# AMBER-DYES: Characterization of Charge Fluctuations and Force Field Parameterization of Fluorescent Dyes for Molecular Dynamics Simulations

Timo Graen, Martin Hoefling, and Helmut Grubmüller*

Max Planck Institute for Biophysical Chemistry, Am Faßberg 11, 37077 Göttingen, Germany

Ⓢ Supporting Information

ABSTRACT: Recent advances in single molecule fluorescence experiments and theory allow a direct comparison and improved interpretation of experiment and simulation. To this end, force fields for a larger number of dyes are required which are compatible with and can be integrated into existing biomolecular force fields. Here, we developed, characterized, and implemented AMBER-DYES, a modular fluorescent label force field, for a set of 22 fluorescent dyes and their linkers from the Alexa, Atto, and Cy families, which are in common use for single molecule spectroscopy experiments. The force field is compatible with the AMBER protein force fields and the GROMACS molecular dynamics simulation program. The high electronic polarizability of the delocalized π-electron orbitals, as found in many fluorescent dyes, poses a particular challenge to point charge based force fields such as AMBER. To quantify the charge fluctuations due to the electronic polarizability, we simulated the 22 dyes in explicit solvent and sampled the charge fluctuations using QM/MM simulations at the B3LYP/6-31G*//TIP3P level of theory. The analysis of the simulations enabled us to derive ensemble fitted RESP charges from the solvated charge distributions of multiple trajectories. We observed broad, single peaked charge distributions for the conjugated ring atoms with well-defined mean values. The charge fitting procedure was validated against published charges of the dyelike amino acid tryptophan, which showed good agreement with existing tryptophan parameters from the AMBER, CHARMM, and OPLS force field families. A principal component analysis of the charge fluctuations revealed that a small number of collective coordinates suffices to describe most of the in-plane dye polarizability. The AMBER-DYES force field allows the rapid preparation of all atom molecular dynamics simulations of fluorescent systems for state of the art multi microsecond trajectories.
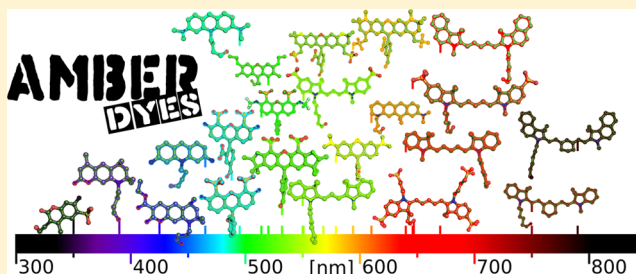
1. INTRODUCTION

Detailed knowledge of protein dynamics is key toward understanding functional mechanisms such as protein folding. A potent single molecule method[1] to probe dynamics on millisecond time scales and below is molecular spectroscopy.[2] Because only very few biomolecules exhibit natural fluorescence in the visible range, labels are required that are typically covalently bound to the biomolecule. Two widely used single molecule fluorescence techniques are Photoinduced Electron Transfer (PET)[3] and Fluorescence Resonance Energy Transfer (FRET).[4]

PET measures the van der Waals contact formation rates of a fluorescent dye with a fluorescence quencher. Upon contact formation, photon excitations of the donor HOMOrbital lead to an electron transfer from the quencher dye. The quencher electron refills the HOMO valence hole and donor fluorescence is lost. Upon dissociation of the dye and the quencher, the dye regains its fluorescence. The PET transfer efficiency decays exponentially with the dye distance and can be observed for dye distances below 1 nm.

FRET is sensitive over a larger range of distances between 2 and 10 nm and has been suggested as a spectroscopic ruler.[5−7] FRET is a nonradiative energy transfer mechanism by which energy is transferred from an excited donor dye to a ground state acceptor dye. The efficiency of this process depends on the distance as well as on the relative orientation factor κ^2 of the two electronic transition dipole vectors. The efficiency can be determined either from the change in lifetime or directly from the ratio of donor to acceptor fluorescence intensity and contains the distance information between the two dyes.[8] However, the distance information cannot be calculated from the efficiency data alone without knowledge of the relative orientation factor.[9] Assuming that the two dyes move isotropically within the donor lifetime, the orientation factor is often approximated by $\kappa_{iso}^2 = {}^2/_3$. Whereas this approximation yields good results for free, noninteracting dyes in solution, protein bound dyes show large deviations from the ideal isotropic motion,[10] thus rendering accurate distance determination challenging.

The shortcomings of both PET and FRET, namely, the on—off character of PET[11] data as well as the unknown orientation factor κ^2 in Förster theory, have recently been addressed by combining the information from experiments and molecular

dynamics simulations.[12–15] It has been shown that by explicitly including the attached dyes in the molecular dynamics simulation, accurate anisotropy values can be obtained,[16,17] which enable to derive distances with markedly enhanced accuracy.

A particular challenge of FRET and PET simulations is the proper choice of force field parameters for the fluorescent labels. Existing parametrizations, among others, include a coarse grained model for Alexa Fluor 488 by Allen et al.,[18] which has been used to investigate the influence of large fluorescent labels on the folding behavior of proteins. A CHARMM27 all atom parameter set for Alexa Fluor 488 and Alexa Fluor 568 has been introduced by Corry et al.[16] who observed only minor differences of 0.05 e between the ground and excited state charge distributions on the aromatic ring atoms. Vaiana et al.[11,19] introduced CHARMM27 parameters for MR121 and a new atom type to describe the orthogonal ring systems of rhodamine 6G, which are also common for many of the Alexa and ATTO dyes. Schröder et al.[17,20] parametrized Cy5 and Alexa Fluor 488 for the GROMOS force field and Lima et. al[13] presented OPLS-AA parameters for Alexa Fluor 594 and Alexa Fluor 488 with approximate excited state charges for the FRET donor based on offsetting the ground state charges by the S1-GS energy gap. Brent Krueger's lab developed parameters for several fluorescent labels for the AMBER force field, such as the DACM donor and eosin,[10] fluorescein and pacific blue,[21] as well as many dyes from the Alexa and Cy dye families.

