

PCCP

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

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

Katarzyna Walczewska-Szewc^{*ab} and Ben Corry^{*a}

Received Xth XXXXXXXXXXX 200X, Accepted Xth XXXXXXXXXXX 20XX

First published on the web Xth XXXXXXXXXXX 200X

DOI: 10.1039/b000000x

Förster resonance energy transfer (FRET) allows in principal for the structural changes of biological systems to be revealed by monitoring distributions and distance fluctuations between parts of individual molecules. However, because flexible probes usually have to be attached to the macromolecule to conduct these experiments, they suffer from uncertainty in probe positions and orientations. One of the way to address this issue is to use molecular dynamics simulations to explicitly model the likely positions of the probes, but, this is still not widely accessible because of the large computational effort required. Here we compare three simpler methods that can potentially replace MD simulations in FRET data interpretation. In the first, the volume accessible for dye movement is calculated using a fast, geometrical algorithm. The next method, adapted from the analysis of electron paramagnetic studies, utilises a library of rotamers describing probe conformations. The last method uses preliminary MD simulations of fluorescent dyes in solution, to identify all conformational states of dyes and overlays this on the macromolecular system. A comparison of these methods in the simple system of dye-labelled polyproline, shows that in the case of lack of interaction between the dye and host, all give results comparable with MD simulations but require much less time. Differences between these three methods and their ability to compete with MD simulations in the analysis of real experiment are demonstrated and discussed using the examples of Cold Shock Protein and Leucine Transporter systems.

1 Introduction

Förster Resonance Energy Transfer (FRET) is a radiationless transfer of energy occurring between excited molecule (donor) and ground-state molecule (acceptor) via long range dipole-dipole interaction¹. This kind of transfer plays an important role in many fields of science that utilize fluorescence phenomena, ranging from characterisation of nano-materials to understanding molecular biophysics². One of the features making FRET so widely used is a strong dependence of energy transfer on donor-acceptor separation. Due to this dependency, it can be used as spectroscopic ruler to measure intermolecular distance in a range of 10 – 100Å. Changes in the distances between molecules or in the shape of a macromolecule can be observed by attaching fluorescent dyes to known locations and determining changes in the transfer efficiency³. In most cases, the transfer efficiency may be directly related to the distance between donor and acceptor, r , using the simple formula

$$E = \frac{R_0^6}{R_0^6 + r^6}, \quad (1)$$

where R_0 is the so called Förster radius determining the distance at which FRET and donor fluorescence are equally probable⁴,

$$R_0^6 = 8.79 \times 10^{-5} \cdot \frac{\kappa^2 \Phi_D}{n^4} \cdot J, \quad (2)$$

where n is the refractive index, Φ_D is the quantum yield of the donor in the absence of acceptor, J is the spectral overlap of the pair of fluorophores and κ^2 describes the relative orientation of the donor and acceptor transition dipoles. Experiments performed in single molecule setup (smFRET) can provide time resolved information about distance fluctuations of individual molecules, which is the key to revealing the dynamics of many biological processes, such as protein folding and conformational changes^{2,5–7}.

A significant hurdle in using resonance energy transfer to measure distances is that such measurements are complicated by the motion of the fluorophores relative to the macromolecule and the inherent dependence of FRET upon the relative orientation of the fluorophores. In other words, the transfer measurements suffer from uncertainty in probe position and orientation.

The first of these problems arises from the fact that the commonly used fluorescent labels are characterized by high conformational flexibility. This flexibility, which is localised particularly in the linker, allows for the introduction of the label with minimal perturbation to the system, because the

[†] Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/

^a Research School of Biology, Australian National University, Australia. E-mail: fizkws@ug.edu.pl (for KWS), ben.corry@anu.edu.au (for BC)

^b Institute of Experimental Physics, University of Gdansk, Poland.

chain can simply adapt to the free space in the macromolecular structure. Nevertheless, the more flexible the dye is, the more it can diffuse around the attachment point, which increases the width of the measured distance distribution. Depending on the length of the linker, the uncertainty can be significant. The use of short linkers can somewhat reduce the impact of dye diffusion⁸, however this solution is not always applicable and can lead to restraints on fluorophore rotation.

The next problem is the result of the transfer dependency on the relative orientation of the donor and acceptor via the orientation factor, κ^2 , that forms part of the definition of the Förster radius (eq. 2). It can be defined as

$$\kappa^2 = (\cos \theta_T - 3 \cos \theta_D \cos \theta_A)^2 \quad (3)$$

where θ_T is the angle formed by donor and acceptor transition moments, θ_D and θ_A are the angles between each transition moment and the vector joining the middle point of each dipole. Using FRET as a spectroscopic ruler is particularly successful in situations where the orientation of donor and acceptor is isotropic, dynamic and uncorrelated. In this case the orientation factor can be approximated by $\kappa^2 = 2/3$ and the simple relation between the distance and the transfer efficiency can be applied (Eq 1)⁴. Unfortunately the presence of the linker and interaction with the macromolecule can strongly affect the orientational distribution of the dye. In such cases, the commonly used approximation of $\kappa^2 = 2/3$ is no longer valid and the orientation factor can assume values in the range of 0 to 4⁴. Furthermore, direct measurements of the dye orientations are not currently possible in most situations⁹.

Directly simulating the motion of the probes on the macromolecule using molecular dynamics (MD) can potentially elucidate the degree of dye motion and orientations^{10–12}. This type of simulation provides us with detailed, time resolved information about the position and orientation of the dye, as well as about the dynamics of the entire molecular system, provided that we already have atomic resolution structural data of the macromolecule to which the dyes are attached. Ideally, the transfer efficiency distribution obtained from such simulations can be compared to the experimental data to help reveal the molecular structure consistent with the measured efficiency. But, to obtain reliable information about the distribution of FRET dyes, simulations have to be conducted for a very long period of time to sample all the conformations of the dye and macromolecule¹³. Thus, despite the undoubted advantages of MD simulations, their usefulness in FRET data analysis is limited because of the large computational effort underlying such calculations.

