

Design and photophysical characterization of quasi-intrinsic fluorescent probes utilized in DNA sequencing

Yaning Zhang(张雅宁), Yongkang Lyu(吕永康), Zhizheng Cao(曹智正), Xiaolin Chen(陈晓琳), Qingtian Meng(孟庆田)†, and Changzhe Zhang(张常哲)‡

Citation : Chin. Phys. B, 2025, 34 (**2**): 023203. DOI: 10.1088/1674-1056/ad9a9a
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

L-shell x-ray fluorescence relative intensities for elements with 62 ≤ *Z* ≤ 83 at 18 keV and 23 keV by synchrotron radiation

M Alqadi, S AL-Humaidi, H Alkhateeb, and F Alzoubi
Chin. Phys. B, 2023, 32 (**8**): 083201. DOI: 10.1088/1674-1056/accb46

Highly sensitive detection of Rydberg atoms with fluorescence loss spectrum in cold atoms

Xuerong Shi(师雪荣), Hao Zhang(张好), Mingyong Jing(景明勇), Linjie Zhang(张临杰), Liantuan Xiao(肖连团), Suotang Jia(贾锁堂)
Chin. Phys. B, 2020, 29 (**1**): 013201. DOI: 10.1088/1674-1056/ab593b

Effect of chloride introduction on the optical properties in Eu^{3+}-doped fluorozirconate glasses

Han Jin(金含), Zhao-Jun Mo(莫兆军), Xiao-Song Zhang(张晓松), Lin-Lin Yuan(苑琳琳), Ming Yan(晏明), Lan Li(李岚)
Chin. Phys. B, 2016, 25 (**10**): 103201. DOI: 10.1088/1674-1056/25/10/103201

Two-photon spectrum of ^{87}Rb using optical frequency comb

Wang Li-Rong (汪丽蓉), Zhang Yi-Chi (张一驰), Xiang Shao-Shan (向少山), Cao Shu-Kai (曹书凯), Xiao Lian-Tuan (肖连团), Jia Suo-Tang (贾锁堂)
Chin. Phys. B, 2015, 24 (**6**): 063201. DOI: 10.1088/1674-1056/24/6/063201

Theoretical study of amplified spontaneous emission intensity and bandwidth reduction in polymer

A. Hariri, S. Sarikhani
Chin. Phys. B, 2015, 24 (**4**): 043201. DOI: 10.1088/1674-1056/24/4/043201

Design and photophysical characterization of quasi-intrinsic fluorescent probes utilized in DNA sequencing

Yaning Zhang(张雅宁), Yongkang Lyu(吕永康), Zhizheng Cao(曹智正), Xiaolin Chen(陈晓琳), Qingtian Meng(孟庆田)†, and Changzhe Zhang(张常哲)‡

School of Physics and Electronics, Shandong Normal University, Jinan 250358, *China*

(Received 19 September 2024; revised manuscript received 20 November 2024; accepted manuscript online 5 December 2024)

To understand the gene-based biological processes in-depth, the single-molecule real-time sequencing has drawn increasing attention with promoted by the Human Genome Project. Herein, a set of newly designed canonical fluorescent bases (A_y, tC, G_b, T_p) are proposed for four-color DNA sequencing. These quasi-intrinsic probes are derived from the fluorophore replacement and ring expansion on natural bases, which still keep the pyrimidine or purine underlying skeleton and Watson–Crick hydrogen bonding face to allow minimal perturbation to the native DNA duplex. More importantly, these nucleobase analogues possess red-shifted absorption and efficient photoluminescence due to the enhanced π-conjugation in character. Meanwhile, the four analogues could generate distinct emission wavelength ($\Delta\lambda \sim 50$ nm) for real-time sequencing. To assess the biological employment of the proposed biosensors, the effects of base pairing and linking deoxyribose are also considered.

Keywords: DNA sequencing, base analogues, fluorescent probe

PACS: 32.50.+d, 31.15.E–, 33.50.Dq, 34.70.+e

DOI: 10.1088/1674-1056/ad9a9a **CSTR:** 32038.14.CPB.ad9a9a

1. Introduction

Deoxyribonucleic acid (DNA) is one of the most vital biopolymers in the midst of living organisms. The arrangements of the four nucleobases (A, T, C, G) pairing in DNA double helix are the essence of encoding, transmitting and expressing genetic information.[1,2] To understand the gene based biological processes in-depth, the DNA sequencing technology has been greatly promoted by the Human Genome Project.[3–8] The first generation sequencing (FGS) is performed on microfabricated capillary electrophoresis chips and the sequencing fragment ladders were fluorescently labeled by utilizing the energy transfer dye-labeled primers.[9] In next-generation sequencing (NGS) methods, four nucleotides (A, C, G, T) are produced as invertible terminators by linking a cleavable fluorophore as substrates.[10] The DNA templates composing of homopolymer regions could be precisely sequenced on a DNA chip by using four-color fluorescent scanner.[11] Recently, the single-molecule real-time sequencing (SMRT) become the available third generation sequencing (TGS) technology, which employs four labeled nucleotides to generate distinct emission spectra during the sequencing process.[12–15] In general, the SMRT is realized upon two-base pattern with considering the fact that single-stranded DNA template to be sequenced could combined with artificial bases to form DNA double-strand with help of DNA polymerase immobilized at the bottom of zero-mode waveguide (ZMW).[16,17] The ZMW is narrower than the excitation wavelength, which makes it to allow the selective photo-excitation only of the incorporated fluorescent nucleotides and provide distinct time trace of four-color fluorescence (see Fig. S1 in supporting information). It should be noticed that, however, the above extrinsic fluorescent dyes could disturb the native structure of DNA and bring poor incorporation in consecutive positions by DNA polymerase due to their large steric hindrance during DNA synthesis phase.[18] Hence, there is a necessity to develop intrinsic fluorescence labelling strategy for real-time DNA sequencing with minimal disturbance with the native structure.

In recent years, the quasi-intrinsic building blocks have achieved increasing attention due to their geometric similarity with natural counterparts and accurate self-assembles.[19–25] For example, Kool *et al.*[26] reported the modified DNA oligomers containing fluorogenic nucleobases to measure the oxidative 8-oxoguanine by utilizing itself as a fluorescence quencher. Moreover, Tan and co-workers reported a new photoresponsive T-analogue (zT) that could provide versatility required for constructing desired DNA nanomolecules and nanodevices.[27] Recently, the artificial ferrocene-base (Fe-base) was synthesized for monitoring single-nucleobase changes and charge transfer rates in DNA.[28] Additionally, the fluorescent thymidine analogue (^{DMA}T) is designed to study the site-specific metal binding reactions in duplex DNA.[29] In 2023, the modified C-analogues are proposed as optical

†Corresponding author. E-mail: qtmeng@sdnu.edu.cn

‡Corresponding author. E-mail: zhe852456@126.com

© 2025 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

probes to detect the carcinogenic DNA adducts (^{ABP}dG) based on the selective base pairing induced excited-state intermolecular charge transfer (ESICT) mechanism.[30]

