Determining Protein Complex Structures Based on a Bayesian Model of *in Vivo* Förster Resonance Energy Transfer (FRET) Data*[S]

Massimiliano Bonomi‡§¶, Riccardo Pellarin‡, Seung Joong Kim‡, Daniel Russel‡, Bryan A. Sundin‖, Michael Riffle‖, Daniel Jaschob‖, Richard Ramsden‖, Trisha N. Davis‖, Eric G. D. Muller¶‖, and Andrej Sali‡¶**

The use of *in vivo* Förster resonance energy transfer (FRET) data to determine the molecular architecture of a protein complex in living cells is challenging due to data sparseness, sample heterogeneity, signal contributions from multiple donors and acceptors, unequal fluorophore brightness, photobleaching, flexibility of the linker connecting the fluorophore to the tagged protein, and spectral cross-talk. We addressed these challenges by using a Bayesian approach that produces the posterior probability of a model, given the input data. The posterior probability is defined as a function of the dependence of our FRET metric $FRET_R$ on a structure (forward model), a model of noise in the data, as well as prior information about the structure, relative populations of distinct states in the sample, forward model parameters, and data noise. The forward model was validated against kinetic Monte Carlo simulations and *in vivo* experimental data collected on nine systems of known structure. In addition, our Bayesian approach was validated by a benchmark of 16 protein complexes of known structure. Given the structures of each subunit of the complexes, models were computed from synthetic $FRET_R$ data with a distance root-mean-squared deviation error of 14 to 17 Å. The approach is implemented in the open-source Integrative Modeling Platform, allowing us to determine macromolecular structures through a combination of *in vivo* $FRET_R$ data and data from other sources, such as electron microscopy and chemical cross-linking. *Molecular & Cellular Proteomics 13: 10.1074/mcp.M114.040824, 2812–2823, 2014.*

From the ‡Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, California Institute for Quantitative Biosciences, University of California, San Francisco, California 94158; §Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom; ‖Department of Biochemistry, University of Washington, Seattle, Washington 98195

Received May 5, 2014, and in revised form, August 13, 2014

Published, MCP Papers in Press, August 19, 2014, DOI 10.1074/mcp.M114.040824

Author contributions: M.B., T.N.D., E.G.M., and A.S. designed research; M.B. and E.G.M. performed research; M.B., B.A.S., M.R., D.J., R.R., and E.G.M. contributed new reagents or analytic tools; M.B., R.P., S.K., D.R., and E.G.M. analyzed data; M.B., T.N.D., E.G.M., and A.S. wrote the paper.

Mapping the organization and function of the cell requires characterization of the structure and dynamics of biological assemblies (1, 2). However, the construction of models consistent with experimental data is often hampered by data sparseness due to incomplete measurements, data noise due to measurement errors, data ambiguity due to multiple copies of the same component in the assembly, and data mixture due to multiple structural states in a compositionally and conformationally heterogeneous sample.

Traditional modeling aims to find a single structural model by minimizing the difference between the data computed from the model and the experimental data. The noise in the data is typically not modeled accurately and thus biases the estimate of model precision. In contrast, Bayesian structural modeling (3, 4) interprets experimental data more objectively by explicitly accounting for data noise and prior knowledge about the system. Here, we developed a Bayesian approach that converts data from *in vivo* Förster resonance energy transfer (FRET)[1] spectroscopy into quantitative distance restraints suitable for structural modeling. The approach is available as part of the open-source Integrative Modeling Platform (IMP) (5, 6). IMP is a platform for integrative structure determination of macromolecular assemblies, based on a variety of experimental data, such as electron microscopy images and density maps, chemically cross-linked residue pairs, small angle x-ray scattering profiles, and various proteomics data (2, 7–10).

FRET is a powerful technique for studying protein–protein interactions both *in vitro* and in living cells (11, 12). FRET occurs when two spectrally matched fluorescent molecules are in close proximity and excitation energy is transferred from the donor to the acceptor fluorophore through nonradiative dipole–dipole coupling (Fig. 1*A*). The efficiency of this process (13) is a common experimentally derived variable of *in vitro* single-molecule experiments (14). It has been used to

[1] The abbreviations used are: FRET, Förster resonance energy transfer; $FRET_R$, index of relative FRET in cells; IMP, Integrative Modeling Platform; dRMS, distance-root-mean-square; GFP, green fluorescent protein; CFP, cyan fluorescent protein; YFP, yellow fluorescent protein; KMC, kinetic Monte Carlo; MD, Molecular Dynamics; AOI, area of interest.

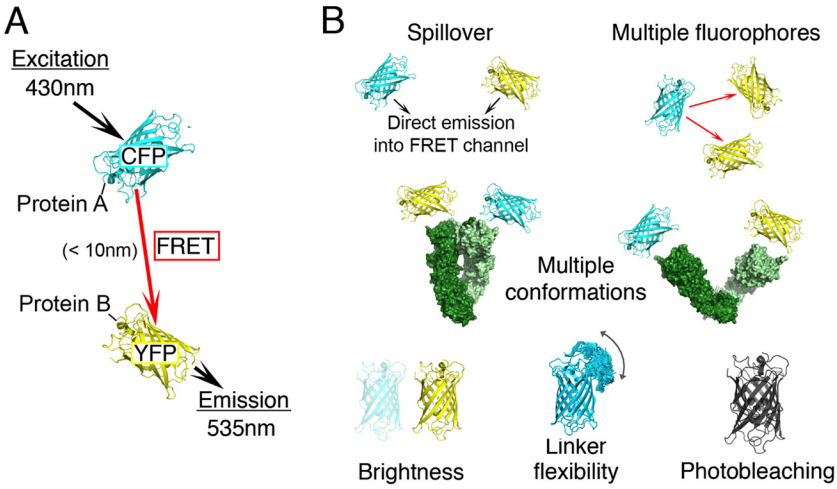


FIG. 1. ***In vivo* FRET microscopy.** *A*, two proteins of interest are tagged with pairs of colored variants of GFP, typically CFP and YFP. When CFP is illuminated (λ_{ex} = 430 nm) and the two fluorophores are sufficiently close, energy transfer occurs and fluorescence is measured at both CFP (λ_{em} = 470 nm) and YFP (λ_{em} = 535 nm) excitation wavelengths. The efficiency of energy transfer can be used to measure the distance between the two proteins. Typically, only the protein termini of each subunit are tagged with GFP; the total number of $FRET_R$ data points per complex that can be used in structural modeling is thus $N(2N - 1)$, where N is the number of subunits in the complex. *B*, a quantitative use of *in vivo* FRET data is complicated by data sparseness, multiple conformations, signal contributions from multiple donors and acceptors, uneven fluorophore brightness, photobleaching, flexibility of the linker connecting the fluorophore to the tagged protein, and spillover of donor and acceptor fluorescence.

probe distances over the range of 1 to 10 nm, resulting in spatial restraints for modeling the structure of the studied complex (15, 16).

Compared with *in vitro* FRET, *in vivo* FRET measurements present several additional challenges (17) (Fig. 1*B*) that mainly originate from the use of donor–acceptor pairs of color variants of the green fluorescent protein (GFP) (18, 19). Despite significant progress (20), these proteins are not ideal FRET partners, and four sources of noise that affect *in vitro* FRET are amplified. First, the unequal brightness of the two fluorophores can lead to different saturation levels in the donor and acceptor images. Second, the emission and excitation wavelengths of the GFP variants are broad and lead to contamination of the emission from energy transfer with light derived from direct emission from both donor and acceptor (direct acceptor excitation and spectral cross-talk). Third, in the case of the common FRET pair CFP–YFP, YFP is photobleached with exposure to the CFP excitation light and thus becomes gradually inactive during data collection. Fourth, fluorescent proteins are often attached to the tagged protein by means of long, flexible linkers that increase the structural variability of the system. In addition, some complexes may be composed of proteins that do not have 1:1 stoichiometry, and this complicates the interpretation of FRET data in terms of distances between individual components. Many of these problems can be overcome with the use of an experimental approach that measures fluorescence lifetimes of FRET donors (21). However, in many situations in live cells in which a complex is in low abundance, fluorescence lifetime measurements are not feasible (22).

The measurement of additional observables has been proposed to supplement the FRET efficiency as a way to address some of these problems (23). Among these observables is the $FRET_R$ index (24–26), a ratio that measures the fluorescence intensity at donor excitation and acceptor emission wavelengths relative to a calculated baseline expected in the absence of FRET. Our Bayesian approach computes this observable for a given structure while accounting for all sources of uncertainty of the *in vivo* $FRET_R$ data listed above, as well as for the presence of multiple distinct conformations in the sample (28, 29).[2] As a result, we can now use $FRET_R$ data to determine the molecular architectures of protein complexes *in vivo*.

Computational Methods and Experimental Procedures—

The $FRET_R$ Index—$FRET_R$ (24, 25) is an index of relative FRET in cells, based on the measurement of fluorescence intensities I_{YFP}, I_{FRET}, and I_{CFP} by an epifluorescence microscope configured with three filter set combinations. In this work, we used filter sets from Chroma® that yielded the YFP (excitation filter at λ_{ex} = 500 nm, emission filter at λ_{em} = 535 nm), FRET (λ_{ex} = 430 nm, λ_{em} = 535 nm), and CFP (λ_{ex} = 430 nm, λ_{em} = 470 nm) images. The baseline fluorescence detected in the FRET image that is not the result of FRET is quantified by the spillover factors S_d and S_d, measured in two separate experiments where YFP and CFP are expressed individually. The S_d factor quantifies the cross-talk between

[2] Bonomi, M., Pellarin, R., Spill, Y., Nilges, M., DeGrado, W., and Sali, A., in preparation.

donor and acceptor emission spectra in the filter sets, and the S_a factor quantifies the direct excitation of the acceptor. In an experiment in which YFP and CFP are co-expressed and energy transfer is measured, $FRET_R$ measures the fold-increase in the intensities in the FRET image relative to a computed and expected baseline.

$$FRET_R = \frac{I_{FRET}}{S_{tot}}, \quad \text{(Eq. 1)}$$

where $S_{tot} = S_d \cdot I_{CFP} + S_a \cdot I_{YFP}$.

Bayesian Model of $FRET_R$ Data—The Bayesian approach (3, 4) estimates the probability of a model given information available about the system, including prior knowledge and newly acquired experimental data. In the multi-state modeling of $FRET_R$ data, the model M consists of a set of N modeled structures $X = \{X_k\}$, their relative populations in the sample $\{w_k\}$, and additional parameters defined below. The posterior probability $p(M|D, I)$ of model M, given data D and prior information I, is

$$P(M|D, I) \propto P(D|M, I) \cdot P(M|I), \quad \text{(Eq. 2)}$$

where the likelihood function $p(D|M, I)$ is the probability of observing data D given M and I, and the prior $p(M|I)$ is the probability of model M given I. To define the likelihood function, one needs a forward model $f(X)$ that predicts the data point that would have been observed for structure(s) X and a noise model that specifies the distribution of the deviation between the observed and predicted data points. The Bayesian scoring function $S(M)$ is defined as $S(M) = -\log [p(D|M, I) \cdot p(M|I)]$ which ranks alternative models the same as the posterior probability.

Forward Model—An ensemble of CFPs and YFPs that are continuously excited by external radiation can return to the ground state through different independent decay pathways, including fluorescence and energy transfer from excited donors to non-excited acceptors. Following Förster theory (13), the rate of energy transfer between donor i and acceptor j is conveniently written as $k_{ij}^{ET} = \frac{k_d^F}{Q_d}\left(\frac{R_0}{R_{ij}}\right)^6$, where R_{ij} is the distance between the two fluorophores and R_0 is the Förster radius. The donor fluorescence quantum yield Q_d is the ratio between the fluorescence rate k_d^F and the total rate of decay and is proportional to the donor brightness. In general, R_0 depends on the orientation factor κ^2 of the interacting dipoles. We adopt the common assumption that donor and acceptor sample their orientations randomly on the time scale of the measurement (30), so that $\kappa^2 = 2/3$. This is considered particularly valid for fluorescent proteins attached by long, flexible linkers to targeted proteins. The linkers do not adopt a fixed conformation. Finally, the MD simulations described in "Results" showed that the linkers were sufficiently long to allow for orientational averaging during the time of image acquisition.

In the limit of rapid de-excitation and slow excitation rate (SI), the donor and acceptor fluorescence intensities are $I_d^F = Q_d \cdot k_d^X \cdot g(X)$ and $I_a^F = Q_a \cdot \{k_a^X \cdot [A] + k_d^X \cdot ([D] - g(X))\}$ where $g(X) = \sum_i \frac{1}{1 + F_i}[D_i]$ quantifies the donor fluorescent intensity in terms of CFP and YFP concentrations and relative proximities. F_i is computed from the Förster expression that relates the rate of energy transfer and distance R_{ij} between the two fluorophores i and j (13): $F_i = \sum_j (R_0/R_{ij})^6 [A_j] \cdot [D]$ and $[A]$ are the CFP donor and YFP acceptor concentrations, respectively, and k_d^X and k_d^X are their excitation rates. The $FRET_R$ forward model (supplemental Fig. S1*A*) is

$$f(X, I_{da}, k_{da}) = 1 + \frac{k_{da} \cdot \{[D] - g(X)\}}{I_{da} \cdot g(X) + [A]}, \quad \text{(Eq. 3)}$$

where I_{da} is the ratio of CFP and YFP fluorescence in two FRET images when each fluorescent protein is expressed individually at equal levels in separate cells. This quantity is treated as a free parameter, but its value is restrained by the experimental measurement (I_{da}^{exp} and σI_{da}^{exp}). $k_{da} = k_d^{X,430}/k_a^{X,430}$ is the ratio between donor and acceptor excitation rates at $\lambda_{ex} = 430$ nm; it is determined by the ratio between CFP and YFP absorption cross-sections at 430 nm. However, because each fluorescent protein has a different absorption spectrum and the excitation wavelength varies with the filter set, k_{da} is treated as a free parameter and is inferred along with the coordinates and the other unknown parameters.

Multi-state Forward Model—For FRET measurements of complexes within living cells, the observed $FRET_R$ may arise from multiple conformations of the complex. In such a case, $FRET_R$ should be expressed in terms of partial contributions resulting from the individual conformations X_k and proportional to their relative populations w_k. The single-state forward model (Eq. 3) can be generalized to take into account multiple states.

$$f(\{X_k, w_k\}, I_{da}, k_{da}) = 1 + \frac{k_{da} \cdot \{[D] - \langle g(X)\rangle\}}{I_{da} \cdot \langle g(X)\rangle + [A]}, \quad \text{(Eq. 4)}$$

where $\langle g(X)\rangle = \Sigma_k w_k g(X_k)$.

Photobleaching—YFP fluorophores are photochemically destroyed by prolonged exposure to radiation at wavelengths near the CFP absorption peak. For *in vivo* measurements, the observed $FRET_R$ is thus averaged over multiple copies of the system in which photobleached fluorophores do not contribute to the signal. Thus, the same multi-state forward model described above (Eq. 4) can be used, except that w_k corresponds to the proportion of molecules that are both non-photobleached and in state X_k.

Likelihood Function—The likelihood function $p(D|M, I)$ for dataset $D = \{d_n\}$ of N_F independently measured $FRET_R$ values is a product of likelihood functions $p(d_n|\{X_k, w_k\}, I_{da}, k_{da}, \sigma_n)$ for each data point. Because the observed $FRET_R$ values were strictly positive and unbounded, we modeled the uncertainty with a log-normal distribution:

$$p(d_n|\{X_k, w_k\}, l_{da}, k_{da}, \sigma_n) = \frac{1}{d_n\sigma_n\sqrt{2\pi}} \cdot \exp\left[-\frac{\log^2(d_n/f(\{X_k, w_k\}, l_{da}, k_{da}))}{2\sigma_n^2}\right]. \quad \text{(Eq. 5)}$$

To account for varying levels of noise in the data, each data point has an individual uncertainty σ_n.

Prior—The prior distribution $p(M|I)$ is a product of priors on the state coordinates X_k, relative populations w_k, forward model parameters l_{da} and k_{da}, and uncertainties σ_n. The priors on the coordinates $p(X_k)$ include terms to maintain the correct stereochemistry of the system, to avoid steric clashes between components, and to incorporate information other than $FRET_R$ data. The priors $p(w_k)$ are uniform distributions over the range from 0 to 1, with the constraint $\Sigma_k w_k = 1$. The priors $p(\sigma_n)$ are unimodal distributions (31):

$$p(\sigma_n|\sigma_0) = \frac{2\sigma_0}{\sqrt{\pi\sigma_n^2}}\exp\left(-\frac{\sigma_0^2}{\sigma_n^2}\right), \quad \text{(Eq. 6)}$$

where σ_0 corresponds to an unknown experimental uncertainty; the heavy tail of the distribution allows for outliers (supplemental Fig. S1*C*). The prior $p(\sigma_0)$ is a uniform distribution over the range from 0.001 to 0.01. If all $FRET_R$ values are measured with the same filter sets and fluorescent proteins, the same values of l_{da} and k_{da} can be used for all data points. The prior $p(l_{da}|l_{da}^{exp}, \sigma_{l_{da}^{exp}})$ is a normal distribution in which l_{da}^{exp} and $\sigma_{l_{da}^{ex}}$ are the average and standard error of the experimental measurements. The prior $p(k_{da})$ is a uniform distribution over the range from 1 to 15, based on typical ratios of CFP to YFP absorption cross-sections (32).

