Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

Tautomers of a Fluorescent G Surrogate and Their Distinct Photophysics Provide Additional Information Channels

Marianna Sholokh[+], Roberto Improta[+], Mattia Mori[+], Rajhans Sharma, Cyril Kenfack, Dongwon Shin, Karine Voltz, Roland H. Stote, Olga A. Zaporozhets, Maurizio Botta, Yitzhak Tor, and Yves Mély**

anie_201601688_sm_miscellaneous_information.pdf

Table of contents

1. Materials

Thienoguanosine (thG) and 2′-deoxy-thienoguanosine (dthG)-labeled (–)PBS DNA were synthesized as previously described by Shin et al.[1] and Sholokh et al.,[2] respectively. Stock solutions of thG were prepared in spectroscopic grade DMSO. The complementary native and mismatched (+)PBS DNA were purchased from IBA Nucleic Acids Product Supply (Germany). dthG-labeled duplexes (–)/(+)PBS were prepared by hybridization of dthG7(–)PBS and (+)PBS samples (concentration ratio 1:3), denatured at 85 °C for 3 min and then slowly cooled down to the room temperature. NC(11-55) peptide was synthesized on a Applied Biosystems A433 peptide synthesizer, as described by de Rocquigny[3] and prepared using 2.2 equivalents of Zn(II). All experiments with aqueous solutions were done in 25 mM TRIS-HCl buffer (pH=7.5), 30 mM NaCl and 0.2 mM $MgCl_2$.

2. Spectroscopic measurements

Spectroscopic grade solvents were used for absorption and fluorescence spectroscopy measurements. Absorption spectra were recorded on a Cary 4000 UV-visible spectrophotometer (Varian). Fluorescence excitation and emission spectra were recorded on a FluoroMax 4 spectrofluorimeter (Jobin Yvon) equipped with a thermostated cell compartment at 20 ± 0.5 °C. Fluorescence spectra were corrected for Raman scattering, lamp fluctuations and instrumental wavelength-dependent bias. thG concentration in the various solvents was 6 µM, with a final DMSO concentration of 0.1 v/v %. Photostability measurements were performed in cuvettes with 50 µL total volume under continuous illumination at 325, 350, 360, 370 or 380 nm during 2000 s.

Deconvolution procedure

As the emission spectrum of the red-shifted tautomer can be obtained at excitation wavelengths > 350 nm, the emission spectrum of the blue-shifted isomer in protic solvents was extracted from the emission spectrum recorded at λ_{ex} = 283 nm, by subtracting the emission spectrum of the red-shifted tautomer normalized at wavelengths > 525 nm. The individual absorption spectra of the two tautomers were deduced from the absorption spectrum of thG, by normalizing the excitation spectra at wavelengths (> 350 nm) where only the red-shifted form absorbs.

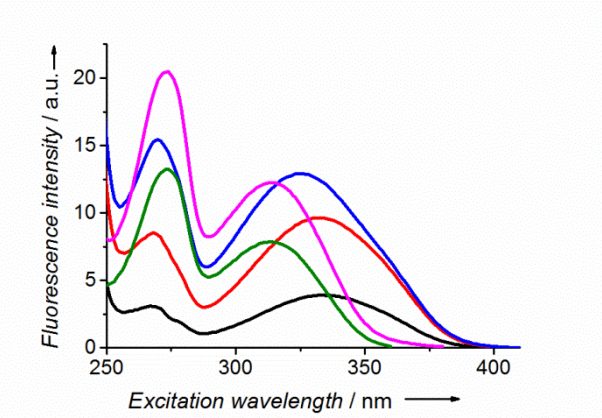

Figure S1. Excitation spectra of thG in TRIS-HCl buffer (25 mM, pH=7.5, 30 mM NaCl, 0.2 mM MgCl2) at different emission wavelengths: 550 nm (black line), 500 nm (red), 450 nm (blue), 400 nm (magenta), 375 nm (green). A maximum at 334 nm is observed when emission is recorded in the 500–550 nm range, and the excitation maximum is progressively blue-shifted down to 313 nm, for shorter wavelength emission.

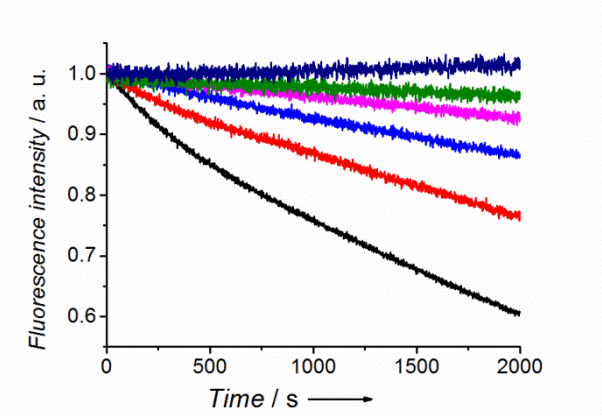

Figure S2. Photostability of the thG ground-state tautomers in buffer. Kinetics at different excitation and emission wavelengths, respectively: 325 and 400 nm (black); 325 and 525 nm (red); 350 and 525 nm (blue); 360 and 525 nm (magenta); 370 and 525 nm (green); 380 and 525 nm (dark blue). Concentration of thG was 1 µM.

Table S1. Photophysical Properties of thG in Buffer and Organic Solvents [a]

Solvent	$E_T(30)$	ε	α	β	λ_{abs}	λ_{em}^{325}	λ_{em}^{380}
Buffer	63.1	78.35	1.17	0.18	322	454	468
Methanol	55.4	32.61	0.93	0.62	326	458	459
Ethanol	51.9	24.85	0.83	0.77	329	458	458
n-Butanol	49.7	17.33	0.79	0.88	330	457	457
2-Methyl-2-butanol	41.0	5.82	0.28	0.93	335	446	445
N,N-dimethylformamide	43.2	37.21	0.00	0.69	338	440	440
Acetonitrile	45.6	35.68	0.19	0.31	330	432	432
Ethyl acetate	38.1	5.98	0.00	0.45	331	426	427
1,4-Dioxane	36.0	2.20	0.00	0.37	330	424	426

[a] $E_T(30)$ is the empiric polarity index[4] reported in kcal mol^{-1}; ε is the dielectric constant at 298 K; α is the Kamlet-Taft solvent hydrogen bond acidity; β is the Kamlet-Taft solvent hydrogen bond basicity;[5] λ_{abs} is the absorption maxima in nm; λ_{em}^{325} and λ_{em}^{380} are the fluorescence emission maxima at 325 nm and 380 nm excitation wavelengths, respectively, reported in nm. TRIS-HCl buffer 25 mM, pH=7.5, 30 mM NaCl, 0.2 mM $MgCl_2$ was used.