Inspired by the advantages of the artificial nucleobases, we systemically investigated the spectral characters of newly designed canonical fluorescent bases for four-color DNA sequencing in solution. Herein, two kinds of modifications were taken into account including the substitute for fluorophore replacement[31] and ring expansion.[32,33] Compared to the natural bases, these nucleobase analogues keep the pyrimidine or purine underlying skeleton and Watson–Crick (WC) hydrogen bonding face to allow minimal perturbation to the native DNA duplex.[32] Meanwhile, the modified bases possess red-shifted absorption and high-efficiency fluorescence due to the enhanced π-conjugation. More importantly, the four analogues (A_y, tC, G_b, T_p) could generate distinct emission wavelength ($\Delta\lambda \sim 50$ nm) for real-time sequencing. To evaluate the employment of the biosensors in biological environment, the effects of base pairing and linking deoxyribose are also considered. This work puts forward a class of quasi-intrinsic candidates that could be recognized by four-color optical scanner and DNA polymerase as substrates, which provides new understandings into the design of probes for SMRT.

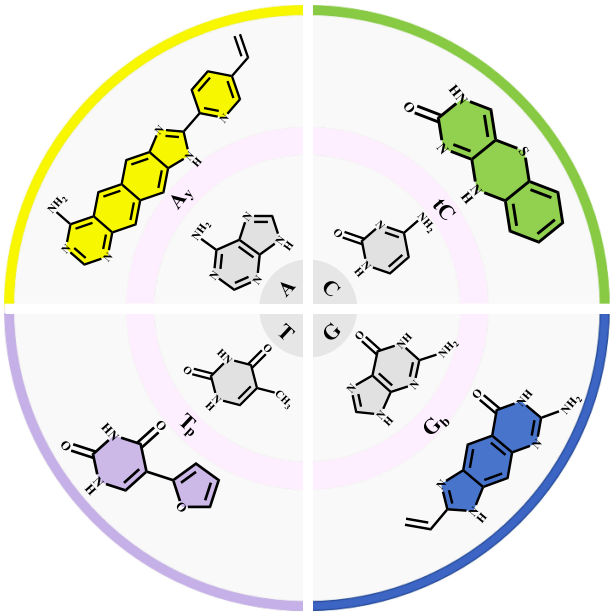


Fig. 1. The schematic structures of A_y, tC, G_b, T_p.

2. Theoretical methods

Gaussian 16 software[34] was used for the density functional theory (DFT) and the time-dependent density functional theory (TDDFT) calculations. The related frontier molecular orbitals (FMOs) were visualized by the VMD program[35] and Multiwfn package.[36] Herein, the molecular optimizations of all analogues as well as Watson–Crick base pairs in the ground state (S_0) and the first excited state (S_1) were obtained by using the 6-311++G(d,p) basis set[37,38] with the B3LYP-D3 function.[39–41] Meanwhile, the polarized continuum model with the integral equation formalism (IEFPCM)[42–44] is used to account for the solvent effect (water). To confirm the optimized structures under solvation equilibrium are the minima of the potential energy surface (PES), the vibrational frequency analyses[45] are also employed. On the basis of the optimized configurations in S_0 and S_1 states, the optical spectra were performed with non-equilibrium solvation and broadened through the following Gaussian function:

$$G(w) = \frac{1}{c\sqrt{2\pi}} \exp\left[-\frac{(w-w_n)^2}{2c^2}\right], \tag{1}$$

where the full width at half maximum (FWHM) is specified to be 0.4 eV and c denotes FWHM/$2\sqrt{2\ln 2}$. The w and w_n were used in this equation on behalf of the abscissa of the spectrum and the excitation energy corresponding to the transition from S_1 to S_n. The above structural and optical results were also examined by utilizing the PBE0D3 functional[46–50] that has been shown to valuably predict the spectra properties of similar systems.[51]

3. Results and discussion

Herein, the four quasi-intrinsic fluorescent probes are proposed for application in four-color sequencing. The utilized fluorescent cytosine-analogue is tC, which has been synthesized experimentally based on the sodium salt glycosylation procedure.[52,53] For comparison, the absorption and fluorescence wavelength of tC are predicted at 362 nm and 488 nm, which agree with experimental observations ($\lambda_{absorption} = 375$ nm, $\lambda_{fluorescence} = 505$ nm[33]). Besides, as depicted in Fig. 2, three newly designed analogues (A_y, G_b, T_p) are constructed from the fluorophore replacement and ring expansion that has been widely utilized in modifications of nucleobases.[19,54] These surrogates resemble one of the natural nucleobases in terms of structural ability to WC base pairs without serious perturbation to the native DNA double helix. To demonstrate their fluorescent recognition of sequencing information, the photophysical characterization of all considered monomers and the effect of base pairing and linking deoxyribose are comprehensively investigated in the following discussion.

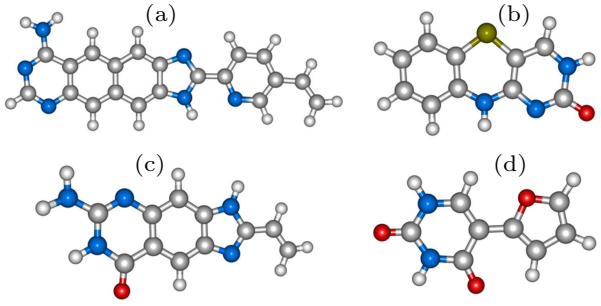


Fig. 2. The geometry structures of A_y (a), tC (b), G_b (c), and T_p (d).

3.1. Photophysical properties of base monomers

The geometries of all considered analogues are optimized adequately with confirmation by harmonic vibrational frequency analyses. The selected main structural parameters are listed in Fig. S3. It is shown that these surrogates (A_y, T_p, tC, G_b) could keep co-planar characteristic, WC hydrogen bonding faces and has negligible changes ($\Delta d_{Ay} < 0.04$ Å, $\Delta d_{Tp} < 0.05$ Å, $\Delta d_{tC} < 0.02$ Å, $\Delta d_{Gb} < 0.04$ Å) due to the modifications, in which Δd stands for the maximum difference of bond lengths between natural bases and modified monomers (Fig. S3). Clearly, the artificial bases are close to natural bases with considering the size and shape, which could ensure less interference with the helical structure and function of nucleic acids.