To facilitate sampling of the posterior distribution, we eliminate its dependence on the uncertainties σ_n by integrating the likelihood function and prior $p(\sigma_n\ \sigma_0)$ with respect to σ_n. Thus, the marginal likelihood function (supplemental Fig. S1*B*) is

$$p(d_n|\{X_k, w_k\}, l_{da}, k_{da}, \sigma_0) = \frac{\sqrt{2}\sigma_0}{\pi d_n} \cdot \frac{1}{log(d_n/f(\{X_k, w_k\}, l_{da}, k_{da}))^2 + 2\sigma_0^2}. \quad \text{(Eq. 7)}$$

A detailed description is provided in the supplemental material.

Kinetic Monte Carlo—KMC simulations (33, 34) were performed on *in silico* models of multiple CFP donors and YFP acceptors (one CFP–one YFP, two CFP–one YFP, and one CFP–two YFP). At each KMC step, one of the following reactions was randomly chosen on the basis of their rates: (a) excitation of either a single non-excited YFP (k_a^X) or (b) CFP (k_d^X); (c) de-excitation of a single excited YFP by either fluorescence (k_a^F) or (d) other pathways; or (e) de-excitation of a single excited CFP by fluorescence (k_d^F), (f) energy transfer to a non-excited YFP (k_{ij}^{ET}), or (g) other pathways. The rate of decay via pathways other than fluorescence was defined by the CFP and YFP quantum yields of fluorescence Q_d and Q_a, which were both set at 0.5. The factor k_{ij}^{ET} was equal to $\frac{k_d^F}{Q_d} \cdot \left(\frac{R_0}{R_{ij}}\right)^6$ where the Förster radius R_0 was set at 4.9 nm. k_d^F and k_a^F were set (35) at 0.4 ns^{-1}. Simulations were run for multiple values of $k_d^{X,430}$ and $k_a^{X,430}$, and $k_a^{X,500}$ was calculated from supplemental Eq. S1. The distance between CFP and YFP was varied between 3 and 10 nm in steps of 0.5 nm. For each choice of the parameters, $FRET_R$ was calculated from Eq. 1 based on the results of three 0.1-s KMC runs used to simulate imaging experiments with 0.1-s exposures. The intensities in the CFP, FRET, and YFP images were calculated from the number of reactions of a given type occurring during the simulations. Based on experimental measurements, S_d and S_a were set at 0.831 and 0.249, respectively. To account for photobleaching, YFPs were randomly labeled as inactive during the acquisition of the CFP image (with the probability set at 0.3) and then removed from the list of possible reactions. $FRET_R$ was thus calculated by averaging quantities over 3200 independent KMC simulations.

Molecular Dynamics—MD simulations were performed with GROMACS4 (36) and PLUMED (37, 38), using the AMBER99SB-ILDN (39) all-atom force field. An implicit solvent based on the Generalized Born formalism combined with the Still method (40) for calculating the Born radii was used. Temperature was controlled by the Bussi–Donadio–Parrinello (41) thermostat. A cutoff of 1.5 nm was used for electrostatic and Lennard–Jones interactions. The parallel tempering algorithm (42) was used to accelerate sampling.

Parallel Tempering Simulation of GFP and Linker—The crystal structure of recombinant wild-type green fluorescent protein (PDB code 1GFL (43)) was used as a template. Modeler 9v8 (44) was used to model the C-terminal residues (HGMDELYKGA) present in the GFP sequence, but not in the crystal structure, and the GlyAla motif at the N terminus. The first 7 and the last 14 residues were treated as flexible segments based on the fluctuations observed in a preliminary MD run. The positions of the other heavy atoms of the protein were restrained by a harmonic potential, with the spring constant equal to 9×10^3 kJ · mol^{-1} · nm^{-2}. 32 replicas were distributed over a temperature range from 300 to 500 K. Simulations were carried out for an aggregate time of 1 μs.

Combined Parallel Tempering and Metadynamics Simulations of Polyprolines—The polyproline constructs YFP–$(PRO)_n$–CFP with n = (0, 5, 10, 15, 20) were simulated through a combination of parallel tempering and metadynamics (45–47). 16 to 40 replicas were used to span a temperature range from 300 to 600 K. A collective variable measuring the number of prolines in *cis* and *trans* conformations was used to accelerate proline *cis–trans* isomerization. For an n-mer peptide, this collective variable was defined (48) as $\Omega = \sum_{i=1}^{n-1} cos\omega_i$ where the torsional angle ω formed by the quadruplet Cα–C–N–Cα was equal to 0° for the *cis* isomer and to 180° for the *trans* isomer. The well-tempered (49) variant of metadynamics was used, with a bias factor equal to 30 and an initial depo-

sition rate of 1 kJ · mol^{-1} · ps^{-1}. YFPs and CFPs were not simulated at atomistic resolution; only the residues belonging to the flexible N- and C-terminal fragments defined in the previous paragraph were explicitly modeled. The fluorescent proteins were instead represented as virtual atoms defined in the fixed reference frame of the first and last modeled residues. Restraints on all distances between virtual and other atoms were used to enforce steric repulsion. A reweighting algorithm (50) was applied to obtain the unbiased distribution of distances between the two virtual atoms representing the center of the fluorophores. Simulations were carried out for an aggregate time ranging from 1 to 8 μs.

Parallel Tempering Simulations of Other Proteins—The NMR structures of the THP12-carrier protein from yellow meal worm (PDB code 1C3Y (51)) and the fourth LIM domain of PINCH protein (PDB code 1NYP (52)), as well as the crystal structures of the human TBP-associated factor hTAF(II)28/hTAF(II)18 heterodimer (here abbreviated as TAF28-TAF18) (PDB code 1BH8 (53)) and the ferrodoxin:thioredoxin reductase (PDB code 1DJ7 (54)), were used as templates. Modeler was used to model the flexible linkers at the N and C termini. Preliminary short MD simulations at 300 K were carried out to measure the fluctuations in terms of distance-root-mean-square (dRMS) deviation from the native state. A restraint on the dRMS was then used during the parallel tempering simulations to avoid unfolding at high temperatures. The terminal flexible residues were not considered in the dRMS calculation. Multiple replicas (from 16 to 64) were used to span a temperature range from 300 to 600 K. YFPs and CFPs were not simulated explicitly (see previous paragraph).

Benchmark—The benchmark was carried out with the open-source IMP (5, 6), version develop-c47408c. The benchmark results and scripts are available online. The method was tested on 11 ternary and 5 quaternary complexes of known structure, selected from 3D Complex (55). For each pair of subunits in the complex, simulated data were generated for all combinations of the N and C termini of the pair, corresponding to 12 and 24 data points for ternary and quaternary complexes, respectively. Low- and high-noise datasets were generated by setting σ_0 equal to 0.001 and 0.01, respectively. The average of 50 different random extractions from the marginal likelihood distribution (Eq. 7) was used to simulate the average from repeated experiments, with the typical standard deviation equal to 0.04 and 0.19 for low- and high-noise data, respectively. The typical standard deviation for *in vivo* data is 0.15. Different percentages (100% and 50%) of the total amount of data were used to assess the role of data sparseness in modeling accuracy. To model linker flexibility, a Gaussian mixture model was fit on a set of 5000 probes of radius equal to 10 Å using 10 Gaussian components. The conformation of each subunit was obtained from the crystal structure of the entire complex; it was represented with Cα atoms for each residue and treated as an independent rigid body. An excluded volume potential was used to avoid steric clashes between subunits. Coordinates, forward model, and likelihood parameters were sampled via a Gibbs sampling scheme combined with a simulated annealing Monte Carlo algorithm. A Monte Carlo move of each rigid subunit consisted of a random rotation and translation of at most 17° and 1.0 Å, respectively. A Monte Carlo move of the forward model parameters k_{da}, I_{da}, and σ_0 consisted of a random perturbation of at most 0.3, 0.3, and 0.001, respectively. Temperature was varied between 1.0 and 5.0 k_BT. The initial positions were randomized in a cubic box with dimensions of 100 Å. For each structure and choice of parameters, 20 independent simulated annealing Monte Carlo runs were performed. A total of 2560 tests were conducted, each for a total of 3×10^7 simulated annealing Monte Carlo steps (supplemental Fig. S9).

In Vivo $FRET_R$ Measurements—Saccharomyces cerevisiae strains expressing the YFP and CFP tagged proteins were grown and imaged as previously described (25). The fluorescent proteins were linked to the target proteins through unstructured linkers. Exposure times were either 0.08 or 0.1 s for each image, allowing for a prolonged sampling of an ensemble of proteins such that each can adopt different relative orientations of the fluorescent proteins. Expression of all constructs was driven by the strong TEF promoter. Importantly, all constructs were engineered with a nuclear localization signal, resulting in two advantages. First, the uniform nuclear fluorescence was used as an indication of proper protein folding, and second, nuclear localization allowed the cytoplasm to be used to measure a local background in the cell. All constructs were integrated into the host genome to ensure uniform cell-to-cell gene expression. Plasmids used for integrating the constructs are described in supplemental Table S1.

Image analysis was performed with FRETSCAL, an integrated collection of MATLAB scripts with a graphical user interface. FRETSCAL identifies an area of interest (AOI) within the images and calculates $FRET_R$ for each AOI. FRETSCAL has user-controlled selection criteria that (i) define the size of the AOI, (ii) set a maximum pixel intensity of the AOI to ensure that selected AOIs are within the linear range of the image acquisition CCD camera, (iii) set a minimum signal-to-background ratio, (iv) set a maximum cutoff value for the width of a Gaussian fit of the intensity values within the AOI, and (v) define other parameters that automate AOI selection and analysis. The software is open source and is available online at the MATLAB Central website.

A single value of $FRET_R$ is calculated as a ratio of the mean background subtracted value of the whole nuclear region in the FRET image divided by the projected value if there was no energy transfer. The projected value is calculated from the corresponding nuclei in the YFP and CFP images of the same field. The projected value is the sum of the mean background subtracted value of the whole nuclear region in the YFP image multiplied by the YFP spillover factor plus the mean background subtracted value of the whole nuclear region in the

CFP image multiplied by the CFP spillover factor. The spillover factors are determined as described above under the $FRET_R$ heading.

All images used in this study are available online from the YRC Public Image Repository. In addition, a composite image is shown that displays the FRETSCAL output. In the online composite image, the nuclei that satisfied the selection criteria used in FRETSCAL are framed in yellow. The corresponding background pixels are shown in gray.

RESULTS

Our Bayesian approach for determining a macromolecular architecture from *in vivo* FRET data is based on a microscopic interpretation (forward model) of the experimental observable $FRET_R$ in terms of structural models and other parameters. It is thus crucial to first assess the validity of the forward model. To do so, we began with computational validation by means of KMC simulations (33, 34) of *in silico* models of multiple CFP donors and YFP acceptors. We then proceeded with comparisons of $FRET_R$ predictions from molecular dynamics simulations to *in vivo* experimental data that were collected from yeast cells expressing constructs of CFP and YFP separated by any one of nine defined linkers and protein structures. Finally, the accuracy of structural modeling using synthetic $FRET_R$ data and the structures of each individual subunit was assessed via comparison of native molecular architectures of 16 protein complexes with their models computed with our Bayesian approach.

Kinetic Monte Carlo Validation of the Forward Model—Based on the physics of fluorescent molecules, we derived master equations that express the excitation and emission of an ensemble of FRET donors and acceptors as visualized with a fluorescent microscope (supplemental Eqs. S2*A* and S2*B*). The $FRET_R$ forward model (Eq. 4) is derived from an approximate solution of these master equations in the limit of rapid de-excitation and slow excitation rate. As a validation of this approximation, the value of the $FRET_R$ predicted by Eq. 4 was compared with the results of KMC simulations governed by the master equations S2*A* and S2*B*. The KMC simulations described the evolution of an *in silico* model of multiple CFP donors and YFP acceptors and computed $FRET_R$ in every excitation/de-excitation regime. For this comparison, we represented CFP and YFP as dimensionless points whose distance and other parameters were varied ("Computational Methods and Experimental Procedures").

$FRET_R$ changed smoothly with the distance between a single CFP and YFP over the range from 3 to 10 nm (Fig. 2). When the CFP excitation rate k_a^X was much smaller than its fluorescent rate k_d^F ($k_a^X/k_D^F < 0.05$), excellent agreement was found between $FRET_R$ from the forward model and KMC simulations, with deviations of less than 1% under all conditions (supplemental Fig. S2*A*).

$FRET_R$ was also computed from KMC simulations of systems of two CFPs and one YFP (supplemental Fig. S3*A*) and of one CFP and two YFPs (supplemental Fig. S3*B*). The behavior of $FRET_R$ differs in the two cases. When multiple donors surround a single acceptor, adjacent donors compete for non-excited acceptors. In contrast, a relative abundance of acceptors increases the chance of energy transfer. However, the effect on energy transfer is shaped by the relative rates of excitation and emission of the donor and acceptor (supplemental Fig. S3*C*). In the limit of rapid de-excitation and slow excitation rate, the agreement between the forward model and KMC simulations was still excellent in both cases, with deviations of less than 1% under all conditions (supplemental Figs. S2*B* and S2*C*).

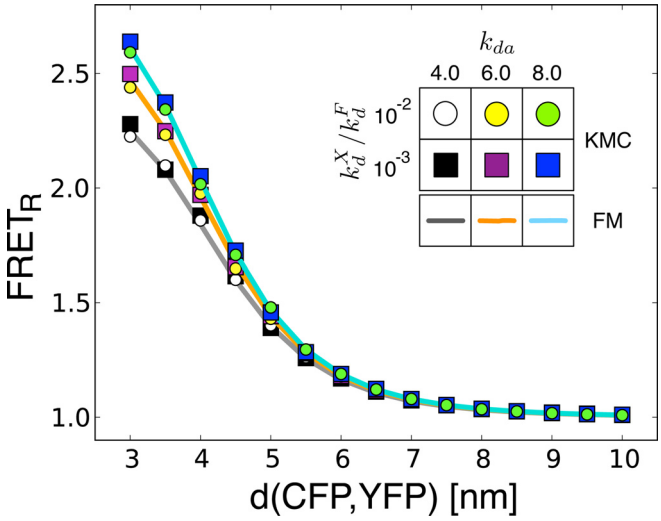


FIG. 2. **KMC validation of the forward model.** $FRET_R$ was calculated from KMC simulations of a system consisting of one CFP donor and one YFP acceptor. CFP and YFP were represented as dimensionless points that correspond to the centers of the fluorophore. $FRET_R$ was computed with varying distance between fluorophores, donor excitation rate k_d^X/k_d^F, and donor/acceptor excitation rate ratio $k_{da} = k_d^{X,430}/k_a^{X,430}$ (circles and squares). k_d^F was set (35) at 0.4 nm^{-1}. The YFP photobleaching ratio during the acquisition of the CFP image was set at 0.3. In the regime of low excitation and fast de-excitation ($k_d^X/k_d^F < 0.05$), the value of $FRET_R$ predicted by the forward model (lines) was in excellent agreement with KMC simulations (error analysis in supplemental Fig. S2).

In all the KMC simulations mentioned above, we included the effect on YFP photobleaching during the experiment. To examine this effect directly, we investigated a model system of multiple YFP acceptors. As expected, with fewer acceptors available because of photobleaching, energy transfer was attenuated at all CFP–YFP distances (compare value in supplemental Fig. S4*A* with that in supplemental Fig. S4*B*); again, the $FRET_R$ computed by the forward model, which included the effect of YFP photobleaching (supplemental Fig. S4*C*), agreed with that from the KMC simulations that included photobleaching (supplemental Fig. S4*B*).

These comparisons demonstrate that the approximate expression for $FRET_R$ given by the forward model (Eq. 4) agrees well with more complex (and far more computationally expen-

sive) simulations based on a more comprehensive physical treatment.

