Supporting Information:

CHARMM-DYES: Parameterization of fluorescent dyes for use with the CHARMM forcefield

Robert A. Shaw,[†,¶] Tristan Johnston-Wood,[†,‡] Benjamin Ambrose,[†] Timothy D. Craggs,[†] and J. Grant Hill[*,†]

†*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.*

¶*Present address: ARC Centre of Excellence in Exciton Science, School of Science, RMIT University, Melbourne, VIC 3000, Australia*

E-mail: grant.hill@sheffield.ac.uk

Contents

1 Force field files

The residues and parameters that make up the CHARMM-DYES force field can be downloaded in machine readable format from `https://dx.doi.org/10.15131/shef.data.12593681`. These are in GROMACS format, and structured as follows. The GROMACS-compatible force field is in the directory charmm36-dyes.ff, which should be copied either to your working directory or directly into your GMXLIB folder before use. Within this directory you can find:

- new atom types in the atomtypes.atp file;

- new residues in merged.rtp, for each dye by itself, with a linker, and attached to thymine;

- new bonded parameters in ffdyesbonded.itp;

- new-by-analogy non-bonded parameters in ffdyesnonbonded.itp.

In the parent directory, there is also a dyes.rtp file containing just the dye residues, for reference, and a residuetypes.dat file which needs to be added to the global GMXDATA/top directory. In addition, we have provided a pdbs folder containing correctly-labelled PDB files for each of the dyes by themselves, and with the linker attached.

If you wish to create new residues, perhaps by attaching the dyes to different nucleobases or amino acids, you will need to update each of the above files. Guidance on how to do this can be found in the GROMACS documentation.

2   Atom types

Below is a table of the new atom types, and figures showing how these are assigned to each dye, and to the linker.

Table S1: List of new atom types, and their corresponding numeric label for the diagrams below and in the main text.

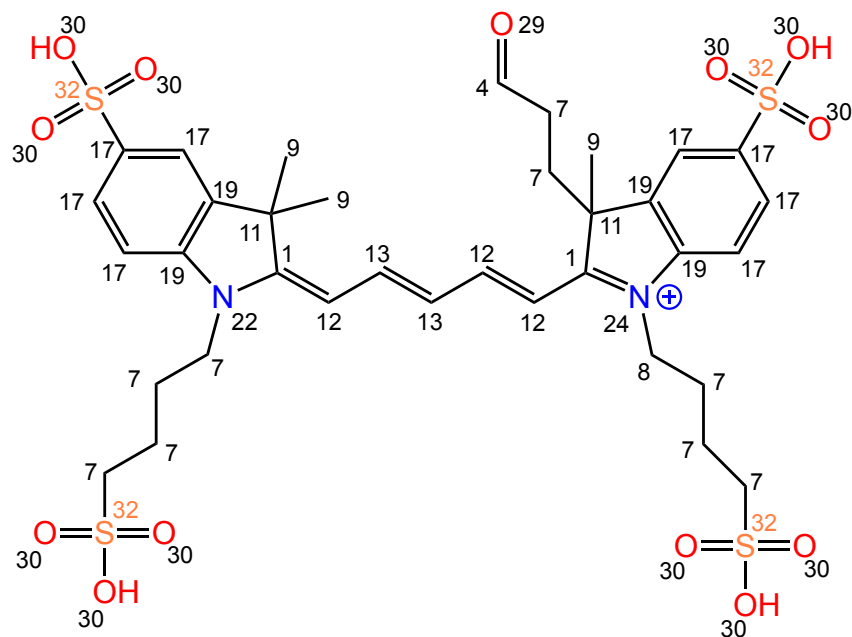| Atom Type	Index
C2510	1
C2520	2
CA201	3
CA204	4
CA301	5
CA311	6
CA321	7
CA324	8
CA331	9
CA334	10
CC350	11
CDC21	12
CDC22	13
CO22	14
CR251	15
CR252	16
CR261	17
CR267	18
CRC20	19
N301	20
N311	21
NC351	22
NP21	23
NR252	24
NS20	25
NS21	26
O311	27
OC361	28
OD21	29
OP21	30
OR360	31
SO31	32
HD1	–
HD2	–
HD3	–
HD4	–
HP1	–
HP2	–
HPAM1	–
HR52	–
HR61	–

