relative orientations can be extracted in addition to the distance distribution, albeit with substantial additional effort (20,34,35).

Complementary to EPR, fluorescence spectroscopy offers the advantage of being performed near or at physiological temperatures, at which dynamics can be probed; moreover, it reaches single-molecule sensitivity (36–38). Distances are most commonly measured by FRET, the nonradiative energy transfer from a donor to an acceptor fluorophore. The rate of energy transfer depends on the coupling between the transition dipole moments of both labels and hence has a distance dependence of r^{-6} (39). To minimize the uncertainty from the orientation dependence of the dipole coupling, fluorescence labels typically have linkers that are flexible enough to ensure rapid rotational averaging. Single-molecule fluorescence experiments can be used to probe nanometer distances and conformational dynamics over a wide range of timescales (37,40,41). However, the conformational dynamics are not necessarily separated in time from other contributions such as rotational relaxation, fluorescence lifetimes, and other photophysics, which complicates the analysis in terms of the underlying distance distributions (42–45).

If long-range distance constraints between fluorescence or spin labels are to be used for structure modeling, it can be essential to explicitly account for the labels in modeling. For spin labels used in pulse EPR spectroscopy, this requirement has led to the development of different molecular modeling approaches. These are either atomistic molecular dynamics (MD) simulations of spin-labeled proteins (23,46–48) or, alternatively, less computationally demanding molecular models that explicitly take the linker structure and energetics into account. Making use of explicit chemical structure rather than an accessible volume-based approach allows these molecular models to predict the often highly anisotropic conformational distributions of protein-bound nitroxide spin labels. These molecular models, namely rotamer libraries (49,50), MtsslWizzard (51), and ALLNOX (52), have become a standard analysis technique for comparing EPR-derived distance distributions to biomolecular structures (28). Rotamer libraries are a set of pre-calculated conformers (or rotamers) for each label that are representative of the labels' conformational space. For any given labeling site in a biomolecule, screening of the libraries allows us to find a set of rotamers that avoids clashes with the protein along with the respective populations for each rotamer that identify energetically preferred states within the labels' conformational space. This approach has not only been used on single structures but also allows for the rapid analysis of series of structures along normal modes (53,54) or of MD trajectories (55). Simulated distance distributions between two sets of spin-label conformations show an average accuracy of 2.5 Å with respect to

experimental distance distributions obtained by DEER (51,56). For integrative structure modeling, the spin-label rotamer libraries have been used successfully together with NMR and crystallographic constraints for modeling flexible protein termini or extended loops and relative positions of flexibly linked domains (57).

For fluorescence labels, a well-established and efficient modeling approach is to calculate the sterically accessible volume (AV) of the dyes (58). The labeled side chain is described by a small number of parameters, such as linker length and width and chromophore radii. Interdye distance distributions or average distances are then calculated from the discretized dye positions within their respective AVs, in which all volume elements are considered with equal weights (59,60). This approach is computationally inexpensive and has been applied in many studies on biomolecules (37,61,62). Rotamer libraries tailored to fluorescence labels have been developed and applied for proteins and nucleic acids (14,63–65). More detailed molecular models that rely on MD simulations (66–70) or Monte Carlo sampling (23,71) are computationally more demanding; hence, they lack the efficiency required for analyzing large ensembles of structures and may require a system-specific optimization of force field parameters, especially regarding fluorophore-protein interactions (72).

Using both FRET and EPR to observe biomolecules under different experimental conditions and with or without ensemble averaging yields complementary data on structure and dynamics (11,25). However, directly combining the two methods in integrative structural modeling has been complicated by the lack of a common framework for treating the labels. To overcome this issue, we adopt the approach commonly used in EPR spectroscopy and establish rotamer libraries for FRET, with an optimized force field as a computationally efficient approach for obtaining distance distributions between two protein-linked fluorophores given a model of the structure. For the subsequent calculation of average FRET efficiencies, we account for the label dynamics by treating them in terms of diffusion in a potential of mean force obtained from the distance distributions. To test the methodology, we evaluate the accuracy with which small distance variations can be determined by the two label-based techniques FRET and EPR. For this purpose, we used six double-cysteine variants of a rigid and well-defined protein system consisting of two RNA recognition motifs (RRMs) of the human polypyrimidine tract binding protein 1 (PTBP1) (73–75). The labeling positions in the two domains (RRM3/4) are shifted by three to seven residues (one or two α -helix turns). The single-molecule FRET efficiencies resolved all but one of the distance variations involving a single helix turn. We find that taking into account the motion of the fluorophores improves the accuracy of the absolute distances recovered. For EPR spectroscopy, the resulting distance distributions between spin labels were well suited for revealing both the absolute distances and the

relative distance shifts by one or two helix turns for five out of six variants. Hence, both techniques can detect distance variations with a precision corresponding to one α -helix turn. The common framework of rotamer library simulations establishes a way of combining experimental data from FRET and EPR spectroscopies in a unified approach suitable for integrative modeling.

MATERIALS AND METHODS

All chemicals were purchased from Sigma-Aldrich (Buchs, Switzerland), unless stated otherwise.

Labeling sites and protein expression

The two-domain construct RRM3/4 of PTBP1 represents a well-structured and conformationally rigid system (74). For distance measurements between spin labels or fluorescence labels, we selected positions separated by distances in a range of 1.5–8.0 nm. As described earlier, the site pair Q388C-S475C delivered narrow distance distributions when investigated with a set of different spin labels (75). Based on this site pair, we selected solvent-exposed α -helical sites separated by four to seven amino acids from Q388C and S475C, respectively, which translates to one or two helix turns. In total, five positions were chosen for mutagenesis, two on RRM3 (Q388C and S392C) and three on RRM4 (E468C, V472C, and S475C). The insertion of cysteine residues at the respective positions was achieved by site-directed mutagenesis, and the constructs of RRM3/4 bearing double-cysteine mutations were expressed and purified as previously described (75).

Spin labeling

The different mutated and purified (75) proteins were spin-labeled with a 10-fold excess of 3-maleimido proxyl (MAP) at ambient temperature. The labeling reaction was performed overnight in labeling buffer (50 mM 3-(N-morpholino)propanesulfonic acid (MOPS), 25 mM NaCl (pH 6.5)) under gentle shaking. Residual unbound spin label was removed by washing the samples with a low-salt buffer (10 mM NaPO₄, 20 mM NaCl (pH 6.5)) via PD10 desalting columns (GE Healthcare, Glattbrugg, Switzerland). Eluted proteins were concentrated to a final volume of ~200 μ L with a concentration of 100 μ M. Sample quality was checked by SDS-PAGE under nonreducing conditions and labeling efficiencies were determined by continuous wave (cw) EPR spectroscopy.

Fluorescence labeling

For site-specific labeling with Cy3b (GE Healthcare, Little Chalfont, UK) and CF660R (Sigma-Aldrich, St. Louis, MO), the double-cysteine constructs of RRM3/4 were reduced with 10 mM dithiothreitol (DTT) and further purified by cation exchange chromatography using a 1-mL MonoS column (GE Healthcare; 20 mM KPO₃/PO₄, 10% glycerol (pH 7.2), gradient: 0–500 mM KCl). The freshly reduced protein was incubated with CF660R maleimide overnight at 4°C at a molar ratio of dye:protein of 0.7:1 (76). Cation exchange chromatography was used to enrich unlabeled and single- and double-labeled species after quenching the reaction with 10 mM DTT. In most cases, these labeling reaction products eluted from the column in four peaks, with some peaks partially overlapping, namely the double-labeled species with one labeling permutant and the other labeling permutant with the unlabeled species. Nevertheless, the labeling permutants with the label attached at either of the cysteines eluted in clearly separated peaks, thus allowing site-specific labeling of RRM3/4. Single-labeled RRM3/4 was subsequently incubated with an excess of Cy3b maleimide overnight at 4°C to label the second cysteine. Free label

was removed by cation exchange chromatography after quenching the reaction with 10 mM DTT. Because attaching Cy3b did not cause a significant shift in elution of the protein from the column, this purification step did not lead to further separation of donor-acceptor-labeled protein from the double-donor- or double-acceptor-labeled species, so that depending on the labeling permutant, the resulting material contains either a small amount of double-donor- or double-acceptor-labeled species in addition to the donor-acceptor-labeled species. In two cases, CF660R maleimide reacted only with one of the cysteine residues in the first labeling step. To produce both permutants of those variants, the labeling procedure had to be reversed. In those cases, Cy3b maleimide was added to the unlabeled protein first, followed by CF660R maleimide in the second labeling step. With this procedure, the dyes were site-specifically attached at the two labeling positions, resulting in 12 donor-acceptor-labeled variants altogether, including two permutants per construct. The attachment positions of the dyes were identified by trypsin digest followed by mass spectrometry.

EPR spectroscopy

For cw EPR experiments, 20 μL of sample with a protein concentration of $\sim 25 \mu\text{M}$ were transferred into glass capillaries with 0.9 mm outer diameter (Blaubrand micropipettes; Brand, Wertheim, Germany). cw EPR measurements were carried out in the X band ($\sim 9.5 \text{ GHz}$) on an EleXsys E500 EPR spectrometer (Bruker Biospin, Rheinstetten, Germany) equipped with a Bruker super-high-Q resonator at room temperature. Spectra were recorded with 100 kHz magnetic field modulation, 0.1 mT modulation amplitude, and a lock-in time constant and conversion time of 10.24 and 40.96 ms, respectively. The power was attenuated by 25 dB of 200 mW incident microwave power. The spin labeling efficiency was determined by digital double integration of the cw EPR spectra using a reference solution of 100 μM 3-(2-iodoacetamido)-proxyl and comparing to the protein concentration.

For pulse EPR experiments, protein samples were first diluted with D_2O to a concentration of $\sim 75 \mu\text{M}$ and then mixed with d_8 -glycerol in a ratio of 1:1. 30 μL sample was added to 3 mm quartz capillaries (Aachener Quartzglas, Aachen, Germany) and shock frozen in liquid nitrogen.

Pulse EPR experiments were carried out at Q-band frequencies ($\sim 34.5 \text{ GHz}$) using a home-built spectrometer (77), equipped with a traveling wave tube amplifier with 200 W nominal microwave power, as well as a home-built TE102 dielectric resonator (78) and a helium flow cryostat (Oxford Instruments, Oxfordshire, UK) to stabilize the temperature to 50 K. DEER data were acquired using the four-pulse DEER sequence (30), $\pi/2 - \tau_1 - \pi - (\tau_1 + t) - \pi_{\text{pump}} - (\tau_2 - t) - \pi - \tau_2 - \text{echo}$, by incrementing the pump pulse delay t in steps of 12 or 16 ns (for $\tau_2 > 4 \mu\text{s}$) with a two-step phase cycle of the first $\pi/2$ pulse to cancel receiver offsets. All pulse lengths were set to 12 ns, and the offset between pump and observer frequencies was set to 100 MHz, with the pump pulse positioned at the maximum of the nitroxide spectrum (78). Nuclear modulation was averaged out by stepping τ_1 in eight steps of 16 ns starting from $\tau_1 = 400 \text{ ns}$ (33,78). The second delay time τ_2 was set between 4 and 7 μs according to the expected distances, and the dead-time delay was 280 ns.

The resulting time traces were analyzed in DEERAnalysis 2016 (79) using a 3D-homogeneous background function. Subsequently, distance distributions were obtained from the model-free Tikhonov regularization (80,81) with the regularization parameter determined according to the L-curve criterion (82).

Single-molecule FRET measurements

For single-molecule experiments, the donor-acceptor-labeled protein variants were diluted to 100 pM in 20 mM Tris, 125 mM KCl (pH 7.4) with 0.001% Tween 20, 10 mM DTT, and 5 nM unlabeled RRM3/4 as additives. The experiments were conducted at 22°C using chambered cover slides (μ -Slide; ibidi, Gräfelfing, Germany) on a custom-built confocal instrument described previously (83), equipped with a supercontinuum source (SuperK

EXTREME EXW-12; NKT Photonics, Birkerød, Denmark) to excite the donor dye and a 635-nm diode laser operated in pulsed mode (LDHD-C-635M; PicoQuant, Berlin, Germany) to excite the acceptor dye. Both lasers were operated at a pulse repetition rate of 20 MHz. The light from the NKT source was filtered with a bandpass filter (BrightLine HC 520/5; Semrock, Rochester, NY). Fluorescence photons were collected through a high-numerical-aperture objective (UPlanApo 60 \times /1.20-W; Olympus, Tokyo, Japan), subsequently separated from the scattered photons with a triple-band mirror (zt405/530/630rpc; Chroma, Bellows Falls, VT), and distributed onto four channels according to their wavelength and polarization. Dichroic mirrors were used to separate donor and acceptor emission (T635LPXR; Chroma). Donor photons were selected with ET585/65m bandpass filters (Chroma) before detection on one of two single-photon avalanche photodiodes (τ -SPAD, PicoQuant). Acceptor photons were selected with LP647RU longpass filters (Chroma) and detected with SPCM-AQRH-14 single-photon avalanche photodiodes (PerkinElmer, Waltham, MA). To remove the contribution of molecules with inactive acceptor, pulsed interleaved excitation was used (84). Time bins of 1 ms containing more than 50 photons (after background correction) emitted upon donor or acceptor excitation were regarded as photon bursts corresponding to a single protein diffusing through the confocal volume. Photon counts were corrected for background, differences in quantum yields of the dyes, different detection efficiencies, spectral cross talk, and direct excitation of the acceptor dye by the donor excitation light. The necessary correction factors were inferred from the measurement of a set of calibration samples (45) labeled with Cy3b and CF660R following the procedure given in (45,85,86). For each burst, the transfer efficiency $E = N_A^d / (N_A^d + N_D^d)$ and the stoichiometry ratio $S = (N_A^d + N_D^d) / (N_A^d + N_D^d + N_A^a)$ were calculated, where N_A^d and N_D^d are the corrected numbers of acceptor and donor photons emitted upon donor excitation and N_A^a the corrected numbers of acceptor photons emitted upon acceptor excitation. The mean values $\langle E \rangle$ and $\langle S \rangle$ of subpopulations were determined via 2D-Gaussian fits to 2D S vs. E histograms (see Fig. S3). Based on the polarization sensitivity of the four-channel detection system, fluorescence anisotropies were quantified for all samples from the fluorescence emission of the donor- and acceptor-only subpopulations. Because the resulting steady-state anisotropies for all samples were below 0.12, we assume for our analysis that the rotational correlation time of the dyes is sufficiently short for a rotationally averaged orientation factor of 2/3 in Förster theory (86,87).

The FRET pair Cy3b/CF660R was chosen in part because of its relatively low sensitivity to changes in local environment upon labeling (45), which is supported by the low root mean-square deviation of 0.016 from $S = 0.5$ averaged over all variants, with no individual deviation being greater than 0.02. Moreover, the fluorescence lifetimes exhibited only a small variability among protein labeling variants. The average donor lifetime of all variants (obtained from the donor-only populations) was 2.85 ns, with a standard deviation of 0.08 ns, close to the estimated uncertainty of the donor lifetime measurements of 0.06 ns. The average acceptor lifetime of all variants (obtained by direct excitation of the acceptor) was 3.24 ns, with a standard deviation of 0.05 ns, close to the estimated uncertainty of the acceptor lifetime measurements of 0.03 ns. Moreover, we observe no correlation between the deviations from S and the fluorescence lifetimes of the individual variants, indicating that the small residual deviations from $S = 0.5$ for individual variants cannot be accounted for by differences in quantum yields reflected by altered fluorescence lifetimes. Possible contributions may originate from residual static quenching, which can affect observed transfer efficiencies without affecting fluorescence lifetimes (45,88).

Rotamer library generation and simulation of distance distributions

Rotamer libraries contain precalculated preferred conformations of the spin or fluorescence label of interest. The precalculation is carried out once for a new label and allows for generating a rotamer library that can subsequently be integrated into the open-source software-tool Multiscale Modeling of

Macromolecules (MMM) (57), available on www.epr.ethz.ch and on [GitHub.com/gjeschke](https://github.com/gjeschke). Once such a library is available, rotamer library analysis (RLA) and calculation of distance distributions can be carried out in MMM for any biomolecular structure (details below).

The generation of a rotamer library consists of three steps: geometry optimization, Monte Carlo sampling, and clustering.

Geometry optimization

The labeled side chains of cysteine-maleimido-CF660R or cysteine-maleimido-Cy3b were first drawn in 3D using the open-source molecular editor Avogadro 1.2.0 (89), which also provided an initial geometry optimization using the universal force field (UFF) (90). Further geometry optimization was carried out by density functional theory in Orca 3 (91) at the restricted Kohn-Sham level of theory using the functional BP86 (92,93) and the triple- ζ basis set def2-TZVP (94), with the resolution of identity (RI) approximation (95,96). Density functional theory calculations further included the conductor-like screening model (97,98), a continuum model to mimic bulk electrostatics of the solvent water.

Monte Carlo sampling

The total conformational space of the labeled side chains is in good approximation given by combining all possible torsion states of the rotatable dihedral angles, of which there are 8 and 12 for Cy3b and CF660R, respectively. Hence, choosing the dihedral angles as the only degrees of freedom, we used a variant of Metropolis Monte Carlo sampling (99) to generate ensembles of 500,000 structures representative of all side-chain conformations using custom-written scripts in MATLAB 2018b (The MathWorks, Natick, MA). The energies taken into account are the dihedral angle and Lennard-Jones potentials from the UFF force field (90) with parameters from the Towhee implementation (100). The Lennard-Jones potential is of the form

$$E_{LJ} = \sum_{ij} \varepsilon_{ij} \left[\left(\frac{f \sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{f \sigma_{ij}}{r_{ij}} \right)^6 \right], \quad (1)$$

where E_{LJ} is the total Lennard-Jones energy summed over all pairs of atoms with indices i and j . ε_{ij} and σ_{ij} are the depth of the potential well and the van der Waals radius, respectively, specific to atoms i and j as taken from the UFF force field. The f factor, also known as the “forgive factor,” is an empirical softening parameter for the potential that we used for small spin labels only when computing the interaction of the label with the protein (50). A different modified Lennard-Jones potential was used in construction of rotamer libraries for native amino acid side chains (101). For chromophores as well as for large spin labels, it is important to soften the potential also for intralabel atom pairs, when only dihedral degrees of freedom are considered, to account for the combined effects of small bond and angle variations that sum up with the increasing length of a label. The f factor also allows us to tune the attractive term in the potential in the absence of a solvent model, whereas a cutoff distance was used for native side chains in SCWRL4 (101). Because there is currently no method available to predict an optimal f value, rotamer libraries with different f factors were compared to the experimental data. Based on this comparison (vide infra), we used $f = 0.175$ unless stated otherwise. This value is much lower than the ones used previously for generating rotamer libraries of small spin labels ($f = 0.7 \dots 1.0$) or for computing spin label-protein interactions ($f = 0.4 \dots 0.6$).

For Monte Carlo sampling, random values for the dihedral angles are drawn from a set of values distributed according to the dihedral potentials. The Lennard-Jones energy ε_i of the new conformer is calculated and Boltzmann-weighted with a temperature $T = 298$ K and the Boltzmann constant k_B to generate a population $p_i^{MC} = \exp(-(\varepsilon_i - \varepsilon_0)/(k_B T))$ with respect to a minimal Lennard-Jones energy ε_0 determined before by 5×10^7 Monte Carlo trials. If the normalized population p_i^{MC} is $\geq 1\%$, the new conformer is accepted into the resulting Monte Carlo ensemble. The threshold of 1%

ensures sufficient sampling of the large number of canonical rotamers. The ensemble size of 500,000 used here is at the limit of what the subsequent clustering step could process in our case. Sampling appears sufficiently converged at this size for both labels, as judged from the dihedral-angle histograms (Fig. S6).

Clustering

The resulting Monte Carlo ensembles were reduced to a smaller number of N_{rot} structures by hierarchical clustering using custom-written MATLAB scripts, where N_{rot} should be large enough to sufficiently represent the full conformational space of the label and small enough for computational efficiency. The clustering algorithm calculates the pairwise similarity of all members of the Monte Carlo ensemble using distance in dihedral angle space as the metric. This procedure divides the initial ensemble into N_{rot} clusters of similar ensemble members. For each cluster, one average structure was determined to represent the whole group in the final library of N_{rot} rotamers. The average was taken in torsion angle space to avoid unphysical conformers. For each rotamer, its population p_i^{rot} was calculated as $p_i^{rot} = \sum_j p_j^{MC}$ by summing up the populations p_j of the cluster members, which results in normalized probabilities with $\sum_i p_i^{rot} = 1$.

