

# Supplementary Information

## FRET guided modeling of nucleic acids

Fabio D. Steffen<sup>1</sup>, Richard A. Cunha<sup>1</sup>, Roland K.O. Sigel<sup>1</sup>, Richard Börner<sup>1,2\*</sup>

<sup>1</sup> Department of Chemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

<sup>2</sup> Laserinstitut Mittweida, Technikumplatz 17, Mittweida, Germany

\* author to whom correspondence should be addressed

[richard.boerner@hs-mittweida.de](mailto:richard.boerner@hs-mittweida.de)

## Supplementary Methods

### *In vitro* nucleic acid preparation and dye labeling

DNA oligonucleotides were purchased with C5-amino-modified deoxythymidines at three donor (T19, T23, T31) and one acceptor (T31) label site. The dyes were coupled in-house to the DNA strand using N-hydroxysuccinimide chemistry. In brief, 5  $\mu$ L of concentrated DNA (1 mM) were mixed with 3  $\mu$ L of dye (few crystals of NHS-dye ester dissolved in 12  $\mu$ L DMSO, >10x excess of dye) in 2  $\mu$ L of potassium phosphate buffer (500mM, pH 8.7) and stirred at 37 °C and 400 rpm for 3 h (1). The DNA was precipitated overnight at -20°C in 50  $\mu$ L (2.5 eq) of ice-cold EtOH supplemented with 10  $\mu$ L KCl (1M). The pellet was washed 3x with 70% EtOH, resuspended in water. Excess salts and phosphate buffer were removed by filtering over a VivaSpin (5000 MWCO) and purity was checked by analytical HPLC.

### Ensemble time-correlated single photon counting (eTCSPC)

Fluorescence lifetime and anisotropy decays were recorded on a Fluorolog FL3-222 (Horiba) spectrophotometer equipped with a double-grating emission monochromator and a PPD-900 photon counting detector. Samples were excited at 504 nm (DD-510L, Horiba) and 633 nm (DD-635L, Horiba) with a repetition rate of 10 MHz. Decays were fitted to a series of exponentials by iterative reconvolution with the instrument response function (IRF) using a custom written Python package named *LifeFit* (<https://github.com/RNA-FRETools/Lifefit>). The dynamic anisotropy was calculated from the polarization resolved decays

$$r(t) = \frac{I_{\parallel}(t) - G I_{\perp}(t)}{I_{\parallel}(t) + 2G I_{\perp}(t)} \quad (1)$$

and analyzed with a local-global rotation model (2, 3)

$$r(t) = [(r_0 - r_{\infty}) e^{-t/\tau_{r,loc}} + r_{\infty}] e^{-t/\tau_{r,glob}}. \quad (2)$$

### Single-molecule confocal experiments

DNA double strands were imaged free in solution in the presence of 100 mM KCl and 20 mM MgCl<sub>2</sub>, buffered in 50 mM Tris-HCl at pH 7.5. Single-molecule measurements were performed on an upright MicroTime 100 confocal microscope (Picoquant) with pulsed interleaved excitation (PIE) (4) at 518 nm and 640 nm (LDH-D-C-640 and LDH-D-C-520, Picoquant) and a repetition rate of 40 MHz. The instrument was equipped with a UPlanSApo, 60x, NA=1.20 water-immersion objective (Olympus) and a TimeHarp 260 TCSPC card (Picoquant) (5). Photons were spectrally separated (Semrock FF01-582/64 and Chroma H690/70m filters), directed through fiber optics (pinhole diameter: 50  $\mu$ m) onto two SPCM-AQRH avalanche photodiodes (Excelitas). The data was analyzed with the Matlab software package *PAM* (6). A sliding time-window with two-color, all photon burst search (APBS) identified bursts with a total

count of at least 100 photons ( $N_D + N_A$ ) (7). Double labeled molecules were selected by applying an intensity threshold of red photons after acceptor excitation ( $F_{A|A} > 50$ ) to remove the donor-only population and a stoichiometry limit  $S > 0.2$  to eliminate any acceptor-only species (8). The three DNA species were measured individually and computationally mixed *post hoc* in order to determine and apply correction factors. In particular the  $\gamma$ -factor requires more than one species to reliably fit a regression line. The FRET histograms were corrected for donor leakage into the acceptor channel ( $\alpha = 0.1$  of donor signal), direct excitation of the acceptor ( $\delta = 0.02$  of acceptor signal) as well as differences in detection and quantum efficiency of the two dyes ( $\gamma = 0.7$ ) by non-linear regression in the 2D FRET-stoichiometry ( $E-S$ ) histogram (9, 10). Mean FRET values and standard deviations were determined by a Gaussian mixture model with seed values initialized near the expected single species.

Single-molecule TIRF experiments of the  $Mn^{2+}$  riboswitch were carried out and published previously by the lab of Nils Walter (11). The RNA was measured in 50 mM HEPES (pH 7.2), 100 mM KCl and 1 mM  $Mg^{2+}$  using prism-based total internal reflection fluorescence (TIRF) microscopy. Selected single-molecule traces were recorded at 16 Hz and corrected for background and spectral cross-talk. FRET histograms were fitted with a two-component Gaussian mixture model.

### **Derivation of fluorophore and linker parameters for *in silico* labeling with *FRETlabel***

FRET labels were composed of three parts: (i) a fluorescent moiety with a delocalized  $\pi$ -electron system, (ii) a flexible linker and (iii) a functional group for conjugation with the biomolecule of interest. We constructed and parameterized suitable adapters for the most common nucleic acid labeling strategies (12). Bonded and non-bonded parameters for the fluorophores were taken from the AMBER-DYES package (13, 14). Parameters for the linker atoms were derived by analogy to similar terms in the AMBER force field using the Antechamber Python Parser interface (Acpype) (15). Linker geometries were optimized with GAUSSIAN 09 (16) at the B3LYP/6-31G\* level of theory and followed by electrostatic potential (ESP) calculations with Hartree-Fock. Partial charges were calculated by two-stage restrained electrostatic potential (RESP) (17) fitting, as implemented in Antechamber (18). For this purpose, appropriate capping groups and group constraints were introduced, mimicking the partial charge distribution in the linker when ligated to the dyes and nucleotides (Supplementary Fig. 2). Step-by-step protocols for linker building, RESP fitting and force field parameterization can be found at [github.com/RNA-FRETtools/fretlabel](https://github.com/RNA-FRETtools/fretlabel). Dyes and linkers were fused to the RNA at the desired position using the PyMOL plugin named *FRETlabel*. Through a graphical user interface (GUI) *FRETlabel* exposes ready-to-use dye-linker adapters for 3'/5'-end labeling as well as internal labeling at deoxythymidines or etheno-bridged adenosines and cytidines (Supplementary Fig. 1).