On the basis of the optimized geometries in S_0 and S_1 states, the spectra properties of the four analogues are presented in Fig. 3. It can be seen that the maximum absorption wavelength of tC is predicted at 362 nm ($f_{abs.} = 0.07$) corresponding to the $S_0 \rightarrow S_1$ transition (HOMO→LUMO, CI = 95.6%), which agree with the experimental observations (375 nm).[33] The absorption of G_b in water is characterized by a structured spectra with a sharp peak at 281 nm ($f_{abs.} = 0.76$) and a weaker one around 362 nm ($f_{abs.} = 0.33$), which are dominated by the transitions of $S_0 \rightarrow S_3$ (HOMO-1→LUMO, CI = 55.9% and HOMO→LUMO+1, CI = 36.3%), and $S_0 \rightarrow S_1$ (HOMO→LUMO, CI = 96.8%). Similarly, the absorption spectra of T_p exhibit two distinct absorption peaks around 264 nm ($S_0 \rightarrow S_2$, $f_{abs.} = 0.44$) and 328 nm ($S_0 \rightarrow S_1$, $f_{abs.} = 0.28$) related to HOMO→LUMO+1 (CI = 97.3%) and HOMO→LUMO (CI = 97.5%), respectively. While, for the A_y, a strong absorption peak centred at 365 nm ($f_{abs.} = 1.00$) can be primarily ascribed to the $S_0 \rightarrow S_3$ (HOMO-1→LUMO, CI = 59.9% and HOMO→LUMO+1, CI = 33.4%). Compared with the case of natural bases (A: 251 nm, T: 265 nm, C: 257 nm, G: 243 nm),[55] the absorption wavelengths of all considered base analogues (A_y, T_p, tC, G_b) are obviously red-shifted due to the narrowed energy gaps between involved orbitals (Fig. 4). Clearly, these analogues are easy to allow for selective photo-excitation resolved from the natural nucleobases.

Following the vertical photo-excitation, the electronic emission spectra at the fully relaxed S_1-configuration are performed in Fig. 3(b). For free tC, the calculated fluorescence peak is observed at 488 nm, which is in agreement with the experimentally determined ones (505 nm[33]). Meanwhile, the three newly designed analogues (A_y, G_b, T_p) are found to possess enhanced photoluminescence in order of decreasing wavelength as A_y (576 nm, $f_{flu.} = 0.45$) > tC (488 nm, $f_{flu.} = 0.12$) > G_b (450 nm, $f_{flu.} = 0.60$) > T_p (406 nm, $f_{flu.} = 0.37$). As given in Table 1, the fluorescence of the four analogues (A_y, tC, G_b, T_p) are dominated by LUMO→HOMO (CI = 99.8%), LUMO→HOMO (CI = 98.8%), LUMO→HOMO (CI = 99.4%), and LUMO→HOMO (CI = 99.4%) transitions, respectively. The inspection of Fig. 4(b) reveals that the S_1 states are $\pi^* \rightarrow \pi$ in character that allows for bright emission, and the above distinct emission wavelength could be applied for real-time sequencing.

In addition, the radiative lifetimes are examined by employing the Einstein transition probability formula as follows:[56,57]

$$\tau = \frac{3}{2 f_{flu.} \nu^2}, \tag{2}$$

in which τ, $f_{flu.}$, and ν stand for radiative lifetime, fluorescent oscillator strength, and emission energy, respectively. As we known, the short radiative lifetime facilitates the high light-emitting efficiency, while the long lifetime makes for the energy and electron transfer. Our calculations reveal that the radiative lifetime are respectively 11.06 ns (A_y), 28.56 ns (tC), 5.09 ns (G_b), and 6.60 ns (T_p) that accord with the measurable scope of organic fluorophore ($\tau < 30$ ns).[58] Especially, compared with the previous proposed fluorescent analogues,[59,60] the shorter lifetimes of newly designed bright nucleobases (A_y, G_b, T_p) are conductive to their application in the four-color sequencing.

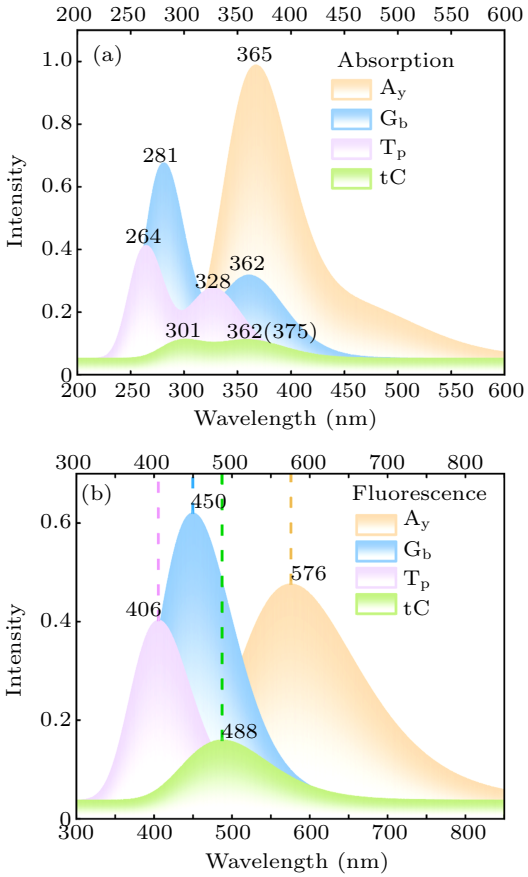


Fig. 3. Absorption (a) and fluorescence emission (b) spectra of the four analogues. The experiment value is shown in parentheses.[33]

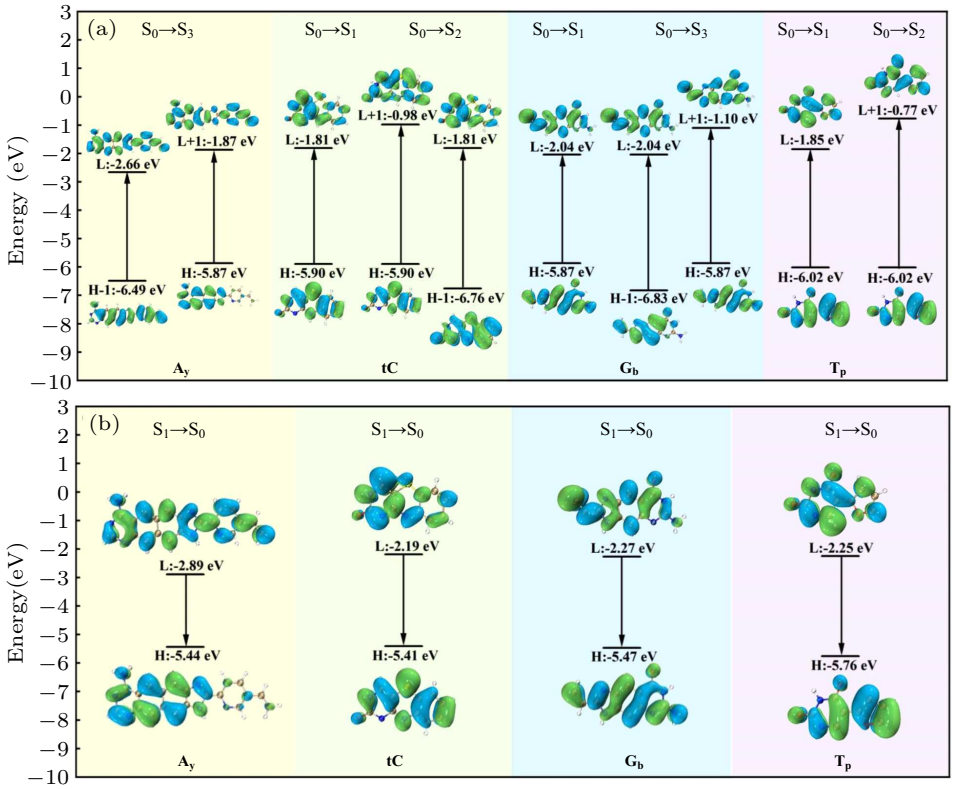


Fig. 4. FMOs (isovalue = 0.02 a.u.) involved in the excitation transitions (a) and the decay processes (b).