Simulation of distance distributions

Simulations of the spatial distribution of the individual spin or fluorescence labels as well as of the distance distributions were carried out using MMM (57). First, loading a structure file, we used either all 20 structures of the solution NMR ensemble (Protein Data Bank (PDB): 2ADC (74)) or, when stated, only the first NMR structure with truncation of the flexible N-terminus (RRM3/4- Δ N without residues G324 to N336). In both cases, only the protein in the structure file was considered. Then, all labeling sites (residues) and one label, such as MAP in the case of the spin label maleimido proxyl, were selected. For each site, the populations p_i for all i rotamers are calculated as $p_i = p_i^{clash} \times p_i^{rot}$, where p_i^{clash} is the i -th rotamer population due to clashes of the label with other residues of the protein. The clash energies ε_i^{clash} were determined by the Lennard-Jones potential as above. Subsequent Boltzmann weighting yields populations $p_i^{clash} = \exp(-\varepsilon_i^{clash}/k_B T)/Z_{clash}$ with $T = 298$ K and $Z_{clash} = \sum_j \exp(-\varepsilon_j^{clash}/k_B T)$ such that $\sum_i p_i^{clash} = 1$. Finally, we also normalize p_i to $\sum_i p_i = 1$.

For each pair of labeling sites, distance distributions were then calculated as population-weighted histograms of all pairwise distances r_{ij} between populated rotamers on the two sites i and j , i.e., a histogram of $p_i \times p_j \times r_{ij}$ for all pairs $\{i, j\}$. For the ensemble distance distributions, we considered only site pairs within the same conformer. Hence, for each of the 20 structures, a distance distribution was calculated, and subsequently these distributions were summed up and renormalized to $\sum_r P(r) = 1$ to form the final ensemble distribution.

The rotamer libraries for the fluorescence labels maleimido-Cy3b and maleimido-CF660R bound to cysteine (called Cy3 and CF6, respectively, in MMM) were made available as custom rotamer libraries under the names Cy3_298K_UFF_N_r1 and CF6_298K_UFF_N_r1, where N is the library size with 1024, 2048, 4096, or 8192 rotamers (see [Supporting materials and methods](#) for further details). The distance r_{ij} between two fluorescence labels was approximated by the distance between the centers of the two chromophores using the midpoint between the central oxygen and the opposite carbon atom. After selection of the rotamer libraries, the same workflow (see [Supporting materials and methods](#)) was used as for spin labels.

Calculation of mean transfer efficiencies from interdyer distance dynamics in a potential of mean force

From peak fitting of the measured transfer efficiency histograms, we obtained the mean transfer efficiencies (E). For comparing these measured values with the theoretical values derived from a distance distribution $P(r)$, we need to take into account that the interdyer distance r fluctuates

on a timescale similar to the fluorescence lifetime of the donor (2.7 ns). Hence, we model the interdyne dynamics in terms of diffusive motion in a potential of mean force corresponding to $P(r)$, as recently described (102). Briefly, we used the general relation (103)

$$\langle E \rangle = 1 - \langle \tau^* \rangle / \tau_D, \quad (2)$$

where $\tau_D = 1/k_D$ is the mean fluorescence lifetime of the donor in the absence of an acceptor and $\langle \tau^* \rangle = \int_0^\infty p_{D^*}(t) dt$ is the mean lifetime of the excited state of the donor dye in the presence of the acceptor. $p_{D^*}(t)$ is the survival probability of the excited state, which we calculate as $p_{D^*}(t) = \int_0^\infty p_{D^*}^*(r, t) dr$ (103) from the solution of the Smoluchowski equation:

$$\frac{\partial p_{D^*}(r, t)}{\partial t} = \left[\mathcal{L} - k_D - k_D \left(\frac{R_0}{r} \right)^6 \right] p_{D^*}(r, t) \quad (3)$$

with the diffusion operator

$$\mathcal{L} = D \frac{\partial}{\partial r} P(r) \frac{\partial}{\partial r} (P(r))^{-1}, \quad (4)$$

where D is the effective diffusion coefficient for the relative translational motion of the fluorophores attached to the protein, given by the sum of the diffusion coefficients of the individual attached dyes. The Smoluchowski equation needs to be solved with the initial condition $p_{D^*}^*(r, t=0) = P(r)$, assuming that the donor was excited at $t=0$ and that the average interphoton time is much longer than the relaxation time of the interdyne distance.

For the calculations, we discretized the Smoluchowski equation with respect to r over an interval (r_a, r_b) outside which $P(r)$ is zero to good approximation. As a result, Eq. 3 becomes a rate equation:

$$\frac{d}{dt} \mathbf{p}^* = (\mathbf{K}^{\text{diff}} + \mathbf{K}^{\text{depop}}) \mathbf{p}^*, \quad (5)$$

for the vector \mathbf{p}^* of components $p_i^*(t)$, with $i = 1 \dots N$ indicating the discretization steps. The $N \times N$ matrix \mathbf{K}^{diff} describes the dye diffusion and contains the following nonzero matrix elements:

$$\mathbf{K}_{i,i \pm 1}^{\text{diff}} = \frac{D}{\delta r^2} \frac{P(r_{i \pm 1/2})}{P(r_{i \pm 1})} \quad \text{and} \quad \mathbf{K}_{i,i}^{\text{diff}} = -\mathbf{K}_{i-1,i}^{\text{diff}} - \mathbf{K}_{i+1,i}^{\text{diff}}, \quad (6)$$

where $r_i = r_a + (i-1)\delta r$ with $\delta r = (r_b - r_a)/(N-1)$. The matrix $\mathbf{K}^{\text{depop}}$ is diagonal with elements

$$\mathbf{K}_{i,i}^{\text{depop}} = -k_D \left(1 + (R_0/r_i)^6 \right) \quad (7)$$

and describes the depopulation of the excited donor state. The mean lifetime of the excited state of the donor dye is then calculated from

$$\langle \tau^* \rangle = \int_0^\infty \mathbf{1}^T e^{(\mathbf{K}^{\text{diff}} + \mathbf{K}^{\text{depop}})t} \mathbf{p}_{eq} dt, \quad (8)$$

where $\mathbf{1}^T = (1, 1, \dots)$ is the transposed vector of ones, \mathbf{p}_{eq} is the vector defined by $\mathbf{K}^{\text{diff}} \mathbf{p}_0 = 0$, and $\mathbf{1}^T \mathbf{p}_0 = 1$. The effective diffusion coefficient D is difficult to measure. Here, we use recently published all-atom MD simulations of polyproline-11 labeled at its ends with Alexa 594 and Alexa 488

(72) to obtain an interdyne distance distribution $P(r)$ and the time correlation of the distance (see Fig. S4), which we fit with (104)

$$G(\tau) = \frac{\mathbf{1}^T \mathbf{R} e^{\mathbf{K}^{\text{diff}} \tau} \mathbf{R} \mathbf{p}_0}{(\mathbf{R} \mathbf{p}_0)^2}, \quad (9)$$

where D is the only fit parameter. \mathbf{R} is the diagonal matrix with elements $\mathbf{R}_{ii} = r_i$. All calculations were done with $N = 50$. We used $\tau_D = 2.7$ ns (105) for Cy3b (derived from the subpopulation with a stoichiometry ratio of one, i.e., the population corresponding to molecules lacking an active acceptor). The Förster radius for the Cy3b/CF660R dye pair, $R_0 = 6.0$ nm, was calculated using the emission spectrum of Cy3b, the absorption spectrum of CF660R (both of the free dye), a fluorescence quantum yield of 0.67 for Cy3b, an excitation coefficient of $100,000 \text{ M}^{-1} \text{ cm}^{-1}$ for CF660R (manufacturer's specifications), a refractive index of 1.334 (100 pM in 20 mM Tris, 125 mM KCl (pH 7.4) with 0.001% Tween 20), and $\kappa^2 = 2/3$.

RESULTS

To probe both absolute distances and small distance variations, we designed six double-cysteine variants of a protein construct containing the two domains RRM3 and RRM4 of PTBP1 (RRM3/4), which provides a well-structured scaffold for positioning the labels (74,75). In each variant, one cysteine is located on α -helix 1 of RRM4 (α_1) and the other on α -helix 2 on RRM3 (α_2 , see Fig. 1). On α_1 , one of the following residues are replaced by cysteine: E468 (*blue*), V472 (*red*), or S475 (*green*); on α_2 , one of two positions: Q388 (*magenta*) or S392 (*orange*). In this way, the neighboring label positions are separated by one helix turn, which allows us to probe small yet well-defined differences in intramolecular distances by comparing these labeling positions (Fig. 1 A). The selected residues are located at solvent-exposed sites that allow for accessibility and sufficient steric freedom of the labels. The described choice of positions is based on recently published EPR distance distribution measurements (Q388C-S475C) (75). We labeled the six double-cysteine variants of RRM3/4 either with pairs of identical spin labels for DEER distance distribution measurements or with donor-acceptor dye pairs for single-molecule FRET measurements, using maleimide-based coupling chemistry in both cases. Fig. 1 B shows the chemical structures of the labels including their linkers, the maleimide group, and the cysteine. (The representation is drawn to scale with respect to the structure of RRM3/4 in Fig. 1 A.) The linker of the nitroxide spin label MAP is just long enough to reduce interference with the protein structure; the fluorescence labels, Cy3b and CF660R, have longer linkers to reduce fluorescence quenching and provide the rotational flexibility required for averaging the orientation factor to $\kappa^2 = 2/3$ (36). The structure of Cy3b has been published (106), and the structure of CF660R was determined based on a patent (107) and experimental analysis (see Supporting materials and methods).

Successful spin labeling was verified by cw EPR spectroscopy (Fig. S1); the labeling efficiencies are given in

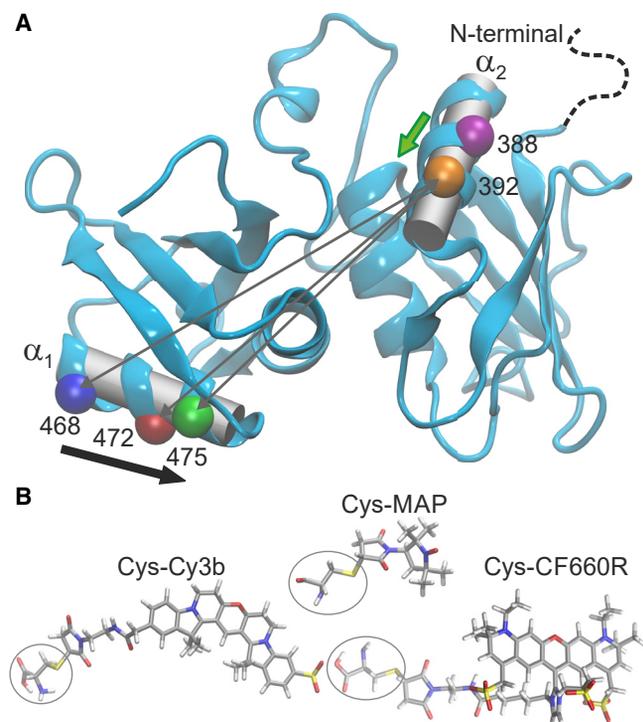


FIGURE 1 Cysteine positions for fluorescence and spin labeling. (A) PTBP1-RRM3/4 (blue ribbons) with labeling positions highlighted (C_α atoms as colored spheres, connected by gray arrows for S392C). Positions are offset by single turns along the two helices α_1 and α_2 (gray cylinders), resulting in a shift to smaller distance (black arrow) from E468C (blue) to S475C (green), or from Q388C (magenta) to S392C (orange) (green arrow). The flexible N-terminus is indicated by a dashed line. (B) Fluorescence- and spin-labeled cysteine side chains (stick representation) are drawn to scale with respect to (A). To see this figure in color, go online.

Table S1. Distance distributions between spin labels measured by four-pulse DEER (30) with subsequent analysis by model-free Tikhonov regularization (80,81) are shown in Fig. 2 A (see Fig. S2 for time-domain data) and report on distances and, because the signal/noise ratio achieved here is sufficient, also on distribution widths between the N-O groups present in the sample. The widths of the distance distributions arise from the conformational distributions of both the protein and the spin-labeled side chains. These are the conformations present at the glass transition temperature of the matrix (108), which are trapped upon rapid freezing. In case of the rather rigid RRM3/4 (75), the width of the distance distribution is dominated by the conformational distribution of the spin labels. Accordingly, we model the distance distributions by RLA, in which the protein structure is kept fixed, and all label conformations are considered that do not clash with the protein, computed using a softened Lennard-Jones potential as described in Materials and methods. The distance distributions resulting from these RLA simulations for all 20 RRM3/4 conformations available in the NMR structure ensemble (PDB: 2ADC (74)) are shown in Fig. 2 B. The conformational distributions of the spin labels, visualized

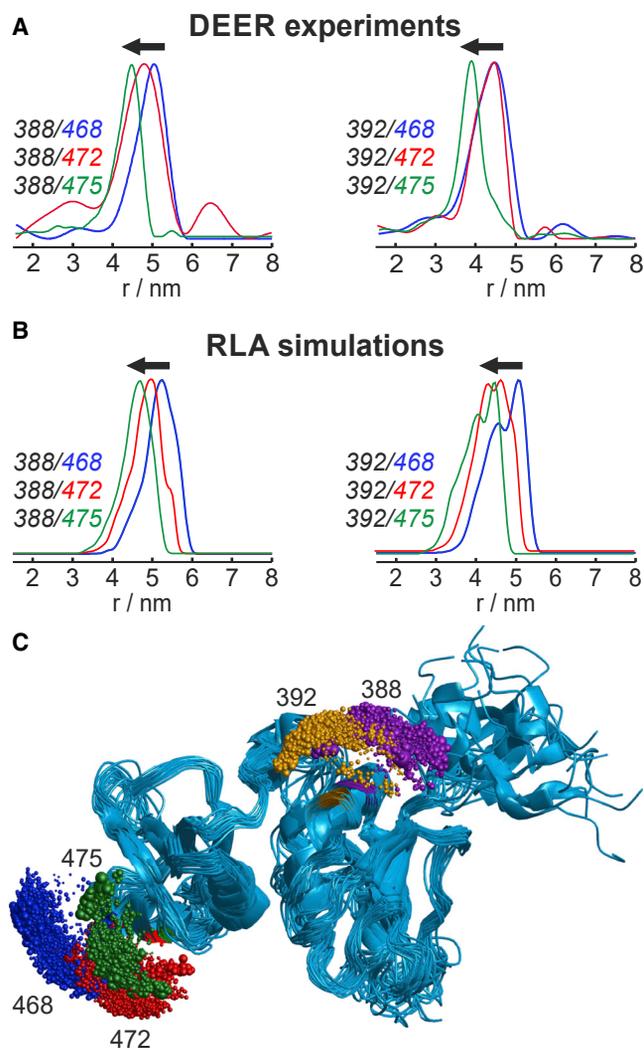


FIGURE 2 EPR distance determination by DEER and simulations. (A) Experimental DEER distance distributions (Fig. S2 shows primary data) between RRM3/4 cysteine positions as indicated in the legend, labeled with maleimido proxyl. (B) Rotamer library simulations for maleimido proxyl for all 20 conformations in the NMR ensemble (PDB: 2ADC (74)). Arrows (in A and B) indicate the shift to smaller distances from E468C (blue) to S475C (green). (C) NMR ensemble of RRM3/4 (blue ribbons, PDB: 2ADC (74)) with point clouds indicating the conformational distributions of maleimido proxyl-labeled side chains from the RLA simulations (colored spheres indicate label positions, size indicates population). Distances between the point clouds are distributed as shown in (B). To see this figure in color, go online.

in Fig. 2 C, illustrate the different states occupied with different probabilities (sphere sizes) that give rise to the anisotropic conformational space of the spin labels, as well as the partial spatial overlap of conformations at adjacent labeling positions. However, the center of gravity is clearly shifted when comparing labels at the different positions. This behavior is also reflected quantitatively by the simulated distance distributions (Fig. 2 B); they overlap, yet their centers are visibly shifted for all positions if one label is moved by a single helix turn (cf. Table S1).

The experimental distance distributions show the same trend as the simulations. The three distributions in which position 388 is labeled on α_2 are clearly shifted to larger distances compared to the corresponding distributions in which position 392 is labeled. The predicted distance shifts for positions 468 \rightarrow 472 \rightarrow 475 are also observed in the DEER distance distributions, with the exception of 468 \rightarrow 472 if position 392 is labeled. In this case, the measured distributions overlap almost completely, and mean distances differ by only 0.1 nm. The similarity of the DEER time-domain data (Fig. S2) underlines that spin labels at these two positions exhibit very similar distances. Taken together, the comparison between the DEER distance distributions and RLA simulations demonstrates that, in most cases, PDS is sensitive enough to measure distance differences as small as a shift of the labeling position by a single helix turn. In a few cases, the difference is obscured by the conformational distribution of the spin labels, an effect whose contribution is expected to depend on the relative orientation of the two α -helices. In terms of absolute distances, we find good agreement to within 2.2 Å on average (Table S1) between the experimental and simulated distance distributions, similar to the 2.5 Å reported for other nitroxide spin labels (51,56).

EPR distance measurements are carried out between identical spin labels, but distance measurements based on FRET require a pair of different fluorescence labels. We employed substoichiometric labeling with the first dye, followed by chromatographic separation of RRM3/4 singly labeled at one cysteine or the other and subsequent labeling of the corresponding other site with the second label (see Materials and methods for details). This approach led to site-specific labeling of RRM3/4 with the donor Cy3b and the acceptor CF660R and allowed for separate measurements of both label permutations for each pair of sites. The resulting single-molecule FRET efficiency histograms of all 12 RRM3/4 labeling variants, measured in free diffusion at 22°C, are shown in Figs. 3 A and S3 (Fig. 3 shows only the peak functions used to fit the histograms to make the small differences in peak positions visible). Note that the widths of the peaks in the FRET efficiency histograms are dominated by shot-noise broadening owing to the limited number of photons detected while single protein molecules diffuse through the confocal volume (109–111). Because the diffusion time through the focus is in the millisecond range but the interdye distance dynamics occur on the nanosecond timescale, only the mean transfer efficiencies $\langle E \rangle$ are obtained from the peak positions in the histograms (see Table S3). The transfer efficiencies clearly shift to higher values when the α_2 labeling positions are moved from 388 to 392 and also for α_1 positions 468 \rightarrow 472, 475, whereas for 472 \rightarrow 475 we observe a clear shift only for two out of four RRM3/4 variants. Notably, $\langle E \rangle$ for the donor-acceptor labeling permutations on the same double-cysteine constructs reveals small but significant

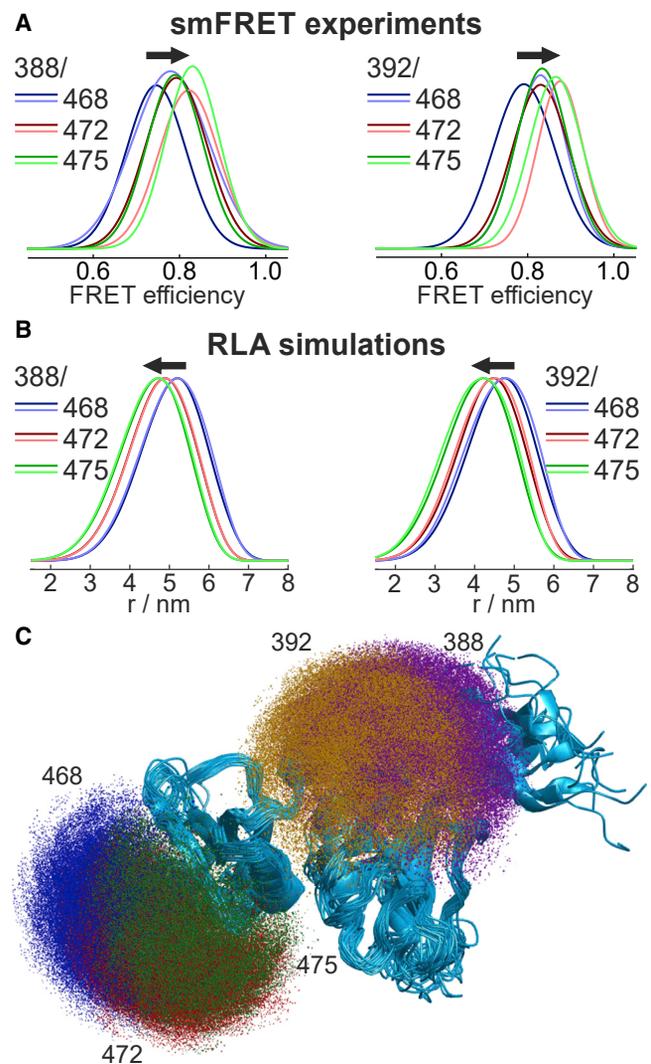


FIGURE 3 Single-molecule FRET measurements and comparison with simulations. (A) Experimental single-molecule FRET efficiency histograms for RRM3/4 represented by the Gaussian fits of the peaks (solid lines; cf. Fig. S3 for the original histograms, which are omitted here to improve visualization of the small differences). The labeling positions are indicated (see legend), with darker or lighter colors for helix α_2 labeled with the acceptor or the donor dye, respectively. (B) Interdye distance distributions obtained from RLA simulations for all 20 conformations of the NMR ensemble (PDB: 2ADC (74)) with the same color code as in (A). Arrows in (A) and (B) indicate the shift to smaller distances or higher FRET efficiencies from E468C (blue) to S475C (green). (C) NMR ensemble of RRM3/4 (blue ribbons, PDB: 2ADC (74)) with spatial distributions of the chromophore centers (point clouds) from rotamer library simulations for fluorescence label Cy3b on Q388C and S392C and for CF660R on E468C, V472C, and S475C for all 20 conformations in the NMR ensemble. To see this figure in color, go online.

differences, with consistently lower $\langle E \rangle$ if CF660R is attached at positions 388 or 392. Static acceptor quenching by surface residues close to helix α_2 might contribute to the observed differences (88) (Table S3). We note that the average difference in $\langle E \rangle$ between the permuted variants of 0.039 is in a similar range as the uncertainty in $\langle E \rangle$ from a recent

multilaboratory benchmark study (86) and would result in a difference in inferred distance of ~ 0.2 nm.

