This is a repository copy of *CHARMM-DYES : Parameterization of fluorescent dyes for use with the CHARMM force field*.

JCTC
Journal of Chemical Theory and Computation

Article

CHARMM-DYES: Parameterization of Fluorescent Dyes for Use with the CHARMM Force Field

Robert A. Shaw, Tristan Johnston-Wood, Benjamin Ambrose, Timothy D. Craggs, and J. Grant Hill*

Cite This: *J. Chem. Theory Comput.* 2020, 16, 7817−7824

Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: We present CHARMM-compatible force field parameters for a series of fluorescent dyes from the Alexa, Atto, and Cy families, commonly used in Förster resonance energy transfer (FRET) experiments. These dyes are routinely used in experiments to resolve the dynamics of proteins and nucleic acids at the nanoscale. However, little is known about the accuracy of the theoretical approximations used in determining the dynamics from the spectroscopic data. Molecular dynamics simulations can provide valuable insights into these dynamics at an atomistic level, but this requires accurate parameters for the dyes. The complex structure of the dyes and the importance of this in determining their spectroscopic properties mean that parameters generated by analogy to existing parameters do not give meaningful results. Through validation relative to quantum chemical calculation and experiments, the new parameters are shown to significantly outperform those that can be generated automatically, giving better agreement in both the charge distributions and structural properties. These improvements, in particular with regard to orientation of the dipole moments on the dyes, are vital for accurate simulation of FRET processes.
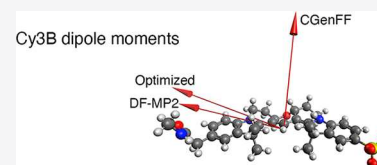
1. INTRODUCTION

The study of the structure and dynamics of biochemical systems, such as DNA and proteins, is a complex but important field of research. There are numerous methods for studying such systems, which encompass both experimental and computational approaches. One popular and successful method is based on the photophysical process, Förster resonance energy transfer (FRET).[1,2]

FRET occurs between a donor and acceptor fluorophore at sufficiently close separation, typically 1−10 nm.[3] The energy transfer also requires that the two fluorophores' absorption and emission spectra overlap and their electric dipoles are non-orthogonal. Typically, an electronically excited donor (D) will then transfer energy non-radiatively to an acceptor (A), which will relax to its ground state by radiating energy, and this process has been used to determine when two molecules are in close proximity.[4] Some studies have also used non-emissive acceptors as quenchers in FRET experiments. Donor and acceptor fluorophores can be covalently bound to sites in DNA or proteins and the FRET efficiency determined from the intensity of the fluorescence emissions. The experimental data can be used to determine the fluorophore separation[5] and thus both structural and dynamical information with the following equation

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

where E is the FRET efficiency—the quantum yield of the energy transfer—while r is the dye-pair separation. The value R_0 is the Förster radius, the separation of the dyes at which there is exactly 50% FRET efficiency. Since R_0 is kept fixed for

a specific donor−acceptor pair in a given solvent, the accurate determination of this value is essential in producing distance information. The expression for R_0 is given by[3]

$$R_0^6 \propto \frac{\phi_D \kappa^2 J(\lambda)}{n^4}$$

where ϕ_D is the fluorescence quantum yield of the donor without the acceptor present, κ describes the relative orientation of D and A (see below), n is the refractive index of the intervening medium, and $J(\lambda)$ is the overlap integral of the D emission spectrum with the A absorption spectrum.

κ^2 is often termed the dipole orientation factor, as κ is defined as

$$\kappa = \vec{\mu}_D \cdot \vec{\mu}_A - 3(\vec{\mu}_D \cdot \vec{R}_{DA})(\vec{\mu}_A \cdot \vec{R}_{DA}) \tag{1}$$

where $\vec{\mu}_i$ is the normalized transition dipole moment of a fluorophore and $\vec{R}_{DA}$ is the normalized displacement between the centers of D and A. The majority of FRET experiments make the so-called "κ^2 approximation", which assumes that the average value of κ^2 is 2/3, corresponding to the value expected if the dipoles orient themselves isotropically over a sphere.[6] There is also an inherent assumption that the rate of rotational motion of the dyes is fast compared to the excited state

lifetime.[7,8] Nonetheless, the FRET technique has been shown to give accurate distance measurements within a 3 Å precision for separations between 30 and 100 Å,[9] which has allowed the use of such measurements to determine biomolecular structure and dynamics.[8,10−12] However, the reliability of FRET below 40 Å is unknown. This is thought to be partially due to the validity of assuming the isotropic limit of κ^2 in this regime.[13,14]

Computational methods, such as molecular dynamics (MD), have also been used to study the dynamics of FRET dyes.[15,16] These MD simulations have helped elucidate the dynamics between the fluorophores and the biochemical system of interest. A study by Shoura *et al.*, investigating the behavior of two fluorophore dyes, Atto 594 and Atto 647N, found that κ^2 was not 2/3 when the dyes were separated by 2 nm but was in fact 0.33.[15] The authors of the said paper note, however, that force field parameterization will strongly affect the observed behavior of the dyes; hence, accurate parameterization is key to obtaining reliable simulation data. In particular, the charge parameters will directly change the predicted dipole moments and thus the dipole orientations of eq 1, while the relative rigidity of the dyes—reflected in bonded parameters—will affect their rotational freedom. More careful parameterization of these dyes is thus essential.