Compared to other biomolecules, the parametrization of dyes is particularly challenging because of the extended, delocalized, and highly polarizable dye π-electron systems and a lack of experimental data, for example, on their solvation free energies. As a result of the former, large-scale fluctuations of the π-electron density due to changes in the protein and solvent environment are expected. This challenges the fixed point charge approximation used in most major protein force fields, such as AMBER,[22] OPLS-AA,[23] CHARMM,[24] and GRO-MOS.[25] To characterize these fluctuations is therefore essential when assessing the quality of dye extensions to these existing protein force fields.

Charge fluctuations also exist in amino acids, though to a lesser extend, and are implicitly accounted for in protein force fields such as AMBER99sb.[26–28] In AMBER99sb, RESP charges of important amino acid conformers are derived on energy minimized vacuum structures. In this procedure, the overpolarization artifact of the Hartree−Fock method proved to provide a suitable and computationally efficient implicit scaling factor to approximate the polarization of amino acids in mixed protein/solvent environments from vacuum calculations. Dyes, in contrast to amino acids, have a lower conformational flexibility due to the rigidity of the π-electron systems. Therefore, the most relevant conformer is flat with small structural fluctuations due to thermal bending. The large charge fluctuations of this conformer are therefore mostly due to changes in the solvent and cannot immediately be accounted for through scaling factors alone. Here, we therefore have developed a charge fitting method which samples the explicitly solvated dye and derives optimal RESP fits to the ensemble of solvated structures from multiple simulation trajectories, following advances in deriving condensed phase point charges from QM/MM simulation schemes.[29,30]

We used this approach to derive AMBER force field parameters for 22 common fluorescent labels from the Invitrogen Alexa family (Alexa Fluor 350, Alexa Fluor 488,

Alexa Fluor 532, Alexa Fluor 568, Alexa Fluor 594, Alexa Fluor 647), the Atto-Tec Atto family (ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 514, ATTO 520, ATTO 610, ATTO Thio12), and the Lumiprobe Cy family (Cy3, Cy3 (water-soluble), Cy5, Cy5 (water-soluble), Cy5.5, Cy7, Cy7.5).

2. METHODS

Commercially available fluorescent protein labels consist of a dye and multiple linker options, which attach the dye to an amino acid of the protein. To reduce the combinatorial complexity of the many possible dye/linker/amino acid combinations, we have created a modular parameter set which consists of 22 dyes and linkers for cysteine, lysine, and an artificial ketone amino acid.[31] All labels were separated into linkers and dyes between peptide bonds. The open bonds were capped using NME and ACE charge caps during all calculations. Modified parameters for the C- and N-termininally attached dyes are also provided. The parameter set supports a total of 126 fluorescent label combinations, which reflects the commercially available options from the respective manufacturers; see Table 1 for details. Energy minimized molecular

Table 1. Overview of AMBER-DYES Dye and Linker Pairs

commercial dye name	residue name	amino acid + linker
Alexa Fluor 350	A35	CYS C5, LYS
Alexa Fluor 488	A48	CYS C5, LYS, Keto
Alexa Fluor 532	A53	CYS C5, LYS
Alexa Fluor 568	A56	CYS C5, LYS
Alexa Fluor 594	A59	CYS C5, LYS
Alexa Fluor 647	A64	CYS C2, LYS, Keto
ATTO 390	T39	CYS C2, LYS
ATTO 425	T42	CYS C2, LYS
ATTO 465	T46	CYS C2, LYS
ATTO 488	T48	CYS C2, LYS
ATTO 495	T49	CYS C2, LYS
ATTO 514	T51	CYS C2, LYS
ATTO 520	T52	CYS C2, LYS
ATTO 610	T61	CYS C2, LYS
ATTO Thio12	Tth	CYS C2, LYS
Lumiprope Cy3	C3N	CYS C2, LYS
Lumiprope Cy3 (water-soluble)	C3W	LYS
Lumiprope Cy5	C5N	CYS C2, LYS
Lumiprope Cy5 (water-soluble)	C5W	LYS
Lumiprope Cy5.5	C55	CYS C2, LYS
Lumiprope Cy7	C7N	LYS
Lumiprope Cy7.5	C75	LYS

structures and GROMACS4.6[32] force field parameters of all labels are provided in the Supporting Information. The structures, parameters and the git repository are also available on Github (www.github.com/t-/amber-dyes).

2.1. Bonded Parameters. Following the AMBER force field parametrization procedure, we used quantum chemical calculations for the parametrization. The dye and linker structures were optimized in vacuum using the HF/6-31G* level of theory. The optimized structures were used to assign atom types and bonded parameters from the Generalized AMBER Force Field[33] (GAFF). The parameters were assigned by comparing known atom configurations from the GAFF force field[34,35] to the configurations found in the optimized structures. The parameters for each molecule were created

using ANTECHAMBER[34] and subsequently converted into a single GROMACS force field file[36] compatible with the GROMACS preprocessors and virtual interaction sites.

2.2. Nonbonded Parameters. The extended π−electron systems of fluorescent dyes have multiple, equally valid, solutions within the RESP[37] point charge approximation which is used in the AMBER protein force fields. In this study, we simulated an ensemble of Boltzmann weighted RESP point charge samples for each dye and reduced it into a single optimized ensemble RESP charge set. All charges were derived from quantum classical QM/MM simulations which allowed explicit polarization of the dyes at the B3LYP/6-31G*//TIP3P[38] level of theory.