For comparing the experimental $\langle E \rangle$ with the structure-based predictions, we first calculate distance distributions between the fluorophores and subsequently use these to predict FRET efficiencies, taking into account the dynamics of the flexible linkers. To calculate distance distributions, we generated rotamer libraries for the fluorescence labels by extending our previous approach for spin labels (50) to significantly larger sampling and clustering capacities required for the larger fluorophores (see **Materials and methods** for details). The libraries ideally cover all label conformations allowed by dihedral rotations. Depending on the local symmetry, the potentials of the dihedral angles have multiple minima, the canonical dihedral angles. The total number of possible label conformations is given by the product of the numbers of minima of all dihedral angles. Cysteine labeled with the common nixtroxide methanethiosulfonate spin label has five dihedral angles, χ_1 to χ_5 , which have 3,3,2,3,4-fold symmetry and hence give rise to $3 \times 3 \times 2 \times 3 \times 4 = 216$ canonical conformers (50); cysteine labeled with MAP has five rotatable dihedral angles, with 3,3,3,1,1-fold symmetry (27 canonical rotamers). In addition, for the MAP library, we had considered two diastereomeric pairs of enantiomers with equal probability and had hence ended up with 108 conformers. This is in stark contrast to the fluorescence labels, which have many more dihedral angles, namely 8 and 12 for Cy3b and CF660R, respectively, leading to 26,244 and 2,125,764 canonical conformers for the two labels. This illustrates how the number of configurations grows exponentially with the number of dihedral angles, which is a challenge both for sampling and for representing these by increasingly large rotamer libraries.

Accordingly, to generate the rotamer libraries for the two fluorescence labels Cy3b and CF660R (see Fig. S5 for flow-chart), we started from structures of the fluorophore-labeled cysteines that were geometry-optimized using quantum chemistry calculations (see **Materials and methods** for details). Subsequently, we generated conformational ensembles with 500,000 structures of the free labels using parallel Monte Carlo sampling in dihedral angle space applying a softened Lennard-Jones potential combined with the dihedral angle potentials provided by the UFF force field (90,100). The f factor tunes two effects: it accounts for otherwise neglected flexibility due to small variations in bond lengths and angles that accumulate with increasing chain length, and it adjusts the attractive term of the label-protein potential to model effective surface interactions. Because the optimal f factor is not known a priori, we scanned a range of values (see **Materials and methods**) for comparison to the experimental data. The resulting Monte Carlo ensembles were checked for convergence (Fig. S6), clustered in dihedral angle space (Figs. S7 and S8), and we checked the clustering (Fig. S9) to find a reduced subset

of representative rotamers. This set of representative conformers forms a rotamer library.

The advantage of precalculating rotamer libraries is that subsequent calculations of energy-weighted label conformations at a specific site on any protein become computationally much less expensive and hence more readily accessible compared to Monte Carlo sampling of the label conformations directly on the protein. These RLA simulations of RRM3/4 labeled at different positions with Cy3b and CF660R give rise to distance distributions calculated between the centers of the two chromophores (Fig. 3 B). The distance distributions calculated with differently softened Lennard-Jones potentials reveal a nonlinear dependence of distances and distribution widths on the f factor (Fig. S10) because of a balance of the two effects it tunes, which underlines the importance of validating the model against experimental data (see below). The volumes over which the chromophore centers are distributed on RRM3/4 are shown in Fig. 3 C and illustrate both the extent of the spatial delocalization due to the linkers and the rather smooth and isotropic nature of the conformational space accessible to the fluorescence labels compared to the spin labels with their shorter linkers (Fig. 2 C), which renders the rotamer libraries of Cy3b and CF660R less sensitive to the library size (Fig. S10). Distributions of the orientation factor κ^2 (72) for the different label pairs yield average κ^2 -values of 0.644–0.674 (data not shown), close to the ideal isotropic value of $\kappa^2 = 2/3$, and the extent of the accessible conformational space gives rise to distance distributions between Cy3b and CF660R that are between 1.99 and 2.25 nm wide (full width at half maximum (FWHM)) (Table S2). Despite this considerable width, the mean values are clearly shifted to shorter distances along the series 468 \rightarrow 472 \rightarrow 475 in α_1 as well as when comparing α_2 positions 388 \rightarrow 392 for all three positions of the second label (Fig. 3 B). When exchanging the donor and acceptor labels in the simulations, we do not observe significant changes in the distance distributions for any of these combinations (Fig. 3 B; Table S2).

For comparing with the experimental FRET results, we need to calculate the expected mean transfer efficiencies from the simulated interdye distance distributions. This procedure is, however, complicated by the relative diffusive motion of the dyes, which occurs on the same timescale (nanoseconds) as the excited state lifetime of the donor (43,44,72). This contribution is ignored when assuming a static distance distribution and calculating the mean transfer efficiency from $\langle E \rangle = \int E(r)P(r)dr$, where $E(r)$ is the transfer efficiency at the (instantaneous) interdye distance r (42–44,102,112). Furthermore, we assume rotational diffusion to be faster than translational diffusion (63) and accordingly κ^2 averaging to be fast. We thus account for the dynamics by describing the fluctuations in interdye distance in terms of diffusive motion in a potential of mean force, with the potential calculated from the distance distribution obtained by the RLA simulations (see **Materials and**

methods for details). In this diffusional averaging procedure, the largest uncertainty originates from the assumed effective diffusion coefficient D , which is difficult to measure. An upper limit is given by $D = D_1 + D_2$, where D_1 and D_2 are the translational diffusion coefficients of the free dyes. The diffusion coefficient of Cy3b at room temperature in water is $0.44 \text{ nm}^2/\text{ns}$ (105). Assuming this value for both dyes results in $\sim 0.88 \text{ nm}^2/\text{ns}$ as an upper limit for D . However, the dyes' motion is expected to be slowed down when they are attached to the protein. Peulen et al. (44) recently obtained a diffusion coefficient of Alexa488-C5-maleimide attached to the human guanylate binding protein 1 of $0.1 \text{ nm}^2/\text{ns}$ by comparing simulated fluorescence decays for various diffusion coefficients with the corresponding experimental fluorescence decay. The simulated decays were obtained from Brownian dynamics simulations of the dye in its AV and included quenching of the dye upon collision with amino acids on the protein surface. Assuming the same diffusion coefficient for both dyes used here would result in $D = 0.2 \text{ nm}^2/\text{ns}$. We obtained a very similar value from recently published all-atom MD simulations of polyproline-11 labeled with Alexa594 via an N-terminal Gly and with Alexa488 via a C-terminal Cys (72). Using the interdye distance distribution and time correlation derived from the MD simulations, we determined an effective diffusion coefficient of $D = 0.22 \text{ nm}^2/\text{ns}$ (see Materials and methods and Fig. S4 for details). We are not aware of such values for the specific combination of Cy3b and CF660R, but in view of the similarity in fluorophore size to Alexa488/594, we assume $D \approx 0.2 \text{ nm}^2/\text{ns}$.

The Förster radius, too, can only be determined to limited accuracy, with a recent estimate of its uncertainty of $\sim 7\%$ (86). Because of the described uncertainties in D and R_0 , the systematic uncertainties for the transfer efficiencies $\langle E \rangle_{\text{sim}}$ calculated from the RLA distributions are much greater than the statistical errors in the measured transfer

efficiencies $\langle E \rangle_{\text{exp}}$ (less than ± 0.01 ; see Table S3). To visualize the effect of the uncertainties in D and R_0 on $\langle E \rangle_{\text{sim}}$, we thus show in the contour plot of Fig. 4 A the root mean-square deviation (RMSD) between $\langle E \rangle_{\text{sim}}$ and $\langle E \rangle_{\text{exp}}$ averaged over all 12 RRM3/4 variants (including labeling permutations) calculated for R_0 ranging from 5.6 to 6.4 nm and D ranging from 0 (no diffusion) to $0.9 \text{ nm}^2/\text{ns}$ (free-dye diffusion). The combination of values we consider most likely according to the discussion above, $R_0 = 6.0 \text{ nm}$ and $D = 0.2 \text{ nm}^2/\text{ns}$, is indicated with a red cross. The simulated values $\langle E \rangle_{\text{sim}}$ from diffusional averaging, calculated for these parameters ($R_0 = 6.0 \text{ nm}$ and $D = 0.2 \text{ nm}^2/\text{ns}$) are compared to the experimental values $\langle E \rangle_{\text{exp}}$ in Fig. 4 B (green) for all 12 RRM3/4 variants. The shaded bands reflect the uncertainty in R_0 and D that are calculated as standard deviations of all transfer efficiencies for the R_0 , D combinations shown in Fig. 4 A, i.e., for $R_0 = 6.0 \text{ nm} \pm 7\%$ and $D = 0 \dots 0.9 \text{ nm}^2/\text{ns}$ for all 12 RRM3/4 variants. The comparison of simulated and experimental transfer efficiencies allows us to determine the optimal Lennard-Jones softening f factor, because for plausible values of R_0 and D , agreement between simulated and experimental FRET efficiencies is obtained only in a narrow range for the f factor, around $f = 0.175$ (Fig. S11). For this value, the standard deviation owing to R_0 and D overlaps with the ideal 1:1 correlation, and the RMSD shows a clear minimum and thus underlines the optimal choice of f for this particular set of rotamer libraries for Cy3b and CF660R (Fig. S11). For f factors outside the range of $f = 0.15\text{--}0.2$, the discrepancy between $\langle E \rangle_{\text{exp}}$ and $\langle E \rangle_{\text{sim}}$ cannot be compensated by varying R_0 and D in reasonable ranges. Hence, we consider the decrease in RMSD around this interval a significant improvement. Note that the optimal f factor for the fluorescence labels is smaller than the values of $f = 1.0 \dots 0.7$ usually applied to generate rotamer libraries of spin labels. This finding points to two different regimes: one for the short

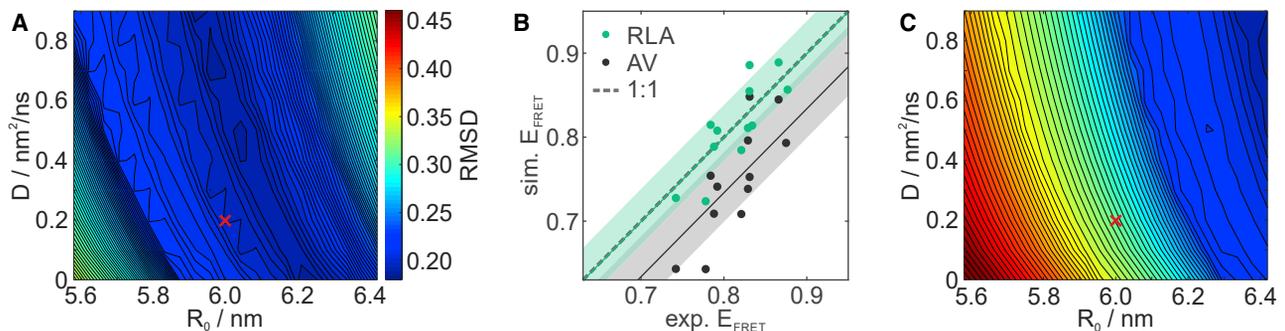


FIGURE 4 Comparison of experimental smFRET efficiencies to predictions using diffusional averaging over distance distributions. (A and C) RMSD between experimental and simulated FRET efficiencies averaged over all 12 data sets (see Table S3) for different Förster radii ($R_0 \pm 7\%$ (86)) and effective diffusion constants of the fluorescence labels based on the distance distributions (see Figs. 3 and S12) of the RLA simulations (A) and AV simulations (C) using the same color scale given in (A). The red crosses mark $R_0 = 6.0 \text{ nm}$ and $D = 0.2 \text{ nm}^2/\text{ns}$ and correspond to the circles in (B). (B) Experimental versus simulated FRET efficiencies from RLA (green) and AV analysis (black) with shaded bands corresponding to uncertainties in Förster radius R_0 and effective diffusion constant D (see A). Linear fits to the data with slope 1 (green and black lines for RLA and AV, respectively) show deviations from the ideal 1:1 correlation (gray dashed line). To see this figure in color, go online.

spin labels, in which particular conformations of the linker dominate, and one for the longer fluorophore linkers, in which a coarse-grained model with populations assigned per unit volume rather than per rotamer may become a suitable representation for the spatial distribution. This interpretation also explains the decreased sensitivity of the distance distributions we observed for rotamer library sizes smaller than the number of canonical rotamers for the fluorescence labels (Fig. S10). For the optimal rotamer libraries, we found that the lowest RMSD between experimental and predicted FRET efficiencies is close to the most likely values of $R_0 = 6.0$ nm and $D = 0.2$ nm²/ns. Using these parameters for the modeling of FRET efficiencies by diffusional averaging, we find agreement with the experimental results to within ± 0.025 in half of the cases and to within ± 0.045 in 11 out of our 12 cases (Table S3).

For comparison, we also calculated the distance distributions resulting from AV simulations (37,58) for the N-terminal truncation variant RRM3/4- Δ N (Fig. S12). We found the single RRM3/4- Δ N structure to be a suitable representation for the ensemble of all 20 NMR structures, as indicated by the similarity of the resulting RLA distance distributions in both cases (Fig. S13; Table S2). In analogy to Fig. 4 A, we also calculated, for the distance distributions derived from the AV analysis, the RMSD to the experimental transfer efficiencies as a function of R_0 and D (Fig. 4 C). The result shows that a larger Förster radius of $R_0 \approx 6.3$ nm and/or a higher value for D would need to be assumed to reach similarly low RMSD values as in the RLA simulations. Nevertheless, simulated transfer efficiencies based on the AV distance distributions that include diffusional averaging with $D = 0.2$ nm²/ns (see Fig. 4 B, black) are closer to the experimental values than transfer efficiencies based on the static FRET approximation (see Fig. S11; Table S3). Notably, AV-based transfer efficiencies calculated both with and without diffusional averaging showed systematically lower values than the experiment. Hence, the prediction closest to the experimental results for our test set of 12 RRM3/4 variants are the transfer efficiencies calculated with diffusional averaging using the RLA-derived distance distributions (Figs. 4 B and S11; Table S3), suggesting that the spatial distributions of the fluorophores are better represented by the rotamer distributions using this optimization (Fig. 3 C) than by the AV distributions with parameters derived from the label structure (Fig. S12).

Our observations from both EPR and single-molecule FRET are summarized in Fig. 5. Shifting the labeling position on helix α_2 from 388 to 392 reduces the distance to helix α_1 and thus leads to increased transfer efficiencies, as calculated using diffusional averaging and the RLA distributions (see Fig. 5, green arrows). The same behavior is observed in the experiments for all three positions on α_2 and all corresponding label permutations (Fig. 5, A and B, green arrows). Similarly, shifting the labels from position 468 on α_1 by a single helix turn to position 472 (see

Fig. 5, black arrows) results in consistent shifts in $\langle E \rangle$. In contrast, the shift from position 472 to 475 cannot be resolved reliably. Overall, the FRET efficiency shifts predicted based on the RLA distance distributions are larger than those found experimentally. The AV simulations give results very similar to the RLA simulations for these smaller shifts (cf. Fig. S12; Table S3), and both simulation methods show no relevant difference upon exchange of the two fluorophores. Therefore, the most likely cause for the discrepancies between model and experiment is different local environments of the labeling sites, whose influence on the photophysics or translational and rotational diffusivity of the dyes is not taken into account in the simulations. Our results suggest that the benchmark we have chosen is close to the limits of distance variations that can be resolved by single-molecule FRET.

In the DEER experiments (Figs. 2 and 5 C), the expected shifts of the mean distances are clearly visible in all cases except 392/468 \rightarrow 472. Yet, in both experimental and simulated distance distributions, the shifts are smaller than the widths of the distributions. On average, the FWHM is 0.97 nm, whereas the theoretical distance change for 392/468 \rightarrow 472 is 0.25 nm. Evidently, this distance change is close to the resolution limit accessible for nitroxide spin labels at solvent-exposed helical sites.

The simulated average distances from the RLA distance distributions exhibit a remarkably high correlation between spin and fluorescence labels (Figs. 5 D and S14). Although the widths of the distributions are clearly larger for the fluorescence labels (Tables S1 and S3), the correlation suggests that the average positions of the labels relative to the protein backbone in these cases are surprisingly similar for both types of labels. In contrast, the backbone C_α - C_α distances (Fig. S14) reveal a clear offset from the average interlabel distances. Because all analyzed distances are between the same two α -helices, the offset happens to be constant here, which is not generally the case. This discrepancy emphasizes the benefit of taking the labels into account.

DISCUSSION

In various applications, experimental distance distribution constraints between spin labels together with rotamer library simulations have contributed to determining structures or structural models of biomolecules or their complexes by EPR (16,57,113–120). Similarly, single-molecule FRET has increasingly been used for modeling biomolecular structures (37,61). To facilitate integrative structure modeling that combines experimental data from both EPR spectroscopy and single-molecule FRET, we established here a common framework for comparing structural models to experimental long-range distance constraints from both techniques. The common challenge when applying these techniques to structure determination is that distances are measured between the active centers of the site-specifically

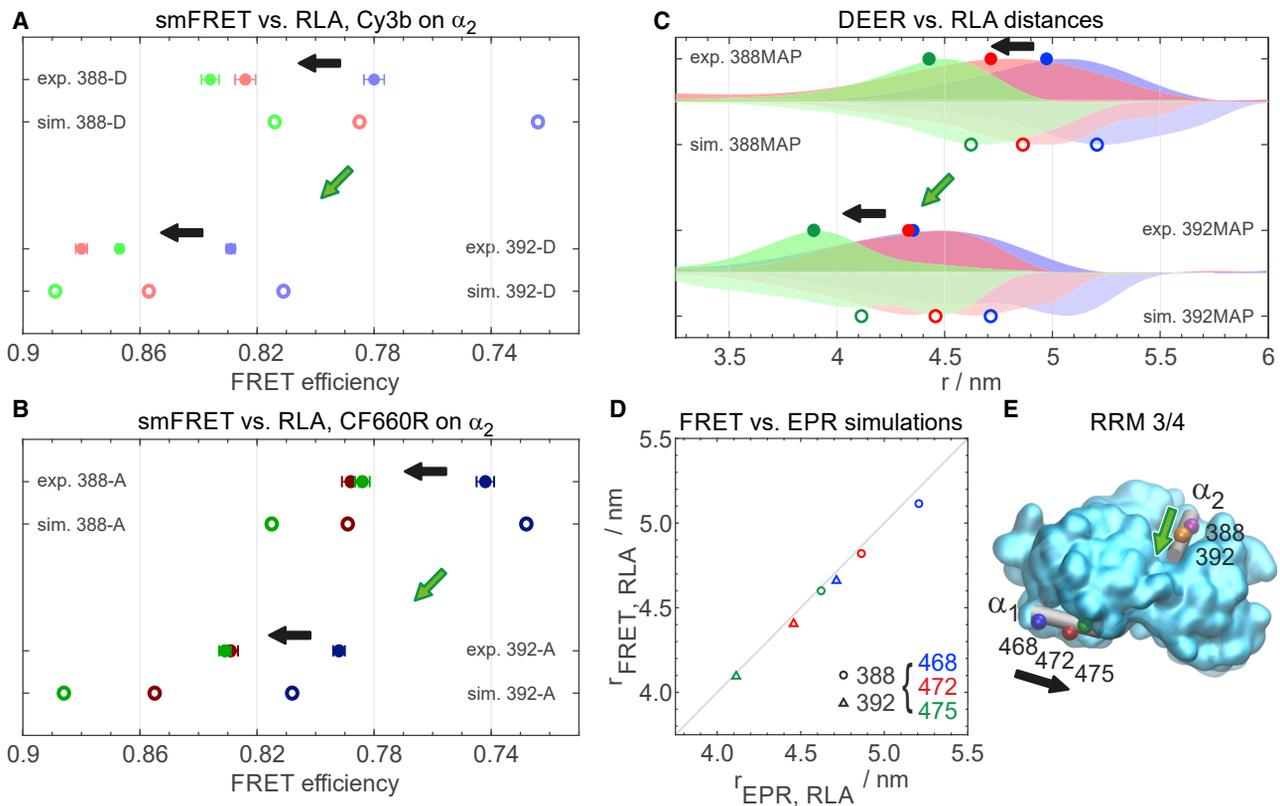


FIGURE 5 Resolving distances and distance variations by EPR and single-molecule FRET (smFRET) using RLA simulations. (A) Comparison of experimental FRET efficiencies (*filled symbols*, error bars represent standard deviations estimated from multiple measurements (Table S3)) with simulated results using RLA distance distributions and diffusional averaging (*open symbols*) for positions 388 and 392 on helix α_2 labeled with Cy3b with the second label at positions 468, 472, and 475 in blue, red, or green, respectively. (B) is analogous to (A) but for CF660R attached to α_2 . (C) Comparison of DEER and RLA distance distributions oriented upwards or downwards, respectively. Labeling positions along α_2 are indicated, and positions 468, 472, and 475 on α_1 are shown in blue, red, or green, respectively. The symbols indicate the average distances. (D) Correlation between RLA average distances calculated for EPR and FRET (see Tables S1–S3 for data). (E) Labeling positions on RRM3/4 with arrows, which indicate distance variations between positions along α_1 (black) or α_2 (green) in all panels. To see this figure in color, go online.

incorporated labels: either the unpaired electrons for EPR, or the transition dipoles of the fluorophores for FRET. Hence, to compare the experimental distance constraints with structures, simulations of the labels at the respective protein sites are required. To have a common simulation technique for both types of labels, we have here transferred the RLA simulations, which are capable of predicting even anisotropic spatial distributions (46,56), to fluorescence labels.