In Vivo Experimental Validation of the Forward Model—We further validated the $FRET_R$ forward model by comparing the predictions from MD simulations to *in vivo* experimental data that we collected on nine proteins of known structure that were expressed in *S. cerevisiae* (supplemental Table S1). These nine systems included a tandem YFP–CFP; YFP–$[Pro]_n$–CFP in which n was equal to 5, 10, 15, or 20 prolines; and four constructs in which CFP and YFP were attached to the N or C termini of proteins of known structure. The latter four constructs were as follows: (i) YFP-THP12-CFP; (ii) YFP-Lim4-CFP; (iii) YFP -TAF28-CFP co-expressed with TAF18; and (iv) FTR117-CFP co-expressed with FTR74-YFP. Finally, a control measurement on the co-expressed but unlinked YFP and CFP pair showed no energy transfer ($FRET_R = 1.04$). In each case hundreds of images of hundreds of cells were acquired. A sample set of images is shown in Fig. 3*A*. All the images used in the dataset are available online at the YRC Public Image Repository. Automated processing of the images was accomplished with the software FRETSCAL. The large number ($n \geq 200$) of identified AOIs provided a strong statistical foundation for the $FRET_R$ measurements used in the Bayesian analysis.

In comparing our forward model against experimental data, we took into account the dependence of the measured $FRET_R$ on the presence of multiple conformations in the sample. To do so, we used MD simulations combined with advanced sampling techniques to explore the conformational landscape of the test structures. Although polyproline peptides have often been employed as a spectroscopic ruler, several experimental (56–58) and computational (48, 57) studies have questioned the role of polyproline as a "rigid rod" in a single dominant conformation. Prolyl isomerization from the *trans* to *cis* isomer, whose activation energy is on the order of 10 to 20 kcal/mol (59, 60), converts the left-handed polyproline II helix (PPII) to the more compact right-handed polyproline I helix (PPI). Thus, a heterogeneous population of structures with distinct patterns of *cis* and *trans* isomers of proline is expected to be present in a cell.

The conformational landscape of polyprolines in solution was predicted by all-atom MD simulations in implicit solvent using parallel tempering (42) and metadynamics (45, 46). These techniques allow (i) exhaustive sampling by accelerating proline *trans-cis* isomerization and (ii) estimates of the equilibrium relative populations $\{w_k\}$ of the conformers (Eq. 4). The polyproline II helix was favored over the polyproline I helix across all lengths studied (supplemental Fig. S5), in agreement with previous computational (48) and experimental results (61). The conformational landscape of the other constructs was also explored using similar computational approaches. Finally, simulations of the tandem YFP–CFP showed that the linkers at the N and C termini were sufficiently long to allow for orientational averaging of the fluorophores on the time scale of the FRET experiment (supplemental Fig. S6).

To compare the $FRET_R$ forward model with experimental data, we calculated the weighted average of $g(X)$, which depends on the model coordinates (Eq. 4), as the ensemble average over the MD conformations (supplemental Figs. S7*A* and S7*B*). We inferred the forward model parameters k_{da} and l_{da}, along with the uncertainty σ_0, by maximizing the posterior distribution, which was defined based on all nine data points using the mean experimental value $k_{da}^{exp} = 6.0$ and standard error $\sigma_{l\,da}^{exp} = 2.0$. Using the inferred parameters ($k_{da} = 7.7$, $l_{da} = 6.6$, and $\sigma_0 = 0.05$), we found good agreement between the forward model and measured $FRET_R$ values (Fig. 3*B*, white and black bars, respectively), except for one outlier, TAF28-TAF18. When the procedure was repeated without the outlier (Fig. 3*B*, gray bars), the inferred parameter values $k_{da} = 7.5$ and $l_{da} = 6.2$ changed minimally, and the data uncertainty σ_0 dropped from 0.05 to 0.03, as expected upon removal of an outlier data point. Thus, the forward model and associated parameters can effectively account for the influence of components of wide-field fluorescence microscopy, such as installed filter sets and illumination intensity, on the measurement of the efficiency of fluorescence energy transfer. The $FRET_R$ forward model can accurately relate $FRET_R$ values and fluorophore distances.

Finally, to improve the computational efficiency of the forward model, we fit an efficient Gaussian mixture model to the expensive all-atom MD simulations of the linker (SI), without a significant decrease in the accuracy of the forward model (supplemental Fig. S8).

Benchmark of Modeling Accuracy—The accuracy of the molecular architectures modeled based on synthetic $FRET_R$ data, given the knowledge of the structure of each subunit, was mapped with the aid of known structures for 16 protein complexes of three and four subunits (55). For this benchmark, we used synthetic $FRET_R$ data that were computed by first applying our $FRET_R$ forward model (Eq. 4) to all pairs of N and C termini of each subunit in the native structures and then adding noise (Eq. 7). The accuracy was defined as the Cα dRMS deviation between the native structure and the most probable model found by the sampling algorithm in IMP, averaged on 20 independent runs. The use of synthetic data in this benchmark allowed us to map the accuracy of structural modeling from $FRET_R$ data as a function of the level of data noise and sparseness, with (supplemental Table S2) and without (supplemental Table S3) taking the linker flexibility into account. A flowchart explaining the different steps of the benchmark is presented in supplemental Fig. S9. It is conceivable, however, that the accuracy of models computed from real $FRET_R$ data might be worse than that from the simulated data, despite our effort to include noise in the simulated data. Real $FRET_R$ data for the FTR117-FTR74 case were not used as a benchmark case, because the flexibility

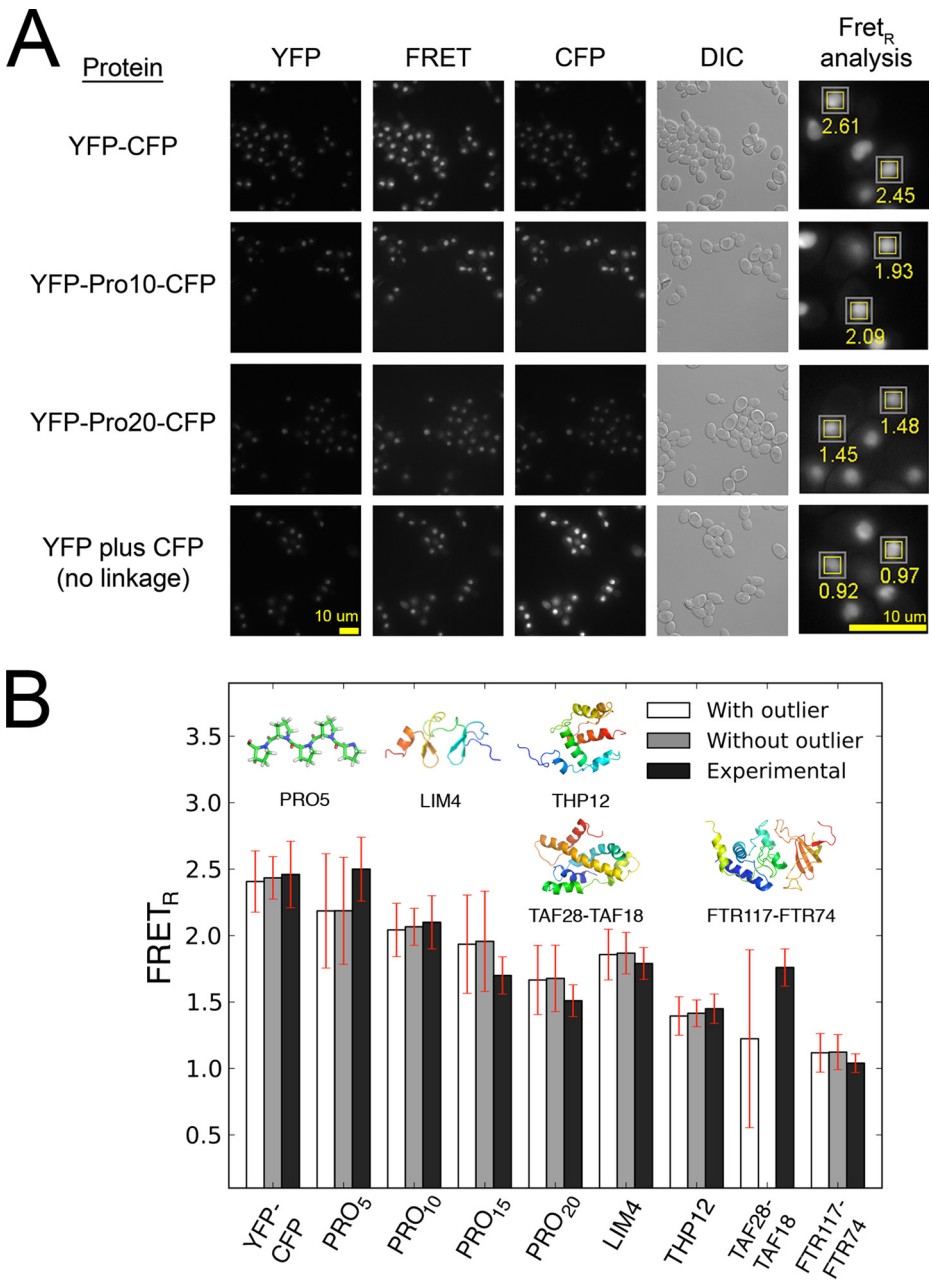


FIG. 3. **Experimental validation of the forward model.** FRET$_R$ values were determined for nine proteins of known structure. The proteins were stably expressed in *S. cerevisiae* with nuclear localization signals (see "Computational Methods and Experimental Procedures"). *A*, a sample of captured images. A 4× enlargement of one region shows the FRET$_R$ values determined by FRETSCAL. *B*, FRET$_R$ values measured *in vivo* on nine proteins of known structure (black bars) compared with the values predicted by the forward model (white bars). The Bayesian parameters were inferred by maximizing the posterior probability on the set of nine measurements. The fit was repeated excluding the outlier data point (TAF28-TAF18) and yielded similar results (gray bars), demonstrating the ability of the Bayesian approach to tolerate outliers. Red error bars indicate experimental standard error and inferred uncertainty.

and the resulting demand on sampling made it difficult to run the benchmark a very large number of times.

When 100% of the data points were used, the accuracy of the predicted structure of the complex was 13.9 Å and 14.8 Å for ternary and quaternary complexes, respectively. This accuracy was marginally reduced to 16.1 Å and 17.4 Å when noisy data were used. The weak dependence on the noise level resulted from the small standard error of FRET$_R$ obtained by averaging FRET$_R$ over many (~100) independent experiments. In contrast, the accuracy was strongly dependent on data sparseness. When only 50% of the data points were used, the accuracy decreased to a range from 20.4 Å to 21.5

Å, depending on the number of subunits and the noise level. This result emphasizes the need to compile as much information as possible from *in vivo* measurements.

Because $FRET_R$ data provide information about the distance between the protein termini, we expected much greater accuracy in determining the positions of the terminal residues ($dRMS_T$ in supplemental Tables S2 and S3). Indeed, the accuracy was 5.2 Å to 9.3 Å for ternary complexes and 7.1 Å to 11.6 Å for quaternary complexes, depending on the noise level.

Finally, the accuracy was also affected by the linker flexibility (supplemental Table S3). In particular, the positions of the tagged termini were inferred with greater accuracy ($\langle \Delta dRMS_T \rangle = 2.2$ Å) when the simulated data were created and the sampling was performed without the linker flexibility. However, the inclusion of the linker flexibility had a relatively small effect on the accuracy ($\langle \Delta dRMS \rangle = 1.1$ Å). Thus, the presence of a flexible linker, while allowing orientational averaging of the fluorophores (supplemental Fig. S6), does not dramatically affect the accuracy of our approach.

DISCUSSION

Many observables have been introduced to quantify *in vivo* FRET (23). Fluorescence lifetime microscopy overcomes many of the problems associated with epifluorescence microscopy, but it is technically challenging and applicable only for complexes with a robust fluorescence signal (21, 22, 62–65). Many FRET indexes have successfully processed steady-state epifluorescence images to yield significant insights into the dynamics of protein associations in live cells (22, 23, 66). However, this work represents the first case in which the supporting theory and structural predictions from a FRET metric have been modeled and tested both *in silico*, with molecular dynamic simulations, and *in vivo*, with benchmark protein complexes.

Although our Bayesian approach could be adapted to incorporate other FRET metrics, or even FRET efficiencies derived from fluorescence lifetime microscopy, we chose the metric $FRET_R$. To our knowledge this is the only live-cell FRET metric in which structural arrangements predicted from *in vivo* measurements were directly confirmed *in vitro* by means of single particle analysis. $FRET_R$ measurements of the γ-tubulin complex in yeast predicted the location of the N and C termini of two proteins, Spc97 and Spc98, in the complex (25). Fluorescent proteins linked to these ends were later directly visualized at the predicted locations via electron microscopy (67). $FRET_R$ has also been used to analyze the structure of the yeast spindle pole body (24, 68) and cohesion architecture (69), and more recently the organization of the yeast kinetochore (26). Of course $FRET_R$ also has limitations, and it is most appropriate for experimental conditions in which the proteins in a complex are uniformly tagged with a fluorescent protein, gene expression is tightly regulated and typically driven from native promoters, and free unincorporated proteins do not interfere with the FRET measurements (17, 23–25). We showed that our $FRET_R$ forward model is accurate, first by comparing the predicted value (Eq. 4) with that computed from KMC simulations of an *in silico* model of multiple CFP donors and YFP acceptors. Excellent agreement was found for typical conditions of fluorescence microscopy,[3] where CFPs and YFPs were not saturated by the incident illumination. In addition, KMC simulations on systems of multiple donors and acceptors (supplemental Fig. S3) illustrated the expected asymmetry of the one CFP–two YFP and two CFP–one YFP experiments and suggested that data from experiments in which the positions of YFP and CFP are swapped provide independent and thus useful information and should not be averaged (24).

We also validated the forward model using experimental data by comparing predicted $FRET_R$ to *in vivo* data collected on nine proteins of known structure, including fluorescent proteins separated by polyproline peptides of different lengths (Fig. 3*B*). Accurate modeling of the experimental data required explicit modeling of multiple conformations in the sample (supplemental Figs. S5 and S7). Although in this study the relative populations $\{w_k\}$ were predetermined by MD simulations, in general they can be inferred along with the coordinates of the system and other parameters using multi-state Bayesian scoring functions (27–29).

We demonstrated that the Bayesian approach is robust with respect to the presence of outlier data points. Collecting $FRET_R$ data in living cells requires tagging a complex with CFP–YFP pairs that might perturb the system and affect its structure. As a result, a data point might not correctly represent the native structure of the complex and thus might be inconsistent with other information, including other $FRET_R$ measurements. For example, the $FRET_R$ value predicted for TAF28-TAF18 was significantly different from the observed one (Fig. 3*B*). This discrepancy might arise from several other factors besides structural changes due to the insertion of the fluorophores, such as non-converged MD simulations and inaccuracy of the molecular mechanics force field. Importantly, for each data point, an uncertainty parameter is either inferred or marginalized (31), allowing those points that are not consistent with the bulk of the data to be properly down-weighted in the construction of the model.

The results of the benchmark (Fig. 4 and supplemental Table S2) indicated the importance of using multiple data points to model a structure. Synthetic $FRET_R$ data between all pairs of subunit N and C termini determined the structure of ternary and quaternary complexes with an accuracy of ~15 Å

[3] For example, when collecting data for the yeast spindle pole body, 1.5 mW of light from the source illuminates the sample, corresponding to a photon per fluorophore every ~50 ns. The excitation rate is of course smaller than implied by this photon flux ($k_a^X/k_d^F < 0.05$), because the YFP and CFP absorption cross-sections are typically much smaller than the fluorophore area.

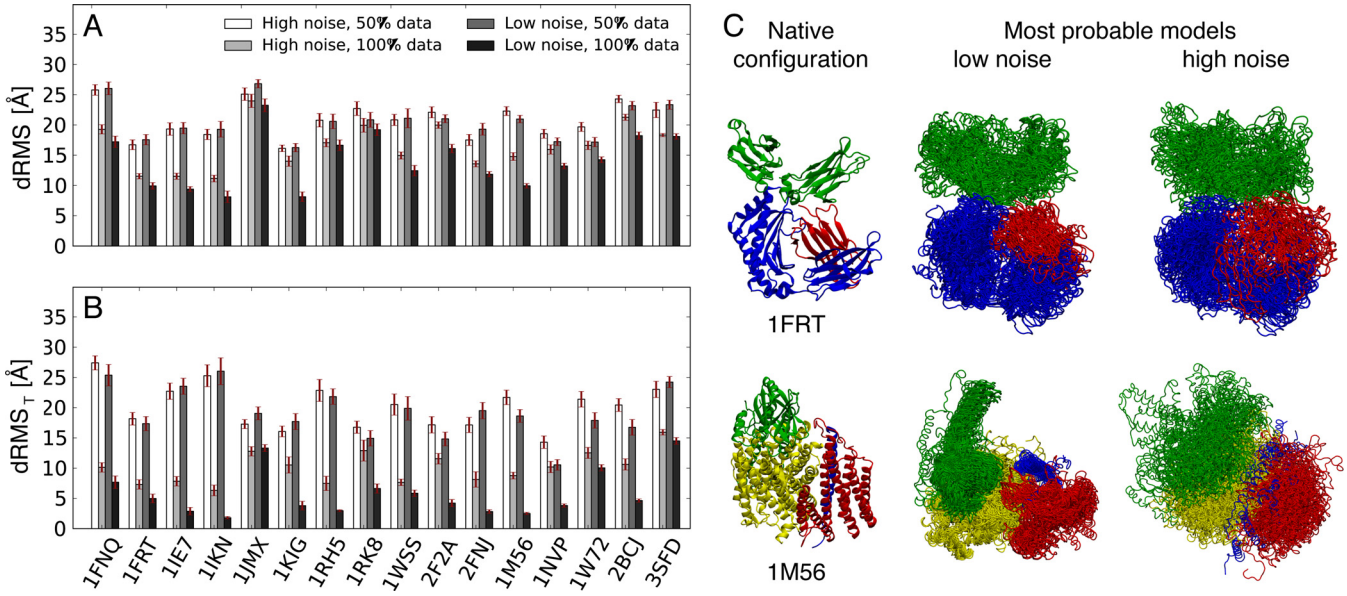


FIG. 4. **Benchmark of a set of complexes modeled using simulated FRET$_R$ data.** *A*, *B*, accuracy of the modeled structures as a function of data sparseness and level of noise. Accuracy was quantified using the average Cα dRMS deviation between the crystallographic structure and the 20 most probable models, each one from an independent run, calculated on the entire complex (*A*) and on the N- and C-terminal residues (*B*); red error bars indicate standard error of the mean. *C*, most probable models of the complex of rat neonatal Fc receptor with Fc (PDB code 1FRT (73)) and of cytochrome *c* oxidase from *Rhodobacter sphaeroides* (PDB code 1M56 (27)), with low- and high-noise datasets, and with all pairwise FRET$_R$ indices.

