Fundamental photophysics of isomorphic and expanded fluorescent nucleoside analogues

Dmytro Dziuba, Pascal Didier, Stefano Ciaco, Anders Barth, Claus A. M. Seidel, Yves Mély

Fundamental photophysics of isomorphic and expanded fluorescent nucleoside analogues

Dmytro Dziuba[1,†], Pascal Didier[1,†], Stefano Ciaco[1,2], Anders Barth[3], Claus A. M. Seidel[3], Yves Mély[1, *]

[1]Laboratoire de Bioimagerie et Pathologies, UMR 7021, Université de Strasbourg, 74 route du Rhin, 67401 ILLKIRCH, France

[2]Department of Biotechnology, Chemistry and Pharmacy, University of Siena, via Aldo Moro 2, 53100 Siena, Italy.

[3]Institut für Physikalische Chemie, Lehrstuhl für Molekulare Physikalische Chemie, Heinrich-Heine-Universität, 40225 Düsseldorf, Germany

Abstract: Fluorescent nucleoside analogues (FNAs) are structurally diverse mimics of the natural essentially non-fluorescent nucleosides which have found numerous applications in probing the structure and dynamics of nucleic acids as well as their interactions with various biomolecules. In order to minimize disturbance in the labelled nucleic acid sequences, the FNA chromophoric groups should resemble the natural nucleobases in size and hydrogen-bonding patterns. Isomorphic and expanded FNAs are the two groups that best meet the criteria of non-perturbing fluorescent labels for DNA and RNA. Significant progress has been made over the past decades in understanding the fundamental photophysics that governs the spectroscopic and environmentally sensitive properties of these FNAs. Herein, we review recent advances in the spectroscopic and computational studies of selected isomorphic and expanded FNAs. We also show how this information can be used as a rational basis to design new FNAs, select appropriate sequences for optimal spectroscopic changes and interpret fluorescence data in FNA applications.

† These authors contributed equally to this work

* Correspondence to: yves.mely@unistra.fr

1. Introduction

Interactions of nucleic acids with sequence-specific DNA- and RNA-binding proteins play a key role in transcription, translation, regulation of RNA stability, localization and turnover, and many other cellular processes [1–3]. Structures of hundreds of nucleic acid–protein complexes have become available as a result of recent advances in high-throughput protein crystallography and cryo-electron microscopy. However, structural biology methods mostly provide static snapshots of highly dynamic supramolecular complexes. Complementary experimental techniques are thus needed to study the binding in real time and to probe the dynamics of nucleic acid–protein complexes, its role in molecular recognition and its biological implications. Fluorescence is undoubtedly one of the most powerful and widely used spectroscopic modalities for studying the dynamics and interactions of biological molecules in solution. The dominant role of fluorescence spectroscopy and fluorescence-based methods in emerging and established interdisciplinary research areas, such as dynamic structural biology, biophysical and bioanalytical chemistry [4–9] is based on the unique combination of features offered by fluorescence emission. One of the most important features is the high sensitivity of fluorescence detection which guarantees acceptable signal-to-noise ratios even when small amounts of material is used. Moreover, several fluorescence parameters are available for quantitative measurements, such as the total intensity, the energy distribution of the emitted photons (i.e., the emission spectrum), the maximum emission wavelength, the anisotropy, quantum yield and time-resolved fluorescence parameters, to name the major ones. These easy-to-measure quantitative parameters in combination with spectral multiplexing unlock the possibility of multiparametric probing and selective analysis of sub-populations within heterogeneous molecular ensembles [10]. A broad range of instrumental setups, well-developed methods, user-friendly protocols, and reliable statistical data analysis pipelines are readily available for recording and interpreting fluorescence data from samples of various size, heterogeneity, and molecular complexity. As a result, fluorescence-based techniques are ideally suited for monitoring conformational dynamics, intermolecular interactions, biochemical transformations of individual biomolecules and multicomponent supramolecular complexes with nanoscale spatial and nanosecond temporal resolution, a level of precision that can barely be reached with other spectroscopic techniques. However, the success of fluorescence-based experiments largely depends on the selection of appropriate fluorescent probes.

Studying the natural nucleic acids DNA and RNA with fluorescence methods relies almost entirely on extrinsic small-molecule organic fluorescent reporters. Indeed, unlike the natural aromatic amino acids in proteins, the natural nucleobases in nucleic acids are virtually non-fluorescent under ambient conditions [11]. The need for extrinsic fluorescent reporters for DNA and RNA labelling had initiated a fifty-year research effort on the design of tailor-made fluorescent reporters compatible with nucleic acids, and the parallel search for reliable chemical and biochemical methods for their site-selective incorporation into polynucleotides [12]. Numerous DNA and RNA labelling strategies have been developed and detailed in recent dedicated reviews [13–23]. Moreover, the applications of fluorescently labelled nucleic acids in bioanalysis and biosensing have been also extensively described and reviewed [24–52].

A facile and straightforward labelling strategy consists in attaching a small-molecule organic fluorescent dye (such as rhodamine, BODIPY, etc.) via a non-nucleosidic linkage to the termini of oligonucleotides chemically synthesized by the solid-phase phosphoramidite method [53]. Nucleic acids can also be labelled internally by introducing nucleotides with reactive groups during the solid-phase synthesis, allowing to target the phosphate, sugar or nucleobase moieties [12,54,55]. End-labelled fluorescent oligonucleotide probes have been utilized in many applications that include the detection

of specific nucleic acid sequences in cells via fluorescence hybridization (FISH and related methods) and the monitoring of the binding of DNA or RNA to proteins or ligands via fluorescence anisotropy measurements [48,56,57]. Fluorophore-labelled oligonucleotides are also instrumental in ensemble and single-molecule FRET studies of the conformational heterogeneity of nucleic acids and the conformational changes accompanying nucleic acid – protein interactions [58–65].

A more sophisticated, but also more synthetically demanding labelling strategy consists in the position-specific replacement of a natural nucleobase with a chemically modified fluorescent nucleobase analogue. These fluorescent biomolecular building blocks, dubbed fluorescent nucleoside analogues (FNAs), are indispensable when an internal fluorescent reporter is needed, for instance to report on site-specific conformational and microenvironment changes via an interpretable change of the fluorescence signal. Hundreds of FNAs have been synthesized and position-selectively incorporated by solid phase synthesis or enzymatic approaches into DNA and RNA oligonucleotides mimicking natural nucleic acid sequences of interest. The optimal synthetic approach depends on the amounts needed and the length of the sequences, knowing that the practical limits of phosphoramidite-based oligonucleotide synthesis are about 100 nt and 60 nt for DNA and RNA, respectively [66,67].

A detailed survey on the structural diversity and applications of FNAs have been summarized in dedicated reviews [14,68–78]. However, the huge variety of FNAs makes it difficult to choose an optimal reporter for a particular application. An ideal FNA should fulfil a number of criteria needed for rational design of the experiment and interpretation of the spectroscopic data. Firstly, it should be able to mimic the natural nucleosides when incorporated into the nucleic acid sequence of interest without perturbing the native structure of the labelled nucleic acid and its interactions. Secondly, the FNA should possess optimal spectroscopic properties, such as a red-shifted absorption band compared to the absorption of the natural nucleic acids and a high fluorescence quantum yield (Φ_F) when incorporated in the nucleic acids. Finally, the mechanisms underlying FNA photophysics must be well understood, so that any spectroscopic change could be easily interpreted in terms of structural or environmental changes.

The photophysics of FNAs should therefore differ from that of the natural nucleobases, which are all weakly fluorescent (for more details, the reader is referred to excellent review articles [11,79–81]). The extremely low quantum yield ($\Phi_F < 10^{-4}$) of the nucleobases is a result of highly efficient non-radiative relaxation pathways from the excited $^1\pi\pi^*$ state to the ground state [82,83]. In the free nucleobases, fast internal conversion (IC) occurs on the picosecond timescale [82,84] through a conical intersection that involves a ring distortion [81,85–90], resulting in the formation of vibrationally hot molecules in the ground state. A competing pathway has also been identified for pyrimidine bases that involves a longer-lived (~10-150 ps) excited state with $^1n\pi^*$ character [80,91]. Moreover, intersystem crossing to triplet states has been reported for all nucleobases as another non-radiative pathway [92]. In DNA, additional decay pathways promoted by the proximity to other bases have been described, such as exciplexes of stacked bases with strong charge transfer character [80,93–96] or photo-induced proton transfer on Watson-Crick base pairs [97–99]. Of special importance are light-induced reactions from the excited singlet-state or the longer-lived $^1n\pi^*$ or triplet states. Resulting DNA photolesions, such as thymine dimers, are responsible for point mutations in the genetic material upon UV exposure [79].

Based on their molecular structure and relations to the natural nucleobases, FNAs can be divided into isomorphic, expanded, extended, and chromophoric base analogues (Figure 1).[75] Isomorphic nucleoside analogues differ from the natural nucleosides by the presence of small substituents or by the number and position of heteroatoms in the heterocyclic core. Expanded nucleosides contain

additional aromatic rings annealed to the purine or pyrimidine core. Extended nucleoside analogues bear a fluorophore tethered to the nucleobase via a linker. Finally, in chromophoric nucleoside analogues the entire nucleobase is replaced by a structurally distinct, and often bulky, aromatic chromophore. In this review, we focus on the photophysics of the isomorphic and expanded families, while the photophysical properties of the extended and chromophoric base analogues will be described by us in a forthcoming review. An important advantage offered by isomorphic and expanded fluorescent nucleosides as probes for nucleic acids is their small molecular size and similarity with the natural nucleobases in the ability to participate in non-covalent interactions. These factors minimize disturbances in nucleic acid folding and stability, and interference with the ability of the labelled nucleic acids to participate in intermolecular interactions. On the other hand, these classes of FNAs possess a number of drawbacks limiting their applications. For instance, their excitation spectral range is fundamentally limited by the small size of the conjugated electronic system. Excitation in the UV region can cause a number of undesired photochemical processes, leading to the formation of photo-lesions, such as covalent adducts and strand breaks. Another fundamental limitation is their relatively low brightness, due to their low molar absorption coefficient and low quantum yields in nucleic acids. The advantages and drawbacks must be carefully considered when choosing the class of FNA (Figure 1) and the specific reporter for a particular application. In this review, we will give an overview of the pro- and cons of using these minimalist FNAs. In addition to reviewing their photophysical properties in the form of free nucleosides and incorporated in nucleic acids as well as their recent applications, we will emphasize the following questions. 1) Why do small structural differences give FNAs remarkably different photophysical properties compared to naturally occuring nucleosides? 2) What are the main electronic excited-states and photophysical processes involved in absorption and emission properties of these FNAs? 3) How can the knowledge of the excited-state processes and photophysics of FNAs be exploited for the rational development of bioanalytical and biophysical applications, and for straightforward interpretation of the obtained data? Obviously, this approach limits the number of particular FNAs that can be discussed, so that we selected representative examples in each family. We first start by discussing, in section 2, the isomorphic FNAs of the purine and pyrimidine sub-groups. In the next section, we move on to expanded FNAs and their photophysics. The review ends with a brief outlook section, outlining current and future trends in the field.

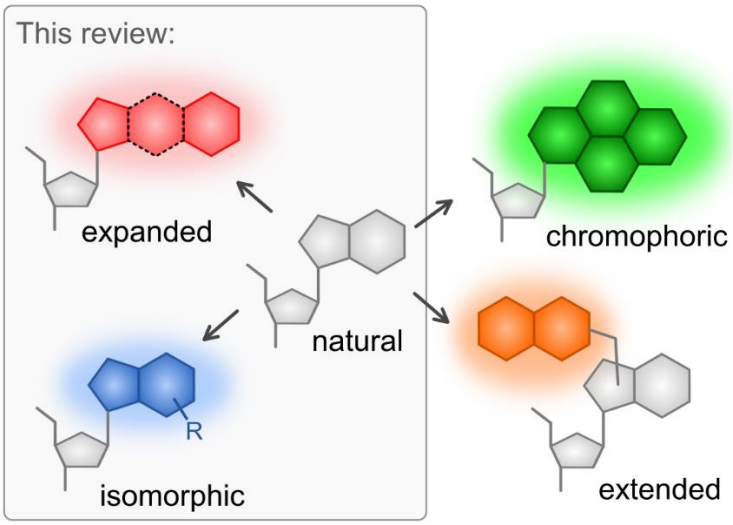

Figure 1. Classification of fluorescent nucleoside analogues (FNAs) based on their structural relationship with the natural essentially non-fluorescent nucleosides.

2. Isomorphic nucleoside analogues

With the exception of rare hyper-modified nucleosides in tRNAs such as wyosine [100–102], no intrinsic reporters are available for fluorescence probing of nucleic acids. An interesting option is to transform the structures of natural nucleobases in an atom-economic way to endow them with fluorescence properties while preserving their molecular size, and their ability to form base pairs and participate in intra- and intermolecular interactions. Several nucleoside analogues have been designed in this way. In the literature, they are referred to as isomorphic fluorescent nucleoside analogues. The most prominent members of this family are discussed in this section.

2.1. 2-Aminopurine (2AP)

2-Aminopurine (**2AP**, Figure 2A) was first described by Stryer and co-workers in the late 1960s as a fluorescent reporter for spectroscopic studies of structure and interactions of nucleic acids [103]. Since then, **2AP** has become the most widely used fluorescent nucleoside analogue of purines (Figure 2B) for studying DNA and RNA dynamics, folding and interactions [69,104]. **2AP** is a constitutional isomer of adenine (Figure 2, A–C) which differs from adenine by the position of the exocyclic amino group. **2AP** forms a Watson-Crick base pair with thymine (Figure 2D), which is thermodynamically less stable than the canonical A·T pair [105,106], and can isomerize into a wobble pair [107] (Figure 2E). **2AP** can also hydrogen bond to cytosine, forming a mismatch that exists as a mixture of a neutral wobble pair (Figure 2E) and a protonated pair in the Watson−Crick geometry (not shown) [108–110]. Moreover, **2AP** can also form a wobble mismatch with adenine that distorts the local B-DNA conformation [111] as well as a Hoogsteen-type base pair with guanine [112]. The relative stability of the modified base pairs decreases in the following order: **2AP**·T > **2AP**·C > **2AP**·A > **2AP**·G [106]. As shown by calorimetric and spectroscopic measurements, **2AP** slightly destabilizes double-stranded DNA [106,113,114]. Double-stranded DNA incorporating a **2AP**·T pair preserves the overall B-DNA conformation but causes local geometry alterations, such as narrowing of the minor grove [105,115] and local destabilization [105,116117,118]. The **2AP**·T base pair (Figure 2D) has an about 80% shorter lifetime as compared to the natural A·T base pair [116,119]. The flanking base pairs are also destabilized but to a lesser extent (20–45%), resulting in changes of the twist angle and transient openings of the helix near the site of modification. Despite that, multiple studies have shown that **2AP** can be used as a reliable fluorescent probe for DNA and RNA, since the conclusions obtained in a large number of applications, and in particular those related to nucleic acid annealing and nucleic acid–protein recognition are not significantly affected by the small perturbations caused by **2AP**.

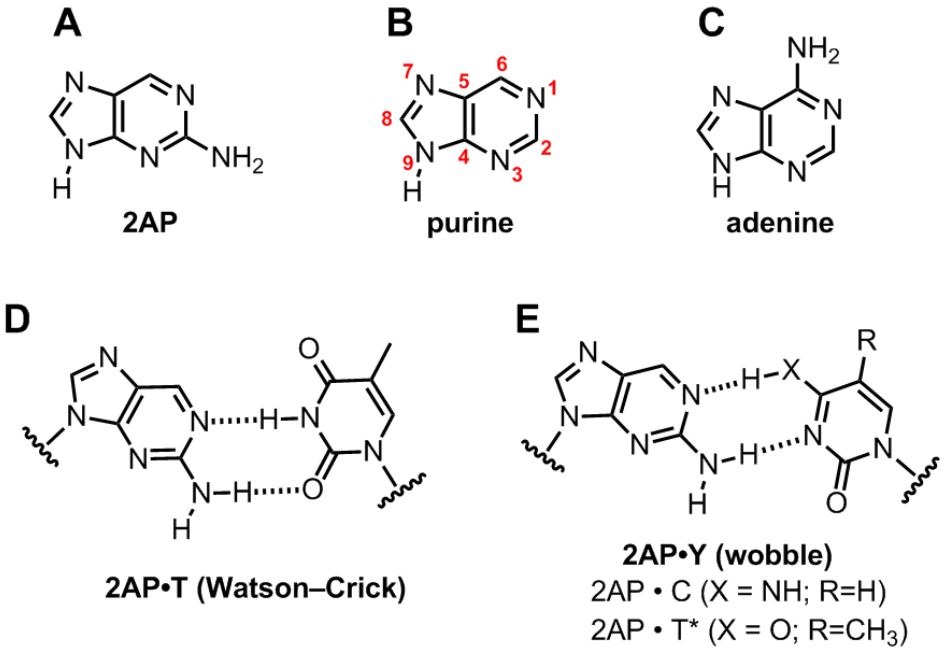

Figure 2. Chemical structures of *9H*-**2AP** (A), *9H*-purine (B), and *9H*-adenine (C); conventional purine numbering is shown in red. Hydrogen bonding of **2AP** to T in Watson–Crick geometry (D) and to pyrimidines (Y) in the wobble geometry (E); T* stands for thymine in its enol form.

2.1.1. Photophysics of the free 2AP chromophore

The **2AP** chromophore has been studied in the forms of free nucleobase, 9-alkyl-2-aminopurine, and the corresponding riboside and 2'-deoxyriboside derivatives using a wide range of spectroscopic and computational methods. As a general trend, the substituent at the position 9 has a marginal effect on the absorption and fluorescence properties of the chromophore [120,121]. The main exception is the free nucleobase, where the absence of substituents leads to a 9*H*-**2AP** ⇆ 7*H*-**2AP** tautomerism with subtle differences in the electronic structures and relaxation pathways of the two tautomers [122].

Desolvated **2AP** chromophore in vacuum is practically non-fluorescent with a Φ_F of 0.005 [123]. Conversely, **2AP** is highly fluorescent in aqueous solutions exhibiting a Φ_F of 0.66 – 0.72, a single emission band with a maximum at ~370 nm (Table 1) and a minimal sensitivity to pH [103,121]. The absorption with a maximum at ~303 nm enables selective photoexcitation of the **2AP** fluorophore in the presence of proteins and non-labelled nucleic acids. Φ_F of **2AP** decreases markedly in apolar solvents (5- to 68-fold) as compared to water [103,120,124]. This Φ_F decrease in apolar solvents is accompanied by a 10–15 nm blue-shift of the emission maximum [120]. Time-resolved fluorescence decays of **2AP** nucleoside in aqueous media are monoexponential with a lifetime of ~ 11 ns which decreases in less polar solvents [122,124,125]. The fluorescence properties of **2AP** differ dramatically from those of the barely fluorescent adenine and the free purine base, despite their nearly identical conjugated π-electronic systems (Figure 2) [11,126,127].

Understanding the relaxation pathways of **2AP** is instrumental for its use as a fluorescent probe for DNA and RNA. A large number of spectroscopic and computational studies have been performed to answer three central questions: i) how does the interplay of a multitude of electronic, structural and environmental factors makes **2AP** to stand out as a bright environmentally sensitive fluorophore with respect to its non-emissive purine relatives? ii) what mechanisms are responsible for the unique

sensitivity of the fluorophore to environmental factors, such as the polarity, hydration and base stacking? iii) how should one interpret the complex multicomponent fluorescence decays obtained from a **2AP** probe site-selectively incorporated into nucleic acids? A number of excellent reviews have focused on the photophysics and applications of **2AP** [69,104,128]. In this section, we will describe the electronic excited states and relaxation pathways of **2AP**, focusing on a few prominent papers that have influenced the field, in the chronological order.

The free purine base, **2AP** and adenine share a similar set of low-lying excited electronic states involved in the absorption of light, excited state dynamics, and competitive relaxation pathways. These states include two $^1\pi\pi^*$ states (historically designated as ^1L_a and ^1L_b), a low-lying $^3\pi\pi^*$ state, and a $^1n\pi^*$ state [126]. The key differences between the three electronic systems resulting in their dramatically different spectroscopic properties are the relative ordering of the excited states at different molecular geometries, the local topography (presence of local minima, energetic profile of minimum energy pathways) and the global topology (presence of conical intersections) of the potential energy surfaces of the electronic states.

The Lim proximity model [129–131] was first used to explain the environmental sensitivity of **2AP** emission. According to this phenomenological model, the small energy gap between the $\pi\pi^*$ and $n\pi^*$ excited states in aromatic molecules with non-bonding electrons facilitates internal conversion. The two excited states couple through low-energy, out-of-plane vibrational distortions. This in turn leads to an overlap with ground state vibrations and thus promotes non-radiative relaxation to the ground state. The vibronic coupling between the $^1\pi\pi^*$ and $^1n\pi^*$ levels was used by several authors to explain the decrease of Φ_F of **2AP** in non-polar solvents. They argued that stabilization of the $^1n\pi^*$ state and destabilization of the $^1\pi\pi^*$ state in non-polar solvents decreased the energy gap separating them, and thus enhanced the internal conversion [123,132]. However, this scenario was later controverted by *ab initio* calculations showing that both **2AP** and adenine have similar energy gaps between $^1\pi\pi^*$ and $^1n\pi^*$ excited states, which could not explain the large differences in internal conversion efficiency between the two nucleobases [133]. A more comprehensive view was obtained from *ab initio* calculations of the minimum energy paths (MEPs) on the potential energy surfaces (PESs) of the low-lying singlet excited states. In their milestone papers, Serrano-Andres and co-authors [133,134] analysed the PESs of **2AP** and adenine, including the accessible conical intersections. For **2AP**, the low-energy absorption and fluorescence emission bands can be attributed to the lowest of the two excited $\pi\pi^*$ states (^1L_a and ^1L_b). However, the exact ordering of $^1(\pi\pi^*\ L_a)$ and $^1(\pi\pi^*\ L_b)$ levels has been controversial [133,135]. After excitation, the $^1(\pi\pi^*)$ state evolves from its Franck-Condon (FC) conformation to a $^1(\pi\pi^*)_{min}$ planar minimum, before releasing the energy by photon emission. Alternately, $^1(\pi\pi^*)_{min}$ can convert into a high-energy transition state, $^1(\pi\pi^*)_{TS}$, associated with a 5 kcal/mol energy barrier and characterized by an out-of-plane deformation of the N3C2 bond, which moves the amino group in a perpendicular direction with respect to the plane of the nucleobase ring. The transition state converts then to the conical intersection (CI) with the ground state $(gs/\pi\pi^*)_{CI}$, which non-radiatively relaxes the molecule back to the ground state. As a consequence of the high energy barrier on the major radiationless relaxation path via the $(gs/\pi\pi^*)_{CI}$, the relaxation via fluorescence is favoured in **2AP**, explaining its high Φ_F. In this model, Φ_F of **2AP** is thought to be enhanced in polar media, due to the solvent-induced stabilization of the $^1\pi\pi^*$ state as compared with the CI, which further increases the barrier to the CI by ~3 kcal/mol and thus further disadvantages the non-radiative relaxation [136].

For adenine, the MES on the MEP of the $^1\pi\pi^*$ state connecting the FC state with a low-lying $(gs/\pi\pi^*)_{CI}$ has no apparent energy barrier [133,134]. Similarly to **2AP**, formation of the CI in adenine

results from the twisting and wagging of the N3C2 bond, but with only one H atom moving perpendicularly in respect to the nucleobase plane. In contrast to **2AP**, this out-of-plane deformation in adenine is barrierless, so that the CI is easily accessed. This barrierless process results in an ultrafast nonradiative internal conversion that very effectively competes with the radiative transition, resulting in an experimentally observed fluorescence lifetime of < 100 fs.

Though logical and attractive, this model grounded on *ab initio* calculations failed to predict that **2AP** is virtually non-fluorescent in its desolvated form in gas phase and becomes highly fluorescent in water. Experiments with jet-cooled solvent-free **2AP** revealed a low Φ_F of 0.005 associated to very short lifetimes of 77–210 ps [123,137,138]. In sharp contrast, supramolecular assemblies of **2AP** with few water molecules exhibited enhanced fluorescence. Binding of one or two water molecules at the sugar-edge site increased the **2AP** lifetime by 4 to 16 times. An even higher increase (95-fold) was achieved when a water molecule was H-bonded with N1 and the 2-amino group at the trans-amino site, giving a lifetime τ = 14.5 ns. A slightly lower increase (30-fold, τ = 6.5 ns) was observed when the water molecule was H-bonded with N3 and the 2-amino group. Finally, the longest lifetime (τ = 17 ns) was obtained when both the trans-amino and sugar-edge sites were hydrated. The control of **2AP** emission by site-selective microhydration was rationalized using the Lim proximity model. An excellent correlation was found between the energy difference of the calculated $^1n\pi^*$ and $^1\pi\pi^*$ energy minima and the lifetime values of the different **2AP** species. The fluorescence lifetimes were short when the $^1n\pi^*$ state was lower in energy than the $^1\pi\pi^*$ state, intermediate when both states were of similar energy and long-lived when the $^1n\pi^*$ state was above the $^1\pi\pi^*$ state [123]. The last scenario also takes place in bulk aqueous media because of the simultaneous destabilization of the $^1n\pi^*$ state and the stabilization of the $^1\pi\pi^*$ state by the H bonding with water molecules. The vibronic coupling between the bright $^1\pi\pi^*$ state and the dark $^1n\pi^*$ state was further supported by the rotational contour analysis of the 0_0^0 vibronic band [138] and out-of-plane vibrational overtone bands seen in the two-colour two-photon ionisation spectra of **2AP** and its hydrated complexes with relatively short fluorescence lifetimes [123]. Finally, a later computational study confirmed that the fluorescence increase for the hydrated **2AP** clusters is due to a higher energy barrier to internal conversion at the C6-puckered conical intersection [139].

In the unsubstituted purine base, intersystem crossing from the local minimum of the dark $^1n\pi^*$ state was shown to be an essential non-radiative relaxation channel [126]. A series of recent work has indicated that this mechanism is also relevant for the free **2AP** in solution. Indeed, a long-lived (~ 1–5 μs) triplet population has been directly observed in the transient absorption spectra of photoexcited **2AP** and two-color pump-probe spectroscopy [138]. In fact, intersystem crossing to the lowest triplet $^3\pi\pi^*$ state appears as a major non-radiative channel competing with the fluorescence emission. This radiationless pathway channel involves up to ~40% of the initial excited-state population [140,141].

Considering all the aforementioned data, the excited state processes in the monomeric **2AP** chromophore can be summarized in a scheme shown in Figure 3. Vertical excitation of the **2AP** chromophore mainly populates the lowest optically bright $^1\pi\pi^*$ excited electronic state (transition A→B in the figure) [121]. The initially excited chromophore undergoes fast vibrational cooling from the Franck-Condon region and reaches the local minimum of the $^1\pi\pi^*$ level (D) within a few picoseconds following the excitation. The topography of the potential energy hypersurface near the local minimum is unfavorable for ultrafast radiationless relaxation. Indeed, the CI with the ground electronic state in **2AP** (C) is separated from the local $^1\pi\pi^*$ minimum by a high energy barrier (D–C). In contrast to adenine where the $(gs/\pi\pi^*)_{CI}$ is readily accessible and responsible for the ultrafast relaxation dynamics, the CI in **2AP** can be directly reached only by a small sub-fraction of

vibrationally hot excited molecules. The inefficiency of the CI pathway and the presence of the well-shaped local $^1\pi\pi^*$ minimum enables fluorescence emission according to the Kasha principle (transition D→A).

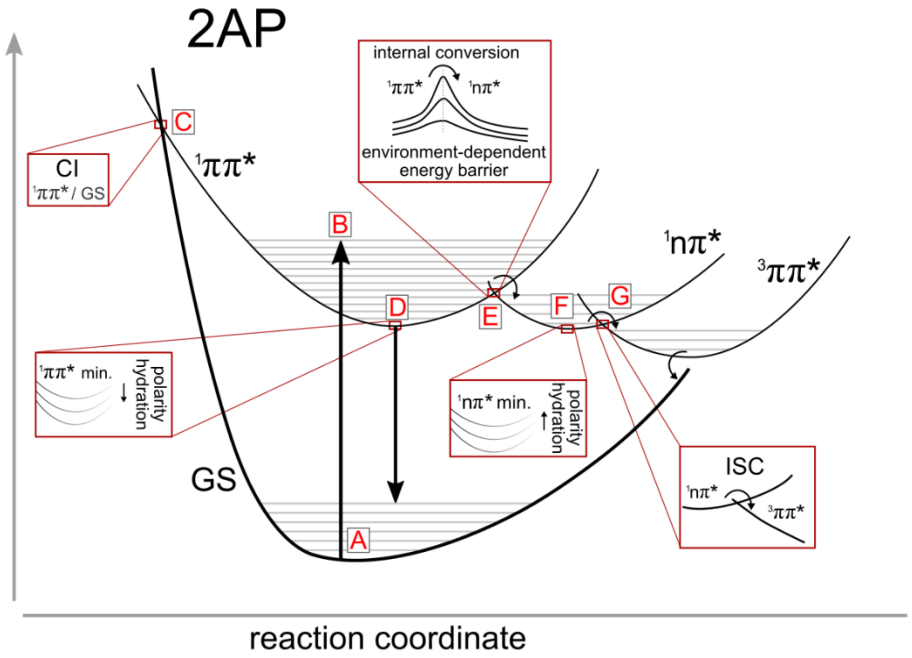

Figure 3. A summary of relaxation pathways in the photoexcited monomeric **2AP** chromophore. (A) Minimum of the ground electronic state; (B) the Frank-Condon excited state; (C) CI of the $^1\pi\pi^*$ level with the ground electronic state; (D) local minimum of the $^1\pi\pi^*$ level; (E) energy barrier on the internal conversion pathway to the $^1n\pi^*$ level; (F) local minimum of the $^1n\pi^*$ level; (G) energy barrier on the intersystem crossing pathway to the $^3\pi\pi^*$ level.