In generating the rotamer libraries, we overcame the sampling problem that grows exponentially with the number of flexible dihedral angles in the linker and found that despite the much longer linkers and thus larger number of canonical conformations of the fluorescence labels, representative sets of not more than 1024 rotamers are sufficient for the analysis (Fig. S10). The likely cause of this finding is that the distributions are relatively smooth for the flexible labels, and hence, fewer discrete points become sufficiently representative. In comparison to the established AV approach (58), we found the predictions by RLA simulations to be slightly more precise. This comes at the cost of increased

computational effort, but once the rotamer libraries are available, RLA simulations are still possible within seconds on a desktop computer. This is efficient enough for screening large numbers of models, as required for integrative structure modeling.

Comparing the uncertainty in the constraints derived from EPR spectroscopy and single-molecule FRET, i.e., the deviations of the RLA simulations from the experimental results, we find these to be considerably larger in terms of absolute distances for the chromophores. There are several reasons for these uncertainties, some of which we have included in our model. An important contribution is the uncertainty in the Förster radius R_0 . Furthermore, protein and fluorophore dynamics make absolute distances more challenging to predict for FRET than for EPR. The simplest approach for FRET uses the static limit (see Results), i.e., it neglects translational diffusion, which causes fluctuations in the interdyer distance on a timescale comparable to the fluorescence lifetime, while at the same time the orientation factor is approximated by $\kappa^2 \approx 2/3$ because of the fast rotational diffusion of the chromophores. To include dynamics

in the model, we use the simulated distance distributions for obtaining potentials of mean force and include the relative translational diffusion of the dyes on the timescale of the donor-excited state lifetime in the analysis. Our calculations (Fig. 4) show that taking fluorophore dynamics into account has a clear influence when evaluating how well a given model fits a set of constraints. The better the Förster radius R_0 and the diffusion constant D can be ascertained experimentally, the more the resulting uncertainties will be reduced until dye quenching and sticking dominate as sources of uncertainty. Sometimes, the latter two effects can be reduced by selecting favorable labeling positions. Despite these uncertainties that give rise to the rather large intervals of D and R_0 (Fig. 4), we found that the extent of overlap of the predicted and experimental transfer efficiencies is distinctly different for the different libraries and hence, we were able to select an optimal set of rotamer libraries (in terms of f factor, see Fig. S11). The f factor determined here is likely to be similar for other dyes with similar linkers that lead to a comparable extent of effective compaction of the linkers and fluorophores, which is affected by the application of the optimal f factor in the rotamer library generation (Fig. S10). The generalizability of the optimal f factor would ideally be tested for additional proteins. However, because the amino acid composition of the protein used here is not unusual, it is reasonable to assume that similar values of f will provide a good approximation also for other folded proteins.

Although the RLA approach provides good agreement with experimental FRET efficiencies on average, the individual values differ from the prediction by <0.025 in half of the cases, and permuting the donor and acceptor results in average differences in the FRET efficiency of 0.04. These deviations are similar to the errors for independent FRET measurements in multiple laboratories obtained in a recent large-scale study (86), suggesting that we are approaching limits in terms of feasible accuracy and precision. Factors that remain untreated in our model, as in most other models, are specific interactions of protein residues with the labels that influence their conformations or, most importantly for FRET, the photophysical properties or orientational distributions of the fluorophores. Accounting for these effects would require a much more detailed treatment with a residue-specific interaction potential. Although the latter is feasible in specifically optimized all-atom force fields (72), aspects such as changes in dye photophysics are beyond current reach. Nevertheless, at this level of accuracy, the predictions are sufficient for quantitative use in integrative structure modeling if the remaining uncertainties are taken into account (86). The approach presented here might work particularly well for modeling conformational changes, in which the local environment at the labeling sites remains the same so that variability in quantum yields and steric restrictions is minimal. If the distance change to be observed is small, site-selective labeling for FRET will be

advantageous to avoid averaging over slightly different transfer efficiencies as observed here for the permutants. Hence, the high sensitivity for distance differences we observed here for both EPR and FRET can best be exploited for modeling conformational changes.

Systematic integration of FRET and EPR restraints would be valuable, e.g., because FRET can be carried out at physiological temperatures and can thus be used to test whether and to what extent the conformations of biomolecules are affected by the shock freezing that is required for EPR distance measurements. Conversely, the combination of the two techniques can potentially reveal whether an ensemble of structures derived from the more accurate EPR restraints is representative of the state present in solution, which is accessible with FRET. Using the RLA approach both for EPR and FRET also makes it easier to relate the detailed shape of distance distributions obtained in EPR to the specific distances and dynamics detected by FRET at ambient temperatures. Because both techniques are capable of detecting the mean label-to-label distances, RLA can also help correlate them to the protein backbone C_α - C_α distances and evaluate the offset between label-to-label and C_α - C_α distances. For the solvent-exposed α -helical sites used here, the mean label-to-label distances are very similar for both label types, despite the significant difference in length between fluorescent and paramagnetic labels (Figs. 5 D and S14), as also observed previously (121). The origin of this similarity is that the spatial distributions of both fluorescence and spin labels are sufficiently uniform such that the centroids are close to each other for the two types of labels (Fig. S14). Most likely, this is not a general result, and deviations between the two RLA-based mean distances for sites with partial steric hindrance or higher anisotropy of the rotamer distributions around the attachment sites could be larger.

CONCLUSIONS

Although it is challenging to use FRET and EPR for revealing small distance variations on the order of one or two turns of an α -helix, we showed here that this resolution could be achieved in most of the cases we probed. This result demonstrates the sensitivity of constraints based on FRET and EPR experiments and RLA for detecting small structural changes for integrative structure modeling. The rotamer library approach could be transferred from spin labels to the significantly larger fluorescence labels. We found that despite the large number of potential rotamers of the fluorescent labels with their long linkers, the approach is computationally feasible because reduced-size libraries of only a few thousand rotamers reproduce the experimental FRET efficiencies well, especially when combined with diffusional averaging that takes the translational motion of the two dyes into account. Using rotamer libraries for both spin and fluorescence labels thus provides a promising

perspective for future applications in integrative structure modeling and enhances the synergy between the two complementary methods.

SUPPORTING MATERIAL

Supporting material can be found online at <https://doi.org/10.1016/j.bpj.2021.09.021>.

AUTHOR CONTRIBUTIONS

D.K., A.H., C.G., D.N., M.Y., F.H.-T.A., B.S., and G.J. designed the research. C.G. and I.R. prepared protein samples and carried out EPR experiments. A.H. carried out fluorescence labeling and single-molecule FRET experiments, analyzed by A.H. and D.N. N.B. conducted NMR experiments and analysis. D.K. and G.J. extended MMM with FRET functionality. D.K. generated rotamer libraries and carried out all simulations, analyzed the data, and wrote the manuscript with contributions from all authors.

ACKNOWLEDGMENTS

We thank Robert Best for helpful discussions and for providing the simulation data for Fig. S4 and Oliver Zerbe for discussions and help with NMR data analysis.

We acknowledge support from a Sinergia Grant of the Swiss National Science Foundation (CRSII5_170976/1 to B.S., G.J., and F.H.-T.A.), and we thank the High-Performance Computing Group (ETH Zurich) for support and access to large-memory machines. We thank the Functional Genomics Center Zurich for performing mass spectrometry.

REFERENCES

- Kim, S. J., J. Fernandez-Martinez, ..., M. P. Rout. 2018. Integrative structure and functional anatomy of a nuclear pore complex. *Nature*. 555:475–482.
- Knockenbauer, K. E., and T. U. Schwartz. 2016. The nuclear pore complex as a flexible and dynamic gate. *Cell*. 164:1162–1171.
- Schlundt, A., J.-N. Tants, and M. Sattler. 2017. Integrated structural biology to unravel molecular mechanisms of protein-RNA recognition. *Methods*. 118–119:119–136.
- Dimitrova-Paternoga, L., P. K. A. Jagtap, ..., J. Hennig. 2020. Integrative structural biology of protein-RNA complexes. *Structure*. 28:6–28.
- Hilger, D., M. Masurel, and B. K. Kobilka. 2018. Structure and dynamics of GPCR signaling complexes. *Nat. Struct. Mol. Biol.* 25:4–12.
- Latorraca, N. R., A. J. Venkatakrishnan, and R. O. Dror. 2017. GPCR dynamics: structures in motion. *Chem. Rev.* 117:139–155.
- Wei, G., W. Xi, ..., B. Ma. 2016. Protein ensembles: how does nature harness thermodynamic fluctuations for life? The diverse functional roles of conformational ensembles in the cell. *Chem. Rev.* 116:6516–6551.
- Steven, A. C., and W. Baumeister. 2008. The future is hybrid. *J. Struct. Biol.* 163:186–195.
- Braitbard, M., D. Schneidman-Duhovny, and N. Kalisman. 2019. Integrative structure modeling: overview and assessment. *Annu. Rev. Biochem.* 88:113–135.
- Tang, H. Y. H., J. A. Tainer, and G. L. Hura. 2017. High resolution distance distributions determined by X-ray and neutron scattering. *In Biological Small Angle Scattering: Techniques, Strategies and Tips* B. Chaudhuri, I. G. Muñoz, S. Qian, and V. S. Urban, eds.. Springer, pp. 167–181. https://doi.org/10.1007/978-981-10-6038-0_10.
- Boura, E., B. Rózycki, ..., J. H. Hurley. 2011. Solution structure of the ESCRT-I complex by small-angle X-ray scattering, EPR, and FRET spectroscopy. *Proc. Natl. Acad. Sci. USA*. 108:9437–9442.
- Boura, E., B. Rózycki, ..., J. H. Hurley. 2012. Solution structure of the ESCRT-I and -II supercomplex: implications for membrane budding and scission. *Structure*. 20:874–886.
- Borgia, A., W. Zheng, ..., B. Schuler. 2016. Consistent view of polypeptide chain expansion in chemical denaturants from multiple experimental methods. *J. Am. Chem. Soc.* 138:11714–11726.
- Fuertes, G., N. Banterle, ..., E. A. Lemke. 2017. Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs. FRET measurements. *Proc. Natl. Acad. Sci. USA*. 114:E6342–E6351.
- Rout, M. P., and A. Sali. 2019. Principles for integrative structural biology studies. *Cell*. 177:1384–1403.
- Alonso-García, N., I. García-Rubio, ..., J. M. de Pereda. 2015. Combination of X-ray crystallography, SAXS and DEER to obtain the structure of the FnIII-3,4 domains of integrin $\alpha 6\beta 4$. *Acta Crystallogr. D Biol. Crystallogr.* 71:969–985.
- Bengtson, T., V. L. Holm, ..., K. Lindorff-Larsen. 2020. Structure and dynamics of a nanodisc by integrating NMR, SAXS and SANS experiments with molecular dynamics simulations. *eLife*. 9:e56518.
- Bottaro, S., T. Bengtson, and K. Lindorff-Larsen. 2020. Integrating molecular simulation and experimental data: a Bayesian/maximum entropy reweighting approach. *In Structural Bioinformatics, Methods in Molecular Biology Z. Gáspári, ed.* Springer, pp. 219–240. https://doi.org/10.1007/978-1-0716-0270-6_15.
- Gomes, G. W., M. Krzeminski, ..., C. C. Gradinaru. 2020. Conformational ensembles of an intrinsically disordered protein consistent with NMR, SAXS, and single-molecule FRET. *J. Am. Chem. Soc.* 142:15697–15710.
- Schiemann, O., and T. F. Prisner. 2007. Long-range distance determinations in biomacromolecules by EPR spectroscopy. *Q. Rev. Biophys.* 40:1–53.
- Czar, M. F., F. D. Breitgoff, ..., R. Zenobi. 2019. Linear and kinked oligo(phenyleneethynylene)s as ideal molecular calibrants for Förster resonance energy transfer. *J. Phys. Chem. Lett.* 10:6942–6947.
- Vöpel, T., C. S. Hengstenberg, ..., J. P. Klare. 2014. Triphosphate induced dimerization of human guanylate binding protein 1 involves association of the C-terminal helices: a joint double electron-electron resonance and FRET study. *Biochemistry*. 53:4590–4600.
- Klose, D., J. P. Klare, ..., H.-J. Steinhoff. 2012. Simulation vs. reality: a comparison of in silico distance predictions with DEER and FRET measurements. *PLoS One*. 7:e39492.
- Grohmann, D., D. Klose, ..., F. Werner. 2010. RNA-binding to archaeal RNA polymerase subunits F/E: a DEER and FRET study. *J. Am. Chem. Soc.* 132:5954–5955.
- Peter, M. F., C. Gebhardt, ..., G. Hagelueken. 2020. Cross-validation of distance measurements in proteins by PELDOR/DEER and single-molecule FRET. *bioRxiv* <https://doi.org/10.1101/2020.11.23.394080>.
- Jeschke, G. 2012. DEER distance measurements on proteins. *Annu. Rev. Phys. Chem.* 63:419–446.
- Schmidt, T., M. A. Wälti, ..., G. M. Clore. 2016. Long distance measurements up to 160 Å in the GroEL tetradecamer using Q-band DEER EPR spectroscopy. *Angew. Chem. Int.Engl.* 55:15905–15909.
- Jeschke, G. 2018. The contribution of modern EPR to structural biology. *Emerg. Top. Life Sci.* 2:9–18.
- Milov, A. D., A. B. Ponomarev, and Y. D. Tsvetkov. 1984. Electron-electron double resonance in electron spin echo: model biradical systems and the sensitized photolysis of decalin. *Chem. Phys. Lett.* 110:67–72.
- Pannier, M., S. Veit, ..., H. W. Spiess. 2000. Dead-time free measurement of dipole-dipole interactions between electron spins. *J. Magn. Reson.* 142:331–340.

31. Kulik, L. V., S. A. Dzuba, ..., Y. D. Tsvetkov. 2001. Electron dipole-dipole interaction in ESEEM of nitroxide biradicals. *Chem. Phys. Lett.* 343:315–324.
32. Borbat, P. P., and J. H. Freed. 1999. Multiple-quantum ESR and distance measurements. *Chem. Phys. Lett.* 313:145–154.
33. Jeschke, G., M. Pannier, ..., H. W. Spiess. 2000. Dipolar spectroscopy and spin alignment in electron paramagnetic resonance. *Chem. Phys. Lett.* 331:243–252.
34. Denysenkov, V. P., T. F. Prisner, ..., M. Bennati. 2006. High-field pulsed electron-electron double resonance spectroscopy to determine the orientation of the tyrosyl radicals in ribonucleotide reductase. *Proc. Natl. Acad. Sci. USA.* 103:13386–13390.
35. Polyhach, Y., A. Godt, ..., G. Jeschke. 2007. Spin pair geometry revealed by high-field DEER in the presence of conformational distributions. *J. Magn. Reson.* 185:118–129.
36. J. R. Lakowicz, ed 2006. Principles of Fluorescence Spectroscopy, Third Edition Springer US, New York <https://doi.org/10.1007/978-0-387-46312-4>.
37. Dimura, M., T. O. Peulen, ..., C. A. M. Seidel. 2016. Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems. *Curr. Opin. Struct. Biol.* 40:163–185.
38. Algar, W. R., N. Hildebrandt, ..., I. L. Medintz. 2019. FRET as a biomolecular research tool - understanding its potential while avoiding pitfalls. *Nat. Methods.* 16:815–829.
39. Förster, T. 1948. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys.* 437:55–75.
40. Schuler, B., and H. Hofmann. 2013. Single-molecule spectroscopy of protein folding dynamics—expanding scope and timescales. *Curr. Opin. Struct. Biol.* 23:36–47.
41. Lerner, E., T. Cordes, ..., S. Weiss. 2018. Toward dynamic structural biology: two decades of single-molecule Förster resonance energy transfer. *Science.* 359:eaan1133.
42. Schuler, B., E. A. Lipman, ..., W. A. Eaton. 2005. Polyproline and the “spectroscopic ruler” revisited with single-molecule fluorescence. *Proc. Natl. Acad. Sci. USA.* 102:2754–2759.
43. Ingarciola, A., S. Weiss, and E. Lerner. 2018. Monte Carlo diffusion-enhanced photon inference: distance distributions and conformational dynamics in single-molecule FRET. *J. Phys. Chem. B.* 122:11598–11615.
44. Peulen, T.-O., O. Opanasyuk, and C. A. M. Seidel. 2017. Combining graphical and analytical methods with molecular simulations to analyze time-resolved FRET measurements of labeled macromolecules accurately. *J. Phys. Chem. B.* 121:8211–8241.
45. Holmstrom, E. D., A. Holla, ..., B. Schuler. 2018. Accurate transfer efficiencies, distance distributions, and ensembles of unfolded and intrinsically disordered proteins from single-molecule FRET. *In Methods in Enzymology* E. Rhoades, ed.. Elsevier, pp. 287–325. <https://doi.org/10.1016/bs.mie.2018.09.030>.
46. Sezer, D., J. H. Freed, and B. Roux. 2008. Parametrization, molecular dynamics simulation, and calculation of electron spin resonance spectra of a nitroxide spin label on a polyaniline α -helix. *J. Phys. Chem. B.* 112:5755–5767.
47. Sezer, D., J. H. Freed, and B. Roux. 2009. Multifrequency electron spin resonance spectra of a spin-labeled protein calculated from molecular dynamics simulations. *J. Am. Chem. Soc.* 131:2597–2605.
48. Beier, C., and H.-J. Steinhoff. 2006. A structure-based simulation approach for electron paramagnetic resonance spectra using molecular and stochastic dynamics simulations. *Biophys. J.* 91:2647–2664.
49. Hilger, D., Y. Polyhach, ..., G. Jeschke. 2007. High-resolution structure of a Na⁺/H⁺ antiporter dimer obtained by pulsed electron paramagnetic resonance distance measurements. *Biophys. J.* 93:3675–3683.
50. Polyhach, Y., E. Bordignon, and G. Jeschke. 2011. Rotamer libraries of spin labelled cysteines for protein studies. *Phys. Chem. Chem. Phys.* 13:2356–2366.
51. Hagelueken, G., R. Ward, ..., O. Schiemann. 2012. MtsslWizard: in silico spin-labeling and generation of distance distributions in PyMOL. *Appl. Magn. Reson.* 42:377–391.
52. Beasley, K. N., B. T. Sutch, ..., I. S. Haworth. 2015. Computer modeling of spin labels: NASNOX, PRONOX, and ALLNOX. *In Methods in Enzymology* P. Z. Qin and K. Warncke, eds.. Elsevier, pp. 569–593. <https://doi.org/10.1016/bs.mie.2015.07.021>.
53. Jeschke, G. 2013. A comparative study of structures and structural transitions of secondary transporters with the LeuT fold. *Eur. Biophys. J.* 42:181–197.
54. Sippach, M., D. Weidlich, ..., H.-J. Steinhoff. 2014. Conformational changes of the histidine ATP-binding cassette transporter studied by double electron–electron resonance spectroscopy. *Biochim. Biophys. Acta.* 1838:1760–1768.
55. Tesei, G., J. M. Martins, ..., K. Lindorff-Larsen. 2021. DEER-PREDict: software for efficient calculation of spin-labeling EPR and NMR data from conformational ensembles. *PLoS Comput. Biol.* 17:e1008551.
56. Jeschke, G. 2013. Conformational dynamics and distribution of nitroxide spin labels. *Prog. Nucl. Magn. Reson. Spectrosc.* 72:42–60.
57. Jeschke, G. 2018. MMM: a toolbox for integrative structure modeling. *Protein Sci.* 27:76–85.
58. Kalinin, S., T. Peulen, ..., C. A. M. Seidel. 2012. A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat. Methods.* 9:1218–1225.
59. Sindbert, S., S. Kalinin, ..., C. A. M. Seidel. 2011. Accurate distance determination of nucleic acids via Förster resonance energy transfer: implications of dye linker length and rigidity. *J. Am. Chem. Soc.* 133:2463–2480.
60. Muschielok, A., J. Andrecka, ..., J. Michaelis. 2008. A nano-positioning system for macromolecular structural analysis. *Nat. Methods.* 5:965–971.
61. Beckers, M., F. Drechsler, ..., J. Michaelis. 2015. Quantitative structural information from single-molecule FRET. *Faraday Discuss.* 184:117–129.
62. Metskas, L. A., and E. Rhoades. 2015. Conformation and dynamics of the troponin I C-terminal domain: combining single-molecule and computational approaches for a disordered protein region. *J. Am. Chem. Soc.* 137:11962–11969.
63. Walczewska-Szewc, K., and B. Corry. 2014. Accounting for dye diffusion and orientation when relating FRET measurements to distances: three simple computational methods. *Phys. Chem. Chem. Phys.* 16:12317–12326.
64. Warner, J. B., IV, K. M. Ruff, ..., H. A. Lashuel. 2017. Monomeric Huntingtin exon 1 has similar overall structural features for wild-type and pathological polyglutamine lengths. *J. Am. Chem. Soc.* 139:14456–14469.
65. Grotz, K. K., M. F. Nueesch, ..., G. Hummer. 2018. Dispersion correction alleviates dye stacking of single-stranded DNA and RNA in simulations of single-molecule fluorescence experiments. *J. Phys. Chem. B.* 122:11626–11639.
66. Merchant, K. A., R. B. Best, ..., W. A. Eaton. 2007. Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc. Natl. Acad. Sci. USA.* 104:1528–1533.
67. Wozniak, A. K., G. F. Schröder, ..., F. Oesterheld. 2008. Single-molecule FRET measures bends and kinks in DNA. *Proc. Natl. Acad. Sci. USA.* 105:18337–18342.
68. Gabba, M., S. Poblete, ..., J. Fitter. 2014. Conformational state distributions and catalytically relevant dynamics of a hinge-bending enzyme studied by single-molecule FRET and a coarse-grained simulation. *Biophys. J.* 107:1913–1923.
69. Walczewska-Szewc, K., E. Deplazes, and B. Corry. 2015. Comparing the ability of enhanced sampling molecular dynamics methods to reproduce the behavior of fluorescent labels on proteins. *J. Chem. Theory Comput.* 11:3455–3465.