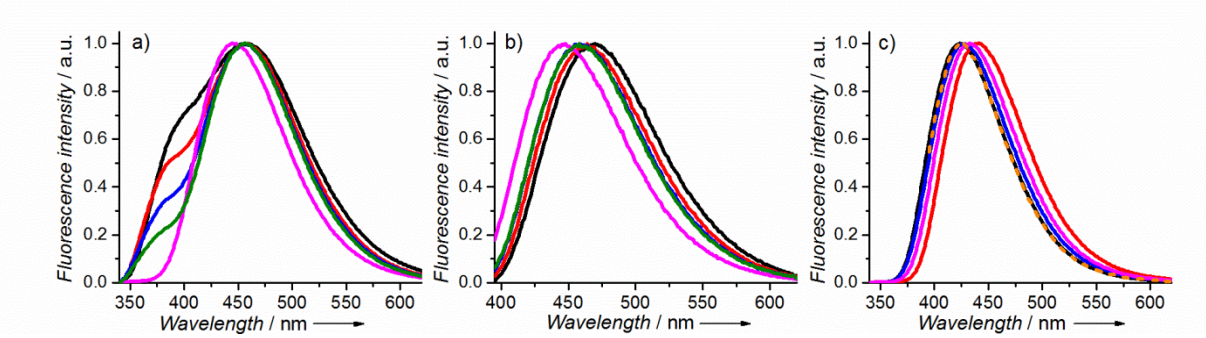

Figure S3. Normalized emission spectra of thG in various solvents. (a) Buffer (black), methanol (red), ethanol (blue), *n*-butanol (green), 2-methyl-2-butanol (magenta), λ_{ex} = 325 nm; (b) Emission spectra in the same solvents as in (a), but with λ_{ex} = 380 nm; (c) Emission spectra in 1,4-dioxane (black), *N,N*-dimethylformamide (red), ethyl acetate (blue), acetonitrile

(magenta), λ_{ex} = 325 nm; superimposed normalized thG emission spectra in 1,4-dioxane obtained at λ_{ex} = 380 nm (orange dash).

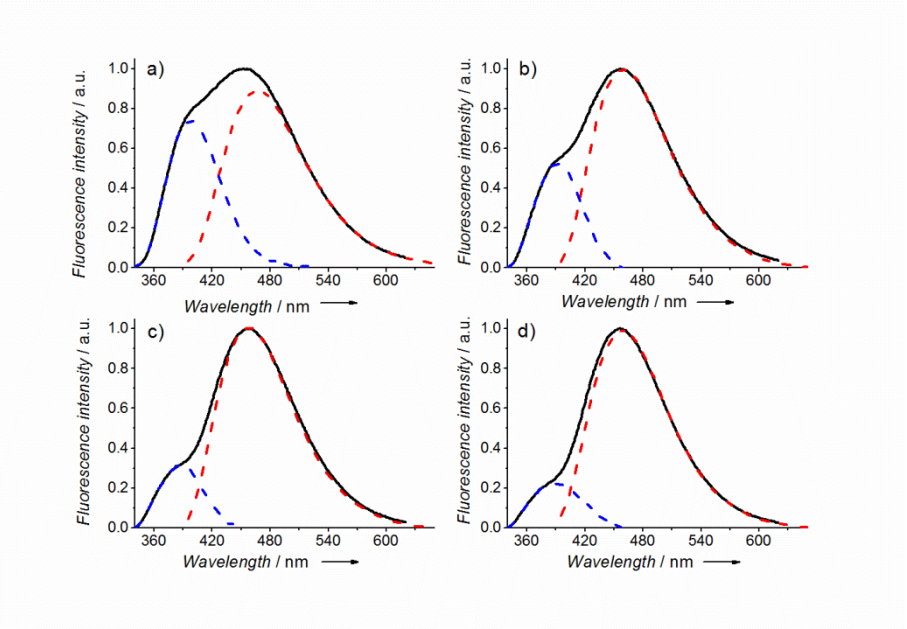

Figure S4. Deconvolution of the emission spectra of thG nucleoside in buffer (a), methanol (b), ethanol (c), and *n*-butanol (d). Excitation wavelength was 325 nm.

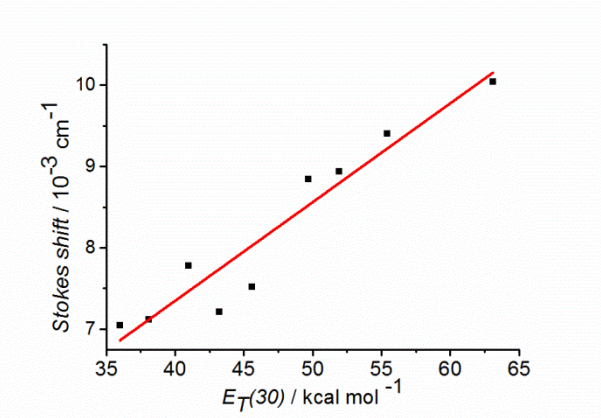

Figure S5. Dependence of the Stokes shift of the red-shifted tautomer on the empiric polarity index $E_T(30)$ (symbols). The red line represents the linear fit to the data with a slope of 0.12 ± 0.01 and an intercept of 2.4 ± 0.6, R^2 = 0.90. The $E_T(30)$ values of the solvents are given in

Table S1. The Stokes shift was calculated from the absorption and emission maxima, after correction of the emission spectra according to: Intensity ($\bar{\upsilon}$) = Intensity (λ) × λ^2.[6]

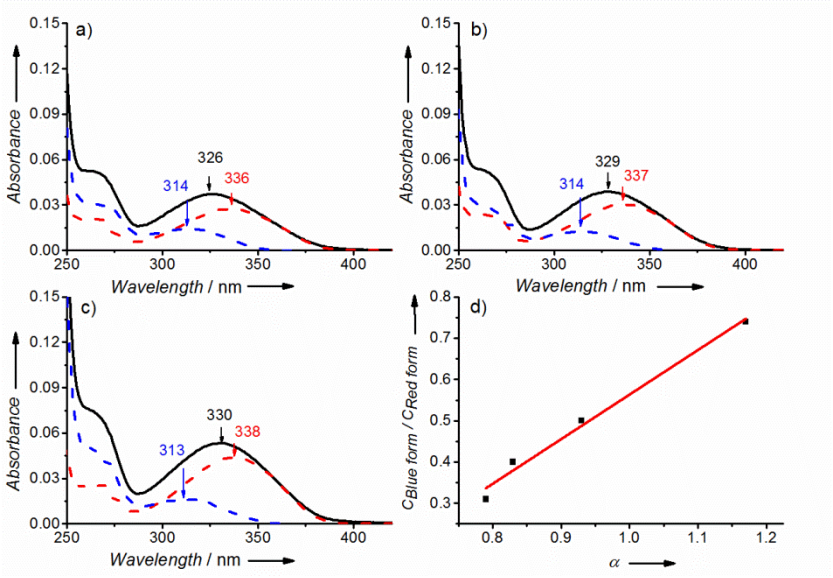

Figure S6. Deconvolution of the absorption spectra of thG in methanol (a), ethanol (b) and *n*-butanol (c) (black line) in its two ground state forms. Absorption spectra of the blue-shifted (blue dashed line) and red-shifted (red dashed line) forms. (d) Dependence of the concentration ratio of the two forms on the Kamlet-Taft's hydrogen bond acidity α. The straight line which was fitted to the experimental points shows a slope of 1.08 ± 0.08 and an intercept of -0.51 ± 0.08, R^2 = 0.98.