(Cα dRMS), whereas using only 50% of the data decreased the accuracy to ~20 Å. The greatest structural uncertainty is in the orientation between the subunits. The accuracy can thus be improved if further data are collected. Typically, only the protein termini of each subunit are tagged with GFP; the total number of FRET$_R$ data points per complex that can be used in structural modeling is thus $N(2N - 1)$, where N is the number of subunits of the complex. However, in principle fluorescent proteins can be inserted at positions other than the protein termini, although such insertions might be more likely to alter the structure of the complex.

Like any search-based approach, our method requires a sufficiently thorough configurational sampling algorithm. Here, we used advanced sampling techniques, including Gibbs sampler MC with simulated annealing (70) and MD combined with parallel tempering and metadynamics (47). We explicitly assessed whether sampling was sufficiently thorough by demonstrating the convergence of the model as a function of the number of sampled models (supplemental Fig. S7).

Compared with other methods that mostly deal with *in vitro* FRET data (15, 16), our approach treats all noise sources that characterize measurements in living cells, accounts for sample heterogeneity, and is robust to outlier data points. Furthermore, our approach is more general, because it allows the use of *in vivo* data collected in both bulk experiments, where multiple CFP and YFP contribute to the measured FRET$_R$, and single-molecule experiments (71), in which a single CFP–YFP pair is present; in the latter application, the observed FRET$_R$ is not the ratio of average intensities in the different images (Eq. 4), but the average of FRET$_R$ measured on samples in which the YFP is either active or photobleached.

Finally, we implemented our method in IMP, an open-source platform for integrative structural modeling of macromolecular systems (5). Through IMP, FRET$_R$ data can be combined with information obtained via other methods, such as electron microscopy, chemical and cysteine cross-linking, small angle x-ray scattering, proteomics, and other theoretical or statistical analyses, in an integrative or hybrid approach (5, 72). The uncertainty in the orientation of the subunits based on FRET$_R$ data alone could thus be resolved by considering additional complementary data, even if sparse and noisy. The Bayesian approach is expected to be even more useful in integrative modeling than modeling based on FRET$_R$ data alone, because data from different experiments can in principle be properly weighted and thus seamlessly integrated.

Acknowledgments—We are grateful to David Sivak and Charles Asbury for commenting on the manuscript and to Ben Webb for help in setting up the benchmark. We also thank Peter Schurmann for the clone of FTR, Peter L. Davies for the clone of THP12, Christophe Romier for the hTAF clones, and Jun Qin for the clone of the Lim4.

* This work was funded by NIH Grant Nos. R01 GM083960 (A.S.), U54 RR022220 (A.S.), and P41 GM103533 (E.M. and T.D.) and was supported by SNSF through Grant Nos. PBZHP3-133388 and PA00P3_139727 (R.P.).

[S] This article contains supplemental material.

¶ To whom correspondence should be addressed: Andrej Sali, Ph.D., Professor, Department of Bioengineering and Therapeutic Sci-

ences, Department of Pharmaceutical Chemistry, California Institute for Quantitative Biosciences, University of California, San Francisco, 1700 4th Street, San Francisco, CA 94158-2330, Tel.: 1-415-514-4227, E-mail: sali@salilab.org; Massimiliano Bonomi, Ph.D., Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom, Tel.: 44-0-1223-336366, E-mail: mb2006@cam.ac.uk; Eric Muller, Ph.D., Department of Biochemistry, University of Washington, 1705 NE Pacific Street, Seattle, WA 98195-7350, Tel.: 1-206-616-3028, E-mail emuller@u.washington.edu.

** Current address: Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, WA 98195.

REFERENCES

1. Sali, A., Glaeser, R., Earnest, T., and Baumeister, W. (2003) From words to literature in structural proteomics. *Nature* **422,** 216–225
2. Alber, F., Dokudovskaya, S., Veenhoff, L., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B., Rout, M., and Sali, A. (2007) Determining the architectures of macromolecular assemblies. *Nature* **450,** 683–694
3. Rieping, W., Habeck, M., and Nilges, M. (2005) Inferential structure determination. *Science* **309,** 303–306
4. Habeck, M., Nilges, M., and Rieping, W. (2005) Bayesian inference applied to macromolecular structure determination. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **72,** 031912
5. Russel, D., Lasker, K., Webb, B., Velazquez-Muriel, J., Tjioe, E., Schneidman-Duhovny, D., Peterson, B., and Sali, A. (2012) Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol.* **10,** e1001244
6. Alber, F., Forster, F., Korkin, D., Topf, M., and Sali, A. (2008) Integrating diverse data for structure determination of macromolecular assemblies. *Annu. Rev. Biochem.* **77,** 443–477
7. Bau, D., Sanyal, A., Lajoie, B. R., Capriotti, E., Byron, M., Lawrence, J. B., Dekker, J., and Marti-Renom, M. A. (2011) The three-dimensional folding of the alpha-globin gene domain reveals formation of chromatin globules. *Nat. Struct. Mol. Biol.* **18,** 107–114
8. Lasker, K., Forster, F., Bohn, S., Walzthoeni, T., Villa, E., Unverdorben, P., Beck, F., Aebersold, R., Sali, A., and Baumeister, W. (2012) Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 1380–1387
9. Velazquez-Muriel, J., Lasker, K., Russel, D., Phillips, J., Webb, B. M., Schneidman-Duhovny, D., and Sali, A. (2012) Assembly of macromolecular complexes by satisfaction of spatial restraints from electron microscopy images. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 18821–18826
10. Fernandez-Martinez, J., Phillips, J., Sekedat, M. D., Diaz-Avalos, R., Velazquez-Muriel, J., Franke, J. D., Williams, R., Stokes, D. L., Chait, B. T., Sali, A., and Rout, M. P. (2012) Structure-function mapping of a heptameric module in the nuclear pore complex. *J. Cell Biol.* **196,** 419–434
11. Selvin, P. R. (2000) The renaissance of fluorescence resonance energy transfer. *Nat. Struct. Biol.* **7,** 730–734
12. Jares-Erijman, E. A., and Jovin, T. M. (2003) Fret imaging. *Nat. Biotechnol.* **21,** 1387–1395
13. Förster, T. (1948) Zwischenmolekulare Energiewanderung und Fluoreszenz. Intermolecular energy transfer and fluorescence *Annalen der Physik* **437,** 55–75
14. Roy, R., Hohng, S., and Ha, T. (2008) A practical guide to single-molecule FRET. *Nat. Methods* **5,** 507–516
15. Brunger, A. T., Strop, P., Vrljic, M., Chu, S., and Weninger, K. R. (2011) Three-dimensional molecular modeling with single molecule FRET. *J. Struct. Biol.* **173,** 497–505
16. Kalinin, S., Peulen, T., Sindbert, S., Rothwell, P. J., Berger, S., Restle, T., Goody, R. S., Gohlke, H., and Seidel, C. A. (2012) A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat. Methods* **9,** 1218–1225
17. Piston, D. W., and Kremers, G. J. (2007) Fluorescent protein FRET: the good, the bad and the ugly. *Trends Biochem. Sci.* **32,** 407–414
18. Giepmans, B. N., Adams, S. R., Ellisman, M. H., and Tsien, R. Y. (2006) The fluorescent toolbox for assessing protein location and function. *Science* **312,** 217–224
19. Lam, A. J., St-Pierre, F., Gong, Y., Marshall, J. D., Cranfill, P. J., Baird, M. A., McKeown, M. R., Wiedenmann, J., Davidson, M. W., Schnitzer, M. J., Tsien, R. Y., and Lin, M. Z. (2012) Improving FRET dynamic range with bright green and red fluorescent proteins. *Nat. Methods* **9,** 1005–1012
20. Kremers, G. J., Gilbert, S. G., Cranfill, P. J., Davidson, M. W., and Piston, D. W. (2011) Fluorescent proteins at a glance. *J. Cell Sci.* **124,** 157–160
21. Becker, W. (2012) Fluorescence lifetime imaging—techniques and applications. *J. Microsc.* **247,** 119–136
22. Zeug, A., Woehler, A., Neher, E., and Ponimaskin, E. G. (2012) Quantitative intensity-based FRET approaches—a comparative snapshot. *Biophys. J.* **103,** 1821–1827
23. Berney, C., and Danuser, G. (2003) FRET or no FRET: a quantitative comparison. *Biophys. J.* **84,** 3992–4010
24. Muller, E. G., Snydsman, B. E., Novik, I., Hailey, D. W., Gestaut, D. R., Niemann, C. A., O'Toole, E. T., Giddings, T. H., Jr., Sundin, B. A., and Davis, T. N. (2005) The organization of the core proteins of the yeast spindle pole body. *Mol. Biol. Cell* **16,** 3341–3352
25. Kollman, J. M., Zelter, A., Muller, E. G., Fox, B., Rice, L. M., Davis, T. N., and Agard, D. A. (2008) The structure of the gamma-tubulin small complex: implications of its architecture and flexibility for microtubule nucleation. *Mol. Biol. Cell* **19,** 207–215
26. Aravamudhan, P., Felzer-Kim, I., Gurunathan, K., and Joglekar, A. P. (2014) Assembling the protein architecture of the budding yeast kinetochore-microtubule attachment using FRET. *Curr. Biol.* **24,** 1437–1446
27. Svensson-Ek, M., Abramson, J., Larsson, G., Tornroth, S., Brzezinski, P., and Iwata, S. (2002) The X-ray crystal structures of wild-type and EQ(I-286) mutant cytochrome c oxidases from Rhodobacter sphaeroides. *J. Mol. Biol.* **321,** 329–339
28. Street, T. O., Zeng, X., Pellarin, R., Bonomi, M., Sali, A., Kelly, M. J., Chu, F., and Agard, D. A. (2014) Elucidating the mechanism of substrate recognition by the bacterial Hsp90 molecular chaperone. *J. Mol. Biol.* **426,** 2393–2404
29. Molnar, K. S., Bonomi, M., Pellarin, R., Clinthorne, G. D., Gonzalez, G., Goldberg, S. D., Goulian, M., Sali, A., and DeGrado, W. (2014) Cys-scanning disulfide crosslinking and Bayesian modeling probe the transmembrane signaling mechanism of the histidine kinase, PhoQ. *Structure*, **22,** 1239–1251
30. Stryer, L. (1978) Fluorescence energy-transfer as a spectroscopic ruler. *Annu. Rev. Biochem.* **47,** 819–846
31. Sivia, D. S., and Skilling, J. (2006) *Data Analysis: A Bayesian Tutorial*, 2nd ed., Oxford University Press, Oxford, UK
32. Strohhofer, C., Forster, T., Chorvat, D., Kasak, P., Lacik, I., Koukaki, M., Karamanou, S., and Economou, A. (2011) Quantitative analysis of energy transfer between fluorescent proteins in CFP-GBP-YFP and its response to Ca2+. *Phys. Chem. Chem. Phys.* **13,** 17852–17863
33. Bortz, A. B., Kalos, M. H., and Lebowitz, J. L. (1975) New algorithm for Monte-Carlo simulation of Ising spin systems. *J. Comput. Phys.* **17,** 10–18
34. Young, W. M., and Elcock, E. W. (1966) Monte Carlo studies of vacancy migration in binary ordered alloys—I. *P. Phys. Soc. Lond.* **89,** 735
35. Kremers, G. J., Goedhart, J., van Munster, E. B., and Gadella, T. W. J. (2006) Cyan and yellow super fluorescent proteins with improved brightness, protein folding, and FRET Forster radius. *Biochemistry* **45,** 6570–6580
36. Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4,** 435–447
37. Bonomi, M., Branduardi, D., Bussi, G., Camilloni, C., Provasi, D., Raiteri, P., Donadio, D., Marinelli, F., Pietrucci, F., Broglia, R. A., and Parrinello, M. (2009) PLUMED: a portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **180,** 1961–1972
38. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., and Bussi, G. (2014) PLUMED 2: new feathers for an old bird. *Comput. Phys. Commun.* **185,** 604–613
39. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78,** 1950–1958
40. Qiu, D., Shenkin, P. S., Hollinger, F. P., and Still, W. C. (1997) The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A* **101,** 3005–3014
41. Bussi, G., Donadio, D., and Parrinello, M. (2007) Canonical sampling