70. Reinartz, I., C. Sinner, ..., A. Schug. 2018. Simulation of FRET dyes allows quantitative comparison against experimental data. *J. Chem. Phys.* 148:123321.
71. Hoefling, M., N. Lima, ..., H. Grubmüller. 2011. Structural heterogeneity and quantitative FRET efficiency distributions of polyprolines through a hybrid atomistic simulation and Monte Carlo approach. *PLoS One.* 6:e19791.
72. Best, R. B., H. Hofmann, ..., B. Schuler. 2015. Quantitative interpretation of FRET experiments via molecular simulation: force field and validation. *Biophys. J.* 108:2721–2731.
73. Vitali, F., A. Henning, ..., F. H.-T. Allain. 2006. Structure of the two most C-terminal RNA recognition motifs of PTB using segmental isotope labeling. *EMBO J.* 25:150–162.
74. Oberstrass, F. C., S. D. Auweter, ..., F. H. Allain. 2005. Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science.* 309:2054–2057.
75. Gmeiner, C., G. Dorn, ..., M. Yulikov. 2017. Spin labelling for integrative structure modelling: a case study of the polypyrimidine-tract binding protein 1 domains in complexes with short RNAs. *Phys. Chem. Chem. Phys.* 19:28360–28380.
76. Zosel, F., A. Holla, and B. Schuler. 2020. Labeling of proteins for single-molecule fluorescence spectroscopy. *chemRxiv* <https://doi.org/10.26434/chemrxiv.11537913.v1>.
77. Gromov, I., J. Shane, ..., A. Schweiger. 2001. A Q-band pulse EPR/ENDOR spectrometer and the implementation of advanced one- and two-dimensional pulse EPR methodology. *J. Magn. Reson.* 149:196–203.
78. Polyhach, Y., E. Bordignon, ..., G. Jeschke. 2012. High sensitivity and versatility of the DEER experiment on nitroxide radical pairs at Q-band frequencies. *Phys. Chem. Chem. Phys.* 14:10762–10773.
79. Jeschke, G., V. Chechik, ..., H. Jung. 2006. DeerAnalysis2006—a comprehensive software package for analyzing pulsed ELDOR data. *Appl. Magn. Reson.* 30:473–498.
80. Jeschke, G., G. Panek, ..., H. Paulsen. 2004. Data analysis procedures for pulse ELDOR measurements of broad distance distributions. *Appl. Magn. Reson.* 26:223–244.
81. Edwards, T. H., and S. Stoll. 2018. Optimal Tikhonov regularization for DEER spectroscopy. *J. Magn. Reson.* 288:58–68.
82. Chiang, Y.-W., P. P. Borbat, and J. H. Freed. 2005. The determination of pair distance distributions by pulsed ESR using Tikhonov regularization. *J. Magn. Reson.* 172:279–295.
83. Dingfelder, F., B. Wunderlich, ..., B. Schuler. 2017. Rapid microfluidic double-jump mixing device for single-molecule spectroscopy. *J. Am. Chem. Soc.* 139:6062–6065.
84. Müller, B. K., E. Zaychikov, ..., D. C. Lamb. 2005. Pulsed interleaved excitation. *Biophys. J.* 89:3508–3522.
85. Lee, N. K., A. N. Kapanidis, ..., S. Weiss. 2005. Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. *Biophys. J.* 88:2939–2953.
86. Hellenkamp, B., S. Schmid, ..., T. Hugel. 2018. Precision and accuracy of single-molecule FRET measurements—a multi-laboratory benchmark study. *Nat. Methods.* 15:669–676.
87. Hellenkamp, B., P. Wortmann, ..., T. Hugel. 2017. Multidomain structure and correlated dynamics determined by self-consistent FRET networks. *Nat. Methods.* 14:174–180.
88. Haenni, D., F. Zosel, ..., B. Schuler. 2013. Intramolecular distances and dynamics from the combined photon statistics of single-molecule FRET and photoinduced electron transfer. *J. Phys. Chem. B.* 117:13015–13028.
89. Hanwell, M. D., D. E. Curtis, ..., G. R. Hutchison. 2012. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* 4:17.
90. Rappe, A. K., C. J. Casewit, ..., W. M. Skiff. 1992. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* 114:10024–10035.
91. Neese, F. 2012. The ORCA program system. *WIREs Comput. Mol. Sci.* 2:73–78.
92. Becke, A. D. 1988. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A Gen. Phys.* 38:3098–3100.
93. Perdew, J. P. 1986. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B Condens. Matter.* 33:8822–8824.
94. Weigend, F., and R. Ahlrichs. 2005. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: design and assessment of accuracy. *Phys. Chem. Chem. Phys.* 7:3297–3305.
95. Whitten, J. L. 1973. Coulombic potential energy integrals and approximations. *J. Chem. Phys.* 58:4496–4501.
96. Weigend, F. 2002. A fully direct RI-HF algorithm: implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.* 4:4285–4291.
97. Klamt, A., and G. Schüürmann. 1993. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perk. T. 2.* 2:799–805.
98. Sinnecker, S., A. Rajendran, ..., F. Neese. 2006. Calculation of solvent shifts on electronic g-tensors with the conductor-like screening model (COSMO) and its self-consistent generalization to real solvents (direct COSMO-RS). *J. Phys. Chem. A.* 110:2235–2245.
99. Metropolis, N., A. W. Rosenbluth, ..., E. Teller. 1953. Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21:1087–1092.
100. Martin, M. G., A. P. Thompson, and T. M. Nenoff. 2001. Effect of pressure, membrane thickness, and placement of control volumes on the flux of methane through thin silicalite membranes: a dual control volume grand canonical molecular dynamics study. *J. Chem. Phys.* 114:7174–7181.
101. Krivov, G. G., M. V. Shapovalov, and R. L. Dunbrack, Jr. 2009. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins.* 77:778–795.
102. Dingfelder, F., S. Benke, ..., B. Schuler. 2018. Mapping an equilibrium folding intermediate of the cytolytic pore toxin ClyA with single-molecule FRET. *J. Phys. Chem. B.* 122:11251–11261.
103. Gopich, I. V., and A. Szabo. 2012. Theory of the energy transfer efficiency and fluorescence lifetime distribution in single-molecule FRET. *Proc. Natl. Acad. Sci. USA.* 109:7747–7752.
104. Gopich, I. V., D. Nettels, ..., A. Szabo. 2009. Protein dynamics from single-molecule fluorescence intensity correlation functions. *J. Chem. Phys.* 131:095102.
105. Ranjit, S., and M. Levitus. 2012. Probing the interaction between fluorophores and DNA nucleotides by fluorescence correlation spectroscopy and fluorescence quenching. *Photochem. Photobiol.* 88:782–791.
106. Stennett, E. M. S., M. A. Ciuba, and M. Levitus. 2014. Photophysical processes in single molecule organic fluorescent probes. *Chem. Soc. Rev.* 43:1057–1075.
107. Mao, F., W.-Y. Leung, ..., H. E. Hoover. 2012. Fluorescent dyes, fluorescent dye kits, and methods of preparing labelled molecules. US Patent US2012/0329068 A1.
108. Jeschke, G., M. Sajid, ..., A. Godt. 2010. Flexibility of shape-persistent molecular building blocks composed of p-phenylene and ethynylene units. *J. Am. Chem. Soc.* 132:10107–10117.
109. Gopich, I., and A. Szabo. 2005. Theory of photon statistics in single-molecule Förster resonance energy transfer. *J. Chem. Phys.* 122:14707.
110. Nir, E., X. Michalet, ..., S. Weiss. 2006. Shot-noise limited single-molecule FRET histograms: comparison between theory and experiments. *J. Phys. Chem. B.* 110:22103–22124.
111. Kalinin, S., S. Felekyan, ..., C. A. M. Seidel. 2007. Probability distribution analysis of single-molecule fluorescence anisotropy and resonance energy transfer. *J. Phys. Chem. B.* 111:10253–10262.

112. Eilert, T., E. Kallis, ..., J. Michaelis. 2018. Complete kinetic theory of FRET. *J. Phys. Chem. B.* 122:11677–11694.
113. Duss, O., E. Michel, ..., F. H.-T. Allain. 2014. Structural basis of the non-coding RNA RsmZ acting as a protein sponge. *Nature.* 509:588–592.
114. Duss, O., M. Yulikov, ..., G. Jeschke. 2015. Combining NMR and EPR to determine structures of large RNAs and protein–RNA complexes in solution. *In Methods in Enzymology* S. A. Woodson and F. H. T. Allain, eds.. Elsevier, pp. 279–331.
115. Bleicken, S., G. Jeschke, ..., E. Bordignon. 2014. Structural model of active Bax at the membrane. *Mol. Cell.* 56:496–505.
116. Puljung, M. C., H. A. DeBerg, ..., S. Stoll. 2014. Double electron-resonance reveals cAMP-induced conformational change in HCN channels. *Proc. Natl. Acad. Sci. USA.* 111:9816–9821.
117. Jeschke, G. 2016. Ensemble models of proteins and protein domains based on distance distribution restraints. *Proteins.* 84:544–560.
118. Masliah, G., C. Maris, ..., F. H.-T. Allain. 2018. Structural basis of siRNA recognition by TRBP double-stranded RNA binding domains. *EMBO J.* 37:e97089.
119. Lawless, M. J., J. R. Pettersson, ..., S. Saxena. 2018. ESR resolves the C terminus structure of the ligand-free human glutathione S-transferase A1-1. *Biophys. J.* 114:592–601.
120. Kachooei, E., N. M. Cordina, and L. J. Brown. 2019. Constructing a structural model of troponin using site-directed spin labeling: EPR and PRE-NMR. *Biophys. Rev.* 11:621–639.
121. Burger, M., S. Rein, ..., S. Kacprzak. 2020. Distance measurements in the F_0F_1 -ATP synthase from *E. coli* using smFRET and PELDOR spectroscopy. *Eur. Biophys. J.* 49:1–10.

Biophysical Journal, Volume 120

Supplemental information

Resolving distance variations by single-molecule FRET and EPR spectroscopy using rotamer libraries

Daniel Klose, Andrea Holla, Christoph Gmeiner, Daniel Nettels, Irina Ritsch, Nadja Bross, Maxim Yulikov, Frédéric H.-T. Allain, Benjamin Schuler, and Gunnar Jeschke

Supporting Information: Resolving small distance variations by single-molecule FRET & EPR spectroscopy using rotamer libraries

Daniel Klose^{1,*}, Andrea Holla², Christoph Gmeiner¹, Daniel Nettels², Irina Ritsch¹, Nadja Bross⁵, Maxim Yulikov¹, Frédéric H.-T. Allain³, Benjamin Schuler^{2,4,*}, and Gunnar Jeschke¹

¹Department of Chemistry and Applied Biosciences, ETH Zurich, Zurich, Switzerland

²Department of Biochemistry, University of Zurich, Zurich, Switzerland

³Institute of Biochemistry, ETH Zurich, Zurich, Switzerland

⁴Department of Physics, University of Zurich, Zurich, Switzerland

⁵Department of Chemistry, University of Zurich, Zurich, Switzerland

*Correspondence: daniel.klose@phys.chem.ethz.ch or schuler@bioc.uzh.ch

SI METHODS

Structure of CF660R maleimide

The chemical structure of CF660R maleimide (Fig. 1) is based on patent US2012/0329068A1 (Biotium, Inc., Hayward, CA, US) (107). Its mass (1010.18 Da) is within experimental error of the value obtained by mass spectrometry (1011.34 Da). The structure was further confirmed by NMR spectroscopy based on the following experiments: ¹H; ¹³C{¹H}; ¹H,¹H-DQF-COSY; ¹H,¹H-TOCSY; ¹H,¹H-ROESY; ¹H,¹³C-HSQC; ¹H,¹³C-HSQC-edited and ¹H,¹³C-HMBC.

RLA simulations of fluorescence labels

The general workflow for the rotamer analysis and calculations of distance distributions is illustrated in Fig. S5 (bottom half) and proceeds in analogy to spin labeling described in the manual of MMM. The steps used here are the following for the two fluorescence labels Cy3b and CF660R that are abbreviated in MMM (Version 2021.1, available as an open source Matlab program on www.epr.ethz.ch and on [GitHub.com/gjeschke](https://github.com/gjeschke) (incl. source code to generate rotamer libraries)) by the three-letter codes Cy3 and CF6.

- Load a structure in PDB format into MMM by *File* → *New from PDB/local*.
- Select residues to be labeled by label 1 by *Display* → *Hierarchy*.
- Calculate rotamer populations for label 1 by *FRET* → *Chromophore site scan*. Leave default options in *Site scan setup*. In *Set labeling conditions* select the chromophore library for label 1, here either *Cy3b_2048* or *CF660R_2048*. This selects the default libraries with 2048 rotamers. The rotamer libraries are provided in sizes of $N_{rot} = 1024, 2048, 4096$ and 8192 (downloaded automatically upon first use). After confirmation the rotamer search for label 1 is performed.
- Select residues to be labeled by label 2 by *Display* → *Hierarchy*.
- Calculate rotamer populations for label 2 by *FRET* → *Chromophore site scan* as above.
- Calculate distance distributions by *FRET* → *Distance distribution/FRET efficiency*. Select two labels one by one and click "+" after each. The distance distribution between the two selected labels will be calculated and shown automatically.
- After having selected two labels the distance distribution and static FRET efficiencies will be calculated. Diffusional averaging with different parameters can be turned on in the bottom left panel.
- The results, i.e. the distance distribution and the FRET efficiencies, can be saved in ascii format using the *Save* button.

In MMM there is a distinction of first calculating rotamer populations and later (optional) attachment, which allows rotamer populations to be calculated for multiple labels at close distances without leading to clashes between the labels. Hence only after attachment, labels are "present" in the structure and therefore considered in clash tests. Here this attachment feature is not used as chromophore pairs are far enough apart not to overlap, however for visualization of the label distributions on the protein structure prior attachment of the labels is required.

AV simulations

AV simulations were carried out using the program FPS 1.1 (58) from the group of Prof. Claus Seidel (University of Düsseldorf, Germany). The required parameters for each label are the linker length l measured from C_β to the center of the chromophore and linker width w as well as three radii that describe the chromophore as an ellipsoid. The parameters were chosen according to the manual of the software using DFT-optimized structures with elongated linkers. For maleimido-Cy3b, this lead to $l = 18.5$ Å, as well as 3.4, 8.2 and 3.0 Å for the three radii, respectively. For maleimido-CF660R, we derived $l = 17.6$ Å, as well as 8.1, 4.2 and 2.1 Å for the three radii, respectively. The linker width w was set to 4.5 Å in both cases, which is the same as for other maleimide-functionalized fluorescence labels.

Since in FPS only a single structure is processed from an input PDB file, we loaded the single RRM3/4-ΔN structures, where atoms beyond C_β (attachment point) of the side chains to be labeled had been removed. The resulting AVs for each label (in .xyz file format) were evaluated both via FPS as well as by custom-written Matlab scripts.

SI TABLES AND FIGURES

Table S1: **Summary of the distance distributions of spin labels.** Experimental values from DEER are colored in red. RLA simulations are given in black. r_{cog} is the center of gravity distance, r_{max} indicates the maximum of the distance distributions and FWHM is the full width at half maximum. λ is the DEER modulation depth and "Lab.eff." denotes the spin labeling efficiency according to cw EPR (gray).

Positions	EPR					
	r_{cog} ^a [nm]	r_{max} [nm]	FWHM [nm]	λ	Lab.eff. ^b	
388/468	4.97 5.21	5.06 5.25	0.87 0.89	0.42	67%	
388/472	4.71 4.86	4.79 4.95	1.14 0.79	0.26	66%	
388/475	4.43 4.62	4.50 4.70	0.62 0.83	0.37	78%	
392/468	4.35 4.71	4.48 5.05	1.06 1.11	0.29	62%	
392/472	4.33 4.46	4.50 4.65	0.90 1.15	0.35	100%	
392/475	3.89 4.11	3.92 4.50	0.60 1.07	0.45	77%	

^a Center of gravity of the upper 75% of the distributions to avoid influence from baseline.

^b Variations in labeling efficiency are partly also due to residual amounts of unbound spin label (see Fig. S1), which does not contribute to the modulation depth in DEER.

Table S2: **Summary of the distance distributions of fluorescence labels from RLA and AV simulations.** RLA simulations were conducted either using all 20 structures of the NMR ensemble (PDB: 2ADC) (shown in black) or using RRM3/4- Δ N (shown in gray). The strong similarity of the results shows that for distance distributions between fluorescence labels the entire ensemble is well-represented by the RRM3/4- Δ N structure. AV simulations carried out for RRM3/4- Δ N (shown in gray). r_{cog} is the center of gravity distance, r_{max} indicates the maximum of the distance distributions and FWHM is the full width at half maximum.

Positions +Labels	FRET - RLA			FRET - AV ^b			
	r_{cog} ^a [nm]	r_{max} [nm]	FWHM [nm]	r_{cog} ^a [nm]	r_{max} [nm]	FWHM [nm]	R_{mp} ^c [nm]
388-A/468-D	5.11 5.11	5.20 5.20	1.99 2.03	5.53	5.60	1.87	5.29
388-D/468-A	5.12 5.13	5.20 5.20	2.10 2.14	5.53	5.55	1.88	5.30
388-A/472-D	4.82 4.79	4.90 4.80	2.01 2.05	5.23	5.20	1.85	4.99
388-D/472-A	4.82 4.82	4.90 4.90	2.01 2.05	5.22	5.40	1.88	5.00
388-A/475-D	4.60 4.63	4.75 4.70	2.04 2.03	4.99	5.10	1.98	4.76
388-D/475-A	4.60 4.64	4.70 4.75	2.09 2.14	4.99	5.20	1.99	4.76
392-A/468-D	4.66 4.66	4.75 4.70	2.05 2.06	5.05	5.15	1.95	4.81
392-D/468-A	4.66 4.64	4.75 4.70	2.23 2.20	5.07	5.10	1.93	4.82
392-A/472-D	4.41 4.36	4.50 4.40	2.01 2.04	4.77	4.80	1.84	4.51
392-D/472-A	4.40 4.35	4.50 4.50	2.13 2.11	4.78	4.80	1.87	4.53
392-A/475-D	4.10 4.11	4.20 4.20	2.10 2.05	4.42	4.50	1.98	4.17
392-D/475-A	4.09 4.08	4.20 4.20	2.25 2.23	4.44	4.60	2.00	4.18

^a Center of gravity of the upper 75% of the distributions to avoid influence from baseline.

