Data Article

Computational refinement of spectroscopic FRET measurements

Alexander Kyrychenko [a,b,*], Mykola V. Rodnin [b],
Chiranjib Ghatak [b], Alexey S. Ladokhin [b,**]

[a] Institute of Chemistry and School of Chemistry, V. N. Karazin Kharkiv National University, 4 Svobody Square,
Kharkiv 61022, Ukraine
[b] Department of Biochemistry and Molecular Biology, The University of Kansas Medical Center Kansas City,
KS66160-7421, USA

ARTICLE INFO

ABSTRACT

This article supplies raw data related to a research article entitled
"Joint refinement of FRET measurements using spectroscopic and
computational tools" (Kyrychenko et al., 2017) [1], in which we
demonstrate the use of molecular dynamics simulations to esti-
mate FRET orientational factors in a benchmark donor-linker-
acceptor system of enhanced cyan (ECFP) and enhanced yellow
(EYFP) fluorescent proteins. This can improve the recalculation of
donor-acceptor distance information from single-molecule FRET
measurements.

© 2017 Published by Elsevier Inc. This is an open access article
under the CC BY license
(http://creativecommons.org/licenses/by/4.0/).

Specifications Table

Subject area	*Chemistry, Biophysics*
More specific subject area	*Computational refinement of Förster resonant energy transfer*

Type of data	*Figure, text*
How data was acquired	*Computer simulations using molecular dynamics method*
Data format	*Raw, analyzed, processed*
Experimental factors	*Fluorescence measurement procedure, molecular dynamics simulation details*
Experimental features	*MD simulations details*
Data source location	*Kansas City, USA; Kharkiv, Ukraine*
Data accessibility	*Data given as Supplemental files in this article*

Value of the data

- Details and flowchart of MD simulations of conformational sampling of ECFP and EYFP proteins linked with a flexible unstructured Gly/Ser peptide are provided.
- MD simulation setup and force field parameterization of ECFP and EYFP are given.
- MD sampling performed in the absence and in the presence of 0.1 M buffering ions of Na^+ and Cl^- revealed that the increased ionic strength had no effect on the equilibrium distance distributions between EFCP and EYFP moieties.
- The distribution of the FRET orientational factor, κ^2, estimated from MD simulations, is found to correspond well to the theoretical curve for an ideal, unrestricted, isotropic re-orientation of dipoles.
- The comparison of the effect of the standard assumption about the unique value of the Förster radius R_0 on the accuracy of determination of the Donor-to-Acceptor distance distributions is presented, revealing that both *apparent* distributions calculated from either MD-generated or experimentally measured FRET efficiencies systematically overestimate the width of the *true* distance distribution R_{DA}.

1. Data

MD simulations for a benchmark donor-linker-acceptor system composed of enhanced cyan (ECFP) and enhanced yellow (EYFP) fluorescent proteins were used to estimate FRET orientational factors, which can improve the recalculation of donor–acceptor distance information from sm-FRET measurements. ECFP and EYFP proteins were linked with a flexible unstructured Gly/Ser peptide linker composed of 3 repeating units of GlyGlySerGlyGlySer (GGSGGS) as shown in Fig. 1.

1.1. MD sampling of ECFP-l_3-EYFP by multiple discrete MD simulations

Proper conformation sampling of l_3 linker is essential for the accurate analysis of MD simulations in ECFP-l_3-EYFP. In general, the convergence of the MD sampling of long-chain polypeptides is challenging, because they have multidimensional energy surfaces characterized by many local minima separated by potentially high free-energy barriers, which can lead to kinetic trapping. In our system, however, we did not expect MD sampling to become restricted to a series of localized metastable conformations, because the flexible linker connecting two rigid FPs is not expected to
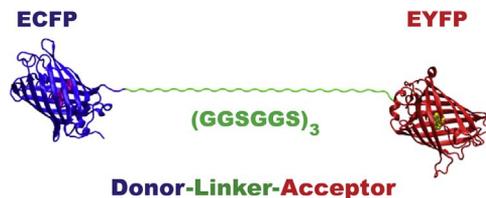