To this end, the following simulation scheme was used (Figure 1) for the dyes. We created a thermal ensemble of the
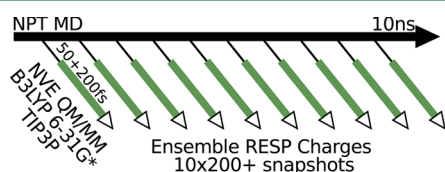

Figure 1. Simulation setup for the calculation of the solvated QM/MM RESP charges. The green arrows indicate independent NVE QM/MM simulations, which were spawned off an initial 10 ns MD NPT trajectory. Each of the forked trajectories was equilibrated to the new potential for 50 fs, followed by 200 fs of productive sampling for the ensemble RESP fit.

solvated dyes in an 10 ns NPT molecular dynamics simulation. The simulation was started from the energy minimized structure with vacuum RESP charges of the NME capped dye. Lennard-Jones parameters were taken from the generalized AMBER force field atom types which are identical with the AMBER99sb values for most of the atoms parametrized in this study. No further optimizations to the Lennard-Jones parameters were made. The dye was solvated in a TIP3P waterbox with 2 nm spacing to the box boundaries in each direction. The productive runs were simulated at 300 K and 1 bar using the v-rescaling thermostat[39] and Berendsen pressure coupling. Both the thermostat and pressure coupling were only active in the NPT run and turned off to generate NVE conditions for the QM/MM simulations. A cutoff of 1.4 nm was used for Coulomb and Lennard-Jones interactions for both the NPT MD and NVE QM/MM B3LYP/6-31G*//TIP3P simulations. No PME was used to make the transition between MD and QM/MM description as smooth as possible, thereby reducing the expensive relaxation time to the new Hamiltonian. The leapfrog algorithm was used with a time step of 1 fs to integrate the equations of motion without bond constraints. From the 10 ns NPT initial simulation, 10 snapshots were extracted as starting configurations for the NVE QM/MM simulations using the positions and finite temperature velocities from the NPT trajectory. The spacing of 1 ns was chosen larger than the sub ps rotation autocorrelation time of the water model to reduce correlation effects. The charges for the linkers were derived using the procedure of the AMBER99sb force field at the HF/6-31G* level of theory without solvated ensemble fits.

All dye simulations were performed using a slightly modified GAUSSIAN09[40,41]/GROMACS 4.5[32,42,43] QM/MM interface that calculated the ESP grid during every time step. Each grid consisted of ten concentric layers with a total number of grid

points between 8×10^4 (Alexa 350) and 1.8×10^5 (Alexa 647). The set of QM/MM simulations was relaxed to the new level of theory for 50 fs, followed by 200 fs of production QM/MM configuration sampling. For all dyes, all snapshots from the productive runs with 1 fs spacing were used to calculate the ESP and RESP charge distributions and ensemble fits. The whole ensemble of frames was analyzed to improve statistics, even thought part of the ensemble included correlated frames. The influence of outliers and correlated frames on the ensemble fits, where shown, was estimated by bootstrapping the confidence intervals. For this estimate, sets of eight trajectories (45 possibilities) were used, omitting data from two trajectories in each ensemble fit. The sample size of eight was chosen to obtain enough configurations to test the stability of the fit while remaining computationally efficient.

The QM/MM ensemble charge fits for the dyes were performed using our Cython[44] implementation of the original two stage RESP algorithm described in refs 37 and 45. The partial charges q_j were calculated with the restraining potential

$$\chi_{\text{rstr}}^2 = a \sum_j^N \left((q_{0_j} - q_j)^2 + b^2 \right)^{1/2} - b \tag{1}$$

applied to all heavy atoms. The variables in the first stage fit were $b = 0.1$, $a = 0.0005$ with a target charge of $q_{0j} = 0.0$. The restraining potential was increased in the second stage fit to $b = 0.1$, $a = 0.001$, and $q_{0j} = 0.0$, but only applied to the aliphatic carbons and methyl group centers; all other charges remained fixed to their stage one values. This procedure was applied to reduce the partial charges of these buried carbons while still keeping the flexibility of the all atom stage one fit as described in the original Kollman papers.[37,46] Lagrange multipliers were used to constrain the total charge Q_{mol} of the molecule

$$\sum_I^N q_i - Q_{\text{mol}} = 0$$

and the total charge Q_{nme} of the NME charge caps

$$\sum_i^N q_i - Q_{\text{nme}} = 0$$

Each snapshot from the productive sampling was used to construct a separate RESP matrix. The optimal RESP fit was then calculated on all realizations of the system simultaneously. In this fit, the local dye symmetry of symmetric groups was automatically determined using a neighbor tree algorithm which calculated a neighbor fingerprint for each atom. From this fingerprint data, the RESP matrices were reduced by removing redundant degrees of freedom from the fit. The resulting nonlinear set of equations was solved using the Numpy[47] Newton Krylov solver.