### **Molecular dynamics simulations**

Molecular dynamics simulations were run on Piz Daint (Swiss National Supercomputing center CSCS, Lugano) with Gromacs 2019.4 and 2020.3 (19) using the AMBER force field (20) with parmbsc0 (21) and  $\chi$ OL3 (22, 23) corrections for RNA and  $\epsilon/\zeta$ OL1 (24),  $\chi$ OL4 (25) and  $\beta$ OL1 (26) corrections (together named OL15) for DNA. Ion parameters for  $K^+$  and  $Cl^-$  are taken from Joung and Cheatham (27). The fluorophore labeled RNA was solvated in a dodecahedral water box with TIP4P-ew (28), charge neutralized with KCl and equilibrated at 298 K and 1 bar with the velocity rescale thermostat and the Berendsen barostat for 10 ns. Bonded interactions were constrained by the LINCS algorithm (29) using an integration time step of 2 fs while non-bonded interactions were treated with the Verlet scheme and a cut-off of 1.4 nm. Long-range electrostatics were computed with the particle mesh Ewald method (30). Periodic boundary correction (PBC) was performed with Plumed 2.6.0 (31) reconstituting and translating molecules to align the center of mass. Only a translational (`TYPE=SIMPLE`) but no roto-

translational fit (TYPE=OPTIMAL) was performed to preserve the global rotational component of the biomolecule for subsequent anisotropy calculations.

```
plumed driver --plumed plumed_pbc.dat --mf_xtc traj.xtc
```

where the configuration file `plumed_pbc.dat` contained

```
nucleic_acid: GROUP NDX_FILE=nucleic_acid.ndx NDX_GROUP=nucleic_acid
WHOLEMOLECULES ENTITY0=dna
FIT_TO_TEMPLATE STRIDE=1 REFERENCE=reference.pdb TYPE=SIMPLE
DUMPATOMS FILE=traj_pbc_translation.xtc ATOMS=nucleic_acid STRIDE=1
```

Distances between the central carbon atoms in the polymethine chain of the donor and acceptor fluorophore were calculated every picosecond. Likewise, the time-dependent orientation factor  $\kappa^2(t)$  was computed for each snapshot using `gmx dyecoupl` (32)

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

where  $\theta_D$  and  $\theta_A$  are angles between the vector connecting the central carbons and the transition dipole of donor or acceptor respectively, and  $\theta_T$  is the angle between the two transition dipoles (2).

The DNA bending angle  $\varphi_{\text{bend}}$  was calculated using Plumed 2.6.0 as the deviation from the helical axis defined by atoms DC1(N4) – DA19(N6) – DC38(N4)

$$\varphi_{\text{bend}} = 180 - \angle_{\text{DC1(N4)-DA19(N6)-DC38(N4)}} \quad (4)$$

### Simulation of photon emission events

Photons were sampled from a Markov chain illustrated in Fig. 1e and characterized by the photon emission rates  $k_{D,f}$  and  $k_{A,f}$ , the internal conversion rates  $k_{D,i}$  and  $k_{A,i}$  and the time-dependent transfer rate  $k_T(t)$ . The transition probabilities between the ground state, the donor-excited state and the acceptor-excited state are given as

$$P = \begin{pmatrix} 0 & 1 & 0 \\ p_{D,f} + p_{D,ic} & 1 - (p_{D,f} + p_{D,ic} + p_T) & p_T \\ p_{A,f} + p_{A,ic} & 0 & 1 - (p_{A,f} + p_{A,ic}) \end{pmatrix} \quad (5)$$

The transfer rate was updated at each time-step of the Monte-Carlo simulation by the inter-dye distance  $R_{DA}(t)$  and the orientation factor  $\kappa^2(t)$  of the underlying MD trajectory

$$k_T(t) = (k_{D,f} + k_{D,ic}) \left( \frac{R_0'(t)}{R_{DA}(t)} \right)^6 \quad (6)$$

$$\text{with} \quad k_{D,f} = \frac{Q_D}{\tau_D}, \quad k_{D,ic} = \frac{1 - Q_D}{\tau_D} \quad (7)$$

$$\text{and} \quad R_0'(t)^6 = \left( \frac{R_0}{2/3} \right)^6 \kappa^2(t) \quad (8)$$

where  $Q_D$  and  $\tau_D$  are the fluorescence quantum yield and lifetime of the donor dye (32, 33).  $R_0$  is the time-independent Förster radius (54 Å, for Cy3-Cy5 assuming isotropic rotation) while  $R_0'(t)$  is time-dependent and is a function of the instantaneous  $\kappa^2(t)$ . For ACV simulations isotropic rotation of the dye was assumed ( $\kappa^2=2/3$ ).

Burst sizes were drawn from an experimental, single-molecule distribution  $P(N_D + N_A)$  or an analytical power law distribution

$$P(x) = x^\lambda \quad \text{with } x \in \mathbb{N} \text{ and } N_{T,\min} < x < N_{T,\max} \quad (9)$$

where  $x$  is the burst size,  $\lambda$  an empirical parameter to approximate the experimental distribution and  $N_{T,\min}$  and  $N_{T,\max}$  lower and upper size cutoffs (32).

For each burst, photons were accumulated until the total photon count  $N_T = N_D + N_A$  had reached the current burst size.

Each burst can be computed from a Markov process involving (i) only one MD trajectory or (ii) all trajectories as described previously (32, 33). The first case reflects a situation where interconversion between conformation is slow compared to the duration of the burst (i.e. each photon burst averages only over fast dye motions within one trajectory). In the second case, the entire trajectory ensemble is represented in one burst (i.e. all conformational dynamics of dyes and biomolecule are averaged in one burst). In either way, a transfer efficiency  $E$  is computed for each burst

$$E = \frac{N_{A,\text{tot}}}{N_{A,\text{tot}} + N_{D,\text{tot}}} \quad (10)$$

where  $N_{A,\text{tot}} = N_A/Q_A$  and  $N_{D,\text{tot}} = N_D/Q_D$ . The resulting histograms can be compared to a fully  $\gamma$ -corrected FRET experiment. Here we calculated FRET distributions for each donor-acceptor DNA strand pair separately to get their respective mean FRET efficiency. Additionally, we reproduced the multiplexed FRET experiment by combining all three 1 ns simulations (Cy-low, Cy-mid and Cy-high) using the “one-trajectory” averaging regime described above.

By calculating photon emission events, we account for shot noise broadening observed in experimental FRET histograms. The variance of the resulting distribution is therefore the sum of the shot-noise contribution and the variance induced by structural dynamics ( $\sigma_{\text{dyn}}$ ) (34)

$$\sigma_{\text{sn+dyn}}^2 = \frac{\langle E \rangle_{\text{exp}} (1 - \langle E \rangle_{\text{exp}})}{N_A + N_D} + \sigma_{\text{dyn}}^2 \quad (11)$$

There are two notable limitations to the burst generation method outline herein. Firstly, donor and acceptor brightness are assumed to be constant and independent of the instantaneous transfer efficiency. Secondly, the photon simulation does not include background noise and therefore should be compared to background corrected FRET experiments.