^b AV simulations carried out on RRM3/4- Δ N.

^c Midpoint distance between fluorophores from output of AV simulation program FPS.

Table S3: **Summary of FRET efficiencies from experiments and simulations.** Experimental mean transfer efficiencies $\langle E \rangle$ from smFRET, the standard deviation $\sigma(\langle E \rangle)$ from repeated experiments and stoichiometry $\langle S \rangle$. Simulated transfer efficiencies were calculated from distance distributions for both RLA and AV simulations by dynamical averaging (see methods) using an excited donor lifetime of 2.7 ns, a diffusion constant for the interdye distance of $D = 0.2 \text{ nm}^2/\text{ns}$ and a Förster radius of $R_0 = 6.0 \text{ nm}$ (FRET-RLA, FRET-AV). The standard deviation in these cases is obtained by variation of R_0 by $\pm 7\%$, an error estimate reported by a multi-laboratory comparison study (86), and D is varied in the interval $[0 \ 2D]$. For comparison, the transfer efficiency E from the AV simulations is given when evaluated with the FPS software (AV-FPS), which uses the static approximation of the Förster equation (58).

Positions +Labels	exp. smFRET			FRET - RLA ^a		FRET - AV ^a		AV - FPS ^a
	$\langle E \rangle$	$\sigma(\langle E \rangle)$	$\langle S \rangle$	$\langle E \rangle$	$\sigma(\langle E \rangle)$	$\langle E \rangle$	$\sigma(\langle E \rangle)$	E
388-A/468-D	0.745	0.008	0.524	0.728	0.044	0.643	0.052	0.614
388-D/468-A	0.780	0.007	0.492	0.724	0.044	0.643	0.052	0.612
388-A/472-D	0.788	0.006	0.517	0.789	0.037	0.709	0.047	0.680
388-D/472-A	0.824	0.007	0.485	0.785	0.038	0.708	0.047	0.678
388-A/475-D	0.784	0.005	0.522	0.815	0.034	0.754	0.042	0.727
388-D/475-A	0.836	0.006	0.484	0.814	0.034	0.753	0.042	0.726
392-A/468-D	0.792	0.004	0.520	0.808	0.034	0.741	0.043	0.714
392-D/468-A	0.829	0.003	0.496	0.811	0.034	0.738	0.043	0.711
392-A/472-D	0.831	0.005	0.517	0.855	0.028	0.796	0.037	0.771
392-D/472-A	0.880	0.004	0.497	0.857	0.027	0.793	0.037	0.770
392-A/475-D	0.831	0.004	0.520	0.886	0.023	0.848	0.029	0.829
392-D/475-A	0.867	0.002	0.485	0.889	0.022	0.845	0.029	0.826

^a Simulations carried out using RRM3/4- Δ N.

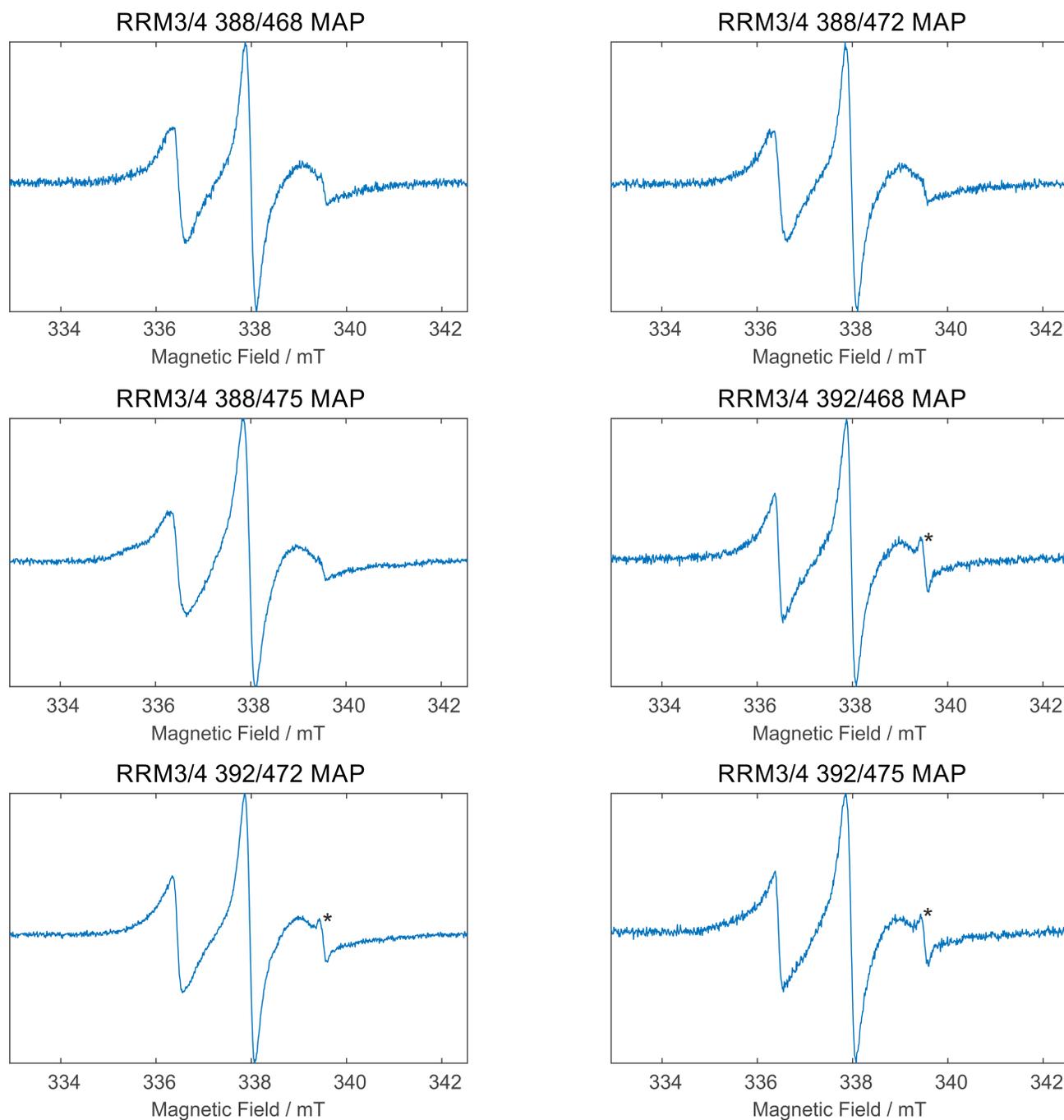


Figure S1: **Verifying spin labeling by cw EPR spectroscopy.** Cw EPR spectra at 9.5 GHz, room temperature, show the amount of bound MAF spin label on RRM3/4. The labeling efficiency determined by double integration is given in Tab. S1. (The asterisks mark putative contributions of residual free label.)

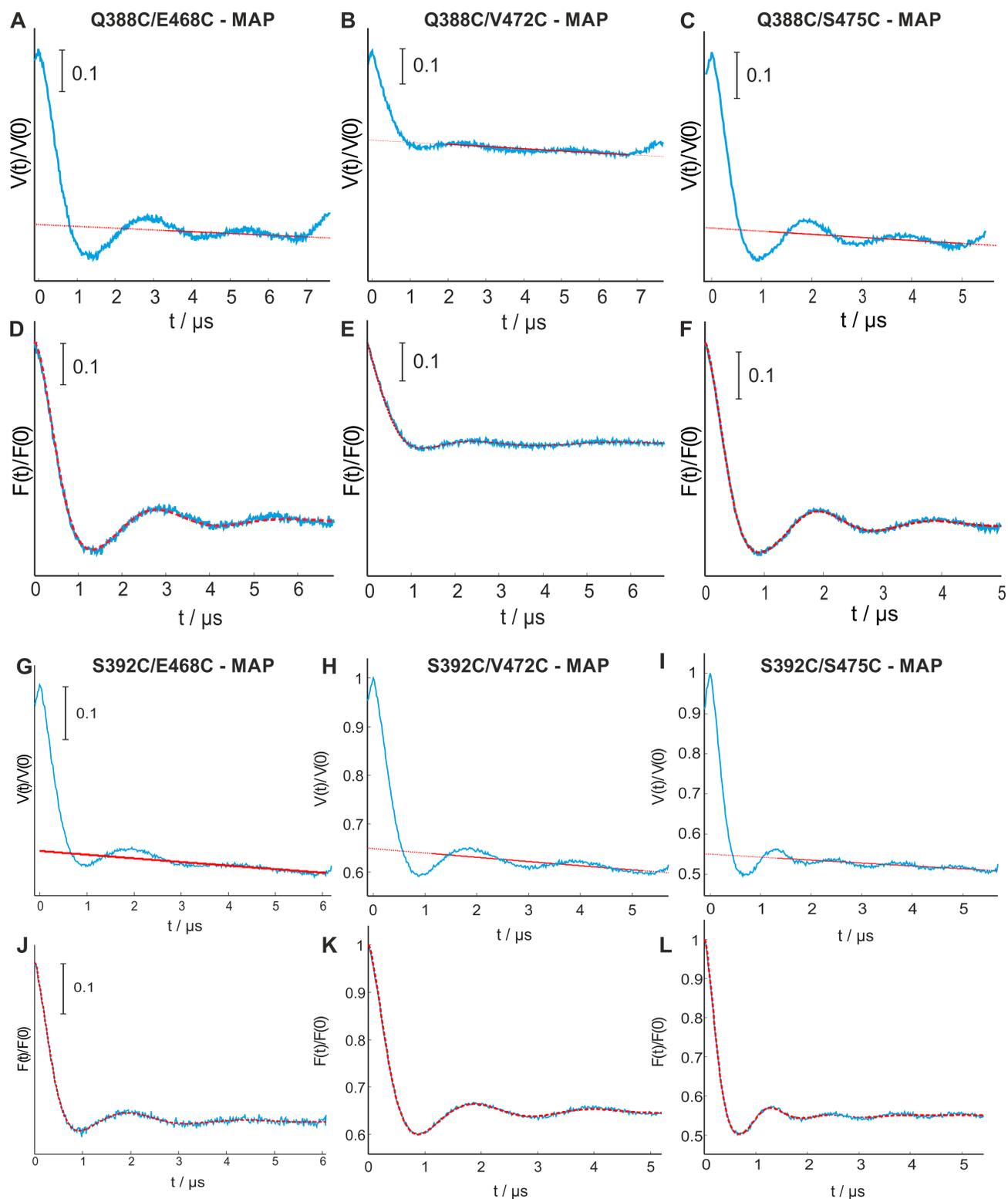


Figure S2: **EPR distance determination by DEER.** A-C & G-I Experimental DEER time traces (normalized) with 3D-homogeneous background fits (red). Cysteine positions in RRM3/4 spin-labeled with maleimido proxyl are given in the legend. D-E & J-L DEER form factors (time traces divided by background) with fits (red) that correspond to the distance distributions shown in Fig. 2A. Distance distributions were determined by Tikhonov regularization with regularization parameters determined by the L-curve criterion in DeerAnalysis2016.

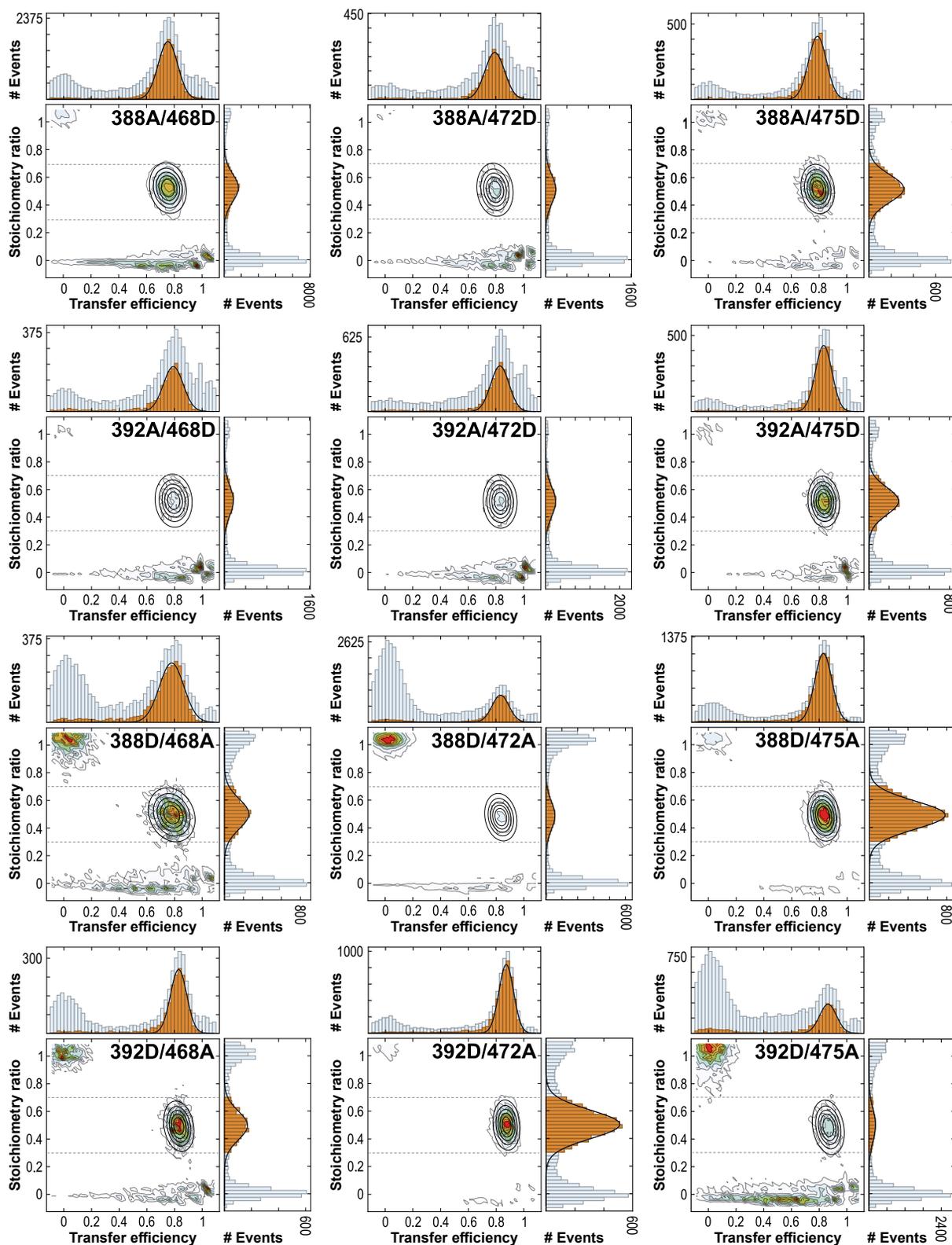


Figure S3: **Single-molecule FRET 2D histograms.** 2D histograms of transfer efficiencies (E) versus stoichiometry ratios (S) from single-molecule FRET. S allows donor-only and acceptor-only subpopulations (top left and bottom of each histogram, respectively) to be separated from donor-/acceptor-labeled sample ($0.3 < S < 0.7$). Labeling positions are shown in the histograms, with D for donor (Cy3b) and A for acceptor (CF660R). Projections along both dimensions show histograms of the full dataset (gray) and the stoichiometry-selected donor-acceptor labeled subpopulations (orange), with selection bounds shown as dashed lines in the 2D histograms.

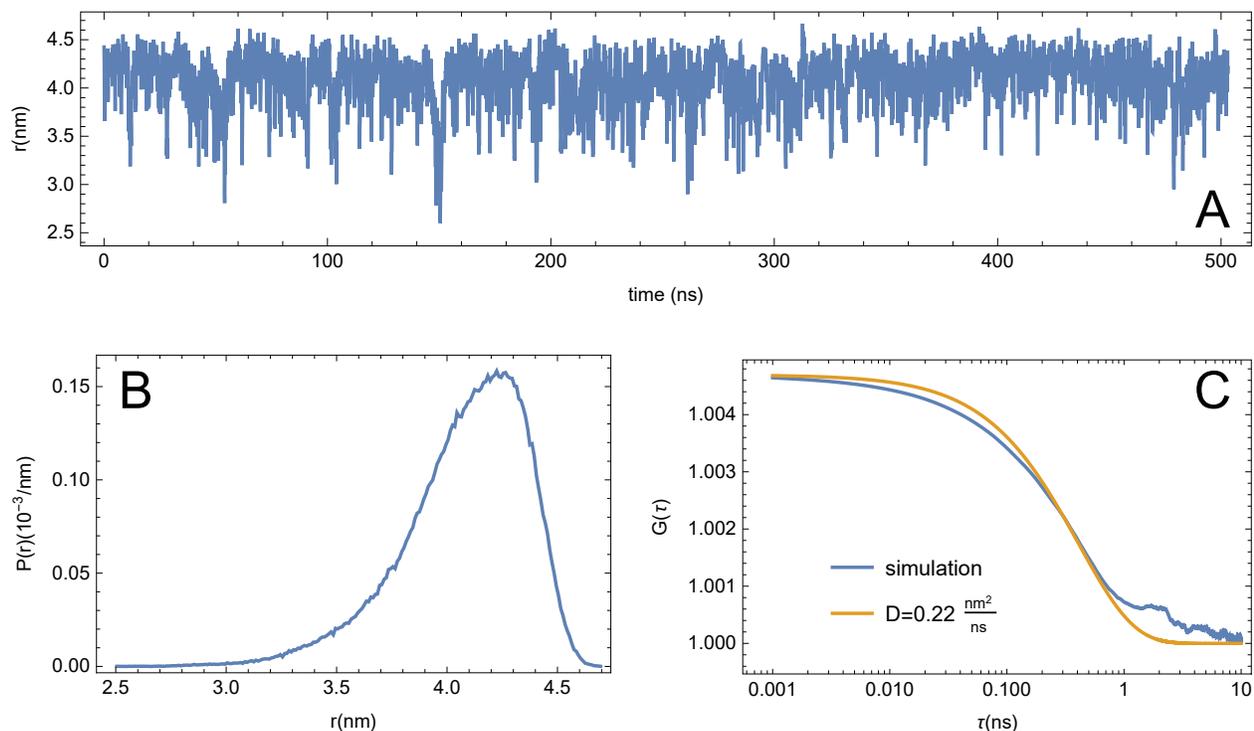


Figure S4: **Inter-dye distance time-trajectory** from an all-atom MD simulation of polyproline-11 labeled via an N-terminal Gly with Alexa594 and via a C-terminal Cys with Alexa488-C5-maleimide (72). **A** Fluctuation of the inter-dye distance r over the time course of 500 ns. **B** Distance distribution $P(r)$ obtained from the trajectory shown in A. **C** Time-correlation of $r(t)$ calculated from the trajectory shown in A (blue). The yellow correlation curve is calculated from Eq. 9 with $P(r)$ from B assuming an effective diffusion coefficient of $D = 0.22 \text{ nm}^2/\text{ns}$. A single diffusion coefficient is not sufficient for fully describing the correlation curve obtained from the simulation, which comprises a major fast component (~ 0.3 ns) and a minor slow component (~ 3 ns). However, both timescales are comparable to the lifetime of the excited state of the donor (4 ns in the absence of the acceptor).