There has been work by Graen *et al.*[17] giving parameters for some of the most popular dyes, but for the AMBER force field.[18] These built upon previous work on determining charges for specific dyes for both the AMBER and CHARMM27 force fields.[19−22] However, these are either not directly transferable to other force fields, or with the exception of the AMBER-DYES force field, done by analogy. As will be discussed later, the CHARMM parameterization protocol is perhaps more well-suited than the general AMBER force field (GAFF) procedure to the physics of the dye interactions. In the present work, extensive force field parameterization of a series of fluorescent dyes commonly used in FRET experiments has been carried out, for use with the CHARMM protein and nucleic acid force fields. FRET dyes are large, presenting a difficult problem to parameterize, and the accurate description of their charge distributions in particular is vitally important for meaningfully studying their dynamics. We demonstrate that due to the high anisotropy of the electron density, automatic generation of parameters by analogy using standard tools does not provide sufficient accuracy and present new parameters that show much better agreement with quantum-mechanical results.

2. PARAMETERIZATION

The CHARMM force field,[23] as with most of the standard point-charge-based classical force fields, divides the parameters into two classes: bonded and non-bonded. The former comprises harmonic bond-stretching, torsional, and improper dihedral terms, mediated by force constant parameters k, and equilibrium values. Added to this are dihedral terms described by a cosine-based Fourier series, controlled by force constants, frequencies n, and phases δ. The non-bonded terms comprise van der Waals forces in the form of Lennard-Jones potentials, with the usual parameters of well-depth and equilibrium bond distance, and the Coulombic interaction between electrostatic point charges, q.

In this work, we follow the well-documented CHARMM general force-field (CGenFF) scheme for optimizing these parameters,[23,24] with some adaptations, which will be described in this section. As a starting point, we have taken the parameters automatically generated by the CGenFF program; all results will be compared to these. It should be noted that the heuristic measure of confidence in the generated values given by the program indicated a strong need for reparameterization of almost all of the bonded parameters. Following convention,[23] we do not reoptimize the Lennard-Jones parameters but take those by analogy from the existing CGenFF parameters.

2.1. Dyes and Atom Types. The fluorescent dyes used experimentally are varied and many, depending on the absorption and emission frequencies needed for the particular experiment. In addition, these will usually be attached to a protein or nucleic acid via some kind of organic linker. Providing parameters for every possible combination of dye, linker, and attachment point would require considerable effort and would not make use of the commonalities between the different systems. To this end, we restrict our attention here to a representative subset, which can easily be extended by analogy to other similar dye systems.

In broad terms, the commercially available dyes can be divided into three families: the Alexa, Atto, and Cy dyes. Within these, we present parameters for Alexa Fluor 647, Atto 550, Atto 647N, Cy3, Cy3B, Cy5, and Cy7. Two of these structures are given in Figure 1 for reference, along with the classification of their atoms into types. These types were based on those generated by the CGenFF program then adapted to fully exploit the similarities in structure between the different
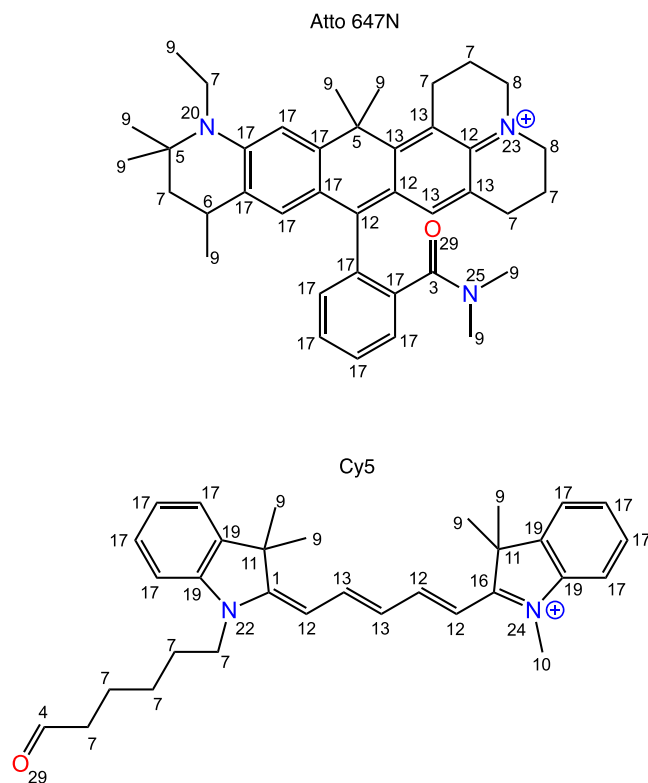
Figure 1. 2D structures of two of the dyes parameterized herein: Atto 647N and Cy5. Non-hydrogen atoms are numbered by their assigned atom type. There are also five hydrogen types, depending on their attachment: aromatic; aliphatic on a primary, secondary or tertiary carbon; on an alkene. The other dyes and a table of which atom types the numbers correspond to can be found in the Supporting Information.

fluorophores. Similar diagrams for the dyes not shown in the figure are given in the Supporting Information.

The linkers used vary greatly in length and the method of attachment. However, they are all structurally simple in comparison, involving only standard functional groups which are already well-parameterized by CGenFF. As such, we only give one such linker with our parameters, using a common attachment to the dyes, and other variants can be generated as needed from this. Similarly, while the linker may be attached to any given amino acid or nucleobase, the point of attachment with the linker is common, so that we give only one example of each attachment. The focus of this work is on the dyes themselves.