Fig. 1. Scheme of an ECFP-donor and EYFP-acceptor FRET pair linked by a flexible peptide bridge (GGSGGS)$_3$ referred to as ECFP-l_3-EYFP.

create high-energy barriers. Nevertheless, in order to ensure sufficient conformational MD sampling, we carried out a series of multiple discrete MD runs, each initiated with random initial structures. Such an approach has been shown to enhance conformational sampling [2–5]. To investigate the role of starting configurations in the equilibrium structure of EFCP-l_3-EYFP, we used a series of four independent parallel MD simulations, each of which was carried out with a different initial conformation of l_3 linker.

Fig. 2 show MD snapshots of the initial and final conformations of EFCP-l_3-EYFP for the four different MD runs referred to as runs 1–4. In runs 1–2, EFCP-l_3-EYFP was constructed as having the extended conformation of l_3, so that the initial value of the donor-to-acceptor distance R_{DA} was about 90 Å. In runs 3–4, the EFCP and EYFP moieties were set in different orientations to avoid local conformation trapping during MD sampling. The large-scale MD parallel simulations of runs 1–4 revealed that in all studied systems R_{DA} converged to some plateau at ~42–46 Å (Fig. 3). From these MD runs the last 100–200 ns were used for the analysis, during which the average orientation factor approached the theoretical value for unrestricted diffusion, confirming that the time range of our simulations is sufficient for the proper conformational sampling of the studied system.

1.2. Effect of ionic strength

Most MD samplings were carried out in water with sodium ions added to counterbalance the protein charge. To estimate the role of ionic strength in conformation equilibrium of EFCP-l_3-EYFP, the control MD simulation (100 ns) was also performed in the presence of 0.1 M buffering ions of Na^+
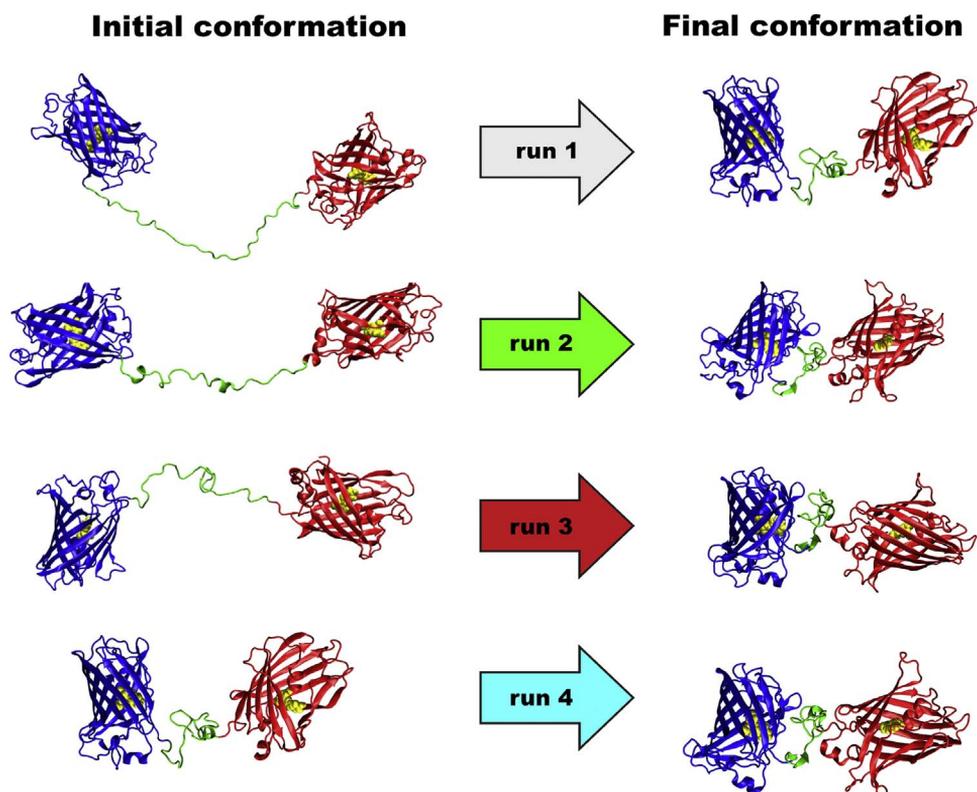

