Accounting for dye diffusion and orientation when relating FRET measurements to distances: three simple computational methods

Katarzyna Walczewska-Szewc and Ben Corry

Supplementary Information

1 Testing the algorithms

1.1 Available Volume

To asses the accuracy of the algorithm and to estimate the number of steps needed to obtain a reliable dye position distribution, the AV calculations were performed in the absence of macromolecule. As expected, choosing an appropriate number of steps, results in a spherical volume with a radius equal the maximum length of the linker.

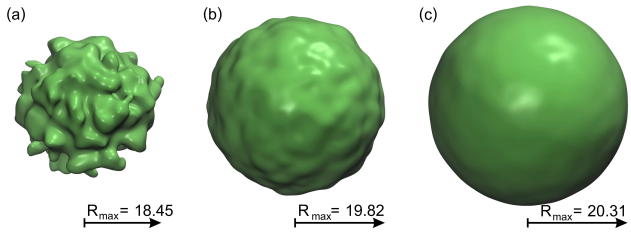

Figure S1: Comparison of AV for AF488 with three segments of the linker in the absence of a macromolecule. Calculations were proceeded for (a) one thousand, (b) ten thousands and (c) one hundred thousands steps.

Depending on the number of linker elements the minimum number of steps needed to obtain proper distribution varies greatly (figure S1) . For the linker composed of only one element about 2000 steps is sufficient. If the structure of the linker is more complex, the minimum number of steps increases to 100000 for three, 500000 for four, and more than 2000000 for five segments. In further calculations, it was decided to use three and four linker elements (depends of the length of the linker), as this is a good compromise between linker flexibility and computational time demands.

1.2 MD completeness tests

To make sure that the preliminary MD simulations, used in both rotamer and free dye snapshot methods, were run for sufficiently long period of time to sample all dye conformations, we needed a method of validation.

Table S1: An example of dimensions for some popular fluorescence dyes: maleimine derivative of AlexaFluor 488 and 594, cyanines Cy3 and Cy5, and succinimide ester of ALexaFluor 594. R_1, R_2 and R_3 denotes the width, height, and depth of the headgroup respectively.

Name	r_{max}	R_1	R_2	R_3
AF488	21.00	6.8	3.9	1.5
AF594m	21.00	7.6	4.1	2.2
AF594s	7.08	7.6	4.1	2.2
CY3	21.12	9.5	3.8	2.3
CY5	21.12	10.3	3.8	2.3

The first method was to check how much a distribution of the orientation factor for each pair of dyes differ from the analytical distribution for the isotropic approximation [1]. Because of the geometry of the simulation systems, which was the freely rotating dyes in the water box, we did not expect huge differences from the analytical distribution. Nevertheless the existence of restraint, the presence of fixed amino acid residue to which the dye is fixed, could couse small deviation to appear.

The results of our tests are shown on the Figure S2a where we combined four separate simulations into two pairs of dyes. In the case of pair AF488-AF594 the distribution of orientation factor is very similar to the isotropic behaviour. For the pair of cyanines, κ^2 distribution deviates slightly from the analytical one, which demonstrates that the rotation of these dyes is more sensitive to the steric obstacles of the residue to which they are fixed. The other possible explanation is the existence of favourable dye conformations.

The next validation method is to check the smoothness of the κ^2 distribution and convergence to an average value. No matter how close the rotation is to the isotropic case, good convergence means that κ^2, as well as other quantities, is well sampled and simulations can be used as a base in our methods. In figure S2b, the cumulative average of orientation factor for two pairs of fluorescent dyes is shown. Both of them are characterized by excellent convergence to the average value. Notably, $< \kappa^2 >$ deviates significantly from 2/3 for the Cy dyes due to the steric hindrance of the single residue to which they are attached.

1.3 RL completeness tests

To check how well the rotamer library reconstructs the MD data, we compared the probability distributions of distance and κ^2 obtained from both rotamer libraries and the MD data underlying the RL construction. To do this, two fluorescence dyes, AF488 and AF594 were separated by 50 Å in the y-axis. Next, the distributions of the distance and orientation factor were calculated.