As can be seen in Figure 1, each dye has regions of planarity and extended conjugation. This causes a rigidity in the structure that is not well described by the default parameterizations and hence the need for refitting of the bonded parameters. Moreover, the electronic delocalization caused by these structural features means that the electron density is highly anisotropic, with regions of concentrated charge, which are not described well by the generic point charges. In this respect, the use of CHARMM over, for example, AMBER is prudent, as the former models the point charges based on interactions with water molecules, which allows for a more accurate description of this anisotropy. In contrast, a similar study parameterizing other dyes for the AMBER force field needed to adapt the charge optimization process to account for this.[17]

2.2. Charges. The accuracy of the point charges is particularly important as the intended application is in the study of the interactions between the dyes and modeling spectroscopic events based on these dynamics. Overly isotropic parameters will lead to a poor description of the interactions between dye pairs that mediate energy transfer, resulting in erroneous dynamics. On the other hand, unbalanced, overly strong charges would lead to the dye pairs being held tightly together indefinitely, whereas experimentally it is expected that there should remain a fairly high degree of rotational and translation freedom in the diffusion of the dyes.[25]

The standard procedure in CHARMM is to optimize the structures of the new residues at the Møller–Plesset second-order perturbation theory (MP2) with the 6-31G* split-valence basis set. However, as will be discussed later, the need to generate a Hessian makes this prohibitively expensive, and the Hessian needs to be generated at the same level of theory as the optimization to be valid. As such, we optimized the dye structures using density-fitted (DF) MP2[26] with the 6-31G* basis[27,28] and the associated auxiliary MP2 fitting sets[29] in the MOLPRO suite of programs.[30,31] Density fitting reduces the scaling of the calculation by an order of magnitude such that even a numerical Hessian takes considerably less resource to calculate. To the authors' knowledge, it has not been used before in a force-field parameterization, but previous benchmark studies have demonstrated that structural parameters and force constants from DF-MP2 only deviate from full MP2 on the order of less than a single percent.[32,33] In particular, a study of the effects of density fitting on harmonic vibrational frequencies (which are determined entirely by the Hessian) showed that frequencies agreed with canonical MP2 to well within a single reciprocal centimeter.[34] This is therefore expected to be a negligible source of error compared to the error inherent in the fitting procedure.

The charge optimizations then proceeded by generating potential energy curves at the Hartree–Fock (HF)/6-31G* level of theory for a water molecule interacting with each non-hydrogen point of contact, using Gaussian '09.[35] These points of contact were generated automatically by the force field toolkit (ffTK) program[36] and then pruned so as to remove interactions where the water molecule would not have access to the site. Following the CGenFF philosophy, the curves were generated as idealized hydrogen-bond interactions between the DF-MP2-optimized geometry of the dye and a single TIP3P geometry water molecule. The intermolecular separation was scanned, keeping all other geometrical parameters fixed, to determine the interaction energy minimum, without a counterpoise correction. The force field charges were then optimized by a constrained least-squares minimization from the molecular mechanics (MM)-calculated potential energies compared to the quantum mechanical (QM) potential energies. An L^2 regularization term was added to avoid overfitting the charges, avoiding potential problems with overly localized charges. This process was then repeated iteratively until the overall objective function changed by less than 0.1%. To maintain compatibility with the CHARMM force field, the aliphatic and aromatic hydrogen charges were constrained to be +0.09 and +0.15e, respectively.

2.3. Bonded Parameters. For the harmonic terms, we followed the analytical partial Hessian fitting approach of Wang and coworkers, using their ParmHess code.[37] We adapted the code to read the outputs from the DF-MP2 numerical Hessian as calculated in MOLPRO. The dyes are fairly large molecules for MP2 (at around 90 atoms each), so it would be considerably cheaper to use a density functional instead. However, the standard CHARMM procedure is to use MP2, and it is vitally important that any new parameters are compatible with the rest of the force field. While density fitting introduces an error, this is relatively small both in terms of geometries and frequencies. It is therefore likely that the DF-MP2 results will be much closer to the full MP2 results than if we used a different method and, as a result, will fit better with the rest of the force field.

For each bond-stretch, angle torsion, and improper dihedral, an MM Hessian for just the two terminating atoms in the term is calculated and the corresponding parameters fitted to the equivalent portion of the QM Hessian. This is repeated iteratively through the list of terms until the parameters change by less than one percent. The procedure is highly dependent on the choice of starting guess: we found that in particular for the bonds involved in the conjugated portions of the dye, the guesses generated by CGenFF are too large and lead to unphysical rigidity (i.e., very large force constants). This problem could be avoided by using a starting guess 0.6 to 0.8 times the original guess, as determined by careful reoptimization at each stage.