3. Quantum chemical calculations

The analysis is based on Density Functional Theory (DFT) and on its time dependent extension (TD-DFT), by using two different functionals, PBE0[7] and M052X.[8] PBE0 is a parameter-free functional, which accurately describes the bright states of guanine[9] and of other nucleobases,[10] providing vertical excitation and emission energies within ~0.15 eV of the corresponding experimental absorption maxima. Since PBE0 might overestimate the stability of charge transfer (CT) transitions[11] and to avoid any possible artifacts due to small

contribution of solute–solvent CT excitations, we have verified the results by using M052X functional, which is particularly effective for the treatment of non-bonding interactions and is not biased by the traditional failure of TD-DFT in describing CT states.[12]

Geometry optimizations have been performed at the less computationally demanding 6-31G(d) level, refining the vertical absorption νA and emission energies νE by single-point calculations employing more extended 6-31+G(d,p) and 6-311+G(2d,2p) basis sets.

Table S2. Relative Energy (in eV) of the thG-H3 Tautomer with respect to the thG-H1 Tautomer (taken as 0) Computed at the PCM/DFT/6-311+G(2d,2p) Level by Using Different Basis Sets and Geometries Optimized at the PCM/DFT/6-31G(d) Level [a]

	PBE0	M052X
	Gas phase	
6-31G(d)	0.441 (0.411)	0.439 (0.400)
6-311+G(2d,2p)	0.399	0.396
	Dioxane (PCM)	
6-31G(d)	0.349 (0.318)	0.345 (0.331)
6-311+G(2d,2p)	0.297	0.292
	Dioxane (PCM + 1 Dioxane)	
6-31G(d)	0.321 (0.299)	0.291 (0.226)
6-311+G(2d,2p)	0.277	0.250
	Water (PCM)	
6-31G(d)	0.199 (0.160)	0.181 (0.186)
6-311+G(2d,2p)	0.114	0.103
	Water (PCM+ $2H_2O$)	
6-31G(d)	0.214 (0.124)	0.179 (0.088)
6-311+G(2d,2p)	0.067	0.054
	Water (PCM + $6H_2O$)	
6-31G(d)	0.138 (0.141)	0.105 (0.141)

6-311+G(2d,2p)	0.049 [b]	0.023 [c]

[a] The values obtained by including vibrational corrections are in parentheses. At 300 K, the calculated molar fraction of the thG-H3 tautomer in water is 0.13[b] and 0.29 [c].

Table S3. Relative Energy (in eV) of the Different thG Tautomers Computed at the PCM/PBE0/level by Using Different Basis Sets and Geometries Optimized at the PCM/PBE0/6-31G(d) Level [a]

	thG-H3	Enol-amino	Keto-imino	Enol-imino
		Gas Phase		
6-31G(d)	0.44 (0.41)	0.33 (0.33)	0.31 (0.31)	0.83 (0.83)
6-31+G(d,p)	0.43	0.25	0.31	0.75
6-311+G(2d,2p)	0.40	0.24	0.29	0.72
		Dioxane (only PCM)		
6-31G(d)	0.35 (0.32)	0.39 (0.38)	0.31 (0.29)	0.85 (0.86)
6-31+G(d,p)	0.32	0.32	0.29	0.78
6-311+G(2d,2p)	0.30	0.31	0.27	0.75
		Water (only PCM)		
6-31G(d)	0.20 (0.16)	0.46 (0.45)	0.32 (0.35)	0.89 (0.91)
6-31+G(d,p)	0.13	0.41	0.30	0.82
6-311+G(2d,2p)	0.11	0.39	0.29	0.79

[a] The values obtained by including vibrational corrections are in parentheses. The energy of the thG-H1 tautomer is taken as 0.

Bulk solvent effects on the electronic states are accounted for with the polarizable continuum model (PCM).[13] The excitation and emission energies, νA and νE, are computed with the "standard" linear response (LR) implementation of PCM/TD-DFT, which has been also used in the excited-state geometry optimizations.[14] To verify the effect of the explicit inclusion of solute–solvent interaction on the optical properties of thG, we have included six

H_2O (Figure S7a, b) or one dioxane (Figure S7c) molecule. In the excited-state geometry, the first solvation shell was fully optimized, i.e., was treated like solute degrees of freedom.

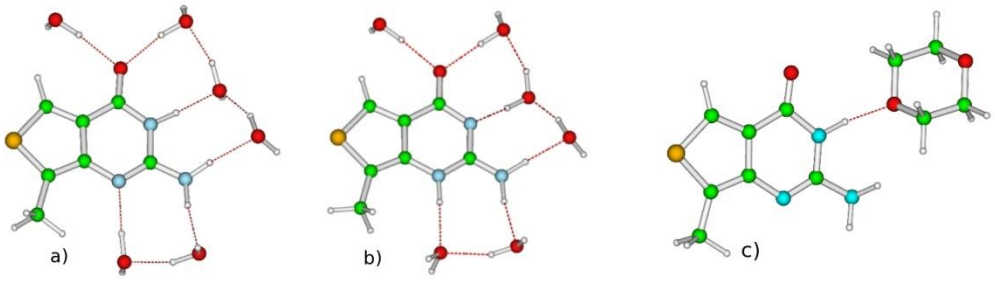

Figure S7. Computational models used to study the absorption and emission spectrum of the thG-H1 and thG-H3 tautomers in water (with six water molecules of the first solvation shell) (a and b, respectively) and thG-H1 in dioxane, with one explicit dioxane molecule (c).

In order to support the solidity of our computational analysis in water, we checked that our prediction do not qualitatively depend on the number of water molecules explicitly included in the computational model. In Table S2, we report the data obtained when only two water molecules, namely those H-bonded to the carbonyl group, are included in our model. In comparison with the calculations using six water molecules, the minimal model with two molecules of water already accounts for ~70% of the difference with respect to pure PCM calculations. Thus, the inclusion of only two water molecules, whose presence in the first solvation shell is extremely likely, is sufficient to provide similar stabilities for thG-H3 and thG-H1 tautomers.