In some cases, especially if the crystallographic structure of the macromolecule is well known, it is possible to replace the computationally demanding MD with a simpler methods to obtain information about the positions that can be occupied by the dyes. In this study we develop and compare the use

of three such simplified methods for understanding the influence of dye position and orientation on the transfer efficiency. For example, the so called Available Volume (AV) method reconstructs the volume of all positions possibly occupied by the dyes that do not overlap with the macromolecule, using a simple, geometrical algorithm. This method has previously been applied to examine the potential motions of dyes attached to DNA⁸. Although the volume should be calculated separately for each dye and each macro-molecule structure, the method is still much more practical in data analysis than MD, because it allows the prediction of the mean distance between donor and acceptor dye using a common personal computer in a reasonable computation time. One of the most important assumptions underlying this method is that all possible positions within the available volume are equally probable.

Another method which can take into account both orientation factor and the existence of favourable conformation of the dyes is the so called Rotamer Library (RL) approach. This method was successfully used in the analysis of spin labels in electron paramagnetic studies^{14–16} but has not yet been adapted to use in FRET. Here, we build a rotamer library for a number of common fluorescent dyes, based on the MD simulations of a free dye in water. The main advantage of this method is that the one library can be applied to many different systems to reconstruct the distance and orientation distribution of the fluorophores. Using this method, we have to assume that conformations of each FRET label are not affected by the interaction with the macromolecule or with another dye (except for steric clashes). The lack of dye to dye interaction approximation is reasonable as long as the fluorescent labels are well separated from one another and do not collide.

The last approach we call the Free Dye snapshots (FD) method. It is very similar to the rotamer approach, but instead of using several representative conformations of the free dye, the entire MD trajectory of the dye in water is taken into account. After removing all conformations that overlap with the molecular system, the remaining frames can be used to calculate the distance and orientation distributions of the fluorescent dyes. This approach is more straightforward, because we do not have to analyze MD trajectories to construct the rotamer libraries, however the computational time needed to apply this method to the system is longer, due to the larger number of snapshots stored in the library.

Here we develop simple protocols for applying each of these methods to interpret FRET experiments, and make them easily accessible to researchers in the field. All three methods are compared to experimental or molecular dynamics data in three representative systems to assess how well they replicate simulated or experimental data. These systems start from the simplest case of fixed polyproline¹¹, move to a small molecular system of cold shock protein¹⁷, and finish on the more complex case of leucine transporter embedded into a lipid

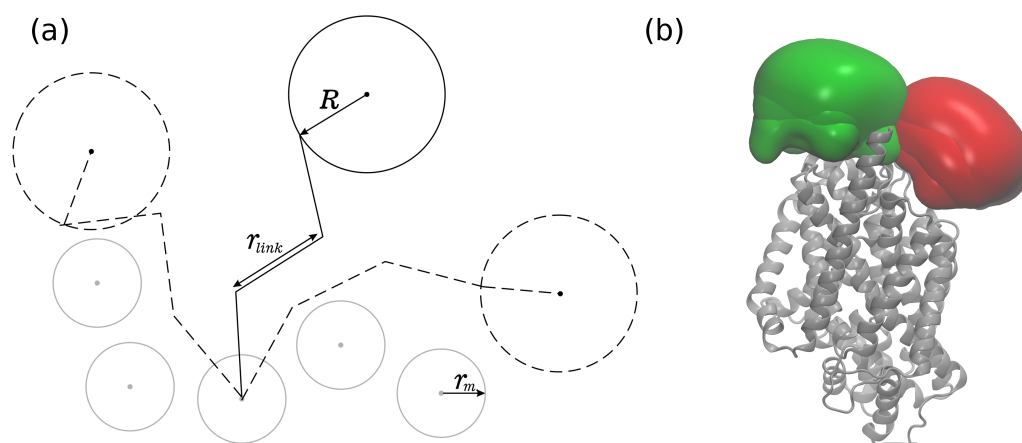


Fig. 1 (a) Schematic representation of the AV algorithm. R denotes one of three radii that approximate the geometry of the dye headgroup. The atoms in the macromolecule are shown in grey and three possible dye positions are indicated. (b) A graphical representation of the available volumes for the LeuT system. Red and green volumes represent the possible positions of donor (CY3) and acceptor (CY5) respectively, attached to specific protein residues.

membrane¹⁸.

2 Methods

2.1 Available Volume

The Available Volume method lets us determine the approximate space in which the dye is able to move. Because of the assumption that all dye positions are equally probable, using this method is strictly restricted to the cases in which none of the possible dye conformations are favoured and there are no interactions between the dyes and macromolecule (including sticking or trapping). What is more, the available volume contains all geometrically possible dye positions, which means that areas which are not explored by the real dye are also likely to be included.

To calculate all geometrically accessible dye positions, it was necessary to choose an algorithm which is relatively simple, provides a uniform distribution over all possible positions within the AV, and takes into account the flexibility of the linker. To ensure this, the calculations were performed using the following protocol, adapted and modified from the original one published by Sindbert, Kalinin et al⁸.

The dye was divided into two sections: a head group, represented as a sphere with radius R , and a linker, which was divided into n segments of equal length, r_{link} (figure 1a). The value of r_{link} was chosen to reconstruct the maximum length of the linker (if all elements are straight in one line), $r_{link} = \frac{r_{max} - R}{n}$.

In each independent calculation step, the orientation of each linker segment and head group was randomly chosen from

the uniform distribution over a sphere. When overlap with the atoms from the macromolecule (approximated by a sphere with a given van der Waals radius for each atom type) occurred, the direction of the segment was changed until steric clashes were avoided. When the dye was successfully reconstructed, the position of the center of the head group was recorded.

To accurately reproduce the available volume for a given dye, three different values of the dye head group radius, R , have to be considered to account for its non-spherical nature. Each represents different dimension of the head group: length, width and depth. The available volume calculated for each of them can differ, because they represent the possible positions of the dye oriented in a different way. By combining the regions accessible for the three different dye orientations, we can reproduce the entire available volume, including the space in cavities in the macromolecular structure. An example of dimensions for some popular dyes are gathered in table S1 in the supplementary material[†].

For an accurate AV calculation, it is crucial to provide a good random distribution of the orientation of the linker elements. This was obtained using the Marsenne Twister pseudo random number generator, which gives a high quality uniform distribution with very long period $2^{19937} - 1$ ¹⁹. To be sure that all possible dye positions are well sampled, we tested how many steps were required, as described in supplementary material 1.1[†]. All data presented here involves at least 500000 independent calculations for each of three R values.