Finally, the dihedral terms represent a particular challenge, due to the simultaneous optimization of multiple parameters in a variable length Fourier series. Potential energy scans of the proper dihedrals were performed at the DF-MP2/6-31G* level. We then took the multiplicity of each Fourier series by analogy with similar dihedrals already in CGenFF, usually as generated automatically. As with the charge optimization, MM scans were then performed with the new parameters, which are then iteratively fitted in a least-squares optimization against the QM results. This was performed by direct calculation of the dihedral term in Python, minimized with the standard SciPy
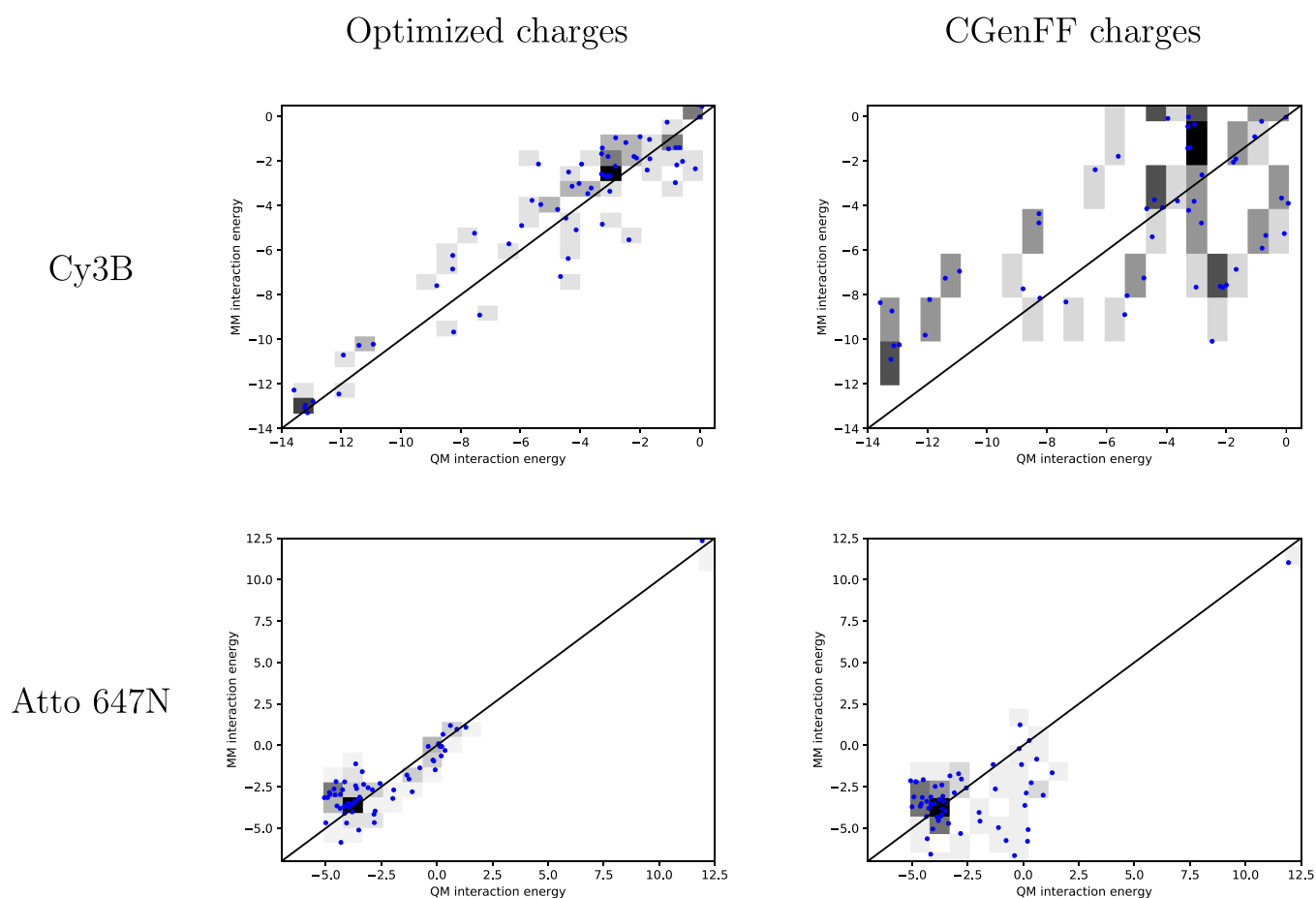
Figure 2. Interaction energies (in kJ mol^{-1}) of Cy3B and Atto 647N with water as calculated using the CHARMM force field with either the new, optimized charges or those from the initial CGenFF guess. These are compared to those calculated at the HF/6-31G* level, overlaid with the ideal line, $y = x$.

BFGS algorithm.[38] After each iteration, we inspected the total error in each dihedral term and increased the multiplicity of any terms that showed large errors. These were then refitted until all parameters changed by less than one percent.

2.4. Simulation Details. All MD simulations were carried out in GROMACS 2016.4,[39] with the TIP3P model for water.[40] There is evidence that the choice of water model used with the CHARMM force field can have a large effect on results, especially binding affinities.[41] However, it is not clear that different models are better or worse, with TIP3P often giving closer agreement to experiments than the more expensive TIP5P model, for example.[42] As the CHARMM force field was parametrized with TIP3P specifically, we choose to use this model.

Positively charged dyes (both ATTO dyes, Cy3, Cy5, and Cy7) were neutralized with chloride ions, while Alexa 647 (−3 charge) was neutralized with sodium ions. The box dimensions and number of water molecules for each simulation are given in the Supporting Information. The parameters developed here are designed to be used with the CHARMM36 force field, although none of the MD simulations herein required additional parameters beyond the water and ion parameters. Temperature was controlled with a Langevin thermostat to a target temperature of 300 K, while a pressure of 1 bar was maintained with a Parinello−Rahman barostat with a time constant of 2 ps. A time step of 2 fs was used throughout. A short-range van der Waals cutoff of 1.4 nm was employed, and

the particle-mesh Ewald method used to calculate long−range electrostatic interactions.

In each MD simulation, we minimized the energy before equilibrating the water and ions for 500 ps to 300 K, before equilibrating the dyes for a further 500 ps. The production runs were all 50 ns long. Diffusion coefficients have been calculated by calculating the mean-square deviation of the bare dye in a cubic box of water expanded 1 nm on all sides around the centered dye. A straight line was then fitted to the mean-square deviation results from 10 ns onward, to ensure equilibration had been achieved. Convergence was tested by measuring the discrepancy between diffusion coefficients measured in 10 ns intervals, and in each case, this was found to be less than the uncertainty in the least-squares fit. IR spectra were calculated using VMD's spectral analysis tool[43] on the last 40 ns of the simulation and compared to the DF-MP2 spectra of the gas-phase dye to determine an empirical shift for the calculated frequencies.