through velocity rescaling. *J. Chem. Phys.* **126** 014101
42. Sugita, Y., and Okamoto, Y. (1999) Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314,** 141–151
43. Yang, F., Moss, L. G., and Phillips, G. N. (1996) The molecular structure of green fluorescent protein. *Nat. Biotechnol.* **14,** 1246–1251
44. Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234,** 779–815
45. Laio, A., and Parrinello, M. (2002) Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 12562–12566
46. Barducci, A., Bonomi, M., and Parrinello, M. (2011) Metadynamics. *Wires Comput. Mol. Sci.* **1,** 826–843
47. Bussi, G., Gervasio, F. L., Laio, A., and Parrinello, M. (2006) Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics. *J. Am. Chem. Soc.* **128,** 13435–13441
48. Moradi, M., Babin, V., Roland, C., Darden, T. A., and Sagui, C. (2009) Conformations and free energy landscapes of polyproline peptides. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 20746–20751
49. Barducci, A., Bussi, G., and Parrinello, M. (2008) Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100** 020603
50. Bonomi, M., Barducci, A., and Parrinello, M. (2009) Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. *J. Comput. Chem.* **30,** 1615–1621
51. Rothemund, S., Liou, Y. C., Davies, P. L., Krause, E., and Sonnichsen, F. D. (1999) A new class of hexahelical insect proteins revealed as putative carriers of small hydrophobic ligands. *Struct. Fold. Des.* **7,** 1325–1332
52. Velyvis, A., Vaynberg, J., Yang, Y. W., Vinogradova, O., Zhang, Y. J., Wu, C. Y., and Qin, J. (2003) Structural and functional insights into PINCH LIM4 domain-mediated integrin signaling. *Nat. Struct. Biol.* **10,** 558–564
53. Birck, C., Poch, O., Romier, C., Ruff, M., Mengus, G., Lavigne, A. C., Davidson, I., and Moras, D. (1998) Human TAF(II)28 and TAF(II)18 interact through a histone fold encoded by atypical evolutionary conserved motifs also found in the SPT3 family. *Cell* **94,** 239–249
54. Dai, S. D., Schwendtmayer, C., Schurmann, P., Ramaswamy, S., and Eklund, H. (2000) Redox signaling in chloroplasts: cleavage of disulfides by an iron-sulfur cluster. *Science* **287,** 655–658
55. Levy, E. D., Pereira-Leal, J. B., Chothia, C., and Teichmann, S. A. (2006) 3D complex: a structural classification of protein complexes. *PLoS Comput. Biol.* **2,** 1395–1406
56. Schuler, B., Lipman, E. A., Steinbach, P. J., Kumke, M., and Eaton, W. A. (2005) Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 2754–2759
57. Best, R. B., Merchant, K. A., Gopich, I. V., Schuler, B., Bax, A., and Eaton, W. A. (2007) Effect of flexibility and cis residues in single-molecule FRET studies of polyproline. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 19064–19066
58. Doose, S., Neuweiler, H., Barsch, H., and Sauer, M. (2007) Probing polyproline structure and dynamics by photoinduced electron transfer provides evidence for deviations from a regular polyproline type II helix. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 17400–17405
59. Fischer, S., Dunbrack, R. L., and Karplus, M. (1994) Cis-trans imide isomerization of the proline dipeptide. *J. Am. Chem. Soc.* **116,** 11931–11937
60. Jhon, J. S., and Kang, Y. K. (1999) Imide cis-trans isomerization of N-acetyl-N′-methylprolineamide and solvent effects. *J. Phys. Chem. A* **103,** 5436–5439
61. Kakinoki, S., Hirano, Y., and Oka, M. (2005) On the stability of polyproline-I and II structures of proline oligopeptides. *Polym. Bull.* **53,** 109–115
62. Sikor, M., Mapa, K., von Voithenberg, L. V., Mokranjac, D., and Lamb, D. C. (2013) Real-time observation of the conformational dynamics of mitochondrial Hsp70 by spFRET. *EMBO J.* **32,** 1639–1649
63. Alber, F., Chait, B. T., Rout, M. P., and Sali, A. (2008) Integrative structure determination of protein assemblies by satisfaction of spatial restraints. In *Protein-Protein Interactions and Networks: Identification, Characterization and Prediction* (Panchenko, A., and Przytycka, T., Eds), pp. 99–114, Springer-Verlag, London, UK
64. Visser, A. J. W. G., Laptenok, S. P., Visser, N. V., van Hoek, A., Birch, D. J. S., Brochon, J. C., and Borst, J. W. (2010) Time-resolved FRET fluorescence spectroscopy of visible fluorescent protein pairs. *Eur. Biophys. J. Biophys.* **39,** 241–253
65. Raicu, V., and Singh, D. R. (2013) FRET spectrometry: a new tool for the determination of protein quaternary structure in living cells. *Biophys. J.* **105,** 1937–1945
66. Hoppe, A. D., Scott, B. L., Welliver, T. P., Straight, S. W., and Swanson, J. A. (2013) N-way FRET microscopy of multiple protein-protein interactions in live cells. *PLoS One* **8** e64760
67. Choy, R. M., Kollman, J. M., Zelter, A., Davis, T. N., and Agard, D. A. (2009) Localization and orientation of the gamma-tubulin small complex components using protein tags as labels for single particle EM. *J. Struct. Biol.* **168,** 571–574
68. Mathieson, E. M., Suda, Y., Nickas, M., Snydsman, B., Davis, T. N., Muller, E. G., and Neiman, A. M. (2010) Vesicle docking to the spindle pole body is necessary to recruit the exocyst during membrane formation in Saccharomyces cerevisiae. *Mol. Biol. Cell* **21,** 3693–3707
69. McIntyre, J., Muller, E. G., Weitzer, S., Snydsman, B. E., Davis, T. N., and Uhlmann, F. (2007) In vivo analysis of cohesin architecture using FRET in the budding yeast Saccharomyces cerevisiae. *EMBO J.* **26,** 3783–3793
70. Kirkpatrick, S., Gelatt, C. D., Jr., and Vecchi, M. P. (1983) Optimization by simulated annealing. *Science* **220,** 671–680
71. Coelho, M., Maghelli, N., and Tolic-Norrelykke, I. M. (2013) Single-molecule imaging in vivo: the dancing building blocks of the cell. *Integr. Biol.* **5,** 748–758
72. Ward, A. B., Sali, A., and Wilson, I. A. (2013) Integrative structural biology. *Science* **339,** 913–915
73. Burmeister, W. P., Huber, A. H., and Bjorkman, P. J. (1994) Crystal-structure of the complex of rat neonatal Fc receptor with Fc. *Nature* **372,** 379–383

Supplementary Computational methods

The $FRET_R$ forward model. The fluorescence intensities measured in 3 different steady-state experiments are needed to define $FRET_R$: in the first two, CFP and YFP are expressed individually and the spillover factors are calculated, in the third CFP and YFP are co-expressed and $FRET_R$ is measured. Therefore, we need a microscopic interpretation of these 3 experiments to write the $FRET_R$ forward model. A detailed derivation is provided in the following paragraphs.

1) YFP individually expressed

A YFP sample excited by an external radiation can return to the ground state through different independent decay pathways: fluorescence and other non-radiative processes, such as dynamic collisional quenching, near-field dipole-dipole interaction, internal conversion, and intersystem crossing. The quantum yield of fluorescence Q_a is defined as the ratio $Q_a = {k_a^F}/{\sum k_a^i}$, where k_a^F is the acceptor fluorescence rate and the sum is over all the decay pathways, each with rate k_a^i. The *master equation* for this system is:

$$\frac{\partial [A^*]}{\partial t} = -\frac{k_a^F}{Q_a}[A^*] + k_a^X([A] - [A^*]),$$

where $[A^*]$ is the concentration of excited acceptors, $[A]$ is the concentration of acceptors, and k_a^X is the excitation rate. The stationary solution for $[A^*]$ is:

$$[A^*] = \frac{k_a^X \cdot [A]}{\frac{k_a^F}{Q_a} + k_a^X},$$

and the fluorescence intensity is:

$$I_a^F = k_a^F \cdot [A^*] = \frac{k_a^F \cdot k_a^X \cdot [A]}{\frac{k_a^F}{Q_a} + k_a^X}.$$

In the limit of rapid de-excitation and slow excitation, $k_a^F \gg k_a^X$:

$$I_a^F = Q_a \cdot k_a^X \cdot [A].$$

This intensity is spread over a broad emission spectrum:

$$I_a^F = \int d\lambda\, I_a^F(\lambda) = (1 + \Lambda_a) \cdot I_a^F\,(\lambda_{em} = 535\text{nm}),$$

where Λ_a is the fraction of the total fluorescence that is emitted at wavelengths not detected by the emission filter used in the experiment (*lost fluorescence*). The emission and excitation wavelengths come from the center of the bandpass filter sets used in our *in vivo* experiments. The acceptor spillover factor S_a is defined as the ratio between the fluorescence intensities measured at λ_{em}=535nm when the sample is illuminated at λ_{ex}=430nm and λ_{ex}=500nm:

$$S_a = \frac{I_a^F(\lambda_{em} = 535\text{nm},\ \lambda_{ex} = 430\text{nm})}{I_a^F(\lambda_{em} = 535\text{nm},\ \lambda_{ex} = 500\text{nm})} = \frac{k_a^{X,430}}{\frac{k_a^F}{Q_a} + k_a^{X,430}} \cdot \frac{\frac{k_a^F}{Q_a} + k_a^{X,500}}{k_a^{X,500}}, \quad (S1)$$

where $k_a^{X,430}$ and $k_a^{X,500}$ are the acceptor excitation rates with incident radiation at λ_{ex}=430nm and λ_{ex}=500nm, respectively. In the limit of rapid de-excitation and slow excitation $k_a^F \gg k_a^X$, the acceptor spillover factor can be approximated by:

$$S_a \simeq \frac{k_a^{X,430}}{k_a^{X,500}}.$$

2) CFP individually expressed

Similarly to the previous experiment, the donor fluorescence intensity I_d^F can be written as:

$$I_d^F = \frac{k_d^F \cdot k_d^X \cdot [D]}{\frac{k_d^F}{Q_d} + k_d^X},$$

where k_d^F is the donor fluorescence rate, k_d^X is the donor excitation rate, Q_d is the donor quantum yield of fluorescence, and $[D]$ is the donor concentration. This fluorescence is spread over a spectrum of emission. The intensities at λ_{em}=470nm and λ_{em}=535nm are measured, and the donor spillover factor is defined as the ratio:

$$S_d = \frac{I_d^F(\lambda_{em} = 535\text{nm}, \lambda_{ex} = 430\text{nm})}{I_d^F(\lambda_{em} = 470\text{nm}, \lambda_{ex} = 430\text{nm})}.$$

The total donor fluorescence can thus be written as:

$$I_d^F = \int d\lambda\, I_d^F(\lambda) = (1 + S_d + \Lambda_d) \cdot I_d^F(\lambda_{em} = 470\text{nm}).$$

where Λ_d is the donor lost fluorescence. The intensities $I_d^F(\lambda_{em} = 470\text{nm})$ and $I_d^F(\lambda_{em} = 535\text{nm})$ can thus be written in terms of the total fluorescence as:

$$I_d^F(\lambda_{em} = 470\text{nm}) = \frac{1}{(1 + S_d + \Lambda_d)} \cdot I_d^F,$$

$$I_d^F(\lambda_{em} = 535\text{nm}) = \frac{S_d}{(1 + S_d + \Lambda_d)} \cdot I_d^F.$$

3) YFP and CFP co-expressed

The master equations of a system of fluorophores that are excited by an external radiation and can transfer energy from an excited donor to a non-excited acceptor is:

$$\frac{\partial [D_i^*]}{\partial t} = -\frac{k_d^F}{Q_d}[D_i^*] + k_d^X([D_i] - [D_i^*]) - \sum_j k_{ij}^{ET}[D_i^*]\left([A_j] - [A_j^*]\right), (S2a)$$

$$\frac{\partial [A_j^*]}{\partial t} = -\frac{k_a^F}{Q_a}[A_j^*] + k_a^X([A_j] - [A_j^*]) + \sum_i k_{ij}^{ET}[D_i^*]\left([A_j] - [A_j^*]\right), \quad (S2b)$$

where $[D_i^*]$ is the concentration of the excited donor *i* and k_{ij}^{ET} is the rate constant of energy transfer between donor *i* and acceptor *j*. Following Förster theory, k_{ij}^{ET} can be written in terms of the distance R_{ij} between the two fluorophores:

$$k_{ij}^{ET} = \frac{k_d^F}{Q_d} \cdot \left(\frac{R_0}{R_{ij}}\right)^6,$$

where R_0 is the Förster radius. In general, this radius depends on the orientation factor κ^2 of the interacting dipoles. Here, we adopt the common assumption that donor and acceptor sample their orientations randomly on the time scale of the measurement, so that[1] κ^2=2/3. The equations above form a system of coupled differential equations. In the limit of rapid de-excitation and slow excitation, it is fair to assume that $([A_j] - [A_j^*]) \sim [A_j]$.

In this regime, the stationary solution for $[D_i^*]$ can be written as:

$$[D_i^*] = \frac{k_d^X \cdot [D_i]}{\frac{k_d^F}{Q_d} \cdot (1 + F_j) + k_d^X},$$

where $F_i = \sum_j \left({R_0}/{R_{ij}}\right)^6 [A_j]$. The total intensity in the donor fluorescence channel is:

$$I_d^F = \sum_i k_d^F \cdot [D_i^*] = \sum_i \frac{k_d^F \cdot k_d^X \cdot [D_i]}{\frac{k_d^F}{Q_d} \cdot (1 + F_i) + k_d^X}.$$

In the limit of rapid de-excitation and slow excitation, $k_d^F \gg k_d^X$:

$$I_d^F = Q_d \cdot k_d^X \cdot \sum_i \frac{1}{1 + F_i} [D_i].$$

The stationary solution for $[A_j^*]$ is:

$$[A_j^*] = \frac{k_a^X \cdot [A_j] + \sum_i \frac{\frac{k_d^F}{Q_d} \cdot \left(\frac{R_0}{R_{ij}}\right)^6 \cdot k_d^X \cdot [D_i] \cdot [A_j]}{\frac{k_d^F}{Q_d} \cdot (1 + F_i) + k_d^X}}{\frac{k_a^F}{Q_a} + k_a^X},$$

and the total intensity in the acceptor fluorescence channel is:

$$I_a^F = \sum_j k_a^F \cdot [A_j^*] = \frac{\sum_j k_a^F \cdot \left\{ k_a^X \cdot [A_j] + \sum_i \frac{\frac{k_d^F}{Q_d} \cdot \left(\frac{R_0}{R_{ij}}\right)^6 \cdot k_d^X \cdot [D_i] \cdot [A_j]}{\frac{k_d^F}{Q_d} \cdot (1 + F_i) + k_d^X} \right\}}{\frac{k_a^F}{Q_a} + k_a^X}.$$

In the limit of rapid de-excitation and slow excitation, $k_d^F \gg k_d^X$ and $k_a^F \gg k_a^X$:

$$I_a^F = Q_a \cdot \left\{ k_a^X \cdot [A] + k_d^X \cdot \sum_i \frac{F_i}{1 + F_i} [D_i] \right\},$$

where $[A]$ is the total concentration of acceptors.

Derivation of the forward model

Using the microscopic quantities defined above, the forward model for $FRET_R$ can be written as:

$$f(X) = \frac{I_{\mathrm{FRET}}}{S_d \cdot I_{\mathrm{CFP}} + S_a \cdot I_{\mathrm{YFP}}} = \frac{\frac{S_d}{1+S_d+\Lambda_d} \cdot Q_d \cdot k_d^{X,430} \cdot \sum_i \frac{1}{1+F_i}[D_i] + \frac{1}{1+\Lambda_a} \cdot Q_a \cdot \left\{ k_a^{X,430} \cdot [A] + k_d^{X,430} \cdot \sum_i \frac{F_i}{1+F_i}[D_i] \right\}}{\frac{S_d}{1+S_d+\Lambda_d} \cdot Q_d \cdot k_d^{X,430} \cdot \sum_i \frac{1}{1+F_i}[D_i] + S_a \cdot \frac{1}{1+\Lambda_a} \cdot Q_a \cdot k_a^{X,500} \cdot [A]} = 1 + \frac{\frac{Q_a \cdot k_d^{X,430}}{1+\Lambda_a} \cdot \sum_i \frac{F_i}{1+F_i}[D_i]}{\frac{S_d \cdot Q_d \cdot k_d^{X,430}}{1+S_d+\Lambda_d} \cdot \sum_i \frac{1}{1+F_i}[D_i] + \frac{Q_a \cdot k_a^{X,430}}{1+\Lambda_a}[A]}. \quad (S3)$$

Eq. S3 can be simplified by introducing the ratio I_{da} of fluorescence intensities in the FRET channel when CFP and YFP are expressed alone and at equal concentrations:

$$I_{da} = \frac{I_d^F(\lambda_{em}=535\mathrm{nm}, \lambda_{ex}=430\mathrm{nm})}{I_a^F(\lambda_{em}=535\mathrm{nm}, \lambda_{ex}=430\mathrm{nm})} = \frac{S_d \cdot Q_d \cdot k_d^{X,430}}{1+S_d+\Lambda_d} \cdot \frac{1+\Lambda_a}{Q_a \cdot k_a^{X,430}}. \quad (S4)$$

This quantity can be calculated from the data collected in the two experiments measuring the spillover factors S_d and S_a. After substituting I_{da} from Eq. S4 in Eq. S3, the forward model can be written as:

$$f(X, I_{da}, k_{da}) = 1 + \frac{k_{da} \cdot \{[D] - g(X)\}}{I_{da} \cdot g(X) + [A]},$$

where $k_{da} = {k_d^{X,430}}\Big/{k_a^{X,430}}$ and $g(X) = \sum_i \frac{1}{1+F_i}[D_i]$.

Model of linker flexibility. Typically, fluorescent proteins are attached to the N- or C-terminus of a protein by means of a flexible linker. As a consequence, conformational heterogeneity in the complex can be present even when the conformations, positions, and orientations of the two tagged proteins remain relatively fixed. In this situation, the multi-state forward model of Eq. 4 (Online Methods**)** is still applicable, provided that the fluorescent proteins and linkers are explicitly represented. However, in the single-molecule case and when the structures of the tagged proteins are known and kept rigid, the forward model can be extended to account directly for the linker flexibility, without representing the fluorescent proteins explicitly. The probability distribution $p(r_{FT})$ as a function of the distance r_{FT} between a fluorophore and the terminal residue of the linker was first calculated with MD simulations of a GFP and linker alone (below). Each protein of the complex was then taken individually and probes representing possible fluorophore positions were distributed close to either the N or C terminus, according to $p(r_{FT})$. Probes that clashed with the protein structure were not retained. Finally, the probability of having the fluorophore in position x_F when the tagged protein is in state X was estimated by fitting the positions of the probes with a Gaussian Mixture Model (GMM):

$$p(x_F|X) = \sum_{i=1}^{N_G} \omega_i \, G(x_F|x_i, \sigma_i),$$

where G is a normalized Gaussian centered in x_i with standard deviation equal to σ_i, ω_i is the weight of the *i*-th GMM component, N_G is the number of components, and $\sum_{i=1}^{N_G} \omega_i = 1$. When calculating $FRET_R$ for a pair of tagged proteins in a given position (X^d, X^a), we assume that the two attached fluorescent proteins populate all the positions allowed by the linkers. The average over multiple conformations of Eq. 4 (Online Methods**)** can thus be written as:

$$\sum_k w_k g(X_k) = \int dx_F^d \; p\left(x_F^d \middle| X^d\right) \int dx_F^a \; p(x_F^a | X^a) \; \frac{1}{1 + \left({R_0}/{\| \; x_F^d - x_F^a \; \|}\right)^6}$$

$$= \sum_{i=1}^{N_G^d} \omega_i \sum_{j=1}^{N_G^a} \omega_j \int dr_F \; p(r_F | r_{ij}, \sigma_{ij}) \frac{1}{1 + \left({R_0}/{r_F}\right)^6},$$

where $r_F = \| \; x_F^d - x_F^a \; \|$, $r_{ij} = \| \; x_i - x_j \; \|$, $\sigma_{ij} = \sqrt{\sigma_i^2 + \sigma_j^2}$, and [2,3]

$$p\left(r_F \middle| r_{ij}, \sigma_{ij}\right) = \frac{r_F}{\sqrt{2\pi} \; \sigma_{ij} \; r_{ij}} \exp\left(-\frac{r_F^2 \; + r_{ij}^2}{2\sigma_{ij}^2}\right) \sinh\left(\frac{r_F \; r_{ij}}{\sigma_{ij}^2}\right).$$

Provided that the tagged proteins are kept rigid, the positions of the GMM centers $\{x_i\}$ and $\{x_j\}$ can be transformed along with the coordinates X^d and X^a and no further GMM fit is required during sampling. Furthermore, the integral in the above equation was evaluated numerically and tabulated for different values of r_{ij} and σ_{ij} prior to sampling.