The dark $^1n\pi^*$ level (F) which is close in energy to the $^1\pi\pi^*$ level (D) is critical for the radiationless relaxation that competes with the fluorescence emission. The $^1\pi\pi^*$ and $^1n\pi^*$ levels are connected by a CI (E). The relative energy and the ordering of these levels, but also the height of the potential energy barrier separating them dramatically depend on the dielectric properties of the medium and the site-specific hydration of the chromophore. An increase in polarity and the presence of water molecules H-bonded to the exocyclic amino group stabilize the $^1\pi\pi^*$ level (D) and destabilize the $^1n\pi^*$ level (F). Accordingly, the energy barrier at the CI becomes higher in polar media, so that emission from the $^1(\pi\pi^*)_{min}$ is favoured. In contrast, the lowering of the energy barrier in low polar solvents and in aprotic/non-polar environment favors the internal conversion from the $^1\pi\pi^*$ state to the dark $^1n\pi^*$ state (path D→E→F) and leads to a low Φ_F and fluorescence lifetime. Noteworthy, the $^1n\pi^*$ level is likely populated only transiently due to inter-system crossing to the long-lived triplet state $^3\pi\pi^*$, which also non radiatively brings **2AP** back to the ground state. Thus, the exact energetic profile of the MEPs and the contributions of the emissive and non-emissive processes in the relaxation depend on the vertical excitation energy and environmental factors, such as solvent polarity and H-bonding with solvent.

2.1.2. Photophysics of 2AP in nucleic acids

2AP can be site-specifically incorporated into synthetic DNA and RNA oligonucleotides using standard solid-phase phosphoramidite chemistry [142,143]. This incorporation decreases Φ_F by up to two

orders of magnitude as compared with **2AP** in aqueous solution due to additional relaxation mechanisms enabled by the stacking and electronic interactions with the natural nucleobases [144]. Recently, the impact of the opposite base was investigated by using an isolated **2AP**·T base pair in Watson−Crick (WC) and Hoogsteen conformations in $CHCl_3$ that mimics the dielectric environment of the DNA double helix [145]. While the WC conformation was shown to provide slowed-down, monomer-like electronic relaxation ($\tau \sim 1.6$ ns) toward ground-state recovery and triplet formation, the **2AP**·T pair in the Hoogsten conformation exhibited faster deactivation ($\tau \sim 70$ ps). The quenching efficiency also depends on the nature of the neighbouring nucleobases, the local duplex conformation and mobility as well as the duplex form [104,144,146]. Computational and experimental studies revealed that an excited-state interbase electron transfer resulting in a charge-transfer (CT) state is the dominant quenching mechanism [147,148]. Photoexcited **2AP** acts as an electron acceptor, and natural nucleobases as electron donors for the CT, with G being the most efficient quencher [149]. Excited-state interactions between **2AP** and guanine or other nucleobases do not require physical interactions of their molecular orbitals. Quenching can take place when the interacting bases are separated by up to ~14 Å, which corresponds to a few base pairs. Nucleobases in between can affect the CT efficiency. For instance, adenines between guanine and **2AP** further increase the efficiency while pyrimidine bases significantly reduce it [104]. Fluorescence measurements of **2AP**-labeled DNA at cryogenic temperatures have shown a suppression of the CT quenching, indicating that the conformational dynamics and heterogeneity of **2AP** gate the process under physiological conditions [146,150]. The dependence of quenching on local base stacking is the central principle of **2AP**-based sensing. A decrease in the stiffness of the local nucleobase packing reduces the CT transfer efficiency and therefore results in a higher fluorescence intensity (Figure 4). This property of **2AP** has been widely utilized for probing interactions, transformations, and dynamics of nucleic acids (section 2.1.3). In addition to CT-mediated quenching, conical intersections with **2AP** ground state or dark states have been also identified in computational studies as important quenching mechanisms [148,151–153].

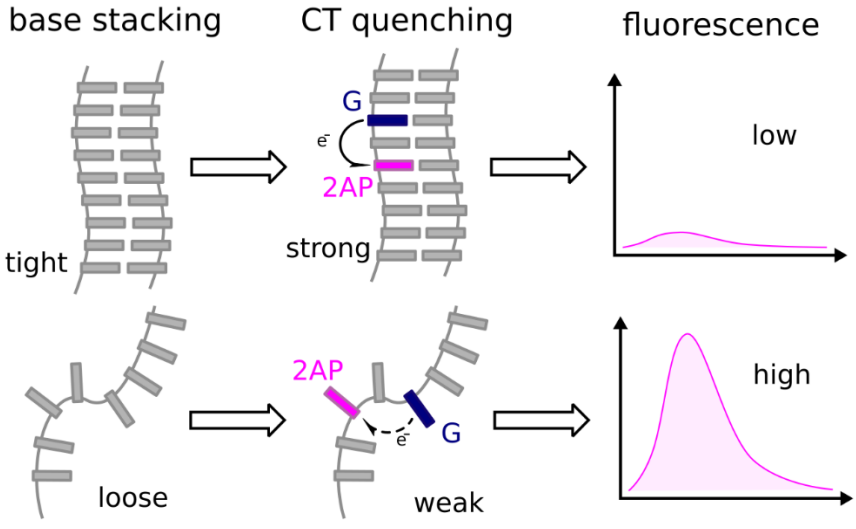

Figure 4. Quenching of **2AP** depends on the stacking interactions with the surrounding nucleobases. Highly stacked conformations enable facile excited-state electron transfer resulting in an ultrafast quenching and low fluorescence intensity (top). Conversely, unstacked conformations prevent CT quenching and elicit higher fluorescence intensity (bottom).

The various quenching mechanisms of **2AP** affect dramatically its time-resolved fluorescence decay. In contrast to the single exponential decay of free **2AP** in solution, the fluorescence decay of **2AP** in oligonucleotides is intrinsically complex. The decays of **2AP**-labeled dsDNA can typically be fitted with four components (Table 1) [105,144,154,155]. For decades, the paradigm was that the four components correspond to distinct conformational subpopulations where **2AP** has different quenching rates due to different orientation and stacking with the surrounding nucleobases [104]. The longest component was assigned to a fully solvent-exposed fluorophore conformation, whereas the shortest component was attributed to a highly stacked conformation, which is dominant in dsDNA [104]. The two other lifetime components were attributed to imperfectly stacked conformations, where **2AP** partially protruded into the solvent.

This paradigm was challenged by Kohler and co-workers [156], due to the observed discrepancies between Φ_F and fluorescence lifetime values. They concluded that the fully stacked conformation should not contribute significantly to the fluorescence emission recorded more than 50 ps after the photoexcitation, and should thus be barely detectable with the conventional TCSPC setups used in the majority of studies with **2AP**. The existence of these conformations with lifetimes less than the 30 ps detection limit of TCSPC measurements is consistent with a number of previous reports, where their existence with amplitudes as high as 90% was deduced from a comparison of the mean lifetimes and Φ_Fs between **2AP**-labelled oligonucleotides and free **2AP** [157,158]. Therefore, the fluorescence recorded by TCSPC setups likely originates exclusively from a multitude of unstacked conformations [156]. The authors also found out that the fluorescence decay kinetics of **2AP** can be modelled using a continuous lifetime distribution instead of discrete states. Moreover, molecular dynamics simulations and quantum chemistry calculations evidenced that the conformational space sampled by **2AP**-containing dinucleotides was pretty large and resulted in several metastable structures with various degrees of stacking [159–161]. In another study, Voltz and co-authors investigated the conformational heterogeneity of **2AP**-labeled DNA hairpins by MD simulations and cross-correlated the computed conformational distribution with fluorescence decay measurements covering five orders of magnitude in timescale [118]. This study showed that the fluorescence decay kinetics exhibited sequence-dependent fluorescence lifetime distributions, with major contributions from sub-50 ps fluorescence lifetimes, faster than the diffusional motions of the **2AP** fluorophore (~ 60 ps). These lifetimes likely originate from a population undergoing ultrafast (instant) quenching in a fully stacked conformation, as suggested by the MD simulations.

To conclude, the multiexponential decay of **2AP** in nucleic acids can be related to the existence of i) multiple conformers with different orientations of the nucleobase rings in respect to each other (structural heterogeneity) and ii) dynamic transitions between the conformations in the excited-state relaxation timescale (conformational dynamics) [159]. Accordingly, the structural heterogeneity plays a major role in the fast relaxation kinetics of **2AP** (sub-picosecond and picosecond time scale), whereas conformational dynamics likely contributes to the long decay components (nanosecond time scale). These studies show that an integrated approach combining ultrafast spectroscopy, available structural data and computer simulations is required for a reliable fluorescence lifetime-based probing of biomolecular complexes using **2AP**.

2.1.3. Selected applications of 2AP for studying dynamics and interactions of nucleic acids

A huge number of **2AP**-labeled synthetic mimics of natural nucleic acid sequences have been synthesized and used in a broad range of applications [104]. Most of the time, measurement of their

fluorescence intensity has proven as a simple and reliable method for monitoring structural and conformational alterations caused by changes in the environment or biochemical transformations. Most applications are based on the fact that an increase in the local base stacking and conformational stiffness results in a decrease in fluorescence, and vice versa. (Figure 4). Applications can be grouped in a several classes and illustrated by representative examples.

Probing the activity of processing enzymes. One of the simplest applications is the monitoring of the enzymatic activities of nucleic acid polymerases, nucleases, and helicases. For instance, **2AP** has been applied to investigate the activity of the Klenow Fragment of *Escherichia coli* DNA Polymerase I [162,163], bacteriophage T4 DNA Polymerase [162,164], DNA polymerase β [165], *E.coli* and T7 RNA polymerases [166,167], λ exonuclease [168], and T4 ATP-dependent DNA helicase [169]. More recently, **2AP** has been utilized to study the molecular mechanisms of target recognition and DNA cleavage by the bacterial RNA-guided endonuclease Cas9, a part of the adaptive immune defense of bacteria that has been repurposed into a revolutionary gene-editing tool [170,171]. Cleavage or formation of a phosphodiester bond in close proximity to the **2AP** can result in a dramatically altered fluorescence intensity. Another molecular event that can increase the fluorescence of **2AP** is the unwinding of dsDNA substrates by enzymes, such as DNA helicases, RNA polymerases, and Cas9 endonuclease (Figure 5). In general, an increase in fluorescence intensity is an indication of the conformational rearrangement of the substrate as well as a measure of the progress of the enzymatic reaction. Additional information regarding kinetics, thermodynamics and mechanism of the enzyme-DNA recognition can be obtained by introducing **2AP** at different positions relative to the site recognized by the enzyme, by using additional chemical modifications in the substrate, and by varying the concentration of salts and cofactors [163,164]. When combined with kinetic studies by fluorescence anisotropy, fluorescence intensity information from the **2AP** probe can directly detect transient intermediates along the enzymatic reaction pathway and measure the kinetics of individual conformational and chemical steps of the reaction. Because of their distinct absorption and emission bands in the UV, the fluorescence of **2AP** can be monitored simultaneously with the fluorescence of tryptophan for a comprehensive conformational analysis of the protein and DNA components of a catalytic complex [165]. Moreover, **2AP** can be combined with other fluorescent nucleosides such as pyrrolo-cytosine (section 3.1.1) for multiplexed enzymatic assays [172]. The utility of **2AP** for monitoring the cleavage of the sugar-phosphate backbone is not restricted to proteins and can also be used to monitor the catalytic activity of ribozymes [173].

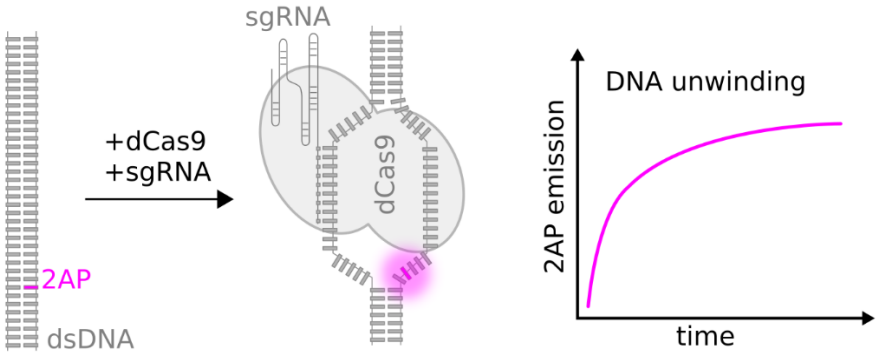

Figure 5. Binding of the inactivated Cas9 endonuclease (dCas9) complexed with a guide RNA (sgRNA) causes unwinding of the **2AP**-labelled dsDNA substrate. The unwinding kinetics can be monitored by the increase of **2AP** fluorescence in a stopped-flow setup [171].

Probing nucleic acid folding. Many non-coding RNAs (ncRNAs) fold into complex tertiary structures that are important for their cellular functions. Understanding the impact of physical and molecular factors on the conformational heterogeneity and folding pathways is critical for elucidating the biological functions of these RNAs [128,174–178]. Many ncRNAs fold and perform their cellular functions only when bound to metal ions. **2AP** has been used to study the Mg^{2+}-induced conformational changes of the hammerhead ribozyme [179] and the binding of Fe^{2+} to the iron responsive element (IRE), an RNA hairpin involved in the post-transcriptional regulation of iron metabolism genes [180]. **2AP** was also used to determine the number and relative affinity of the metal binding sites. In addition, **2AP** incorporation at several strategic positions allowed monitoring conformational rearrangements within specific RNA structural motifs, such as loops and base-paired stems for instance.

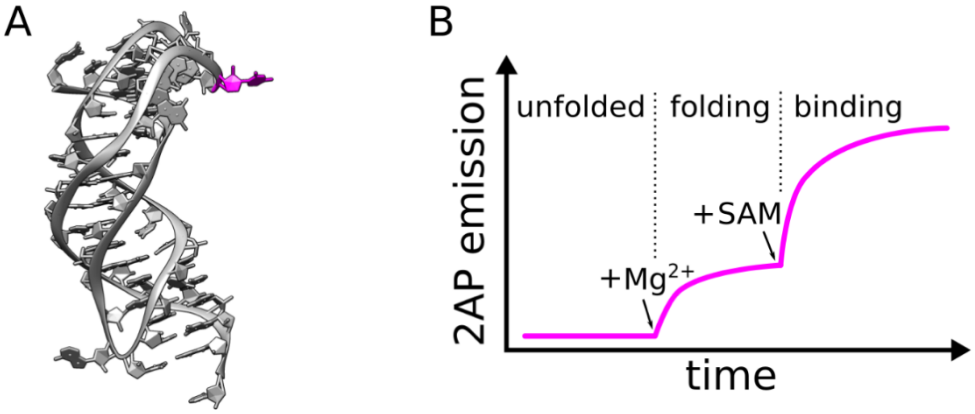

Figure 6. (A) Tertiary structure of the SAM-II riboswitch (PDB 2QWY). One of the positions selected for **2AP** incorporation based on the SHAPE data is shown in magenta. (B) Fluorescence kinetics measurements of **2AP**-labeled riboswitches show folding- and SAM binding-induced conformational rearrangements of the RNA structure. Adapted from [181].

2AP has also been used to analyse the folding pathways and kinetics of short DNA and RNA hairpins [182], pseudoknots [183], and G-quadruplexes [184–186]. Combinations of fluorescence kinetics measurements with temperature jump and computer simulation data can distinguish alternative folding pathways for a sequence of interest [183]. Selection of appropriate labelling positions providing reasonable fluorescence intensity changes upon folding is a crucial aspect of the experimental design [182]. This selection is challenging for relatively long structured RNAs, especially in the absence of reliable structural data. To facilitate this process, Micura and co-authors have developed a universal approach applicable to a broad range of RNAs [181]. They used selective 2′-hydroxyl acylation analysed by primer extension (SHAPE) probing [187,188], which measures the relative local conformational mobility of every nucleotide position in an RNA of interest, as a reliable predictor for the efficiency of **2AP** labelling. Using the S-adenosylmethionine–responsive riboswitch SAM-II [189] as a model, they showed that the positions selected on the basis on their SHAPE reactivity gave the best fluorescence response upon conformational transitions of the riboswitch. This allowed direct observations of the Mg^{2+}-induced RNA folding and S-adenosylmethionine binding steps [181] (Figure 6).

Probing binding of small ligands to RNA. Short structured RNA motifs can specifically interact with small-molecule and peptide ligands, and thus, emerge as new therapeutic targets [190–192]. Position-specific labelling with **2AP** can be used to directly monitor these binding reactions in solution. This has been notably demonstrated for the binding of ligands to riboswitches [181], antibiotics to the decoding site of bacterial ribosomes [193], and small-molecule and peptide ligands to structured viral

RNA elements,[194,195] to name a few. Their dissociation constant (K_d) can directly be extracted from fluorescence titration curves. Therefore, **2AP**-based fluorescence binding assays can be used for discovering and optimizing RNA-targeting small molecules and peptides.

Probing interaction with nucleic acid-binding proteins. Typically, interacting proteins affect the overall conformation of their nucleic acid targets, and thus alter the effective stacking of nucleobases. Therefore, fluorescence measurements with of **2AP**-labelled sequences constitutes an important method for investigating the binding of DNA- and RNA-binding proteins to their targets.

2AP has notably been used to monitor the binding of EcoRI DNA methyltransferase to its target. This enzyme flips out target adenine residues to perform an enzymatic methylation. When a target adenine in a model dsDNA was replaced by **2AP**, the increase in **2AP** fluorescence intensity as well as the blue shift of its emission band showed that the base of interest was flipped into the hydrophobic pocket of the enzyme [196]. This increase of **2AP** fluorescence intensity resulting from a base-flipping event was also used with other methyltransferases [197]. Fluorescence lifetime measurements were also used to study the base flipping induced by the M.HhaI and M.TaqI methyltransferases [112,198]. For M.HhaI-induced base flipping, the fluorescence response strongly depended on the **2AP** position relative to the active site of the enzyme. The fluorescence decay of **2AP** located outside of the active site was multiexponential, requiring the classical set of four lifetime components characteristic of **2AP** within dsDNA. In contrast, for **2AP** within the enzyme recognition site, base flipping resulted in the disappearance of the shortest component (< 100 ps) and a large increase of the amplitude of the longest component (~ 10 ns). This decreased **2AP** quenching was attributed to the loss of stacking with the flanking nucleobases. The response for base flipping was consistent in crystals and solution, confirming that base stacking is the main factor affecting the fluorescence decay of **2AP**-labeled nucleic acids [112].

For M.TaqI-induced base flipping, high-resolution X-ray diffraction data of the native and **2AP**-labeled duplexes revealed a nearly identical configuration of **2AP** and adenine in the active site of the enzyme [198], demonstrating the ability of **2AP** to accurately mimic natural purines in DNA–protein interactions. The M.TaqI-induced base flipping resulted in a loss of the shortest lifetime component, which was more pronounced than in the case of M.HhaI. However, **2AP** in the active site of M.TaqI depended on additional quenching factors, such as the putative π-stacking-based quenching with a tyrosine residue. In addition, acrylamide quenching experiments indicated that a fraction of the extra-helical **2AP** population was located outside of the enzyme binding pocket and was attributed to a putative base flipping intermediate.

DNA hairpins play an essential role in the replication of retroviruses. For instance, (-)PBS, the primary binding sequence of the HIV-1 genome, undergoes multistep conformational rearrangements chaperoned by the nucleocapsid protein (NC) in the course of the reverse transcription of the viral genome. **2AP** has been instrumental for studying the dynamics of the (-)PBS hairpin and the nucleic acid chaperoning function of the NC. Time-resolved fluorescence spectroscopy has been used in combination with MD simulations to characterize the conformational heterogeneity of isolated (-)PBS hairpins, the kinetics of their interactions with the NC protein and the conformational intermediates of the annealing reaction [118,157].

2.2. Nucleobase analogues of the thieno[3,4-d]-pyrimidine family

Atomic mutagenesis has become one of the most effective strategy to improve the photophysical properties of the natural non-emissive purine nucleobases. When the heteroaromatic nature of the five-membered ring of purines was changed, a series of analogues with dramatically improved fluorescence properties was obtained. The synthesized FNAs closely resembled the corresponding natural purines in terms of molecular geometry, van der Waals volume, and ability to form Watson-Crick base pairs. This strategy has been successfully implemented by Tor's group. While the first generation of purine analogues based on thieno[3,2-d]pyrimidine derivatives exhibited moderate Φ_F values (0.037 – 0.058) [199], the newer generations based on thieno[3,4-d]-pyrimidine and isothiazolo[4,3-d]-pyrimidine derivatives, and named respectively **thN** and **^{tz}N**, were particularly promising [200,201]. Both the **thN** and **^{tz}N** families (Figure 7) were initially developed as ribonucleosides. Later, **thG** was also successfully synthesized as a 2'-deoxyribonucleoside [117,202]. Both modified nucleoside families were successfully incorporated into DNA and RNA oligonucleotides and used in a number of applications.

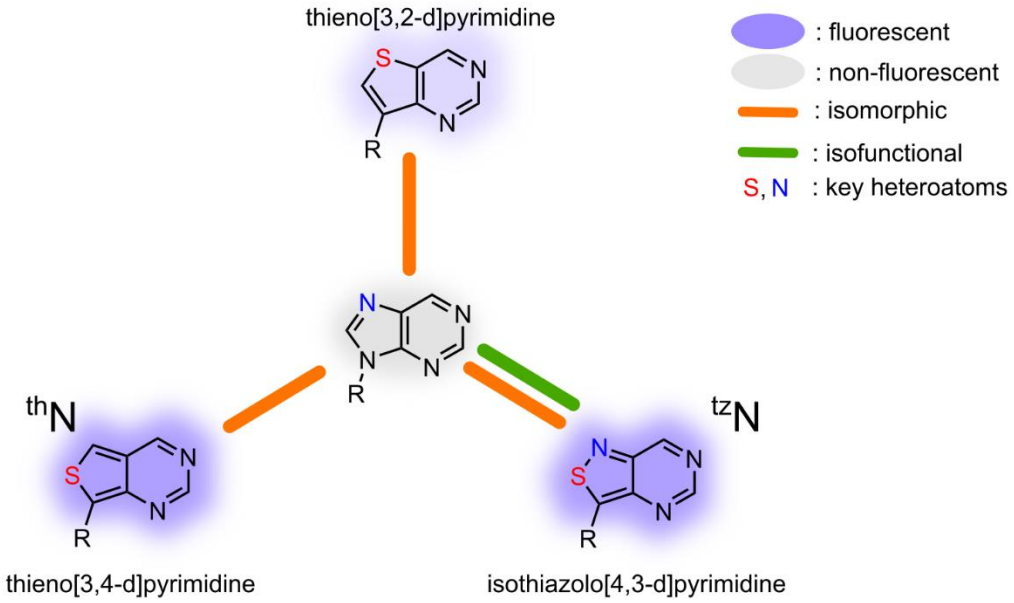

Figure 7. Three classes of isomorphic emissive purine analogues resulting from chemical mutagenesis of the natural purine (in the centre). R = ribose or 2'-deoxyribose (Adapted from [199,201]).

2.2.1. Fluorescent nucleobases of the thieno[3,4-d]-pyrimidine family

The thieno[3,4-d]-pyrimidine heterocycle can be viewed both as a 5,6-modified pyrimidine and a purine mimic. Glycosylation of the heterocycle with D-ribose at the N3 position (corresponds to N1 in the pyrimidine numbering) yields expanded pyrimidine nucleoside analogues (see section 3). At the same time, glycosylation at the position C9 of the heterocycle (corresponds to the position C2 of the thiophene ring) yields isomorphic purine nucleoside analogues. Appropriate modifications of the H-bonding interface of the heterocycle resulted in a complete set of nucleoside analogues (an RNA alphabet, Figure 8). The ribonucleoside analogues **thA**, **thG**, **thC**, and **thU** are all able to form Watson-Crick base pairs with appropriate partners. Note that **thC**, and **thU** actually belong to the class of the expanded pyrimidine analogues, but it is more convenient to discuss their photophysical properties together with the isomorphic purine analogues. The properties and applications of **thU** and **thC** will be also discussed in section 3.1.4.

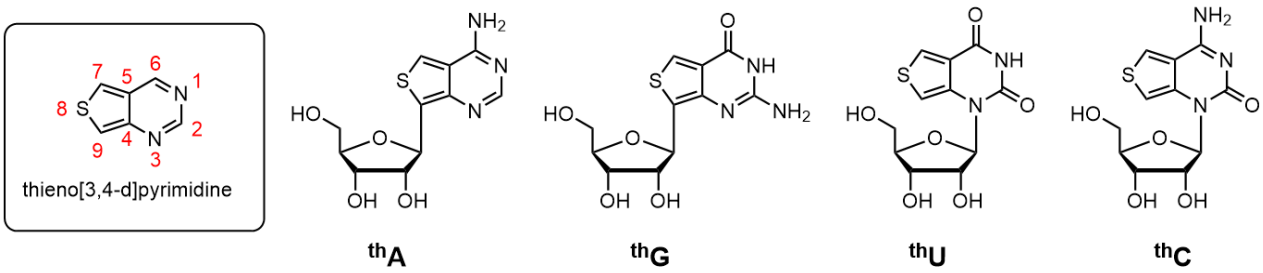

Figure 8. Chemical structure of the thieno[3,4-d]-pyrimidine core and the emissive RNA alphabet derived from it. Conventional purine numbering used throughout this review is shown in red. Note that a different numbering scheme based on pyrimidines was used in the original publication.

The photophysical properties of the **thN** family are summarized in Tables 1 and 3. The **thN** ribonucleoside analogues in water have an absorption maximum ranging from 304 nm (**thU**) to 341 nm (**thA**). Their red-shifted absorption with respect to the native nucleobases is a direct consequence of the inclusion of the S-heteroatom, in which the $3p_x$ electrons conjugate poorly with the proximal C-centred $2p_x$ electrons in the thiophene moiety. This leads to a poorly conjugated π-system, and thus, less stable π orbitals as compared to the native nucleobases. The destabilization of the π orbitals is directly connected to an even larger stabilization of the π* orbitals, which reduces the energy gap between the ground and excited states at the origin of the red-shift[203]. Their molar absorption coefficient at the absorbance maximum ranges from 3160 (**thU**) to 7440 (**thA**) $M^{-1}\cdot cm^{-1}$. Their fluorescence maxima range between 409 to 453 nm, with the most red-shifted derivative being **thG**. The emission of all **thN** derivatives exhibits sensitivity to polarity, as shown by the dependence of their Stokes shifts on the $E_T(30)$ solvent polarity values. The most solvent-sensitive derivative is **thG**. Importantly, all **thN** derivatives are well fluorescent in aqueous buffers, with Φ_F ranging from 0.21 (**thA**) to 0.46 (**thG**). Of note, the exact Φ_F value of **thG** is still debated since somewhat higher values (0.49 and 0.58) have been reported later [202,204]. Interestingly, the Φ_F values of both **thC** and **thU** dramatically drop in less polar solvents, while the Φ_F of **thG** is even higher in less polar solvent than in water [200,202]. Correlated with Φ_F, the observed fluorescence lifetime of **thA** is 3.9 ns, the lowest value among the **thN** family. For the three other nucleosides, the observed lifetimes are above ten nanoseconds, reaching values as high as 15.2 ns for **thC** [200] and 20.5 ns for **thG** [202,204]. The combination of their relatively high molar absorption coefficients with their good Φ_F results in a high brightness ($\varepsilon \bullet \Phi_F$) in water, ranging from ~1300 (**thU**) to ~2000 $M^{-1}\cdot cm^{-1}$ (**thG**). Among the **thN** nucleosides, **thG** stands out by combining high structural similarity with guanine, perfect thermodynamic stability of labelled duplexes, high brightness, high sensitivity of its emission maximum but not of its Φ_F to environmental polarity and exceptionally high fluorescence lifetime values. This stimulated thorough characterisation of **thG** photophysics in search for an isomorphic fluorescent guanosine analogue with outstanding properties.

Table 1. Steady-state and time-resolved fluorescence properties of the isomorphic FNAs

Nucleoside analogue	$\lambda_{abs,m}$ (nm)			ε (10^{-3} M^{-1} cm^{-1})	$\lambda_{F,m}$ (nm)			Φ_F			Lifetime (ns)			References
	nucleoside	ssNA	dsNA	nucleoside	nucleoside	ssNA	dsNA	nucleoside	ssNA	dsNA	nucleoside	ssNA	dsNA	
2AP	303	303	303	3.6	370	370	370	0.68	0.02–0.09	10^{-3} – 0.02	10.2	≤0.1 0.4–1 2–4 7–10	≤0.1 0.4–1 2–4 7–10	103,104
thG	313–334[a]	313–334[a]	330–350	4.15	400–470[a]	400–470[a]	453–463	0.46–0.58	0.10[b]	0.07–0.48	12.8–20.5[a]	0.5[b] 2.8 12.3	1–5.5 9–29	117,200,202,204,205
thA	341	341	341	7.44	453	453	453	0.21	0.07[b]	0.07[b]	3.9	n.d.	n.d.	200,206
^{tz}A	338	n.d.	n.d.	7.79	410	n.d.	n.d.	0.053	n.d.	n.d.	n.d.	n.d.	n.d.	201,207
^{tz}G	333	n.d.	n.d.	4.87	459	n.d.	n.d.	0.247	n.d.	n.d.	n.d.	n.d.	n.d.	201,207
^{tz}I	316	n.d.	n.d.	7.63	377	n.d.	n.d.	0.01	n.d.	n.d.	n.d.	n.d.	n.d.	201,207

^{tz}isoG	334	n.d.	n.d.	9.14	413	n.d.	n.d.	0.05	n.d.	n.d.	n.d.	n.d.	n.d.	[201,207]
^{tz}X	321	n.d.	n.d.	6.61	472	n.d.	n.d.	0.04	n.d.	n.d.	n.d.	n.d.	n.d.	[201,207]
^{tz}2-AA	346	n.d.	n.d.	2.85	447	n.d.	n.d.	0.27	n.d.	n.d.	n.d.	n.d.	n.d.	[201,207]
8vA	290	n.d.	n.d.	12.6	382	388	383	0.65	0.01–0.04	0.005–0.03	4.7	0.05–0.2 0.2–0.9 1.2–2.0	0.02–0.15 0.15–0.4 0.6–1.4 3–4	[208,209]
8vG	260	n.d.	n.d.	0.72	396			0.72	0.06–0.08	0.03–0.07	4.35	n.d.	n.d.	[210,211]
3MI	350	350	350	4.13	430	430	430	0.88	0.03[c] – 0.3[d]	0.03[c] – 0.28[d]	6.5	2.3–2.5 5.2–6.1	0.21 1.9–2.9 5.3–6.3	[212–214]
6MI	340	340	340	11.0	431	431	431	0.70	0.03[c] – 0.3[d]	0.03[c] – 0.25[d]	6.35	0.30 2.2–3.5 7.0–7.3	0.21–0.26 1.0–2.2 5.9–6.3	[212–214]
6MAP	310	329	329	8.51	430	430	430	0.39	0.07[c] – 0.16[d]	0.05[c] – 0.04[d]	3.8	0.17–0.69 1.1–2.8 2.6	n.d.	[213,215,216]
DMAP	333	330	330	8.9	430	430	430	0.48	> 0.01[c] – 0.10[d]	> 0.01[c] – 0.11[d]	4.8	0.28–0.77 2.5–2.8	n.d.	[213,215,217]

dm^{5}K	315	323–326	n.d.	n.d.	385	390–392	n.d.	n.d.	n.d.	n.d.	4	0.6–4	n.d.	218,219
^{FU}dU	316	320	315	11.0	431	434	430	0. 03	n.d.	n.d.	n.d.	n.d.	n.d.	220–222
^{SE}dU	324	n.d.	n.d.	n.d.	452	n.d.	445	0.01	n.d.	n.d.	0.16 3.25	n.d.	n.d.	223
$^{TH\text{-}6A}$dU	332	n.d.	n.d.	11.0	455	n.d.	n.d.	0.2	n.d.	n.d.	4.9	n.d.	n.d.	224
$^{FU\text{-}6A}$dU	320	n.d.	n.d.	n.d.	443	n.d.	n.d.	0.05	n.d.	n.d.	3.1[a]	n.d.	n.d.	224

[a] The range of absorbance and emission wavelength maxima and lifetimes in the free nucleoside and ssDNA is a result of the presence of two tautomers.