Fig. 2. (**A**) The series of the four different starting configurations of ECFP-l_3-EYFP (runs 1–4) were run in parallel to achieve adequate convergence statistics. (*Left*) MD snapshots of the initial conformations of ECFP-l_3-EYFP. (*Right*) MD snapshots of final equilibrium conformation of ECFP-l_3-EYFP, taken at the end of the MD sampling.
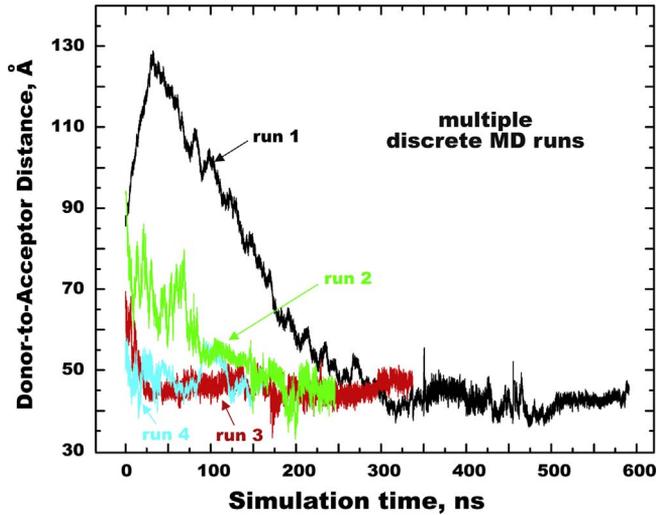
Fig. 3. Parallel MD sampling of ECFP-l$_3$-EYFP. The series of the four different starting configurations of ECFP-l$_3$-EYFP (run 1–4) were run in parallel to achieve adequate convergence statistics.

and Cl$^-$ (Fig. 4A), which revealed only a small effect of ionic strength on equilibrium distances between bulk EFCP and EYFP moieties (Fig. 4B).

1.3. Orientation factor κ^2

The orientation factor κ^2 that provides the dependence of the interaction between two electric dipoles on their orientations as shown in Fig. 5. κ^2 can be defined as shown in Fig. 5B. To calculate instantaneous $\kappa^2(t)$ from MD trajectories, the angles θ_D, θ_A and distance R_{DA} were calculated by using GROMACS analysis utilities *g_angle* and *g_sgangle*, respectively.