Multi-state Bayesian scoring function. The multi-states posterior probability can be written as:

$$p(\{X_k, w_k\}, I_{da}, k_{da}, \{\sigma_n\}|\{d_n\})$$
$$\propto p\left(I_{da}\middle|I_{da}^{exp}, \sigma_{da}^{exp}\right)p(k_{da})\prod_{k=1}^{N}p(X_k)p(w_k)\prod_{n=1}^{N_F}p(\sigma_n|\sigma_0)p(d_n|\{X_k, w_k\}, I_{da}, k_{da}, \sigma_n).$$

Finally, to reduce the number of parameters, it is useful to marginalize all σ_n:

$$p(\{X_k, w_k\}, I_{da}, k_{da}, \sigma_0|\{d_n\})$$
$$\propto p\left(I_{da}\middle|I_{da}^{exp}, \sigma_{da}^{exp}\right)p(k_{da})\prod_{k=1}^{N}p(X_k)p(w_k)\prod_{n=1}^{N_F}\int d\sigma_n\, p(\sigma_n|\sigma_0)p(d_n|\{X_k, w_k\}, I_{da}, k_{da}, \sigma_n)$$
$$= p\left(I_{da}\middle|I_{da}^{exp}, \sigma_{da}^{exp}\right)p(k_{da})\prod_{k=1}^{N}p(X_k)p(w_k)\prod_{n=1}^{N_F}p(d_n|\{X_k, w_k\}, I_{da}, k_{da}, \sigma_0),$$

where the marginal likelihood is:

$$p(d_n|\{X_k, w_k\}, I_{da}, k_{da}, \sigma_0) = \frac{\sqrt{2}\sigma_0}{\pi d_n}\cdot\frac{1}{\log(d_n/f(\{X_k, w_k\}, I_{da}, k_{da}))^2 + 2\sigma_0^2}.$$

Supplementary Experimental Methods

In vivo $FRET_R$ measurements *S. cerevisiae* strain BSY9 (MAT**a**/MATα , ade2-1oc/ade2-1oc, ADE3/ade3Δ100, can1-100/can1-100, CYH2s/cyh2r, his3-11,15/his3-11,15, leu2-3,112/leu2-3,112, trp1-1/ trp1-1, ura3-1/ura3-1) was the host of a series of plasmids that were integrated at either the URA3 or LEU2 locus. Descriptions of the plasmids, along with their Genbank accession numbers, are given in supplemental Table S1. All plasmids are yeast integrative plasmids based on either pRS305 (GenBank: U03437.1) or pRS306 (GenBank: U03438.1).

Microscopy and FRET analysis were performed as described in [4]. In brief, images were captured using a Deltavision microscope equipped with CFP/YFP filter set 89002-ET from Chroma Technology Corp and a mercury HBO 100W light source. Independent emission and excitation filter wheels were used to position filters. Images were captured on a Coolsnap HQ camera with 0.08 or 0.1 second exposure times. The order of image acquisition is important and all images were captured in the sequence YFP, then FRET and finally the CFP channel. A DIC image was captured at the end of the experiment to provide cellular context to the fluorescence images. Images were then analyzed for FRET using Fretscal. Fretscal is an integrated set of custom Matlab scripts and is available online at no cost at the MATLAB Central file exchange, http://www.mathworks.com/matlabcentral/.

References

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

2 Churchman, L. S., Flyvbjerg, H. & Spudich, J. A. A non-Gaussian distribution quantifies distances measured with fluorescence localization techniques. *Biophys J* **90**, 668-671, (2006).

3 Kohler, S. D., Spitzbarth, M., Diederichs, K., Exner, T. E. & Drescher, M. A short note on the analysis of distance measurements by electron paramagnetic resonance. *J Magn Reson* **208**, 167-170, (2011).

4 Kollman, J. M. *et al.* The structure of the gamma-tubulin small complex: implications of its architecture and flexibility for microtubule nucleation. *Mol Biol Cell* **19**, 207-215, (2008).

5 Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc Natl Acad Sci U S A* **99**, 12562-12566, (2002).

6 Barducci, A., Bonomi, M. & Parrinello, M. Metadynamics. *Wires Comput Mol Sci* **1**, 826-843, (2011).

7 Bussi, G., Gervasio, F. L., Laio, A. & Parrinello, M. Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics. *J Am Chem Soc* **128**, 13435-13441, (2006).

8 Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* **314**, 141-151, (1999).

9 Bonomi, M., Barducci, A. & Parrinello, M. Reconstructing the Equilibrium Boltzmann Distribution from Well-Tempered Metadynamics. *J Comput Chem* **30**, 1615-1621, (2009).

Figure Legends

Fig. S1. Bayesian model of $FRET_R$ data. (A) $FRET_R$ calculated from the forward model of Eq. 1 (Online Methods) as a function of the distance between donor and acceptor. $FRET_R$ is calculated for different values of the ratio between donor and acceptor excitation rates $\boldsymbol{k_{da}}$ and the ratio of donor and acceptor intensities in the FRET channel $\boldsymbol{I_{da}}$, measured when CFP and YFP are expressed separately. (B) Negative logarithm of the marginal likelihood function (Eq. 7 Online Methods) for 3 values of the uncertainty $\boldsymbol{\sigma_0}$, plotted against the $FRET_R$ value of data point d_n and given a prediction by the forward model of a $FRET_R$ value of 2.0. (C) The unimodal distribution used as the prior for uncertainty $\boldsymbol{\sigma_n}$ (Eq. 6 Online Methods) for 3 values of $\boldsymbol{\sigma_0}$.

Fig. S2. Accuracy of the $FRET_R$ forward model. The relative deviation of the $FRET_R$ forward model values from KMC simulations is calculated as a function of the parameters k_d^X/k_d^F and k_{da}, for systems of (A) one CFP-one YFP, (B) two CFP-one YFP, and (C) one CFP-two YFP. The relative deviation is averaged over a distance range from 3 to 10 nm. The YFP photobleaching ratio during the acquisition of the CFP channel is set to 0.3.

Fig. S3. $FRET_R$ from KMC simulations of systems of (A) two CFP-one YFP and (B) one CFP-two YFP, as a function of the distances between fluorophores. k_{da} is set to 8.0, k_d^X/k_d^F to 0.001. The YFP photobleaching ratio during the acquisition of the CFP channel is set to 0.3. (C) The ratio of $FRET_R$ of the two CFP-one YFP to the one CFP-two YFP systems, calculated from the forward model with donor-acceptor distances set both to R_0, as a function of k_{da} and I_{da}.

Fig. S4. Effect of YFP photobleaching on $FRET_R$. $FRET_R$ calculated from KMC simulations of a system of one CFP-two YFP as a function of the distances between fluorophores, when (A) YFP are not photobleached during the acquisition of the CFP channel, and (B) the YFP photobleaching ratio is equal to 0.3. The values of k_{da} and

k_d^X/k_d^F are the same as in Fig. S3. (C) $FRET_R$ of the photobleached system calculated from the multi-state forward model of Eq. 4 (Online Methods**)**. The average relative deviation from the KMC simulations of panel (B) is 0.8%**.**

Fig. S5. Free energies of a series of CFP-YFP pairs separated by polyproline peptides of length 5 (A), 10 (B), 15 (C), and 20 (D). The free energies are calculated by all-atom molecular dynamics simulations as a function of $\Omega = \sum_{i=1}^{n-1} \cos \omega_i$, where ω is the torsional angle formed by the proline quadruplet Cα-C-N-Cα. ω is equal to 0˚ and 180˚ for the *cis* and *trans* proline isomers, respectively. For a peptide of length *n*, Ω=-(*n*-1) indicates an ideal left-handed polyproline II helix, Ω=*n*-1 an ideal right-handed polyproline I helix. Sampling is accelerated by a combination of Parallel Tempering and metadynamics[5-7].

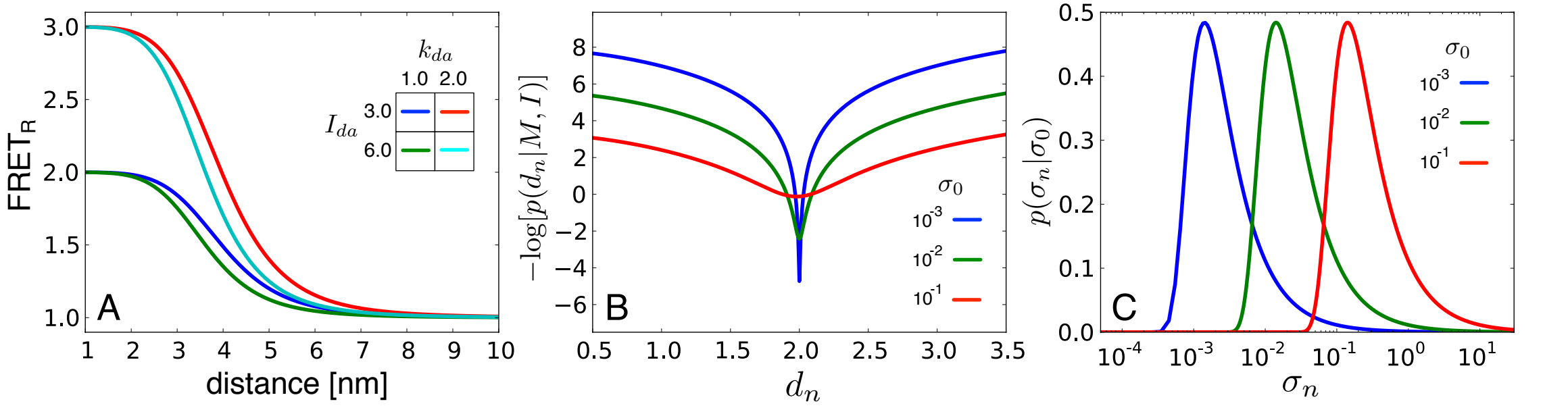
Fig. S6. Free energy of a CFP-YFP pair separated by a flexible linker as a function of the distance between the centers of the two fluorophores (x axis) and the angle formed by the major axes of inertia of the two fluorescent proteins (y axis). The free energy is calculated by all-atom molecular dynamics simulations, accelerated by Parallel Tempering[8]. Isoenergy lines are drawn every 1 k_BT.

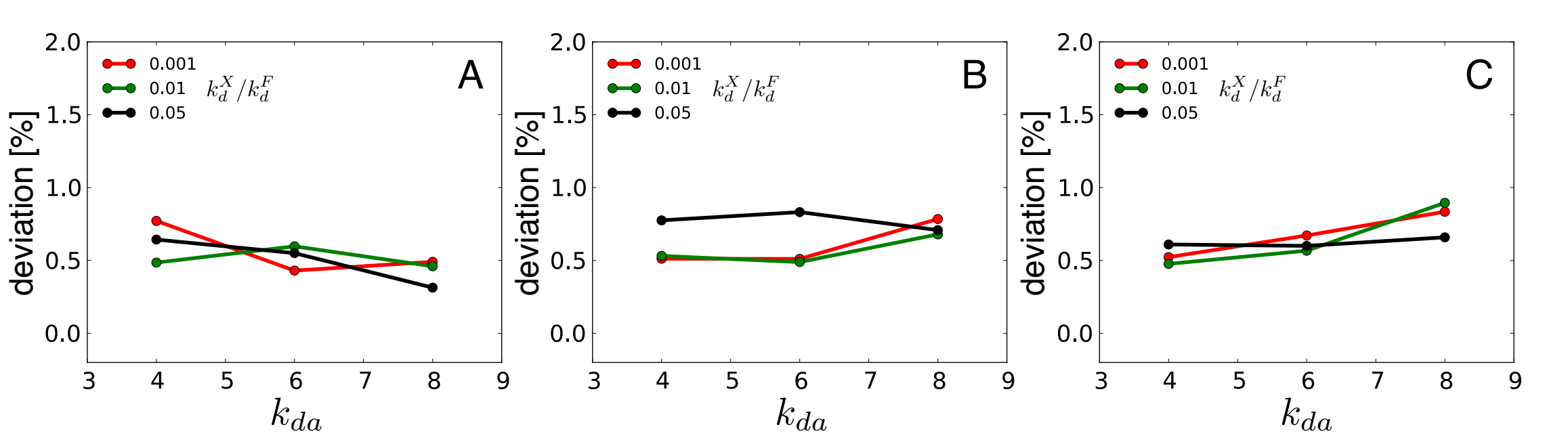
Fig. S7. Assessment of sampling convergence. (AB) Convergence of the MD simulations of YFP–CFP pairs separated by (A) polyproline peptides of different length and (B) a flexible linker or other proteins of known structure. The weighted average $\langle g(X) \rangle$ needed by the multi-state forward model of Eq. 4 (Online Methods**)** was calculated as the ensemble average of $g(X)$ over the MD conformers, with R_0=4.9nm, after discarding the first 10% of the trajectory. To assess the convergence, $\langle g(X) \rangle$ is calculated and plotted as a function of the simulation time. A reweighting algorithm[9] was used to obtain from the biased PTMetaD simulations canonical averages. (CD) Convergence of the benchmark accuracy as a function of the number of tests. Accuracy is defined as the average Cα dRMS between the crystallographic structure and the most

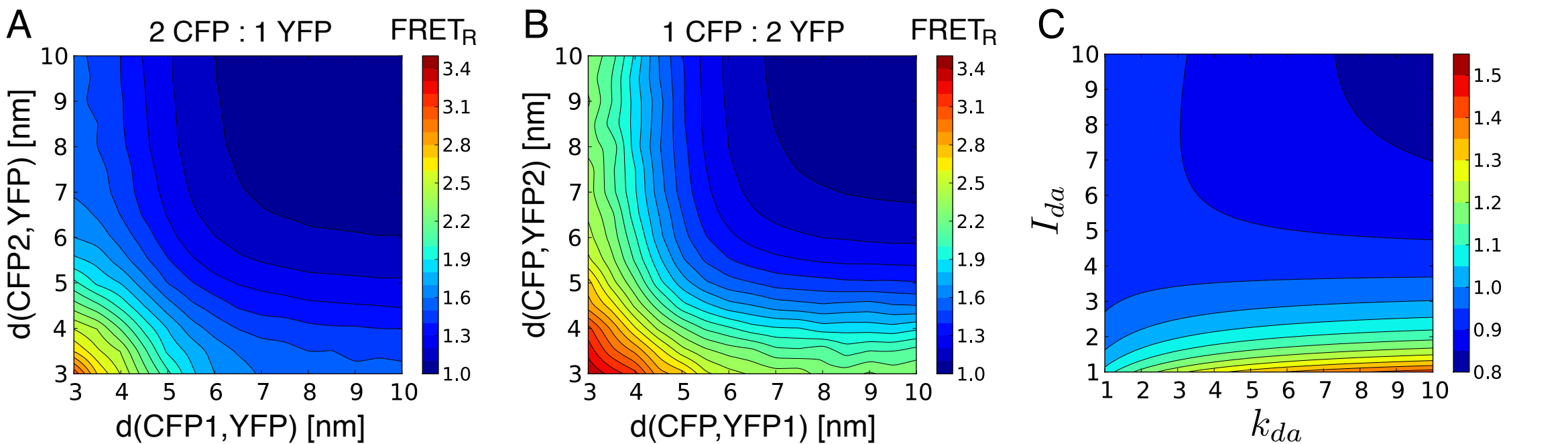
probable model, calculated (C) on the entire complex and (D) on the N- and C-terminal residues. Results are with low-noise data and using 100% of the data points.