Furthermore, analysis of solute-solvent H-bond interactions performed by G, dthG-H1 or dthG-H3 within the context of the (–)PBS molecule (Figure S12 structure #10, solvent exposed base) clearly confirmed that the solvation shell used in QM calculations may reflect the solvation shell observed in the DNA environment. Indeed, analysis of the solvation shell issuing from the MD simulations in the (–)PBS molecule shows the existence of an average of

~3.5 and ~2.8 H-bonds between the solvent molecules and the dthG-H1 and dthG-H3 tautomers, respectively (Figure S8). Considering that the thresholds for the definition of the solute-solvent H-bonds are rather strict (distance between the heavy atoms < 3 Å and angle > 135° (CPPTRAJ default values, see Roe et al. [15]) and that our QM calculation shows that 2 solute-solvent hydrogen bonds are sufficient to get the stability of the two tautomers very close, we can infer that the conclusions provided by QM calculations on the isolated bases apply also to the study of dthG in a DNA environment, such as in (–)PBS.

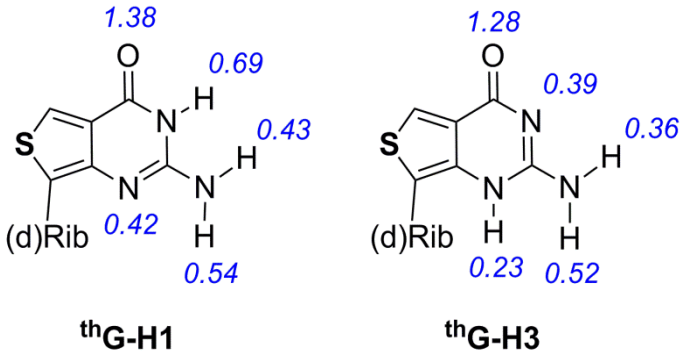

Figure S8. Average number of H-bonds calculated along MD trajectories for dthG-H1 and dthG-H3 in (–)PBS.

Finally, to further confirm the relevance of the computational approaches used for thG, we applied them to 9-methyl-guanine for which estimates based on alternative solvation models (MD/FEP) are available. For 9-methyl-guanine•5H_2O, including bulk solvent effect at the PCM level, the keto-amino (G-H1) was found to be more stable in aqueous solution than the enol (G-OH) tautomer by 7.1 kcal/mol according to PCM/PBE0/6-311+G(2d,2p)//PCM/PBE0/6-31G(d) calculations and by 6.2 kcal/mol according to PCM/M052X/6-311+G(2d,2p)//PCM/M052X/6-31G(d) calculations. Both values are very close to the estimates (5.7 - 7.5 kcal/mol) obtained by computational methods adopting MD/TI approaches to compute hydration energy. [16]

Table S4. Vertical Absorption and Emission Wavelengths Computed for the thG-H1 and thG-H3 Tautomers at the PCM/TD-DFT/6-311+G(2d,2p) Level in Different Solvents and Solvation Models by Using M052X and PBE0 Functionals [a]

thG-H1 tautomer				thG-H3 tautomer			
PBE0		M052X		PBE0		M052X	
Abs	Em	Abs	Em	Abs	Em	Abs	Em
Gas phase							
317 (0.08)	406 (0.06)	288 (0.11)	368 (0.08)	286 (0.11)	340 (0.08) [b]	261 (0.15)	330 (0.09) [b]
Dioxane (PCM)							
325 (0.10)	416 (0.08)	292 (0.14)	376 (0.11)	292 (0.14)	352 (0.11)	267 (0.19)	323 (0.16)
Dioxane (PCM + 1 Dioxane)							
330 (0.11)	418 (0.08)	295 (0.15)	378 (0.11)	293 (0.13)	353 (0.12)	269 (0.20)	332 (0.14)
Water (PCM)							
331 (0.09)	422 (0.08)	296 (0.16)	381 (0.11)	299 (0.12)	362 (0.10)	273 (0.16)	333 (0.13)
Water (PCM + $6H_2O$)							
350 (0.08)	448 (0.07)	308 (0.12)	402 (0.10)	309 (0.10)	383 (0.08)	279 (0.14)	349 (0.11)

[a] Oscillator strengths are given in parentheses. Geometry optimizations are at the PCM/TD-DFT/6-31G(d) level. [b] 6-31+G(d,p) optimized geometries were used.

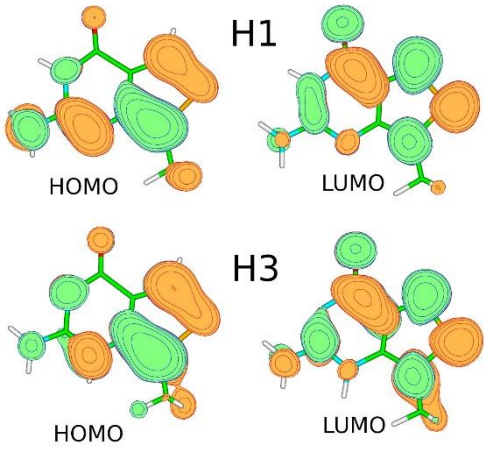

Figure S9. Frontier orbitals involved in the $S_0 \rightarrow S_1$ transition of the thG-H1 and thG-H3 tautomers.

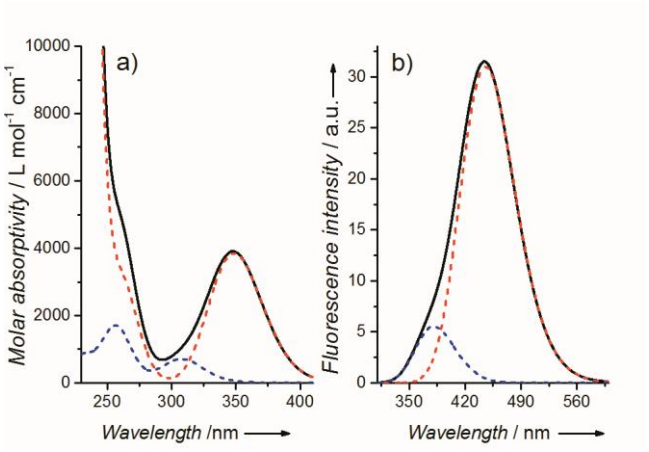

Figure S10. Absorption (a) and emission (b) spectra of the thG-H1 (red) and thG-H3 (blue) tautomers. The individual spectra and their sum weighted for their molar fraction were computed at PCM/TD-PBE0/6-311+G(2d,2p)/PCM/ PBE0/6-31G(d) level. The transitions have been convoluted by a phenomenological Gaussian with a HWHM of 0.25 eV.

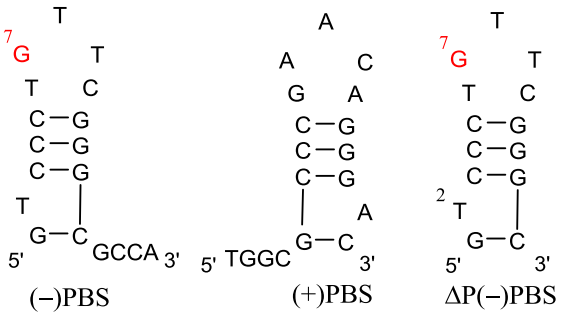

Figure S11. Structures of the primer binding site (PBS) oligonucleotides used in this study.