The instantaneous orientation factor $\kappa^2(t)$ for the FRET ECFP-EYFP molecular pair was calculated from the instant position coordinates of the MD simulation trajectory, which allowed us
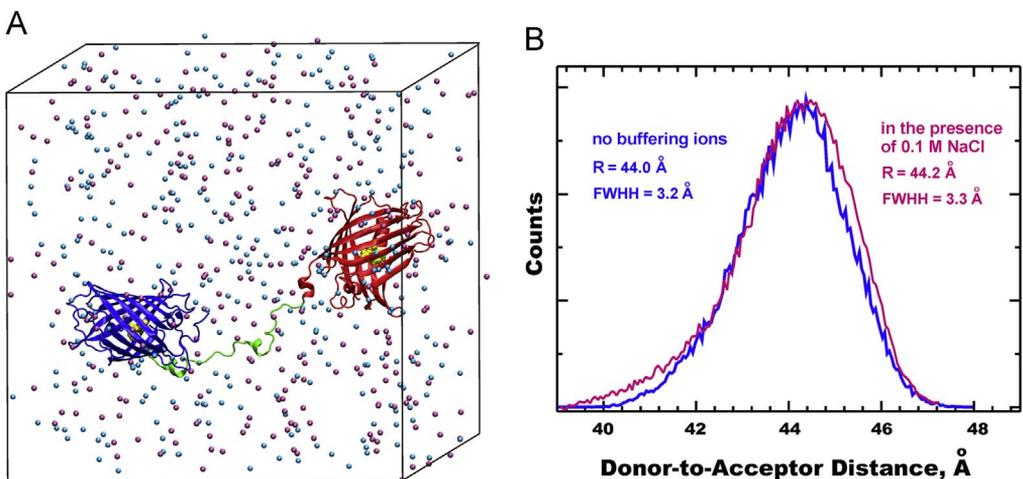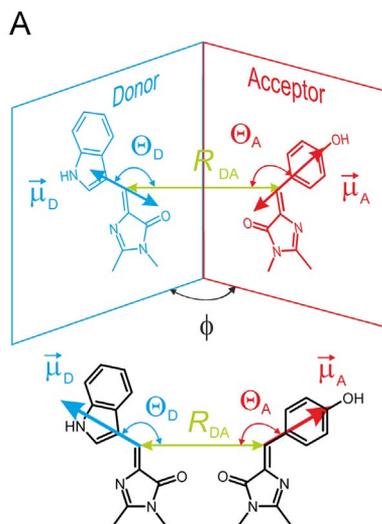

Fig. 4. MD simulations of ECFP-to-EYFP distance. (**A**) Snapshot of ECFP-l$_3$-EYFP in a MD box in the presence of 0.1 M buffering ions Na$^+$ (*cyan balls*) and Cl$^-$ (*mauve balls*). (**B**) Comparison of ECFP-to-EYFP distance distributions calculated in the absence (*blue*) and in the presence of 0.1 M NaCl (*magenta*).

Fig. 5. (**A**) The transition dipole moments of the ECFP donor and EYFP acceptor chromophores are shown as vectors μ_D and μ_A, respectively. The angles Θ_D and Θ_A define the dipole orientations with respect to the connecting unit distance vector R_{DA}, whereas φ is the angle between the two planes. (**B**) The definition of the orientation factor κ^2.

to estimate the distribution of the donor and acceptor transition dipoles (Fig. 6). The MD-estimated average value of $<\kappa^2> = 0.69$ is in close agreement with the isotropic limit of 2/3, serving as an additional proof of sufficient conformational sampling. Comparison of the simulated distribution of κ^2 with the theoretical angular distribution for isotropic dipoles in a three-dimensional system [6,7] is shown in Fig. 6. While overall agreement was observed, it might be noted that the donor–acceptor pair linked by a series of chemical bonds still shows some minor deviations from ideal angular orientations of random non-interacting dipoles, also observed for other systems [8–11].

1.4. MD-based calculation of FRET efficiency

The apparent FRET efficiency E_{DA} for the ECFP-l₃-EYFP system was calculated from MD-generated time traces by using Eqs. (1) and (2), the MD-estimated instantaneous value of the ECFP-to-EYFP distance R_{DA}, and the instantaneous orientation factor $\kappa^2(t)$, both evaluated at each MD trajectory step [1].