The condition of equally probable positions within the available volume was fulfilled by projecting all output coordinates on a discrete grid of coordinates with resolution of 0.5Å. Thus, each position on the grid can be sampled only one time.

Once the full set of possible positions is determined, distance distributions between pairs of dyes can be obtained as well as distributions of FRET efficiencies. Because the dye orientation is not considered in this approach we assume $\kappa^2 = 2/3$ in all AV results presented here. The typical run of this method for a single CPU (3.40GHz) takes about 1 hour.

2.2 Rotamer Library

The rotamer library method is an adaptation to the case of FRET of the method successfully used by Polyhach, Bordignon and Jeschke in electron paramagnetic studies^{14–16}. The main advantage of this approach in comparison with AV is that it provides information about relative orientations and favourable dyes conformations.

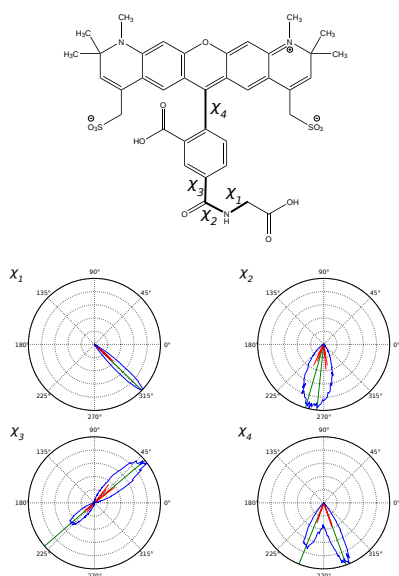


Fig. 2 Constructing the rotamer library. Probability distribution of dihedral angles ($\chi_1 - \chi_4$) for AF592 calculated from the MD trajectory (blue). The distance from the centre of each plot indicates the frequency at which each value arises. Canonical dihedral values are shown in green. The red lines show superimposed weights of real dihedral angle values for each rotamer.

2.2.1 Preliminary MD simulations. To identify the favourable conformational states of the dyes, we used NAMD Molecular Dynamic Software version 2.8²⁰ to perform molecular dynamics simulations of five fluorescent dyes solvated in water and 300mM NaCl. The total length of each simulation was 500 ns. The MD simulations for the five fluorescent dyes: maleimide derivative of AlexaFluor 488 and 594, cyanines Cy3 and Cy5, and succinimide ester of AlexaFluor 594, were run using previously developed parameters based on the CHARMM27 force field^{10,21,22}. Because our aim was to align

these dyes to different proteins, each dye in the MD simulation has been pre-attached to a single amino acid. Bonds between each hydrogen and the atom to which it is bonded were held rigid, allowing the use of a 2 fs time step. We applied periodic boundary condition in three dimensions, to avoid the influence of edges on our system. The electrostatic interactions were calculated using the particle mesh Ewald (PME) algorithm. The simulations were performed with constant temperature 298K maintained using Langevin dynamics and constant pressure of 1atm maintained using a modified Nosé-Hoover method in which Langevin dynamics is used to control fluctuations in the barostat. All simulations began with an initial minimization period of 1000 steps.

2.2.2 Rotamer library construction. The sets of possible dihedral angle values (rotamers) for the popular fluorescent dyes were built analogously to the rotamer libraries of spin labels. Each rotamer state represents one of the most frequently chosen conformations of the dye during the MD simulation. The statistical weights of each conformation in the library correspond to the number of MD frames assigned to given rotamer.

The first step to build the rotamer library was to extract data about dihedral angles of the dye from the MD trajectory. To obtain this, we used MDAnalysis, a python library to analyse and manipulate molecular dynamics trajectories²³.

Based on histograms plotting the frequency at which each dihedral is adopted by the flexible linker (fig.2, blue curve), a few of the most frequently sampled values of each angle were chosen as "canonical dihedral angles" (fig 2, green lines). Then, all possible combinations of canonical angles for each dihedral, so called rotamers, were generated and used as a base for further calculations (fig 2, red lines). For each frame in the MD trajectory, one of the canonical angle configurations was assigned based on a least square deviation to the actual dihedral values in the frame. The dye conformations described by the same rotamer, were then averaged to obtain the mean values of each dihedral angle in this particular rotamer. Next, the assignment of the canonical angles configurations was repeated using the mean values calculated in previous step and the average values were recalculated.

The frame with the dihedral angles configuration closest to the mean values is chosen (using least square deviation) to represent each single rotamer and its coordinates are recorded in the rotamer library. The number of MD frames assigned to each rotamer, normalized by a total trajectory length, is the statistical weight of given configuration of dihedral angle in the underlying MD trajectory, representing how favourable it is for the dye to adopt that rotamer. This entire process is written into easily run Python scripts. The rotamer library construction for maleimide derivative of AlexaFluor 488 (11 dihedral angles, 15 552 possible rotamers) using 500ns of MD

data takes about 12 hours for a single CPU. In the case of succinimide ester of AlexaFluor 594, which conformational changes can be represented as a set of 8 rotamers (figure 2), this time will be much shorter (less than 30s). Once the libraries are built, the time needed to apply them to a molecular system depends on their size- from 5 seconds for two small libraries to 3 hours for two large ones.

2.2.3 Application of the RL method in the molecular systems. To apply the rotamer library to a molecular system, the base of the dye has to be aligned to the dye binding site on the macromolecule. Since the library output was recorded as a dcd file, we used the python MDAnalysis tool to align the set of rotamers coordinates to the protein coordinate system. All conformations which cause steric clashes were removed from the library and the weights of remaining rotamers were renormalized to yield a set of possible positions of the dye attached to the macromolecule. After applying the rotamer libraries to each of the molecular systems, the weights of all possible combinations of the rotamer pairs were calculated. Distributions of distance, κ^2 and FRET efficiency for each system were then created by generating a random set of 10^5 dye configurations based upon the rotamer positions and weights.