For dynamic anisotropy decays, the photon polarization probability was calculated as

$$P(p) = \frac{2 \cos^2(\theta)}{\cos^2(\theta) + 1} \quad \text{and} \quad P(s) = 1 - P(p) \quad (12)$$

where  $\theta$  is the angle between the transition dipole at the time points of excitation and emission respectively. Apart from depolarization due to rotational diffusion, the angular displacement  $\beta$  between the absorption and emission dipole as well as photoselection (prefactor: 2/5) also decrease the anisotropy (33). The time-resolved anisotropy  $r(t)$  is thus expressed as the product of three terms

$$r(t) = \frac{2}{5} \left( \frac{3 \cos^2(\beta) - 1}{2} \right) \left( \frac{I_p(t) - I_s(t)}{I_p(t) + 2I_s(t)} \right) \quad (13)$$

where  $I_p$  and  $I_s$  are the intensities of the polarization resolved lifetime decays.

## Riboswitch structure prediction with Rosetta

For Rosetta modeling of the Mn<sup>2+</sup> sensing *yybP-ykoY* riboswitch aptamer from *Xanthomonas oryzae* (11) we used the FARFAR2 protocol (35) which is part of Rosetta 3.12 (<https://www.rosettacommons.org/>) and includes improved base-pair sampling as well as an updated fragment library (36) and scoring function ("res4") (37). Models were calculated with `rna_denovo`, executed on multiple cores in parallel:

```
rna_denovo.linuxgccrelease -nstruct 1000 -fasta mn_riboswitch.fasta
-secstruct_file mn_riboswitch_secstruct.txt -silent mn_riboswitch.out
-minimize_rna true -cycles 20000
```

As inputs only the FASTA sequence of the FRET construct (11) and its secondary structure annotation were provided:

```
>mn_riboswitch A:1-57 B:1-48
auccuugggggaguagccugcuuucucggaaagcgccuguaucaacauacucggcua,uagccguggu
gcaggcaacggcgaaagccgucuggcgagaccagggau

((((((((.....(((.....))))))((((((((.....(((.....)))))))))
))))).((((.....)))..))....))))))
```

Supplementary Fig. 6a shows an alignment of the FRET construct and the crystal structure (PDB entry 6n2v).

The top-500 lowest energy models out of a total of 5000 structures calculated by FARFAR2 were extracted with Rosetta's `extract_pdbs`

```
grep "^SCORE:" mn_riboswitch.out | grep -v description | sort -nk2 |
head -n 500 | awk '{print $NF ": " $2}' | tee top_models.txt

extract_pdbs.linuxgccrelease -in:file:silent mn_riboswitch.out -tags
`cat top_models.txt | cut -f1 -d':'`
```

The extracted PDBs were merged into a single file `mn_riboswitch.pdb`

```
for pdb in `cat top_models.txt | cut -f1 -d':'`; do
  echo MODEL $i >> mn_riboswitch.pdb
  echo TITLE "$pdb" >> mn_riboswitch.pdb
  cat "$pdb.pdb" >> mn_riboswitch.pdb
  echo -e "ENDMDL\n\n" >> mn_riboswitch.pdb
  i=$((i+1))
done
```

Sequence agnostic RMSD was calculated for all top-500 RNA models after superimposition to the crystal structure using PyMOL's `align` function

```
rmsd = []
for i in range(1,501):
  rmsd[i]=cmd.align(f"mn_riboswitch and state {i}", "6n2v_conformer",
  mobile_state=i, cycles=0, transform=0)[0]
```

To calculate metrics relying on exact sequence identity, the sequence of the FRET model was threaded onto the crystal structure (i.e. mutating residues of the crystal structure to match the nucleotides of the FRET construct) using Rosetta's `rna_thread`

```
rna_thread -in:file:fasta alignment.fasta -s 6n2v_conformer.pdb -o
6n2v_with_FRETsequence.pdb
```

where `alignment.fasta` contained the following sequence alignment:

```
>mn_riboswitch A:2-58 B:1-15 B:18-19 B:22-27 B:30-47
uccuuggggagugagccugcuuuc-uuc-ggaaagcgccuguaucaacauacucggcua,
uagccguggugcaggaccgaaaggucuggcgagaccaggga
```

```
>6n2v_conformer1.pdb B:1-99
ggcuuggggagugagccugcuuucggaaacgaaagcgccuguaucaacauacucggcgaaagccguggu
gcaggaccgaaaggucuggcgagaccaggcc
```

The threaded target structure was renumbered with `pdb_resi_renumber` from *rosettascripts* (<https://github.com/RNA-FRETtools/rosettascripts>) to conform to the FRET construct sequence in Supplementary Fig. 6a/b.

```
pdb_resi_renumber -pdb 6n2v_with_FRETsequence.pdb -e 'A:1-97>A:2-
57,B:1-15,B:18-19,B:22-27,B:30-47'
```

RMSD and Interaction Network Fidelity (INF) metrics were calculated over residues A:2-24, A:28-57, B:1-15 and B:30-47 using `rna_tools` (38) akin to quality assessment in RNA puzzles

```
rna_calc_rmsd.py -t 6n2v_with_FRETsequence_renum.pdb --target-selection
A:2-24+A:28-57+B:1-15+B:30-47 --model-selection A:2-24+A:28-57+B:1-
15+B:30-47 top500_models/*.pdb
```

```
rna_calc_inf.py -t 6n2v_with_FRETsequence_renum.pdb --target-
selection A:2-24+A:28-57+B:1-15+B:30-47 --model-selection A:2-
24+A:28-57+B:1-15+B:30-47 top500_models/*.pdb
```

The deformation index was calculated as  $DI = RMSD / INF_{all}$ .

FRET was predicted using donor and acceptor ACVs computed with *FRETraj* (39) as described in the main text.

## Supplementary Tables

**Supplementary Table 1.** Labeling chemistries for nucleic acids and fluorescent dyes currently implemented by *FRETlabel* (12, 40–43).

Position	Chemistry	Base	Dye
Internal	U/dT-C5	U dT	sCy3 sCy5
	Etheno-adduct	A C dA dC	Cy5.5 Cy7 Cy7.5 Alexa350
5'-end	Phosphate	A U G C	Alexa488 Alexa532 Alexa568 Alexa594
3'-end	Phosphate	dA dT dG dC	Alexa647 Atto390 Atto425 Atto465
	Hydrazide	A U G C	Atto488 Atto495 Atto514 Atto520 Atto610

**Supplementary Table 2.** Dye and linker parameters used for ACV simulations.

Parameter	symbol	value
Dye radii	$R_{1,Cy3}$ (nm)	8.0
	$R_{2,Cy3}$ (nm)	3.0
	$R_{3,Cy3}$ (nm)	1.5
	$R_{1,Cy5}$ (nm)	9.5
	$R_{2,Cy5}$ (nm)	3.0
	$R_{3,Cy5}$ (nm)	1.5
Linker length	$L_{linker}$ (nm)	21.0
Linker width	$W_{linker}$ (nm)	3.0
Diameter of contact volume	$d_{CV}$ (nm)	3.0

**Supplementary Table 3.** *FRETraj* parameters used for calculating shot-noise broadened FRET efficiency histograms from MD simulations with explicit all-atom dyes or multi-ACVs.