Table 1. Wavelengths (λ, in unit nm), oscillator strengths ($f_{abs.}$ and $f_{flu.}$), state assignments for monomers (A_y, tC, G_b, T_p) of absorption processes. The CI represents the abbreviation of composition index of the orbitals of the corresponding transitions. The gap and type denote the energy gaps and characters involved in orbital transitions, respectively.

Optical spectra	Monomer	Transition	λ	$f_{abs.}$	Assignment (CI/Gap (eV)/Type)
Absorption	A_y	$S_0 \rightarrow S_1$	471.11	0.1816	H→L(98.9%/3.00/$\pi \rightarrow \pi^*$)
		$S_0 \rightarrow S_2$	376.11	0.1625	H→L+1(64.2%/3.79/$\pi \rightarrow \pi^*$)
		$S_0 \rightarrow S_3$	365.44	0.9974	H-1→L(59.9%/3.83/$\pi \rightarrow \pi^*$) H→L+1(33.4%/3.79/$\pi \rightarrow \pi^*$)
	tC	$S_0 \rightarrow S_1$	362.46	0.0702	H→L(95.6%/4.09/$\pi \rightarrow \pi^*$)
		$S_0 \rightarrow S_2$	301.06	0.0538	H→L+1(84.6%/4.92/$\pi \rightarrow \pi^*$)
		$S_0 \rightarrow S_3$	292.92	0.0193	H→L+2(64.7%/4.98/$\pi \rightarrow \pi^*$) H-1→L(22.6%/4.95/$\pi \rightarrow \pi^*$)
	G_b	$S_0 \rightarrow S_1$	361.58	0.3255	H→L(96.8%/3.83/$\pi \rightarrow \pi^*$)
		$S_0 \rightarrow S_2$	303.75	0.0142	H→L+1(59.8%/4.77/$\pi \rightarrow \pi^*$) H-1→L(37.6%/4.79/$\pi \rightarrow \pi^*$)
		$S_0 \rightarrow S_3$	280.76	0.7637	H-1→L(55.9%/4.79/$\pi \rightarrow \pi^*$) H→L+1(36.3%/4.77/$\pi \rightarrow \pi^*$)
	T_p	$S_0 \rightarrow S_1$	327.93	0.2757	H→L(97.5%/4.17/$\pi \rightarrow \pi^*$)
		$S_0 \rightarrow S_2$	264.21	0.4441	H→L+1(97.3%/5.25/$\pi \rightarrow \pi^*$)
		$S_0 \rightarrow S_3$	253.10	0.0000	H-1→L(88.9%/5.24/$\pi \rightarrow \pi^*$)
Fluorescence	A_y	$S_1 \rightarrow S_0$	576	0.45	L→H(99.8%/2.55/$\pi \rightarrow \pi^*$)
	tC	$S_1 \rightarrow S_0$	488(505)	0.12	L→H(98.8%/3.22/$\pi \rightarrow \pi^*$)
	G_b	$S_1 \rightarrow S_0$	450	0.60	L→H(99.4%/3.20/$\pi \rightarrow \pi^*$)
	T_p	$S_1 \rightarrow S_0$	406	0.37	L→H(99.4%/3.51/$\pi \rightarrow \pi^*$)

3.2. Structural and spectra analyses of base pairs

Considering the requirement of SMRT with two-base sequence pattern, the effects of base pairing are further investigated. The base pairs including modified analogues are shown to maintain the co-planar characteristic in geometry and therefore enable minimal perturbation to the canonical DNA (Fig. S5). To examine the structural stability of modified analogues paired with their complementary bases, the interaction energies are calculated with BSSE correction as shown in the following equation:[61]

$$\Delta E = E_{AB} - (E_A + E_B) + E_{BSSE}, \quad (3)$$

where ΔE denotes the energy difference between the complex (E_{AB}) and the individuals (E_A and E_B). The calculated interaction energy of the four analogues is respectively -11.32 kcal/mol (A_y:T), -17.82 kcal/mol (tC:G), -16.66 kcal/mol (G_b:C), and -11.19 kcal/mol (T_p:A), which is comparable with native base pairs (-11.01 kcal/mol $\sim$ -17.19 kcal/mol). Moreover, the strength of intermolecular interactions is visualized through the reduced density gradient (RDG) versus $\text{sign}(\lambda_2)\rho$ were shown via the VMD and the Multiwfn programs by utilizing the formulas

$$\text{RDG}(r) = \frac{1}{2(3\pi^2)^{1/3}} \frac{|\nabla\rho(r)|}{\rho(r)^{4/3}}, \tag{4}$$

$$\Omega(r) = \text{sign}(\lambda_2(r))\rho(r). \tag{5}$$

The atomic theory in molecules (AIM) demonstrate that the strength and type of interaction are reflected respectively by the RDG(r) and $\Omega(r)$ with respect to the electron density $\rho(r)$ and eigenvalue λ_2.[62,63] As shown in Fig. 5, the negative $\text{sign}(\lambda_2(r))\rho(r)$ isosurfaces (blue color) between paired bases mean that there are distinct H-bond interactions between paired bases in S_0/S_1 states, thus generating size-altered stable helical structure in DNA.

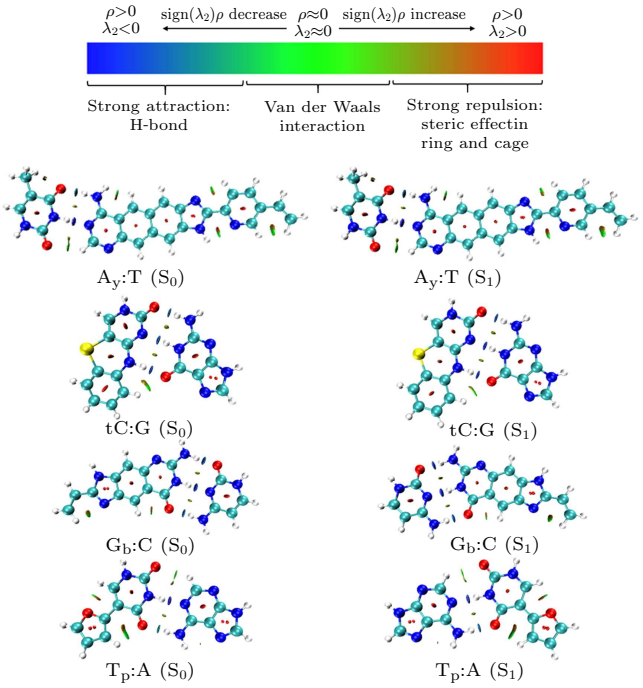


Fig. 5. The color-filled RDG isosurface (contour value = 0.5) of A_y:T, tC:G, G_b:C, and T_p:A in S_0 and S_1 states.