2.3 Free dye snapshots

To implement this method the same set of preliminary MD simulations has been used as for the RL method (fig. 3). This time, instead of clustering the frames into a set of rotamers, all possible conformations of the dye were regarded separately with each snapshot of the MD simulation stored as a possible dye conformation in the FD snapshot library. No weighting is required as favourable configurations will appear more frequently in the library. This approach allows explicit calculations of dyes positions and orientation, however, extracting conformations which do not overlap with macromolecule can be more time consuming than previous methods as many more snapshots are stored in the library. Nevertheless the reward for this is higher accuracy, as the sampling of possible dye configurations is better.

As previously, the position of dyes in the FD snapshot library was aligned to the binding sites of proteins using the MDAnalysis tool. After removing overlapping conformations, the distances, κ^2 and FRET efficiencies for the remaining combinations of fluorophore pairs were calculated. The whole process of applying the snapshots library to a molecular system takes about 12 hours for a single CPU when using a 1000 frame trajectory for each dye. Results of similar quality to those shown here can be obtained in less time (≥ 2 hours) by storing less frames in the snapshot library (see fig S4).

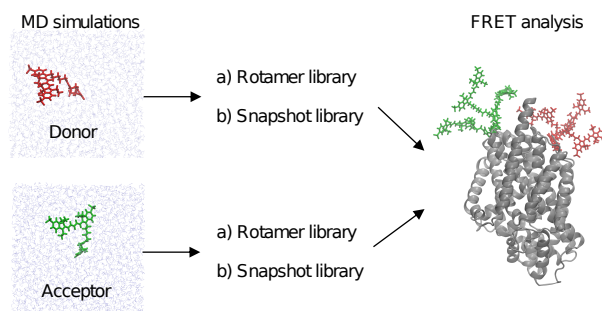


Fig. 3 Scheme representing the principle of the Rotamer and Free Dye snapshot methods. The preliminary MD simulations of donor and acceptor (on the left side) are analysed to extract the libraries of rotamers or snapshots, which then are applied to one of the molecular systems (right).

2.4 Burst averaging

In real experiments a recorded signal usually comes from averaging a bunch of several photons, emitted by fluorophores of different positions and orientation, which can not be simply compared to results from single photon obtained from the simulation. To reconstruct this effect in simulations we apply a ‘burst averaging scheme’ similar to that suggested by Hoefling et al.²⁴. In this, each new data point is simply created by averaging N randomly chosen data points from the efficiency histogram. The final shape of the distribution depend strongly on the number of photons in a single burst as can be seen on figure 4. The number of photons in single burst will vary depending on the experimental setup (principally on the laser power and detector settings), but usually it is more than ten²⁵. A value of 10 is used in all data shown here.

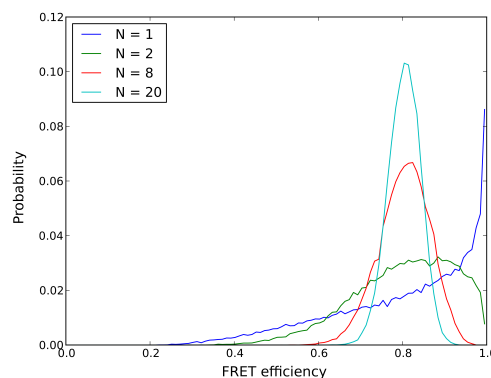


Fig. 4 Effect of bursts averaging on the final distribution of FRET efficiency for different numbers of photons in burst, N .

3 Results and discussion

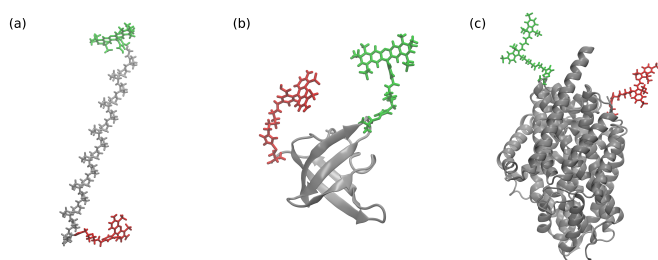


Fig. 5 The molecular systems studied here. (a) Polyproline 20 labeled with AF488 as a donor and AF594 as an acceptor. (b) Folded state of Cold Shock Protein labeled with AF488 and AF594. (c) Leucine transporter labeled with CY3 and CY5 as the donor and the acceptor respectively. The three images are not in the same scale.

3.1 Test in a simple system

To demonstrate the ability of the approaches to reconstruct the distance distribution between two dyes attached to a macromolecule, first their results were compared with experimentally relevant MD simulations of a polyproline system. This system has been studied repeatedly over the last years both in experiments and simulations^{3,11,26,27}. A chain of 20 prolines was labelled with two FRET dyes (Alexa 488 and 594), attached to additional glycine and cysteine at the polyproline termini (fig 5a). The chain of all trans prolines is regarded to be quite rigid²⁶, hence, to focus only on the dynamics of the attached dyes and since the simplified methods do not include dynamics of the macromolecule, we fixed the chain position during the MD simulations. The MD simulations were run for 200 ns with the dye and polyproline solvated in a waterbox with 300 mM NaCl.

The distributions of the distances between the dyes, obtained using the three simplified methods and MD, are shown in figure 6a. The AV method (red line) is the most approximate, because of the assumption of freely diffusing dyes, without any preferred dye conformation. Although the distance distribution provided by this method deviates slightly from the one obtained from MD (cyan line), it still captures the overall shape in this simple system.

The two remaining methods show the compromise between accuracy and computational effort. The distribution obtained using the rotamer library (blue line) suffers slightly from the limited sampling of possible positions. The curve is very sharp and fluctuates strongly. However, the calculations are very fast. On the other hand, the FD method gives the most accurate reconstruction, but the price is higher computational cost.

It is worth noting, that the distance between terminal cysteine and glycine (shown as black, dashed line) nearly overlap

with the middle of dyes distance distributions for each method. As will be shown later, this situation is not very common and means that the polyproline provides only a small restraint on the positions explored by the dyes.