[b] Only reported for one sequence.

[c]: Highly quenched environment by flanking purines.

[d]: Moderately quenched environment by flanking pyrimidines.

2.2.2. Spectral properties of thienoguanosine (thG) as a free nucleoside

Existence of tautomers. In the seminal paper of Tor's group on **thG** [200], the authors reported an absorption maximum at 321 nm and an emission maximum at 453 nm in water. Later, careful examination of the fluorescence spectra of **thG** in water revealed a progressive shift of the fluorescence to the red when the excitation wavelength was gradually increased from 300 to 380 nm [225]. Combined analysis of the absorption, excitation, and fluorescence spectra revealed that **thG** exists as two ground-state tautomers in equilibrium, with distinct spectra (Figure 9A). The blue-shifted tautomer ($\lambda_{abs,m}$ = 313 nm ε = 4600 $M^{-1}\cdot cm^{-1}$ and $\lambda_{F,m}$ = 400 nm) and red-shifted tautomer ($\lambda_{abs,m}$ = 334 nm, ε = 4530 $M^{-1}\cdot cm^{-1}$ and $\lambda_{F,m}$ = 470 nm) can be investigated individually by properly selecting excitation and emission wavelengths, e.g., the red-shifted tautomer is selectively excited at wavelengths > 350 nm. Both tautomers have a Φ_F of 0.49 [225], but differ in their fluorescence lifetime, which is 12.8 and 20.5 ns for the blue- and red-shifted tautomer, respectively [204]. The red-shifted and blue-shifted species were unambiguously identified by quantum mechanical (QM) calculations as being the keto-amino **thG**-H1 and **thG**-H3 tautomers, respectively (Figure 9B). Alternative tautomeric forms (imino and OH tautomers, not shown) were concluded to be far less stable in water, so that their populations should be marginal [204].

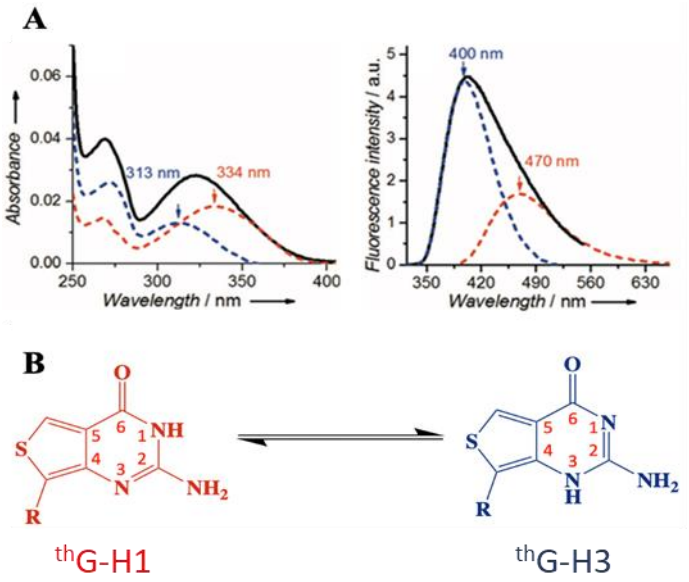

Figure 9. Spectra and structure of the two major tautomers of **thG** . (A) Deconvoluted absorption (left) and emission (right) spectra of the two **thG** tautomers. The excitation wavelength for the fluorescence spectrum was 283 nm. (B) Chemical structures of the **thG**-H1 (left) and **thG**-H3 (right) tautomers (R = ribose or 2′-deoxy- ribose) at the origin of the spectra drawn in red and blue, respectively in panel A (Reproduced from [225] with permission from John Wiley and Sons).

While the two tautomers existed in similar concentrations in water, the population of the H3 tautomer gradually decreased in alcohols when going from methanol to butanol and was fully absent in aprotic solvent [225]. QM calculations of nucleobase analogues methylated at the position C9 rationalized that the linear dependence of the relative concentration of the H3 tautomer with the solvent proticity resulted from the stabilization of the H3 tautomer by H-bond donors. The increased stabilization of **thG**-H3 compared to **thG**-H1 was attributed to its larger dipole moment, as a result of a dipolar resonance structure with a formal charge on the carbonyl ($C6^{+}$–$O6^{-}$) and amino ($C2{=}NH2^{+}$) groups

on the pyrimidine ring. This charge separated **thG**-H3 resonance structure stabilized by H-bond donor solvent molecules was prevented in **thG**-H1 by the proton on the N1 atom. The existence of two tautomers and their different sensitivity to solvent proticity largely contributed to the exquisite sensitivity of **thG** fluorescence to its environment.

In methanol, the **thG**-H3 and **thG**-H1 tautomers were characterized by a fluorescence lifetime of 8.2 ns and 14.5 ns, respectively. In ethanol, these lifetimes shifted to 6.8 ns and 16.0 ns, respectively [204]. However, in both solvents, an additional minor 2.75 ns component at the blue emission side and a 0.5 ns rise component at the red edge were observed, indicating an excited-state reaction.

QM calculations showed that the first excited state (S_1) of the **thG**-H1 tautomer in water corresponded to a $\pi_1\pi^*$ excitation, which can be described as a HOMO→LUMO transition (Figure 10). A second excited state S_2 (corresponding to the second absorption band) as well as a low-lying $n_O\pi^*$ excited state, involving the lone pair of the oxygen atom, were also predicted. For **thG**-H3, the energy of its $\pi_1\pi^*$ state was blue-shifted by ~0.4 eV as compared to **thG**-H1. The S_1 and S_2 excited states were closer in energy in **thG**-H3 as compared to **thG**-H1, and the S_2 excited state of **thG**-H3 was twice as intense as S_1. The $n_O\pi^*$ state was stabilized in the **thG**-H3 tautomer. While the $n_O\pi^*$ state was degenerated (0.06 eV) with the $\pi_1\pi^*$ excited state in dioxane, it was close in energy to S_2 in H-bonding solvents. Since the $n_O\pi^*$ state is much closer in energy to the S_1 state in **thG**-H3 as compared to **thG**-H1, its involvement in the photoactivated dynamics is much more likely for the **thG**-H3 tautomer.

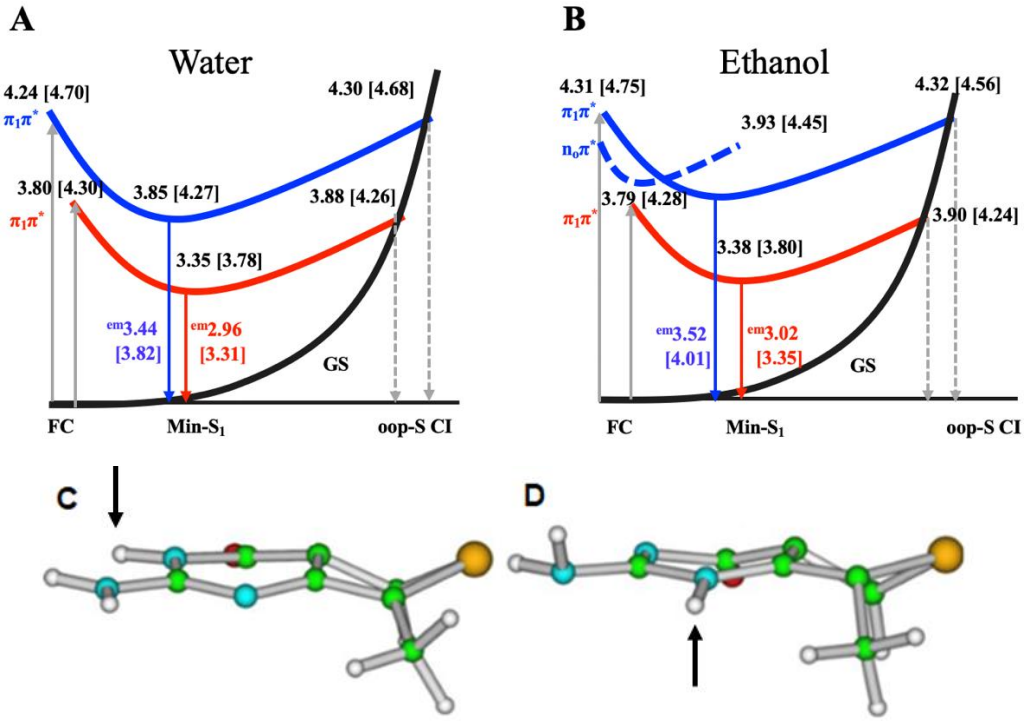

Figure 10. De-excitation pathways of **thG**-H1 (in red) and **thG**-H3 (in blue) in water (A) and ethanol (B). Energies expressed in eV are calculated with TD-PBE0/6-31G(d) and [TD-M052X/6-31G(d)] levels relative to the ground state at the Franck-Condon region. Scheme of the conical intersection CI conformations, computed for **thG**-H1 (C) and **thG**-H3 (D) tautomers. The CI of both tautomers is characterized by an out-of-plane distortion of the sulphur atom (oop-S). The C, N and S atoms are in green, blue and orange, respectively. The H1 and H3 atoms are shown by arrows. The sugar at the

C9 atom was replaced by a methyl group in the calculations (Reproduced from [204] with permission from John Wiley and Sons).

QM calculations further predicted that the **thG**-H1 excited-state $\pi_1\pi^*$ decayed to a minimum in the first adiabatic state with relaxed S_1 geometry (min-$\pi_1\pi^*$), where the N1-C6 and C7-S8 distances were longer by +0.06 Å compared to the S_0 minimum. In water, the calculated vertical emission energy for the min-$\pi_1\pi^*$ was 2.96 eV (PBE0) or 3.31 eV (M052X) (Figure 10A). Both the computed and experimental Stokes shifts were not significantly affected by a change of the solvent from water to ethanol (Figure 10A and 10B). Geometry optimization of the **thG**-H3 $\pi_1\pi^*$ excited state revealed an increase of the C7-S8 distance in the min-$\pi_1\pi^*$ compared to the S_0 minimum. In water, the min-$\pi_1\pi^*$ was very stable and blue-shifted as compared to **thG**-H1. Calculated fluorescence lifetimes were in good agreement with the experimental lifetimes, as well as the 30% shorter radiative lifetime of **thG**-H3 as compared to **thG**-H1. Calculations also identified the lowest energy CI between min-$\pi_1\pi^*$ and the ground state S_0 (Figure 10). The CI conformation for both tautomers was characterized by an out-of-plane distortion of the sulphur atom (oop-S) with a dihedral angle ($dC_9S_8C_7C_5$) of –130° (Figure 10 C,D). The substantial 0.4–0.5 eV energy barrier between the min-$\pi_1\pi^*$ and the CI for the two tautomers in water that had to be overcome to allow the non-radiative de-excitation of min-$\pi_1\pi^*$ through this oop-S channel (Figure 10A), explained well the high Φ_F of **thG**.

In ethanol, the deactivation mechanism of **thG**-H1 was similar to that in water with an energy barrier of 0.44 eV to the CI. For **thG**-H3, the min-$\pi_1\pi^*$ was destabilized, leading to a lower energy barrier (0.1 eV) to the CI. QM calculations further revealed that the imino tautomer of **thG** was more stable than the **thG**-H3 tautomer in alcohols. This imino tautomer could be an intermediate toward the excited-state minimum of the **thG**-H1 tautomer and explain the two additional fluorescence lifetimes (0.5 and 2.75 ns) observed for the fluorescence decay of **thG** in alcohols.

In conclusion, although **thG** and guanine are isomorphic, their electronic structures differ, so that the C_2-NH_2 out-of-plane motion responsible of the non-radiative decay in photoexcited guanine is far less effective in **thG**, explaining the high Φ_F of **thG**. Moreover, in contrast to **2AP** (section 2.1), there is no close-lying dark excited state in **thG** which could decrease its Φ_F. Noteworthy, a similar deactivation mechanism with a significant energy barrier between the min-$\pi_1\pi^*$ and the CI was reported for **thA**, so that its excited population mainly relaxed from the min-$\pi_1\pi^*$ state by emitting fluorescence [203].

pH dependence. Steady-state and time-resolved fluorescence spectroscopy together with theoretical calculations have also been used to describe the ground-state and excited-state equilibria of the unsubstituted **thG** nucleobase as a function of pH in aqueous solution [226]. The **thG**-H1 and **thG**-H3 tautomers are the only emitting species in the 6–9 pH range (Figures 11A and B). They are in equilibrium (K=[H1]/[H3]=1.3) in the ground-state and do not interconvert during their fluorescence lifetime. Moreover, their interconversion does not involve any net protonation or deprotonation step. At basic pH, in their ground-state, both tautomers deprotonate (pK_a = 10.2) into $[^{th}G_b]^-$, which has an absorbance maximum at 340 nm, (ε = 5950 $M^{-1}\cdot cm^{-1}$) with Φ_F = 0.49 and a fluorescence lifetime of 13.5 ns. At acidic pH, the ground-state H1- and H3-tautomers convert (pK_a = 4.4) into the protonated $[H1\text{-}H3]^+$ species, which has an absorbance maximum of 309 nm with an absorption coefficient of 4210 $M^{-1}\cdot cm^{-1}$. In contrast to neutral and basic pH, the fluorescence decays of **thG** at acidic pH were complex, showing rise components typical for excited-state reactions. In the excited-state, $[H1\text{-}H3]^{*+}$ is thought to convert to $[H1\text{-}OH]^{*+}$ through an excited-state intramolecular proton transfer (ESIPT) reaction, before being converted into H1* through a water-mediated ESPT reaction. The $[H1\text{-}OH]^{*+}$

species is fluorescent and more stable than $[H1\text{-}H3]^{+*}$, so that it is likely the main contributor to the fluorescence spectrum with a calculated maximum emission wavelength at 479 nm, in full line with the observed value (482 nm) at pH 2. A fluorescence lifetime of 0.8 ns is attributed to this species. This low lifetime value is in line with the observed Φ_F decrease from 0.49 at pH 5 to 0.11 at pH 2. Of note, the pKa* (1.0) in the excited-state is substantially lower than the pKa (4.4) in the ground-state.

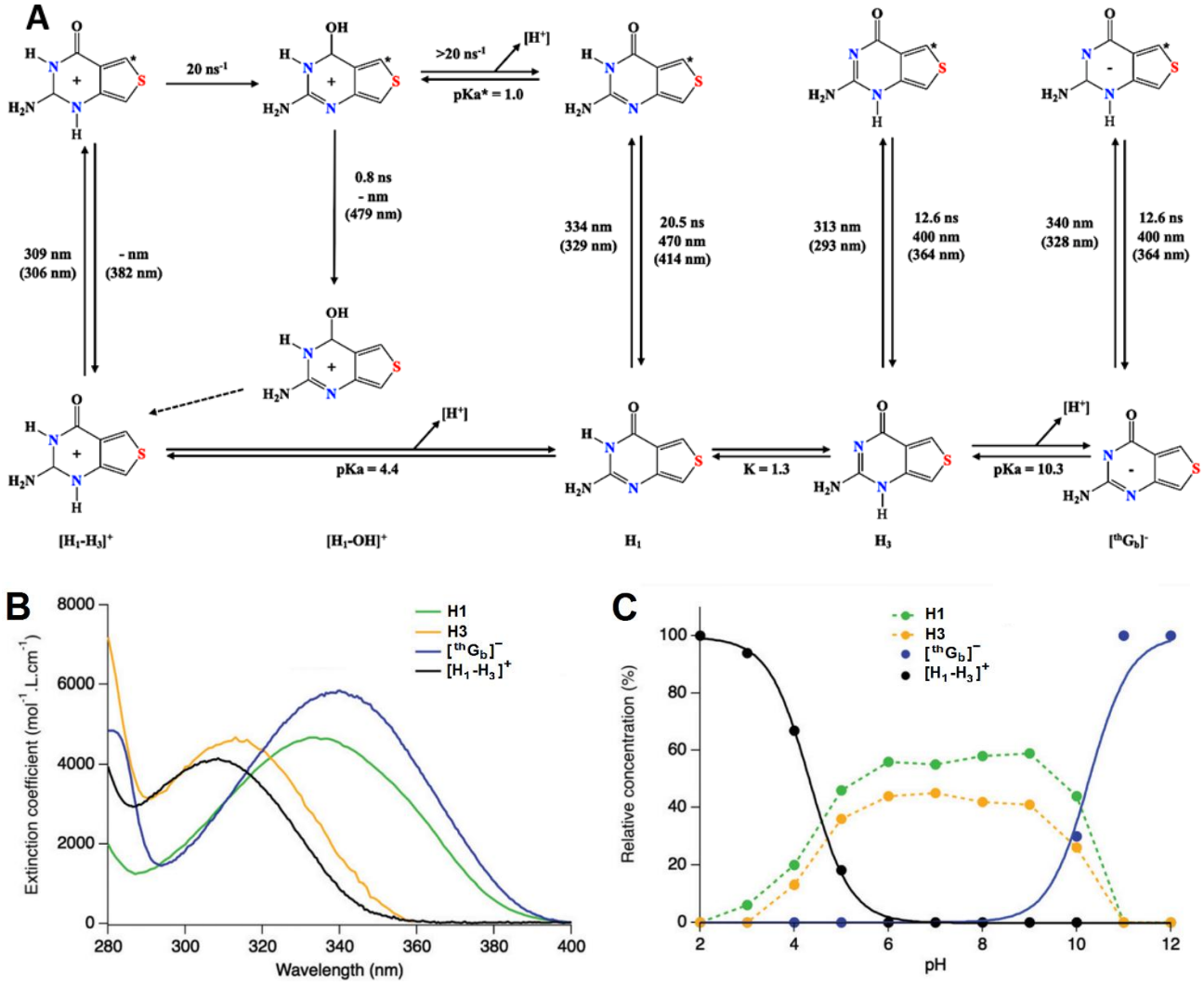

Figure 11. (A) pH-dependent photophysical processes of **thG** in aqueous solutions studied via the unsubstituted base. The ground state and excited state species are displayed in the bottom and the top rows, respectively. The measured wavelengths are reported for each species except for $[H1\text{-}H3]^+$ and $[H1\text{-}OH]^+$, which could not be measured. The theoretical values are reported in parentheses. Absorption spectra (B) and relative molar fractions (C) of the four different **thG** species in the ground state (Reproduced from [226] with permission from Royal Society of Chemistry).

2.2.3. Fluorescence properties of thG in oligonucleotides

The **thG** probe was successfully incorporated into 15–20 nt RNA and DNA oligonucleotides (ONs) using the standard solid-phase phosphoramidite method.[117,200,202,227] For longer RNA sequences, in vitro transcription using T7 RNA polymerase and **thGTP**, as a GTP surrogate, was used to produce

RNA sequences with **thG** replacing all guanosine residues [228] or selectively a single guanosine residue [229]. The single-labelling approach was recently improved using a mutant T7 RNA polymerase, bearing the P266L mutation, that allowed using lower **thG**:GTP ratios (10:1 vs 100:1 used for the wild type) and reduced the number of aborted sequences [229,230]. Moreover, a one-pot approach combining transcription by T7RNAP P266L and post-transcriptional clean-up by a polyphosphatase and an exonuclease, further led to near-homogeneous 5'-**thG**-modified transcripts. Long **thG**-labelled DNA sequences have also been obtained by Sugiyama and co-workers using primer extension and polymerase chain reaction with a number of DNA polymerases (KF polymerase, KOD polymerase, One Taq and Deep Vent) in the presence of **thdGTP**/dGTP mixtures[231].

In the seminal paper of Tor's group [200], **thG** was included at a central position in a 17-mer RNA ONs and further annealed with a complementary sequence to generate matched and mismatched duplexes. The authors pointed out that even when sandwiched between two flanking guanosines, **thG** was not strongly quenched. A series of more thorough investigations was performed by Mély and co-workers [117,225] who replaced the G7 residue by **thG** in the loop of the HIV-1 (-)PBS DNA stem-loop and in the corresponding (+)/(–)PBS duplex. Using these sequences, they investigated the impact of the ON structure on the H1/H3 tautomer equilibrium and performed a back to back comparison with **2AP** inserted at the same position. The H1/H3 tautomer equilibrium was fully preserved when **thG** was inserted in the (–)PBS loop [225]. Molecular Dynamics (MD) rationalized this observation by showing that the natural guanine base and the two **thG** tautomers have the same behaviour in the loop. In contrast, the equilibrium was almost fully shifted toward the H1 tautomer in duplexes, because only **thG**-H1 can form WC base pairs with the opposite cytosine and therefore stably replace guanine. In mismatched duplexes, **thG**-H3 tautomers could be perceived, but their contribution remained low. To avoid any spectroscopic contribution from this minor tautomer and to study only the photophysical properties of the major **thG**-H1 tautomer, duplexes should be excited at wavelengths ≥350 nm, where only the **thG**-H1 tautomer absorbs.

As compared to the free nucleoside, the Φ_F of **2AP** in (–)PBS was observed to drop by 8-fold in the (–)PBS stem-loop and more than 50-fold in the matched (+)/(–)PBS duplex, reaching a value of about 0.01 [117]. In contrast, the Φ_F of **thG** dropped to 0.1 in the stem-loop but increased to 0.2 in the matched duplex and even to 0.38 in a mismatched duplex (Figure 12). Moreover, the fluorescence maximum shifted by 5 and 12 nm, respectively, relative to the free nucleoside when **thG** was in the stem loop and the duplex, respectively, while **2AP** exhibited almost no wavelength shift, underlying the higher sensitivity of **thG** to the environment as compared to **2AP**.

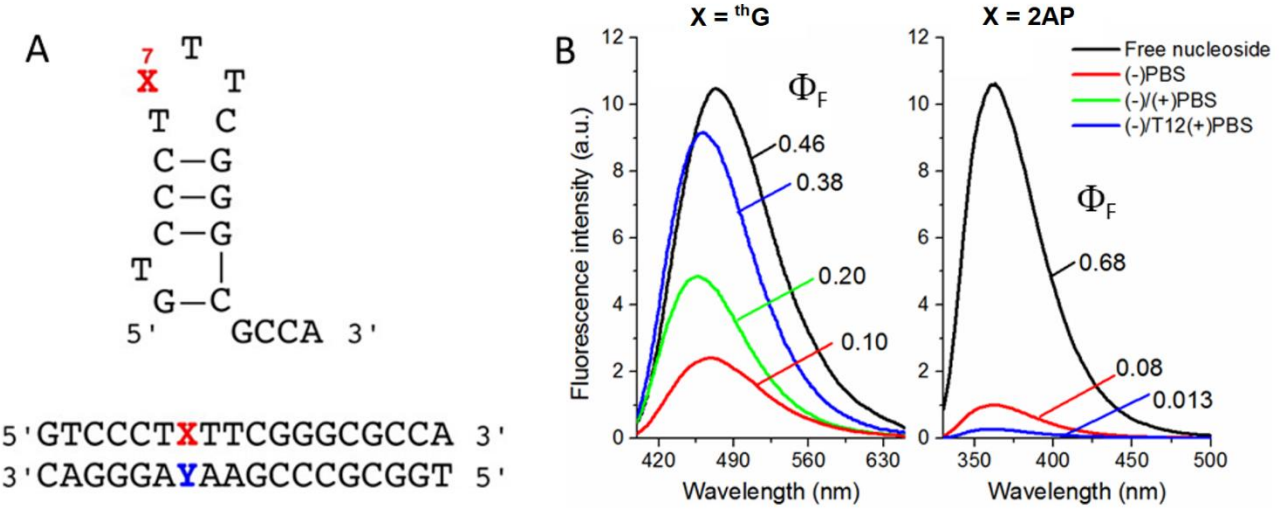

Figure 12. Comparison of **thG**- and **2AP**-labelled (–)PBS oligonucleotides. A) Sequences of (–)PBS stem-loop and (–)/(+)PBS duplex. X corresponds to **thG** or **2AP**. Y corresponds to C or T. B) Spectra of the **thG**- and **2AP**-labelled (–)PBS oligonucleotides. For **thG**, the excitation wavelength was 380 nm. The Φ_F of each species is indicated (Reproduced and adapted from [117] with the permission of ACS publications).

The most remarkable differences between the two probes were seen in their time-resolved fluorescence data. In both the (–)PBS stem-loop and (–)/(+)PBS duplex, **2AP** showed complicated decays with four fluorescence lifetimes and about 50% of dark species with lifetimes shorter than the 30 ps detection limit of the set-up (Table 2). In the duplex, 98% of the **2AP** species had lifetimes shorter than 0.5 ns, making it very difficult to obtain relevant conclusions. In comparison, the fluorescence decays of **thG**-labelled ONs were much simpler with only three lifetimes (0.5 ns, 2.8 ns and 12.3 ns) for the stem-loop and only two (1.1 and 11.3 ns) for the duplex, with no dark species. This limited number of lifetimes observed for **thG** in the duplex was attributed to its ability to form stable WC base pairs and thus, adopt defined conformations. Important differences between the two labels were also observed in the time-resolved fluorescence anisotropy data. Indeed, a rotational correlation time associated with the local mobility of the probe was observed with **2AP**, but not with **thG**. This local mobility was associated with the extrahelical or less stacked conformations of **2AP**, which represent only a small and non-biologically relevant population of conformers but dominate the emission of the **2AP**-labelled ONs. In contrast, only the tumbling motion of the stem-loop (θ_c = 2.4 ns) and duplex (θ_c = 8.1 ns) was perceived with **thG**-labelled ONs, because the major and most emissive **thG** conformations are rigidly maintained in the two structures. Thus, the time-resolved fluorescence spectra of **thG** reliably reflect the predominantly populated conformers and can thus better than **2AP** convey relevant information on the conformations and dynamics of the ONs.

Table 2. Time-resolved fluorescence data of **thG**- and **2AP**-labelled (-)PBS stem-loop and (+)/(-)PBS duplex[a].

stem-loop or duplex	Φ_F	τ_1	α_1	τ_2	α_2	τ_3	α_3	τ_4	α_4	$\langle\tau\rangle$	α_0
d2AP7(-)PBS	0.08	0.15	0.16	0.66	0.10	2.6	0.15	7.4	0.11	2.4	0.48
d2AP7(-) /T12(+)PBS	0.013	0.18	0.20	0.44	0.27	1.4	0.01	5.2	0.01	0.4	0.51
dthG7(-)PBS	0.10			0.5	0.32	2.8	0.40	12.3	0.28	4.7	<0.1
dthG7(-)/(+)PBS	0.20			1.1	0.17			11.3	0.83	9.6	<0.1

[a]Φ_F is the fluorescence quantum yield, τ_i are the fluorescence lifetimes (ns) and α_i are their amplitudes. α_0 is the amplitude of the dark species. $\langle\tau\rangle$ is the mean fluorescence lifetime (ns). Excitation wavelengths were 315 and 370 nm for 2AP and **thG**, respectively. SDs for the lifetimes and amplitudes are <20%. SDs for Φ_F are <10%. Data from [117,205]

To get a full understanding of **thG's** photophysics in DNA, twenty DNA duplexes, where the bases facing and neighbouring **thG** at position 7 in the (–)PBS were systematically varied (Figure 13), were thoroughly studied using fluorescence spectroscopy, MD simulations and mixed quantum mechanical/molecular mechanics (QM/MM) calculations [205]. In matched duplexes, **thG**'s hypochromism due to excitonic coupling was larger for flanking G/C residues (40%) than for A/T residues (30%), but its Φ_F (0.15−0.18) and lifetime values (10.6−12.3 ns for the long-lived lifetime τ_2 and 8.4−10.1 ns for the mean lifetime $<\tau>$) were almost independent of the flanking bases.