Based on the relaxed S_0-configurations, the absorption spectra of considered base pairs are performed in Fig. 6, and the excitation details are given in Table 2. The absorption spectra of T_p:A belongs to $S_0 \rightarrow S_1$ transition that is located at 327 nm ($f_{abs.} = 0.23$) corresponding to HOMO$\rightarrow$LUMO (CI = 97.3%). While, the maximum absorption wavelength of A_y:T and G_b:C are red-shifted to 366 nm ($f_{abs.} = 1.03$) and 377 nm ($f_{abs.} = 0.33$), respectively. The $S_0 \rightarrow S_3$ transition of A_y:T (366 nm) has dominant contributions from HOMO-1$\rightarrow$LUMO (CI = 56.2%) combined with HOMO$\rightarrow$LUMO +1 (CI = 37.6%), and the $S_0 \rightarrow S_1$ transition of G_b:C (377 nm) is associated with HOMO $\rightarrow$LUMO (CI = 97.3%). For tC:G, there are two absorption peaks around 375 nm ($f_{abs.} = 0.08$) and 308 nm ($f_{abs.} = 0.04$), which can be ascribed to the $S_0 \rightarrow S_1$ transition (HOMO$\rightarrow$LUMO with CI = 94.5%) and the $S_0 \rightarrow S_3$ transition (HOMO$\rightarrow$LUMO+1 with CI = 85.2%).

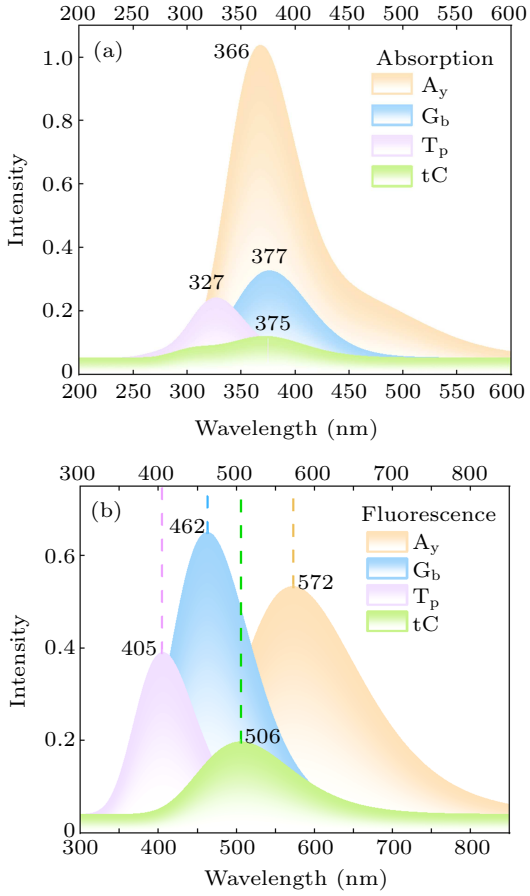


Fig. 6. Absorption (a) and fluorescence emission (b) spectra.

After the vertical photo-excitation, the emissive spectra are calculated at the excited geometries (Fig. 6). Similar to the case of monomer, the fluorescence peak of the four analogues can be ordered by A_y:T (572 nm, $f_{flu.} = 0.53$) > tC:G (506 nm, $f_{flu.} = 0.18$) > G_b:C (462 nm, $f_{flu.} = 0.64$) > T_p:A (405 nm, $f_{flu.} = 0.37$) due to the increased H-L energy gaps as 2.57 eV (A_y:T) < 3.11eV (tC:G) < 3.13 eV (G_b:C) < 3.52 eV (T_p:A). As displayed in Table 2, the fluorescence of the four analogues (A_y:T, tC:G, G_b:C, T_p:A) are all dominated by the LUMO$\rightarrow$HOMO (CI = 99.7%), LUMO$\rightarrow$HOMO (CI = 98.8%), LUMO$\rightarrow$HOMO (CI = 99.4%), and LUMO$\rightarrow$HOMO (CI = 99.4%) transitions, respectively. Accordingly, the distinct emission wavelength ($\Delta\lambda \sim 50$ nm) with efficient oscillator strength allows the four base pairs (A_y:T, tC:G, G_b:C, T_p:A) to serve as quasi-intrinsic biosensors for four-color optical scanner.

Table 2. Wavelengths (λ, in unit nm), oscillator strengths ($f_{abs.}$ and f_{flu}), state assignments for base pairs (A_y:T, tC:G, G_b:C, T_p:A) of absorption processes.

Optical spectra	Base pair	Transition	λ (nm)	$f_{abs.}$	Assignment (CI/Gap(eV)/Type)
Absorption	A_y:T	$S_0 \rightarrow S_1$	469.07	0.1944	H→L(98.8%/3.01/$\pi \rightarrow \pi^*$)
		$S_0 \rightarrow S_2$	376.57	0.1900	H→L+1(60.0%/3.79/$\pi \rightarrow \pi^*$) H-1→L(35.5%/3.82/$\pi \rightarrow \pi^*$)
		$S_0 \rightarrow S_3$	365.79	1.0306	H-1→L(56.2%/3.82/$\pi \rightarrow \pi^*$) H→L+1(37.6%/3.79/$\pi \rightarrow \pi^*$)
	tC:G	$S_0 \rightarrow S_1$	374.86	0.0794	H→L(94.5%/3.96/$\pi \rightarrow \pi^*$)
		$S_0 \rightarrow S_2$	339.32	0.0054	H-1→L(98.5%/4.12/$\pi \rightarrow \pi^*$)
		$S_0 \rightarrow S_3$	308.46	0.0372	H→L+1(85.2%/4.81/$\pi \rightarrow \pi^*$)
	G_b:C	$S_0 \rightarrow S_1$	377.08	0.3347	H→L(97.3%/3.70/$\pi \rightarrow \pi^*$)
		$S_0 \rightarrow S_2$	324.06	0.0274	H→L+1(95.2%/4.24/$\pi \rightarrow \pi^*$)
		$S_0 \rightarrow S_3$	306.54	0.0082	H→L+2(64.2%/4.70/$\pi \rightarrow \pi^*$) H-1→L(28.6%/4.72/$\pi \rightarrow \pi^*$)
	T_p:A	$S_0 \rightarrow S_1$	327.29	0.2307	H→L(97.3%/4.18/$\pi \rightarrow \pi^*$)
		$S_0 \rightarrow S_2$	295.20	0.0060	H-1→L(99.4%/4.65/$\pi \rightarrow \pi^*$)
		$S_0 \rightarrow S_3$	276.26	0.0192	H→L+1(99.6%/4.82/$\pi \rightarrow \pi^*$)
Fluorescence	A_y:T	$S_1 \rightarrow S_0$	572	0.53	L→H(99.7%/2.57/$\pi \rightarrow \pi^*$)
	tC:G	$S_1 \rightarrow S_0$	506	0.18	L→H(98.8%/3.11/$\pi \rightarrow \pi^*$)
	G_b:C	$S_1 \rightarrow S_0$	462	0.64	L→H(99.4%/3.13/$\pi \rightarrow \pi^*$)
	T_p:A	$S_1 \rightarrow S_0$	405	0.37	L→H(99.4%/3.52/$\pi \rightarrow \pi^*$)