Category	Parameter	Symbol	Value
Dyes	Fluorescence lifetimes <sup>1</sup>	$\tau_D$ (ns)	1.0-1.2
	Quantum yields <sup>1</sup>	$\tau_A$ (ns)	1.5
		$QY_D$	0.2
	Angle between absorption and emission dipole <sup>2</sup>	$QY_A$	0.3
		$\beta$ (°)	10.5
Bursts	Number of bursts	$n_{bursts}$	10000
	Minimum photon threshold for analytical burst sizes <sup>2</sup>	$N_{total,min}$	20
	Maximum photon threshold for analytical burst sizes <sup>2</sup>	$N_{total,max}$	150
	Exponent of analytical burst size distribution <sup>2</sup>	$\lambda$	-2.3
	Averaging regime		Ensemble
	QY correction		False
FRET	Förster radius	$R_0$ (nm)	5.4
	Dye quenching distance	$R_{quench}$ (nm)	1.0
	Gamma <sup>3</sup>		True

<sup>1</sup> Measured in this study and/or previously (3, 44, 45).

<sup>2</sup> Derived from the fundamental anisotropy  $r_0 = 0.38$  according to  $r_0 = (3\cos^2(\beta) - 1)/5$  (2).

<sup>2</sup> Analytical burst size distribution from (32).

<sup>3</sup> Set to `True` if simulation is to be compared against a gamma-corrected FRET experiment.

**Supplementary Table 4.** FRET measurements with Atto550/Atto647N compared to the multi-laboratory benchmark study (10).

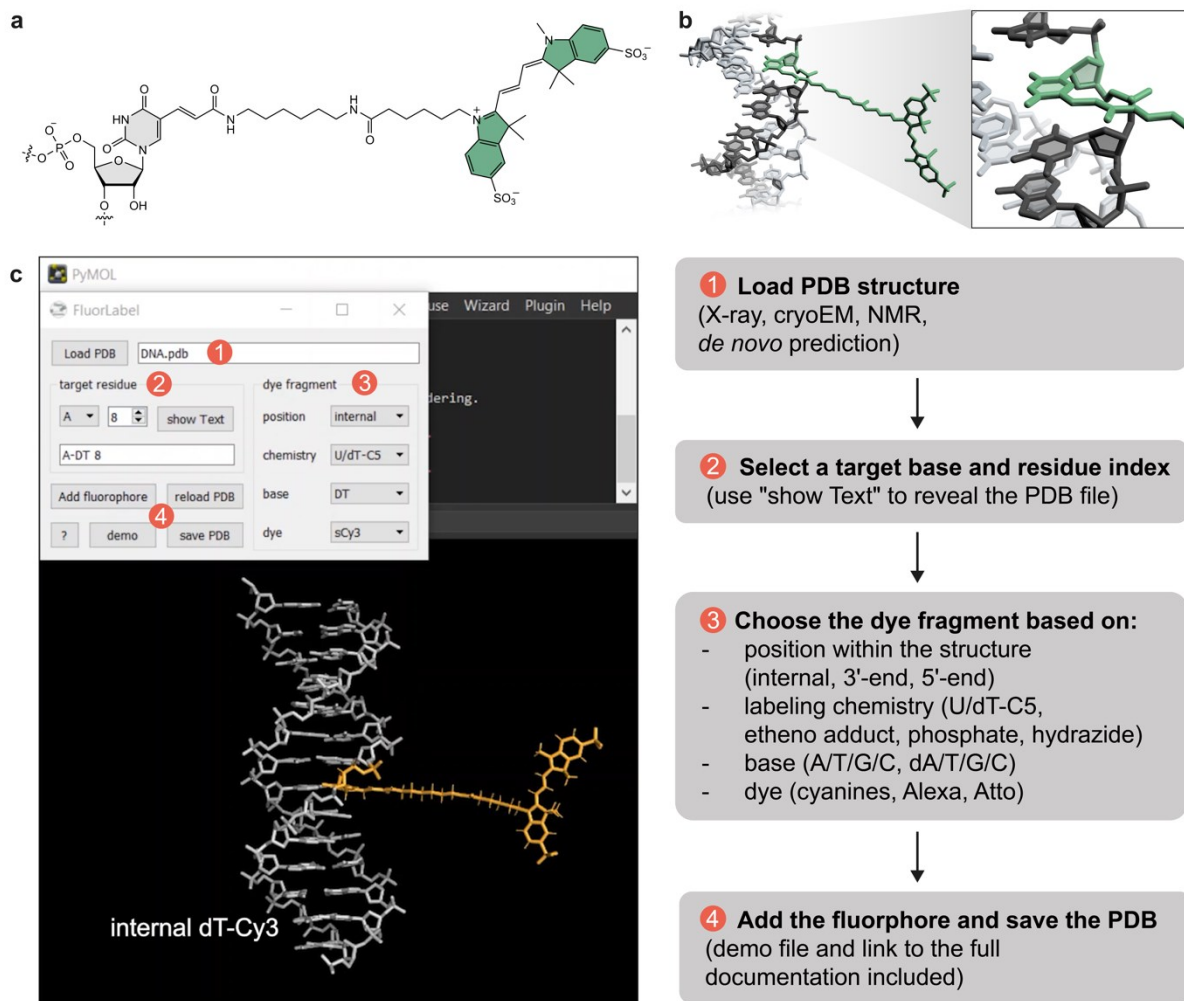
Construct	$E_{DA}$ (this study) <sup>1</sup>	$E_{DA}$ (multi-laboratory benchmark study) <sup>2</sup>
<b>Atto-low</b> T31(Atto550)-T31(Atto647N)	$0.14 \pm 0.14$	$0.15 \pm 0.02$
<b>Atto-mid</b> T23(Atto550)-T31(Atto647N)	$0.55 \pm 0.09$	$0.56 \pm 0.03$
<b>Atto-high</b> T19(Atto550)-T31(Atto647N)	$0.77 \pm 0.09$	$0.76 \pm 0.02$

<sup>1</sup>  $E_{DA}$  are given as the means  $\pm \sigma$  of a three subspecies Gaussian mixture model.