3. RESULTS AND DISCUSSION

As a first test to quantify how well dye charges derived from explicitly solvated simulation ensembles compare to the existing protein force field charges, we benchmarked the ensemble charges against known force field values and especially the AMBER99sb and AMBER03 force fields. The partial charges in these force fields rely on vacuum HF/6-31G* (AMBER99sb) RESP and continuum electrostatic B3LYP/cc-pVTZ//HF/6-31G** (AMBER03) RESP charges, which are practical approximations for amino acids.
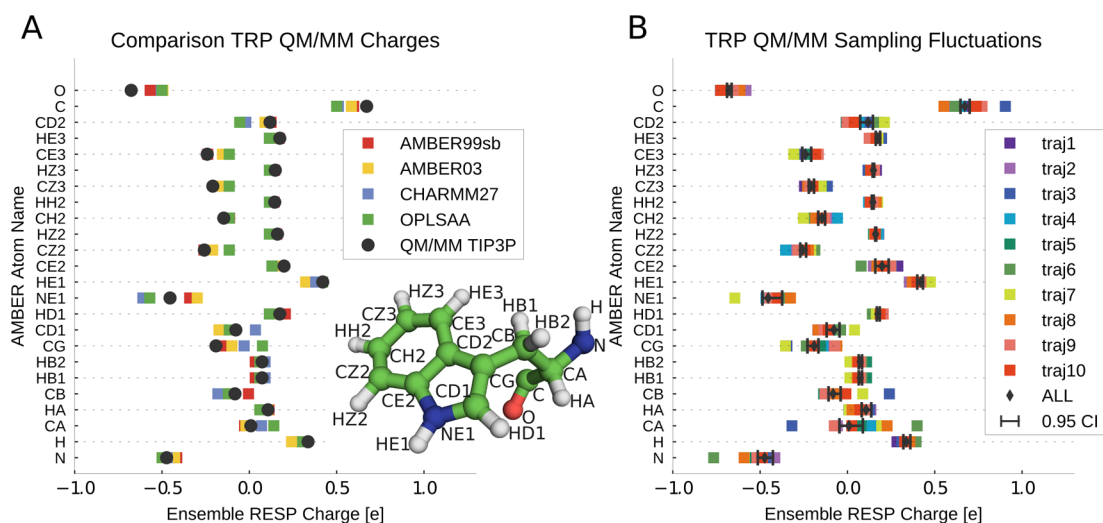
Figure 2. (A) Comparison of the solvated QM/MM ensemble charge scheme to existing tryptophan charges from the AMBER, CHARMM, and OPLS force field families. Filled circles indicate RESP charges derived from the QM/MM ensemble. The colored squares indicate the point charges from the different force fields. (B) The ensemble fit was repeated for the individual trajectories (colored squares). The best fit to the whole ensemble from all trajectories (filled diamonds) is shown with confidence intervals at 0.95, bootstrapped from subsets ($N = 45$) of eight trajectories (error bars).

Due to the physical differences, specifically the large extended π-electron systems, between dyes and amino acids, the artificial overpolarization of the Hartree−Fock method in vacuum cannot immediately be transferred to describe dyes in solution. The overpolarization of Hartree−Fock charges can be improved by using the DFT method in combination with continuum electrostatics at the disadvantage of introducing a homogeneous solvent environment. Here, we parametrized the dyes in explicit solvent to also sample spontaneous polarization effects due to water configurations and thermal motion. This approach is computationally more demanding and convergence needs to be tested. To this end, we compared tryptophan protein force field point charges to point charges from our B3LYP/6-31G*//TIP3P simulation trajectories in explicit solvent. Tryptophan was chosen as a test system for its similarity to the Coumarin, Alexa 350, and Atto 390 fluorescent dyes. We compared these charges to the ones from the AMBER99sb,[26−28] AMBER03,[48] CHARMM27,[24,49,50] and OPLSAA[23] protein force fields (Figure 2A). As can be seen, the ring charges are slightly more polar than the AMBER force field values. The best agreement for the ring carbon and hydrogen atoms was obtained for AMBER99sb. The simulations were repeated with CHARMM27 instead of AMBER Lennard-Jones parameters without major changes in the charge fits, data not shown. Neglecting these small differences in the underlying van der Waals parameters, good agreement was found between the solvated QM/MM charges and the range spanned by the different force fields within the consensus range found among modern protein force fields.[51]

To further compare the charge heterogeneity from the protein force fields to the heterogeneity in the simulation ensembles, we repeated the ensemble fits for each individual tryptophan trajectory (Figure 2B) and estimated the confidence intervals of the ensemble fits as described in the Methods section. The individual trajectories show different extends of charge heterogeneity very similar to the differences across the protein force fields with the exception of the backbone carbonyl group, which is more polar in our fits. The confidence intervals

are largest for the C_α and the amine groups but much narrower than the fluctuations of the individual trajectories.

Our data suggests that there is no single best fixed point charge for a given tryptophan atom but rather a distribution of several equally valid, or invalid representations, depending on the perspective. The microscopic details of these charge fluctuations are likely not the dominant factor for derived macroscopic properties, given the success of fixed point charge force fields at predicting experimental observables. The QM/MM ensemble charges likely still describe derived FRET and PET properties at a level which is comparable to that of other protein force field observables in terms of electrostatic interactions while sampling the dyes π-electron systems in their natural solvated environment.

Next, we characterized the environment induced charge fluctuations for the Alexa 488 dye, which is more representative for the dyes parametrized in this work than tryptophan for its large π-electron system, inherent structural symmetry, and charged solvation groups. We investigated the distribution of the whole ensemble directly for all simulation frames as the distributions do not necessarily have to be single modal or of equal width. To obtain the raw distributions, we performed ESP fits without symmetry constraints or restraining potentials (Figure 3A). The upper plot shows histograms of the observed charge fluctuations; the middle and lower figure show Gaussian fits to the noncarbon and carbon atoms, respectively.

The analysis reveals three important features. First, the point charge distributions are single modal and broadest for the central ring atoms while hydrogen and sulfur oxygens show much narrower charge fluctuation ranges. Second, there is little difference for distributions of symmetric atoms in the absence of symmetry constraints and almost identical mean values. Third, the optimal ESP ensemble fits for symmetrical atoms diverge from each other toward the extremes of the distributions as can be seen for the sulfur and ring noninteracting hydrogen atoms in the central figure.