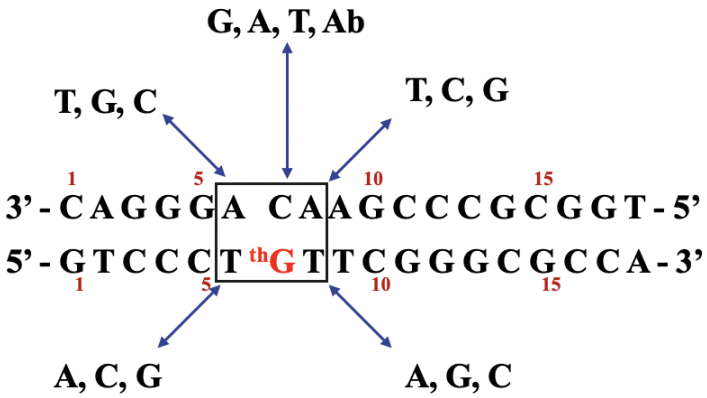

Figure 13. th**G**-labelled (-)/(+)PBS DNA duplexes used to investigate th**G**'s photophysics. The G7 residue of (-)PBS was replaced by th**G**. The flanking and opposite bases were systematically varied. Ab designates an abasic site. The natural flanking thymine residues were systematically changed by two similar residues, and the opposite bases were concomitantly replaced by their matched Watson-Crick residues to minimize duplex destabilization (adapted from [205]).

QM/MM calculations indicated that the high stability of these duplexes placed th**G** with the proper orientation and distance to the flanking bases in order to allow an efficient charge transfer (CT). Thus, in addition to the spectroscopic $\pi\pi^*$ state, essentially localized on th**G**, a relatively close-lying CT transition that differed in energy by less than 1 eV as compared to S_1 was always present. The direction of the CT depended on the flanking base, since th**G** acted as an electron acceptor with guanine and an electron donor with all the other flanking bases (Figure 14). This CT transition opened up a non-emissive, deactivation channel in the matched duplexes, responsible for the decrease in Φ_F and lifetime values as compared to the free nucleoside. Moreover, the relaxed S_1 state min-$\pi_1\pi^*$ also exhibited a low CT character in the ONs, which reduced its oscillator strength and radiative constant with respect to free th**G**-H1, leading to a further decrease in the Φ_F and lifetime values. The poor dependence of the spectroscopic properties of th**G** on its context is a great asset as th**G** can be used to replace any guanine residue in matched DNA and RNA duplexes, while always maintaining consistent photophysical features.

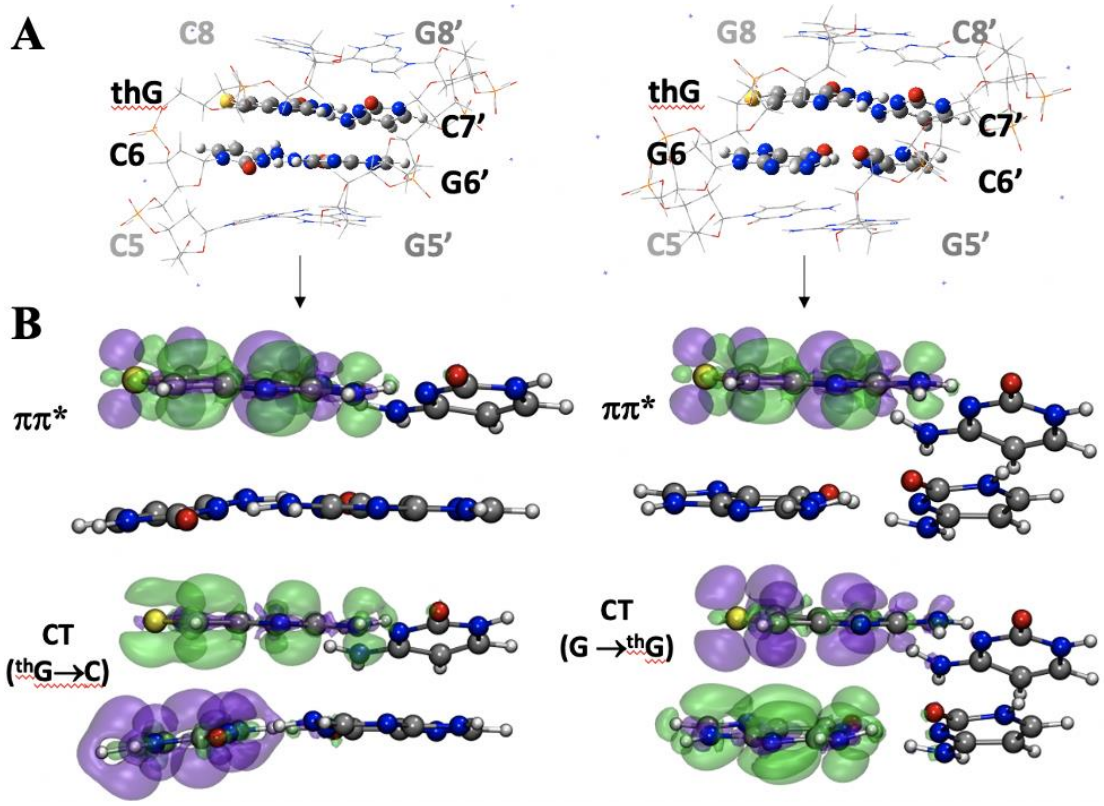

Figure 14. Ground-state geometries (A) and excited-state electronic densities (B) of the **thG**-labelled matched (–)/(+)PBS duplexes. A) Most relevant geometries for the optimized minima at the QM(PCM/M052X/6-31G(d))/MM level of theory. Nucleobases treated at the QM level are in ball and sticks, whereas the MM system is depicted in lines. B) Excited-state density difference, green means negative difference while purple indicates positive difference (Reproduced from [205] with the permission of ACS publications).

In contrast to the matched duplexes, the quantum yield of the mismatched duplexes varied significantly: Φ_F (0.10−0.48). Large changes in the values of the long-lived lifetime (τ_2 = 9.3–28.6 ns) and the mean lifetime (<τ> = 6.3–22.9 ns) were also observed. These changes were highly dependent on the nature of the flanking base, with the following order for **thG** quenching: G>C>T>A that differed from the sequence of oxidation potentials [149]. Thus, quenching was modulated by both electronic and structural effects. The strongest quenching was observed when a CT state (e.g. $^{th}G^{+}$→pyrimidine^{-} or G^{+}→$^{th}G^{-}$) between **thG** and its closest stacked flanking base led to an effective non-radiative decay channel. The efficiency of this non-radiative pathway partly depended on the local stiffness of the duplex, being higher when **thG** was flanked by the more stable G·C base pairs as compared to A·T base pairs. This stiffness was instrumental to preserve a close proximity between **thG** and its closest neighbour. The opposite nucleobase to **thG** played also a role, as MD simulations suggested that an opposite purine created a steric clash, which flipped **thG** toward the solvent and prevented its base stacking with the flanking nucleobases (Figure 15). This explained the poor quenching of **thG** when mismatched with A or G. When **thG** was flanked by A·T base pairs, τ_2 values (23.8–28.6 ns) much higher than the fluorescence lifetime of **thG**-H1 in water (20.5 ns) were observed. These high values could be explained by the low non radiative rate constant k_{nr} value related to the poor quenching environment provided by the A·T base pairs and the low radiative rate constant k_r value, related to the strong hypochromism of **thG** stacked with its neighbours. Finally, the short-lived lifetimes (τ_1 = 2.1–5.5 ns) which were especially abundant in duplexes with an abasic site opposite

to **thG**, were thought to be associated to a dynamic quenching of **thG** by CT to its neighbours, modulated by the conformational fluctuations of DNA.

Obvious applications of the strong dependence of **thG** photophysical properties on its flanking and paired nucleobases, while preserving good Φ_F values, are the detection of single nucleotide polymorphism as well as the monitoring of local structural changes induced by helix-destabilizing, NA-chaperoning, DNA-bending or base-flipping proteins. Another unique feature revealed by this study were the unusually high values of the long-lived excited-state lifetime τ_2 (9–29 ns), which largely exceeded those of the vast majority of small fluorophores, so that selective measurements even in complex mixtures of fluorescently labelled molecules or in autofluorescent media could be envisioned by using a fluorescence lifetime-based or time-gated detection scheme. In conclusion, **thG** appears as an outstanding emissive isomorphic substitute for guanine with good Φ_F, long fluorescence lifetimes and exquisite sensitivity to local structural changes. Moreover, the tautomerism of **thG**, constitutes a useful additional channel to provide local structural and dynamical information on the site of incorporation.

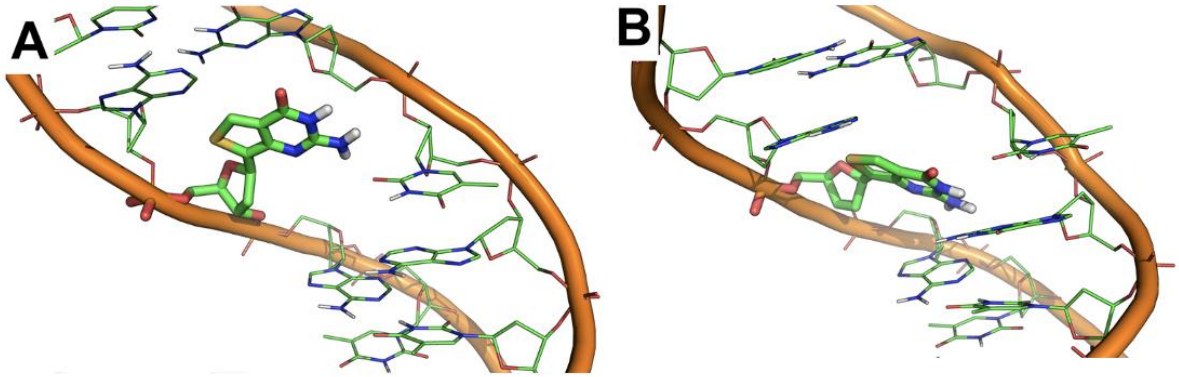

Figure 15. Conformational behaviour of **thG** with opposite A (A) and opposite G (B) in (–)/(+)PBS duplexes. Representative structures from MD simulations are shown. **thG** is shown as sticks, surrounding nucleotides as lines and the phosphate backbone as cartoon. (Reproduced from [205] with the permission of ACS publications).

2.2.4. Applications of thG as a fluorescent reporter in nucleic acids

In one of its first applications, **thG** was used to monitor the transition between B- and Z-DNA using a self-complementary **thG**-labelled dodecamer [202]. Promoting the B–Z transition by high salt concentrations resulted in an observable fluorescence increase. This increase can be explained by the increased distance between bases and the increased rotation by residue in Z-DNA versus B-DNA, which likely decreases the CT between **thG** and its closest neighbour in Z-DNA. A similar fluorescence increase was observed when the same **thG**-labelled B-DNA was titrated by the DNA-binding domain of the double-stranded RNA adenosine deaminase (ADAR1) [202] which promoted the B to Z conversion by specifically binding to the Z form [232].

Taking advantage of the doubling of the Φ_F of the **thG**-labelled (–)PBS stem-loop during annealing with (+)PBS, the annealing reaction was monitored in real time [233]. The reaction was found to follow a two-step process with the formation of an intermediate complex (IC) initiated by the PBS overhangs and stabilized by the complementary loops. The data obtained with the **thG**-labelled (–)PBS were

confirmed by using a (–)PBS sequence labelled in its loop by a chromophoric nucleobase based on 2-thienyl-3-hydroxychromone. The use of the two probes provided the first complete mechanistic description of the (+)/(–)PBS annealing reaction that plays a key role in HIV-1 replication [234]. Moreover, the interaction between the **thG**-labelled (–)PBS and NCp7, a nucleic acid chaperone protein of HIV-1 has been investigated [225]. Addition of NCp7 markedly increased **thG** emission and favoured the H1 tautomer over the H3 one. As NCp7 flipped the substituted G7 residue to the outside of the loop [235], the observed fluorescence increase likely resulted from the increase in the distance between **thG** and its nearest neighbours, preventing the CT pathway.

Moreover, **thG** was also used to monitor the flipping of methylated cytosines (mC) by the "Set and Ring Associated" domain (SRA) of the ubiquitin-like containing PHD and RING finger domains 1 (UHRF1) protein [236]. UHRF1 together with the DNA methyltransferase 1 (DNMT1) plays a key role in the maintenance of the DNA methylation patterns in cells upon proliferation. To monitor the base flipping, the G residue was replaced by **thG** in the CpG motif which is selectively recognized in its hemi-methylated (HM) form by the SRA domain. This substitution was shown to preserve the preferential binding of SRA to HM duplexes over non methylated (NM) duplexes. While only a moderate fluorescence increase accompanied the binding of SRA to NM duplexes, a large fluorescence increase was observed with HM duplexes (Figure 16A). This large fluorescence increase was not observed with a SRA mutant unable to flip mC, indicating that it resulted from the base flip event. This fluorescence increase was related to the loss of stacking between **thG** and mC, as a result of the mC flip. Altogether, the data obtained with the **thG**-labelled duplexes allowed concluding that SRA interacted with the duplexes through a fast “bind-and-slide” mechanism, searching for CpG recognition sites. At these sites, SRA flipped the mC with rate-limiting kinetics that stabilized the binding of SRA and allowed recruiting DNMT1 (Figure 16B).

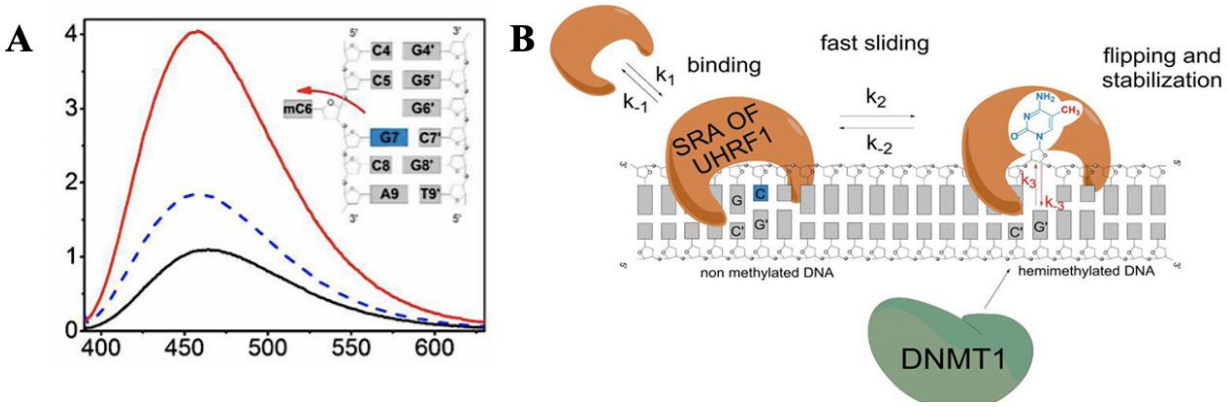

Figure 16. Monitoring mC base flipping by the SRA domain of UHRF1, using **thG**-labelled duplexes. A) Fluorescence spectra of HM and NM duplexes labelled by **thG** at position 7 for an excitation wavelength at 350 nm. Black line: spectra of free labelled duplexes (HM and NM are superimposable), blue line: spectra of NM duplex bound with SRA, red line: spectra of HM duplex bound with SRA. B) Proposed mechanism of base flipping (Reproduced from [236] with the permission of ACS publications).

In addition, **thG** has also been used in combination with another fluorescent nucleoside based on the tricyclic 1,3-diaza-2-oxophenothiazine core (**tC**, section 3.1.2) to measure changes in the distance and orientation between nucleobases in DNA by Förster Resonance Energy Transfer (FRET) [237]. By using duplexes where **tC** and **thG** were at different positions, a decrease in the distance between the two probes changed the emission colour from blue (**thG** emission, low FRET efficiency) to green (**tC**

emission, high FRET efficiency). Duplexes labelled with the two probes were used to monitor B- to Z-DNA transitions. The same **thG-tC** FRET pair was also incorporated in the 145 bp nucleosomal 601 DNA sequence [238]. The doubly labelled DNA was wrapped around a reconstituted nucleosome and the FRET signal was used to investigate the mononucleosome structure [238].

RNA sequences where all guanosine residues or only one specific guanosine were replaced by **thG** have also been annealed to non-labelled RNA sequences to assemble hammerhead ribozymes and monitor their enzymatic activities [228,229]. A large fluorescence increase that allowed real-time monitoring of the cleavage reaction was observed when **thG** specifically replaced the G residue at the cleavage site. The observed fluorescence increase can be easily explained by a decrease in CT efficiency resulting from the loss of one of its flanking nucleobases and the poor stability of the duplex termini.

2.2.5. Applications of thA as a fluorescent reporter

Several applications have also been developed with **thA**. This fluorescent nucleobase surrogate is recognized as a substrate by the adenosine deaminase (ADA), which converts it into **thI**, the fluorescent analogue of inosine [239]. Since in comparison to **thA** the maxima of excitation and fluorescence of **thI** were blue shifted by 20 nm and **thI**'s Φ_F was substantially increased, the **thA→thI** conversion by ADA can be monitored in real time by following the intensity increase at 339 nm, the maximum of emission of **thI**. Moreover, since ADA is a drug target for a number of diseases, this ADA-mediated **thA→thI** fluorogenic reaction can be used to establish a high-throughput screening assay to discover new ADA inhibitors. Noteworthy, though ADA properly recognized **thA**, the **thA→thI** conversion mediated by ADA was markedly slower than the A→I conversion, likely due to the carbon-for-nitrogen substitution at the position 7 of the purine core, which is important for the binding to ADA. The **thA→thI** conversion was also successfully monitored when **thA** was incorporated in the mRNA substrate of ADAR2. In this case, the deamination rates of A and **thA** were fully comparable [240].

2.3. Fluorescent nucleobases of the isothiazolo[4,3-d]-pyrimidine family

One obvious structural flaw of the **thN** family is the lack of the basic N7 atom in the purine skeleton that is essential for a number of intermolecular interactions. To overcome this limitation, Tor's lab has designed and synthesized the **^{tz}N** family based on the isothiazolo[4,3-d]-pyrimidine core that contains the N7 atom present in natural purines (Figure 17) [201]. The molecular geometry of **^{tz}G** and **^{tz}A** obtained from X-ray diffraction analysis closely resembles the natural G and A, indicating that these nucleobases could serve as isomorphic replacements of the natural purines.

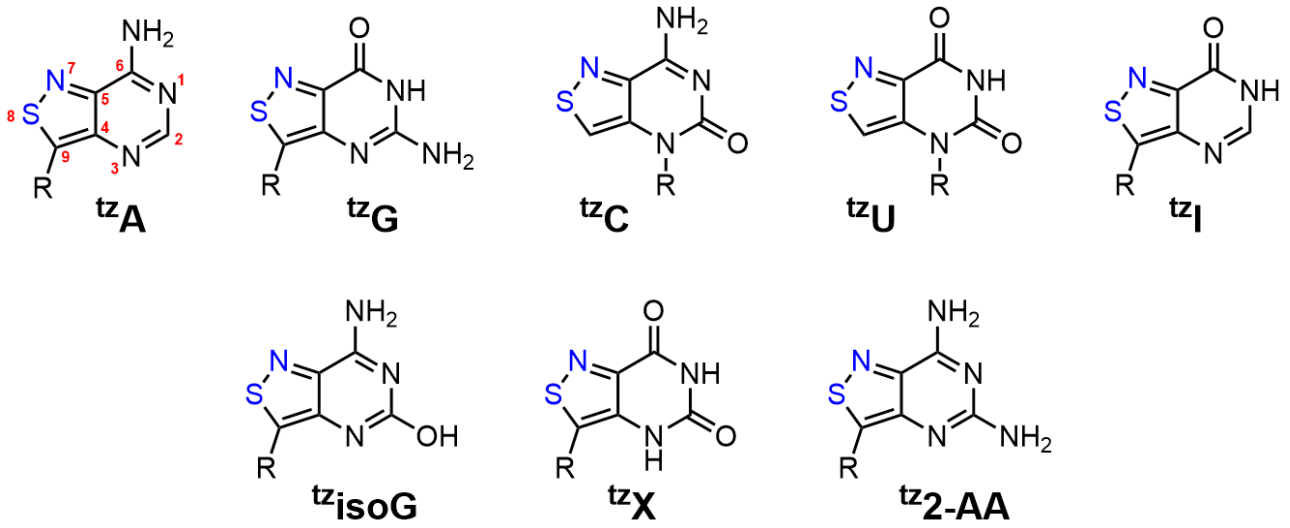

Figure 17. Structure of the isothiazolo[4,3-d]pyrimidine-based nucleoside analogues. Conventional purine numbering used throughout the review is shown in red; the heteroatoms of isothiazole are shown in blue; R = ribose.

2.3.1. Photophysical properties of the ^{tz}N nucleosides

Similar to the th**N** alphabet, the absorbance and emission maxima of the ^{tz}**N** alphabet in water range from 312 to 338 nm and from 377 to 459 nm, respectively (Tables 1 and 3). The absorption coefficients of the two families are comparable as well. However, the two families markedly differ by their Φ_F. While the Φ_Fs of the th**N** family range from 0.21 to 0.49, the Φ_Fs of the ^{tz}**N** family are much lower, being of 0.053 for ^{tz}**A** and ^{tz}**C** and even below 0.01 for ^{tz}**I** and ^{tz}**U**. Only ^{tz}**G** demonstrates a fairly good Φ_F (0.24), although two times lower than the Φ_F of th**G**. As a result, the brightness of the ^{tz}**N** compounds is rather low (40–540 $M^{-1}\cdot cm^{-1}$), with the exception of ^{tz}**G** (1200 $M^{-1}\cdot cm^{-1}$). The ^{tz}**N** compounds also show a linear dependence of their Stokes shifts on solvent polarity $E_T(30)$ values, indicating their responsiveness to the local molecular environment. The ^{tz}**N** compounds are also sensitive to pH changes, generally showing a red shift in their absorption when the pH is increased. These shifts allow determining their pK_a values, which for ^{tz}**A** and ^{tz}**G** correlate well with the pK_a values of adenosine and guanosine, respectively, illustrating the ability of the isothiazole ring to properly mimic the imidazole moiety of the natural purines. The improved mimicking ability of the ^{tz}**N** family over the th**N** family was well illustrated by the similar deamination rates by ADA of ^{tz}**A** and adenosine, while the deamination rate of th**A** was markedly lower. The izothiazole core with its additional nitrogen as compared to the thiophene ring restored thus the native basicity and Hoogsteen H-bonding interface of the native purines. Consequently, in addition to being isomorphic, the ^{tz}**N** family and in particular the ^{tz}**A** and ^{tz}**G** compounds are considered to be "isofunctional" with respect to the native purines [201]. The ^{tz}**N** family has been further expanded by the inclusion of the analogues of isoguanosine (^{tz}**isoG**), xanthosine (^{tz}**X**) and 2-aminoadenosine (^{tz}**2-AA**) [207]. Comparison of their photophysical properties with those of the related ^{tz}**G**, ^{tz}**I** and ^{tz}**A** derivatives provided some insight in the relationship between their structure and photophysical properties. For instance, the 5-fold higher Φ_F and sensitivity to solvent polarity of ^{tz}**G** and ^{tz}**2-AA** as compared to ^{tz}**isoG** and ^{tz}**A** clearly highlighted the key role of the exocyclic amine on the pyrimidine ring to improve the photophysical properties of these compounds. The photophysical properties of the ^{tz}**N** family are summarized in Tables 1 and 3.

QM calculations were performed to further understand the photophysical properties of the ^{tz}**N** derivatives[203,241–243]. The de-excitation pathways were found to be similar to those described for the

thN family, with a competition between the emission from the relaxed S_1 state min-$\pi_1\pi^*$ state and the nonradiative relaxation via a CI with the ground state. For **^{tz}A** and **^{tz}G**, the min-$\pi_1\pi^*$ structures are thought to be mainly planar, while the conical intersection $(\pi_1\pi^*/gs)_{CI}$ between $\pi_1\pi^*$ and electronic ground state is characterized by large distortions of the six-membered ring at the C2 position, with the bonded moiety being perpendicular to the six-membered ring. As a result of these distortions, there is a significant energy barrier to cross in order to reach the $(\pi_1\pi^*/gs)_{CI}$ nuclear configuration, so that emission from the more stable min-$\pi_1\pi^*$ state is favoured[243]. A similar picture applies for both **^{tz}C** and **^{tz}U**, where the $(\pi_1\pi^*/GS)_{CI}$ state cannot be reached, due to steric hindrance of the C5-C6 position, where C5 connects the rings in both molecules[242]. Thus, the origin of the striking differences in the Φ_Fs of the **^{tz}N** and **thN** families remains to be clarified.

2.3.2. Applications of the ^{tz}N nucleosides

As aforementioned, **^{tz}A** was used for real-time monitoring of the **^{tz}A** $\rightarrow$ **^{tz}I** conversion catalyzed by ADA [201]. The deamination activity of ADA was also tested with **^{tz}isoG**, **^{tz}X** and **^{tz}2-AA** to obtain additional insights into the substrate selectivity of ADA [207]. Moreover, **^{tz}A** was also applied as a fluorescent substitute for adenine in the structure of nicotinamide adenine dinucleotides NAD^+ and NADH [244] (Figure 18A). The maximum absorption and emission wavelengths of both substrates **N-^{tz}AD$^+$** and **N-^{tz}ADH** were comparable to those of **^{tz}A**. While Φ_F of **^{tz}A** was only marginally altered in **N-^{tz}AD$^+$** (0.044), it decreased in **N-^{tz}ADH** three-fold (0.015). The difference in Φ_Fs between the two forms was used to follow the **N-^{tz}AD$^+$** $\rightarrow$ **N-^{tz}ADH** conversion by alcohol dehydrogenase (ADH), as well as the back reaction catalyzed by lactate dehydrogenase (LDH) (Figure 18 B, C)). Remarkably, the reaction rate remained unchanged by the substitution, indicating that **^{tz}A** did not alter the recognition by the enzymes. Moreover, both **N-^{tz}AD** forms were also shown to be substrates for the NAD^+ nucleosidase, allowing real-time monitoring of **^{tz}ADP**-ribose release. Recently, **N-^{tz}AD$^+$** has been enzymatically phosphorylated to generate the corresponding **N-^{tz}ADP$^+$** [245]. As for the **N-^{tz}AD$^+$** $\rightarrow$ **N-^{tz}ADH** conversion ,the enzymatic conversion of **N-^{tz}ADP$^+$** to **N-^{tz}ADP** resulted in a decrease of **^{tz}A** fluorescence . Therefore, the **N-^{tz}ADP$^+$**/**N-^{tz}ADP** cycle could be followed in real time by fluorescence spectroscopy. Finally, **N-thAD$^+$** has been compared to **N-^{tz}AD$^+$** as emissive cofactors in a number of enzymatic reactions that play central roles in the cell [206]. **N-thAD$^+$** was found to be substantially more responsive than **N-^{tz}AD$^+$** and thus appeared as a more suited probe for real-time applications. In contrast, **N-thAD$^+$** is less isofunctional than **N-^{tz}AD$^+$**, giving lower conversion rates or increased by-product formation.

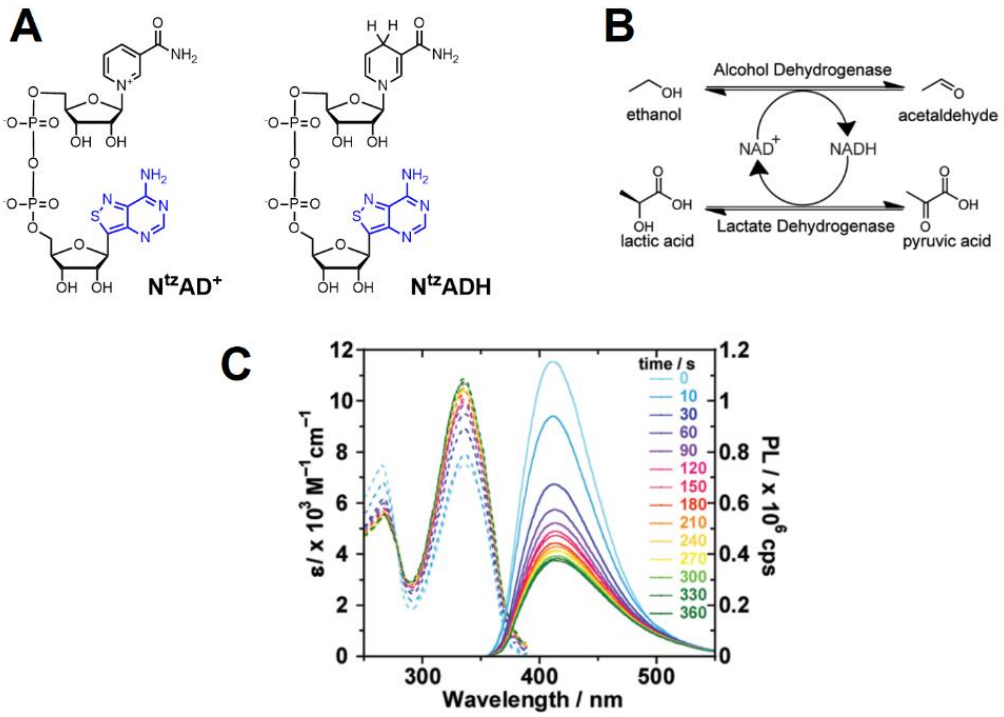

Figure 18. Use of **N-^{tz}AD** to monitor the enzymatic cycle of NAD^+ consumption and regeneration with ADH and LDH. A) Chemical structures of **N-^{tz}AD$^+$** and **N-^{tz}ADH**; the modified nucleobase is shown in blue. B) Enzymatic reactions studied with the fluorescent cofactors. C) Progress of the ADH-mediated oxidation results in consumption of **N-^{tz}AD$^+$**, which can be monitored by a fluorescence intensity decrease (Reproduced from [244] with the permission of ACS publications).