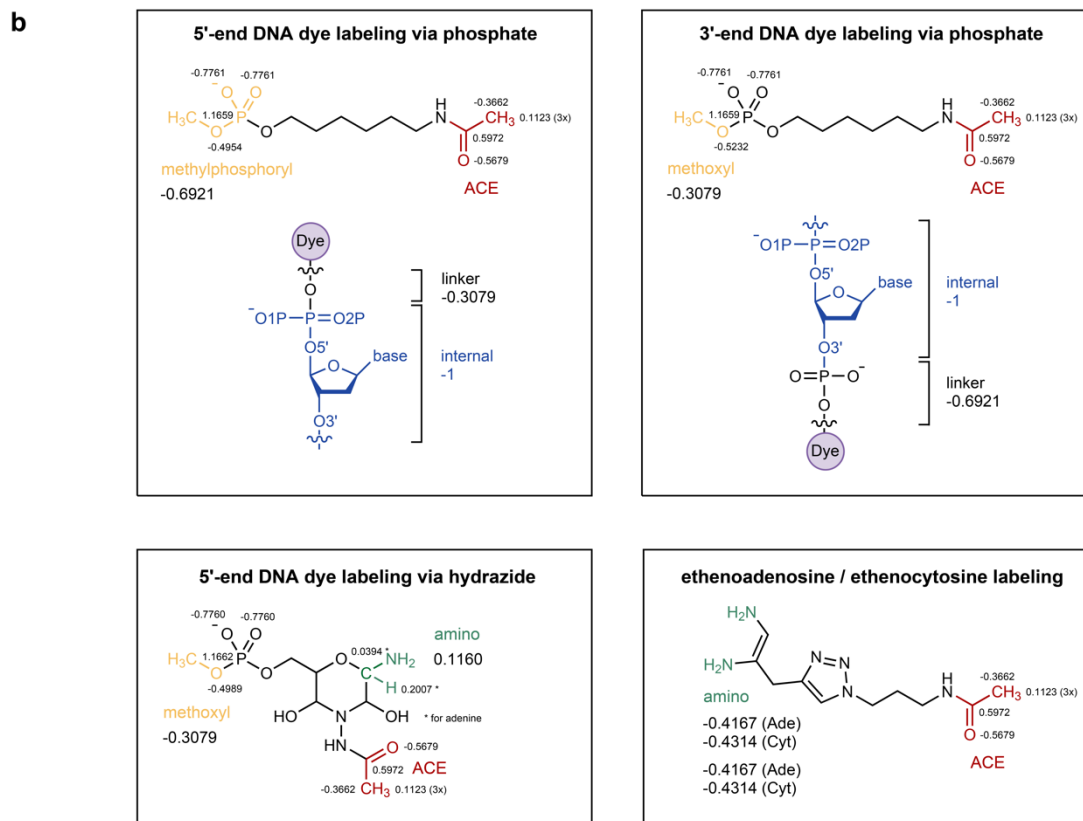
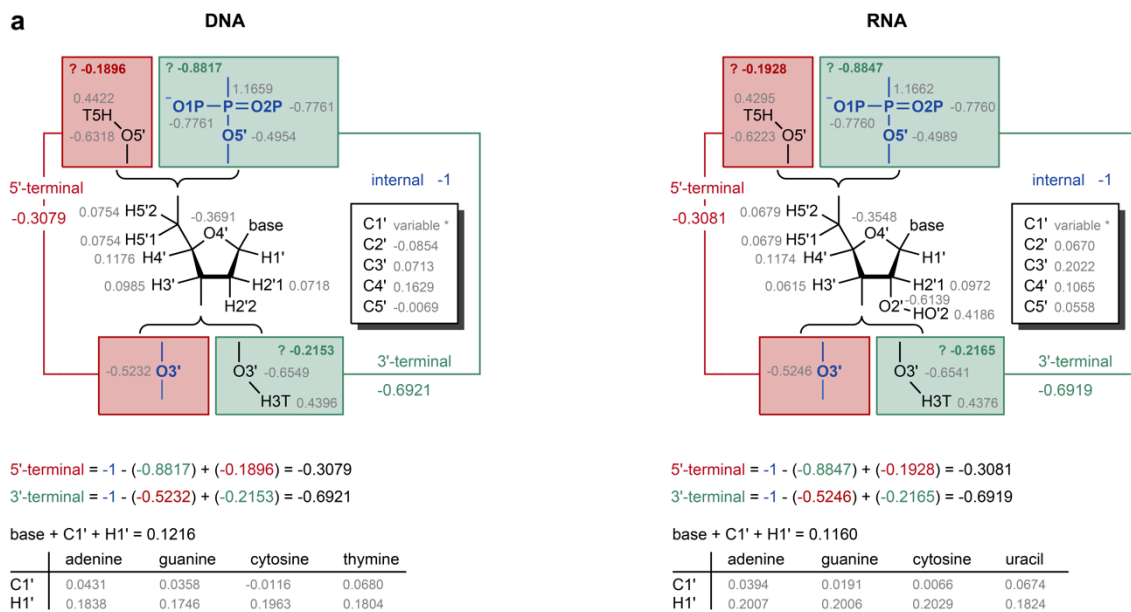
<sup>2</sup>  $E_{DA}$  represent the mean transfer efficiencies with standard deviations reporting on the statistical variations of the FRET measurements and thus their precision, as noted in the benchmark study.