In most FRET studies the κ^2 distribution is replaced by the average value of $\kappa^2 = 2/3$. This is justified in the case of isotropically oriented, rapidly moving free dyes. However, if this condition is not fulfilled, it can lead to significant discrepancies between experiment and theory^{4,9,13}. The normalized probability distribution of κ^2 assuming isotropic, unrestricted distributions of all angles is given by equation⁴.

$$p(\kappa^2) = \begin{cases} \frac{1}{2\sqrt{3}\kappa^2} \ln(2 + \sqrt{3}) & 0 \leq \kappa^2 \leq 1 \\ \frac{1}{2\sqrt{3}\kappa^2} \ln\left(\frac{2 + \sqrt{3}}{\sqrt{\kappa^2 + \sqrt{\kappa^2 - 1}}}\right) & 1 \leq \kappa^2 \leq 4 \end{cases} \quad (4)$$

The probability distributions of κ^2 obtained for the polyproline system using three different methods are shown in figure 6b. Remarkably, the distribution obtained from MD simulations (cyan curve) and the analytical one (magenta line) coincide in the finest detail. This indicates that the dyes diffuse uniformly about the positions to which they are tethered and that the macromolecule has limited influence on the positions sampled by the dye. As will be seen in the following examples, this one is unusual. Because the dyes labelled to the polyproline chain diffuse and rotate almost unrestricted, the 2/3 approximation is justified in this case. The FD method (shown in green) can reconstruct the overall shape of this distribution, but a small discrepancy appears for $\kappa^2 > 1$. Similarly, the RL method shows the main trend of the analytical probability distribution, however the fact that this method is more "discrete" than the other causes the fluctuation around the analytical curve. The AV approach does not provide us any information about κ^2 distribution, as it does not include information about dye orientations.

The transfer efficiency is a function of the distance between dipoles and relative orientation of the dye. The relation between them is shown by equation 1&2. Obviously, both the shape of the distributions of r and κ^2 have a contribution to the shape of the probability distribution of E . Because of the complexity of transfer efficiency, the problem of how to take into account both components arises.

According to the existing literature, at least two averaging regimes can be distinguished- static and dynamic. The difference between them refers to the comparison of the time scale of rotational motion of dyes (τ_{rot}) and the transfer time (τ_{FRET}).

When the donor and acceptor rotate fast enough to sample the entire range of orientations during the transfer time, $\tau_{rot} \ll \tau_{FRET}$, we are in the so called dynamic averaging regime. In this case the orientation factor in equation 1 is replaced by the average value calculated from the κ^2 distribution,

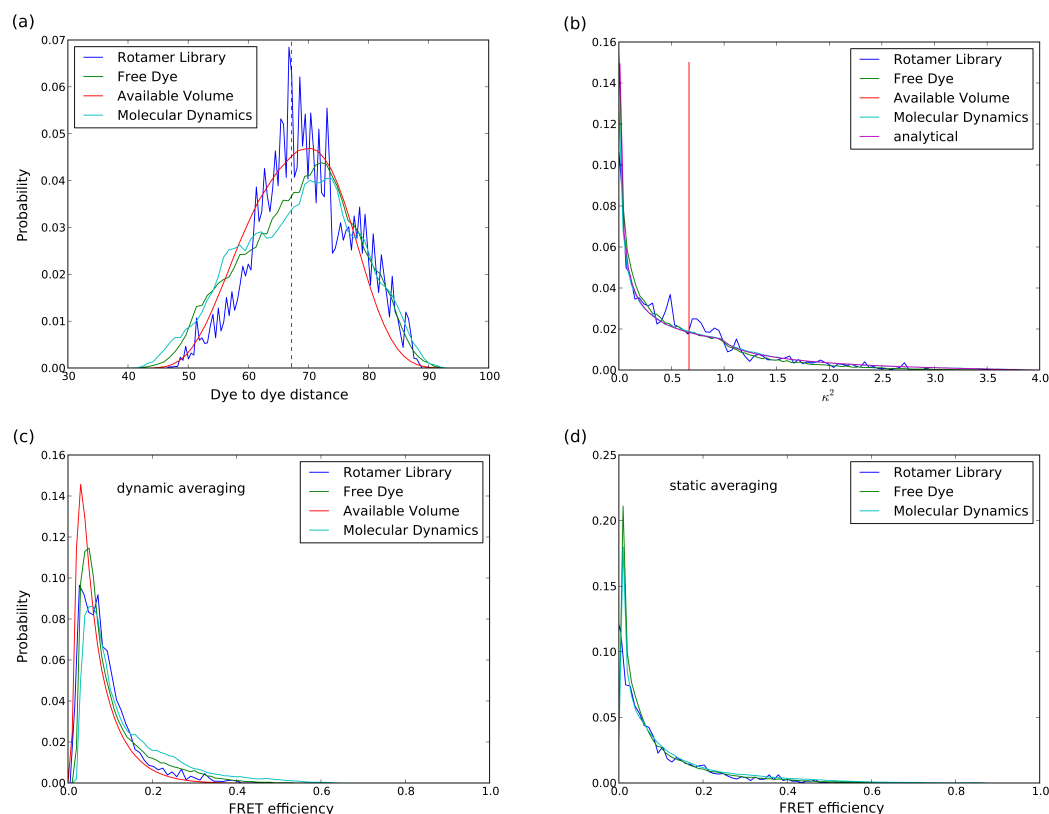


Fig. 6 Tests of the methods in a simple system. The characteristics of dye-labelled polyproline obtained using AV (red), RL (blue), FD (green) and MD method (cyan). The comparison of (a) dye to dye distance, (b) κ^2 , and the FRET efficiency distribution in (c) dynamic and (d) static regime. The black, dashed line on the distance distribution graph represents the distance between polyproline termini.

$$\langle E \rangle_{dynamic} = \frac{\frac{3}{2} \langle \kappa^2 \rangle R_0^6}{\frac{3}{2} \langle \kappa^2 \rangle R_0^6 + r^6}, \quad (5)$$

where R_0 is literature value of Förster radius, determined for $\kappa^2 = 2/3$.

An alternative situation is when dyes change their orientations slowly enough to assume that the energy transfer occurs between dipoles with well defined orientation, $\tau_{rot} \gg \tau_{FRET}$. Hence, the specific orientation factor should be considered for each snapshot of fluorophore positions (equation 3). This is the so called static averaging regime and the FRET efficiency can be calculated as

$$\langle E \rangle_{static} = \left\langle \frac{\frac{3}{2} \kappa^2 R_0^6}{\frac{3}{2} \kappa^2 R_0^6 + r^6} \right\rangle. \quad (6)$$

The transfer efficiency calculated in static, dynamic and any intermediate averaging regime can yield different results, therefore the decision about choosing the averaging regime should be based on the experimental measurements of the characteristic time scales, τ_{rot} and τ_{FRET} .