$$E = \frac{1}{1 + (R_{DA}/R_0)^6} \tag{1}$$

where R_0 is the Förster radius characteristic for the specific dye pair involved in FRET. For ECFP/EYFP pair $R_0 = 48$ Å [12], as calculated from the following

$$R_0^6 = \frac{9000 \cdot (\ln 10) \cdot \kappa^2 Q_D}{128 \cdot \pi^5 N n^4} J(\lambda) \tag{2}$$

where refractive index $n = 1.4$ [12,13], quantum yield of the donor $Q_D = 0.4$, the overlap integral $J(\lambda)$ is calculated using the normalized ECFP emission spectrum and the EYFP excitation spectrum,
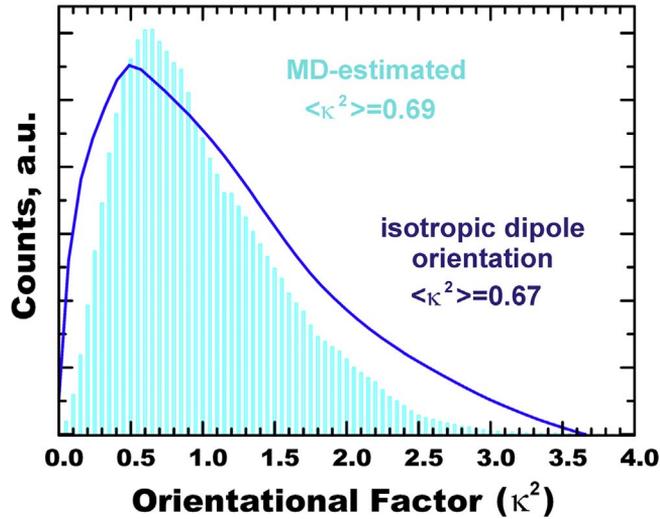
Fig. 6. Comparison of the MD-simulated distribution of $<\kappa^2>$ (*cyan*) with the theoretical curve for an ideal unrestricted isotropic dynamic orientation of dipoles $<\kappa^2> = 2/3$ (*blue*).

normalized to 84,000 M^{-1} cm^{-1} at 514 nm. For κ^2, a value of 2/3 was used, i.e., assuming dynamic averaging of the relative orientations of the ECFP emission and EYFP absorption dipole moments.

The mean value of the MD-estimated apparent FRET efficiency was estimated to be 0.55 and the corresponding E_{DA} histogram reveals a broad distribution (FWHH=0.25), which can be compared to the distribution obtained from the *sm*-FRET measurements (Fig. 7A) [1].

1.5. Reconstructing R_{DA} distance distributions from FRET efficiencies

First, we used MD-generated FRET efficiency E_{DA} to reconstruct the "apparent" Donor–Acceptor distance distribution (R_{App}), assuming a unique Förster distance of $R_0 = 48$ Å that corresponds to the average orientational factor $\kappa^2 = 0.69$ (Eq. 1). Then, we compared it to the "true" distance distribution (R_{True}) directly estimated from the MD trajectory (Fig. 7B).

The comparison of the effect of the standard assumption of the unique R_0 value on the accuracy of determination of the Donor-to-Acceptor distance distributions is presented in Fig. 8. Both *apparent* distributions calculated from either MD-generated or experimentally measured FRET efficiencies, calculated under standard assumptions, systematically overestimate the width of the *true* distance distribution R_{DA}. We conclude that a careful consideration of the orientational dynamics within a FRET pair is crucial for accurate measurements of distance distributions.

2. Experimental design, materials and methods

2.1. Materials

The materials and methods used to prepare the sample and collect the absorption, fluorescence intensity and fluorescence decay and sm-FRET data are given in [1,14].