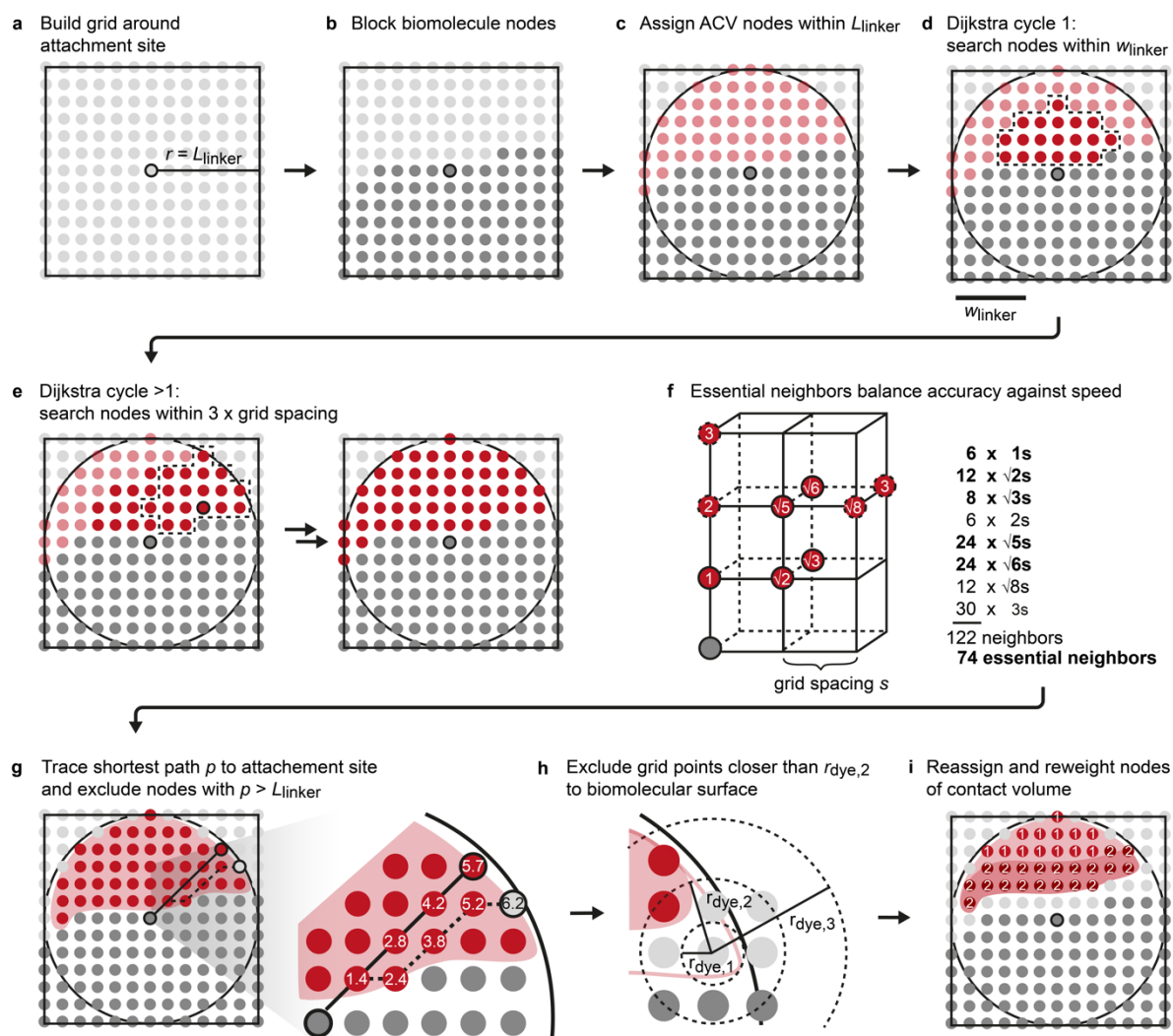
## Supplementary Figures



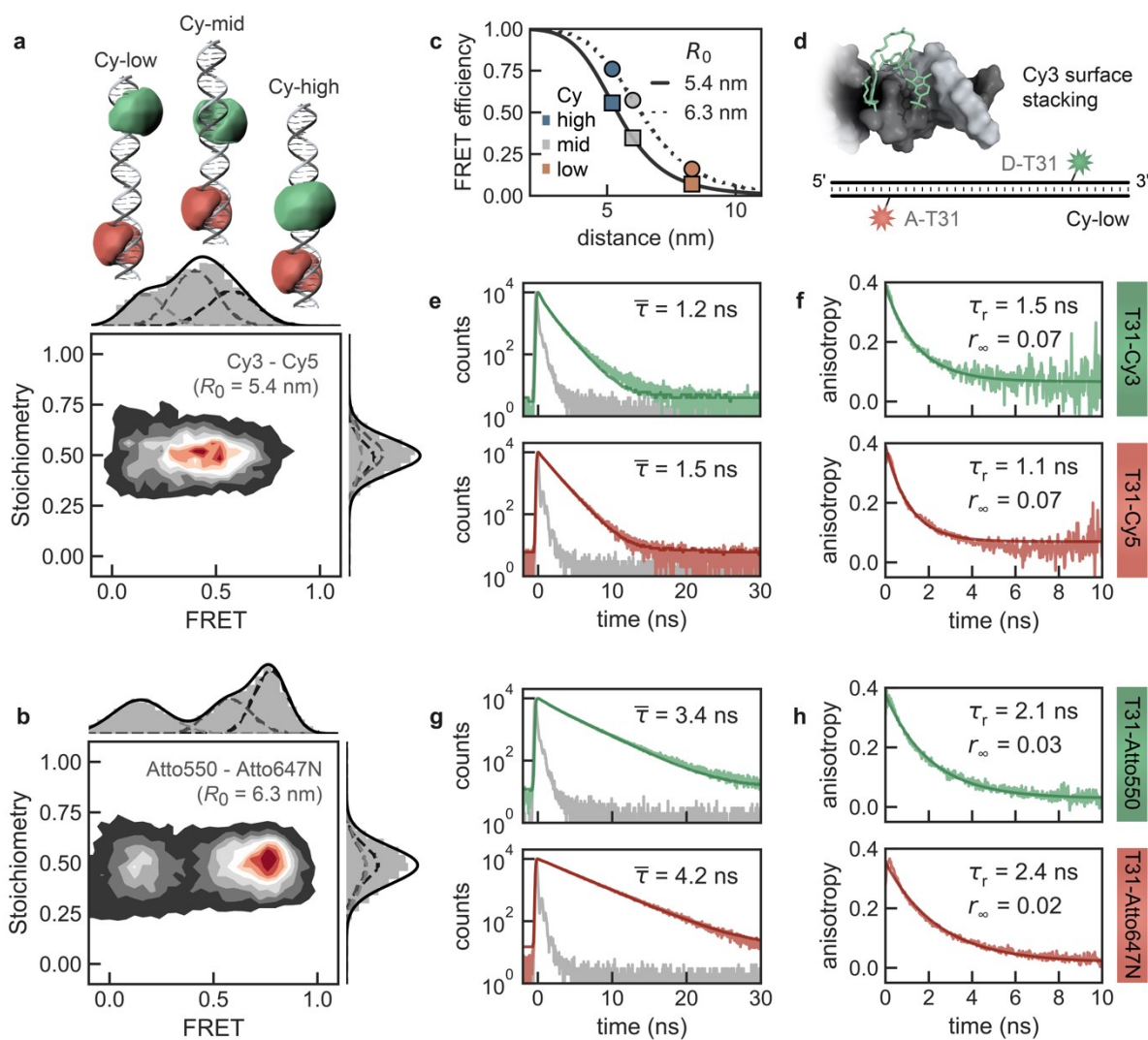
**Supplementary Fig. 1.** *In silico* fluorescence labeling of nucleic acids. **a** Chemical structure of a Cy3 fluorophore which is covalently attached to a C5-modified 2'-deoxyuridine. **b** Licorice representation of the Cy3 dye coupled to a DNA double helix. **c** Implementation of the automated *in silico* labeling as a PyMOL plugin. The user can select the target residue, the position (terminal, internal), the type of chemistry (e.g. dU-C5, etheno-adduct) and the dye (various cyanines, Alexa and Atto fluorophores).



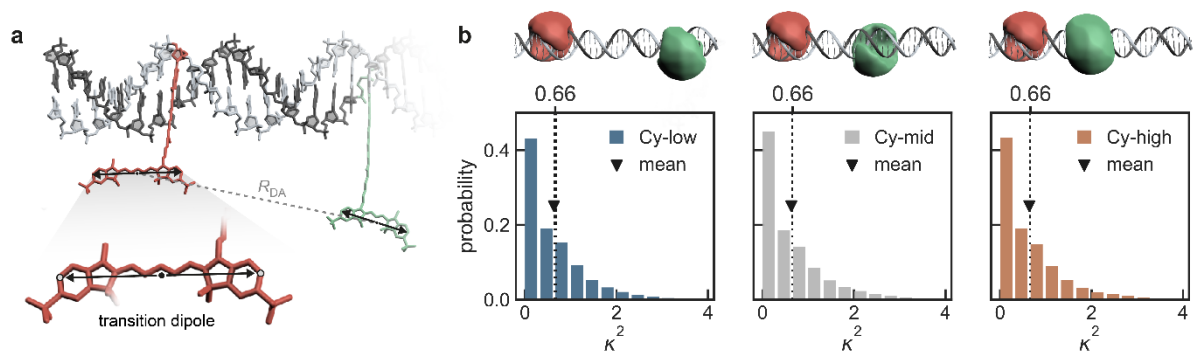
**Supplementary Fig. 2.** Force-field parameterization of linker atoms. **a** Partial charges of RNA and DNA residues in the AMBER force field derived by Cornell *et al.* (20) **b** Partial charges for the capping groups of the dye linker constrained during RESP fitting (17).