The comparison of E distribution obtained using MD and the alternative methods in dynamic and static averaging regimes is shown in figures 6c and 6d respectively. In the dynamic case the average value of κ^2 was calculated based on the orientation factor distribution shown in figure 6a. For the static calculations, the entire κ^2 distribution was taken into account.

All three methods are able to reconstruct the efficiency distribution from MD in the dynamic regime, which is most commonly used among experimentalists. Because the κ^2 distribution only slightly deviates from the isotropic approximation ($\langle \kappa^2 \rangle = 2/3$) even the most approximate method - AV yields the correct result. Obviously this method gives no information about the efficiency distribution in the static regime, which suggests an advantage for the snapshot and rotamer library methods if interpreting experiment conducted in the static regime.

3.2 Comparison to experimental data: CSP

The next step in our analysis was to compare the results of the simplified methods with experimental results. We chose the case of the folded state of the small cold shock protein homolog from the hyperthermophilic bacterium *Thermotoga maritima*, CspTm. This group of proteins are induced into a cell as a response to a rapid temperature decrease^{17,28}. Experimental data, to which we are comparing our numerical results, was obtained by Hoffmann et al. by using smFRET. As a crystal structure of folded state we used those available from the Protein Database (accession number 1G6P)²⁹. Next, to mimic the experimental system we incorporated two cysteine residues at positions 10 and 67 (fig. 5b), to provide functional groups for the attachment of maleimide derivatives of the dyes (Alexa Fluor 488 as donor and Alexa Fluor 594 as acceptor).

In our simulations the primary measured physical quantities are the distances between the dyes and their relative orientation. Then, based on that, the transfer efficiency is calculated using two different averaging regimes, as described above. The real experiment proceeds contrary to this order. There, the measurable quantity is the FRET efficiency, which can be translated into the dye-to-dye distance using some approximations for κ^2 . This explains why the simulated dye distance distributions, shown in the figure 7a, are not always identical to the experimental ones which are calculated using the isotropic approximation ($\kappa^2 = 2/3$).

When relating FRET efficiency distributions to distance distributions a key role is played by the orientation factor. Its distribution can tell us about the relative orientations of two dyes and how much their rotation differs from undisturbed ones. The κ^2 distributions obtained during the simulations can be seen on the figure 7b. The irregular shape of the green and blue curves (RL and FD methods) indicates that the presence of protein affects the rotation of dyes and prevents them from adopting some configurations.

Irregularities of the κ^2 distribution have an influence on the transfer efficiency in the static regime and cause its distribution to be significantly broadened (Fig 7c, dashed lines). Nevertheless the mean value of the orientation factor is still very close to $2/3$, which corresponds with the isotropic approximation. Consequently, the translation from the distance to the transfer efficiency is consistent with the reverse case in which the experimental values of the FRET efficiency are used to determine the distance distribution provided we are in the dynamic regime (Fig 7a,c). It may be possible to tell which regime the experiments were performed in by seeing which modelled efficiency distribution gives the best match to the experimental data. However, it is hard to eliminate other sources of discrepancy such as dye-protein interaction or protein motion.

In most cases, we are interested in the separation of the

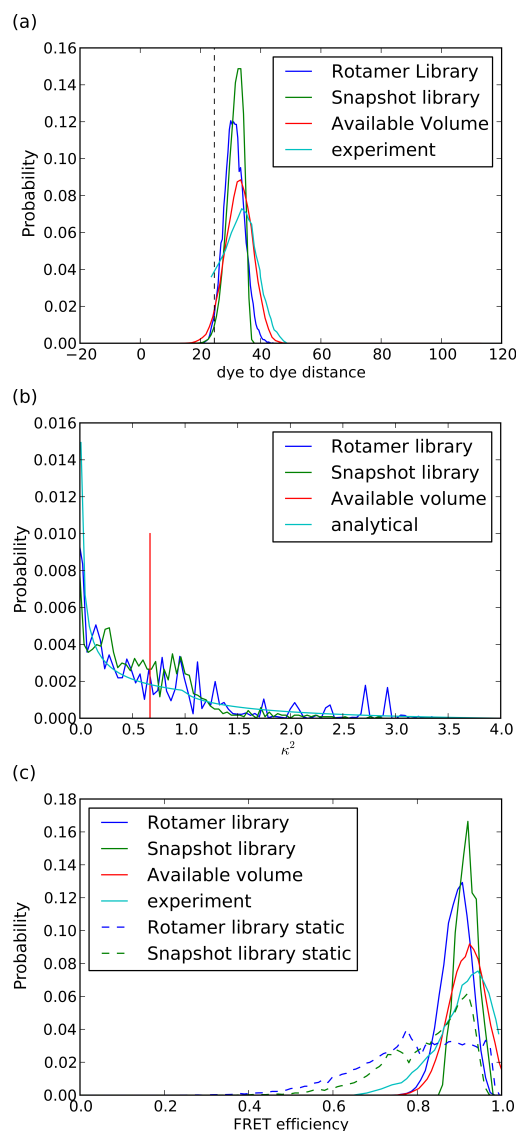


Fig. 7 The results of three methods compared to the experimental data of Hoffmann et al¹⁷. The black, dashed line on the distance distribution graph represents the distance between residues CYS 10 and CYS 67 in the crystal structure.

residues to which the dyes are attached, not the positions of the dyes themselves. In contrast to the polyproline case, here the distance between the residues (dashed line in fig 7a) is significantly different to the mean distance between the dyes. Such a difference is likely to exist in most molecular systems as the bulk of the host macromolecule will force the dyes to one side of the tethering point. This fact highlights the motivation for using computational methods for interpreting FRET data as presented here, as it provides a way to estimate the position of the tethering points from FRET efficiencies collected

from the mobile dyes.