Table S5. Spectroscopic Parameters and Free Energy of Binding of dthG-containing Matched and Mismatched Duplexes [a]

Opposite base to dthG7	I_{375}/I_{550}	QY^{H1}	$\lambda_{em}^{max\ (H1)}$	ΔG^{H1}	ΔG^{H3}
dC	0.13 ± 0.02	0.20 [b]	461 ± 1	-71 ± 1	-58 ± 1
dT	0.14 ± 0.01	0.38 [b]	467 ± 1	-59 ± 1	-48 ± 1
dG	0.35 ± 0.01	-	462 ± 1	-54 ± 1	-53 ± 1
dA	0.65 ± 0.03	-	454 ± 1	-52.5 ± 0.8	-50 ± 1

[a]Fluorescence emission spectra of dthG-containing matched and mismatched duplexes were recorded at 310 nm excitation wavelength; I_{375}/I_{550} is the ratio of emission intensities at 375 and 550 nm; QY^{H1} is the fluorescence quantum yield of the dthG-H1 tautomer; $\lambda_{em}^{max(H1)}$ (in nm) is the fluorescence emission maxima of the dthG-H1 tautomer; $\Delta G^{H1/H3}$ is the free energy of binding (kcal/mol) of (+)PBS to (–)PBS where dG7 was substituted by dthG-H1 or dthG-H3 tautomer estimated using Molecular Mechanics Poisson-Boltzmann Surface Area approach along MD trajectories. The I_{375}/I_{550} and $\lambda_{em}^{max(H1)}$ values are the means from three experiments. [b] Values reported in [2].

4. Molecular Dynamics simulations

4.1 Molecular dynamics simulations using the AMBER force field

The family of ΔP(−)PBS DNA NMR structures[17] was provided by Nelly Morellet (personal communication). Two of these structures, namely structures #1 and #10, were selected as representative examples where the nucleobase at position 7 in the loop was in a π-π stacked or solvent-exposed conformation, respectively (Figure S12). The (−)/(+)PBS duplex structure was built with the Nucleic Acid Builder (NAB) molecular manipulation language, and the dG nucleotide at position 7 was manually replaced with d^{th}G-H1 or d^{th}G-H3.[18] For both d^{th}G tautomers, bond lengths and partial charges were obtained by full geometry optimization at the DFT level, using the B3LYP functional in conjunction with the 6-311++G(d,p) basis set. The remaining parameters were taken from the Amber ff12SB force field[19] which adds novel torsional parameters for backbone and side chain torsions to the ff99SB force field for proteins, and includes the torsional modifications already validated in the ff99bsc0 force field for nucleic acids. Each macromolecular system investigated by MD was solvated by a cubic box of TIP3P-typed water[20] buffering 8 Å from the molecular surface and the total charge was neutralized by the addition of Na^+ counter ions. Water molecules were energy minimized for 500 steps using the Steepest Descent algorithm (SD) and a further 1500 steps using the Conjugate Gradient algorithm (CG), while keeping the solute as fixed. Removing the constraints, the solvated solute was energy minimized for 1000 steps using the SD and 4000 steps using the CG before being heated at constant volume from 0 to 300 K over 100 ps and using the Langevin thermostat. A density equilibration was carried out at constant pressure (NPT ensemble) for 100 ps, before running the production of unbiased MD trajectories for 200 ns. The frame with the lowest Root Mean Square Deviation (RMSD) with respect to the average structure of each system was extracted and used for

graphical representations of (–)/(+)PBS DNA, while the representative structure of the most populated cluster of ΔP(–)PBS DNA trajectories was selected for graphical representation. The binding free energy between (–)PBS and (+)PBS DNA was estimated by means of the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) approach implemented in Amber12.[21] All MD simulations were performed with Amber12 on two Nvidia K20 GPUs.[19b]

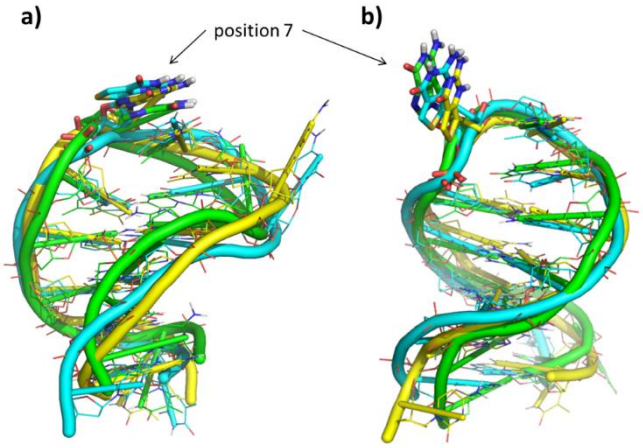

Figure S12. Superimposition of representative structures of the most populated clusters extrapolated from MD trajectories of ΔP(–)PBS single-stranded DNA. Representative structures from MD simulations of structure #1 (a) and structure #10 (b). The nucleotide at position 7 highlighted by a black arrow is shown as sticks. ΔP(–)PBS DNA bearing dG at position 7 is in green, that bearing d^{th}G-H1 is in cyan and that bearing d^{th}G-H3 is in yellow.

MD simulations of single-stranded ΔP(–)PBS DNA. The introduction of d^{th}G-H1 or -H3 at position 7 in the single-stranded loop of the ΔP(–)PBS does not affect the overall conformation of the oligomer (Figure S13a, b). The total energy of each ΔP(–)PBS system was nearly constant in both structures #1 and #10 during MD simulations, whereas structure #1 proved to be thermodynamically more stable than #10 (compare Figure S14a and b). Moreover, we checked whether the d^{th}G tautomers could influence the conformational

population of ΔP(–)PBS DNA. MD trajectories of the single-stranded loop of wild-type ΔP(–)PBS and ΔP(–)PBS bearing dthG-H1 or dthG-H3 at position 7 were clustered. Results clearly show that in all MD simulations one cluster of loop conformations is predominant, occurring between 45 and 73% of MD frames (Figure S12). Notably, the conformation of the nucleoside at position 7 is highly comparable in the representative structures of the predominant cluster of wild-type ΔP(–)PBS and the two dthG-bearing systems, thus suggesting that dG replacement with dthG-H1 or dthG-H3 in single stranded ΔP(–)PBS does not alter the conformational preferences of ΔP(–)PBS.