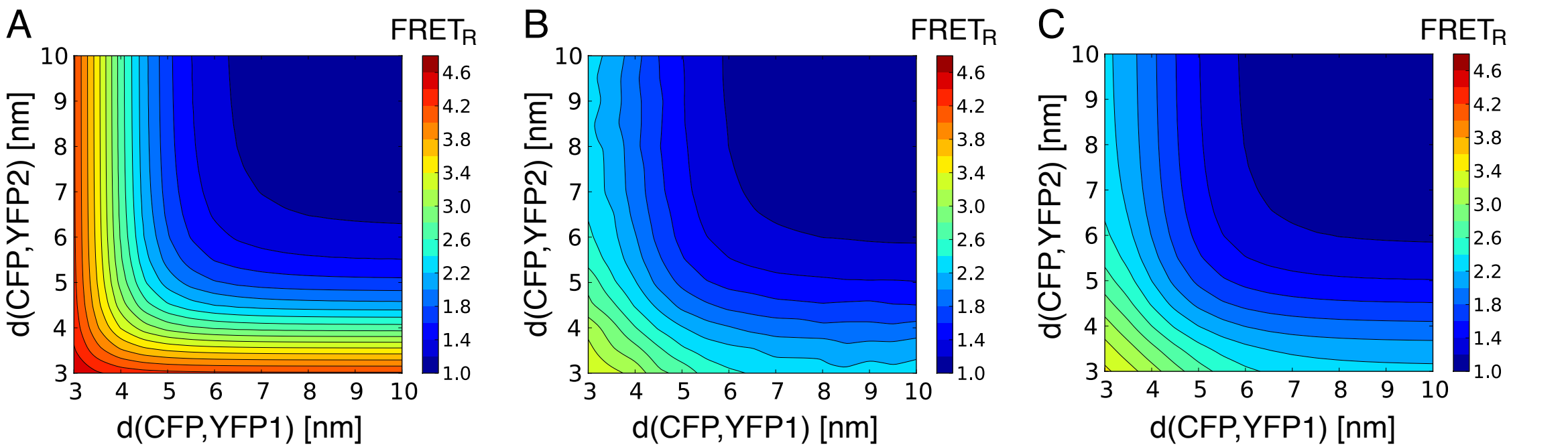
Fig. S8. *In vivo* experimental validation of the forward model. $FRET_R$ values measured on 5 systems of defined structure (black bars) expressed in *S. cerevisiae* are compared to the values predicted by the forward model (white bars), using the model of linker flexibility. The other 4 data points of Fig. 3 (polyproline series) were not used for this comparison, because in these systems heterogeneity was dominated by the presence of conformations with different patterns of proline isomers, rather than the linker flexibility. Red lines indicate experimental and model errors.

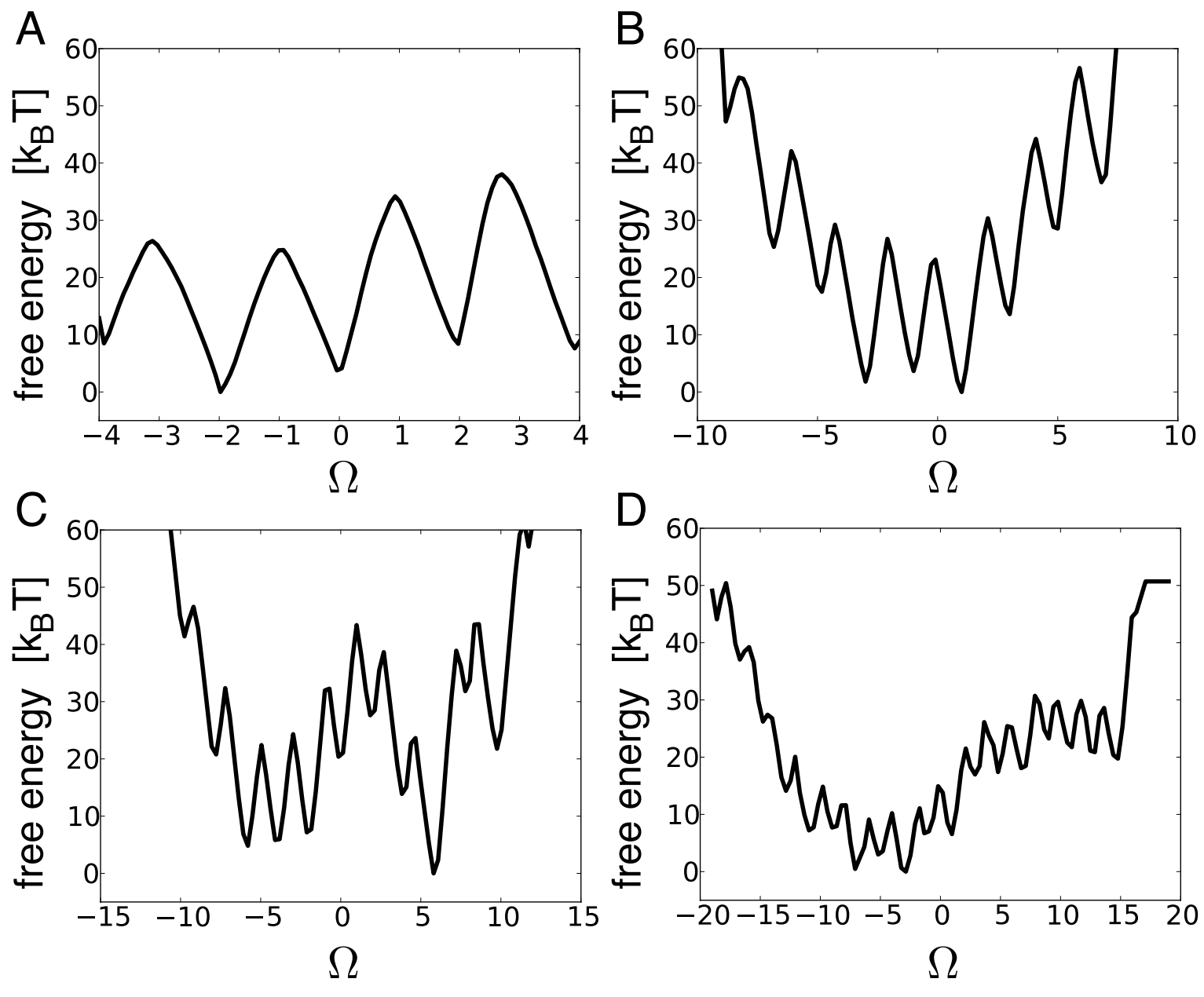
Fig. S9. Benchmark flowchart. (A) A flowchart for benchmarking the accuracy of our Bayesian approach to determining the molecular architecture of a complex from synthetic $FRET_R$ data. The benchmark contains 16 protein complexes. The stages include target selection, system representation, synthetic data generation, sampling, and analysis. (B) Schematics of the sampling algorithm based on a Gibbs sampling Monte Carlo scheme coupled with Simulated Annealing. The flowchart depicts a typical step of our sampling scheme in which random changes of the position and orientation of each subunit, the values of the forward model, and likelihood parameters are proposed, and either rejected or accepted based on the Metropolis criterion. The temperature of the system is cyclically varied from 1.0 to 5.0 $k_B T$ to avoid trapping the system in local energy minima, thus enhancing the sampling.

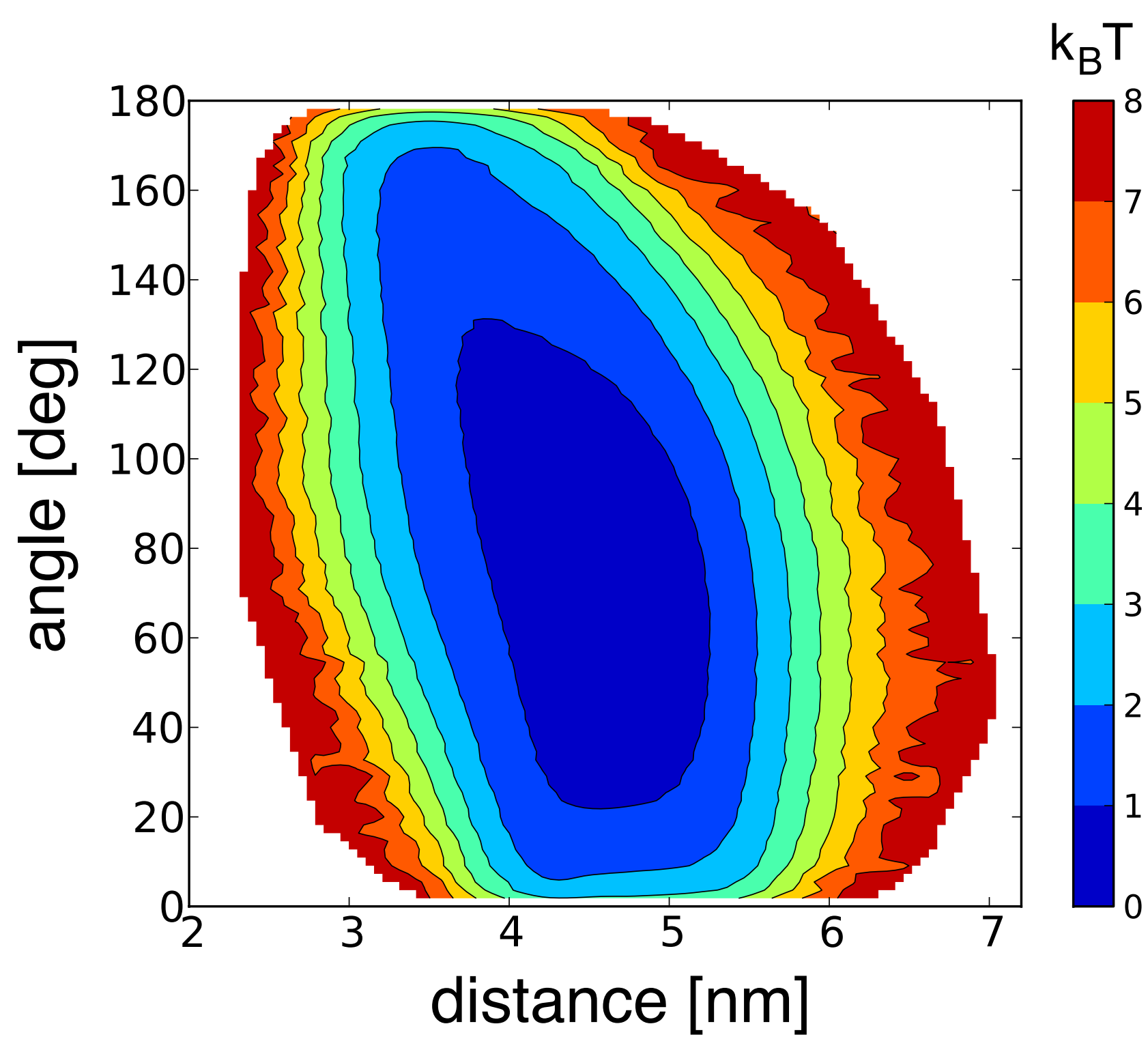

FRET_R
distance [nm]
A
k_{da}
1.0 2.0
I_{da}
3.0
6.0
$-\log[p(d_n|M,I)]$
d_n
B
σ_0
10^{-3}
10^{-2}
10^{-1}
$p(\sigma_n|\sigma_0)$
σ_n
C
σ_0
10^{-3}
10^{-2}
10^{-1}

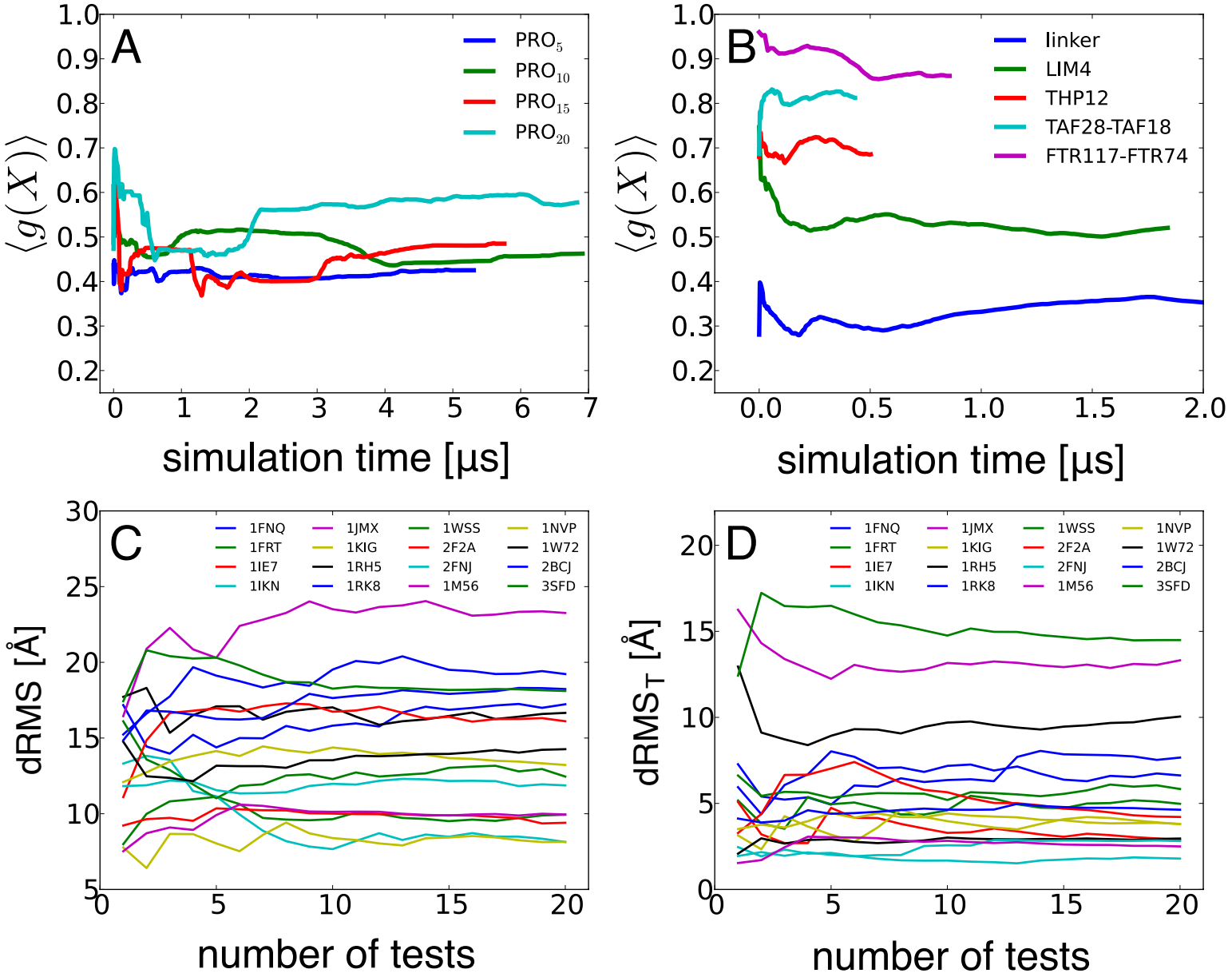

A
B
C
0.001
0.01
0.05
k_d^X/k_d^F
deviation [%]
k_{da}
2.0
1.5
1.0
0.5
0.0
3
4
5
6
7
8
9

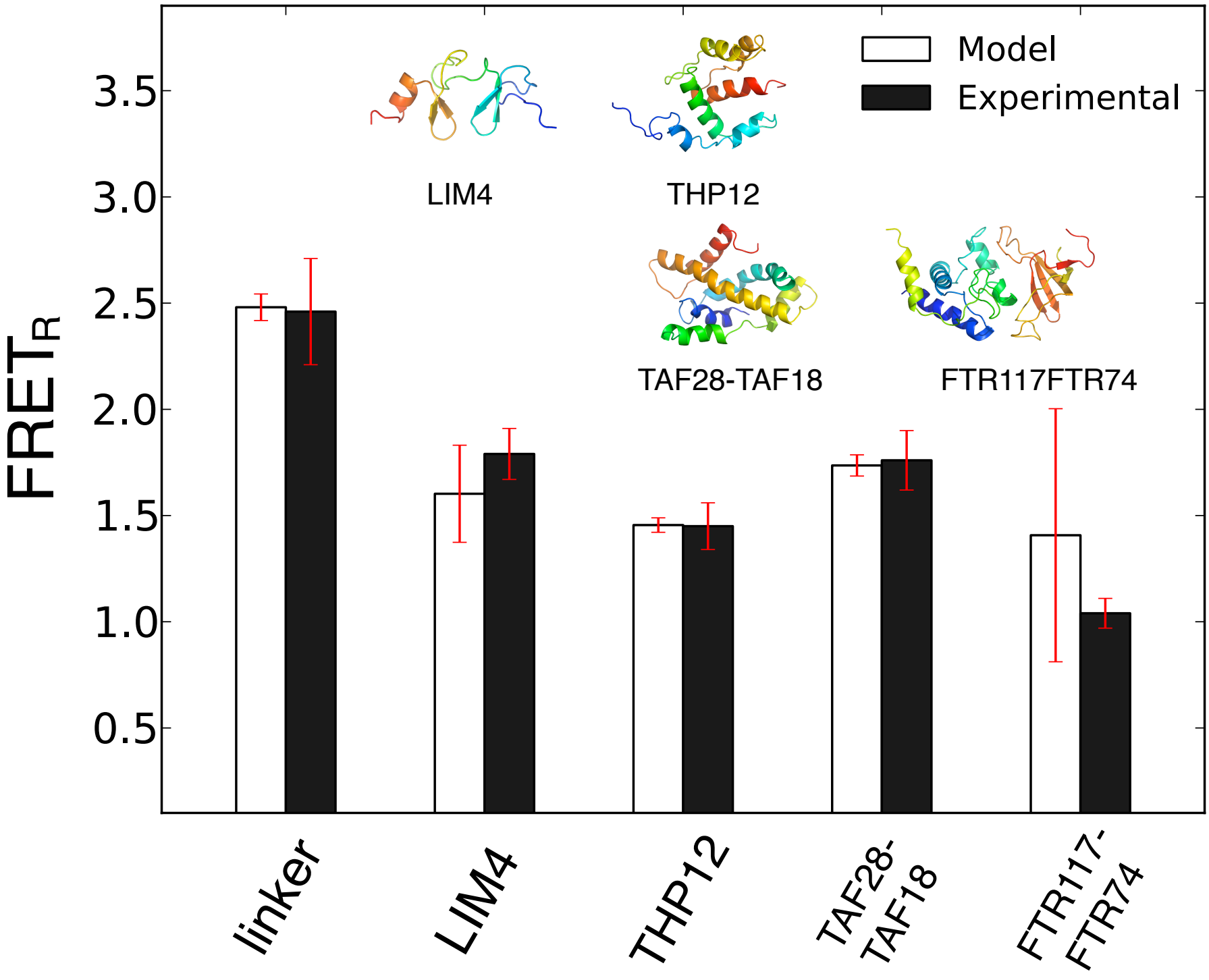

A
2 CFP : 1 YFP
FRET$_R$
d(CFP2,YFP) [nm]
d(CFP1,YFP) [nm]
B
1 CFP : 2 YFP
FRET$_R$
d(CFP,YFP2) [nm]
d(CFP,YFP1) [nm]
C
I_{da}
k_{da}