From the shape of the distributions (Figure 3A) and the trend of the ESP ensemble fits toward the extremes of the distributions, it is tempting to use the mean values of the
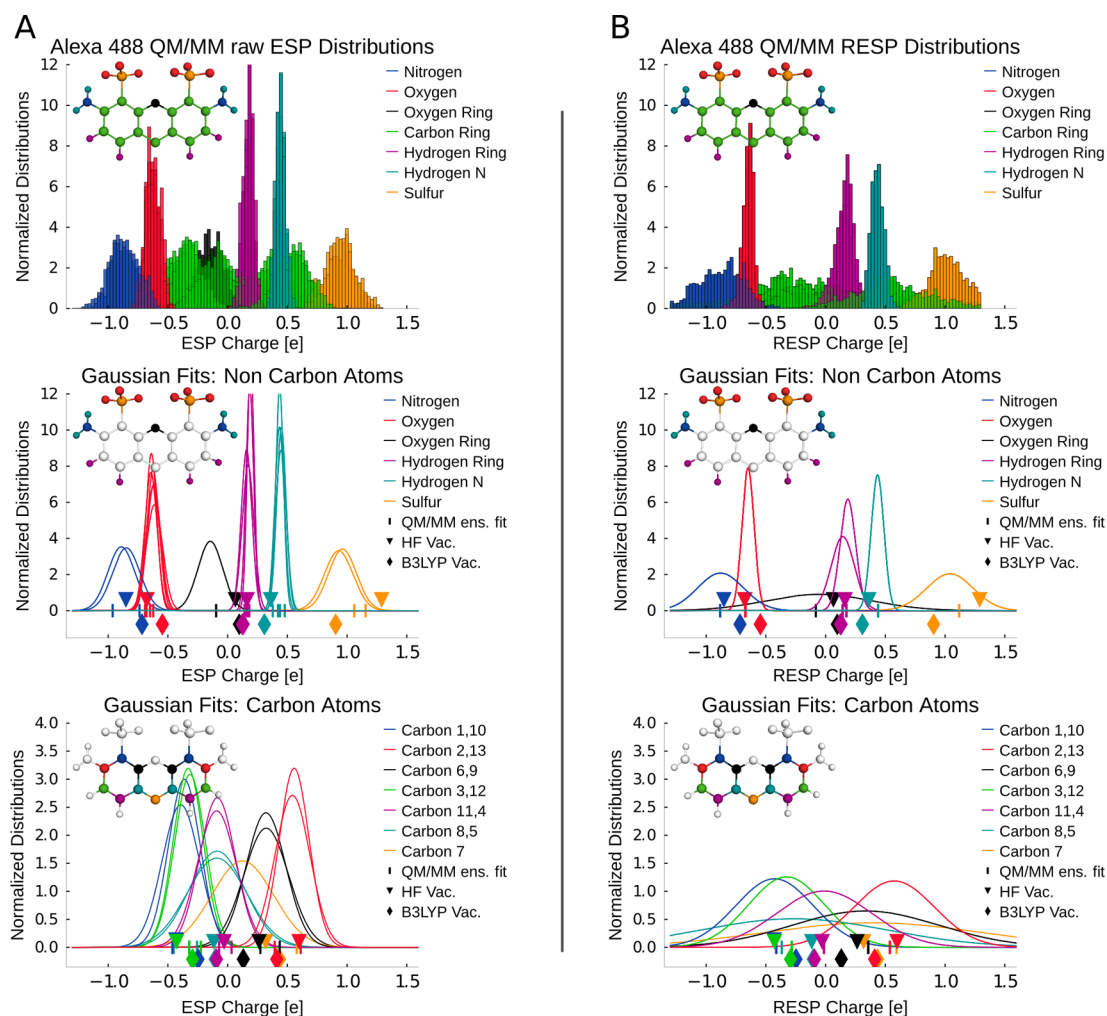
Figure 3. ESP and RESP QM/MM charge distributions for the Alexa 488 dye are presented by atom type and symmetry. Depicted colors in the ball and stick representations match those in the figures. A) The raw ESP distributions are shown (top). Each distribution was fitted by a Gaussian, the non carbon atoms are shown (middle) as well as the ring carbon atoms (bottom). Below each Gaussian fit, the best QM/MM ESP ensemble fit on the data is shown together with vacuum HF and B3LYP RESP fits. B) Same setup as described in A) but for the symmetry constrained two stage RESP fit. Note the change of distribution widths, the fits were performed to the same data set in both cases.

distributions with a small correction for symmetric atoms instead of the ensemble fits. However, this does not address the overpolarization of buried atoms such as methyl and aliphatic carbons that was addressed by Bayly et al.[37,46] and motivated the hyperbolic restraint term in the RESP method. While the magnitude of this term is small and not every dye carries buried carbons, using the RESP symmetry constraints and charge restraints is also closer to the AMBER parametrization protocol, which is why RESP fits were applied here.

RESP fitted charge distributions (Figure 3B) were derived from the same ESP grids as used in Figure 3A. The best RESP fits to the ensemble are shown as colored dashes. The main difference between part A and part B of Figure 3 being the enforced symmetry during the fit and the two stage weak restraining potentials. As can be seen from the figure, we found a vast increase in width of the distributions. Symmetrical atoms entered as a single degree of freedom in the fit and only one fit value to each distribution remains. The charges were compared to B3LYP and HF vacuum fits. The solvated ensemble charges are closer to the HF than the B3LYP values and are, as a whole, more polar than the vacuum B3LYP values. The overall polarity

also increased with the basis set size, and the amount of exact exchange mixed into the functional, data not shown.