Both **^{tz}G** and **thG** have also been used to monitor the enzymatic synthesis and hydrolysis of cyclic dinucleotides (CDNs), [246]which are key second messengers in bacteria, regulating their cellular physiology and host–pathogen interactions, such as biofilm formation, motility and virulence.[247] Tor and co-workers reported four fluorescent CDN analogues, namely **c-di-^{tz}GMP**, **c-di-thGMP**, **c-G-^{tz}GMP** and **c-G-thGMP**, which were synthesized enzymatically using DncV, a cyclic dinucleotide synthetase from *Vibrio cholerae*. The fluorescence of all studied **^{tz}G-** or **thG**-containing CDNs was quenched, probably due to the stacking with the second aromatic nucleobase (Figure 19A). The Φ_Fs ranged from 0.039 for **c-di-^{tz}GMP** to 0.18 for **c-G-thGMP**. Therefore, synthesis of these CDNs from **thGTP** or **^{tz}GTP** by DncV led to a large drop in fluorescence that helped to monitor the progress of the reaction in real time (Figure 19B). Similarly, the hydrolysis of CDNs by phosphodiesterase could also be followed. For both enzymatic reactions, the kinetic rate constants of the enzymatic reactions were effectively retrieved [246].

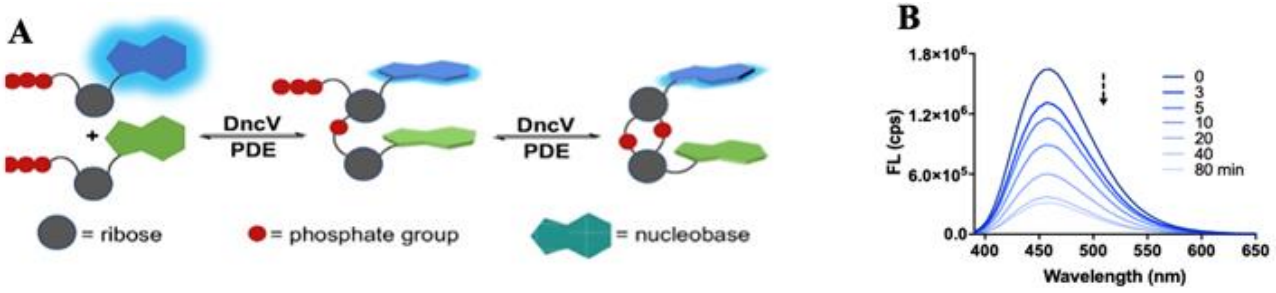

Figure 19. Use of **thG** and **^{tz}G** to monitor the enzymatic synthesis and hydrolysis of CDNs. A) Underlying principle of the fluorescence changes upon CDN formation and hydrolysis. B) DncV-

promoted synthesis of **c-di-^{tz}GMP** as followed from the emission spectra (Reproduced from [246] with the permission from John Wiley and sons).

2.4. 8-vinyl-purines

In order to improve the fluorescence properties of the natural purines, a vinyl group was added at the position 8 of adenine and guanine. This strategy led to two new fluorescent nucleoside analogues, 8-vinyl-adenine (**8vA**) [208] and 8-vinyl-guanine (**8vG**) [210] (Figure 20). These nucleoside analogues preserve the Watson-Crick hydrogen bonding interface of the natural purines, and can be incorporated into ONs using either chemical or enzymatic approaches [208,210,248]. Although the vinyl group at the position 8 slightly influences the *syn-anti* equilibrium of the nucleosides, both analogues can adopt the desired *anti*-conformation when incorporated into dsDNA.

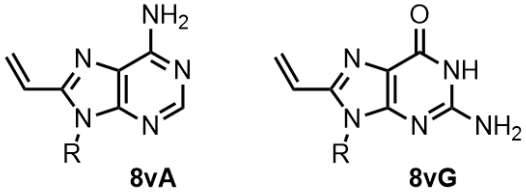

Figure 20. Chemical structures of the 8-vinyl-adenine (8vA) and 8-vinyl-guanine (8vG) nucleosides; R = 2'-deoxyribose.

The thermodynamic stability of the **8vA**-labelled duplexes decreased with the opposite base in the following order: A·T > **8vA**·T > **8vA**·G > **8vA**·A > **8vA**·C [209]. The destabilizing effect depended also on the flanking nucleobases. Duplexes with **8vA** flanked by T or A were destabilized only by 1°C as compared to the unlabelled one. In contrast, the melting temperature decreased by 4 °C when **8vA** was flanked by C. This suggested that flanking thymine or adenine residues favoured the anti-conformation of **8vA**, while the flanking cytosine favoured the *syn*-conformation, which decreased the duplex stability due to the steric penalty of the vinyl substituent. Similar conclusions were drawn for **8vG**, where the melting temperature decreased by 4 °C with flanking C, whereas with flanking T or A, the decrease ranged from 1 to 3 °C [210].

2.4.1. Photophysics of monomeric 8vA

In aqueous solutions, **8vA** nucleoside was found to be very fluorescent (Φ_F = 0.65) with an absorption maximum at 290 nm (ε = 12600 $M^{-1}\cdot cm^{-1}$) and a fluorescence maximum at 386 nm (Table 1). **8vA** can be excited at wavelengths up to 340 nm where the natural nucleobases and aromatic amino acids do not absorb [208]. In methanol, a 7 nm blue-shift in the emission maximum and a 4 nm red-shift in absorption in comparison to water was observed. In methanol, Φ_F decreased to 0.32.

The effect of the vinyl group on the electronic structure and photophysics of the purine core of **8vA** was studied computationally [249]. Three vertical electronic transitions in **8vA**, namely $S_0 \rightarrow S_1$ (L_a $\pi\pi^*$), $S_0 \rightarrow S_3$ (L_b $\pi\pi^*$), and $S_0 \rightarrow S_2$ ($^1n\pi^*$) were characterized. The energy difference between the $S_0 \rightarrow S_1$ and $S_1 \rightarrow S_0$ transitions was 7735 cm^{-1}, after taking into account the re-organization energy associated with the relaxation of the molecular structure at ground- and excited-states of the five-membered ring and the vinyl group (Figure 21). The difference between the calculated and the experimental value (8300 cm^{-1}) was attributed to the solvent effect. The value of the oscillator strength associated with

the $S_0 \rightarrow S_1$ transition was in agreement with the experimental value (0.274) in buffer and was about twice that of **2AP** (0.127). The high Φ_F of **8vA** could be explained by the large energy gap between the low-lying $^1n\pi^*$ and $^1\pi\pi^*$ levels, which hinders the internal conversion according to the Lim proximity model.

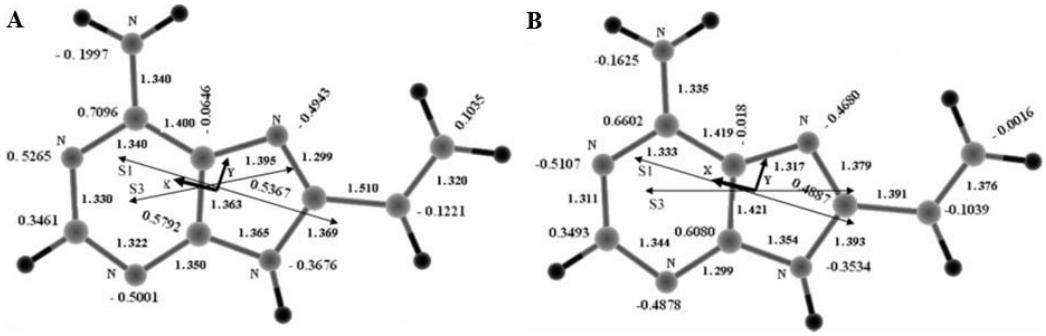

Figure 21. Structure of **8vA** showing the bond lengths, the charge densities, and the directions of the transition dipoles for the $S_0 \rightarrow S_1$ (La $\pi\pi^*$) and $S_0 \rightarrow S_3$ (L_b $\pi\pi^*$) transitions. (A) S_0 minimum and (B) S_1 minimum. The figure was reproduced from [249] with the permission of ACS publications.

To study further the contribution of the vinyl group to the spectroscopic properties of **8vA**, the energetic profiles of the three highest occupied molecular orbitals (HOMO) and the two lowest unoccupied molecular orbitals (LUMO) were calculated (Figure 22). The $^1(L_a$ $\pi\pi^*)$ and $^1(L_b$ $\pi\pi^*)$ transitions have unique contributions from HOMO→LUMO and HOMO–1→LUMO, respectively, whereas the $S_0 \rightarrow S_3$ transition has dominant contributions from HOMO–2→LUMO (44%) and HOMO→LUMO+1 (27%). The electron density in HOMO and HOMO–2 is delocalized on the whole **8vA** molecule, while it is predominantly localized on the purine nitrogen atoms and the vinyl group in HOMO–1 and LUMO, respectively. In LUMO+1, the electron density is localized on the NH group of the five-membered ring. Thus, the $^1(L_a$ $\pi\pi^*)$ transition has a substantial intramolecular charge transfer (ICT) character, as suggested by the modification of the charge distribution between the S_0 and the S_1 states (Figure 22). In the $S_0 \rightarrow S_2$ transition, the charge migration from the purine nitrogen atoms to the vinyl group is consistent with the $^1n\pi^*$ character of this state [249]. Therefore, in line with the calculated barrier (6272 cm^{-1}) between the $^1(n\pi^*)$ and $^1(L_a$ $\pi\pi^*)$ transitions, a local minimum likely traps the energy for long enough to allow a strong fluorescence.

The change in **8vA** electronic structure for its lowest energy bright transition was further evaluated by Stark spectroscopy in frozen ethanol at 77 K to obtain the direction and degree of charge redistribution [250]. These data were completed by QM calculations to calculate the transition energies, oscillator strengths, and dipole moments of the ground and excited states of **8vA**. The obtained data were in good agreement with those of the Mély's group [249] and confirmed that upon excitation, an ICT moves the electron density from the purine ring onto the vinyl group. This charge redistribution provided a significant difference dipole vector, whose direction is less than 40° from the transition dipole moment vector [250].

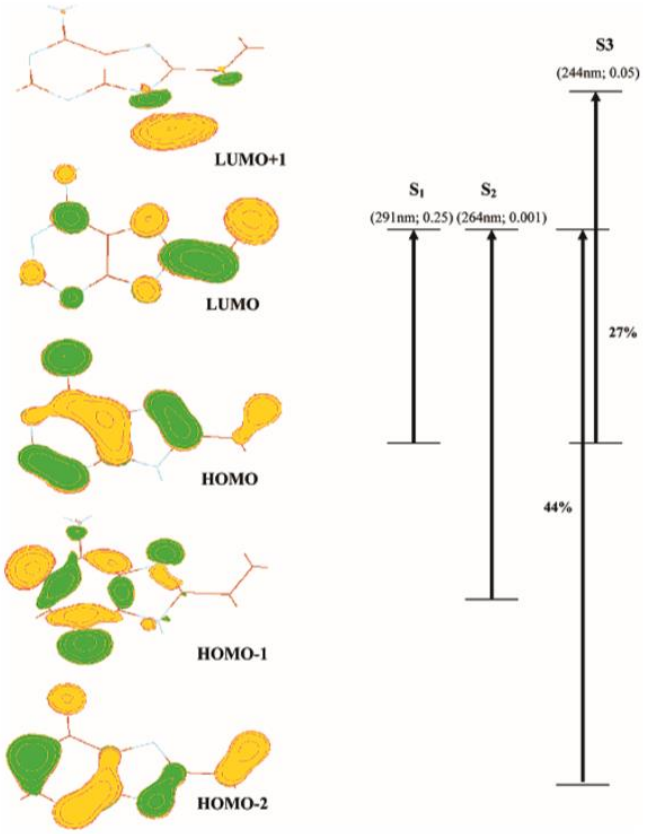

Figure 22. Singlet excited-state transitions of free **8vA** calculated by TDDFT [B3LYP/6-311+G(d)]. Transition wavelengths and oscillator strengths are shown in parentheses. The figure was reproduced from [249] with the permission of ACS publications.

2.4.2. Photophysics of monomeric 8vG

8vG in aqueous solution has an absorption coefficient of 7200 $M^{-1}\cdot cm^{-1}$ at 260 nm and a Φ_F of 0.72 (Table 1) [210]. Similar to **8vA**, **8vG** can be excited at longer wavelengths (up to 325 nm) compared with the natural nucleobases and aromatic amino acids, which enables its selective excitation in ONs and their complexes with proteins. The position of the fluorescence maximum and the value of the fluorescence lifetime of **8vG** were observed to depend on the polarity and H-bonding properties of the solvent [211]. Indeed, by going from THF to water, the fluorescence maximum shifted from to 370 to 396 nm, and the fluorescence lifetime increased from 2.07 ns to 4.35 ns. As for **8vA**, the photophysical properties of **8vG** were interpreted through theoretical electronic calculations [211,251]. The purine moiety of **8vG** was found to adopt a near-planar structure similar to the natural guanine, which suggested that the vinyl substitution at the position C8 did not significantly influence its electronic structure in the ground state. Due to the vinyl substitution, the lowest singlet excited state of **8vG** was the S_1-ICT ππ* state, which involved an electron density redistribution from the purine moiety onto the vinyl group. The S_2 state corresponded to a slightly higher energy ^{1}ππ* state almost entirely localized on the guanine moiety, and thus, very similar to the $^1(L_a$ ππ*) state of 9H-guanine. It was thus referred as the L_a ππ* state of **8vG**. As the oscillator strength of the S_1 (ICT ππ*) state was considerably higher than that of the S_2 (L_a ππ*) state, photoexcitation of **8vG** at 305 nm mainly populated the S_1-ICT ππ* state that subsequently relaxed from the Franck-Condon geometry to the S_1-ICT-min, where it became trapped for long enough for fluorescence to take place (Figure 23).

Alternatively, **8vG** can leave the potential energy basin around S_1-ICTmin by overcoming a low potential energy barrier in order to reach the L_a ππ* region, from where the (L_a ππ*/gs) CI can be easily accessed and provide a radiationless internal conversion pathway to the ground state. The energy barrier increased with solvent polarity, which favoured the fluorescence pathway. An increase in solvent polarity probably also disadvantaged the S_1(ICTππ*) → T_2(ππ*) intersystem crossing process by increasing the energy gap between the two states, which further favoured the fluorescence pathway. Finally, the small fraction of molecules excited into the S_2 (L_a ππ*) state was believed to either undergo internal conversion into the S_1(ICT ππ*) near the Franck−Condon geometry, and became trapped in S_1-ICT-min, or alternatively relax on the L_a ππ* state pathway toward the S_1(L_a ππ*)/S_0 CI seam.

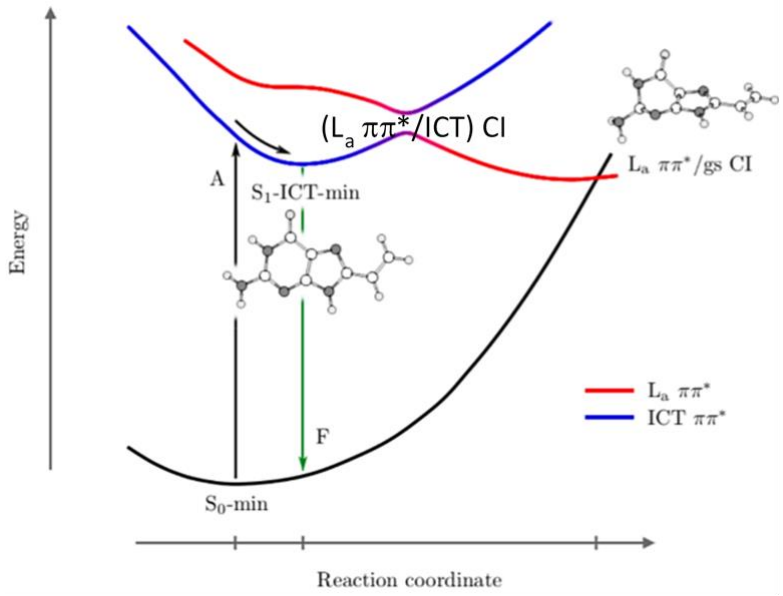

Figure 23. Proposed relaxation pathways in the photoexcited monomeric **8vG** in solution. (Adapted from [251] with the permission of ACS publications)

2.4.3. Steady-state and time-resolved fluorescence of 8vA in oligonucleotides

The fluorescence decays of the monomeric **8vA** nucleoside in pure solvents are mono-exponential with fluorescence lifetimes of 4.7 ns and 2.9 ns for water and methanol, respectively [208]. When **8vA** was incorporated into DNA flanked by T residues, its Φ_F dropped from 0.65 to ~0.025 (Table 1). This Φ_F decrease was accompanied by the appearance of complex multi-exponential fluorescence decays. Three to four lifetime components were observed, ranging from 50 ps to 2 ns and from 20 ps to 4 ns for ssDNA and dsDNA, respectively. This complexity is likely due to a combination of static and dynamic quenching, resulting from the π-stacking with the natural nucleobases and the exposure of the chromophore to the solvent. Interestingly, duplex formation resulted in a blue-shifted emission spectrum of **8vA** by about 5 nm without any significant change in Φ_F or time-resolved parameters, suggesting that the spectroscopic behaviour of **8vA** was mainly governed by intrastrand factors [208]. When flanked by A residues in ssDNA, Φ_F of **8vA** dropped to ~0.015. Marginal changes were observed upon duplex formation, confirming the strong connection of **8vA** photophysical properties with the flanking residues [208]. In ONs with flanking C residues, the Φ_Fs were higher than with A and T neighbours in ssDNA (~0.035). In contrast, in dsDNA, Φ_F decreased significantly (~0.005), likely

due to the hydrogen bonding of **8vA** with the opposite residue which strengthened the π-stacking interactions of **8vA** with its neighbouring C bases [208].

The multiple lifetime components of **8vA** in ONs likely reflected a distribution of local conformations in which **8vA** experienced varying degrees of stacking with the flanking nucleobases of the same strand. The small amplitude (< 1%) of the least-stacked form of **8vA** in the three tested environments suggested that **8vA** experienced only limited displacement relative to its neighbours [249]. Theoretical calculations indicated that stacking of **8vA** with C, T, or A in double stranded dimers or trimers resulted in fluorescence quenching through mechanisms that depended on the nature and relative orientation of the flanking bases. When **8vA** was stacked with thymine residues, the quenching likely resulted from internal conversion from the dipole-allowed **8vA**-like state to low-lying dark states. This was accompanied by an electron transfer to the thymine residues that exhibited the highest electron affinity among the nucleobases [249]. When **8vA** was flanked by adenines, the quenching likely resulted from a strong π-orbital overlap in the ground state that provided static quenching, with a reduction of the radiative rate constant. A dynamic quenching through internal conversion to low-lying states was also expected. Both quenching mechanisms might explain the low Φ_F observed with flanking adenines. Finally, when **8vA** was flanked by C residues, the low-lying S_1 state appeared as an emissive state with a rather high oscillator strength, explaining that **8vA** was more fluorescent when flanked by cytosines as compared to T or A residues [249]. The quenching of **8vA** fluorescence when flanked by C residues might arise mainly from the decrease in the radiative rate constant resulting from the ground-state interactions between the nucleobases.

2.4.4. Steady-state and time-resolved fluorescence of 8vG in oligonucleotides

Compared to free **8vG**, the incorporation of **8vG** in ssONs reduced the fluorescence of **8vG** by 8.7-, 9.8-, 10.9-, 11.6-fold in presence of flanking A, C, T, G, respectively [210]. This indicated a strong intrastrand quenching that depended on the nature of the flanking nucleobase. A further up to two-fold reduction of its fluorescence was observed upon duplex formation, likely caused by the hydrogen bonding to the opposite nucleobase which reinforced the intrastrand π-stacking interactions of **8vG** with its neighbouring bases.

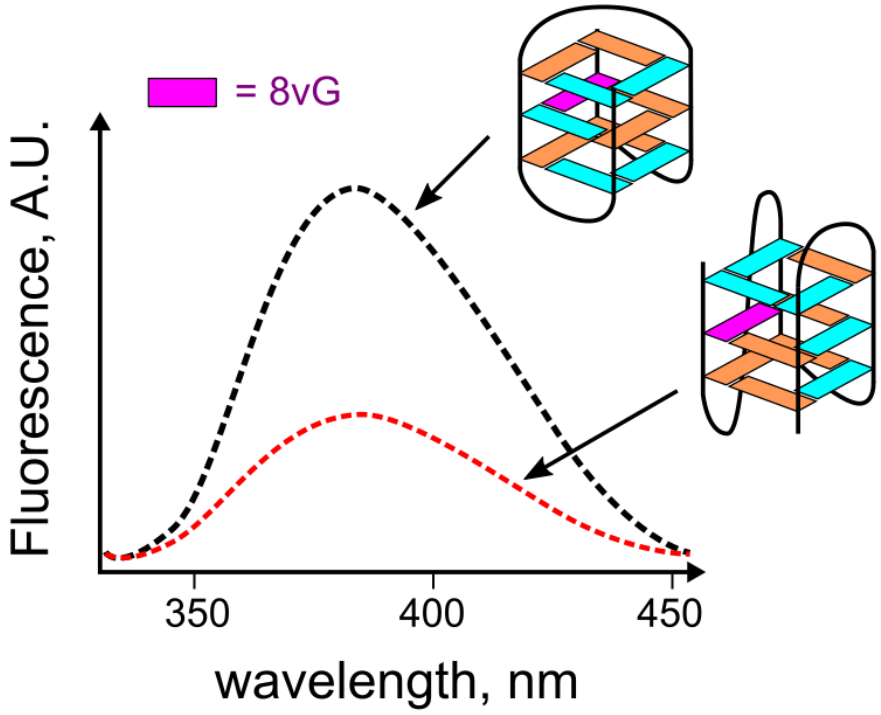

Figure 24. Fluorescence spectra of **8vG**-labeled tel22 G-quadruplex-forming sequence in NaCl (black) and KCl (red). The syn-G residues are in green and the anti-G residues are in red in both quadruplex topologies. The substituted G position is labelled by a star (Redrawn from [210]).

The ability of **8vG** to adopt both *syn*- and *anti*-conformations was also used to monitor the formation of DNA G-quadruplexes. G-rich sequences can fold into quadruplex structures with topologies that depended on the temperature, pH, and the presence of monovalent or divalent cations. The tel22 oligonucleotide derived from the telomeric DNA is an example of G-quadruplex-forming sequence. In the presence of Na^+ ions, tel22 folded into a G-quadruplex topology with three G-tetrads connected by two edgewise and one diagonal loops [252,253]. In the presence of K^+ ions, tel22 mainly folded into a G-quadruplex topology where the G-tetrads are connected by double chain reversal loops [254]. By including **8vG** at three different positions in the tel22 sequence, temperature-dependent UV spectroscopy and circular dichroism (CD) measurements revealed that both topologies were not markedly affected by **8vG** incorporation [210]. The fluorescence signal of **8vG** was found to depend on the quadruplex topology for the **8vG**-substituted positions in the central G quartet, but not in the outer G quartet. For **8vG** in the central G quartet, the Na^+-induced form was three-fold brighter than the K^+-induced form (Figure 24).

Since **8vG** can sense different topologies and structures in DNA, it was incorporated into peptide nucleic acids (PNA) and used as a hybridization-sensitive probe [255]. Melting temperature experiments showed that the vinyl group minimally impacted the stability of the labelled duplexes. The fluorescence of **8vG**-labeled PNA was different for single-stranded PNA and PNA-DNA hybrids. A 72% decrease of fluorescence was observed upon duplex formation at 5 °C. Heating of the PNA duplex above the melting temperature (70°C) provided a fluorescence spectrum that was almost identical to the single-stranded PNA. Furthermore, the fluorescence spectra of the PNA/DNA duplexes were similar to the spectra of the corresponding PNA/PNA duplexes. Therefore, incorporation of **8vG** in PNA sequences made it possible to detect formation of both PNA/PNA and PNA/DNA duplexes [255].

2.5. Pteridines

The group of Hawkins introduced a family of highly emissive nucleoside analogues named pteridines [212,256]. The most studied members of this family are 3-methyl isoxanthopterin (**3MI**), 6-methyl isoxanthopterin (**6MI**), 4-amino-6-methyl-pteridone (**6MAP**) and 4-amino-2,6-dimethyl-pteridone (DMAP) (Figure 25). The guanosine-like pteridines **3MI** and **6MI** and the adenine-like pteridines **6MAP** and **DMAP** differ from the natural purines by their heterocyclic core and the presence of additional methyl groups (Figure 25) [213]. The structural similarity of pteridines with the natural purines makes them good nucleobase mimics upon incorporation into nucleic acids [215,257]. At the same time, the **3MI** pteridine bearing a methyl group at the base-pairing interface cannot form WC-like base pairs with complementary pyrimidines. As expected, melting temperature (T_m) analysis indicated that a **3MI**·C base pair destabilized dsDNA similarly to a single base pair mismatch. In contrast, **6MI**, where this methyl group was absent, could form a WC-like base pair with cytosine. Therefore, T_m values of **6MI**·C duplexes were nearly identical with those of the natural duplexes [212]. The WC base-pairing of **6MAP** and **DMAP** with thymine was similar to that of adenine, resulting in similar T_m values [215].

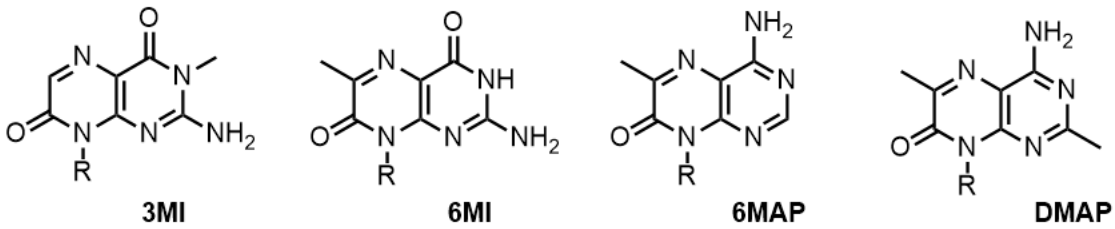

Figure 25. Chemical structures of selected pteridine nucleoside analogues; R= 2'-deoxyribose or 2'-ribose ([213]).

2.5.1. Photophysical properties of monomeric pteridines

In their monomer form in buffered aqueous solutions, the pteridine nucleoside analogues are highly fluorescent in the visible range (380−500 nm), likely due to the increased conjugation resulting from the five- to six-member ring extension. Their absorption (> 300 nm) and fluorescence maxima (430 nm) were significantly red-shifted as compared to canonical nucleobases and amino acids making them useful for studies of protein/DNA complexes (Table 1) [208-210]. Compared to natural nucleobases, their increased conjugation lowered the HOMO–LUMO gap between the ground and excited states, which explained their spectral shifts. The pteridines were characterized by high Φ_Fs (0.39–0.88). Although all four compounds retained their spectral characteristics after incorporation into ONs, they were strongly quenched due to base stacking interactions and base pairing. In ssDNAs or dsDNAs, the Φ_F of **3MI** and **6MI** ranged from 0.03 to 0.3, while the Φ_F of **6MAP** and **DMAP** ranged from 0.01 to 0.16. In general, flanking purines quenched incorporated pteridines stronger than pyrimidines.

In their free nucleoside form, the fluorescence lifetimes of **3MI** and **6MI** were of 6.5 ns and 6.35 ns, respectively. When **3MI** and **6MI** were incorporated in ONs, their fluorescence decays became more complex with the appearance of short-lived lifetime components due to the quenching mechanisms mentioned above [213]. When **3MI** was surrounded by pyrimidines in both ssDNA and dsDNA, the fluorescence decays were bi-exponential with a long-lived lifetime of ~5.3 ns a shorter component (~2.5 ns) representing about 30−40%. When **3MI** was surrounded by purines in dsDNA, the fluorescence decays become even more complex, with the shortest lifetime (0.2 ns) being the most

populated (70%). This short-lived lifetime was thought to be related to an electron transfer allowed by the π-stacking between the probe and the flanking purine. As the methyl group of **3MI** prevented the formation of a WC-like base pair, the fluorophore likely distorted the tertiary structure of the duplex and became partially exposed to the solvent. This likely reduced the quenching with the neighbouring bases, which may result in the 2.9 ns lifetime [213,214]. When incorporated in ssDNAs or dsDNAs with flanking purines, **6MI** was more strongly quenched than **3MI** in dsDNA [216,217], exhibiting a three-component fluorescence decay, with lifetimes of 0.2, 2.2, and 5.9 ns. The two shortest components were the most populated (81−83%, respectively), likely as a result of the perfect base pairing of **6MI** with cytosine that allowed a strong π-stacking, and thus an efficient electron transfer with the flanking nucleobases.

Free **6MAP** and **DMAP** nucleosides in solution had mono-exponential decays with lifetimes of 3.8 and 4.8 ns, respectively [213,215]. When incorporated into ssDNAs, their decay curves became more complex. The dominant component for both probes ranged from 2.5 to 2.9 ns, depending on the nature of the flanking bases. Simultaneously, a shorter-lived lifetime (0.2−0.7 ns) appeared with amplitudes ranging from 22 to 59% for **6MAP** and from 39% to 65% for **DMAP**. As **6MAP** and **DMAP** were structurally close and formed WC-like base pairs similar to A, their decay patterns in dsDNAs with three lifetimes were close to that observed for **6MI**. This suggested that their quenching was due to a combination of π-stacking between the probes and flanking bases (purines being stronger quenchers) and exposure to the solvent. In this context, the shortest and most populated component was ascribed to the probe perfectly stacked with the nucleobases, the middle component to the probe still stacked but partially exposed to the solvent and the longest component to the probe exposed to the solvent [215].

2.5.2. Applications of pteridines

Due to the sensitivity of their fluorescence to subtle changes in base stacking and base pairing, the pteridine analogues were used to probe DNA interactions and interconversions [214]. For instance, pteridines analogues were used to monitor the cleavage activity of the HIV-1 integrase, an enzyme responsible for the integration of the viral DNA into the host cell genome [212]. The first pivotal function of integrase is the cleavage of a specific dinucleotide from the 3'-ends of the long terminal repeat (LTR) sequence at either end of the DNA provirus. To monitor this cleavage, the pteridine probe **3MI** was incorporated at the integrase cleavage site into a double-stranded 21-mer mimicking the U5 or U3 terminus of HIV-1 DNA. After cleavage, the loss of base-stacking increased the fluorescence intensity of **3MI** and allowed monitoring the cleavage in real time. A similar fluorescence increase was observed upon cleavage of pteridine-labeled substrates by other nucleases, such as P1 nuclease and exonuclease III [213].