2.2. Molecular dynamics simulation setup

The structural model of peptide ECFP-l3-EYFP [12] was designed based on available X-ray structures of ECFP (PDB code: 1CV7) and EYFP (PDB code: 1YFP) [15], connected by flexible linker
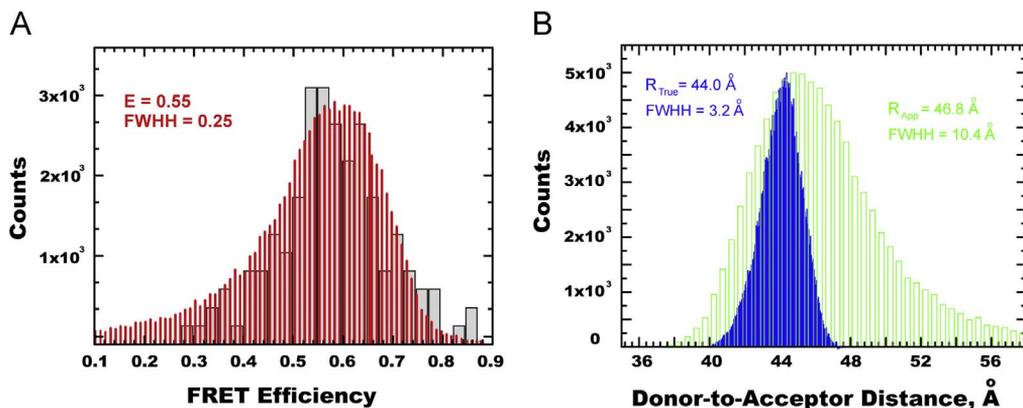
Fig. 7. (**A**) Comparison of the experimentally observed *sm*-FRET efficiency (*grey bars*) and MD-reconstructed distribution of FRET probability. (**B**) Comparison of ECFP-to-EYFP distance distribution histograms calculated by the direct MD sampling (*blue*) and reconstructed from the MD-estimated E_{DA}, under the standard simplifying assumption of unique $R_0 = 48$ Å (*green*) (Raw data associated with this Figure are provided in Supplementary files).
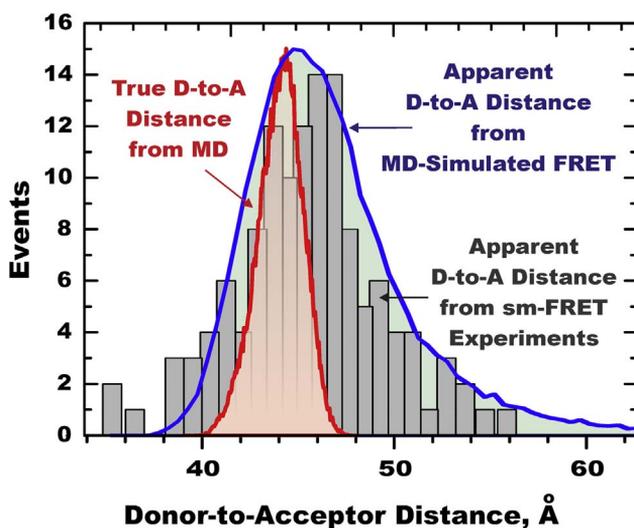

Fig. 8. Comparison of True and Apparent distance distributions in ECFP-l_3-EYFP model FRET system. True Donor-to-Acceptor distance distribution (*red*) is calculated directly from MD trace without any assumptions. Apparent Donor-to-Acceptor distance distributions are calculated from either experimental sm-FRET data (*black*) or MD-generated FRET data, assuming an average orientation factor $<\kappa^2> = 0.69$ (and subsequently unique value of $R_0 = 48$ Å) for the entire population of conformations. (Raw data associated with this Figure are provided in Supplementary files).

TLGMDELYKSGIR(GGSGGS)$_3$-TMVS referred to as l$_3$. The CHARMM27 force field for proteins, recently adopted for the GROMACS package, was used [16]. In this force field, the CHARMM27 CMAP correction term was implemented. The solvent water was modeled using the special CHARMM (TIP3P) model [16]. The bond length and angle parameters for a chromophore residue of ECFP and EYFP were optimized by density functional theory calculations at the B3LYP/cc-pVDZ level and adopted for the CHARMM27 force field. Partial charges needed for Coulomb interactions were derived from the B3LYP/ccpVDZ electron densities by fitting the electrostatic potential to point (ESP) charges [17]. The

$S_0 \rightarrow S_1$ transition dipole moment needed for calculation of the orientation factor κ^2 was derived from time-dependent density functional calculations TD-B3LYP/cc-pVDZ, and the corresponding dipole unit vectors are shown in Fig. 5. The estimated direction of the excited-state transition dipole moments of the ECFP and EYFP chromophores are found to be in agreement with those previously published for the chromopores of the green fluorescent protein family [18]. The MD topology file for polypeptide ECFP-l$_3$-EYFP was built using GROMACS pdb2gmx utility. The CHARMM27 topology building blocks for non-native amino acid residues CRO and CRF were implemented using the corresponding GROMACS RTP library.