Figure S1: Alexa Fluor 647 structure and atom types. We note that the carbon chain between the chromophores and the SO$_3$H groups can vary in length between different dye manufacturers. If the length of the chain requires adjustment, we suggest the charges can be modified using a 'delete and average' approach.
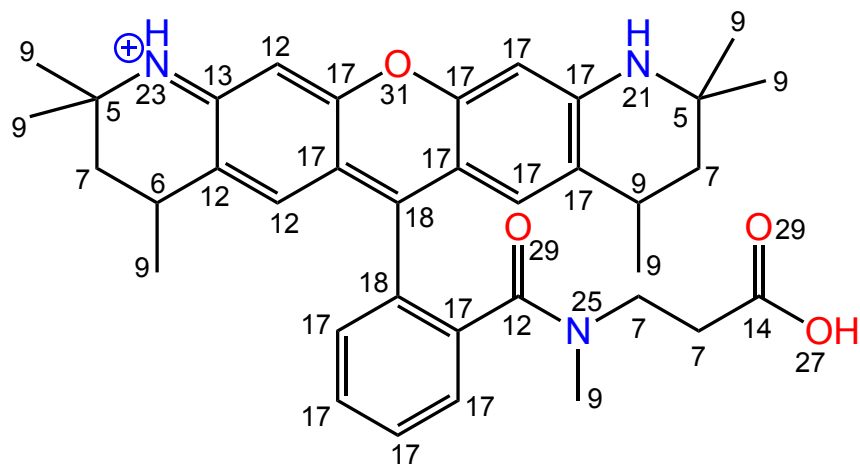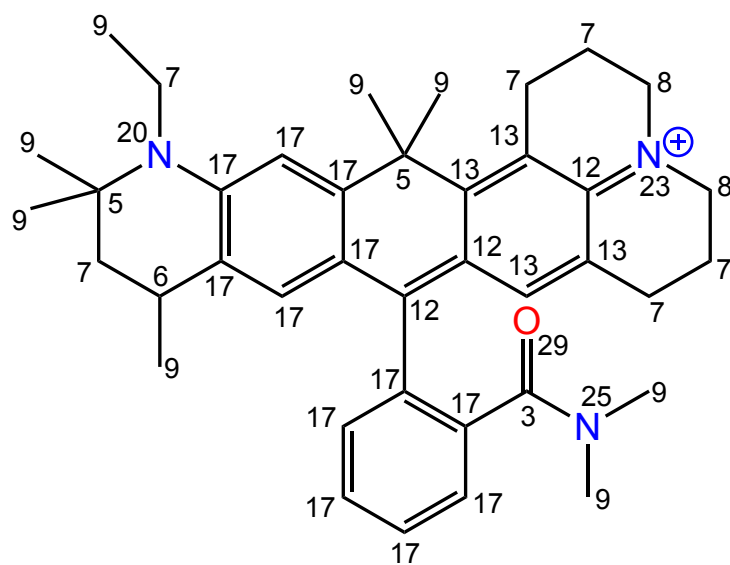
Figure S2: Atto 550 structure and atom types

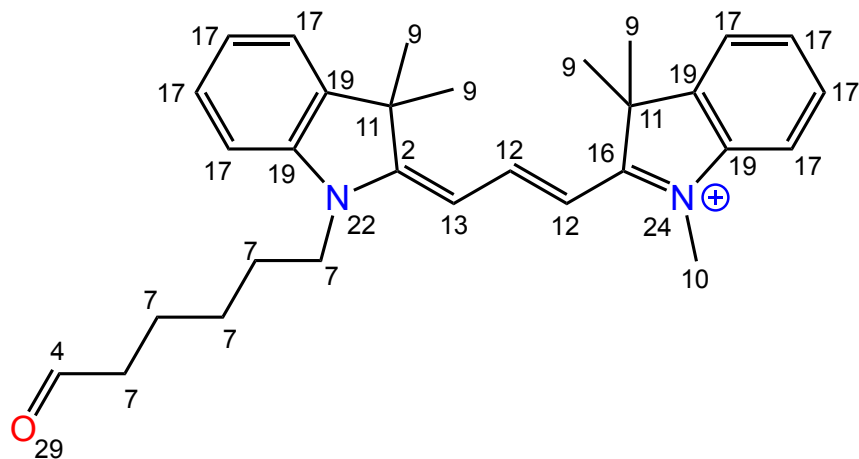Figure S3: Atto 647N structure and atom types