3. RESULTS

3.1. Charges and Electric Dipole Moments. There is a paucity of experimental data on FRET dyes with which we can meaningfully compare the results from the MD simulations. The standard approach would be to compare with known crystal structures in proteins or to compute bulk quantities, neither of which is possible; to the authors' knowledge, no such crystal structures are available, and FRET dyes are far too

costly to study in the bulk. As such, we will instead focus on comparing to quantum-mechanical results, in particular of the charge distributions and structural properties, while also demonstrating the performance compared to the CGenFF-generated parameters.

Figure 2 shows the distribution of interaction energies with water from the charge optimization portion of the parameterization, for two representative dyes. As can be seen from the figure, the CGenFF results vary significantly from the quantum mechanical ones, in particular for Cy3B where there is essentially no correlation. In contrast, the new parameters show clear correlation, with overall much smaller errors. A closer analysis of where the main improvements lie shows that, in both cases, it is the water interactions with carbons in the central rings that decrease in error in the course of the optimization. This is a reflection of the strong electronic conjugation in these regions of the molecule, causing significant anisotropy in the charge distribution. There are no similar systems in CGenFF with which reasonable analogy can be drawn. The resulting electric molecular dipole moments for Cy3B and Atto 647N are shown in Figure 3, where it is
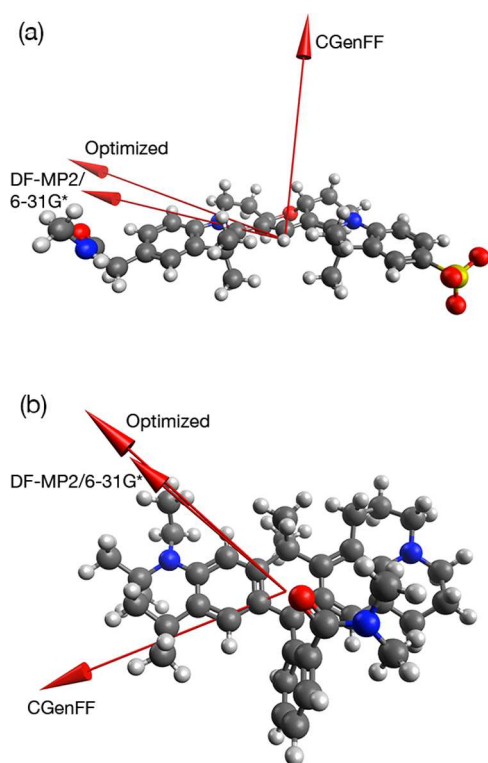

Figure 3. Electric dipole moments, originating from the center of mass, as calculated quantum mechanically using DF-MP2/6-31G*, or from MD simulations using the optimized parameters or the initial CGenFF guess. (a) Cy3B dipole moments scaled by a factor of 0.3. (b) Atto 647N dipole moments scaled by a factor of 0.7. In all cases, the DF-MP2/6-31G*-optimized geometry is shown.

particularly striking that the molecular dipole as calculated with the optimized parameters has the correct orientation, whereas the CGenFF results gave dipoles orientated 60−90° away from the DF-MP2 dipole. Analogous trends were also observed for the other dyes under consideration, with the molecular dipoles calculated with the optimized parameters within 1.2 to 1.4 times the QM magnitude and possessing the correct orientation. The implications for simulations of FRET

dynamics could be significant—inaccurate orientation of the electric dipole moments could lead to erroneous rotational behavior of the dyes, affecting the value of κ^2. As discussed in the Introduction, this will in turn strongly affect computed FRET efficiencies.

A second, less clear, validation of the charges can be found through computation of the dyes' infrared spectra. Theoretically, such spectra depend strongly on both the bond force constants and the local dipoles (thus the charges) in those bonds. However, no experimental IR spectra are available in the literature; while we could instead calculate UV−visible spectra by performing quantum-mechanical calculations on snapshots from the MD, this would be an even less direct validation. As such, we chose two dyes (Atto 647N and Cy3B) to determine IR spectra for and compare to the results from the MD simulations. The comparison is clouded further by the dyes used in the experiment containing N-hydroxysuccinimide (NHS) groups that are required for the reaction to conjugate them to biomolecules. The experimental and computed spectra are compared in Figure S10 of the Supporting Information, alongside details of the experimental procedure, and it can be seen that computation qualitatively reproduces the key peaks around the 3000 and 1000−2000 wavenumber region. It should be noted that the reactive NHS groups leave during the reaction with biomolecules; hence, they are not included in the force field parameters or any of the simulations reported.

3.2. Validation of Bonded Parameters. A clearer validation of the bonded parameters can be found through direct comparison of the QM geometry with the energy-minimized MM geometries, as determined using both the new and CGenFF parameters. Figure 4 shows these for the Atto
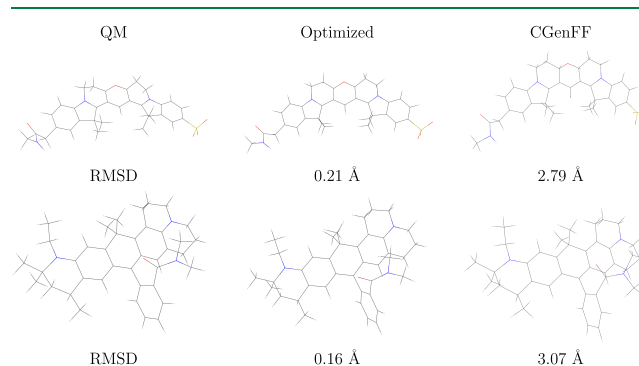
Figure 4. Optimized structures of the dyes Cy3B (top) and Atto 647N (bottom), at the DF-MP2/6-31G* level (left), and from energy minimization with the optimized (middle) and CGenFF (right) parameters. Root-mean-square deviations of the optimized and CGenFF structures from the QM results are given below each structure.