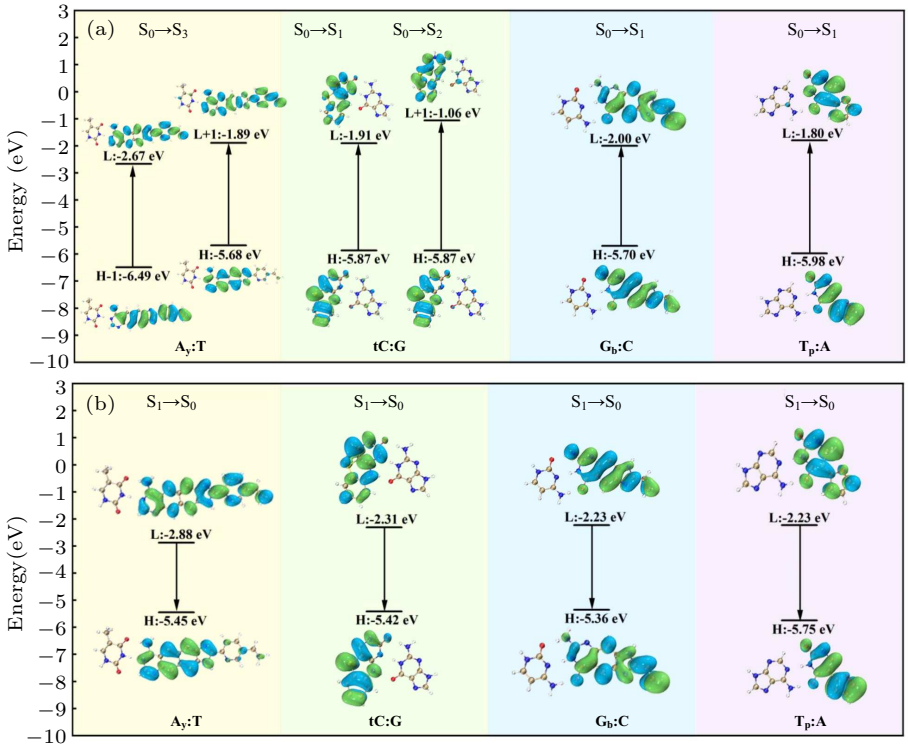


Fig. 7. FMOs (isovalue = 0.02 a.u.) involved in the excitation transitions (a) and the decay processes (b).

3.3. Effects of linking to deoxyribose on spectral characterization

With aim to ensure the biological application of the four-color indicators, we examined the effects of binding deoxyribose on spectral characterization. Geometrically, these bases retain co-planar characteristic after linking with deoxyribose (see Fig. S6). Electronically, the absorption and emission spectra at the relaxed S_0 and S_1 geometries are performed in Fig. 9. The absorption of dA_y is centred at 350 nm ($f_{abs.}$ = 0.72) originated from the $S_0 \rightarrow S_3$ transition. While, the other three analogues possess structured spectra with two absorption peaks. The maximum absorption peaks are located at 368 nm (dtC: $S_0 \rightarrow S_1$, $f_{abs.} = 0.11$), 286 nm (dG_b: $S_0 \rightarrow S_3$, $f_{abs.} = 0.67$), 265 nm (dT_p: $S_0 \rightarrow S_2$, $f_{abs.} = 0.42$), and the side peaks centred around 301 nm (dtC: $S_0 \rightarrow S_2$, $f_{abs.} = 0.08$), 361 nm (dG_b: $S_0 \rightarrow S_1$, $f_{abs.} = 0.35$), 330 nm (dT_p: $S_0 \rightarrow S_1$, $f_{abs.} = 0.31$), respectively. Inspection of the decay process reveals that the four analogues keep efficient fluorescence or-

dered in decreasing wavelength as dA_y: 584 nm ($f_{flu.} = 0.50$) > dtC: 493 nm ($f_{flu.} = 0.16$) > dG_b: 452 nm ($f_{flu.} = 0.65$) > dT_p: 408 nm ($f_{flu.} = 0.40$), which has unique contribution from LUMO to HOMO transition. Compared with the case of free bases, the fluorescence maxima of tC red-shifts from 488 nm to 493 nm with unchanged oscillator strength ($\Delta f < 4\%$) after linking to the deoxyribose. While, the linking with sugar groups could red-shift the fluorescence emission from 576 nm (A_y) to 584 nm (dA_y), 450 nm (G_b) to 452 nm (dG_b), and 406 nm (T_p) to 408 nm (dT_p), corresponding to an increasing oscillator strength from 0.45 (A_y) to 0.50 (dA_y), 0.60 (G_b) to 0.65 (dG_b), and 0.37 (T_p) to 0.40 (dT_p).

For base pairs, the spectra of absorption and fluorescence emission are depicted in Figs. 9(c) and 9(d). The maximum absorption peaks of the four analogues (dA_y:dT, dtC:dG, dGb:dC, dTp:dA) located at 350 nm ($f_{abs.} = 0.79$), 380 nm ($f_{abs.} = 0.12$), 376 nm ($f_{abs.} = 0.38$), and 330 nm ($f_{abs.} = 0.26$) that originated from the $S_0 \rightarrow S_3$ transition, $S_0 \rightarrow S_1$ transition, $S_0 \rightarrow S_1$ transition, and $S_0 \rightarrow S_1$ transition, respectively. Meanwhile, the fluorescence peak of four analogues can be ordered by dA_y:dT > dtC:dG > dG_b:dC > dT_p:dA with distinct emission wavelength ($\Delta\lambda \sim 50$ nm), which is essential to identify the sequencing information. Detailedly, the decay process of the four analogues (dA_y:dT: 581 nm ($f_{flu.} = 0.53$), dtC:dG: 508 nm ($f_{flu.} = 0.18$), dG_b:dC: 464 nm ($f_{flu.} = 0.64$), dT_p:dA: 406 nm ($f_{flu.} = 0.37$)) corresponds to $S_1 \rightarrow S_0$ emission and is dominated by the LUMO→HOMO transition (see Fig. S8 and Table S2 in the supporting information). The conclusions drawn above have been confirmed at the PBE0-D3 /6-311++G(d,p) level (see Table S3), which has also been shown to valuably predict the spectra properties of similar systems.[64]

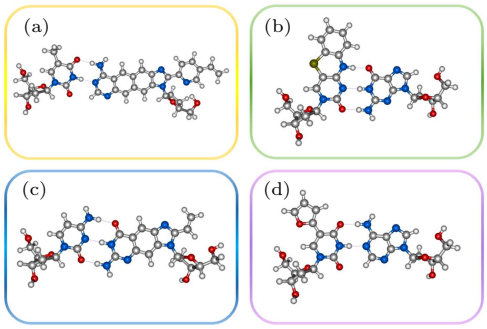


Fig. 8. Optimized geometries of dA_y:dT (a), dtC:dG (b), dG_b:dC (c) and dT_p:dA (d).