**Supplementary Fig. 3.** Simplified two-dimensional scheme of grid node assignment in the ACV algorithm. **a** A grid with radius  $r = L_{\text{linker}}$  and grid spacing  $s$  is constructed around the attachment site (black circle). **b** Grid nodes within the VdW surface of the biomolecule plus a clash radius  $r_{\text{clash}} = w_{\text{linker}}/2$  are assigned to the biomolecule (dark gray spheres). **c** Grid nodes within  $L_{\text{linker}}$  are tentatively allocated to the accessible volume (AV, light red spheres). **d** Dijkstra's algorithm is used to validate the ACV grid nodes, allowing only nodes with a shortest path  $p > L_{\text{linker}}$  to the attachment site (see below). To avoid blocking the algorithm near the attachment site, all nodes within half of the linker width  $w_{\text{linker}}$  are searched in the first round (nodes in dotted envelope). **e** In subsequent iterations, the grid is searched for nodes within a radius of  $3s$ . **f** Hence, nearest neighbors can be located along a rank or on a diagonal but must always be closer than  $3s$  from the current node. This results in a total of 122 neighbors. In practice, only 74 essential neighbors with distances  $1s, \sqrt{2}s, \sqrt{3}s, \sqrt{5}s, \sqrt{6}s$  from the current node are added to the adjacency list to balance accuracy of the final volume against speed. **g** Each node is retraced to the attachment site and only nodes with a shortest path (i.e. cumulative sum) smaller than  $L_{\text{linker}}$  are accepted. **h** Grid points located closer to the biomolecular surface than one of the dye radii (here middle radius  $r_{\text{dye},2}$ ) are excluded. **i** Nodes within  $d_{\text{CV}}$  from the biomolecule's VdW surface are allocated to the contact volume and are assigned a weight according to Equation 3 (see main text).



**Supplementary Fig. 4.** Comparison of different fluorophore pairs for FRET, fluorescence lifetime and anisotropy measurements on double stranded DNA. **a** 2D FRET-stoichiometry histogram of three cyanine labeled DNA double helices mixed in near equimolar amounts and recorded at 20 mM  $Mg^{2+}$  and 5 mM NaCl at pH 7.5. Accessible volume pictograms indicate the different dye positions. **b** Control FRET measurements with Atto-dyes as used in the original benchmark study (10). **c** Expected transfer efficiencies for different Förster radii  $R_0$  (Cy3-Cy5: 5.4 nm; Atto550-Atto647N: 6.3 nm) as a function of the inter-dye distance. **d** Surface interaction of Cy3 in the Cy-low sample used for lifetime and anisotropy experiments. In this duplex construct the donor fluorescence lifetime is least affected by FRET. **e/f** Lifetime and anisotropy decays of the cyanine dyes. **g/h** Same measurements as in e/f but for the Atto samples. The anisotropy decays are very similar to those of the cyanines although with an overall better signal-to-noise ratio due to the longer lifetimes of the Atto labels.



**Supplementary Fig. 5.** Fluorophores rotate isotropically in all-atom MD simulations. **a** Definition of the transition dipole by two peripheral carbon atoms in cyanine dyes. The donor acceptor distance  $R_{DA}$  is given by the distance between the centers of mass of the two atoms on each dye. **b**  $\kappa^2$ -distributions of the three Cy-samples from all-atom MD simulations. The mean of the distributions coincides with the isotropic average ( $\kappa^2 = 2/3$ , dashed line) confirming that the ACV model, which presumes isotropic rotation of the fluorophores, is indeed applicable.



## References

1. Greenfeld, M. and Herschlag, D. (2013) Fluorescently Labeling Synthetic RNAs. In *Methods in Enzymology*. Elsevier, Vol. 530, pp. 281–297.
2. Lakowicz, J.R. (2006) Principles of fluorescence spectroscopy 3rd ed. Springer, New York.
3. Steffen, F.D., Sigel, R.K.O. and Börner, R. (2016) An atomistic view on carbocyanine photophysics in the realm of RNA. *Phys. Chem. Chem. Phys.*, **18**, 29045–29055.
4. Müller, B.K., Zaychikov, E., Bräuchle, C. and Lamb, D.C. (2005) Pulsed Interleaved Excitation. *Biophysical Journal*, **89**, 3508–3522.
5. Wahl, M., Röhlicke, T., Rahn, H.-J., Erdmann, R., Kell, G., Ahlrichs, A., Kernbach, M., Schell, A.W. and Benson, O. (2013) Integrated multichannel photon timing instrument with very short dead time and high throughput. *Review of Scientific Instruments*, **84**, 043102.
6. Schimpf, W., Barth, A., Hendrix, J. and Lamb, D.C. (2018) PAM: A Framework for Integrated Analysis of Imaging, Single-Molecule, and Ensemble Fluorescence Data. *Biophysical Journal*, **114**, 1518–1528.
7. Nir, E., Michalet, X., Hamadani, K.M., Laurence, T.A., Neuhauser, D., Kovchegov, Y. and Weiss, S. (2006) Shot-Noise Limited Single-Molecule FRET Histograms: Comparison between Theory and Experiments. *J. Phys. Chem. B*, **110**, 22103–22124.
8. Kapanidis, A.N., Lee, N.K., Laurence, T.A., Doose, S., Margeat, E. and Weiss, S. (2004) Fluorescence-aided molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 8936–8941.
9. Lee, N.K., Kapanidis, A.N., Wang, Y., Michalet, X., Mukhopadhyay, J., Ebright, R.H. and Weiss, S. (2005) Accurate FRET Measurements within Single Diffusing Biomolecules Using Alternating-Laser Excitation. *Biophysical Journal*, **88**, 2939–2953.
10. Hellenkamp, B., Schmid, S., Doroshenko, O., Opanasyuk, O., Kühnemuth, R., Rezaei Adariani, S., Ambrose, B., Aznauryan, M., Barth, A., Birkedal, V., *et al.* (2018) Precision and accuracy of single-molecule FRET measurements—a multi-laboratory benchmark study. *Nat Methods*, **15**, 669–676.
11. Suddala, K.C., Price, I.R., Dandpat, S.S., Janeček, M., Kührová, P., Šponer, J., Banáš, P., Ke, A. and Walter, N.G. (2019) Local-to-global signal transduction at the core of a Mn<sup>2+</sup> sensing riboswitch. *Nat Commun*, **10**, 4304.
12. Steffen, F.D., Börner, R., Freisinger, E. and Sigel, R.K.O. (2019) Stick, Flick, Click: DNA-guided Fluorescent Labeling of Long RNA for Single-molecule FRET. *Chimia*, **73**, 257.
13. Graen, T., Hoefling, M. and Grubmüller, H. (2014) AMBER-DYES: Characterization of Charge Fluctuations and Force Field Parameterization of Fluorescent Dyes for Molecular Dynamics Simulations. *J. Chem. Theory Comput.*, **10**, 5505–5512.
14. Schepers, B. and Gohlke, H. (2020) AMBER-DYES in AMBER: Implementation of fluorophore and linker parameters into AmberTools. *J. Chem. Phys.*, **152**, 221103.
15. Sousa Da Silva, A.W. and Vranken, W.F. (2012) ACPYPE - AnteChamber PYthon Parser interface. *BMC Res Notes*, **5**, 367.
16. 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., *et al.* (2016) Gaussian 09.
17. Cieplak, P., Cornell, W.D., Bayly, C. and Kollman, P.A. (1995) Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.*, **16**, 1357–1377.
18. Case, D.A., Belfon, K., Ben-Shalom, I.Y., Brozell, S.R., Cerutti, D.S., Cheatham, T.E., Cruzeiro, V.W.D., Darden, T.A., Duke, R.E., Giambasu, G., *et al.* (2020) AMBER 2020.
19. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B. and Lindahl, E. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1–2**, 19–25.
20. 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. and Kollman, P.A. (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
21. Pérez, A., Luque, F.J. and Orozco, M. (2007) Dynamics of B-DNA on the Microsecond Time Scale. *J. Am. Chem. Soc.*, **129**, 14739–14745.