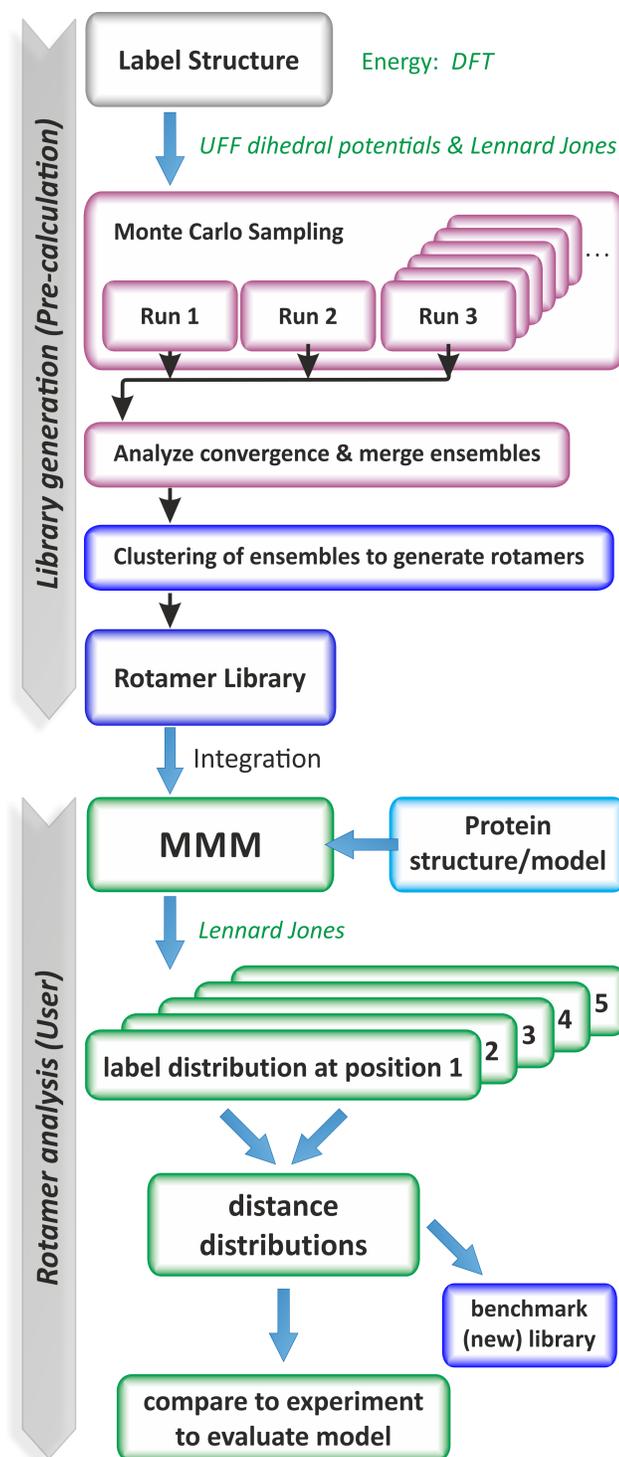


Figure S5: **Flowchart for rotamer library generation and rotamer analysis of a protein structure.** Generation of a rotamer library (*top half*) for a specific spin or fluorescence label is carried out in three stages (*color-coded*). First, a minimum-energy 3D structure of the label is obtained by DFT calculations. Second, the full label conformational space is sampled by Monte Carlo using rotatable dihedral angles as degrees of freedom. Third, clustering of the Monte Carlo ensemble leads to discrete rotamers. These stages are done once, and the resulting library is integrated into MMM for further use. Calculating distance distributions between labeling sites (*bottom half*). For each labeling site in a given protein structure or model, the conformational distribution of the label is calculated using the pre-calculated rotamer library. For any pair of labeling sites, a distance distribution is then calculated as a weighted histogram comprising all possible combinations of label conformations at the two sites. The distance distributions can be benchmarked against experimental data for a known system to validate new rotamer libraries (*as in this study for Cy3b & CF660R*) and then used with any structure for analysis and modeling.

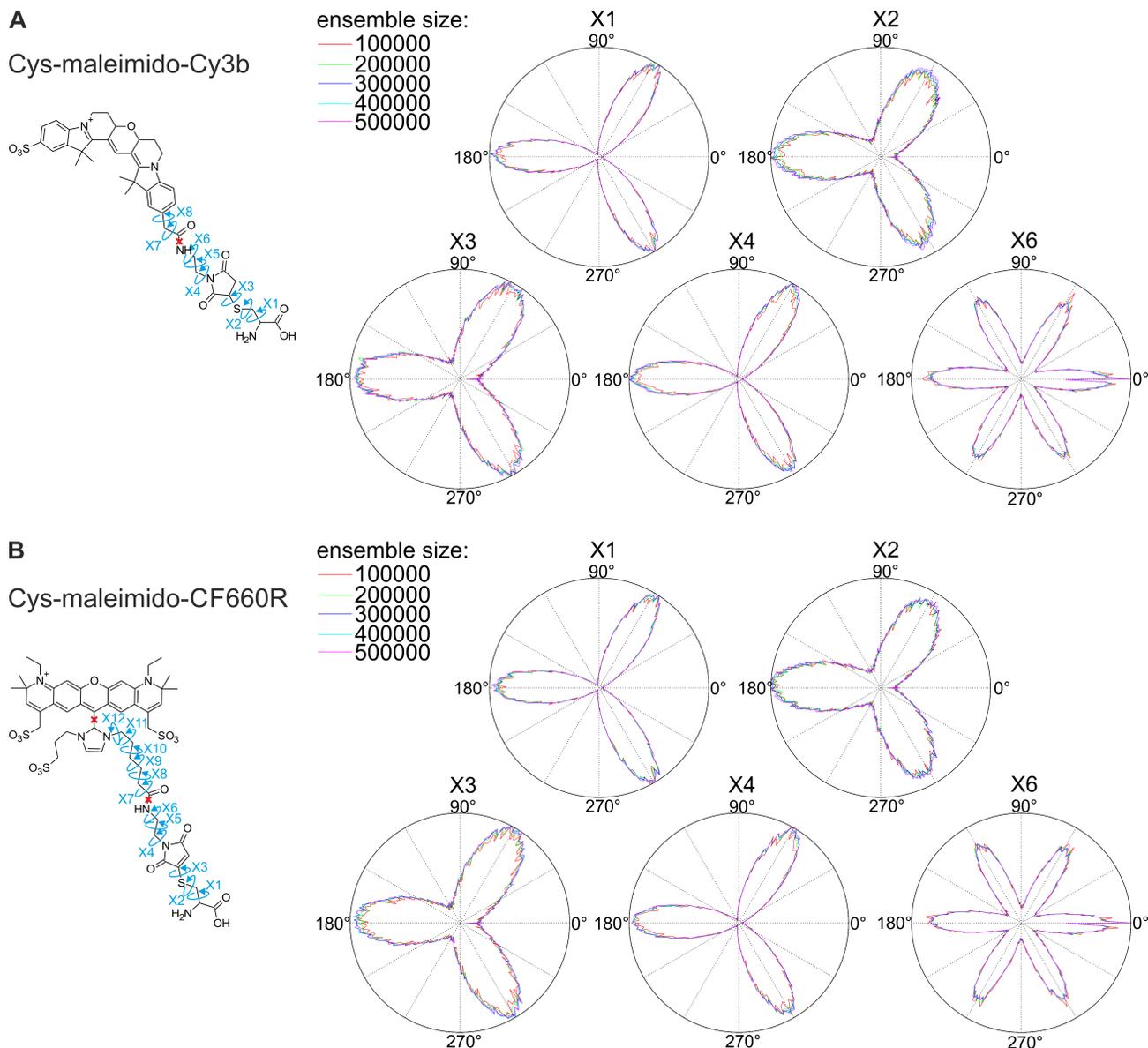


Figure S6: Dihedral angles and convergence of Monte Carlo Sampling. Definition of dihedral angle numbering (left) and dihedral angle histograms (right) for Monte Carlo ensembles of different sizes (see legend) for Cy3b and CF660R labeled side chains in **A** and **B**, respectively. The histograms show that the dihedral angle distributions still contain significant noise due to undersampling at an ensemble size of 100000, while there is barely any change observed in the histograms between ensemble sizes of 400000 and 500000, indicating the onset of convergence.

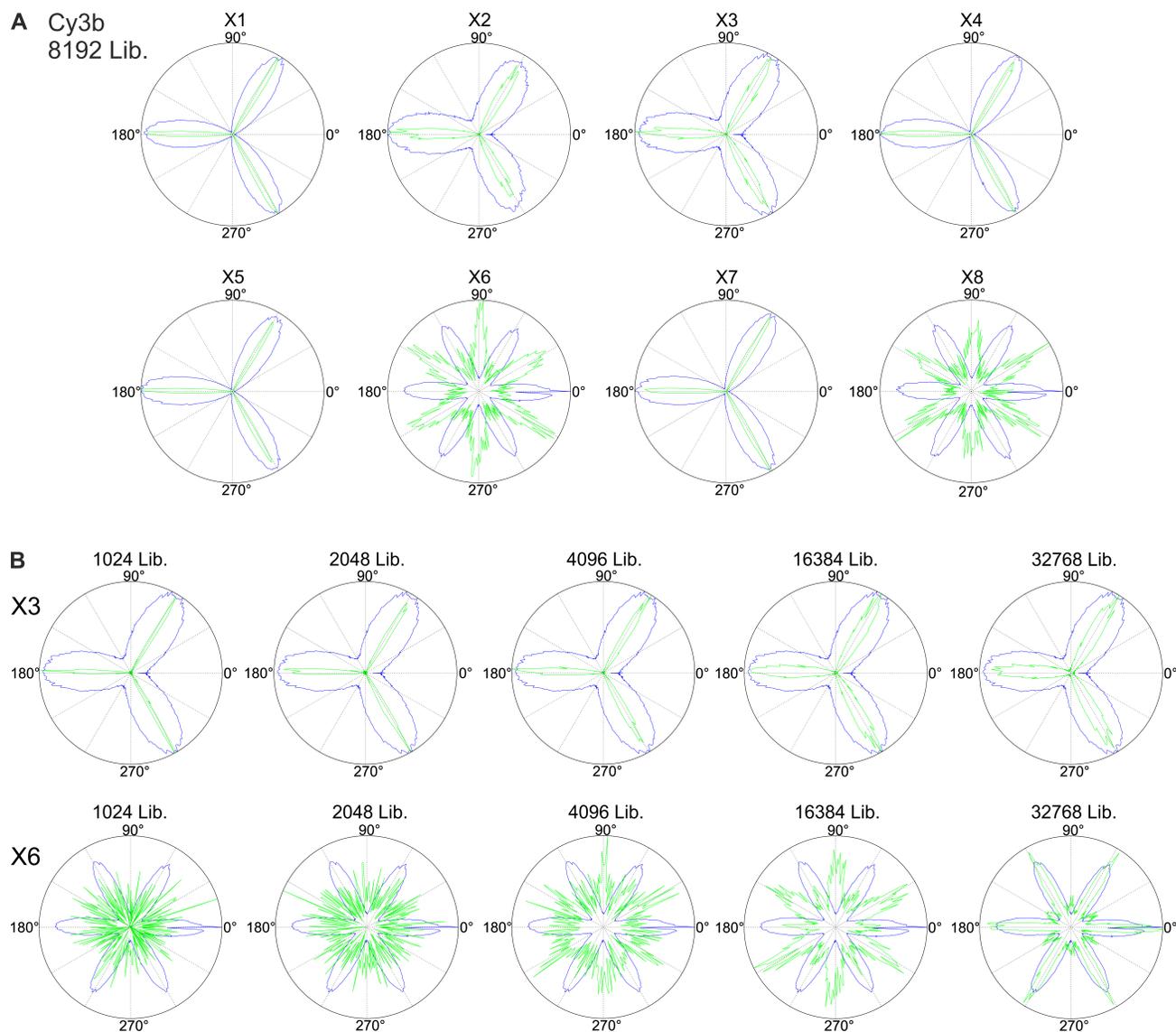


Figure S7: Rotamer library generation: Dihedral sampling of Cy3b. Dihedral angle histograms of Cy3b-labeled side chain (eight rotatable bonds with dihedrals X1-X8). The 500,000 member Monte Carlo ensemble is shown in blue, rotamer libraries in green. **(A)** Histograms of dihedral angles X1 to X8 for a rotamer library size of 8192. **(B)** Histograms of the two dihedral angles X3 and X6 (representative of the threefold and sixfold dihedrals, respectively) for increasing rotamer library sizes from 1024 (left) up to 32768 rotamers (right). The histograms show that dihedrals featuring three-fold rotation symmetry are well-represented in the rotamer libraries, albeit dihedrals are closer to the canonical values in the rotamer libraries versus the Monte Carlo ensemble. In contrast, for the two dihedrals X6 (and X8), strongly limited library sizes lead to averaging across different canonical dihedral angles, which leads to pronounced populations in between the canonical dihedral angles up to 8192 rotamers, while for the largest library, which is sufficiently sized to cover all canonical angle combinations for the eight dihedrals, this effect is almost absent. This effect was, however, found to have no significant influence on the resulting distance distributions calculated using different rotamer library sizes (see Fig. S10).

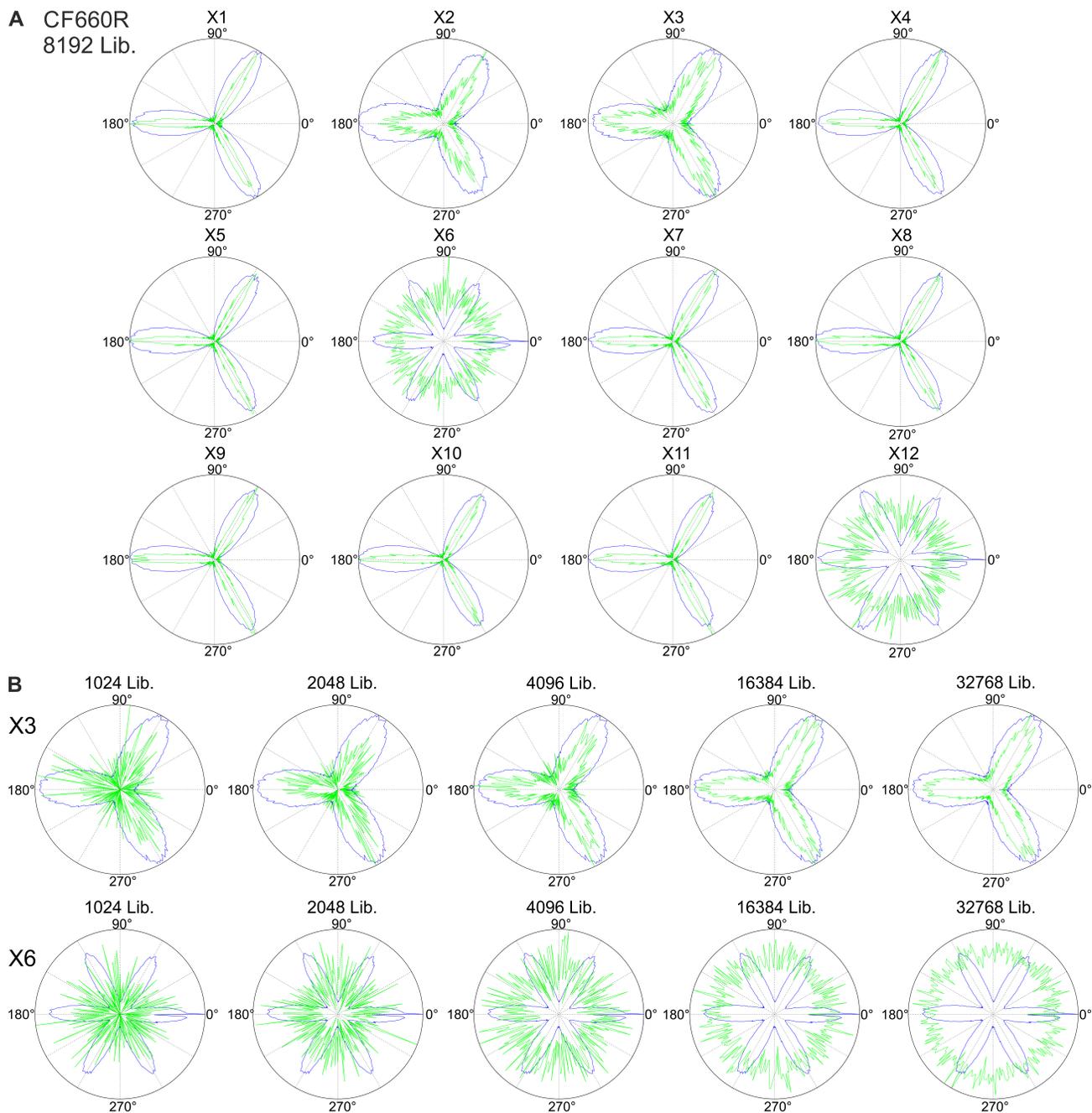


Figure S8: Rotamer library generation: Dihedral sampling of CF660R. Dihedral angle histograms of CF660R-labeled side chain (twelve rotatable bonds with dihedrals X1-X12). The 500,000 member Monte Carlo ensemble is shown in blue, rotamer libraries in green. **(A)** Histograms of dihedral angles X1 to X12 for a rotamer library size of 8192. **(B)** Histograms of the two dihedral angles X3 and X6 (representative of the threefold and sixfold dihedrals, respectively) for increasing rotamer library sizes from 1024 (left) up to 32768 rotamers (right). The histograms show that dihedrals featuring three-fold rotation symmetry are well-represented in the rotamer libraries, albeit dihedrals are closer to the canonical values in the rotamer libraries versus the Monte Carlo ensemble. In contrast, for the two dihedrals X6 (and X12), strongly limited library sizes lead to averaging across different canonical dihedral angles, which leads to pronounced populations in between the canonical dihedral angles even for the largest library of 32768 rotamers. This effect was, however, found to have no significant influence on the resulting distance distributions calculated using different rotamer library sizes (see Fig. S10).

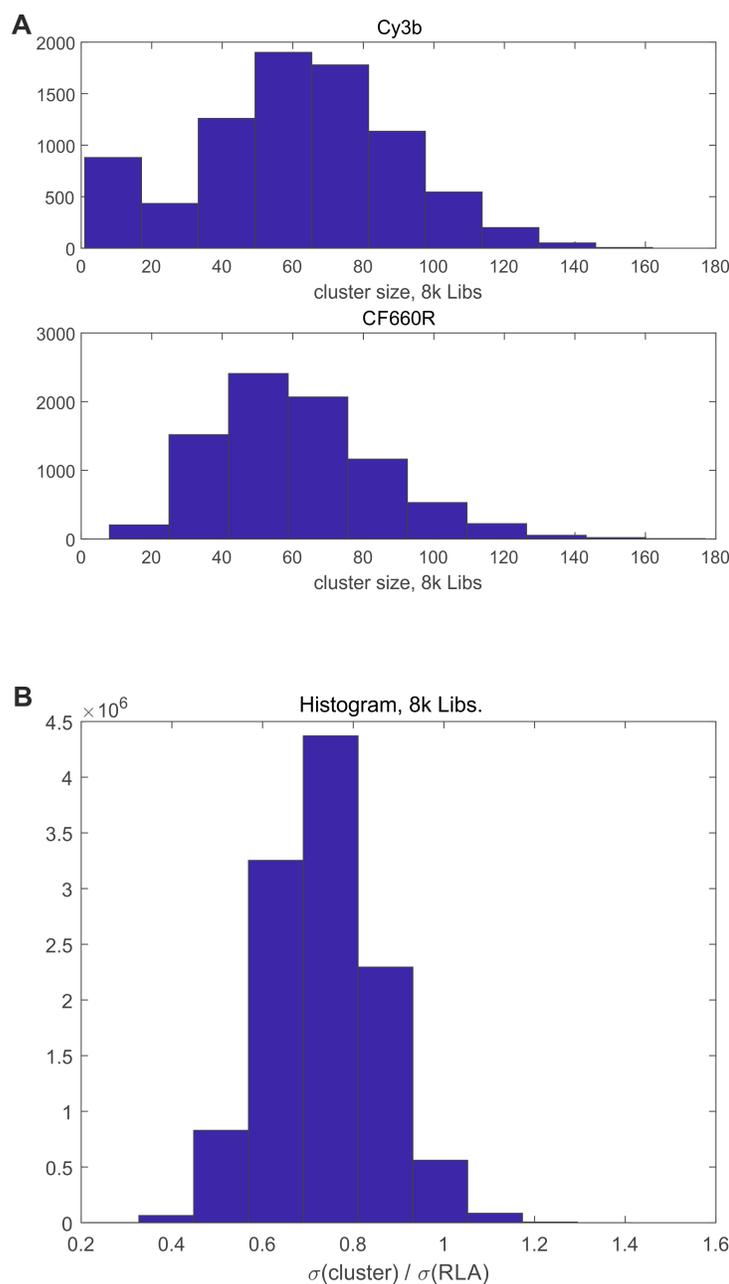


Figure S9: **Clustering test for the fluorescence label rotamer libraries.** **A** Size distributions of the clusters that give rise to one rotamer for a library of size of 8192 for Cy3b (top) and CF660R (bottom). **B** Distance distributions between pairs of clusters for a library size of 8192 were calculated for coordinates corresponding to labeling positions 388-Cy3b/475-CF660R on RRM3/4 (without clash tests and accordingly with uniform weights), and the standard deviation $\sigma(\text{cluster})$ obtained for each pair. The standard deviation of the distance distribution between the two full Monte Carlo ensembles $\sigma(\text{MC}) = 0.943$ nm is obtained by random sampling from the two ensembles. The histogram of the relative widths $\sigma(\text{cluster})/\sigma(\text{MC})$ peaks at ca. 0.75 showing that the many distributions between two clusters are slightly narrower than the distribution between the full ensembles. This ratio becomes smaller at larger library sizes. Due to the heavy computation required these calculations were carried out to convergence rather than exhaustively.