In another application, **3MI** was used to monitor the DNA-binding of the nonspecific histone-like DNA-binding protein (HU) [258]. Fluorescence anisotropy measurements were used to determine the binding of HU to a 13 bp DNA duplex (H1-13-1) and two 34 bp DNA duplexes (H1-34-1 and H1-34-2) that differed in the position of the **3MI** probe (5 bases from the 3′ end for H1-34-1 and in the middle of the duplex for H1-34-2). As expected, in the absence of HU, the shorter 13-bp duplex had a lower initial anisotropy value as compared to the 34 bp duplexes. The H1-34 duplexes exhibited a greater change in anisotropy ($\Delta r \sim 0.067$) upon complex formation with HU than the H1-13-1 duplex ($\Delta r = 0.033$). Moreover, the 34 bp duplexes (H1-34-1 and H1-34-2) exhibited an increase in fluorescence intensity by 23 and 43%, respectively, being consistent with an increased solvent

exposure of the probe. No change in fluorescence intensity was observed upon HU binding to the H1-13-1 duplex. The increase of **3MI** fluorescence intensity upon HU binding to the 34 bp duplexes was indicative of decreased base stacking interactions consistent with local unwinding or bending of the longer DNA helix by HU. In contrast, the 13 bp duplex being below the persistence length was too short to be bent by HU. Pteridine analogues were also used to map the specific binding site of the unwinding protein (UP1) on ssDNA [259]. Finally, **3MI** was incorporated into a sequence designed to bulge out from the dsDNA upon annealing to a target. The bulging out of **3MI** resulted in an increase of fluorescence intensity (up to 27-fold) that was used for sequence-specific detection of PCR amplicons [260].

2.6. Isomorphic fluorescent pyrimidine analogues

Pyrimidine nucleobases in DNA and RNA are represented by thymine, uracil, and cytosine. In addition, 5-methylcytosine (5mC) and its oxidized derivatives are found in DNA and RNA as modified nucleobases of epigenomic and epitranscriptomic significance [261,262]. Several other modified pyrimidines have been found in the ribosomal and transfer RNAs where they impact the functioning of the translational machinery of the cell [252-254]. The vast majority of the natural pyrimidines are virtually non-fluorescent [11]. For instance, thymine, uracil and cytosine undergo a rapid nonadiabatic radiationless relaxation due to easily accessible CIs between the singlet exited state S_1 and the ground state S_0 [85,255-257]. Fast internal conversion results in fluorescence lifetimes of less than 1 ps [82,84,85,267] and very low Φ_Fs of less than 10^{-4} [83,267]. A combination of spectroscopic and computational techniques has been used to study the mechanisms underlying the ultrafast dynamics of electronically excited pyrimidines. The major relaxation pathway via the S_1–S_0 CI required an out-of-plane deformation of the planar six-membered heterocyclic nucleobase ring [80,268]. The key chemical bond involved in this process is the C5–C6 double bond. The change of its hybridization from sp^2 to sp^3 led to a dihedral twist of the bond, which favoured the non-planar conformation of the nucleobase ring needed to reach the CI [86,265,269]. An alternative relaxation pathway in the excited pyrimidines relies on a $^1\pi\pi^* \rightarrow {}^1n\pi^*$ transition involving the lone electron pairs of the nitrogen and oxygen atoms of the pyrimidine core [80,86]. An important practical consequence of these mechanisms is the possibility to rationally modulate and improve the fluorescence properties of the pyrimidine nucleobases by introducing appropriate substituents at the C5 position of the ring and/or by modifying the electron donors involved in the relaxation pathway via the dark $^1n\pi^*$ states [256,258,260-263].
Several approaches combining molecular design with a rational substitution strategy have been employed to convert the structures of practically non-fluorescent natural pyrimidine nucleosides into fluorescent isomorphic nucleoside analogues in an atom-economic way. In most cases, this strategy preserved the compact molecular size of pyrimidines and their ability to interact with other nucleobases via hydrogen bonds and π-π stacking interactions. Here, we review the photophysical properties, together with applications of selected members of this family (Table 3).

2.6.1 Derivatives of 2-pyrimidone

One of the easiest ways to generate isomorphic fluorescent pyrimidines was to eliminate the substituent at the C4 position of the natural pyrimidine nucleobases. This elimination led to derivatives of 2-pyrimidone, a minimal parent fluorophore of a small group of fluorescent heterocyclic nucleobases and nucleoside analogues (Figure 26). A synthetic route leading to the 1-β-D-deoxyribosides of 2-pyrimidinone was first described in the 1960th [274], and since then the properties

of these nucleosides have been studied extensively. Advances in synthetic chemistry of nucleoside analogues made accessible 2-pyrimidinone-based nucleosides such as **dK**, **dm^{5}K**, and **rK** (zebularine), which are derivatives of deoxycytidine, thymidine, and cytidine/uridine, respectively [265-268]. This has enabled phosphoramidite-based chemical synthesis of DNA and RNA incorporating 2-pyrimidinone mimics of thymine or cytosine at a programmed position [275,279,280]. Enzymatic labelling of DNA with 2-pyrimidinone-based nucleosides using primer extension reaction has also been demonstrated [281].

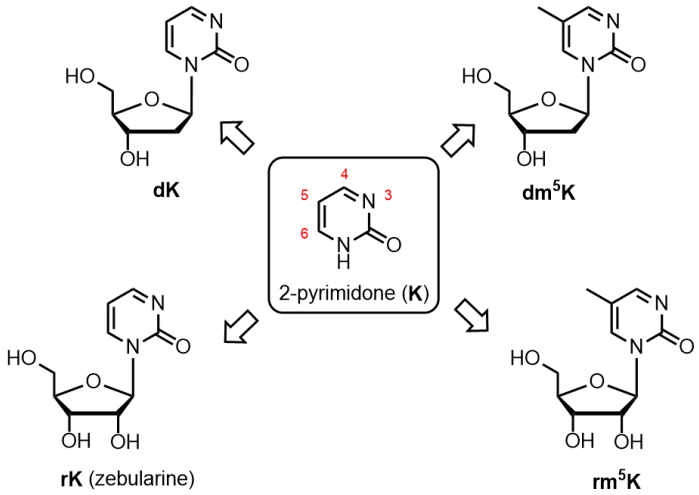

Figure 26. Fluorescent nucleoside analogues of the 2-pyrimidone family. Conventional pyrimidine numbering is shown in red.

As the substituent at the position C4 of pyrimidine nucleobases is involved in WC hydrogen bonding, its absence in 2-pyrimidinone nucleosides affected the stability and dynamics of dsDNAs. The modified nucleosides can form a weak base pair with guanine stabilized by two hydrogen bonds instead of three in the C·G base pair. At the same time, no hydrogen bond could form in a 2-pyrimidinone·A pair, inevitably destabilizing the duplex [276]. A combination of CD spectroscopy and DNA melting curve analysis on 10- and 12-mer duplexes confirmed that dsDNAs harboring a single **K**·G pair preserved the overall B-DNA conformation, but exhibited a dramatic decrease in melting temperature [276,282]. The 2-pyrimidinone·A pair had even a greater destabilizing effect [283]. MD simulations supported the spectroscopic observations, revealing an overall preservation of B-DNA conformation associated with a high local mobility of the artificial nucleoside due to partial disruption of the WC base pair [284].

Compared to the natural pyrimidine nucleobases, the 2-pyrimidinone derivatives exhibited distinct spectroscopic properties. Depending on the pH, the absorption spectra of the free nucleosides **dK** and **dm^{5}K** had maxima at 303–315 nm and 314–321 nm, respectively [274]. Their fluorescence maxima were at 365 nm and 385 nm, respectively [219,279]. The free fluorophore 1-methyl-2(1H)-pyrimidinone has a Φ_F of 0.04 [285], two orders of magnitude higher than the Φ_Fs of the canonical pyrimidine nucleobases. Computational studies helped to explain this increase in Φ_F [286,287]. Although the potential energy surfaces (PES) of the excited state of 5-methyl-2-pyrimidinone and cytosine were qualitatively similar, featuring the same type of CIs, substantial quantitative differences were observed. Due to the lower vertical excitation energy in 5-methyl-2-pyrimidinone, the topography of the PES of its excited electronic state changed in the vicinity of the nearest CI. As a result, an energy barrier appeared on the minimum energy path of this relaxation pathway, making it less accessible and leading to an increase of the excited-state lifetime [268,286,287]. Experimentally, 1-methyl-2-pyrimidinone had a fluorescence lifetime of ~ 400 ps [285] in water, two orders of magnitude higher than for the natural

pyrimidines, but still quite short. This lifetime was likely due to secondary radiationless relaxation pathways, such as intersystem crossing via the T_1 state [285,288].

Incorporation of the 2-pyrimidinone-based nucleosides into DNA was associated with small changes in their fluorescence maxima and Φ_Fs. For **dm^{5}K**, the fluorescence maximum shifted from 385 nm to 390–392 nm, and Φ_F decreased by 1.5–2 fold [219]. A more dramatic change was observed in the time-resolved fluorescence data. While the free nucleoside **dm^{5}K** exhibited a monoexponential decay with a 4 ns lifetime, the unnatural nucleoside incorporated into a 10-mer ssDNA exhibited a multi-exponential decay with three lifetime components of 0.6 ns, 2 ns and 4 ns. These lifetimes were associated to subpopulations of **dm^{5}K** with varying degrees of stacking with the flanking nucleobases [218], with the 4 ns component assigned to a fully unstacked state [218]. The sensitivity of the fluorescent decays of pyrimidinone-based nucleosides to base-stacking interactions resembled that of **2AP** (section 2.1), indicating the potential utility of 2-pyrimidinones as isomorphic pyrimidines to probe local conformational dynamics in DNA. Unlike the significant environmental sensitivity of their fluorescence lifetimes, the response of their fluorescence intensity is generally insufficient for practical applications. As an illustration, the fluorescent ribonucleoside analogue **rK** was used as a reporter of the hammerhead ribozyme functioning [280]. Although site-selective incorporation of **rK** in place of specific cytidine residues proved the functional importance of the exocyclic amino group for the cleavage reaction, no significant fluorescence intensity change useful for monitoring the process of RNA cleavage by the ribozyme was observed. In another study, **dm^{5}K** was examined as a reporter for the interaction of ssDNA with the RecA protein involved in DNA repair in *E.coli*. Upon binding of 30–60 nt ssDNA to RecA, a modest 1.8-fold increase in fluorescence intensity was observed [275]. These two examples illustrated that 2-pyrimidinone nucleosides are not the best tools for intensity-based fluorescence sensing due to their modest environmental sensitivity. Moreover, a potentially harmful photochemical effect of 2-pyrimidinones on labelled DNA has been recently revealed. In cellular DNA, the 2-pyrimidinone chromophore can form pyrimidine-pyrimidone (6-4) photoproducts from two adjacent pyrimidine bases upon exposure to UV light [289]. The 2-pyrimidone chromophore acted as a DNA photosensitizer, facilitating secondary photolesions and phosphodiester backbone breaks [290]. The key intermediate of the secondary photochemical reactions is the T_1 state of the 2-pyrimidone chromophore, which promotes photodamage via triplet transfer mechanism [284]. Altogether, both the photosensitization of DNA damage and the low environmental sensitivity of the 2-pyrimidone fluorophore undermine its usefulness as a probe for DNA dynamics and interactions.

2.6.2 Derivatives of pyrimidines with heterocyclic substituents at C5

A more successful design of isomorphic fluorescent pyrimidines is based on the addition of an electron-rich five-membered heterocyclic ring at the C5 position of the pyrimidine core [291]. This C-5 substitution increases the energetic toll of the dihedral twist of the C5–C6 bond and thus, suppresses the radiationless relaxation in pyrimidines. It also shifts the absorption and fluorescence to the red and improves the light absorption properties by expanding the conjugated electronic system. Moreover, these compounds closely resembled the natural pyrimidines, having similar molecular size and the WC hydrogen bonding interface. Tor and co-workers have shown that heteroaryl-modified nucleosides and their derivatives (Figure 27) were readily accessible synthetically in the form of ribosides and 2'-deoxyribosides in a small number of steps [221,222,292,293]. They also systematically explored the spectroscopic properties of different combinations of the nucleoside core (2'-deoxyuridine, 2'-deoxycytidine) with five-membered heterocycles (furan, thiophene, oxazole, and thiazole) [221,222]. These studies evidenced that 5-(fur-2-yl)-uracil derivatives (**^{FU}dU** and **^{FU}U**) exhibited optimal photophysical properties and thus, could be useful as probes for DNA and RNA [221,293]. Fluorescently labeled DNA and RNA can be prepared using both solid-phase phosphoramidite

chemical synthesis and *in vitro* enzymatic polymerization approaches. Melting temperatures of dsDNA were unaffected when a dT moiety in a perfectly matched duplex was replaced by **^{FU}dU**, confirming its perfect mimicking properties [221,222]. This class of isomorphic fluorescent pyrimidine analogues was further expanded with thiophenyl-, selenophenyl- and triazolyl-pyrimidines (**TH(d)U**, **^{SE}(d)U**, **^{TR}dC**, **^{TR}dU**, Figure 27) [216,218,284-286]. The heteroaryl-pyrimidine nucleosides had two absorption bands in water. In addition to the absorption of the pyrimidine chromophore at ~ 260 nm, there was a well-separated red-shifted absorption band centred at 291–325 nm, depending on the heterocyclic substituent [291,296]. A computational study showed that the red-shifted absorption band arose from π-π* electronic transitions in the planar conformation of the chromophores having expanded conjugated electronic system [297]. The fluorescence spectra featured a single band with maxima ranging from 407 nm for **^{TR}dC** [296], to 431–440 nm for the furan derivatives [221,293] and 452 nm for **^{SE}U** [295]. Notably, these fluorescent nucleosides exhibited very large Stokes shifts of up to 9700 cm^{-1}. Unlike the absorption spectra, the emission spectra of the heteroaryl-pyrimidine nucleosides showed positive solvatochromism, indicating a higher dipole moment of excited states relative to the ground state [291]. The Φ_{FS} of the free nucleosides in solvents were rather low (< 0.035) and their fluorescence decays were biexponential, with average lifetimes in the sub-nanosecond range [222,295,296]. Their low Φ_{FS} were explained by an additional radiationless relaxation mechanism linked to their rotatable biaryl bond connecting the heterocycle and nucleobase rings. Internal conversion via torsional relaxation was likely the main nonradiative decay channel for these compounds [298]. Unlike other isomorphic nucleosides such as **2AP**, 5-heteroaryl-pyrimidines were not quenched upon incorporation into RNA and perfectly matched dsDNA, exhibiting only minor alterations of their fluorescence intensities and spectral shape. Due to their sensitivity to polarity and rigidity, these fluorophores appeared as useful tools for probing nucleic acid structures and discriminating between alternative folds or hybridization states using intensity- and wavelength-based fluorescence sensing.

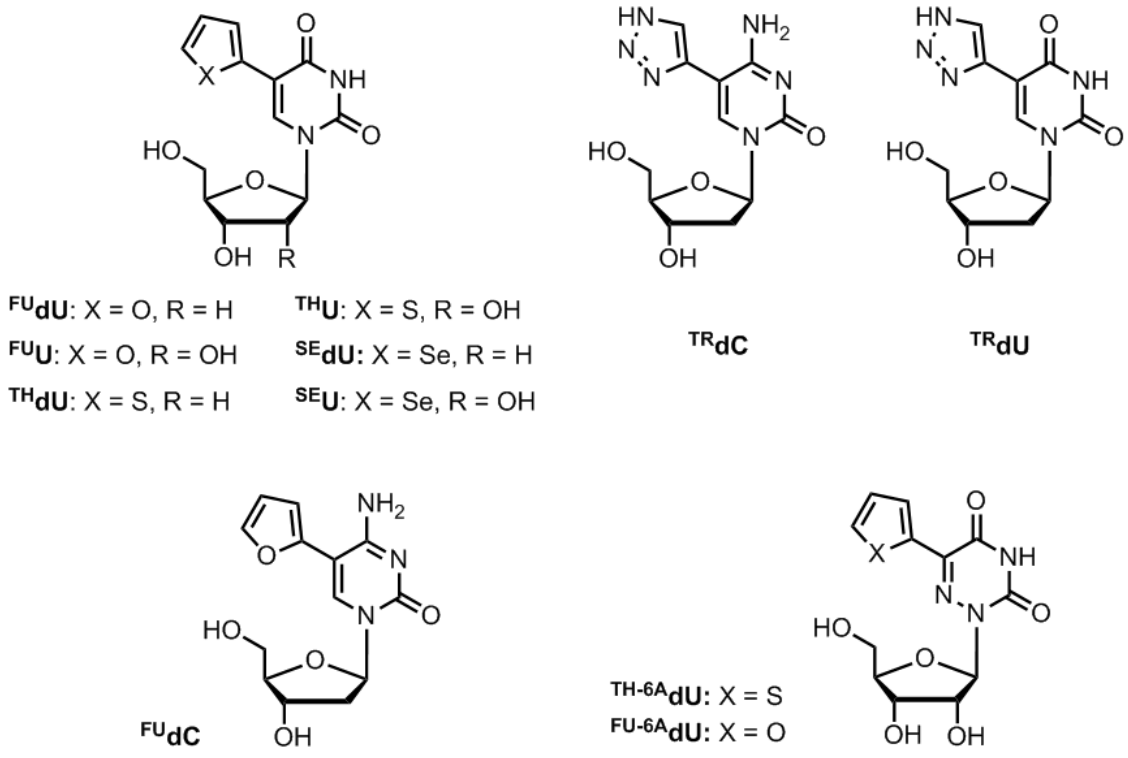

Figure 27. Heteroaryl-modified isomorphic pyrimidine analogues

The sensitivity of 5-(fur-2-yl)-uridines to the rigidity of their molecular environment has been utilized for the detection of abasic sites in DNA. A seven-fold increase in fluorescence intensity was observed when **^{FU}dU** was paired with an abasic site instead of the matched A in a duplex [221]. This light-up response was explained by the ability of the furan ring of the fluorophore to occupy the cavity formed by the abasic site in the duplex. As a result, the flanking nucleobases restricted the torsional relaxation and increased the observed fluorescence intensity [298]. In a similar way, **^{FU}dC**-labeled ONs were used as hybridization-based probes to distinguish between T, G, and its oxidized form 8-oxoG [299]. The sensitivity of the fluorescence maximum of **^{FU}dU** to the dielectric properties of the medium were also used for site-selective probing of the local polarity in the major groove of dsDNA. In this case, binary solvent mixtures were used to calibrate the wavelength response of the probe in order to estimate the local polarity at the labelled site from the emission maxima observed in A- and B-DNA [220]. The dual sensitivity of the heteroaryl-pyrimidine nucleosides to the molecular stiffness and polarity results in characteristic fluorescence emission patterns, which help to distinguish closely related nucleic acid folds. This property was used to study alternative folds of DNA G-quadruplexes with **^{SE}dU** and **^{FU}dU** as fluorescent reporters [223,300]. Finally, the environmental sensitivity of these fluorescent nucleosides allowed monitoring RNA-ligand interactions, such as the binding of aminoglycoside antibiotics to the ribosomal A-site [293,295].

2.6.3 Derivatives of triazines

The still low Φ_F of heteroaryl-pyrimidine fluorophores remained a major obstacle for their applications. Tor and co-authors have elegantly addressed the problem of low Φ_Fs of isomorphic pyrimidines via an atomic mutagenesis approach. They introduced an additional heteroatom into the nucleobase core by replacing the C6 atom with a nitrogen atom to obtain a group of 1,2,4-triazine derivatives. Due to a higher electron deficiency relative to pyrimidines, the obtained isomorphic nucleoside analogue 6-aza uridine (Figure 27, **$^{FU\text{-}6A}$dU** and **$^{TH\text{-}6A}$dU**) possessed higher charge separation in the excited state, which resulted in a notable shift of its spectra towards the red and a marked increase in the Φ_F values [224]. The absorption and fluorescence maxima of **$^{TH\text{-}6A}$dU** were 332 nm (ε = 11000 M^{-1} cm^{-1}) and 455 nm (Φ_F = 0.2 with a fluorescence lifetime of 4.9 ns), respectively. The furan derivative **$^{FU\text{-}6A}$dU** absorbed at 320 nm (ε = 10000 M^{-1} cm^{-1}) and emitted at 443 nm (Φ_F = 0.05) with a 3.1 ns fluorescence lifetime. Both compounds displayed similar response to pH. Due to its pKa (6.7), the thiophene derivative is an attractive probe for investigating (de)protonation events in RNA, as there is no isomorphic pyrimidine analog with pK_a values in this range. The authors also investigated the sensitivity of the optical properties with respect to viscosity as both compounds may isomerize along the single bond between the two π-systems. The properties of the thiophene derivative were almost not modified upon a 2000-fold increase in viscosity, indicating that rotational relaxation was not a major non-radiative pathway. In contrast, the Φ_F of the furan derivative reached 0.46 in pure glycerol suggesting that it could be used as a molecular rotor. Finally, the thiophene nucleoside analogue was incorporated into RNA and used as a reporter for aminoglycoside-RNA interactions [301].

3. Nucleoside analogues with expanded electronic system

In this section, we will focus on the photophysical properties of selected expanded pyrimidine and purine nucleobase analogues, which have been obtained by annealing, inserting, or replacing the natural nucleobase core with an additional aromatic ring. This approach extends the conjugated

electronic system of the nucleobase and therefore shifts the optical properties toward the visible range. Expanded aromatic systems also harbour additional positions for electron-donating and withdrawing substituents, which is beneficial for designing environmental sensitive dyes based on intramolecular charge or proton transfer reactions.

3.1. Expanded pyrimidine analogues

This family of fluorescent pyrimidine nucleobases includes analogues of thymine, cytosine and uracil that preserve WC base pairing with their purine partners. In the following, we highlight a series of representative molecules with increasing complexity in terms of chemical modifications (extended conjugation or ring addition) on the pyrimidine moiety. Note, that thieno[3,4-d]- and isothiazolo[4,3-d]-pyrimidine based nucleoside analogues (Figure 8 and 17, Table 3) were discussed in sections 2.2 and 2.3.

3.1.1. Pyrrolocytosine (pC). Although first synthesized in 1987 [302], **pC** (Figure 28A) found applications as fluorescent probe only in the early 2000s [303]. Methyl-pyrrolocytosine in the form of free nucleobase (pC_H: Figure 28, pC, R=H) had an absorption maximum at 340 nm (ε = 5900 M^{-1} cm^{-1}) and fluorescence maximum at 449 nm with Φ_F = 0.038 and a single fluorescence lifetime of 2.9 ns [304]. Coupling to ribose or deoxyribose shifted the fluorescence maximum 10 nm to the red together with a drop of Φ_F (0.026) and fluorescence lifetime (2.1 ns), due to an increased internal conversion rate associated to the decrease of the energy gap between the S_0 and S_1 levels. Upon incorporation in ssDNA or dsDNA, **pC** was further quenched when flanked by guanine residues. In this case, an additional small lifetime component (0.65 ns) appeared, probably associated with a photo-induced electron transfer with the guanine residues. Apart from guanines, theoretical calculations showed that base stacking does not affect the optical properties of **pC** [304]. More recently, Matsika et al. used theoretical approaches to characterize the optical properties of isolated **pC**, microhydrated **pC**, as well as complexes where **pC** was base-stacked or hydrogen-bonded with guanine [305]. For the free base $\mathbf{pC_H}$, the PES associated to S_1 displayed a very small barrier between the S_1 minimum and the closest CI, resulting in an ultrafast non-radiative decay. In the presence of H-bonded water molecules, the increase of the barrier height correlated with the increase of Φ_F. Using 5'pC-G3' and 5'G-pC3' sequences as models, the authors found that radiative decay rates decreased by 20% due to the π-stacking interactions, in good agreement with experiments [305]. The calculations also revealed a charge transfer state involved in G oxidation, which resulted in a fast non-radiative decay.

One of the first applications of **pC** was the mapping of melted regions during transcription [303]. To characterize the transcription bubble in T7 RNA polymerase elongation complexes, the authors incorporated **pC** in template and non-template sequences. The quenching of **pC** in dsDNA with respect to ssDNA was exploited to monitor the elongation bubble over 8 bases upstream of the pause site, while **2AP** was only able to monitor this process over one nucleotide. More recently, **pC**-labelled ONs were used to monitor the hybridization of modified metal-ion-chelating ONs with their unmodified counterparts [306]. The metal-ion mediated base pair was used to stabilize the hybridization of short ssDNA [307]. The decrease of **pC** fluorescence correlated well with the thermal stability of the double helix. When placed opposite to a sterically demanding base, such as the 2,6-bis(3,5-dimethylpyrazol-1-yl)purine, **pC** flipped out of the base stack, resulting in a fluorescence emission increase. This approach was used to probe the hybridization of short metal-ion-carrying ONs with the **pC**-labelled bulge motif of TAR RNA from HIV-1, but no clear conclusions could be drawn, likely due to the low affinity of TAR RNA for the short sequences used.

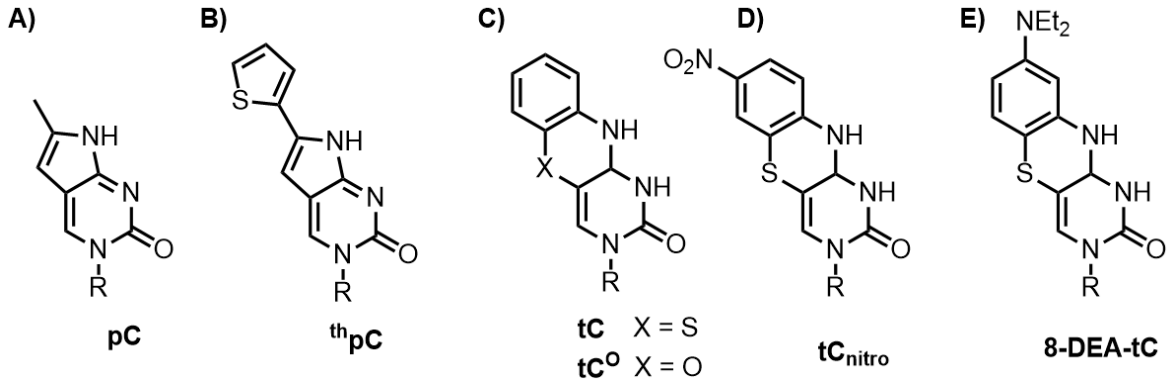

Figure 28. Chemical structures of expanded cytidine derivatives. A) Methylpyrrolocytosine (**pC**); B) thiophen-2-yl pyrrolocytosine (**thpC**), C–E) tricyclic cytosine (**tC**) derivatives; R = H, ribose or 2'-deoxyribose.

To further diversify the library of cytosine derivatives, Tor's group designed an expanded **pC** derivative by conjugating a thiophene to produce the thiophen-2-yl pyrrolocytosine (Figure 28B, **thpC**) [308]. This derivative had improved absorption coefficient (11600 M^{-1} cm^{-1}) and Φ_F (0.41) relative to **pC**, which provided a 24-fold increase in molecular brightness. The absorption and emission maxima were only slightly affected by the thiophene conjugation, showing a 17 nm and 8 nm red shift, respectively.

3.1.2. Tricyclic cytosine analogues (tC and tC^O). The tricyclic cytosine analog **tC** (Figure 28) was synthesized by Matteucci et al. in 1995 [309] to design duplexes with improved stability. The first systematic characterization of the optical properties of **tC** was performed by Wilhelmsson and co-workers in the early 2000th in studies combining experiment and theory [310 311]. The free nucleobase (Figure 28C, **tC**, R=H) had an absorption maximum at 375 nm (ε = 4500 M^{-1} cm^{-1}) that was red shifted by 25 nm upon incorporation in duplexes. Together with the observed hypochromic effect in absorbance (ε = 4000 M^{-1} cm^{-1}), these parameters indicated that **tC** is stacked with adjacent bases. Both free and incorporated base had a fluorescence maximum at 505 nm and a Φ_F of 0.2. Due to its long-lived fluorescence lifetime in ssDNA and dsDNA (5.9–6.3 ns), the authors performed lifetime-based FRET experiments using **tC** as donor. In parallel, pH studies revealed that anionic/neutral and neutral/cationic equilibria of **tC** forms were characterized by pK_a values of 1.0 and 13.2 respectively. Therefore, the optical properties under physiological conditions were dominated by the neutral form of **tC** which formed WC base pairs and exhibited a higher Φ_F than the charged species (Φ_F<0.1). Experiments with organic solvents (Φ_F≈0.5) evidenced the role played by hydrogen bonding in lowering Φ_F of **tC** in aqueous buffer. In addition, using polarization-resolved spectroscopy, the absorption band of **tC** was associated to a single electronic transition polarized at 35° with respect to the long axis of the molecule, which allowed quantitative analysis of FRET experiments quantitatively.

The oxo-homologue of **tC**, denoted **tC^O** (Figure 28C) was also introduced by Matteucci et al [309]. Compared to **tC**, its absorption maximum was blue shifted by 15 nm and its molar absorption coefficient was higher (ε = 9000 M^{-1} cm^{-1})[312]. Its transition dipole was shown to be oriented at 33° with respect to the long axis of the molecule. In phosphate buffer, **tC^O** had an fluorescence maximum at 461 nm with a Φ_F of 0.3 and a fluorescence lifetime of 3.4 ns [312]. Upon incorporation into dsDNA,

the optical properties **tC^O** were moderately altered, showing Φ_F ranging from 0.18 to 0.30 and a single fluorescence lifetime (3.4–4.8 ns) that depended only weakly on the nature of the neighboring bases. The only change from the free nucleoside was the hypochromic effect which decreased by 35–50% the radiative rate constant in accordance with the Strickler-Berg equation. When **tC^O** was incorporated into ssDNA, its photophysics was more complex, being sensitive to the nature of the neighboring bases. When **tC^O** was flanked by a 5' guanine, Φ_F dropped to 0.15 and a short-lived species (2 ns, 80%) dominated its fluorescence decay. This quenching is likely a signature of a photo-induced electron transfer. When **tC^O** was flanked by 5' A, T, or C, its photophysics resembled that of the free nucleoside. However, structural information was missing to rationalize the observed properties. **tC^O** was also used for the detection of individual melting processes in complex nucleic acid structures [313]. **tC^O** can also detect base-pair mismatches, through a decrease of its emission and a structuration of its fluorescence spectrum, particularly with the A mismatch. This likely resulted from hydrophobicity change and/or tautomerization of **tC^O**. Later, Kutcha et al. produced **tC^O**-labeled DNA of high molecular weight using PCR [314]. Unquenching experiments with exonuclease I identified collisional quenching of **tC^O** with coiled secondary structure and long range DNA mediated electron transfer as key factors modulating Φ_F.