All MD simulations were carried out at a constant number of particles, constant pressure ($P=1$ atm), and constant temperature ($T=298$ K, NPT ensemble). Three-dimensional periodic boundary conditions were applied with the z-axis lying along a direction normal to the bilayer. The pressure was controlled isotropically, so that the x, y and z dimensions of the simulation box were allowed to fluctuate independently from each other, keeping the total pressure constant. The reference temperature and pressure were kept constant using the Berendsen weak coupling scheme with a coupling constant of $\tau_T=0.1$ ps for the temperature coupling and $\tau_P=1.0$ ps for the pressure coupling [19]. Electrostatic interactions were simulated with the particle mesh Ewald (PME) approach using the long-range cutoff of 0.8 nm [20]. The cutoff distance of Lennard-Jones interactions was also equal to 0.8 nm. All bond lengths in the protein were kept constant using the LINCS routine [21]. The MD integration time step was 2 fs. The MD simulations were carried out using the GROMACS set of programs, version 4.5.5 [22]. Molecular graphics and visualization were performed using VMD 1.8.6 software packages [23].

2.3. Molecular dynamics simulation flowchart

A series of MD simulation runs of ECFP-l$_3$-EYFP, each starting with a different initial configuration, was performed according to the following flowchart: To initialize MD simulations, the FP protein was placed in a rectangular unit cell with a minimum distance of 2.0 nm from the box edge. The MD unit cell was filled with TIP3P water molecules by using GROMACS *genbox* utility. Next, steepest descent energy minimization was performed followed by an addition of ions in order to counterbalance the protein charge, yielding a final system of 11 Na$^+$ ions. The control MD sampling was also performed in the presence of 0.1 M buffering ions Na$^+$ and Cl$^-$. Before accumulating MD conformational sampling, each starting system was taken through a 10 ns pre-equilibration run. To achieve better MD statistics, four initial configurations of ECFP-l$_3$-EYFP with different distances and orientations between ECFP (donor) and EFYP (acceptor) were sampled. These four conformations of ECFP-l$_3$-EYFP were modeled in MD unit cells of the following sizes: $149.5 \times 149.5 \times 149.5$ Å (107902 waters), $202.4 \times 102.1 \times 102.1$ Å (64973 waters), $116.2 \times 116.2 \times 116.2$ Å (48461 waters) and $141.6 \times 101.4 \times 101.4$ Å (45049 waters). The MD simulations ran for a total time of 1.5 µs, from which the last 500 ns were taken for the equilibrium FRET analysis.

Supporting information

Raw simulation data of various conformations of protein ECFP-l3-EYFP are provided in PDB format. Supplementary raw data shown in Figs. 7 and 8 can be found in the online version of the paper.

Acknowledgements

Transparency document. Supporting information

Transparency data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2017.03.041.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2017.03.041.