647N and Cy3B dyes. The CGenFF structures have root-mean-square deviations from the QM structure an order of magnitude larger than found with the new parameters. In particular, the CGenFF bond-stretching force constants are much larger, resulting in the central rings being poorly described. There is also significant difference in the conformations of the various methyl groups on the peripheries of each dye. The differences are again most likely entangled with the differing charge distributions, and it is impossible to separate the effects. Overall, however, the new parameters give good agreement with the QM results, with RMSD values on the order of 0.2 Å.

Finally, we have calculated diffusion constants for a subset of the dyes, based on the 50 ns simulations in water. The diffusional mobility of the dyes is an important property in determining the FRET efficiency, such that good agreement in diffusion constants is a key requirement on our simulations. Table 1 gives these along with experimentally determined

Table 1. Diffusion Constants for Three of the Dyes in Water, as Calculated from Mean-Square-Deviation Plots of Simulations Using the New, Optimized Parameters, and the Original CGenFF Guesses[a]

	diffusion constant (10^{-6} cm^2 s^{-1})		
dye	experimental	optimized	CGenFF
Alexa Fluor 647	2.8 ± 0.1	2.9 ± 0.7	4.5 ± 1.5
Cy5	3.6 ± 0.1	2.7 ± 0.8	5.0 ± 1.4
Atto 550[b]	4.0 ± 0.1	3.1 ± 0.7	5.2 ± 1.8

[a]These are compared to experimental values from refs 44 and 45, determined in water at 298.15 K. [b]Experimental value for Atto 488-carboxylic acid as Atto 550 data not available.

values from the manufacturers, where available. As each dye family has broadly similar structures, their diffusion constants are also similar, such that we only consider one from each family. This is in part also due to the limited availability of experimental data. It should also be noted that the experimental data was determined at a slightly different temperature (298.15 K compared to 300 K in the simulations), but this is unlikely to make a large difference. However, while the calculated results do not quite agree quantitatively with the experimental ones, the new parameters do appear to perform slightly better than the CGenFF ones, which consistently overestimate. The sample size is too small to make any definitive statements about the significance of this, and there is some overlap in the error distributions between the two. That the uncertainty for the optimized parameters is overall less, on the other hand, is an indirect indicator of improvement, but again it is difficult to characterize whether that is an improvement. This is perhaps due to the differences in charge distributions, with the inaccurate interactions with water from Figure 2 affecting the ability of the dyes to diffuse through water. There is also a possibility that finite size effects could be artificially increasing the diffusion constants.

4. CONCLUSIONS

New molecular mechanics parameter sets, compatible with the CHARMM force field, have been designed, developed, and tested for a group of fluorescent dyes commonly used in FRET experiments. Specifically, this includes dyes from the Alexa (Alexa Fluor 647), Atto (Atto 550 and Atto 647N), and Cy (Cy3, Cy3B, Cy5 and Cy7) families, along with an organic linker of the type used to attach a dye to a protein or nucleic acid. Both the dyes and the linker are presented as a representative subset, and the commonality between systems means that parameters for similar dyes and linkers can be straightforwardly generated by analogy. Validation of the new parameter sets with quantum mechanical benchmark data demonstrates that they produce significantly more accurate results than parameters automatically generated using CGenFF. The accurate description of both the charge distribution and conformational flexibility of the dyes is vital to give realistic dynamics, as these dyes have highly delocalized

electronic structures and are more rigid than the typical biomolecules studied using CHARMM.

The accurate charges presented a particular difficulty, as the optimization of these dyes at a high-level of theory is very computationally expensive. To this end, we used density-fitted MP2 rather than the HF level used previously in optimization of parameters for use with the AMBER force field.[17] The resulting optimized charges much better reproduced the dipole orientations and magnitudes than the analogy-assigned (automatically generated) parameters, with the latter giving electric dipole moments orientated 60–90° away from the quantum mechanical result. As the FRET efficiency directly depends on the orientation factor, κ^2, this has important implications for future simulations of FRET dyes; both the charge distribution and the relative orientation of the dyes will be affected by the force field parameters chosen. Additionally, the new bonded parameters give diffusion constants in better agreement with experiments and better reproduce the DF-MP2 geometry. This should facilitate probing whether the assumptions used in FRET studies—specifically, the rotational diffusion is much faster than the excitation event—are realistic.

There are still assumptions implicit in our model, however. First, we have had to assume that the excited-state electronic structure is not significantly different from that of the ground state. Previous studies have suggested that this assumption is reasonable, though only for specific dyes.[19] If this were not the case, however, then the use of a CHARMM-like force field would perhaps be ill-advised. While it is certainly possible to characterize and thus parameterize the excited state directly, any simulation would depend heavily on the lifetime of that excited state, and simulating the excitation process itself is a considerably more complicated problem.