A comparison between the ESP and RESP ensemble fits is shown in Figure 4, which uses the same color coding and
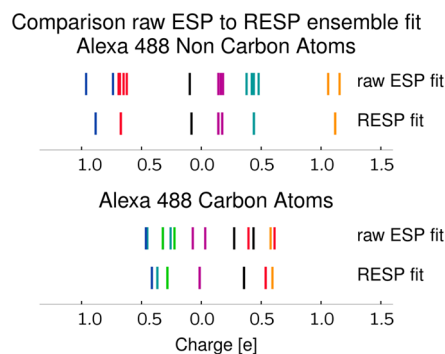

Figure 4. Alexa 488 RESP and ESP ensemble fits separated into carbon and non-carbon atoms. The color coding was adopted from Figure 3 for direct comparison. The RESP fit has fewer data points due to the applied symmetry constraints.

scaling as Figure 3A and B. Even though the symmetry constraint RESP distributions largely differ from their ESP counterparts, the ensemble fits to all trajectory frames are all within range of the ESP values.

The RESP symmetry constraints eliminate all short-lived polarization effects and are therefore only meaningful on the ensemble level with little practical application of the raw RESP distributions. We find that RESP charges can, at least in principle, be closely recovered from averaging the ESP values according to their molecular symmetry after the fit. This, of course, only holds as long as the second stage fit can be neglected as is the case for most fluorescent dyes with few buried carbons.

In a next step, variances and correlations in the observed charge fluctuations were analyzed via principal component analysis of the ESP charge ensemble. As can be seen in Figure 5, the obtained eigenvalue spectrum decays rapidly; therefore, a
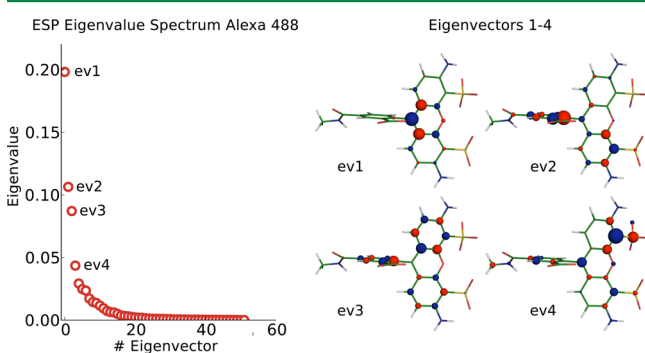

Figure 5. Principal component analysis was performed on the ESP point charge ensemble from the QM/MM trajectories. The eigenvalue spectrum is shown (left). The first four normalized charge fluctuation eigenvectors are shown (right). The size of the spheres indicates the magnitude of the vector components for each atom. A red and blue color coding was applied to indicate the direction of change for each atom along the charge eigenvector.

remarkably large fraction of the charge fluctuations is captured already by the first few eigenvectors. The corresponding eigenvectors describe collective charge fluctuations along the ring system and the sulfate group.

This observation suggests an alternative approach to include the essential contribution of the dyes' electronic polarizability into force field calculations, which would circumvent the notorious problem of conventional polarizable force fields,[52−54] for which a challengingly large number of force field parameters has to be determined. In particular, Drude oscillator based models[55−60] require many parameters to obtain reliable fits to experimental data. Rather than including all atomistic polarizabilities separately, we therefore suggest to only include *collective polarizations* described, in linear response approximation, by the first N eigenvectors v_i of the above charge fluctuation PCA. The respective Drude force constants k_i would be determined from the corresponding matrix of atomic polarizability tensors α by fitting the polarization energy of the PCA Drude particles $U_{pol}^{Drude} = 1/2 \sum_i^N k_i \lambda_i^2$ against the QM polarization energy $U_{pol} = 1/2 \sum_i^N \mu_i \alpha^{-1} \mu_i$ as

$$\min \| U_{pol} - U_{pol}^{Drude} \|_2^2 \qquad (2)$$

Here, $\mu_i = \lambda_i v_i x$ is the induced dipole of PCA oscillator v_i scaled by the extension of the oscillator λ_i and multiplied by the fixed nuclear position vector x.

For the rather inflexible dyes considered here, only a small number of additional polarization degrees of freedom should be required to account for most of the electronic polarization effects in solution. It may even suffice to only consider fluctuations within the molecular plane, as in the PCA analysis above. Implementation and assessment of this approach is, however, beyond the scope of this manuscript.

4. CONCLUSIONS

To enable routine and direct MD simulations of single molecule PET and FRET experiments, we here developed AMBER-DYES, a set of force field parameters for the 22 most commonly used fluorescent labels.

We paid particular attention to the extended, delocalized π-electron systems, which are much more pronounced for fluorescent dyes than for amino acids. A set of ten 200 fs B3LYP/6-31G*//TIP3P QM/MM simulations for each dye was performed, which explicitly described the π-electron polarizability, and to which fixed RESP point charges were fitted. To assess sampling convergence, we calculated the unconstrained raw ESP charge distributions, which showed good convergence of the distributions for symmetrical atoms. To further test the compatibility of this "in solutio" charge procedure to the "in vacuo" AMBER force field, a tryptophan was reparametrized using the same setup and was compared to existing AMBER force field charges. The charges covered the consensus range spanned by existing protein force fields and were closest to the AMBER99sb force field, thus supporting the compatibility of AMBER-DYES with that widely used force field.

The bonded parameters for the fluorescent labels were created using the parameter set derived for druglike ligands by Wang et al.[33] In contrast to most other large biomolecules, most dyes exhibit only low conformational flexibility. We exploited this property in the parametrization as the extended π-electron systems restrict the dyes into a planar conformation. We note that the situation is slightly different for cyanine based dyes such as Alexa647 or Cy3−7, which can undergo marked cis−trans transitions between fluorescent and non fluorescent states on multi-microsecond time scales. In the present study, only fluorescent all-trans states were therefore considered.