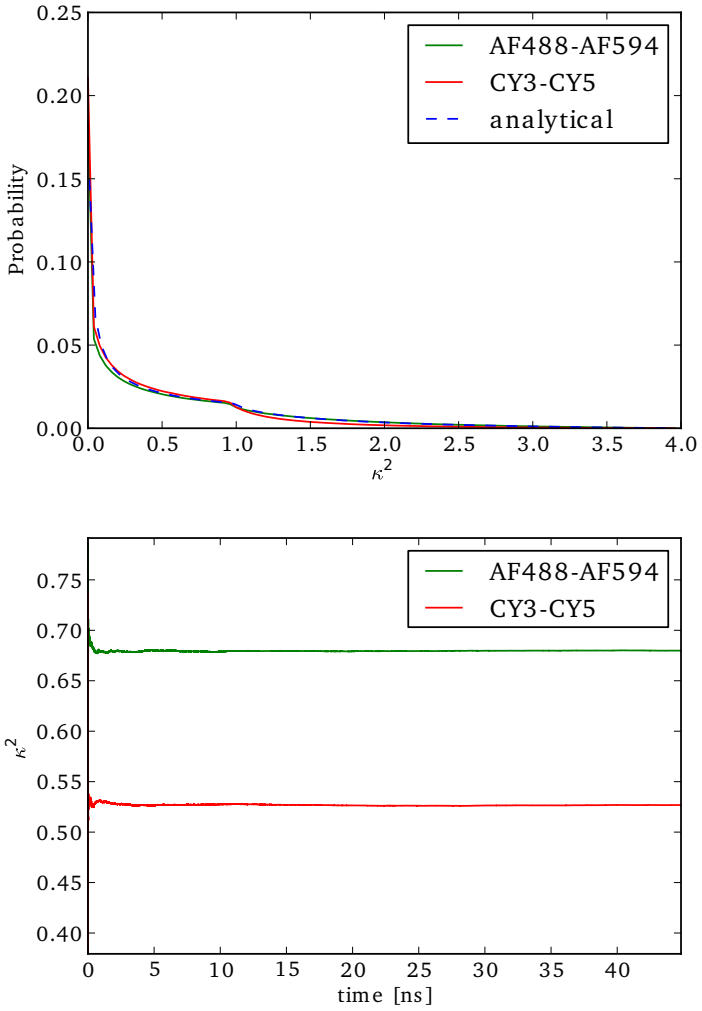

Figure S2: Distribution (upper) and cumulative average of orientation factor for two pairs of fluorophores: AF488-AF594 and CY3-CY5.

In both distance and κ^2 distributions, small differences are seen, however, the overall shape of distribution is conserved (Fig S3 a,b). This is a result of more discrete set of samples in the case of rotamer library. On this base we can suppose, that RL can successfully reconstruct not only the distance distribution, but also the orientation of fluorophores to within the level of discreteness of the library.

1.4 FD library reduction

Th FD method is designed to use the raw output from MD simulations. Usually, a longer trajectory will produce better sampling of allowed dye

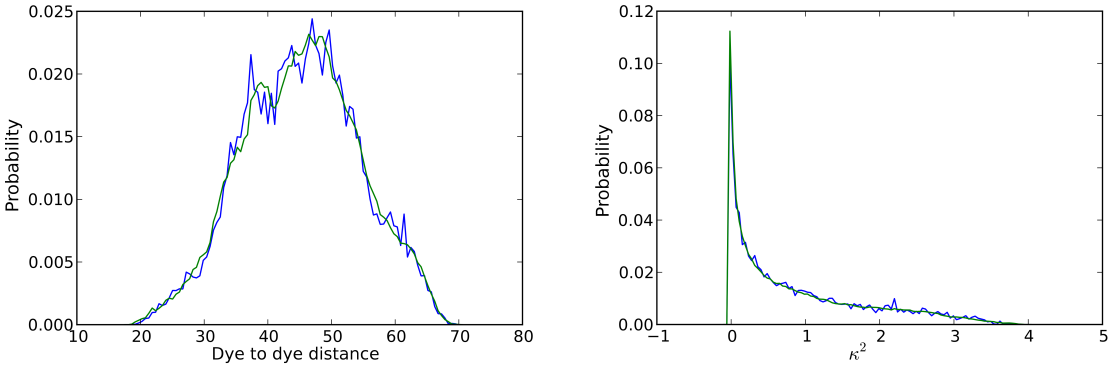

Figure S3: Distance and κ^2 distribution for the AF488-AF594s from rotamer (blue) and snapshot (green) libraries in the absence of a host molecule.

positions. However, since the algorithm for calculating the distance, FRET efficiency, and κ^2 distributions has to loop through each combination of frames from each trajectory of the donor and acceptor, including the entire MD trajectory can by time consuming. It may be possible to achieve equally good results using a lower number of frames and thus less time.

To examine this, in figure S4 we compare the distance and κ^2 distributions obtained using the FD method with less frames. Initially snapshots from the entire 500ns trajectory were stored every 10 ps yielding a trajectory comprising 50,000 frames for each dye. This full trajectory is used in the data shown in the main text. In Figure S4 we show the result for the polyproline system obtained using only a subset of the total trajectory in which only every Sth frame was used.