3.3 Comparison to experimental data: LeuT

The most complex case in which we test the simple methods relates to smFRET experiments on the prokaryotic leucine transporter LeuT. The FRET data of this transporter labelled with Cy3 and Cy5 has been published by Zhao et al¹⁸. The molecular system of the LeuT embedded into the lipid membrane POPC has been built based on the crystal structure proposed by Yamashita et al³⁰ (fig 5c). Similarly to the case of the cold shock protein, to provide the functional groups necessary to attach the maleimide derivatives of dyes, we replaced residues His 7 and Arg 86 with cysteines (fig. 5c).

In the LeuT molecular system the translation between distance and efficiency, or the reverse, is not straightforward. The first reason is the fact that the preliminary MD simulations of dyes in water box showed that the free rotation of dyes is disturbed even if the only factor is the presence of one cysteine residue fixed at the end of the dye linker (see supplementary material†). Hence, any additional steric obstacles provided by the LeuT protein or lipid membrane will only increase a deviation from the isotropic approximation (Fig 8b). That is why the distribution of distances obtained two of the simplified methods, FD and RL, are slightly shifted in comparison with the experimental relation (Fig 8a) even if the transfer efficiency graphs nearly overlap (Fig 8c). The case of the AV method demonstrates that the deviations from the isotropic approximation for κ^2 may lead to an incorrect transition from the distance to the transfer efficiency distribution and cause the shift of the transfer efficiency peak (Fig 8c, red curve). In other words, assuming $\kappa^2 = 2/3$ when interpreting experiment is likely to yield an incorrect distance distribution.

Analogously to the CSP system, one of the features of the distribution of FRET efficiency in the static regime is the significant broadening. As previously, this is the result of considering the whole variety of κ^2 values instead of the mean value.

The difference between the mean dye separation and the separation of the residues to which they are tethered is even greater in this case than for CSP. The bulkiness of the LeuT protein (and the presence of the lipid) presents more constraints on the dye positions which highlights the difficulty in relating efficiency measurements to macromolecular structure.

4 Conclusions

In our studies, we wanted to address two issues which handicap the estimation of distances from FRET efficiency measurements. The first problem is caused by the dyes diffusing about their point of attachment. Using the simplified methods gives us an idea how much dye diffusion broadens the

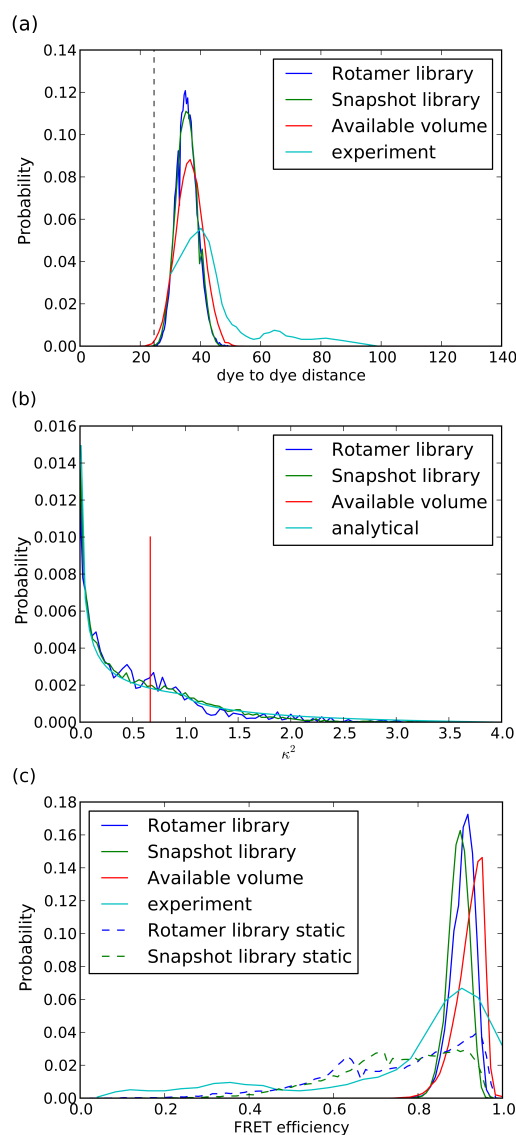


Fig. 8 Tests of the methods in a simple system. The characteristics of dye-labelled LeuT obtained using AV (red), RL (blue) and FD method (green) compared to the experiment of Zhao et al¹⁸ (cyan curve). The black, dashed line on the distance distribution graph represents the distance between residues CYS 7 and CYS 86 in the crystal structure.

distance and FRET distribution. Knowing this enables us to determine more exactly how much of the broadening comes from the dye diffusion and how much comes from dynamics of the host molecule. What is more, the simulations show how the mean separation of the dyes differs from that of the residues to which they are attached due to limitations on the space in which the dyes can move. After having a good agreement between experiment and simulations for a molecular sys-

tem, the real distance between specific parts of molecule can be determined more precisely using one of these approaches.

The second problem was the transfer's dependency on the orientation factor. Because of the limitation of the methods it is hard to address how interaction between dyes and protein influences κ^2 (something that could only be done with MD). Nevertheless we can still have some idea if a shape of protein can restrain the free rotation of the dyes (as in the LeuT case), and how it affects the final FRET efficiency distribution. Moreover our simulations help us to reveal the differences between two averaging regimes, static and dynamic.

The other possible application of simplified methods is to quickly test if a given protein model is consistent with experiments. Because of the short computational time needed in such methods, the elimination of improper models and choosing those which potentially can explain the experimental results can be more effective. It will save a lot of effort and allow us to focus on a smaller number of systems for further MD simulations.

The methods presented here involve many simplifications which mean that they will never be able to fully replicate the experimental situation. Most notably, as stated above, they all lack interactions between the dye and macromolecule or between the dyes themselves. It is likely that in many situations these interactions will be significant, vastly altering the probability of finding the dye in a given position and thus dramatically changing the FRET efficiency distribution. In addition these methods are all applied to a single static structure of the macromolecule, when it will itself be dramatically fluctuating and thus potentially altering the likely dye separations.