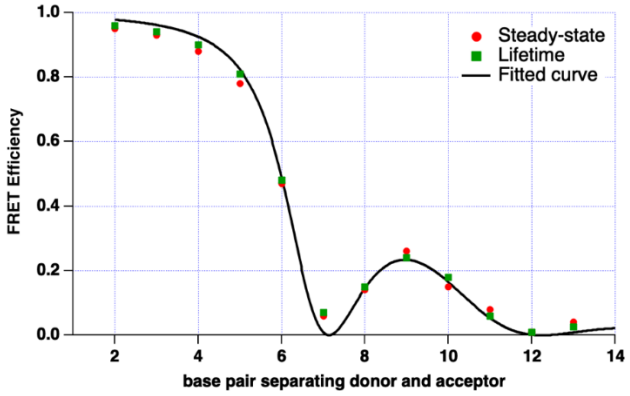

Figure 29. Energy transfer efficiency of the **tC^O**–**tC$_{nitro}$** pair measured using fluorescence lifetime decays and steady state fluorescence intensities. The observed FRET efficiency modulation correlates with changes in the FRET orientation factor κ^2. (Adapted from [315].)

As **tC^O** is characterized by a relatively long fluorescence lifetime when incorporated in dsDNA, an additional **tC** derivative (**tC$_{nitro}$**, Figure 28D) was designed to obtain an acceptor for FRET experiments with **tC^O** as a donor [315]. **tC$_{nitro}$** is essentially non-fluorescent and its absorption with a maximum at 440 nm overlaps well the emission of **tC^O** (Förster radius R_0= 27 Å, Table 4). The **tC^O**–**tC$_{nitro}$** FRET pair can thus be used to get structural information on DNA duplexes (Figure 29).

The group of B. Purse also contributed to the development of **tC** analogues by introducing the diethylamino-modified nucleoside analogue **8-DEA-tC** (Figure 28E)[316]. In PBS buffer, the dye had absorption and fluorescence maxima centered at 395 nm (ε = 2700 M^{-1} cm^{-1}) and 493 nm (Φ_F=0.006), respectively. In contrast to the majority of FNAs, **8-DEA-tC** displayed turn-ON properties upon incorporation in ss and ds DNA. The authors performed a study in which both the nature of the flanking residues and the opposite bases were systematically modified in PBS buffer at pH 7.4. In ss-DNA, the absorption and emission maxima of **8-DEA-tC** ranged between 413–425 nm and 492–499 nm, respectively, while the Φ_F ranged between 0.008 when **8-DEA-tC** is flanked by adenine nucleobases and 0.032 when **8-DEA-tC** is flanked by 5'-G and 3'-C nucleobases. In ds-DNA, the absorption and emission maxima ranged between 348–422 nm and 492–501 nm, respectively with a Φ_F increasing from 0.007 (opposite abasic site) to 0.12 (flanked by 5'-G and 3'-C). In parallel, the

authors rationalized their observations with the help of DFT calculations and showed that the HOMO and LUMO of parent **tC** are distributed across the arene while the HOMO of **8-DEA-tC** is much more polarized towards the diethylaminobenzene ring and the LUMO towards the pyrimidine ring, indicating a push–pull character. Finally, they identified excited state proton transfer towards the solvent as the main quenching mechanism of **8-DEA-tC**.

In 2003, Saito's group designed the benzopyridopyrimidine (**BPP**, Figure 30A), a probe that is strongly quenched upon base pair formation with purines [317]. In ssDNA, **BPP** had an absorption maximum at 347 nm and a fluorescence maximum at 390 nm with Φ_F of 0.04. The fluorescence decay of **BPP**-labeled ssDNA in phosphate buffer was characterized by three lifetimes ranging from 0.16 to 2.4 ns. Upon binding to a complementary sequence in which A was opposite **BPP**, Φ_F decreased only slightly (0.035). In contrast, when forming a dsDNA with an opposite guanine, Φ_F decreased by a factor of 20 and the fluorescence decay was dominated by a short-lived lifetime (37 ps) resulting from the formation of a **BPP·**G base pair stabilized through hydrogen bonding. This sensitivity to the opposite base illustrated the possible use of **BBP** for single nucleotide polymorphism (SNP) typing of genes [318]. However, the general utility of **BBP** in SNP typing was limited by the fact that **BBP** is also sensitive to the flanking nucleobases and notably to the G nucleobase, which dynamically quenches **BPP**. Later, the same group synthesized naphthopyridopyrimidine (**NPP**, Figure 30B), which showed a strong fluorescence when the opposite base was adenine [319]. In aqueous buffer, **NPP** had an absorption maximum at 364 nm (ε = 10200 M^{-1} cm^{-1}) and a fluorescence maximum at 395 nm (Φ_F = 0.13). Upon incorporation in ssDNA, Φ_F dropped to 0.06 probably due to self-aggregation of the single strands. In the dsDNA context, the optical properties of **NPP** are able to distinguish an opposite A from G by a higher Φ_F (0.01) for A than for G where Φ_F fell to 0.

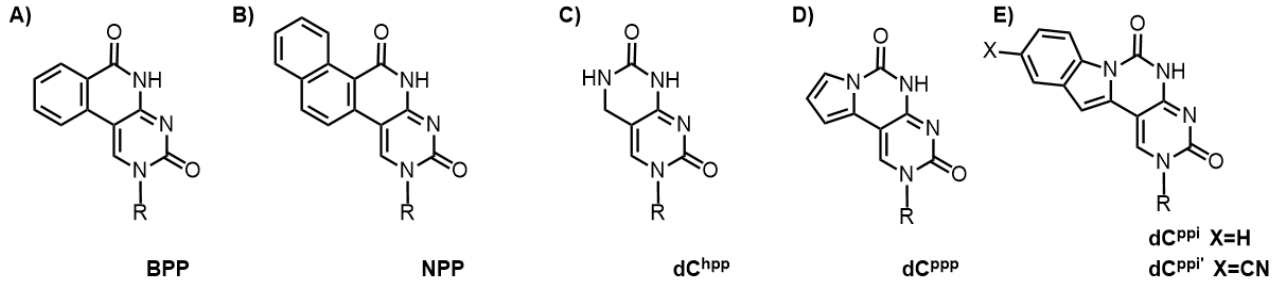

Figure 30. A) Benzopyridopyrimidine (**BPP**), B) Naphthopyridopyrimidine (**NPP**). C-E) Fluorescent cytosine analogues synthesized by the group of Sekine; R = 2'-deoxyribose.

3.1.3. Additional expanded cytosine derivatives. In 2006, Sekine et al. synthesized the bicyclic 4-N-carbamoyldeoxycytidine, **dC^{hpp}** (Figure 30C), a fluorescent cytosine analogue forming stable base pairs both with guanine and adenine. This compound had absorption maximum at 295 nm and fluorescence maximum at 360 nm with Φ_F = 0.12 in aqueous buffer at pH=7. Upon incorporation in dsDNA, its optical properties remained unchanged except for the case where it base paired with guanine, where a drop in Φ_F was observed. However, no systematic study was done to decipher its photophysics. Next, the same group synthesized a series of bi- and tricyclic deoxycytidine derivatives with improved optical properties compared to the parent **dC^{hpp}** nucleobase. Among them, the free pyrrolo-ring derivative (**dC^{ppp}**, Figure 30D) [320] had an absorption maximum at 369 nm (ε = 4760 M^{-1} cm^{-1}) with a strong Stokes shift of about 120 nm and Φ_F = 0.11 in aqueous buffer. However, these derivatives were not incorporated in ONs. Finally, Sekine's group synthesized substituted pyrimidopyrimidoindole cytidine derivatives such as **dC^{ppi}** (Figure 30E) [321] that form stable base pairs with guanine in duplexes. These fluorescent nucleobases had absorption maxima in the 378−390

nm range with fluorescence maxima ranging from 465 nm to 511 nm. The absorption and fluorescence spectra wee respectively, red- and blue-shifted compared to the free nucleoside. The red-shift in the absorption is likely a consequence of its H-bonding and base stacking while the blue shift in emission may result from the hydrophobic environment in the duplexes. In most cases, incorporation of **dC^{ppi}** and its derivatives in ss- and dsDNA resulted in fluorescence quenching relative to the parent **dC^{hpp}** nucleoside. The only exception was **dC$^{ppi'}$** (Figure 30E) which displayed Φ_Fs ranging from 0.02 to 0.26, the lowest value being obtained for a base pair with guanine, where photoinduced electron transfer is involved in the quenching mechanism. **dC^{ppi}** was applied to monitor the formation of DNA triplexes [322], resulting from the association of ssDNA with dsDNA via WC and Hoogsteen base pairing. **dC^{ppi}** was introduced into a duplex with an opposite abasic site containing a propylene linker. Molecular modelling suggested that this design enabled the aromatic ring of **C^{ppi}** to intercalate into the free space of the abasic site. As a consequence, the fluorescence intensity of **dC^{ppi}** increased by 3-fold upon formation of the triplex. Quenching experiments with potassium iodide confirmed that **C^{ppi}** was in the pocket formed by the abasic site. Later, Sekine's group designed ONs bearing both **2AP** and **dC^{ppi}**. These doubly labelled ONs exhibited a dual emission (370 nm for **2AP** and 500 nm for **dC^{ppi}**) upon excitation at 300 nm [323]. Ratiometric experiments with these doubly labelled ONs showed 5 and 2-fold increase for **2AP** and **dC^{ppi}**, respectively during duplex formation with abasic sites facing one or both dyes [323].

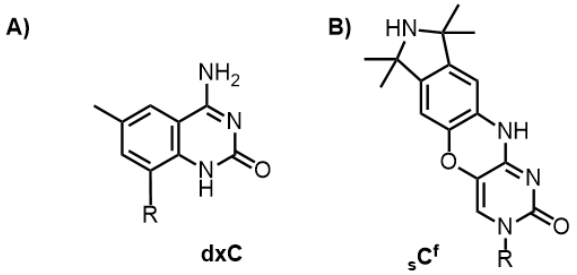

Figure 31. Chemical structures of the benzo-expanded cytosine **dxC** (A) and **$_s$C^f** (B). R = 2'-deoxyribose.

Based on the work of Leonard et al., that first described the synthesis of benzo-expanded nucleosides [324], Kool et al. produced a prospective expanded cytosine derivative (**dxC**, Figure 31A) by annealing an aromatic ring to the nucleobase core [325]. In methanol, **dxC** had an absorption maximum at 330 nm (ε =4100 M^{-1} cm^{-1}) and a fluorescence maximum at 388 nm (Φ_F = 0.52). In aqueous buffer, Φ_F dropped to 0.31. The authors synthesized ONs exclusively composed of (**dxC**)$_n$,. Their Φ_Fs were 0.16 and 0.21 for n=2 and n=4, respectively. In duplexes, the optical properties of **dxC** were unaltered except when **dxC** was paired with guanine, which led to strong quenching by photoinduced electron transfer [326]. As an application, **dxC** was used to monitor enzymatic reactions [327], using template independent polymerases and terminal deoxynucleotidyl transferase (TdT). Using **dxCTP** triphosphate as a TdT substrate, the changes in fluorescence intensity were used to monitor the incorporation of up to 30 nucleotides onto the DNA primer and thus, helped determining the kinetic parameters of the reaction (Figure 32). Finally, the TdT-mediated DNA tailing reaction with **dxCTP** was used in a microscopy assay, to monitor apoptosis-related DNA fragmentation in fixed cells.

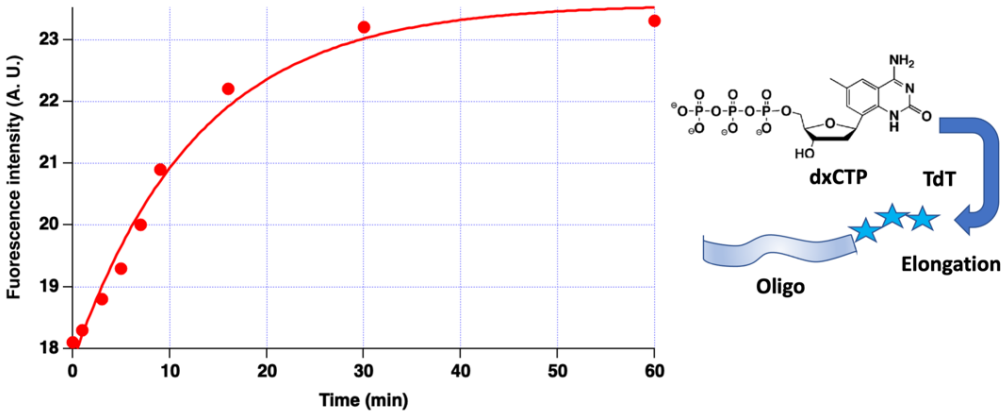

Figure 32. Changes in fluorescence emission during TdT-mediated **dxCTP** incorporation measured at 395 nm (excitation 330 nm). The solid line is an exponential fit used to determine the kinetic rate of the enzymatic reaction. Adapted from ref. [327]

Sigurdsson's group synthesized the cytosine analog $_{s}\mathbf{C}^{\mathbf{f}}$ (Figure 31B) characterized with an absorption maximum at 364 nm (ε = 11600 M^{-1} cm^{-1}) and fluorescence maximum at 446 nm (Φ_F = 0.31 in aqueous buffer) [328]. Upon incorporation in ssDNA, the positions of absorption and emission maxima were marginally affected while the Φ_F decreased by up to 30%, depending on the flanking residues. The highest brightness was observed when $_{s}\mathbf{C}^{\mathbf{f}}$ had a 5' flanking guanine or adenine. The fluorescence spectra of $_{s}\mathbf{C}^{\mathbf{f}}$ in matched and mismatched dsDNA showed that $_{s}\mathbf{C}^{\mathbf{f}}$ can report on the nature of the mismatch especially with opposite adenine for which the fluorescence spectrum appeared more structured. This structured emission probably resulted from vibrational replicates associated to a $_{s}\mathbf{C}^{\mathbf{f}}$ tautomer that stabilized the adenine mismatch through additional H-bonds. Thus, $_{s}\mathbf{C}^{\mathbf{f}}$ was proposed to detect SNPs [329].

3.1.4. Expanded thymine derivatives In 1998, Toulmé et al. synthesized **BgQ**, a very fluorescent thymine analogue (Figure 33A) [330,331]. The free nucleoside had an absorbance maximum at 360 nm (ε = 76000 M^{-1} cm^{-1}) and a fluorescence maximum at 434 nm with Φ_F = 0.82 (pH=7). **BgQ** formed a WC base pair with A. Its Φ_F decreased by 15% when **BgQ** was flanked by cytosines and thymines and by 100-fold when flanked by two guanines. These changes were attributed to base staking and electron transfer. **BgQ** was used to monitor the formation of DNA triplexes [331].

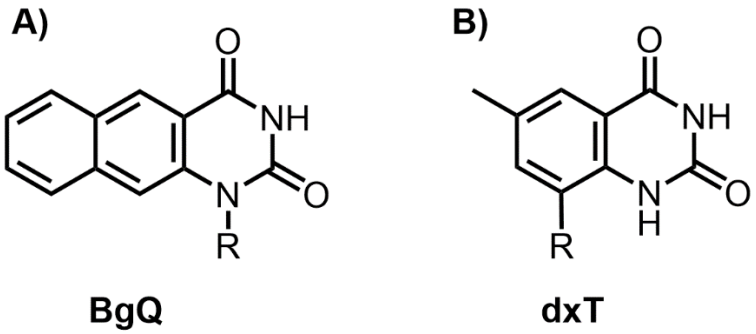

Figure 33. Expanded thymine derivatives. A) Benzo[g]quinazoline-2,4-(lH,3H)-dione (**BgQ**) and B)methylquinazolinedione C-glycoside (**dxT**); R = 2'-deoxyribose.

Using the same strategy as for the expanded cytosine derivative **dxC** (Figure 31), Kool's group developed the expanded dT derivative **dxT** (Figure 33B) [332]. In methanol, **dxT** had an absorption

maximum at 320 nm (ε = 3400 M^{-1} cm^{-1}) and a fluorescence maximum at 377 nm with Φ_F = 0.3. In aqueous buffer, the fluorescence maximum shifted to 385 nm with an unchanged Φ_F [333]. In ssDNA, its Φ_F strongly depended on the nature of its flanking nucleobases (0.01 for 5'-A-**xT**G-3'and 0.33 for 5'-C-**xT**T-3'). Upon duplex formation of 5'-C**xT**T-3' with its complementary ssDNA, the Φ_F ranged between 0.05 (T mismatch) and 0.09 (abasic site). Stability experiments revealed that **dxT** had a destabilizing effect compared to T in A·T pairs but still displayed selective recognition of its hydrogen-bonding complementary partner, suggesting that WC pair can still form despite the local backbone strain.

In 2016, Luedtke's group synthesized ^{DMA}**T**, a thymine derivative (Figure 34A) with the same pK_a and base pairing characteristics as the parent nucleobase [334]. In aqueous buffer, it had an absorbance maximum at 357 nm (ε = 2900 M^{-1} cm^{-1}) and a fluorescence maximum at 522 nm with Φ_F = 0.03. As predicted by DFT calculations, the absorption and emission spectra at high pH were blue shifted as the result of the formation of an anionic species characterized by Φ_F = 0.09. The measured pK_a was 9.5. At both neutral and basic pH, deuteration led to an increase in Φ_F (0.07 and 0.19, respectively) suggesting that proton transfer acted as an efficient non-radiative pathway.

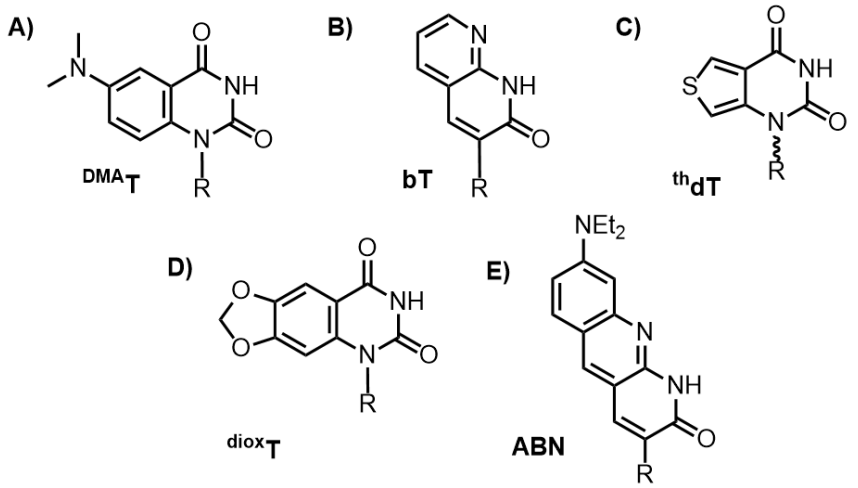

Figure 34. Chemical structures of A) N,N-dimethylaniline-2-deoxy-thymidine (^{DMA}**T**), B) bicyclic thymine (**bT**), C) thieno[3,4-d]-pyrimidine (th**dT**), D) dioxoloquinazoline-based nucleoside^{diox}**T**, and E) 8-(diethylamino)benzo[b][1,8]naphthyridin-2(1H)-one-based nucleoside **ABN**; R = 2'-deoxyribose.

^{DMA}**T** was incorporated in 21-residue DNA sequences both at internal positions and at the 5' end to evaluate the impact of the nature of the flanking nucleobases and DNA end ''breathing'' motions. Φ_F of internal ^{DMA}**T** was in the range 0.05 −0.09 in ssDNA and in the range 0.11−0.20 in matched dsDNA. Φ_F of ^{DMA}**T** was very similar to the free nucleoside when it was at the 5' end of ss- or dsDNA. Moreover, ^{DMA}**T** showed a substantial decrease in Φ_F as well as a red-shifted absorption and fluorescence when ^{DMA}**T** was mismatched with thymine, guanine or cytosine nucleobases, probably due to changes in hydration and base stacking. ^{DMA}**T** was also used to probe Hg^{II} ions that are known to induce T-Hg-T base pairs leading to miscoding during DNA synthesis [334]. Adding increasing Hg^{II} concentrations to ^{DMA}**T**-labeled duplex led to a bi-phasic fluorescence quenching showing site-specific binding of Hg^{II} ions to ^{DMA}**T**·T sites (Figure 35A). The first phase with a steep slope was associated with T-T-specific binding, while the shallow slope was associated to non-specific interactions (Figure 35B).

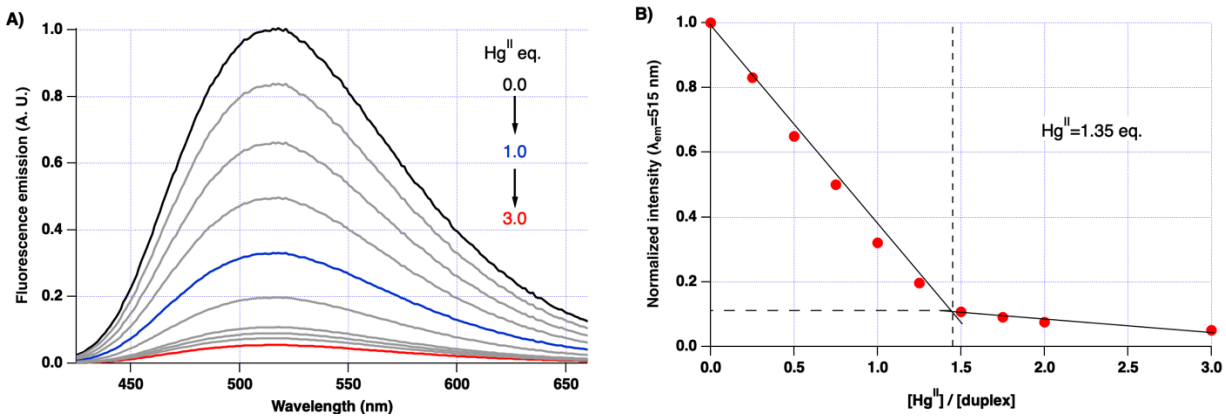

Figure 35. (A) Fluorescence spectra of a ^{DMA}**T**-labeled duplex in the absence (black) and in the presence of increasing $Hg(ClO_4)_2$ concentrations. (B) Plot of the changes in fluorescence intensity (λ_{em} = 515 nm) as a function of the [Hg^{II}]/[duplex] molar ratio. Adapted from ref. [334]

In 2018, Wilhelmsson et al. synthesized and characterized the optical properties of the bicyclic thymine analog **bT** (Figure 34B) [335]. The deoxyribonucleoside **bT** had absorption maxima in water and ethanol of 321 and 323 nm (ε = 15000 M^{-1} cm^{-1}), respectively, while its fluorescence maxima were 368 nm (water) and 370 nm (ethanol) with a nearly identical Φ_F (~0.05). A red shift of both absorption and fluorescence maxima was observed upon incorporation of **bT** in ssDNA and dsDNA. On average, the absorption coefficient decreased to 11500 M^{-1} cm^{-1} for ssDNA and to 6200 M^{-1} cm^{-1} for dsDNA as a consequence of strong base stacking. Φ_F of **bT** decreased to ~0.01 in ssDNA and to an even lower value in dsDNA. In both ssDNA and dsDNA, Φ_F was strongly dependent on the nature of the flanking nucleobase. For instance, Φ_F of **bT** in dsDNA ranged from 0.015 (with flanking thymines) to 0.001 (with a flanking guanosine). This strong quenching by G was thought to be caused by a photoinduced electron transfer from G to **bT**.

A bright U-mimic ribonucleoside (th**U**, Figure 8) and a T-mimic deoxynucleoside (th**dT**, Figure 34C) from the thieno[3,4-d]-pyrimidine family discussed in section 2.2 were reported by the groups of Tor and Sugiyama [294,336]. In water (Table 3), thU and thdT exhibited similar absorption maximum (303 nm and 304 nm, respectively) and extinction coefficient (ε ~3100 M^{-1} cm^{-1}), but differed by their fluorescence maxima (409 nm and 420 nm, respectively), Φ_F (0.41 and 0.64, respectively) and fluorescence lifetimes (11.5 ns and 16 ns, respectively). The origin of these differences has not been discussed but may be related to the nature of the sugar or the presence of tautomers, which can strongly affect the measured fluorescence parameters as has been shown for th**G** [226] . Thermal stability measurements with matched and mismatched duplexes showed that th**dT** was a perfect isomorphic T analogue.[336] Its optical properties were very similar in matched and mismatched duplexes, except for dsDNA containing an abasic site opposite th**dT** where its fluorescence intensity increased, suggesting that it can be used to detect abasic sites in the context of base-excision repair systems. Sugiyama's group also investigated the enzymatic incorporation of th**dT** into DNA in the form of deoxythymidine triphosphate analogue (th**dTTP**). To demonstrate that th**dTTP** can be processed by DNA polymerases, primer extension and polymerase chain reaction were performed using the 3'-5' exonuclease proficient Klenow fragment and KOD Plus polymerase. Despite displaying isomorphic properties, th**dTTP** did not perfectly replace dTTP upon enzymatic amplification of long DNA fragments, so that a mixture of dTTP/th**dTTP** had to be used. A similar conclusion has been reached when the modified ribonucleotide triphosphate th**UTP** was used as a substrate for enzymatic synthesis of RNA with T7 RNA polymerase.[294]

In 2019, Sugiyama's group synthesized a new size-expanded T nucleoside based on the dioxoloquinazoline core: **^{diox}T** (Figure 34D) [337]. In water, the free nucleoside had an absorption maximum at 325 nm (ε = 10900 M^{-1} cm^{-1}) and a fluorescence maximum at 388 nm with Φ_F = 0.36 and a single fluorescence lifetime (3.5 ns). Using solvents of different polarity, a correlated decrease of Φ_F and fluorescence lifetime was observed, indicating a possible charge transfer character of the excited state. **^{diox}T** was incorporated in ss- and dsDNA by systematically changing the 5' and 3' neighbors as well as the opposite base. Replacement of T by **^{diox}T** resulted in a limited destabilization, as suggested by the 1 to 4°C change in T_m values. The absorption maxima were red shifted in dsDNA relative to ssDNA, while the fluorescence maxima exhibited a blue shift upon duplex formation (Figure 36). These observations were consistent with the increased stacking interactions and lower polarity in dsDNA. The lowest Φ_F values (≤0.02) in both ss and dsDNA were observed when **^{diox}T** was flanked by a G residue. Calculations of the HOMO energy indicated that G>**^{diox}T**>A>C>T, indicating that the quenching of **^{diox}T** by guanine probably resulted from an electron transfer from guanine. Interestingly, with flanking A or T bases, Φ_F increased by 1.5-fold upon duplex formation. HOMO/LUMO calculations suggested that a substantial charge transfer could occur from the dioxole ring to the benzene ring fused with the T moiety. This intramolecular charge transfer of **^{diox}T** might contribute to the Φ_F enhancement upon duplex formation. When **^{diox}T** was flanked by T and A residues, its emission was found to be sensitive to the nature of the opposite base in a duplex (Figure 36 A and B), which favored its use for SNP typing [337]. Finally, ss- and dsDNA containing three adjoining or alternating **^{diox}T**s were prepared. Their optical properties revealed that self-quenching decreased the fluorescence intensity for adjoining **^{diox}T**s as compared to alternating ones, opening the route for self-quenching based applications to probe conformational changes in nucleic acids. To further extend its applications, **^{diox}T** was recently used as a metal sensor in a study where Sugiyama et al. demonstrated that **^{diox}T**·T and **^{diox}T**·C mismatched base pairs in duplexes can serve as a nucleic-acid-based mercury sensor.[338] The group of Sugiyama also investigated the possibility of using **^{diox}T** as a FRET donor in combination with **tC** acting as an acceptor. Both were incorporated in double-strand DNA in which the distance between the two dyes was systematically increased [339]. Similarly to the experiments performed by Wilhelmsson et al. (See Figure 29), the authors were able to correlate the change in the FRET efficiency with the modification of the orientation factor (κ^2) and distance between the FRET pair. These results evidenced the fixed orientation of **^{diox}T** and **tC** in the double-strand. With respect to the previously reported FRET pair based on **thdG** and **tC**[237], the **^{diox}T**/tC pair displays two well distinct emission spectra making FRET efficiency determination more accurate using steady-state and time-resolved measurements.