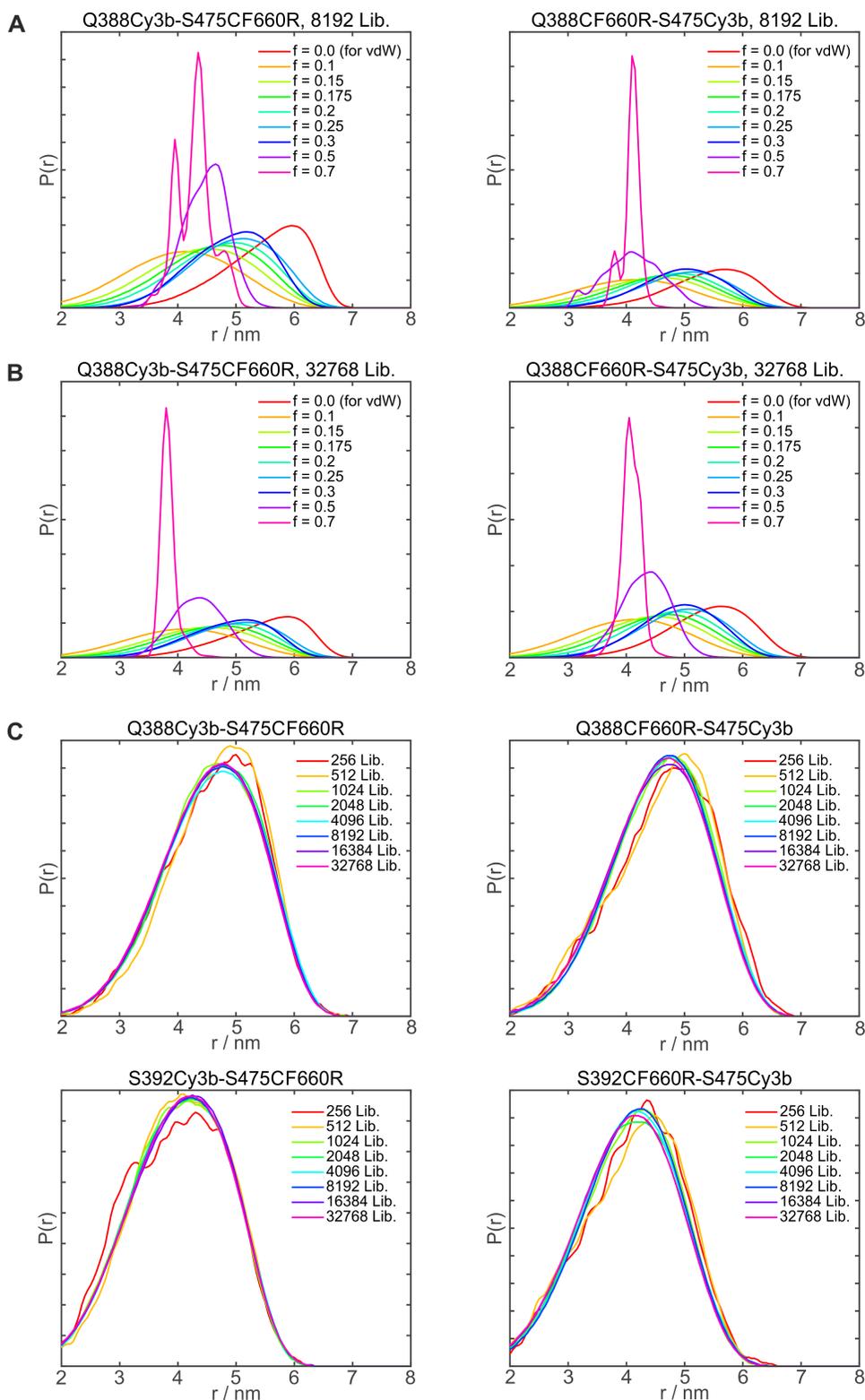


Figure S10: Distance distributions between fluorescence labels from rotamer analysis: Testing for f factor and library size. Representative distance distributions of RRM3/4 (PDB 2ASW, structure1- Δ N-term.) using different f factors (see legend) with rotamer library sizes of 8192 and 32768 rotamers in **A** and **B**, respectively. The distributions show a complex dependence on the f factor and indicate that libraries with a larger f factor require larger sizes for convergence with these labels. Where $f = 0$ for the van der Waals interaction, for the repulsive interaction f is kept at 0.7 for maximum contrast (cf. Eq. 1). **C** Representative distance distributions using the optimal f factor of $f = 0.175$ (see Fig. S11) using different rotamer library sizes (see legend). The strong similarity of the distributions shows that at this low f factor, small rotamer libraries (with respect to the number of canonical rotamers) of 1024 to 2048 rotamers yield results with sufficient sampling.

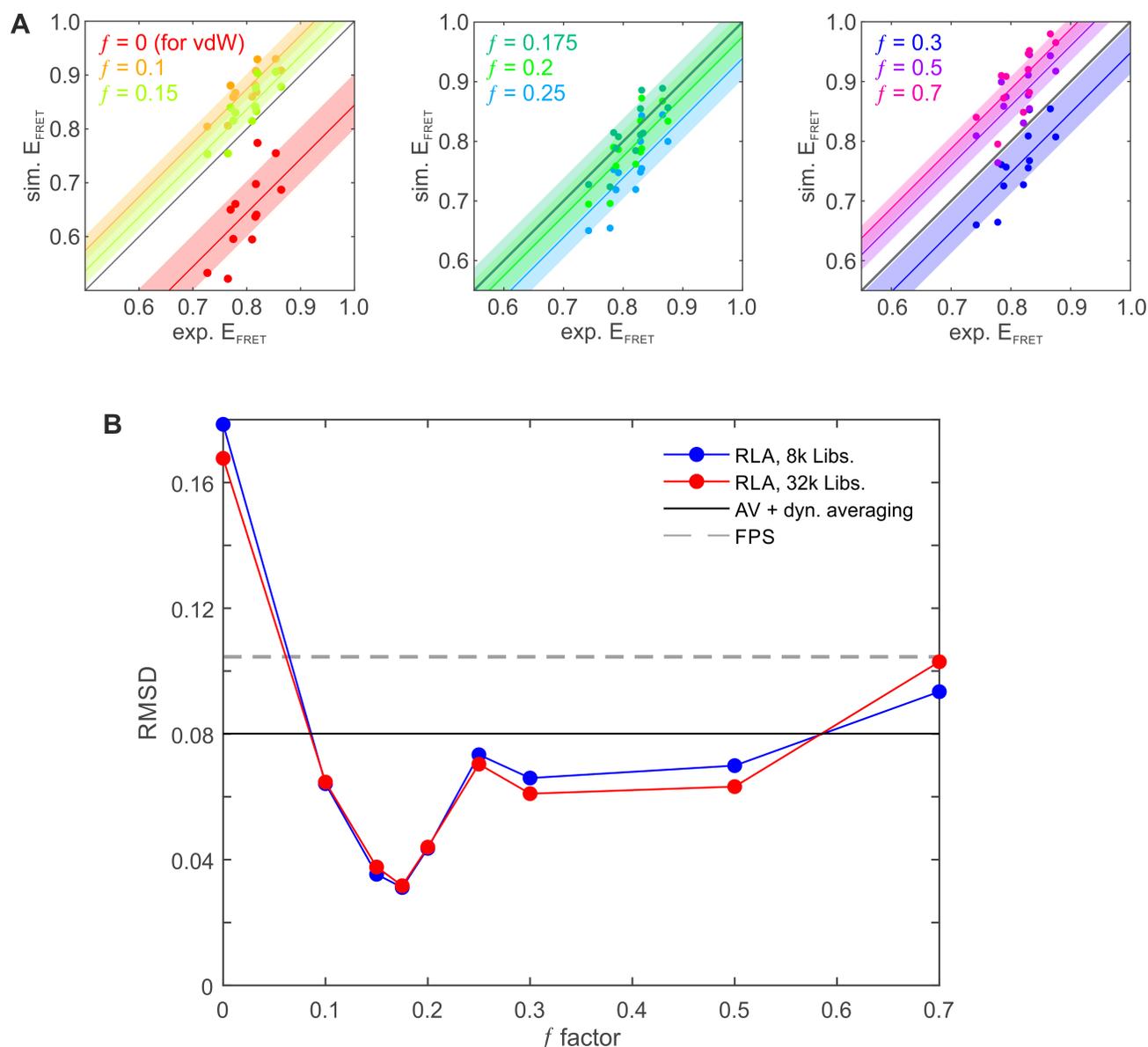


Figure S11: Choice of f factor for the rotamer libraries - RMSD values vs. experiments. **A** Experimental versus simulated FRET efficiencies. Simulated FRET efficiencies (points) from diffusional averaging ($D = 0.2 \text{ nm}^2/\text{ns}$ and $R_0 = 6 \text{ nm}$) using RLA libraries with different f factors, color-coded according to the legend. Standard deviations of simulated FRET efficiencies from twelve RRM3/4 variants due to uncertainties in Förster radius R_0 and translational diffusion constant D are indicated by color-shaded areas. Linear fits with slope 1 to the simulated FRET efficiencies (colored lines) emphasize the deviation from ideal 1:1 correlation (gray line). Note that the influence of the different f factors is larger than that of R_0 and D (given by the color-shaded areas), such that the different f factors can be distinguished to identify the optimum value of f . Further note, where $f = 0$ for the van der Waals interaction, for the repulsive interaction f is kept at 0.7 for maximum contrast (cf. Eq. 1). **B** RMSD values of experimental versus simulated FRET efficiencies for all f factors for 8k and 32k rotamer library sizes (blue and red, respectively) and AV distance distributions (see Fig. S12) after diffusional FRET averaging assuming $D = 0.2 \text{ nm}^2/\text{ns}$ (black), as well as transfer efficiencies from the AV approach simulated with the FPS software (gray), where the inter-dye distance is assumed to be static on the timescale of the excited-state lifetime of the donor. Since lower RMSD values represent better agreement with the experimental results, we obtain $f = 0.175$ as the optimal f factor.

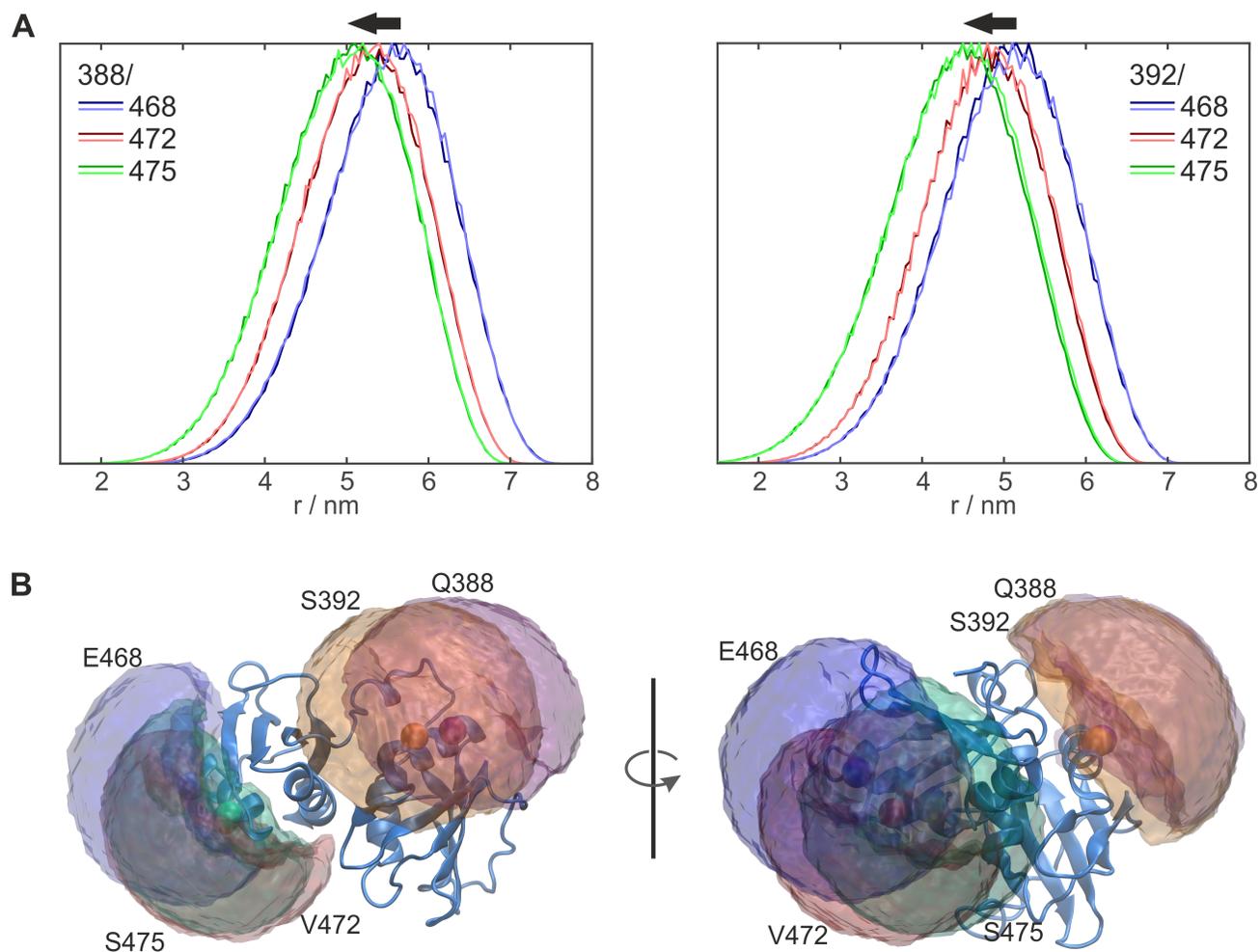


Figure S12: AV simulations: Distance distributions and spatial distributions. **A** Distance distributions calculated from AV simulations as a histogram over all pairwise distances between voxels of two accessible volumes. The labeling positions are indicated by color (see legend), with dark colors indicating the acceptor CF660R and light colors indicating the donor Cy3b at positions Q388C (left) and S392C (right). Arrows symbolize the shifts to smaller distances from E468C (blue) to S475C (green). **B** RRM3/4- Δ N (blue ribbons, PDB: 2ADC with N-terminal truncation) with accessible volumes of Cy3b at positions Q388C and S392C, and CF660R at positions E468C, V472C and S475C. From left to right, the structure is turned in the direction indicated by about 45°.

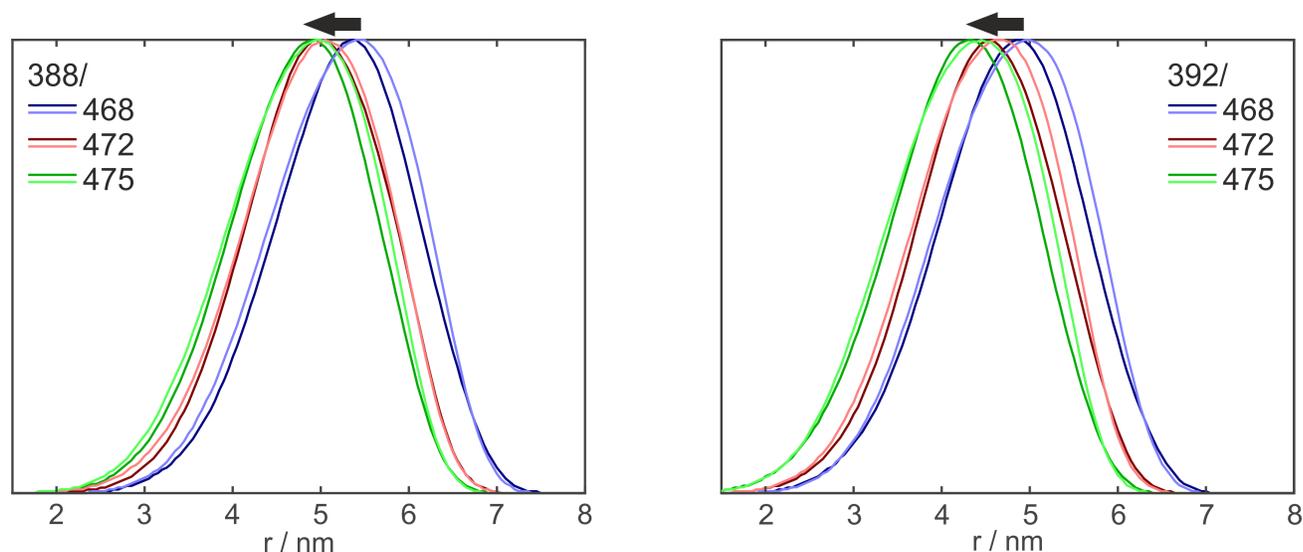


Figure S13: **Distance distributions between fluorescence labels calculated by RLA simulations for RRM3/4- Δ N.** RLA distance distributions based on the structure of RRM3/4- Δ N (PDB: 2ADC, first structure with N-terminal truncation). The labeling positions are indicated by color (see legend), with dark colors indicating the acceptor CF660R and light colors indicating the donor Cy3b at positions Q388C (left) and S392C (right). Arrows symbolize the shifts to smaller distances from E468C (blue) to S475C (green). These calculated distance distributions are virtually unchanged upon exchange of donor and acceptor fluorophores (see also Fig. 3 and Fig. 5 for corresponding FRET efficiencies), and are insensitive to the presence of the flexible N-terminus in the ensemble simulations (Fig. 3) *versus* the simulations based on RRM3/4- Δ N (here).

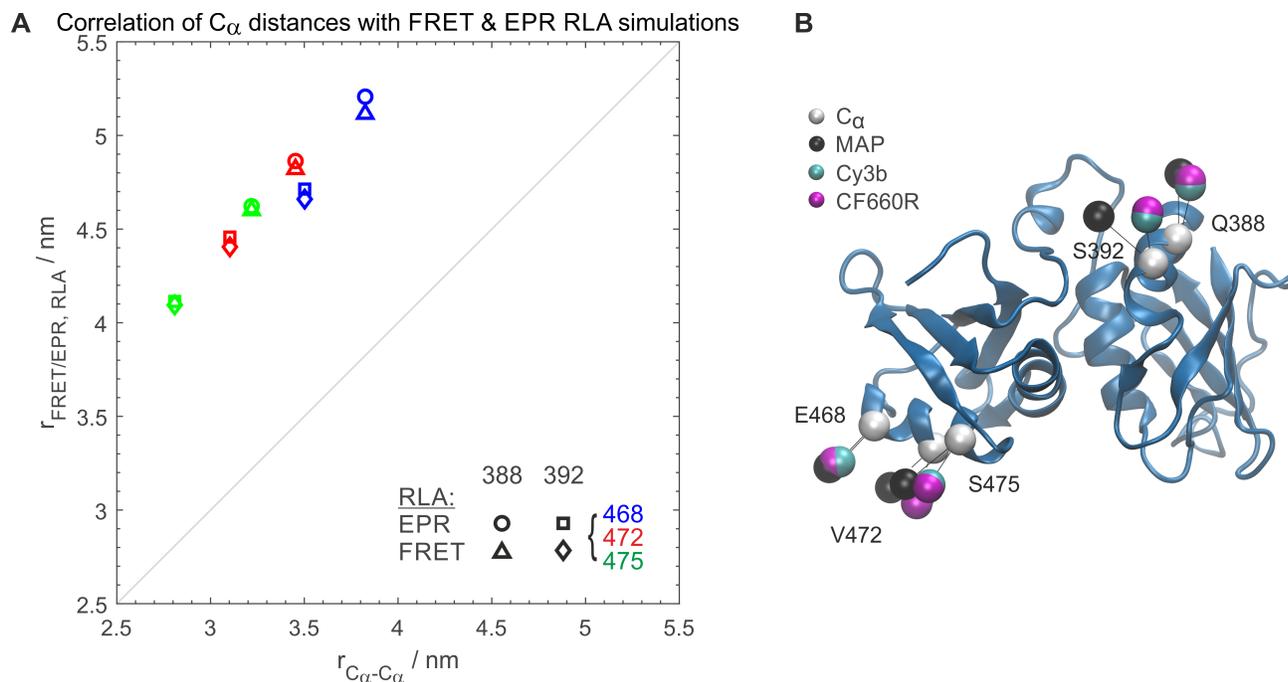


Figure S14: **Backbone Distances *versus* label-label distances.** **A** Correlation of backbone C_{α} - C_{α} distances to distances between labels determined by the RLA center of gravity distances for EPR and FRET (see Tabs. S1 and S2 for data). **B** 3D visualization of spin and fluorescence label midpoints (population-weighted mean of spatial distribution) on RRM3/4 (ribbon representation). Spheres represent the C_{α} atoms and the label midpoints according to the color code in the legend. The midpoints of spin and fluorescence labels on RRM3/4 show an RMSD of 2.1 Å.