A
B
C
FRET$_R$
4.6
4.2
3.8
3.4
3.0
2.6
2.2
1.8
1.4
1.0
d(CFP,YFP2) [nm]
d(CFP,YFP1) [nm]
3
4
5
6
7
8
9
10


A
B
C
D
free energy [k_BT]
Ω
60
50
40
30
20
10
0
−4 −3 −2 −1 0 1 2 3 4
−10 −5 0 5 10
−15 −10 −5 0 5 10 15
−20 −15 −10 −5 0 5 10 15 20


k_BT
8
7
6
5
4
3
2
1
0
180
160
140
120
100
80
60
40
20
0
angle [deg]
2
3
4
5
6
7
distance [nm]


A
PRO5
PRO10
PRO15
PRO20
⟨g(X)⟩
simulation time [μs]
B
linker
LIM4
THP12
TAF28-TAF18
FTR117-FTR74
⟨g(X)⟩
simulation time [μs]
C
1FNQ
1FRT
1IE7
1IKN
1JMX
1KIG
1RH5
1RK8
1WSS
2F2A
2FNJ
1M56
1NVP
1W72
2BCJ
3SFD
dRMS [Å]
number of tests
D
dRMST [Å]
number of tests


FRET$_R$
3.5
3.0
2.5
2.0
1.5
1.0
0.5
Model
Experimental
LIM4
THP12
TAF28-TAF18
FTR117FTR74
linker
LIM4
THP12
TAF28-TAF18
FTR117-FTR74

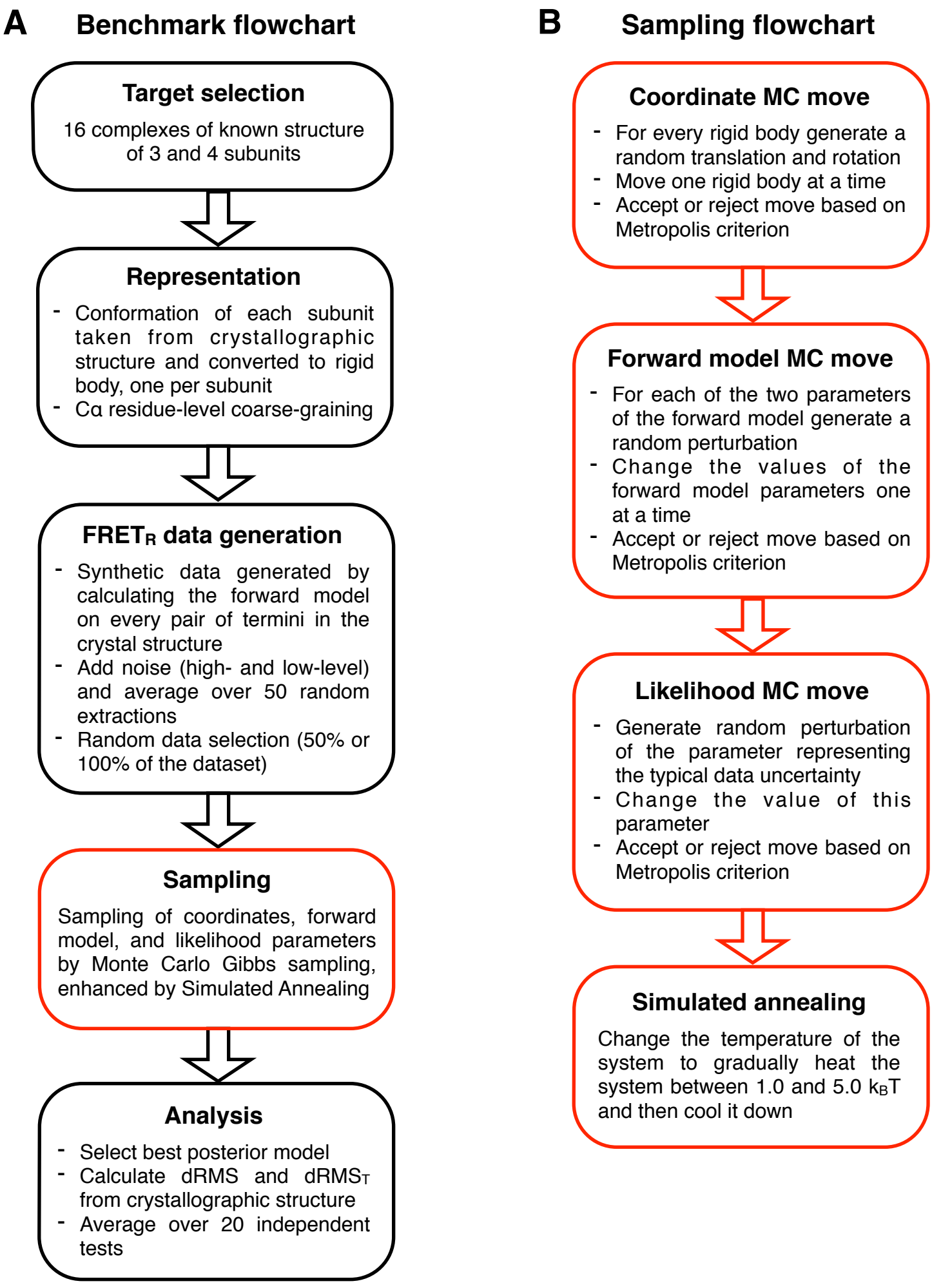

A
Benchmark flowchart
Target selection
16 complexes of known structure of 3 and 4 subunits
Representation
- Conformation of each subunit taken from crystallographic structure and converted to rigid body, one per subunit
- Cα residue-level coarse-graining
FRET$_R$ data generation
- Synthetic data generated by calculating the forward model on every pair of termini in the crystal structure
- Add noise (high- and low-level) and average over 50 random extractions
- Random data selection (50% or 100% of the dataset)
Sampling
Sampling of coordinates, forward model, and likelihood parameters by Monte Carlo Gibbs sampling, enhanced by Simulated Annealing
Analysis
- Select best posterior model
- Calculate dRMS and dRMS$_T$ from crystallographic structure
- Average over 20 independent tests
B
Sampling flowchart
Coordinate MC move
- For every rigid body generate a random translation and rotation
- Move one rigid body at a time
- Accept or reject move based on Metropolis criterion
Forward model MC move
- For each of the two parameters of the forward model generate a random perturbation
- Change the values of the forward model parameters one at a time
- Accept or reject move based on Metropolis criterion
Likelihood MC move
- Generate random perturbation of the parameter representing the typical data uncertainty
- Change the value of this parameter
- Accept or reject move based on Metropolis criterion
Simulated annealing
Change the temperature of the system to gradually heat the system between 1.0 and 5.0 k_BT and then cool it down

Table S1. Plasmids and Strains

Plasmid	Fluorescent Protein[1]	Selective Marker[2]	Genbank Accession	Strain[3]
pBS41	CFP	URA3	KF177452	EMY200
pBS43	YFP	URA3	KF177455	EMY201
pBS49	YFP	LEU2	KF177459	EMY202
pBS42BN[4]	YFP-Pro0-CFP	URA3	KF177454	EMY203
pBS46	YFP-Pro5-CFP	URA3	KF177456	EMY204
pBS47	YFP-Pro10-CFP	URA3	KF177457	EMY205
pBS48	YFP-Pro15-CFP	URA3	KF177458	EMY206
pBS50	YFP-Pro20-CFP	URA3	KF177460	EMY207
pFTR74-YFP	FTR74-YFP	LEU2	KF177461	EMY208
pFTR117-CFP	FTR117-CFP	URA3	KF177462	EMY208
pYFP-TAF28-CFP	YFP-TAF28-CFP	URA3	KF177465	EMY209
pTAF18[5]		LEU2	KF177463	EMY209
pYFP-Lim4-CFP	YFP-Lim4-CFP	URA3	KF177464	EMY210
pYFP-THP112-CFP	YFP-THP112-	URA3	KF177466	EMY211

[1] Expression of all proteins is driven by the TEF promoter. All proteins have an N-terminal nuclear localization signal.

[2] Plasmids were integrated into strain BSY9 (MATa/MATα ade2-1oc/ade2-1oc ADE3/ade3?100 can1-100/can1-100 CYH2s/cyh2r his3-11,15/his3-11,15 leu2-3,112/leu2-3,112 trp1-1/ trp1-1 ura3-1/ura3-1) at URA3 by digesting the plasmids with StuI, and integrated at LEU2 by digestion with BstEII. Since the Lim4 gene contains a StuI site, pYFP-Lim4-CFP was digested with ApaI instead.

[3] Plasmids were integrated at the selective marker.

[4] Contains BamHI and NheI sites between YFP and CFP genes.

[5] TAF18 is not tagged, but co-expressed with YFP-TAF28-CFP.

Table S2. Results for the benchmark set of complexes modeled using simulated $FRET_R$ data, with linker flexibility.

Sub units	Noise	% Data	dRMS [Å]	$dRMS_T$ [Å]	Placement score [Å, °]
3	low	50	20.8	20.0	22.4, 97
		100	13.9	5.2	16.1, 76
	high	50	20.5	20.1	22.4, 100
		100	16.1	9.3	18.6, 87
4	low	50	20.4	17.6	26.7, 103
		100	14.8	7.1	19.2, 82
	high	50	21.5	20.2	26.7, 102
		100	17.4	11.6	23.3, 96

Column 1 shows the number of subunits. The data noise (Column 2) is determined by the uncertainty $\boldsymbol{\sigma_0}$, which is set to 0.001 and 0.01 for low- and high-noise data, respectively. The data sparseness (Column 3) is defined as the percentage of all pairwise combinations of protein termini tagged with fluorescent proteins. The structural accuracy is defined as the average Cα dRMS deviation of the best-scoring posterior model from the native structure, calculated on the entire structure (Column 4) and on the N- and C-terminal residues (Column 5), by the average placement score (26) of the assembly subunits of the best posterior model (Column 6). The placement score for a subunit is defined as the (placement distance in Å, placement angle in degrees) pair, resulting from comparing the model with the native configuration, in the least-squared superposition of all assembly atoms. The placement distance is given by the subunit centroid distance, and the placement angle is the screw rotation that superposes the model subunit on the native subunit.

Table S3. Results for the benchmark set of complexes modeled using simulated $FRET_R$ data, without linker flexibility.

Sub units	Noise	% Data	dRMS [Å]	$dRMS_T$ [Å]	Placement score [Å, °]
3	low	50	18.1	16.5	20.1, 100
		100	13.9	3.8	14.5, 79
	high	50	18.3	16.4	20.4, 102
		100	14.7	5.6	16.5, 86
4	low	50	18.9	15.0	24.6, 97
		100	15.3	6.7	22.5, 86
	high	50	20.5	19.3	24.2, 103
		100	16.4	9.9	23.1, 92

Columns as in Table S2.

Table S4. Table of symbols used in the equations.

Symbol	Description
X	Set of N modeled structures $\{X_k\}$.
X_k	Individual conformation of multiple donors and acceptors in a sample.
w_k	Proportion of an individual conformation, *k*, in the total population.
k_{da}	Ratio of the excitation rates of donor to acceptor at the excitation wavelength of donor.
I_{da}	Ratio of CFP and YFP fluorescence intensities in the FRET channel when each is expressed at equal levels in separate cells at the excitation wavelength of donor.
I_{da}^{exp}	Measured value of I_{da}.
$\sigma_{I_{da}^{exp}}$	Standard deviation of the measured I_{da}^{exp}.
k_{ij}^{ET}	Rate of energy transfer between donor *i* and acceptor *j.*
k^F	Rates of fluorescence. Subscripts “d” or “a” refer to donor or acceptor.
k^X	Rates of excitation. Subscripts “d” or “a” refer to donor or acceptor.
F_i	Term that quantifies the energy transferred from donor *i* to all the acceptors.
R_0	Forster radius.
R_{ij}	Distance between donor *i* and acceptor *j.*
$g(X)$	A function describing the impact of energy transfer on donor fluorescence.
[*D*]	Concentration of donor.
[*A*]	Concentration of acceptor.
k_d^{XF}	Ratio of the rate of donor excitation to the rate of donor fluorescence at the excitation wavelength of donor.
Q	Quantum yield. Subscripts “d” or “a” refer to donor or acceptor.
I^F	Fluorescence intensity. Subscripts “d” or “a” refer to donor or acceptor.
d_n	Data point *n.*
σ_n	Uncertainty for data point *n.*
σ_0	Typical uncertainty of dataset $\{d_n\}$.
S	Spillover factor. Subscripts “d” or “a” refer to donor or acceptor.

Table S5. Table of the main equations.

Key Equations	Description
$\mathrm{FRET_R} = \frac{I_{\mathrm{FRET}}}{S_{\mathrm{tot}}}$	Experimental definition of $\mathrm{FRET_R}$.
$p(M\|D,I) \propto p(D\|M,I) \cdot p(M\|I)$	Posterior probability from Bayes theorem.
$f(X, I_{da}, k_{da}) = 1 + \frac{k_{da} \cdot \{[D] - g(X)\}}{I_{da} \cdot g(X) + [A]}$	Microscopic model of $\mathrm{FRET_R}$. Referred to as the $\mathrm{FRET_R}$ forward model.
$f(\{X_k, w_k\}, I_{da}, k_{da}) = 1 + \frac{k_{da} \cdot \{[D] - \langle g(X)\rangle\}}{I_{da} \cdot \langle g(X)\rangle + [A]}$	Solution of $\mathrm{FRET_R}$ generalized to take into account multiple states. Referred to as the (multi-state) $\mathrm{FRET_R}$ forward model.
$p(d_n\|\{X_k, w_k\}, I_{da}, k_{da}, \sigma_n) = \frac{1}{d_n\, \sigma_n \sqrt{2\pi}} \cdot \exp\left[-\frac{\log^2(d_n / f(\{X_k, w_k\}, I_{da}, k_{da}))}{2\sigma_n^2}\right]$	Data likelihood for data point *n*.
$p(\sigma_n\|\sigma_0) = \frac{2\sigma_0}{\sqrt{\pi}\sigma_n^2} \exp\left(-\frac{\sigma_0^2}{\sigma_n^2}\right)$	Prior distribution for the uncertainty σ_n associated to data point *n*.
$p(d_n\|\{X_k, w_k\}, I_{da}, k_{da}, \sigma_0) = \frac{\sqrt{2}\sigma_0}{\pi d_n} \cdot \frac{1}{\log(d_n / f(\{X_k, w_k\}, I_{da}, k_{da}))^2 + 2\sigma_0^2}$	Marginal data likelihood for data point *n*.
$p(\{X_k, w_k\}, I_{da}, k_{da}, \sigma_0\|\{d_n\}) \propto p\left(I_{da}\middle\|I_{da}^{exp}, \sigma_{da}^{exp}\right) p(k_{da}) \prod_{k=1}^{N} p(X_k) p(w_k) \prod_{n=1}^{N_F} p(d_n\|\{X_k, w_k\}, I_{da}, k_{da}, \sigma_0)$	Multi-state posterior probability.