Calculations of distance, κ^2 and efficiency distribution takes less than one minute for the highest value of S, and respectively, 4, 28, 115, and 460 minutes for lower reduction levels. This can form a guideline for how many frames are necessary (and thus how much time) to gain the desired level of accuracy.

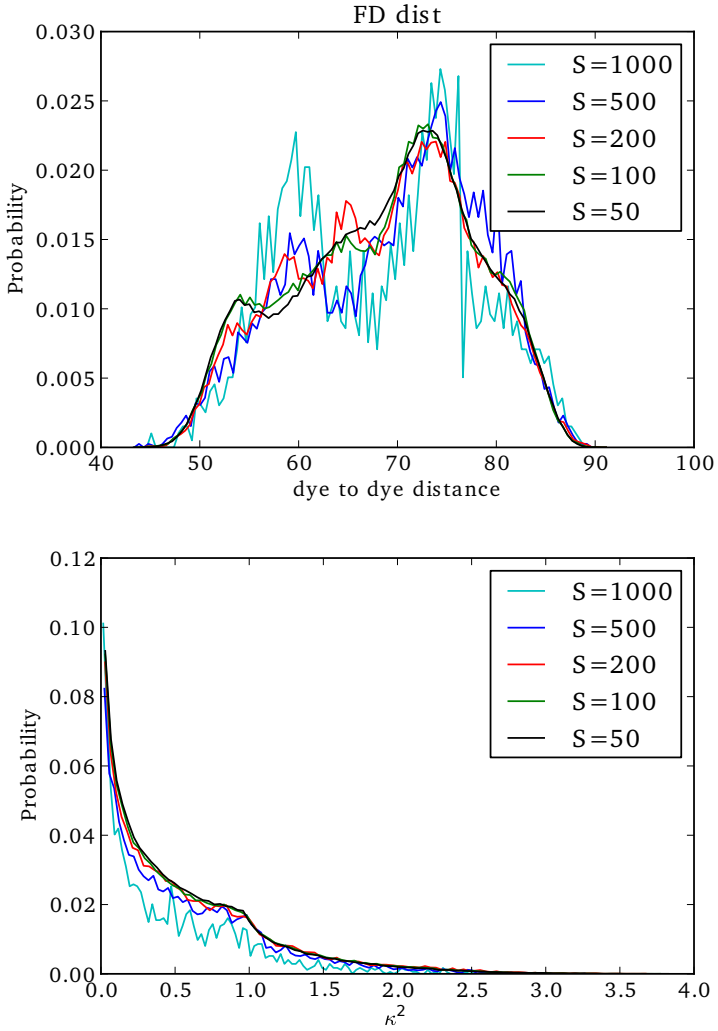

Figure S4: Distance and κ^2 distribution for the AF488-AF594s on the poliproline termini with different level of trajectory reduction, S.

2 Algorithm flow diagrams

2.1 Available Volume

The use of the AV method requires two steps. In the first step we calculate an available volume for each dye (*AVolume.cpp*). Then, we use Python post processing script (*AVanalyse.py*) to extract information about the distance, κ^2 and FRET efficiency distributions.

2.1.1 Imput parameters.

```
int STEPS = 1000000;          // number of runs of main program loop
char outfile[8] = "Dye.dat";  // the name of output file
double L = 21.5;              // the length of the dye linker
double R1 = 7.1;              // half of the length of the dye head group
double R2 = 3.1;              // half of the width of the dye head group
double R3 = 2.2;              // half of the depth of the dye head group
double asize = 2.;            // the average size of dye atoms
int Link_elements = 3;        // the number of elements of the linker
double PO[3] = {-1., 2.3, 3.};// the point in which the dye is tethered
int prot_num = 1060;          // the number of host molecule atoms
char prot_name[9] = "Pro.pdb";// the pdb file of macromolecule
```

2.1.2 AVolume.cpp

The algorithm consists of several steps (figure S4). In the first step all atoms positions from the macromolecule pdb file are read. Then, for each of three dyes head group radii the main loop is started. In the main loop (repeated STEPS times) the position of each linker element is calculated and checked for the possible overlapping with the macromolecule atoms or other segments of the linker.

2.1.3 Postprocessing

To analyse calculated volumes we use Python script, which transform the output of previous program (the AV of each dye) into the distance and FRET efficiency distributions graphs.