References

[1] A. Kyrychenko, M.V. Rodnin, C. Ghatak, A.S. Ladokhin, Joint refinement of FRET measurements using spectroscopic and computational tools, Anal. Biochem. 522 (2017) 1–9.
[2] L.S.D. Caves, J.D. Evanseck, M. Karplus, Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin, Protein Sci. 7 (1998) 649–666.
[3] V. Mirjalili, M. Feig, Protein structure refinement through structure selection and averaging from molecular dynamics ensembles, J. Chem. Theory Comput. 9 (2012) 1294–1303.
[4] V. Mirjalili, K. Noyes, M. Feig, Physics-based protein structure refinement through multiple molecular dynamics trajectories and structure averaging, proteins: structure, Funct., Bioinforma. 82 (2014) 196–207.
[5] A. Kyrychenko, F. Wu, R.P. Thummel, J. Waluk, A.S. Ladokhin, Partitioning and localization of environment-sensitive 2-(2′-pyridyl)- and 2-(2′-pyrimidyl)-indoles in lipid membranes: a joint refinement using fluorescence measurements and molecular dynamics simulations, J. Phys. Chem. B 114 (2010) 13574–13584.
[6] L.M.S. Loura, Simple estimation of Forster Resonance Energy Transfer (FRET) orientation factor distribution in membranes, Int. J. Mol. Sci. 13 (2012) 15252–15270.
[7] B.W. van der Meer, Kappa-squared: from nuisance to new sense, Rev. Mol. Biotechnol. 82 (2002) 181–196.
[8] D.B. VanBeek, M.C. Zwier, J.M. Shorb, B.P. Krueger, Fretting about FRET: correlation between κ and R, Biophys. J. 92 (2007) 4168–4178.
[9] E. Deplazes, D. Jayatilaka, B. Corry, Testing the use of molecular dynamics to simulate fluorophore motions and FRET, Phys. Chem. Chem. Phys. 13 (2011) 11045–11054.
[10] M. Hoefling, N. Lima, D. Haenni, C.A.M. Seidel, B. Schuler, H. Grubmüller, Structural heterogeneity and quantitative FRET efficiency distributions of polyprolines through a hybrid atomistic simulation and Monte Carlo approach, PLoS One 6 (2011) e19791.
[11] K. Walczewska-Szewc, B. Corry, Do bifunctional labels solve the problem of dye diffusion in FRET analysis? Phys. Chem. Chem. Phys. 16 (2014) 18949–18954.
[12] T.H. Evers, E.M.W.M. van Dongen, A.C. Faesen, E.W. Meijer, M. Merkx, Quantitative understanding of the energy transfer between fluorescent proteins connected via flexible peptide linkers, Biochemistry 45 (2006) 13183–13192.
[13] J.R. Lakowicz, Principles of Fluorescence Spectroscopy, 3rd ed., Springer, New York, 2006.
[14] A. Kyrychenko, Y.O. Posokhov, M.V. Rodnin, A.S. Ladokhin, Kinetic intermediate reveals staggered pH-dependent transitions along the membrane insertion pathway of the diphtheria toxin T-domain, Biochemistry 48 (2009) 7584–7594.
[15] R.M. Wachter, M.-A. Elsliger, K. Kallio, G.T. Hanson, S.J. Remington, Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein, Structure 6 (1998) 1267–1277.
[16] Pr Bjelkmar, P. Larsson, M.A. Cuendet, B. Hess, E. Lindahl, Implementation of the CHARMM force field in GROMACS: analysis of protein stability effects from correction maps, virtual interaction sites, and water models, J. Chem. Theory Comput. 6 (2010) 459–466.
[17] A. Kyrychenko, A.S. Ladokhin, Molecular dynamics simulations of depth distribution of spin-labeled phospholipids within lipid bilayer, J. Phys. Chem. B 117 (2013) 5875–5885.
[18] T. Ansbacher, H.K. Srivastava, T. Stein, R. Baer, M. Merkx, A. Shurki, Calculation of transition dipole moment in fluorescent proteins-towards efficient energy transfer, Phys. Chem. Chem. Phys. 14 (2012) 4109–4117.
[19] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684–3690.
[20] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: an Nx log(N) method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089–10092.
[21] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: a linear constraint solver for molecular simulations, J. Comput. Chem. 18 (1997) 1463–1472.
[22] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, GROMACS: fast, flexible, and free, J. Comput. Chem. 26 (2005) 1701–1718.
[23] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38.