Second, we are constrained by the fixed point-charge approximation used in the CHARMM force field. This may not necessarily be appropriate as the delocalized electronic structure of the dyes is likely highly polarizable. One possibility to address this would be to add polarizable centers to the model; this then presents the question of where to put these centers, and how many are needed. For example, in the Drude polarizable force field,[46] centers are added to each polarizable atom by an imaginary spring, modeling the electronic degrees of freedom of that center. This has allowed for reproduction of ab initio polarizabilities at a relatively low cost but is still some four times more computationally intensive than the CHARMM-like force fields considered in this work. It would also require a considerable amount of additional parameterization. Both this and characterization of the excited-state structures would be interesting but challenging avenues of future research. The parameters presented here, in the meantime, will improve our ability to describe the dynamics of dyes attached to biomolecules and how those dynamics may affect FRET events.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.0c00721.

> Additional figures of all the dyes with atom types shown and dye structures and optimized parameters in GROMACS-compatible format (PDF)

■ AUTHOR INFORMATION

Corresponding Author

  J. Grant Hill − *Department of Chemistry, University of Sheffield, Sheffield S3 7HF, U.K.;* ◉ orcid.org/0000-0002-6457-5837; Email: grant.hill@sheffield.ac.uk

Authors

  Robert A. Shaw − *Department of Chemistry, University of Sheffield, Sheffield S3 7HF, U.K.; Present address: ARC Centre of Excellence in Exciton Science, School of Science, RMIT University, Melbourne, VIC 3000, Australia;* ◉ orcid.org/0000-0002-9977-0835

  Tristan Johnston-Wood − *Department of Chemistry, University of Sheffield, Sheffield S3 7HF, U.K.; Present address: Department of Chemistry, Physical and Theoretical Chemistry Laboratory, University of Oxford, Oxford OX1 3QZ, U.K.*

  Benjamin Ambrose − *Department of Chemistry, University of Sheffield, Sheffield S3 7HF, U.K.*

  Timothy D. Craggs − *Department of Chemistry, University of Sheffield, Sheffield S3 7HF, U.K.;* ◉ orcid.org/0000-0002-7121-0609

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jctc.0c00721

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Förster, T. Energy transfer and fluorescence between molecules. *Ann. Phys.* **1948**, *437*, 55−75.

(2) Beljonne, D.; Curutchet, C.; Scholes, G. D.; Silbey, R. J. Beyond Förster resonance energy transfer in biological and nanoscale systems. *J. Phys. Chem. B* **2009**, *113*, 6583−6599.

(3) Hohlbein, J.; Craggs, T. D.; Cordes, T. Alternating-laser excitation: single-molecule FRET and beyond. *Chem. Soc. Rev.* **2014**, *43*, 1156−1171.

(4) Hillisch, A.; Lorenz, M.; Diekmann, S. Recent advances in FRET: distance determination in protein−DNA complexes. *Curr. Opin. Struct. Biol.* **2001**, *11*, 201−207.

(5) Piston, D. W.; Kremers, G.-J. Fluorescent protein FRET: the good, the bad and the ugly. *Trends Biochem. Sci.* **2007**, *32*, 407−414.

(6) Muñoz-Losa, A.; Curutchet, C.; Krueger, B. P.; Hartsell, L. R.; Mennucci, B. Fretting about FRET: failure of the ideal dipole approximation. *Biophys. J.* **2009**, *96*, 4779−4788.

(7) Deplazes, E.; Jayatilaka, D.; Corry, B. Testing the use of molecular dynamics to simulate fluorophore motions and FRET. *Phys. Chem. Chem. Phys.* **2011**, *13*, 11045.

(8) Kalinin, S.; Peulen, T.; Sindbert, S.; Rothwell, P. J.; Berger, S.; Restle, T.; Goody, R. S.; Gohlke, H.; Seidel, C. A. M. A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat. Methods* **2012**, *9*, 1218−1225.

(9) Craggs, T. D. Cool and dynamic: single-molecule fluorescence-based structural biology. *Nat. Methods* **2017**, *14*, 123.

(10) Nagy, J.; Grohmann, D.; Cheung, A. C. M.; Schulz, S.; Smollett, K.; Werner, F.; Michaelis, J. Complete architecture of the archaeal RNA polymerase open complex from single-molecule FRET and NPS. *Nat. Methods* **2015**, *6*, 6161.

(11) Hellenkamp, B.; Wortmann, P.; Kandzia, F.; Zacharias, M.; Hugel, T. Multidomain structure and correlated dynamics determined by self-consistent FRET networks. *Nat. Methods* **2017**, *14*, 174−180.

(12) Craggs, T. D.; Sustarsic, M.; Plochowietz, A.; Mosayebi, M.; Kaju, H.; Cuthbert, A.; Hohlbein, J.; Domicevica, L.; Biggin, P. C.; Doye, J. P. K.; Kapanidis, A. N. Substrate conformational dynamics drive structure-specific recognition of gapped DNA by DNA polymerase. *Nucleic Acids Res.* **2019**, *47*, 10788.

(13) Sanborn, M. E.; Connolly, B. K.; Gurunathan, K.; Levitus, M. Fluorescence properties and photophysics of the sulfoindocyanine Cy3 linked covalently to DNA. *J. Phys. Chem. B* **2007**, *111*, 11064−11074.

(14) VanBeek, D. B.; Zwier, M. C.; Shorb, J. M.; Krueger, B. P. Fretting about FRET: Correlation between κ and R. *Biophys. J.* **2007**, *92*, 4168−4178.

(15) Shoura, M. J.; Ranatunga, R. J. K. U.; Harris, S. A.; Nielsen, S. O.; Levene, S. D. Contribution of fluorophore dynamics and solvation to resonant energy transfer in protein-DNA complexes: a molecular-dynamics study. *Biophys. J.* **2014**, *107*, 700−710.