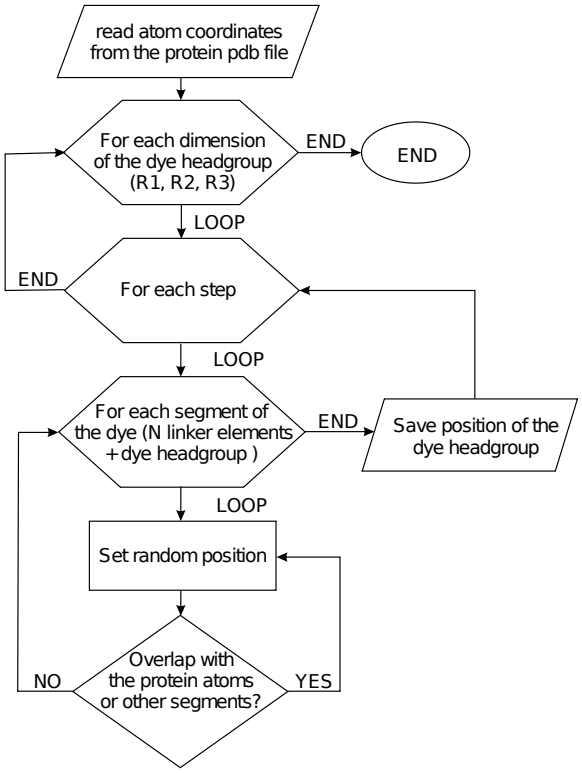

Figure S4: AVolume.cpp

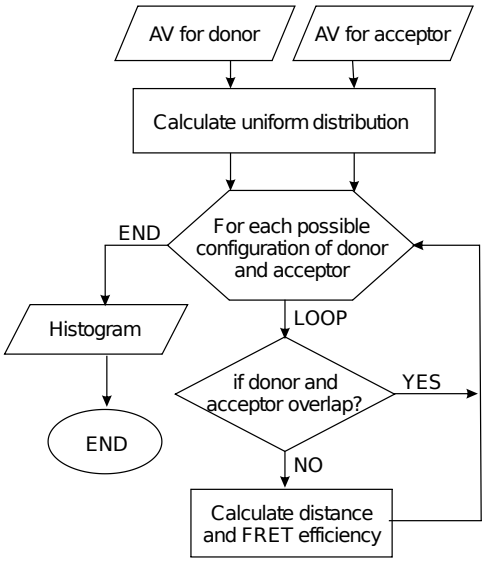

Figure S5: AVanalyse.py

2.2 Rotamer library

2.2.1 Library construction

To build a rotamer library w need the preliminary set of MD simulations of free dyes in water box. These simulations can be transformed into RL library used our Python script RotamerLibrary.py

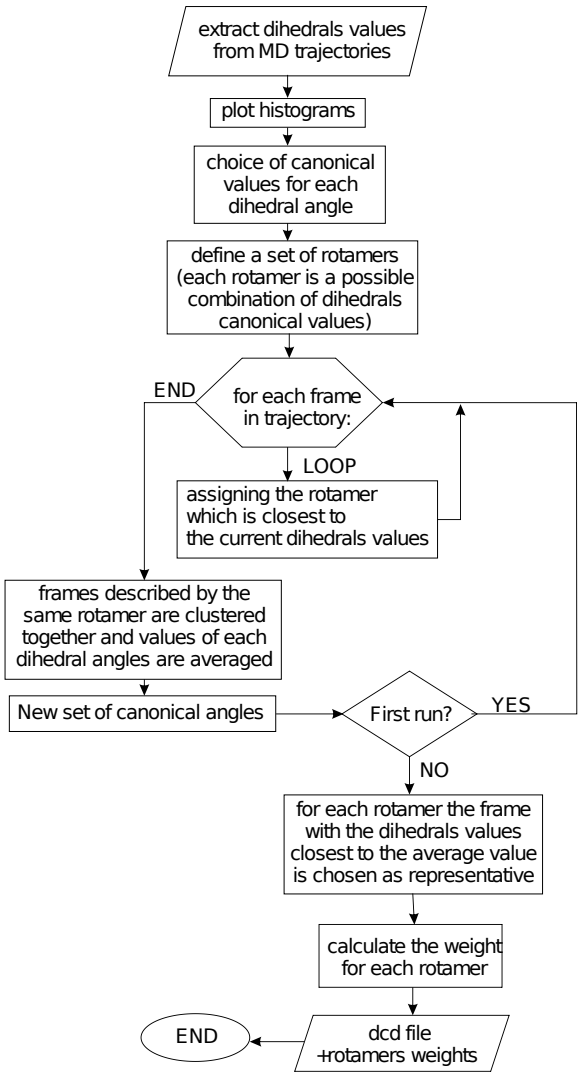

Figure S6: RotamerLibrary.py

2.2.2 Postprocessing

Once the rotamer libraries are built and the macromolecular system is chosen, the analysing script RotamerAnalyse.py should be run to calculate the

distance, orientation and FRET efficiency distribution.