As a consequence of using the GAFF force field, inaccuracies may arise in the fine structure of the bonded parameters such as single bond, angle, or dihedral parameters. For improved accuracy, these can be corrected by fitting the parameters against the QM Hessian, vibrational eigenvectors, and eigenvalues. However, we assume these effects to have a small contribution on derived PET and FRET photon distributions due to the overall rigidity of the molecules.

A further possible concern are the accuracy of thermodynamic properties such as solvation free energies, which are sensitive to protein−dye and solvent−dye interactions. Specifically the former are difficult to assess due to their very heterogeneous nature and are not considered in our quantum mechanical calculations. Although additional validation against experimental data would be highly desirable, we are not aware of suitable thermodynamic data for the dyes considered here. Until such data is available, applications of our force field, therefore, have to rest on the assumption that the solvation free energy of the dyes is dominated by the water shell we have used

in our QM/MM calculations. Due to the polar and charged nature of the dyes, such an assumption will be valid in many cases.

A possible future improvement of AMBER-DYES parameters are explicit polarizable centers, which would represent the charge distributions instead of just the static ensemble values. From our characterization of the charge fluctuations, we conclude that the majority of the charge variance can be described by an unexpectedly small number of collective charge variables. This result suggests that a small number of polarizable centers might already suffice to capture most of the charge variance along the ring systems with only a small computational overhead, thus alleviating the notorious many-parameter problem of polarizable force fields for dyes.

The AMBER-DYES force field relies on the approximation that the dye dynamics do not dramatically differ in the electronic ground and excited states, respectively. Such differences might become relevant for donor dyes, which predominantly exist in the excited state during a FRET or PET experiment. Force fields to describe excited states in MD simulations can, in principle, be parametrized using the same approach as presented here for the ground state charges. However, excited electronic states are much more challenging to calculate sufficiently accurately.

The AMBER-DYES force field can readily be used to simulate a wide variety of fluorescently labeled proteins and fluorescence experiments. The included dyes can also serve as a guideline as to which fluorescent dyes serve best for combined experimental and computational approaches.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Force field parameters and molecular structures for all dyes presented in this study are provided in the GROMACS 4.6 force field format. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: hgrubmu@gwdg.de. Phone: +49 (0)551 201 2300. Fax: +123 (0)123 201 2302.

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Schuler, B.; Hofmann, H. Single-molecule spectroscopy of protein folding dynamics: Expanding scope and timescales. *Curr. Opin. Struct. Biol.* **2013**, *23*, 36−47.

(2) Schuler, B.; Eaton, W. A. Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol.* **2008**, *18*, 16−26.

(3) Lindhoud, S.; Westphal, A. H.; Visser, A. J.; Borst, J. W.; van Mierlo, C. P. Fluorescence of Alexa Fluor dye tracks protein folding. *PloS One* **2012**, *7*, e46838.

(4) Preus, S.; Wilhelmsson, L. M. Advances in quantitative FRET-based methods for studying nucleic acids. *ChemBioChem.* **2012**, *13*, 1990−2001.

(5) Stryer, L.; Haugland, R. P. Energy transfer: A spectroscopic ruler. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *58*, 719.

(6) Schuler, B.; Lipman, E. A.; Steinbach, P. J.; Kumke, M.; Eaton, W. A. Polyproline and the spectroscopic ruler: Revisited with single-molecule fluorescence. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2754−2759.

(7) Best, R. B.; Merchant, K. A.; Gopich, I. V.; Schuler, B.; Bax, A.; Eaton, W. A. Effect of flexibility and cis residues in single-molecule FRET studies of polyproline. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18964−18969.

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

(9) 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.

(10) 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.

(11) Vaiana, A. C.; Neuweiler, H.; Schulz, A.; Wolfrum, J.; Sauer, M.; Smith, J. C. Fluorescence quenching of dyes by tryptophan: Interactions at atomic detail from combination of experiment and computer simulation. *J. Am. Chem. Soc.* **2003**, *125*, 14564−14572.

(12) Hoefling, M.; Grubmüller, H. In silico FRET from simulated dye dynamics. *Comput. Phys. Commun.* **2013**, *184*, 841−852.

(13) Hoefling, M.; Lima, N.; Haenni, D.; Seidel, C. A. M.; Schuler, B.; Grubmüller, H. Structural heterogeneity and quantitative FRET efficiency distributions of polyprolines through a hybrid atomistic simulation and Monte Carlo approach. *PLoS One* **2011**, *6*, e19791.

(14) Woźniak, A. K.; Schröder, G. F.; Grubmüller, H.; Seidel, C. A.; Oesterhelt, F. Single-molecule FRET measures bends and kinks in DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 18337−18342.

(15) Klose, D.; Klare, J. P.; Grohmann, D.; Kay, C. W.; Werner, F.; Steinhoff, H.-J. Simulation vs reality: A comparison of in silico distance predictions with DEER and FRET measurements. *PloS One* **2012**, *7*, e39492.

(16) Corry, B.; Jayatilaka, D. Simulation of structure, orientation, and energy transfer between AlexaFluor molecules attached to MscL. *Biophys. J.* **2008**, *95*, 2711−2721.

(17) Schröder, G. F.; Alexiev, U.; Grubmüller, H. Simulation of fluorescence anisotropy experiments: Probing protein dynamics. *Biophys. J.* **2005**, *89*, 3757−3770.

(18) 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.

(19) 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.