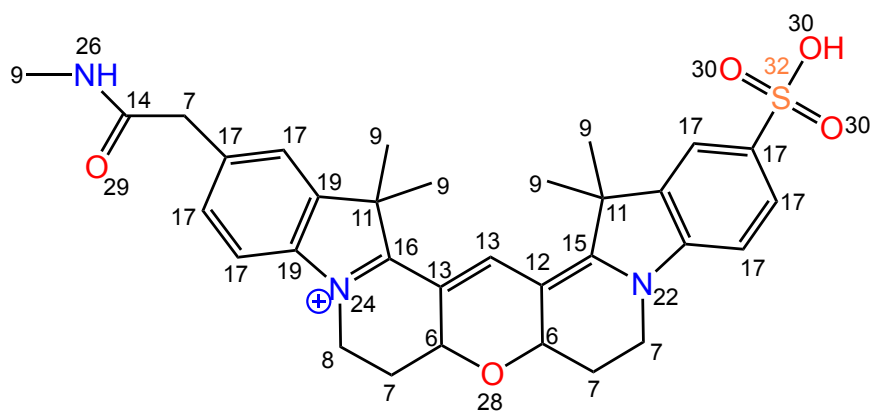Figure S4: Cy3 structure and atom types

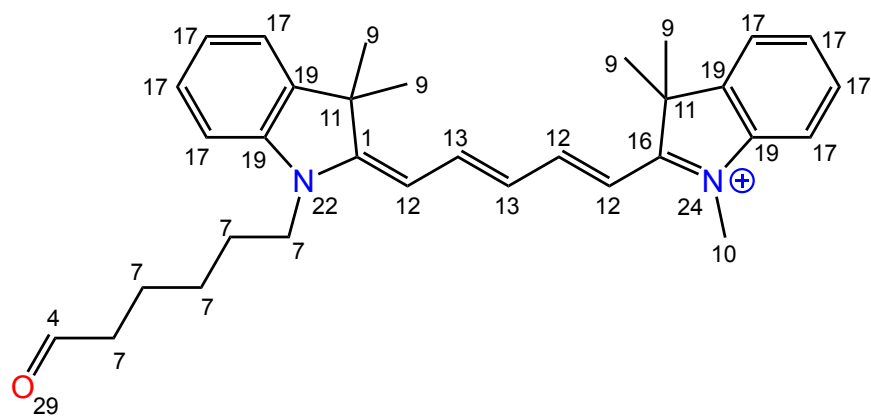Figure S5: Cy3B structure and atom types

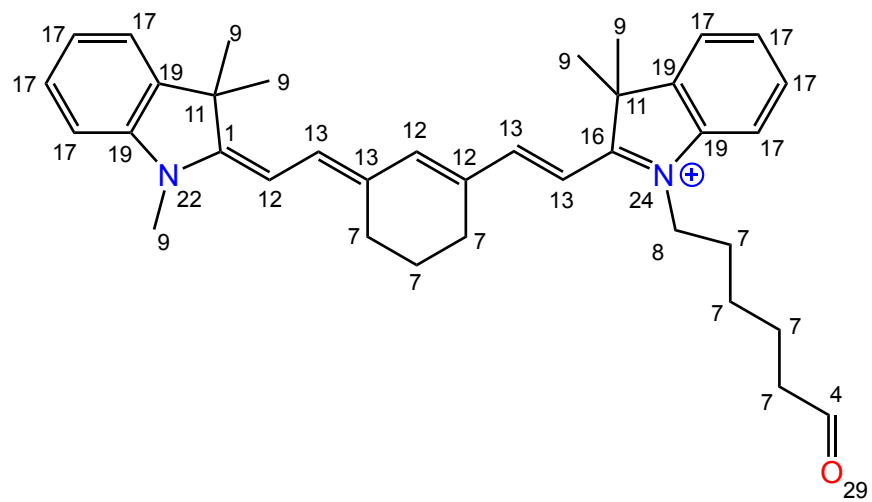
Figure S6: Cy5 structure and atom types
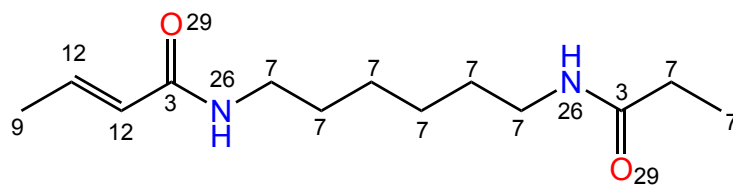
Figure S7: Cy7 structure and atom types

Figure S8: Linker structure and atom types

3   Simulation details

These are the basic simulation details for each dye, used to calculate diffusion constants and RMSD values. Boxes were selected to give 1 nm of space on all sides around the centred dye molecule.