It is possible to extend these methods to account for these issues, for example weighting each dye position by calculated dye-protein interaction energy, or by allowing certain degrees of macromolecular motion. While this may yield small improvements it is always going to be difficult to quickly sample all interactions and conformations accurately and to account for hydration effects etc. If this level of detail is required, a better approach may simply be to run full MD simulations, which could give a good estimation of the intermolecular interactions if the issues of sampling and force field inaccuracies are accorded for. Nevertheless, for some applications it is important to be able to get a rough idea of how the dye diffusion and orientation will alter the relationship between E and r or to know that the dye separations are likely to be larger or smaller than that of the residues to which they are tethered. In this case, rapid methods such as those described here may be appropriate. While the FD method is the most accurate, all three methods produced similar qualitative results. Thus, the choice of which to use will depend on weighing up the computational time involved with the required level of detail. For example, once the libraries are made the RL approach gives a better result than FD for the same amount of computational

effort (i.e. same number of entries in the library), as it is a better reduction of the full MD trajectory. But if computational effort is not an issue then the FD approach can give a slightly better result by including a greater number of representative dye conformations.

To support the use of the simple methods by other researchers, we provide scripts as well as rotamer libraries and free dye snapshots for some of common dyes†.

5 Acknowledgments

The work has been partially supported by the Foundation for Polish Science with the International PhD Project MPD/2009-3/4. The scripts and libraries developed here are available for download at www.karri.anu.edu.au/handy.

References

- 1 T. Förster, *Discuss. Faraday Soc.*, 1959, **27**, 7–17.
- 2 J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Springer, 3rd edn, 2006.
- 3 L. Stryer, *Ann. Rev. Biochem.*, 1978, **48**, 819–846.
- 4 W. van der Meer, *Resonance Energy Transfer Theory and Data*, VCH, New York, 1994.
- 5 Y. Wang, H. DeBerg, T. Nomura, M. Tonks-Hoffman, P. Rohde, B. Martinac and P. R. Selvin, *Biophysical Journal*, 2013, **104**, 385a.
- 6 C.-H. Ma, Y.-T. Liu, C. G. Savva, P. A. Rowley, B. Cannon, H.-F. Fan, R. Russell, A. Holzenburg and M. Jayaram, *Journal of Molecular Biology*, 2014, **426**, 793–815.
- 7 U. B. Choi, P. Strop, M. Vrljic, S. Chu, A. T. Brunger and K. R. Weninger, *Nat Struct Mol Biol*, 2010, **17**, 318–324.
- 8 S. Sindbert, S. Kalinin, H. Nguyen, A. Kienzler, L. Clima, W. Bannwarth, B. Appel, S. Müller and C. A. M. Seidel, *J. Am. Chem. Soc.*, 2011, **133**, 2463–2480.
- 9 B. Corry, D. Jayatilaka, B. Martinac and P. Rigby, *Biophysical Journal*, 2006, **91**, 1032–1045.
- 10 B. A. Corry and D. Jayatilaka, *Biophys. J.*, 2008, **133**, 2711–2721.
- 11 M. Hoefling, N. Lima, D. Haenni, C. A. M. Seidel, B. Schuler and H. Grubmüller, *PLOS ONE*, 2011, **6**, e19791.
- 12 M. I. Fajer, H. Li, W. Yang and P. G. Fajer, *Journal of the American Chemical Society*, 2007, **129**, 13840–13846.
- 13 E. Deplazes, D. Jayatilaka and B. A. Corry, *Phys. Chem. Chem. Phys.*, 2011, **13**, 11045–11054.
- 14 Y. Polyhach, E. Bordignon and G. Jeschke, *Phys. Chem. Chem. Phys.*, 2011, **13**, 2356–2366.
- 15 Y. Polyhach and G. Jeschke, *Spectroscopy*, 2010, **24**, 651–659.
- 16 G. Jeschke, *Progr. Nucl. Magn. Reson. Spectrosc.*, 2013, **72**, 42–60.
- 17 A. Hoffmann, A. Kane, D. Nettels, D. E. Hertzog, P. Baumgärtel, J. Lengefeld, G. Reichardt, D. A. Horsley, R. Seckler, O. Bakajin and B. Schuler, *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 105–110.
- 18 Y. Zhao, D. Terry, L. Shi, H. Weinstein, S. C. Blanchard and J. A. Javitch, *Nature*, 2010, **465**, 188–193.
- 19 M. Matsumoto and T. Nishimura, *ACM Trans. Model. Comput. Simul.*, 1998, **8**, 3–30.
- 20 J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé and K. Schulten, *Journal of Computational Chemistry*, 2005, **26**, 1781–1802.
- 21 A. MacKerell, D. Bashford, M. Bellott, R. Dunbrack, J. Evanseck, M. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy,

- L. Kuchnir, K. Kuczera, F. Lau, C. Mattos, S. Michnick, T. Ngo, D. Nguyen, B. Prodhom, W. Reiher, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorcikiewicz-Kuczera, D. Yin and M. Karplus, *Journal of Physical Chemistry B*, 1998, **102**, 3586–3616.
- 22 S. Søndergaard and B. Corry, *Private communication*.
- 23 N. Michaud-Agrawal, E. J. Denning, T. B. Woolf and O. Beckstein, *Journal of Computational Chemistry*, 2011, **32**, 2319–2327.
- 24 M. Hoefling and H. Grubmüller, *Computer Physics Communications*, 2013, **184**, 841–852.
- 25 D. Badali and C. C. Gradinaru, *The Journal of Chemical Physics*, 2011, **134**, 225102.
- 26 B. Schuler, E. A. Lipman, P. J. Steinbach, M. Kumke and W. A. Eaton, *Proceedings of the National Academy of Sciences of the United States of America*, 2005, **102**, 2754–2759.
- 27 S. Doose, H. Neuweiler, H. Barsch and M. Sauer, *Proceedings of the National Academy of Sciences of the United States of America*, 2007, **104**, 17400–17405.
- 28 G. Horn, R. Hofweber, W. Kremer and H. R. Kalbitzer, *Cell Mol Life Sci*, 2007, **64**, 1457–70.
- 29 W. Kremer, B. Schuler, S. Harrieder, M. Geyer, W. Gronwald, C. Welker, R. Jaenicke and H. Kalbitzer, *Eur. J. Biochem.*, 2001, **268**, 2527–2539.
- 30 A. Yamashita, S. Singh, T. Kawate, Y. Jin and E. Gouaux, *Nature*, 2005, **437**, 215–223.