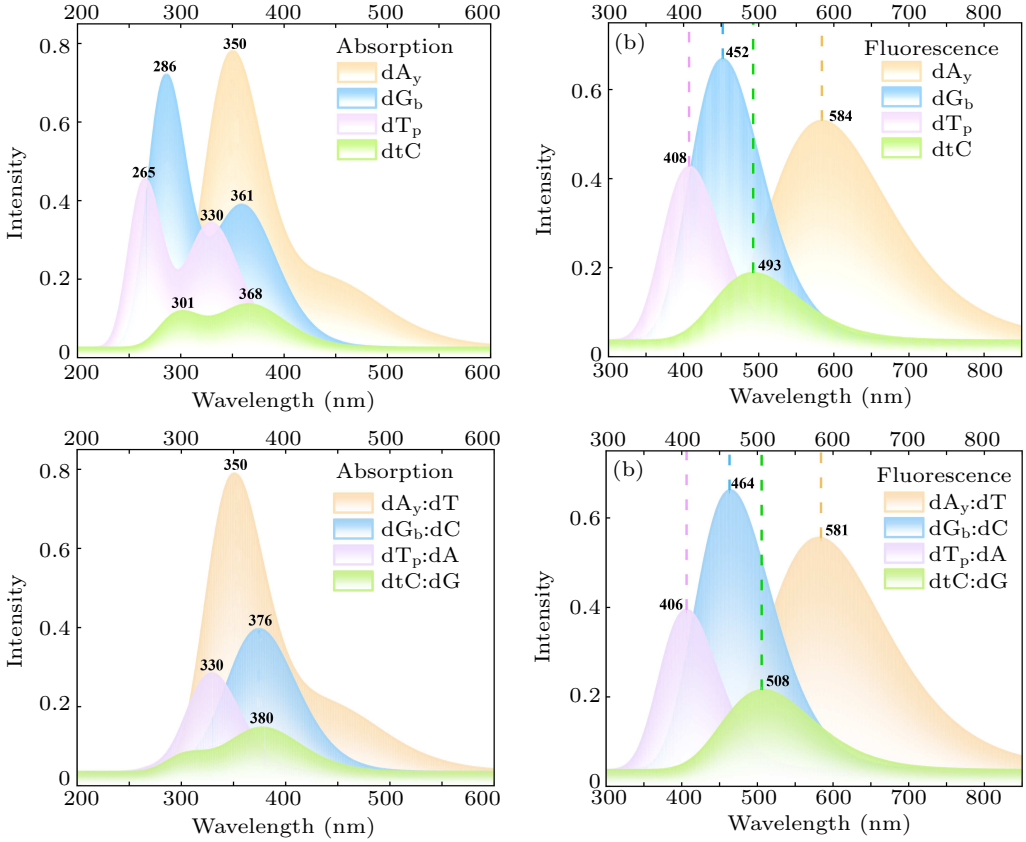


Fig. 9. Absorption (a), (c) and fluorescence emission (b), (d) spectra.

4. Conclusions

In this work, the structural and spectra properties of newly designed fluorescent bases are theoretically characterized to propose quasi-intrinsic biosensors for four-color DNA sequencing. Herein, two kinds of modifications were taken into account including the fluorophore replacement and ring expansion. These surrogates could keep the pyrimidine or purine underlying skeleton and Watson–Crick hydrogen bonding face to allow minimal perturbation to the native DNA duplex. More importantly, compared to the natural bases, these

nucleobase analogues possess red-shifted absorption and efficient fluorescence as a result of the enhanced π-conjugation in character. To ensure the utilization of biosensors for SMRT with two-base artificial sequence pattern, the effect of base pairing is examined. The fluorescence peak of the four analogues can be ordered by A_y:T (572 nm) > tC:G (508 nm) > G_b:C (464 nm) > T_p:A (406 nm) and the distinct emission wavelength ($\Delta\lambda \sim 50$ nm) allows for expected time trace of four-color fluorescence during the single-molecule real-time DNA sequencing. Meanwhile, the similar trends are reproduced with linking deoxyribose. This computational study proposes a set of fluorescent analogues that could serve as quasi-intrinsic biosensors for four-color scanner, thus contributing new insights into the design of probes for SMRT.

Acknowledgements

Project supported by the National Natural Science Foundation of China (Grant Nos. 11804195, 11847224, 11674198, and 12274265) and the Natural Science Foundation of Shandong Province, China (Grant Nos. ZR2018BA034 and ZR2022MA006).