22. Banáš,P., Hollas,D., Zgarbová,M., Jurečka,P., Orozco,M., Cheatham,T.E., Šponer,J. and Otyepka,M. (2010) Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theory Comput.*, **6**, 3836–3849.
23. Zgarbová,M., Otyepka,M., Šponer,J., Mládek,A., Banáš,P., Cheatham,T.E. and Jurečka,P. (2011) Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.*, **7**, 2886–2902.
24. Zgarbová,M., Luque,F.J., Šponer,J., Cheatham,T.E., Otyepka,M. and Jurečka,P. (2013) Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *J. Chem. Theory Comput.*, **9**, 2339–2354.
25. Krepl,M., Zgarbová,M., Stadlbauer,P., Otyepka,M., Banáš,P., Koča,J., Cheatham,T.E., Jurečka,P. and Šponer,J. (2012) Reference Simulations of Noncanonical Nucleic Acids with Different  $\chi$  Variants of the AMBER Force Field: Quadruplex DNA, Quadruplex RNA, and Z-DNA. *J. Chem. Theory Comput.*, **8**, 2506–2520.
26. Zgarbová,M., Šponer,J., Otyepka,M., Cheatham,T.E., Galindo-Murillo,R. and Jurečka,P. (2015) Refinement of the Sugar–Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *J. Chem. Theory Comput.*, **11**, 5723–5736.
27. Joung,I.S. and Cheatham,T.E. (2008) Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B*, **112**, 9020–9041.
28. Horn,H.W., Swope,W.C., Pitner,J.W., Madura,J.D., Dick,T.J., Hura,G.L. and Head-Gordon,T. (2004) Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *The Journal of Chemical Physics*, **120**, 9665–9678.
29. Hess,B., Bekker,H., Berendsen,H.J.C. and Fraaije,J.G.E.M. (1997) LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, **18**, 1463–1472.
30. Essmann,U., Perera,L., Berkowitz,M.L., Darden,T., Lee,H. and Pedersen,L.G. (1995) A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, **103**, 8577–8593.
31. Tribello,G.A., Bonomi,M., Branduardi,D., Camilloni,C. and Bussi,G. (2014) PLUMED 2: New feathers for an old bird. *Computer Physics Communications*, **185**, 604–613.
32. Hoefling,M. and Grubmüller,H. (2013) In silico FRET from simulated dye dynamics. *Computer Physics Communications*, **184**, 841–852.
33. Hoefling,M., Lima,N., Haenni,D., Seidel,C.A.M., Schuler,B. and Grubmüller,H. (2011) Structural Heterogeneity and Quantitative FRET Efficiency Distributions of Polypyrrolines through a Hybrid Atomistic Simulation and Monte Carlo Approach. *PLoS ONE*, **6**, e19791.
34. Gopich,I.V. and Szabo,A. (2011) Theory of Single-Molecule FRET Efficiency Histograms. In Komatsuzaki,T., Kawakami,M., Takahashi,S., Yang,H., Silbey,R.J. (eds), *Advances in Chemical Physics*. John Wiley & Sons, Inc., Hoboken, NJ, USA, pp. 245–297.
35. Watkins,A.M., Rangan,R. and Das,R. (2020) FARFAR2: Improved De Novo Rosetta Prediction of Complex Global RNA Folds. *Structure*, **28**, 963-976.e6.
36. Leontis,N.B. and Zirbel,C.L. (2012) Nonredundant 3D Structure Datasets for RNA Knowledge Extraction and Benchmarking. In Leontis,N., Westhof,E. (eds), *RNA 3D Structure Analysis and Prediction*, Nucleic Acids and Molecular Biology. Springer Berlin Heidelberg, Berlin, Heidelberg, Vol. 27, pp. 281–298.
37. Watkins,A.M., Geniesse,C., Kladwang,W., Zakrevsky,P., Jaeger,L. and Das,R. (2018) Blind prediction of noncanonical RNA structure at atomic accuracy. *Science Advances*.
38. Magnus,M., Antczak,M., Zok,T., Wiedemann,J., Lukasiak,P., Cao,Y., Bujnicki,J.M., Westhof,E., Szachniuk,M. and Miao,Z. (2019) RNA-Puzzles toolkit: a computational resource of RNA 3D structure benchmark datasets, structure manipulation, and evaluation tools. *Nucleic Acids Research*, 10.1093/nar/gkz1108.
39. Steffen,F.D., Sigel,R.K.O. and Börner,R. (2021) FRETraj: integrating single-molecule spectroscopy with molecular dynamics. *Bioinformatics*, **37**, 3953–3955.
40. Zhao,M., Steffen,F.D., Börner,R., Schaffer,M.F., Sigel,R.K.O. and Freisinger,E. (2018) Site-specific dual-color labeling of long RNAs for single-molecule spectroscopy. *Nucleic Acids Research*, **46**, e13–e13.
41. Klöcker,N., Weissenboeck,F.P. and Rentmeister,A. (2020) Covalent labeling of nucleic acids. *Chem. Soc. Rev.*, **49**, 8749–8773.



42. Hanspach,G., Trucks,S. and Hengesbach,M. (2019) Strategic labelling approaches for RNA single-molecule spectroscopy. *RNA Biology*, **16**, 1119–1132.
43. Qin,P.Z. and Pyle,A.M. (1999) Site-Specific Labeling of RNA with Fluorophores and Other Structural Probes. *Methods*, **18**, 60–70.
44. Sanborn,M.E., Connolly,B.K., Gurunathan,K. and Levitus,M. (2007) Fluorescence Properties and Photophysics of the Sulfoindocyanine Cy3 Linked Covalently to DNA. *J. Phys. Chem. B*, **111**, 11064–11074.
45. Harvey,B.J., Perez,C. and Levitus,M. (2009) DNA sequence-dependent enhancement of Cy3 fluorescence. *Photochem Photobiol Sci*, **8**, 1105–1110.