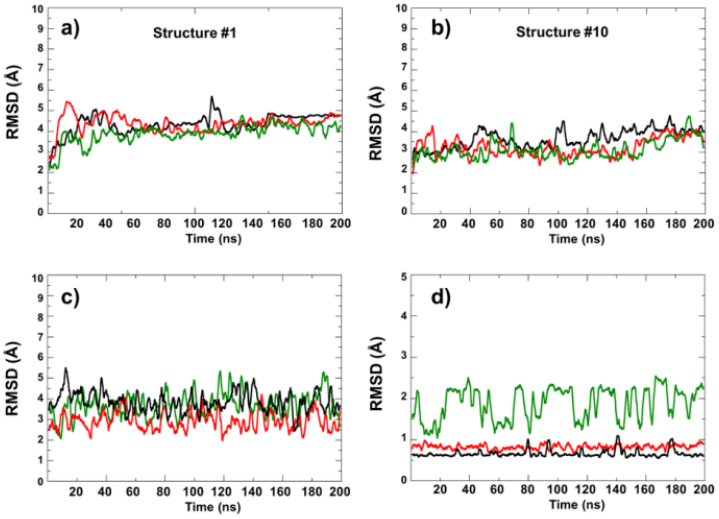

Figure S13. RMSD variation of (–)PBS and (–)/(+)PBS duplex. (a-b) RMSD variation of the single-stranded (–)PBS DNA structure #1 (a) and structure #10 (b) along 200 ns of unbiased MD trajectories. In black: unmodified (–)PBS having dG at position 7; in red: dthG7(–)PBS bearing dthG-H1 tautomer at position 7; in green: dthG7(–)PBS bearing dthG-H3 tautomer at position 7. (c-d) RMSD variation of the whole (–)/(+)PBS DNA duplex (c) and the dG (or dthG)-C Watson-Crick base pair at position 7 (d) along 200 ns of unbiased MD trajectories. Color codes for c-d are the same as for a-b.

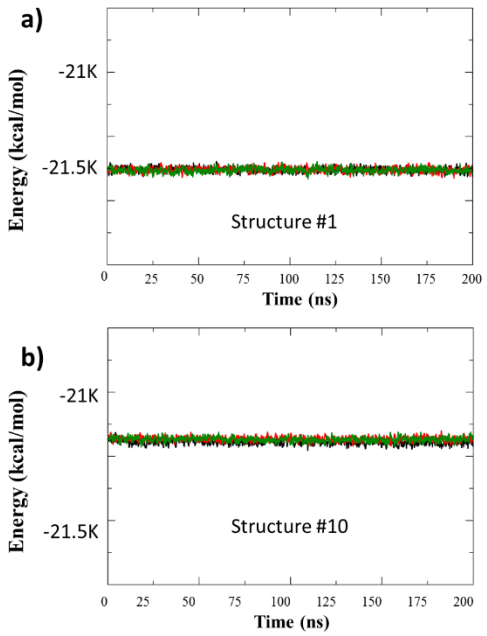

Figure S14. Total energy of the NMR-based structure #1 (a) and structure #10 (b) of ΔP(–)PBS DNA along 200 ns of unbiased MD simulation, bearing dG (black line), d^{th}G-H1 tautomer (red line) or d^{th}G-H3 tautomer (green line) at position 7.

MD simulations of (–)/(+)PBS DNA duplex. In the (–)/(+)PBS DNA duplex, the dG base at position 7 was kept unchanged or was replaced with d^{th}G-H1 or d^{th}G-H3, and each duplex was investigated by means of 200 ns of unbiased MD simulations. Analysis of RMSD along each MD trajectory unequivocally shows that replacing dG with d^{th}G-H1 or d^{th}G-H3 does not impact the overall geometry of the DNA duplex (Figure S13c). To further investigate the effects of replacing dG with d^{th}G on the stability of the duplex, the free energy of binding of (–)PBS to (+)PBS in the duplex was estimated using MM-PBSA approach along MD trajectories. In line with the conformational findings described above and in comparison with the wild type (–)/(+)PBS (ΔG = –71 ± 1 kcal mol^{-1}), the introduction of d^{th}G-H1 does not affect the stability of the DNA duplex (ΔG = –71.3 ± 0.8 kcal mol^{-1}), whereas the replacement of dG at position 7 with d^{th}G-H3 induces a significant change (ΔG = –57.9 ± 0.8 kcal mol^{-1}).

This thermodynamic effect is probably due to the inability of dthG-H3 to establish canonical H-bond interactions with the opposite dC nucleobase in the duplex.

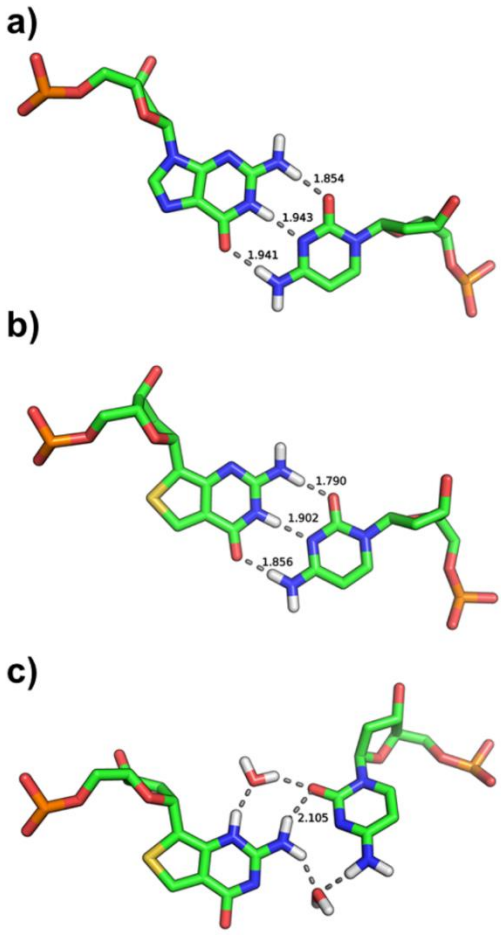

Figure S15. Structure of the base pair involving dG (a), dthG-H1 tautomer (b), and dthG-H3 tautomer (c) in the (–)/(+)PBS DNA duplex. Frames corresponding to the most representative structures of unbiased MD trajectories are shown. H-bond interactions are shown as grey dashed lines. Distances for direct and water-bridged H-bonds between bases are reported (in Å).