(16) Reinartz, I.; Sinner, C.; Nettels, D.; Stucki-Buchli, B.; Stockmar, F.; Panek, P. T.; Jacob, C. R.; Nienhaus, G. U.; Schuler, B.; Schug, A. Simulation of FRET dyes allows quantitative comparison against experimental data. *J. Chem. Phys.* **2018**, *148*, 123321.

(17) Graen, T.; Hoefling, M.; Grubmüller, H. AMBER-DYES: Characterization of Charge Fluctuations and Force Field Parameterization of Fluorescent Dyes for Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 5505−5512.

(18) Schepers, B.; Gohlke, H. AMBER-DYES in AMBER: Implementation of fluorophore and linker parameters into Amber-Tools. *J. Chem. Phys.* **2020**, *152*, 221103.

(19) Corry, B.; Jayatilaka, D. Simulation of Structure, Orientation, and Energy Transfer between AlexaFluor Molecules Attached to MscL. *Biophys. J.* **2008**, *95*, 2711−2721.

(20) Schröder, G. F.; Alexiev, U.; Grubmüller, H. Simulation of Fluorescence Anisotropy Experiments: Probing Protein Dynamics. *Biophys. J.* **2005**, *89*, 3757−3770.

(21) Allen, L. R.; Paci, E. Simulation of fluorescence resonance energy transfer experiments: effect of the dyes on protein folding. *J. Phys.: Condens. Matter* **2010**, *22*, 235103.

(22) Vaiana, A. C.; Schulz, A.; Wolfrum, J.; Sauer, M.; Smith, J. C. Molecular mechanics force field parameterization of the fluorescent probe rhodamine 6G using automated frequency matching. *J. Comput. Chem.* **2003**, *24*, 632−639.

(23) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671−690.

(24) Vanommeslaeghe, K.; MacKerell, A. D., Jr. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **2012**, *52*, 3144−3154.

(25) Wallace, B.; Atzberger, P. J. Förster resonance energy transfer: Role of diffusion of fluorophore orientation and separation in observed shifts of FRET efficiency. *PLoS One* **2017**, *12*, No. e0177122.

(26) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. Use of approximate integrals in ab initio theory. An application in MP2 energy calculations. *Chem. Phys. Lett.* **1993**, *208*, 359−363.

(27) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724−728.

(28) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-

Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257−2261.

(29) Tanaka, M.; Katouda, M.; Nagase, S. Optimization of RI-MP2 auxiliary basis functions for 6-31G** and 6-311G** basis sets for first-, second-, and third-row elements. *J. Comput. Chem.* **2013**, *34*, 2568−2575.

(30) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M. Molpro: a general-purpose quantum chemistry program package. *WIREs Comput. Mol. Sci.* **2012**, *2*, 242−253.

(31) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Celani, P.; Györffy, W.; Kats, D.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Shamasundar, K. R.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; O'Neill, D. P.; Palmieri, P.; Peng, D.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M. *MOLPRO*; version 2015.1, a package of ab initio programs. 2015; see http://www.molpro.net.

(32) Werner, H.-J.; Manby, F. R.; Knowles, P. J. Fast linear scaling second-order Møller-Plesset perturbation theory (MP2) using local and density fitting approximations. *J. Chem. Phys.* **2003**, *118*, 8149−8160.

(33) Lutz, O. M. D.; Rode, B. M.; Bonn, G. K.; Huck, C. W. The performance of RI-MP2 based potential energy surfaces in a vibrational self-consistent field treatment. *Chem. Phys. Lett.* **2015**, *619*, 66−70.

(34) Hrenar, T.; Rauhut, G.; Werner, H.-J. Impact of Local and Density Fitting Approximations on Harmonic Vibrational Frequencies. *J. Phys. Chem. A* **2006**, *110*, 2060−2064.

(35) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*; *Revision D.01*.Gaussian, Inc.: 2009.

(36) Mayne, C. G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J. C. Rapid parameterization of small molecules using the force field toolkit. *J. Comput. Chem.* **2013**, *34*, 2757−2770.

(37) Wang, R.; Ozhgibesov, M.; Hirao, H. Partial hessian fitting for determining force constant parameters in molecular mechanics. *J. Comput. Chem.* **2016**, *37*, 2349−2359.

(38) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, I.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; SciPy 1.0 Contributors. SciPy 1.0−Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261−272.

(39) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19−25.

(40) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(41) Nguyen, T. T.; Viet, M. H.; Li, M. S. Effects of Water Models on Binding Affinity: Evidence from All-Atom Simulation of Binding of Tamiflu to A/H5N1 Neuraminidase. *Sci. World J.* **2014**, *2014*, 1−14.

(42) Onufriev, A. V.; Izadi, S. Water models for biomolecular simulations. *WIREs Comput. Mol. Sci.* **2018**, *8*, No. e1347.

(43) Humphrey, W.; Dalke, A.; Schulten, K. VMD : Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.

(44) Loman, A.; Dertinger, T.; Koberling, F.; Enderlein, J. Comparison of optical saturation effects in conventional and dual-focus fluorescence correlation spectroscopy. *Chem. Phys. Lett.* **2008**, *459*, 18−21.

(45) Jung, C.; Lee, J.; Kang, M.; Kim, S. W. Measurement of the temperature-dependent diffusion properties of nanoparticles by using fluorescence correlation spectroscopy. *J. Korean Phys. Soc.* **2014**, *65*, 1083−1089.

(46) Lamoureux, G.; Roux, B. Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.* **2003**, *119*, 3025−3039.