(20) Woźniak, A. K.; Nottrott, S.; Kühn-Hölsken, E.; Schröder, G. F.; Grubmüller, H.; Lührmann, R.; Am Seidel, C.; Oesterhelt, F. Detecting protein-induced folding of the U4 snRNA kink-turn by single-molecule multiparameter FRET measurements. *RNA* **2005**, *11*, 1545−1554.

(21) Speelman, A. L.; Munoz-Losa, A.; Hinkle, K. L.; VanBeek, D. B.; Mennucci, B.; Krueger, B. P. Using molecular dynamics and quantum mechanics calculations to model fluorescence observables. *J. Phys. Chem. A* **2011**, *115*, 3997−4008.

(22) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(23) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

(24) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R.; Evanseck, J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S. a. All-

atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(25) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843−856.

(26) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple AMBER force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712−725.

(27) Sorin, E. J.; Pande, V. S. Exploring the helix−coil transition via all-atom equilibrium ensemble simulations. *Biophys. J.* **2005**, *88*, 2472−2493.

(28) DePaul, A. J.; Thompson, E. J.; Patel, S. S.; Haldeman, K.; Sorin, E. J. Equilibrium conformational dynamics in an RNA tetraloop from massively parallel molecular dynamics. *Nucleic Acids Res.* **2010**, *38*, 4856−4867.

(29) Gao, J.; Luque, F.; Orozco, M. Induced dipole moment and atomic charges based on average electrostatic potentials in aqueous solution. *J. Chem. Phys.* **1993**, *98*, 2975−2982.

(30) Cerutti, D. S.; Rice, J. E.; Swope, W. C.; Case, D. A. Derivation of fixed partial charges for amino acids accommodating a specific water model and implicit polarization. *J. Phys. Chem. B* **2013**, *117*, 2328−2338.

(31) Brustad, E. M.; Lemke, E. A.; Schultz, P. G.; Deniz, A. A. A general and efficient method for the site-specifﬁc dual-labeling of proteins for single molecule fluorescence resonance energy transfer. *J. Am. Chem. Soc.* **2008**, *130*, 17664−17665.

(32) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.

(33) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.

(34) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Antechamber: An accessory software package for molecular mechanical calculations. *J. Am. Chem. Soc.* **2001**, *222*, U403−U403.

(35) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247−260.

(36) da Silva, A. W. S.; Vranken, W. F. ACPYPE-Antechamber python parser interface. *BMC Res. Notes* **2012**, *5*, 367.

(37) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(38) 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.

(39) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(40) 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, .; Foresman, J. B; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision D.01; Gaussian, Inc.: Wallingford, CT, 2009.

(41) Hertwig, R. H.; Koch, W. On the parameterization of the local correlation functional. What is Becke-3-LYP? *Chem. Phys. Lett.* **1997**, *268*, 345−351.

(42) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.

(43) Pronk, S.; Páall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845−854.

(44) Behnel, S.; Bradshaw, R.; Citro, C.; Dalcin, L.; Seljebotn, D.; Smith, K. Cython: The best of both worlds. *Comput. Sci. Eng.* **2011**, *13*, 31−39.

(45) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **1990**, *11*, 431−439.

(46) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollmann, P. A. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.* **1993**, *115*, 9620−9631.

(47) Oliphant, T. E. Python for scientific computing. *Comput. Sci. Eng.* **2007**, *9*, 10−20.

(48) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999−2012.

(49) MacKerell, A. D.; Banavali, N.; Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* **2000**, *56*, 257−265.

(50) Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. Implementation of the CHARMM force field in GROMACS: Analysis of protein stability effects from correction maps, virtual interaction sites, and water models. *J. Chem. Theory Comput.* **2010**, *6*, 459−466.

(51) Ponder, J. W.; Case, D. A. Force fields for protein simulations. *Adv. Protein Chem.* **2003**, *66*, 27−86.

(52) Gao, J.; Pavelites, J. J.; Habibollazadeh, D. Simulation of liquid amides using a polarizable intermolecular potential function. *J. Phys. Chem.* **1996**, *100*, 2689−2697.

(53) Gao, J.; Habibollazadeh, D.; Shao, L. A polarizable intermolecular potential function for simulation of liquid alcohols. *J. Phys. Chem.* **1995**, *99*, 16460−16467.

(54) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable atomic multipole-based AMOEBA force field for proteins. *J. Chem. Theory and Comp.* **2013**, *9*, 4046−4063.

(55) Schropp, B.; Tavan, P. The polarizability of point-polarizable water models: Density functional theory/molecular mechanics results. *J. Phys. Chem. B* **2008**, *112*, 6233−6240.

(56) Geerke, D. P.; Thiel, S.; Thiel, W.; van Gunsteren, W. F. Combined QM/MM molecular dynamics study on a condensed-phase SN2 reaction at nitrogen: The effect of explicitly including solvent polarization. *J. Chem. Theory Comp.* **2007**, *3*, 1499−1509.

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

(58) Lamoureux, G.; MacKerell, A. D., Jr.; Roux, B. A simple polarizable model of water based on classical Drude oscillators. *J. Chem. Phys.* **2003**, *119*, 5185−5197.

(59) Vosmeer, C. R.; Rustenburg, A. S.; Rice, J. E.; Horn, H. W.; Swope, W. C.; Geerke, D. P. QM/MM-based fitting of atomic polarizabilities for use in condensedphase biomolecular simulation. *J. Chem. Theory Comput.* **2012**, *8*, 3839−3853.

(60) Lopes, P. E.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D., Jr. Polarizable force field for peptides and proteins based on the classical Drude oscillator. *J. Chem. Theory Comput.* **2013**, *9*, 5430−5449.