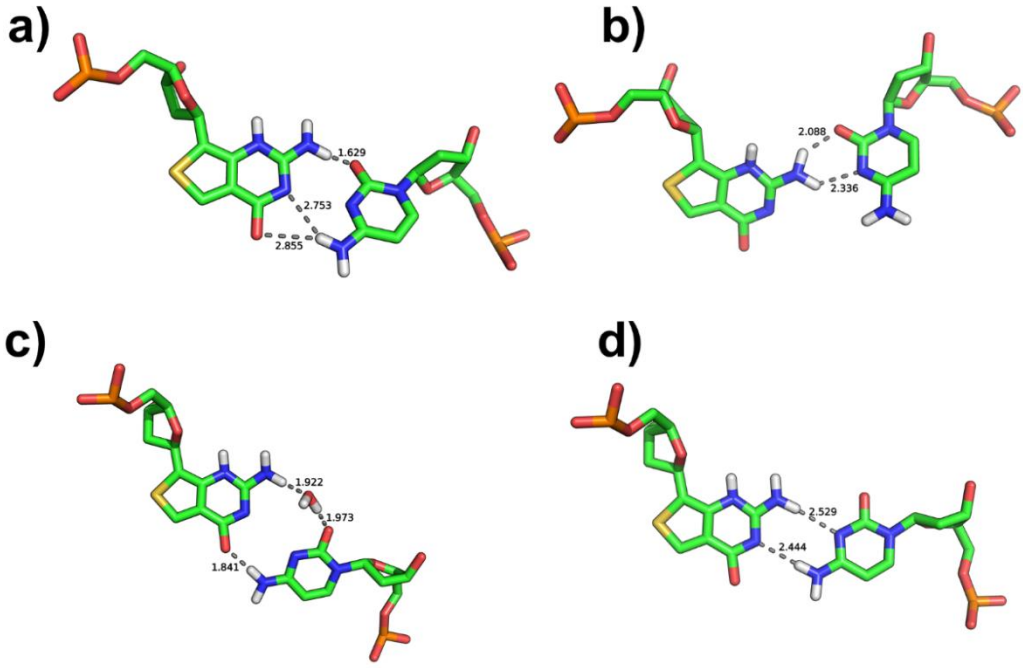

Figure S16. Four clusters of the dthG-H3 tautomer in various base-pair conformations with the opposite dC nucleobase. The most populated and stable cluster is shown in Figure S15c. H-bond interactions are showed as grey dashed lines. The distances of direct and water-bridged H-bonds between dthG and dC are shown.

Syn-anti conformation of the two dthG tautomers

To investigate the conformation adopted by the two dthG tautomers, the dihedral angle at the glycosidic bond was measured in MD simulations for the nucleotide at position 7 (dG, dthG-H1, or dthG-H3) of the (–)/(+)PBS duplex and the single stranded (–)PBS. Dihedral angle values comprised between -90° and +90° were assigned to the *anti* conformation of the nucleotide, whereas values comprised between -180° and -90° or +90° and +180° were assigned to the *syn* conformation. Results of this analysis (Table S6) unequivocally show that the preferential conformation of these nucleotides is the *anti*, particularly for dG, and dthG-H1 in both (–)/(+)PBS and (–)PBS systems. For dthG-H3, a slightly lower abundance of the *anti* nucleotide was observed in the (–)/(+)PBS duplex, whereas a more important variation was

observed in single stranded (–)PBS. In (–)PBS, the NH at position 3 of d^{th}G-H3 can establish H-bond interactions with the phosphate group, which likely enhances the abundance of the *syn* conformation with respect to dG, and d^{th}G-H1.

Table S6. Percentage of the syn or anti conformation of dG, d^{th}G-H1, or d^{th}G-H3 in MD simulations of (–)/(+)PBS and (–)PBS

	(–)/(+)PBS		(–)PBS	
	anti (%)	*syn* (%)	*anti* (%)	*syn* (%)
dG	92.5	7.5	89.4	10.6
d^{th}G-H1	94.3	5.7	91.3	8.7
d^{th}G-H3	87.5	12.5	63.2	36.8

Thus, our data strongly suggest that the *anti* conformation is preferred for d^{th}G-H1 and d^{th}G-H3 in the systems investigated in this work.

4.2 Molecular dynamics simulations using the CHARMM force field

Structure Preparation: To build the initial structures of the wild-type ΔP(–)PBS and ΔP(–)PBS bearing d^{th}G-H1 or d^{th}G-H3 at position 7, common heavy atom positions were retained from the experimental structures and the remaining heavy atoms were placed using the tools present in the CHARMM program.[22] The topology and parameters for the d^{th}G tautomers were constructed based on similarity to groups and parameters existing in the CHARMM27 all-atom nucleic acid force field.[23] An initial energy minimization consisting of 1000 steps using the Steepest Descent method followed by 1000 steps of the Adapted Basis Newton-Raphson minimization method was realized in order to eliminate strong steric contacts prior to system solvation.

All-atom MD simulations set-up: Molecular dynamics simulations of the nucleic acids were done using the all-atom force field CHARMM27.[23] The system preparation and the analysis were done using the CHARMM program, while the simulations themselves were done using the NAMD program.[22] Periodic boundary conditions were used and the long-range electrostatic interactions were treated with the Particle Mesh Ewald (PME) algorithm.[24] All hydrogen-covalent bonds were constrained using the SHAKE algorithm[25] and an atom-based switching function with a cutoff of 12 Å was applied to the van der Waals non-bonded interactions. A 2 fs integration time step was used for all simulations. The water molecules were initially relaxed around a harmonically-constrained DNA by 5000 steps of Conjugate Gradient (CG) energy minimization, followed by 10000 steps without constraints. Subsequently, a molecular dynamics-based heating to 300 K over the course of 600 ps with the DNA harmonically constrained was done, followed by an equilibration phase, where the harmonic constraints were gradually removed over the course of 6 ns (the constraints were reduced every 500 ps). Pressure control was introduced during equilibration using a Berendsen piston[26] with a relaxation time of 400 fs and a rescaling of the atomic positions every 4 fs. The temperature was maintained using Langevin dynamics with a damping coefficient of 1 ps^{-1} applied to each atom. Finally, a 50-ns production simulation was performed in the isothermal–isobaric (NPT) ensemble at 1 bar and 300 K.

Simulations of the wild-type ΔP(–)PBS, as well as the d^{th}G-H1- and d^{th}G-H3-labelled ΔP(–)PBS, were run for 50 ns. The analysis included the calculation of the RMSD in order to quantify the structural behavior of the different tautomers. The same protocol was applied to the (–)/(+)PBS DNA duplex constructed with the same tautomers.

ΔP(–)PBS Simulation Results: The overall structure of the ΔP(–)PBS was reoriented over the backbone of bases 5-9 and the RMSD calculations of the base 7, which was dG, d^{th}G-H1 or d^{th}G-H3 tautomer (excluding the main chain), were performed. Average RMSDs at

the end of the simulations were: wild-type = 0.9 Å, d^{th}G-H1=0.9 Å and d^{th}G-H3=1.0 Å. The RMSD values show for the three constructs that there is no significant preference for the conformation of the base and that both tautomers can be present in the single-stranded ΔP(−)PBS.

(−)/(+)PBS *Simulation Results:* RMSD calculations were done in order to assess the structural deformation introduced by the substitution of dG7 by either d^{th}G-H1 or d^{th}G-H3 tautomers. From the simulations, the RMSDs were calculated in the same manner that is superposition over bases 5–9 and the calculation of the RMSD of base 7. An average RMSD was calculated over the final segment of the simulation yielding the following values for the three constructs: 0.5 Å for the wild-type, 0.6 Å for the d^{th}G-H1 and 0.7 Å for the d^{th}G-H3 tautomer in the final part of the trajectory.