References

[1] Watson J D and Crick F H C 1953 *Nature* **171** 737
[2] Watson J D and Crick F H C 1953 *Nature* **171** 964
[3] Zhu J, Xue J, Zhao W, Zhang C, Feng X Q and Wang K G 2023 *Chin. Phys. B* **32** 118704
[4] Wong K C, Zhang J, Yan S, Li X, Lin Q, Kwong S and Liang C 2019 *ACM Comput. Surv.* **52** 1
[5] Searle B, Müller M, Carell T and Kellett A 2023 *Chem. Int. Ed.* **62** e202215704
[6] LyuY K, Chen S, Zhao Y, Yuan H X, Zhang C Y, Zhang C Z and Meng Q T 2024 *Phys. Chem. Chem. Phys.* **26** 12552
[7] Lavergne T, Degardin M, Malyshev D A, Quach H T, Dhami K, Ordoukhanian P and Romesberg F E 2013 *J. Am. Chem. Soc.* **135** 5408
[8] Hassan S, Bahar R, Johan M F, Hashim E K M, Abdullah W Z, Esa E, Hamid F S A and Zulkafli Z 2023 *Diagnostics* **13** 373
[9] Xie Y, Chan L Y, Cheung M Y, Li M W and Lam H M 2023 *Plant Genome* **16** e20316
[10] Chen F, Gaucher E A, Leal N A, Hutter D, Havemann S A, Govindarajan S, Ortlund E A and Benner S A 2010 *Biochemistry* **107** 1948
[11] Dijk E L, Jaszczyszyn Y, Naquin D and Thermes C 2018 *Trends Genet.* **34** 666
[12] Eid J, Fehr A, Gray J, *et al.* 2009 *Science* **323** 133
[13] Choy L Y L, Peng W L, Jiang P Y, Cheng S H, Yu S C Y, Shang H M, Tse O Y O, Wong J, Wong V W S, Wong G L H, Lam W K J, Chan S L, Chiu R W K, Chan K C A and Lo Y M D 2022 *Clin. Chem.* **68** 1151
[14] Li Z J, Zhou X, Liao D X, Liu R L, Zhao X, Wang J, Zhong Q, Zeng Z, Peng Y Z, Tan Y L and Yang Z C 2023 *Front. Cell. Infect. Mi.* **13** 1180194
[15] Trémeaux P, Latour J, Ranger N, Ferrer V, Harter A, Carcenac R, Boyer P, Demmou S, Nicot F, Raymond S and Izopet J 2023 *Microbiology Spectrum* **11** 4
[16] Levene M J, Korlach J, Turner S W, Foquet M, Craighead H G and Webb W W 2003 *Science* **299** 682
[17] Foquet M, Samiee K T, Kong X, Chauduri B P, Lundquist P M, Turner S W, Freudenthal J and Roitman D B 2008 *J. Appl. Phys.* **103** 034301
[18] Wang Y, Sischka A, Walhorn V, Tonsing K and Anselmetti D 2016 *Biophys. J.* **111** 1604
[19] Xu W, Chan K M and Kool E T 2017 *Nat. Chem.* **9** 1043
[20] Kong J W, Dou S X, Li W, Li Hui and Wang P Y 2023 *Chin. Phys. Lett.* **40** 078701
[21] Zhang L B, Kong X M, Zheng M M and Wang M 2017 *J. Mol. Graph. Model.* **77** 339
[22] Ludford P T, Yang S, Bucardo M S and Tor Y 2022 *Chem. Eur. J.* **28** e202104472
[23] Jung J W, Edwards S K and Kool E T 2013 *ChemBioChem* **14** 440
[24] Parvin R, Jia Q, Ma J B, Xu C H, Lu Y, Ye F F and Li M 2022 *Chin. Phys. B* **31** 088701
[25] Chen Y H, Kong D R, Qiu L P, Wu Y G, Dai C H, Luo S, Huang Z P, Lin Q Y, Chen H, Xie S T, Geng L, Zhao J, Tan W H, Liu Y Q and Wei D C 2023 *Anal. Chem.* **95** 1446
[26] Edwards S K, Ono T, Wang S, Jiang W, Franzini R M, Jung J W, Chan K M and Kool E T 2015 *ChemBioChem* **16** 1637
[27] Wang R W, Jin C, Zhu X Y, Zhou L Y, Xuan W J, Liu Y, Liu Q L and Tan W H 2017 *J. Am. Chem. Soc.* **139** 9104
[28] Abdullah R, Xie S T, Wang R W, Jin C, Du Y L, Fu T, Li J, Tan J, Zhang L L and Tan W H 2019 *Anal. Chem.* **91** 2074
[29] Schmidt O P, Mata G and Luedtke N W 2016 *J. Am. Chem. Soc.* **138** 14733
[30] Duan L J, Zhang X, Zhao Y, Meng Q T and Zhang C Z 2023 *Phys. Chem. Chem. Phys.* **25** 3859
[31] Tanpure A A, Pawar M G and Srivatsan S G 2013 *Isr. J. Chem.* **53** 366
[32] Liu H B, Gao J M, Lystranne M, Y David S and Eric T K 2004 *J. Am. Chem. Soc.* **126** 1102
[33] Wilhelmsson L M, Sandin P, Holmén A, Albinsson B, Lincoln P and Nordén B 2003 *J. Phys. Chem. B* **107** 9094
[34] Frisch M J, Trucks G W, Schlegel H B, *et al.* 2013 *Gaussian* 16 *Revision A*.03 (Gaussian Inc)
[35] Humphrey W, Dalke A and Schulten K 1996 *J. Mol. Graph.* **14** 33
[36] Lu T and Chen F W 2012 *J. Comput. Chem.* **33** 580
[37] Goerigk L, Hansen A, Bauer C, Ehrlich S, Najibi A and Grimme S 2017 *Phys. Chem. Chem. Phys.* **19** 32184
[38] Becke A D 1993 *J. Chem. Phys.* **98** 5648
[39] Zhao G J, YangY F, Zhang C Y, Song Y Z and Li Y Q 2021 *J. Lumin.* **230** 117741
[40] Kumar A and Sevilla M D 2008 *J. Am. Chem. Soc.* **130** 2130
[41] Qi Y T, Wang Y, Tang Z, Liu J Y, Hou Y M, Gao Z Q, Tian J and Fei X 2020 *J. Mol. Liq.* **314** 113614
[42] Cossi M, Scalmani G, Rega N and Barone V 2022 *J. Chem. Phys.* **117** 43
[43] Tomasi J, Mennucci B and Cammi R 2005 *Chem. Rev.* **105** 2999
[44] Zhao J F, Jin B and Tang Z 2022 *Phys. Chem. Chem. Phys.* **24** 27660
[45] Yang Y F, Shi W, Chen Y P, Ma F C and Li Y Q 2021 *J. Lumin.* **229** 117698
[46] Grimme S, Antony J, Ehrlich S and Krieg H 2010 *J. Chem. Phys.* **132** 154104
[47] Perdew J P, Burke K and Ernzerhof M 1996 *Phys. Rev. Lett.* **77** 3865
[48] Zhang Y K and Yang W T 1997 *Phys. Rev. Lett.* **78** 1396
[49] Carlo A and Vincenzo B 1999 *J. Chem. Phys.* **110** 6158
[50] Duan L J, Zhao Y, Zhang X, Cui X X, Meng Q T and Zhang C Z 2022 *Spectrochim. Acta A* **282** 121675
[51] Green J A, Jouybari M Y, Aranda D, Improta R and Santoro F 2021 *Molecules* **26** 1743
[52] Sandin P, Wilhelmsson L M, Lincoln P, Powers V E C, Brown T and Albinsson B 2005 *Nucleic. Acids. Res.* **33** 5019
[53] Sandin P, Lincoln P, Brown T and Wilhelmsson L M 2007 *Nat. Protoc.* **2** 615
[54] Moliner F, Bufi F N and Vendrell M 2024 *Curr. Opin. Chem. Biol.* **80** 102458
[55] Callis P R 1983 *Ann. Rev. Phys. Chern.* **34** 329
[56] Barzilai I L, Bulatov V, Gridin V V and Schechter I 2004 *Anal. Chim. Acta* **501** 151
[57] Lukes V, Aquino A and Lischka H 2005 *J. Phys. Chem. A* **109** 10232
[58] Cao R F, Naivar M A, Wilder M and Houston J P 2014 *Cytom part A* **85** 999
[59] Shin D, Sinkeldam R W and Tor Y 2011 *J. Am. Chem. Soc.* **133** 14912
[60] Park S, Otomo H, Zhenga L and Sugiyama H 2014 *Chem. Commun.* **50** 1573
[61] Duijneveldt F B, Duijneveldt R J G C M and Lenthe J H 1994 *Chem. Rev.* **94** 1873
[62] Lu T and Chen Q X 2021 *Chem. Methods* **1** 231
[63] Johnson E R, Keinan S, Sánchez P M, García J C, Cohen A J and Yang W 2010 *J. Am. Chem. Soc.* **132** 6498
[64] Zhan H B, Wang Y, Li Z X, Tang Z, Tian J and Fei X 2021 *J. Phys. Chem. A* **125** 2866