Table S2: Cubic box length and number of water molecules for each dye simulation

Dye	Box length (nm)	No. of waters
Alexa647	4.15658	2293
ATTO 550	3.62950	1478
ATTO 647N	3.56519	1413
Cy3	3.64959	1518
Cy3B	3.93399	1853
Cy5	3.87775	1699
Cy7	4.08680	2214

4 Constrained and unconstrained hydrogen charges

To examine the effect of the standard CHARMM approach of constraining the aliphatic and aromatic hydrogen charges to be $+0.09$ and $+0.15\,e$, respectively, an unconstrained fit of these charges was carried out for the Atto 647N and Cy3B dyes. The deviation between the optimized and constrained charges are displayed as a histogram in Figure S9, where the mean absolute deviation is $0.030\,e$.
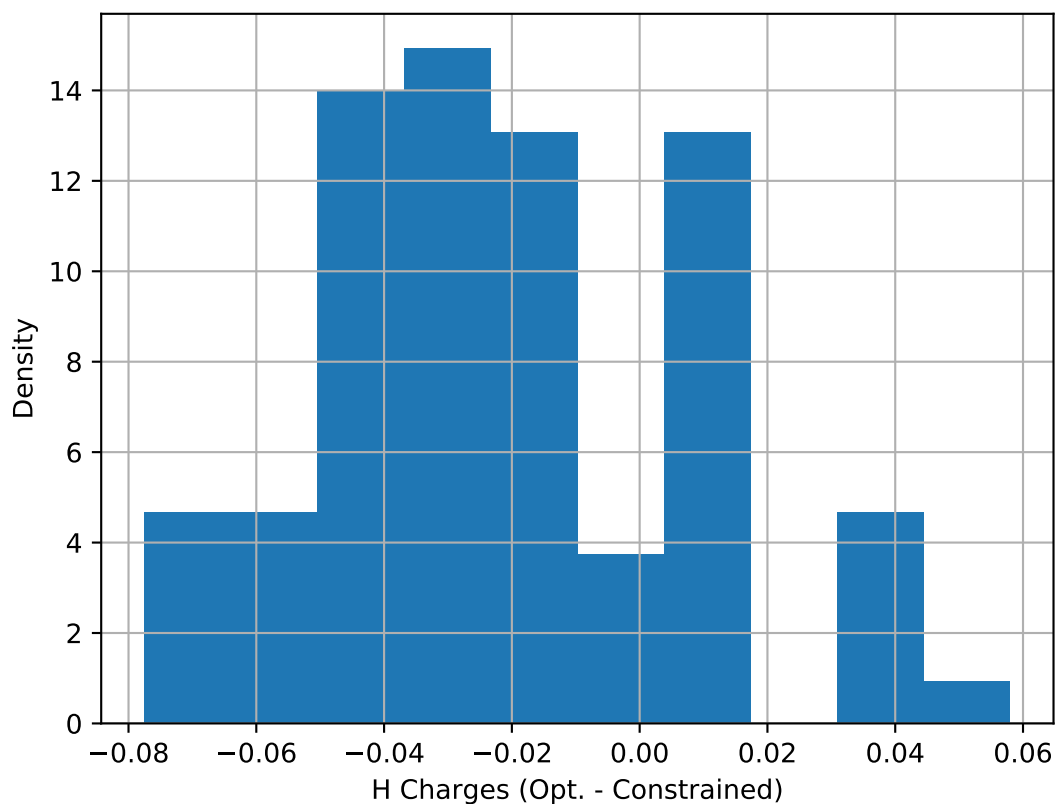

Figure S9: Deviation of the Atto 647N and Cy3B hydrogen atom charges when optimized values are compared to the constrained CHARMM charges of $+0.09$ and $+0.15e$ for aliphatic and aromatic hydrogens, respectively.

5 IR spectra

In Section 3.1 of the main manuscript the performance of the optimized parameters are validated in terms of interaction energies and electric dipole moments. Properties calculated using the CHARMM forcefield with the new, optimized charges were substantially improved when compared to those using the CGenFF initial guess, using the output of QM calculations as a benchmark. A second, less clear, validation of the charges can be found through computation of the dyes' infrared spectra. We chose two dyes (Atto 647N and Cy3B) to determine IR spectra for, and compare to the results from the MD simulations. FTIR measurements were taken using a PerkinElmer Spectrum Two with a UATR module. 100 nanomoles of dye were dissolved in ethanol and transferred to the ATR crystal. The ethanol was allowed to evaporate before the pressure arm was applied, and scans were taken. Spectra shown are an average of 4 scans from 4000 to 600 cm^{-1}, with a subtracted background acquired from a repeat procedure with ethanol alone.