While all the RMSDs are relatively small, the d^{th}G-H3 tautomer shows the largest deviation from the wild-type, which is qualitatively consistent with what was observed in the simulations done using the Amber force field. While the results do not contradict each other, the structural perturbations observed in the Amber force field simulations are somewhat larger, likely due to differences in the force fields.

When examining average energy minimized structures obtained from the end of the simulations, it was interesting to note that the d^{th}G-H3 took on a non-canonical conformation in its base pairing with dC similar to that of Figure S16d, in contrast to dG and d^{th}G-H1, which displayed a canonical H-bond network.

Further analysis was carried out using an MM-PBSA approach for free energy decomposition. We employed a protocol based on the MM-PBSA method described by Lafont et al,[27] which has proven to yield an accurate description of binding energetics. We looked at the by-base contribution to the binding energy. Individual contributions of each base to the complex formation were estimated. From the analysis, the wild-type dG7 and d^{th}G-H1

tautomer make larger contributions to binding, −3.8 ± 1.5 and −3.5 ± 0.7 kcal mol^{-1}, respectively, than the d^{th}G-H3 tautomer which contributed −1.1 ± 0.6 kcal mol^{-1}. This suggests that the alternative d^{th}G-H3 tautomer contributes less to the total binding free energy, in full line with the results presented in the main text, where the AMBER force field was used.

The results presented here confirm that the d^{th}G-H1 tautomer mimics very well the dG base in both single-stranded and double-stranded DNA. The d^{th}G-H3 tautomer does not show any distinct structural dynamics characteristic in the single-stranded DNA, but exhibits a non-canonical hydrogen bonding pattern in its base pairing with dC, so that it contributes less to the overall free energy of duplex formation than either the wild-type dG or the d^{th}G-H1 tautomer.

References:

[1] D. Shin, R. W. Sinkeldam, Y. Tor, *J. Am. Chem. Soc.* **2011**, *133*, 14912.

[2] M. Sholokh, R. Sharma, D. Shin, R. Das, O. A. Zaporozhets, Y. Tor, Y. Mely, *J. Am. Chem. Soc.* **2015**, *137*, 3185.

[3] H. de Rocquigny, D. Ficheux, C. Gabus, M. C. Fournie-Zaluski, J. L. Darlix, B. P. Roques, *Biochem. Biophys. Res. Commun.* **1991**, *180*, 1010.

[4] C. Reichardt, *Chem. Rev.* **1994**, *94*, 2319.

[5] M. J. Kamlet, J. L. M. Abboud, M. H. Abraham, R. W. Taft, *J. Org. Chem.* **1983**, *48*, 2877.

[6] J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*, 3rd ed., Springer, New York, **2006**.

[7] a) C. Adamo, G. E. Scuseria, V. Barone, *J. Chem. Phys.* **1999**, *111*, 2889; b) M. Ernzerhof, G. E. Scuseria, *J. Chem. Phys.* **1999**, *110*, 5029; c) C. Adamo, V. Barone, *J. Chem. Phys.* **1999**, *110*, 6158.

[8] a) Y. Zhao, N. E. Schultz, D. G. Truhlar, *J. Chem. Theory Comput.* **2006**, *2*, 364; b) Y. Zhao, D. G. Truhlar, *Acc. Chem. Res.* **2008**, *41*, 157.

[9] a) V. Karunakaran, K. Kleinermanns, R. Improta, S. A. Kovalenko, *J. Am. Chem. Soc.* **2009**, *131*, 5839; b) Y. Y. Zhang, R. Improta, B. Kohler, *Phys. Chem. Chem. Phys.* **2014**, *16*, 1487.

[10] R. Improta, V. Barone, *Top. Curr. Chem.* **2015**, *355*, 329.

[11] a) A. Dreuw, M. Head-Gordon, *Chem. Rev.* **2005**, *105*, 4009; b) R. Improta, *Phys. Chem. Chem. Phys.* **2008**, *10*, 2656.

[12] a) F. Santoro, V. Barone, R. Improta, *J. Am. Chem. Soc.* **2009**, *131*, 15232; b) R. Improta, V. Barone, *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 12016; *Angew. Chem*. **2011**, *123*, 12222.

[13] a) S. Miertus, E. Scrocco, J. Tomasi, *Chem. Phys.* **1981**, *55*, 117; b) J. Tomasi, B. Mennucci, R. Cammi, *Chem. Rev.* **2005**, *105*, 2999.

[14] G. Scalmani, M. J. Frisch, B. Mennucci, J. Tomasi, R. Cammi, V. Barone, *J. Chem. Phys.* **2006**, *124*.

[15] D. R. Roe, T. E. Cheatham, 3rd, *J. Chem. Theory Comput.* **2013**, *9*, 3084.

[16] a) M. Kabelac, P. Hobza, *Phys. Chem. Chem. Phys.* **2007**, *9*, 903; b) M. Hanus, F. Ryjacek, M. Kabelac, T. Kubar, T. V. Bogdan, S. A. Trygubenko, P. Hobza, *J. Am. Chem. Soc.* **2003**, *125*, 7678; c) C. Colominas, F. J. Luque, M. Orozco, *J. Am. Chem. Soc.* **1996**, *118*, 6811; d) J. R. Blas, F. J. Luque, M. Orozco, *J. Am. Chem. Soc.* **2004**, *126*, 154.

[17] S. Bourbigot, N. Ramalanjaona, C. Boudier, G. F. J. Salgado, B. P. Roques, Y. Mély, S. Bouaziz, N. Morellet, *J. Mol. Biol.* **2008**, *383*, 1112.

[18] T. J. Macke, D. A. Case in *Molecular Modeling of Nucleic Acids*, (Eds.: N. B. Leontes, J. Santa Lucia Jr.), American Chemical Society, Washington DC, **1998**, pp. 379-393.

[19] a) V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins* **2006**, *65*, 712; b) D.A. Case, T. A. Darden, T.E. Cheatham, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Goetz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P. A. Kollman, AMBER12, University of California, San Francisco, **2012**.

[20] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926.

[21] a) M. Mori, U. Dietrich, F. Manetti, M. Botta, *J. Chem. Inf. Model.* **2010**, *50*, 638; b) B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke, A. E. Roitberg, *J. Chem. Theory Comput.* **2012**, *8*, 3314.

[22] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, *J. Comput. Chem.* **2009**, *30*, 1545.

[23] N. Foloppe, A. D. MacKerell, *J. Comput. Chem.* **2000**, *21*, 86.

[24] H. G. Petersen, *J. Chem. Phys.* **1995**, *103*, 3668.

[25] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327.

[26] H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684.

[27] V. Lafont, M. Schaefer, R. H. Stote, D. Altschuh, A. Dejaegere, *Proteins* **2007**, *67*, 418.