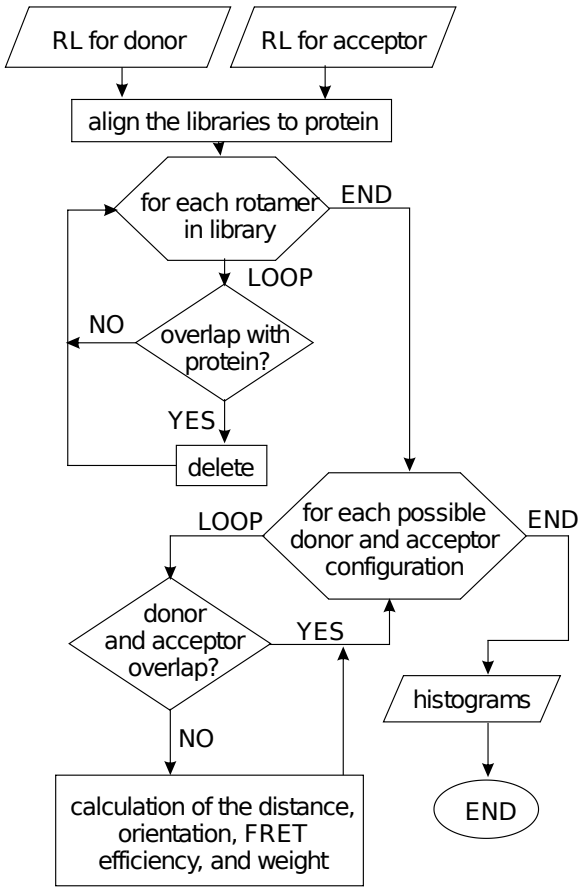

Figure S7: RotamerAnalyse.py

2.3 Free Dye snapshots library

The FD method can be run using raw dcd files obtained during the preliminary MD simulations of free dyes in water box. In the analysing script FDAnalyse.py one should provide an information about molecular system (pdb file) and the residues to which the dyes are tethered. As an output, the distance, orientation and FRET efficiency distribution are calculated.

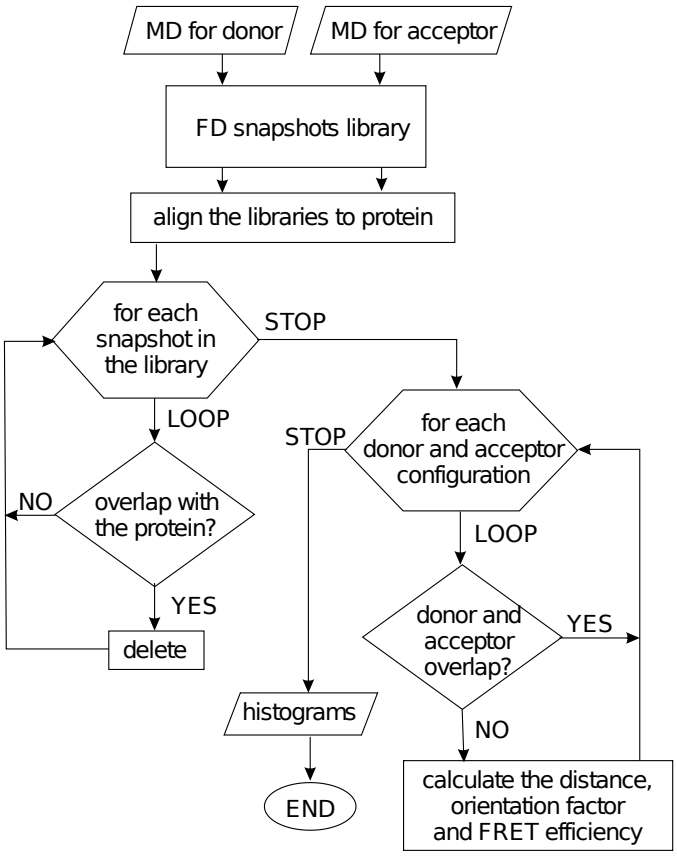

Figure S8: FDAnalyse.py

References

[1] Van Der Meer, B.W., Coker, G.III, and Chen, S.-Y. (1994) *Resonance Energy Transfer Theory and Data*, VCH, New York