Figure S10 shows the resulting spectra, with overlays of both the experimental and computed spectra. Key peaks around the 3000 and 1000-2000 wavenumber regions appear to be qualitatively reproduced by computation. It should be noted that the computed intensities, which rely strongly on the magnitude of dipole moments, are known to be unreliable. The dyes used in the experimental IR spectra contain N-hydroxysuccinimide (NHS) groups that are required for the reaction to conjugate them to biomolecules. There are additional peaks in the experimental spectra due to these groups; they can be clearly seen in both dyes, particularly in the 2000-3000 wavenumber region. The reactive NHS groups leave during the reaction with biomolecules, hence they are not included in the forcefield parameters or any of the simulations reported.
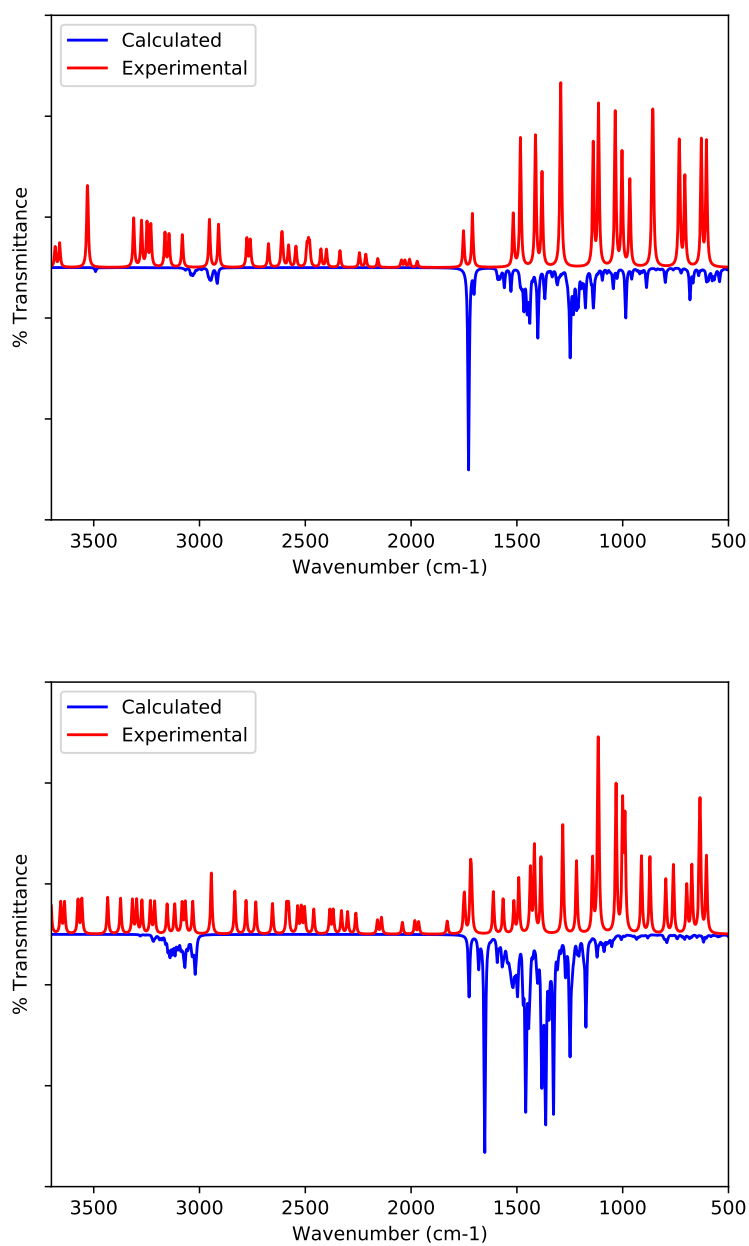
Figure S10: Infrared spectra for Cy3B (top) and Atto 647N (bottom), as calculated from MD simulations using the optimized parameters (blue lines), compared to experimental spectra (red lines). The dyes used in the experiments differed from those in the computation by the addition of reactive NHS groups.