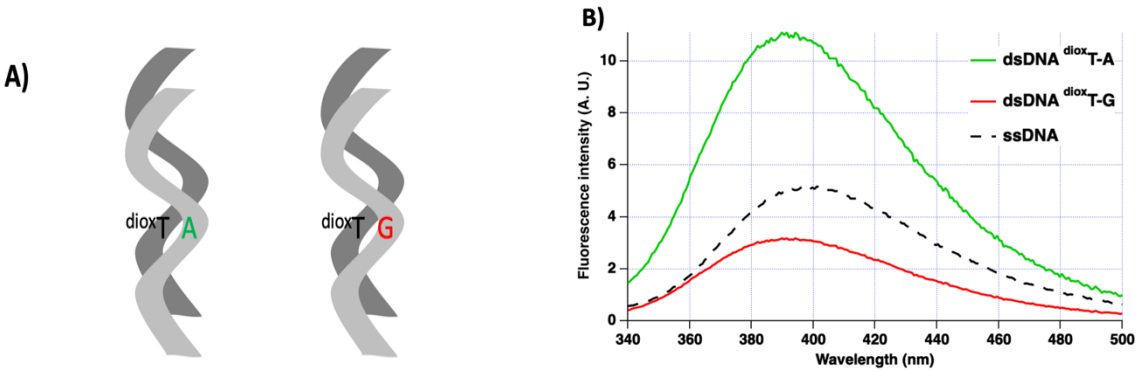

Figure 36. Applications of **^{diox}T** for single nucleotide polymorphism (SNP) typing. A) Schematic illustration of SNP typing with **^{diox}T**-labeled oligonucleotides. B) Fluorescence spectra of dsDNA labeled with **^{diox}T** opposite G (red) or A (green). The **^{diox}T**-labelled oligonucleotide in its single strand

form is given for comparison (dotted line). **^{diox}T** is flanked by T and A residues in the labeled oligonucleotide. Adapted from ref. [337]

More recently, using rhodamine B as a model, Purse et al. further modified the push-pull properties of **tC** by introducing **ABN** (Figure 34E)[340]. The free nucleoside, present only in the T-like tautomeric form, had an absorption and fluorescence maxima centered at 442 nm (ε = 20 000 M^{-1} cm^{-1}) and 540 nm (Φ_F=0.39), respectively. When **ABN** was incorporated in oligonucleotides, the DNA duplexes were destabilized to some extent. The destabilization was most pronounced when **ABN** was flanked by 5'-G and 3'-C neighbors. The optical properties of **ABN** were alsocharacterized in ss and ds-DNA. In ss-DNA, the absorption and fluorescence maxima ranged between 450–452 nm and 530–540 nm, respectively, with Φ_F increasing from 0.49 (**ABN** flanked by two adenines) to 0.62 (**ABN** flanked by G and C). For ds-DNA, the absorption ranged between 440 nm (adenine as the opposite base) and 470 nm (guanine as the opposite base) with a corresponding fluorescence of 530 nm (Φ_F = 0.53) and 523 nm (Φ_F = 0.29). Based on its exceptional brightness, the authors used fluorescence correlation spectroscopy and wide-field imaging to perform single-molecule experiments using one and two-photon excitation. The molecular brightness measured in wide-field imaging was 15 kHz/molecule (power density 0.2 kW/cm^{-2}) with a mean "on" time of 0.35 s. Through its high brightness and red shifted absorption and fluorescence, **ABN** appears thus as a highly promising FNA for single molecule applications.

Table 3. Photophysical properties of expanded pyrimidine and purine fluorescent nucleoside analogues

Name	**$\lambda_{abs,m}$ (nm)**			**ε (10^{-3} $M^{-1}cm^{-1}$)**	**$\lambda_{F,m}$ (nm)**			**Φ_F**			**Lifetimes (ns)**			**Refs**
	Nucleoside	**ssNA**	**dsNA**	**Nucleoside**	**Nucleoside**	**ssNA**	**dsNA**	**Nucleoside**	**ssNA**	**dsNA**	**Nucleoside**	**ssNA**	**dsNA**	
thC	320	320	320	4.53	429	429	429	0.41	n.d.	n.d.	15.2	n.d.	n.d.	[200,204]
thU	304	304	304	3.16	409	409	409	0.41	n.d.	n.d.	11.5	n.d.	n.d.	[200]
thT	303	303	303	3.10	420	420	420	0.64	n.d.	n.d.	16	n.d.	n.d.	[336]
^{tz}C	325	n.d.	n.d.	5.45	411	n.d.	n.d.	0.053	n.d.	n.d.	n.d.	n.d.	n.d.	[201]
^{tz}U	312	n.d.	n.d.	5.17	392	n.d.	n.d.	0.01	n.d.	n.d.	n.d.	n.d.	n.d.	[201]

pC	340	340	n.d.	5.9	449	447–456	n.d.	0.026	0.021–0.044	0.026–0.05	2.1[a]	2-3[a,*]	n.d.	[304]
thpC	357	n.d.	n.d.	11.6	457	n.d.	n.d.	0.41	n.d.	n.d.	n.d.	n.d.	n.d.	[308]
tC	375	400	400	4.5	505	505	505	0.20	0.20	0.20	3.7[a]	6.2[a]	5.9[a]	[311]
tC^{O}	360	363–370	363–369	9	461	446–461	447–461	0.30	0.14–0.41	0.18–0.30	3.4[a]	2.2–3.4[a,*]	3.4–4.8[a]	[312]
8-DEA-tC	395	413-425	348-422	2.7	493	492-499	492-501	0.006	0.008-0.032	0.007–0.12	n.d.	n.d.	n.d.	[316]
BPP	n.d.	347	347	n.d.	n.d.	390	390	n.d.	0.04	0.002–0.035	n.d.	0.16-2.4[b]	0.30-3.3[b]	[317]
NPP	364	364	364	10.2	395	395	395	0.13	0.06	0–0.01	n.d.	n.d.	n.d.	[319]
dC^{hpp}	295	295	295	n.d.	360	360	360	0.12	0.12	0–0.12	n.d.	n.d.	n.d.	[341]

dC$^{ppi'}$	n.d.	371	378	n.d.	n.d.	475	465–477	n.d.	0.12	0.2-0.27	n.d.	n.d.	n.d.	[321]
dxC	330[c]	330	330	4.1[c]	388[c]	392	392	0.52[c]	0.31	0.25	4[a]	n.d.	n.d.	[325]
$_{s}$C^{f}	364	359-364	359-364	11.6	446	442-450	442–450	0.31	0.10–0.25	0.07-0.32	n.d.	n.d.	n.d.	[328]
BgQ	360	360	360	76	434	434	434	0.81	0.008–0.62	0.16–0.81	n.d.	n.d.	n.d.	[330]
dxT	320[a]	320	320	3.4[a]	377[a]	385	385	0.30[a]	0.01–0.33	0.05–0.09	n.d.	n.d.	n.d.	[333]
^{DMA}T	357	355	355-370	2.9	522	486-505	486-508	0.03	0.03-0.20	0.05-0.20	n.d.	n.d.	n.d.	[334]
bT	321	323–327	327–330	15	368	367–371	372–380	0.05	0002–0.026	0001–0.015	n.d.	n.d.	n.d.	[335]
thdT	303	310	310	3.1	420	420–425	415–420	0.64	0.016–0.057	0.021–0.027	16[a]	n.d.	n.d.	[336]

^{dioxT}	325	321–330	324–330	10.9	388	384–393	374–383	0.36	0.01–0.23	<0.01-0.20	3.5[a]	1.8-3.5[d]	0.4-3.2[d]	337
ABN	442	450–452	440–470	20	540	530–540	523–532	0.39	0.49–0.62	0.29–0.55	n.d.	n.d.	n.d.	340
^{eth}A	294	n.d.	n.d.	3.7	410	n.d.	n.d.	0.60	n.d.	n.d.	23[a]	n.d.	n.d.	342
^{MD}I	315	n.d.	n.d.	7.8	442	424	424	0.12	n.d.	0.001–0.08	n.d.	n.d.	n.d.	343
^{MD}A	327	n.d.	n.d.	2.7	427	424	424	0.12	n.d.	0.002–0.1	n.d.	n.d.	n.d.	343
NDA	n.d.	352	352	n.d.	n.d.	375	380	n.d.	0.01	0.005–0.3	n.d.	n.d.	1.1[a]	344
dxA	333[c]	333	333	11.08[c]	393[c]	393	393	0.44[c]	0.08	0.08	3.3[c]	1.4–5[b]	n.d.	325
dxG	320[c]	320	320	3.4[c]	413[c]	393	413	0.40[c]	0.08	0.08	7.8[c]	1.6–10[b]	n.d.	325

A^T	282	n.d.	n.d.	16.5	353	351–358	349–354	0.61	0.005–0.21	0.003–0.05	1.8[a]	0.9–1.5[d]	0.2–1.2[d]	[345]
qA	335	337–338	331–339	5	456	432–443	434–459	0.068	0.002–0.058	0.002–0.006	3.2[a]	0.4–2.0[d]	0.4–1.4[d]	[346]
2-CNqA	356	357–360	355–358	10.8	482	452–457	448–459	0.42	0.1–0.42	0.22–0.32	10	2.8–11	7.4–9.7	[347]
qAN4	356	358	358	7.3	445	414–424	406-414	0.32	0.01–0.29	0.008–0.07	4.8[a]	0.2–5.9[d]	0.2–1.7[d]	[348,349]
pA	387	388–392	388–392	15.2	420	405–409	400–406	0.66	0.024–0.58	0.034–0.22	3.9[d]	0.4–7.4[d]	0.7–3.8[d]	[350,351]

[a]: single component decay; [b]: multiple components decay; [c]: measured in methanol; [d]: mean lifetime.
[*]: apart when G is a neighbor
n.d.: not determined

3.2. Expanded purine analogues

Similar to fluorescent pyrimidine analogues, purines have been also modified using the same principles to provide expanded purine analogues able to sense DNA dynamics and interactions with minimal perturbations. In this section, we will highlight several examples of expanded adenine and guanine derivatives.

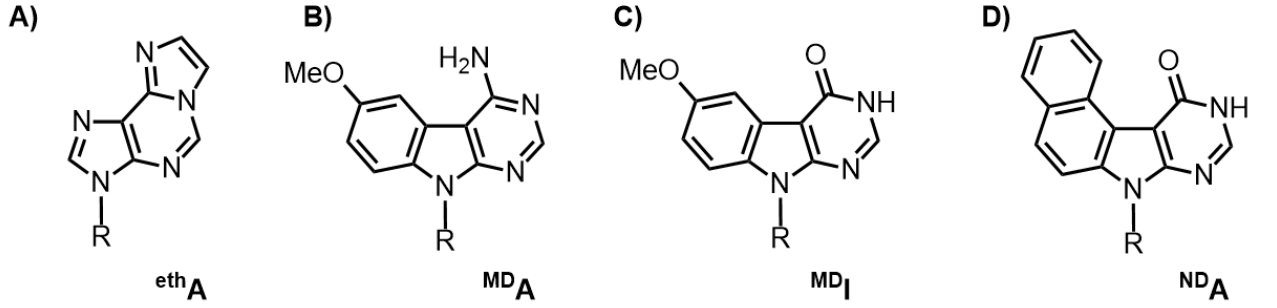

Figure 37. Chemical structure of fluorescent expanded purine analogs. A) etheno-adenosine (**^{eth}A**), B) methoxybenzodeaza-adenosine (**^{MD}A**), C) methoxybenzodeaza-inosine (**^{MD}I**) and D) naphthodeaza-adenosine (**NDA**); R = ribose or 2'-deoxyribose.

In 1972, Leonard et al. reported etheno-adenosine **^{eth}A** (Figure 37A) [342]. In buffer at pH=7, **^{eth}A** had an absorption maximum at 294 nm (ε = 3700 M^{-1} cm^{-1}) and a fluorescence maximum at 410 nm with Φ_F = 0.6 and a long fluorescence lifetime of 23 ns, that makes it highly suited for time-resolved fluorescence applications. **^{eth}A** replaced almost perfectly the adenine moiety in ATP for enzymatic reactions [342]. However, due to the etheno moiety, **^{eth}A** was not able to form WC base pairs with T, perturbing thus the dsDNA structure [352] and limiting its use. Later, Leonard et al. developed a library of adenine derivatives fused to benzene moiety giving rise to linear and bended fluorescent nucleoside analogues [324]. For instance, the linear benzoadenosine had an absorption maximum at 348 nm and a fluorescence maximum at 372 nm with Φ_F = 0.44 and τ = 3.7 ns. In enzymatic assays with adenosine deaminase (ADA), the kinetic parameters with benzoadenosine were very similar to those with adenosine, further supporting their use in this context [324].

In 2003, the group of Saito proposed two fluorescent purine analogs: methoxybenzodeaza-adenosine (**^{MD}A**) and methoxybenzodeaza-inosine (**^{MD}I**) (Figure 37B and C) [343]. **^{MD}I** and **^{MD}A** were characterized by absorption maxima at 315 nm (ε = 7800 M^{-1} cm^{-1}) and 327 nm (ε = 2700 M^{-1} cm^{-1}), respectively. In aqueous buffer, their fluorescence spectra had maxima at 442 nm and 427 nm, respectively, both having a Φ_F of 0.12. Upon incorporation in ssDNA and matched dsDNA (**^{MD}A**·T and **^{MD}I**·C), Φ_F decreased 24- to 100-fold. In mismatched duplexes, the Φ_Fs were somewhat higher: **^{MD}A**·C (Φ_F = 0.08), **^{MD}A**·G (Φ_F = 0.02), **^{MD}I**·T (Φ_F = 0.01) [343]. Thus, **^{MD}I** and **^{MD}A** were used to detect SNP of the T/C SNP sequence of the human breast cancer 1 gene. In their quest of developing base discriminating fluorescent nucleoside analogues, Saito's group synthesized naphthodeaza-adenosine **NDA** (Figure 37D) [344]. In dsDNA, the absorption maximum of **NDA** was 352 nm (with opposite G, T and A) and 354 nm (with opposite C). Its fluorescence maximum was at 380 nm for all dsDNAs. While the fluorescence wavelength was insensitive to the nature of the opposite base, Φ_F was sensitive: 0.03 for the **NDA**·C pairs and less than 0.005 for the remaining base pairs. In ssDNA, Φ_F of **NDA** was 0.01. Finally, fluorescence lifetime measurements validated the use of **NDA** in FRET applications with fluorescein acting as a FRET acceptor. The fluorescence decay of dsDNA with an **NDA**·C pair in the absence of fluorescein was characterized by a single fluorescence lifetime of 1.1

ns. When **NDA** and fluorescein were separated by 3 to 5 bases in the same ssDNA, the average lifetimes dropped to 0.58 ns and 0.69 ns giving FRET efficiencies of 83% to 48%, respectively [344].

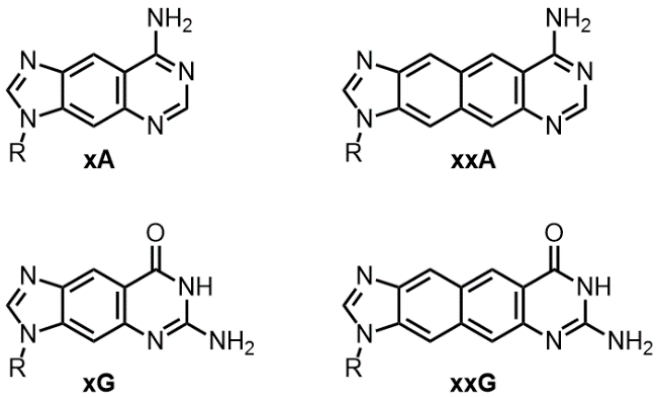

Figure 38. Chemical structures of mono (**x**-) and doubly-expanded (**xx**-) adenine and guanine deoxyribonucleosides; R = 2'-deoxyribose.

3.2.1. Mono (x-) and doubly-expanded (xx-) purine deoxyribonucleosides. In 2008, the group of Kool synthesized nucleosides with expanded purine bases to establish the xDNA family [325] (Figures 31, 33B and 38). In methanol, the free nucleosides, **dxA** and **dxG** (Figure 38) had absorption maxima at 333 nm (ε = 11080 M^{-1} cm^{-1}) and 320 nm (ε = 3400 M^{-1} cm^{-1}), respectively. Their fluorescence had maxima at 393 nm (Φ_F = 0.44, τ_F = 3.3 ns) and 413 nm (Φ_F = 0.4, τ_F = 7.8 ns), respectively. In the ssDNA context, their absorption properties remain unchanged. The fluorescence spectra of **dxA**-containing ssDNAs retained the monomer fluorescence maximum while the spectra of **dxG**-containing ssDNAs shifted 20 nm to the blue. Their Φ_F dropped by 5-fold in both cases. Upon duplex formation, the fluorescence properties ($\lambda_{F,m}$ and Φ_F) of **dxA** and **dxG** were weakly dependent on the nature of the opposite base. Finally, in ssDNA containing 1 to 4 consecutive **dxG**, Φ_F dropped from 0.08 (n=1) to 0.02 (n=4) due to increased self-quenching. In the case of (**dxA**)$_n$, an additional long wavelength fluorescence band (520 nm) associated with excimer formation appeared for $n \geq 2$ [325]. More recently, the possibility of further expanding the purine analogues was theoretically explored [353]. In particular, a comparison between the doubly expanded purines (**xxA** and **xxG**) and the singly expanded or natural purines was performed using QM methods. Analysis of the glycosidic torsional flexibilities in oligonucleotide models showed that xxDNA purines probably form WC base pairs with complementary pyrimidines. In addition, due to the further extension of the purine moiety, a 50% increase of the stacking energy with respect to the natural purines should strongly stabilize the duplexes. Noteworthy, **dxA** and **dxG** were used with DNA polymerase θ (Polθ), that performs translesion synthesis upon DNA repair [354]. Polθ efficiently and selectively incorporated **dxA** and **dxG** into DNA in the form of triphosphates. As Polθ was inhibited after incorporation of multiple **dxG** residues, this opened a route for the use of **dxG**, as a selective inhibitor of Polθ, that is highly expressed in cancer cells.

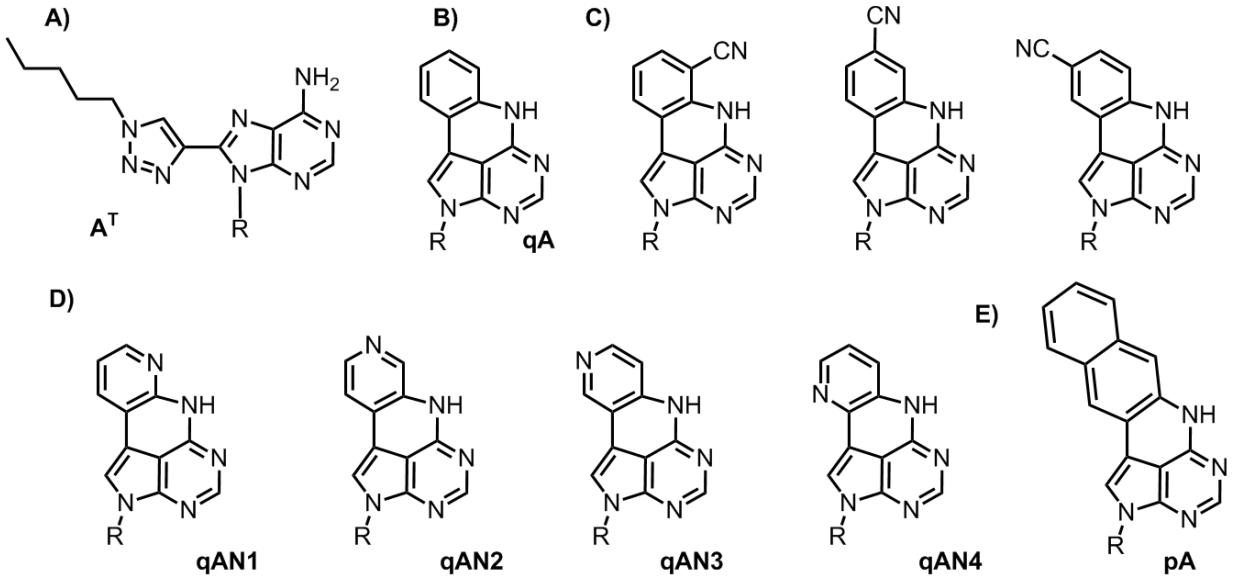

Figure 39. Chemical structures of polycyclic expanded adenine analogues. A) 8-(1-pentyl-1H-1,2,3-triazole-4-yl)-2'-deoxyadenosine ($\mathbf{A^T}$); B) the quadracyclic analogue **qA**; C) the cyano-substituted quadracyclic adenines; D) the aza-substituted quadracyclic adenines; E) the pentacyclic adenine, **pA** (R= Et or 2'-deoxyribose).

3.2.2. The 2'-deoxyadenosine analogue $\mathbf{A^T}$ with a 1,2,3-triazole residue. The group of Wilhelmsson designed and synthesized a large number of fluorescent purine analogues. For instance, they developed 8-(1-pentyl-1H-1,2,3-triazole-4-yl)-2'-deoxyadenosine ($\mathbf{A^T}$, Figure 39A) [345]. In aqueous buffer, $\mathbf{A^T}$ had an absorption maximum at 282 nm (ε = 16500 M^{-1} cm^{-1}) and a fluorescence maximum at 353 nm with Φ_F = 0.61 and a single lifetime of 1.8 ns. Fluorescence anisotropy experiments revealed a single transition dipole moment with parallel absorption and emission transition moments. While its fluorescence maximum was not significantly shifted upon incorporation into ss- and dsDNA, its Φ_F decreased significantly and ranged between 0.005 and 0.21 for ssDNA and 0.003−0.05 for dsDNA. The fluorescence decay of $\mathbf{A^T}$ in both ss- and dsDNA was best fitted with a three components model with an average lifetime ranging from 0.9 to 1.5 ns for ssDNA and 0.2 to 1.2 ns for dsDNA. The only exceptions were the ssDNA with A$\mathbf{A^T}$A and G$\mathbf{A^T}$C triplets, showing a mono-exponential decay. The decrease of Φ_F and the average lifetime was attributed to an increase of the non-radiative rate constant in combination with a decrease in the radiative rate constant. Photoinduced electron transfer from neighboring bases was invoked to explain the increase of the non-radiative rate constant. The decrease in the radiative rate constant was related to stacking interactions leading to a lower absorption coefficient, as expected from the Strickler-Berg equation. The multiexponential decay was associated to the existence of multiple species exhibiting different excited-state geometries (planar vs twisted).

3.2.3. Quadra- and pentacyclic adenine analogues. The quadracyclic 2′-deoxyadenosine analogue **qA**, (Figure 39B), synthesized by Matteucci et al [355], was characterized both spectroscopically and by TDDFT calculations in the group of Wilhelmsson [346]. In aqueous buffer, **qA** had an absorption maximum at 335 nm (ε = 5000 M^{-1} cm^{-1}) and a fluorescence maximum at 456 nm with Φ_F = 0.068 and single fluorescence lifetime of 3.2 ns. **qA** exhibited three distinct absorption transition dipole moments, one around 400 nm that was parallel to the emission transition moment ($S_0 \rightarrow S_1$), and two others ($S_0 \rightarrow S_2$ and $S_0 \rightarrow S_3$, respectively) with different angles relative to the constant emission dipole moment. **qA** was also incorporated in ss- and dsDNA and its optical properties were systematically

investigated as a function of its flanking residues. The fluorescence maximum was blue shifted ($\lambda_{F,m}$ < 440 nm) relative to the free monomer, in good agreement with the decreased polarity in DNA. A decrease of the mean fluorescence lifetime and Φ_F which depended on the nature of the surrounding bases was also observed. To accurately fit the fluorescence decays, two to three lifetime components were required ranging from 0.1 to 4.1 ns. The obtained average lifetimes of **qA** varied between 0.4 and 2.0 ns in ssDNAs and between 0.4 and 1.4 ns in dsDNAs. For ssDNA, the average decrease in Φ_F was about 11-fold and depended on **qA**'s neighbors. The highest Φ_F was measured when **qA** was flanked by two purines (Φ_F > 0.01) and it decreased gradually for sequences with only one purine neighboring **qA** at its 5' side (~0.018) or at its 3' side (~0.008) and finally for sequences where **qA** was flanked by two pyrimidines (< 0.004). For the corresponding duplexes, an additional 4-fold decrease in Φ_F was observed. The observed decrease in Φ_F and lifetimes upon incorporation of **qA** into DNA probably resulted from base-stacking that allowed charge transfer through molecular orbital overlap with the surrounding nucleobases as well as non-radiative relaxation through CI [346].

The available set of quadracyclic adenines was further extended with the synthesis of **qA** derivatives featuring different substituents (-F, -OMe, -CN) in the fused aromatic ring. The cyano-substituted **qA** derivatives are shown as an illustration in Figure 39C [356]. Also, a set of aza-substituted **qA** derivatives **qAN1** to **qAN4** was generated by introducing an additional nitrogen heteroatom at different positions in the annealed ring (Figure 39D) [348]. These derivatives were designed with the help of QM calculations (TDDFT) using the oscillator strength as a discriminating factor. In water, **2-CNqA** had an absorption maximum at 356 nm (ε = 10800 M^{-1} cm^{-1}) and a fluorescence maximum at 482 nm with Φ_F = 0.42 and a fluorescence lifetime of 10 ns. Thus, the brightness of **2-CNqA** (ε Φ_F = 4500 M^{-1} cm^{-1}) was 13-fold increased relative to the parent compound **qA** and was virtually insensitive to its microenvironment. The **2-CNqA** derivative was introduced in ssDNA and dsDNA where its neighbors were systematically varied [347]. **2CNqA** retained a high Φ_F (0.1−0.4) and molecular brightness (1000–4300 M^{-1} cm^{-1}) in ssDNAs and dsDNAs.

Among the four adenine analogues of the **qAN** series (Figure 39D), **qAN4** showed a 5-fold increase in the oscillator strength relative to the free base **qA** and appeared thus as the most promising candidate. In water, **qAN4** had an absorption maximum at 356 nm (ε = 7300 M^{-1} cm^{-1}) and a fluorescence maximum at 445 nm with Φ_F = 0.32 and a single fluorescence lifetime of 4.8 ns. The fluorescence properties of **qAN4** under acidic conditions (pH=3−6) evidenced an acid/base equilibrium paving the way for wavelength-ratiometric pH sensing under acidic conditions. While **qAN1** exhibited slightly weaker fluorescence properties compared to **qAN4**, **qAN2** and **qAN3** were much less bright due to a weak $S_0 \rightarrow S_1$ transition and an excited-state proton-transfer reaction (Φ_F = 0.06 and 0.01). Upon incorporation in ss- and dsDNA, the absorption properties of **qAN4** were not modified while its fluorescence maximum was blue shifted (410 nm) in line with the lower polarity in DNA compared to water [349]. The Φ_F (0.01−0.29 for ssDNA, 0.01−0.07 for dsDNA) and the mean fluorescence lifetime (0.2−5.9 ns for ssDNA and 0.2−1.7 ns for dsDNA) depended in the same way as the parent molecule **qA** on the nature of the neighboring bases, gradually decreasing when going from purine to pyrimidine neighbors. At pH=5 and 7.5, both absorption and fluorescence spectra showed additional bands associated to the protonated form of **qAN4** demonstrating that **qAN4** can be used as a pH-sensitive probe inside nucleic acids. Finally, **qAN4** in combination with $\mathbf{qA_{nitro}}$, a non-fluorescent **qA** analogue bearing a nitro group at the position 3 of the annealed aromatic ring [357] was shown to enable accurate FRET measurements over more than one turn of B-DNA [349].

In 2018, Bood et al. synthesized **pA**, a pentacyclic fluorescent adenine analogue (Figure 39E) [350]. The free nucleoside had an absorption maximum at 387 nm (ε = 15200 M^{-1} cm^{-1}) and fluorescence

maximum at 420 nm (Φ_F = 0.66 and mean τ_F =3.9 ns) in water. Due to its high absorption coefficient and Φ_F, the probe exhibited a brightness of 10000 M^{-1} cm^{-1}. Steady-state and time-resolved spectroscopies revealed the existence of two-ground state tautomers with Φ_F and τ_F differing by more than one order of magnitude [351]. The optical properties of **pA** upon incorporation in ss- and dsDNA were investigated for 16 different combinations with neighboring bases. In aqueous solution, the absorption and emission maxima were red- (≈3 nm) and blue-shifted (≈15 nm) relative to the monomer, as a result of base stacking and polarity decrease. Φ_F of **pA** ranged from 0.024 to 0.56 in ssDNA and from 0.034 to 0.22 in dsDNA, the lowest values being associated to **pA** flanked by pyrimidines, as for **qA**. The mean fluorescence lifetimes followed a similar trend ranging from 0.4 to 7.4 ns in ssDNA and from 0.7 to 3.8 ns in dsDNA. Similar to **qAN4**, **pA** can be used in combination with **qA_{nitro}** for FRET measurements. The FRET efficiency varied periodically as the donor-acceptor distance increased, indicating that **pA** and **qA_{nitro}** were perfectly stacked inside dsDNA. The periodicity was due to the change of the relative orientation of the transition dipole moments that was confirmed by TDDFT calculations. It can be seen that the nucleoside-based FRET pairs discussed throughout this review (summarized in Table 4) cover broad spectral and transfer efficiency ranges, offering researchers flexibility in choosing fluorescent nucleoside substituents for conformation probing experiments.

Table 4. FRET pairs based on fluorescent nucleoside analogues

Donor	Acceptor	R_0 (nm)[a]	Labeling context	References
^{diox}**T**	**tC**	2	dsDNA	339
qAN1	**qA_{nitro}**	2.2	dsDNA	357
pA	**qA_{nitro}**	2.4	dsDNA	350
qAN4	**qA_{nitro}**	2.5	dsDNA	349
tC^O	**tC_{nitro}**	2.7	dsDNA	315
2CNqA	**qA_{nitro}**	2.8[b]	dsDNA and dsRNA	347
th**dG**	**tC**	3.1	dsDNA	237
2CNqA	**tC_{nitro}**	3.4[b]	dsDNA and dsRNA	347

[a] Förster radius, assuming that $\kappa^2 = 2/3$; the quantum yields of the donors and acceptors can be found in Tables 1 and 3. [b] Numerical value has been deduced from the graphical data presented in the papers.

The **pA** nucleoside can also be excited through two-photon absorption (Figure 40A), as its two-photon (2P) cross-section was 6.6 GM at 780 nm [350]. Upon incorporation in ss- and dsDNA, a relatively high 2P cross-section (2.5−3 GM) was retained [351]. The decrease in 2P cross section on incorporation in DNA was tentatively associated to the impact of the local electrostatic field on the change in dipole moment on excitation. Interestingly, the photostability of **pA** under two-photon excitation at 780 nm was significantly enhanced relative to UV excitation. To reach single molecule sensitivity with **pA**, wide-field [350] and confocal [351] microscopies were performed. In the former case, liposomes labeled with five A**pA**A sequences were immobilized and successfully imaged using an inverted TIRF microscope on which UV excitation was provided by a mercury lamp [350]. Confocal fluorescence microscopy with two-photon excitation at 780 nm and fluorescence correlation spectroscopy (Figure 40B) allowed for measuring the brightness of **pA** incorporated in ssDNA, but not in dsDNA [351]. These experiments showed that single-molecule detection of fluorescent base analogues was within reach.

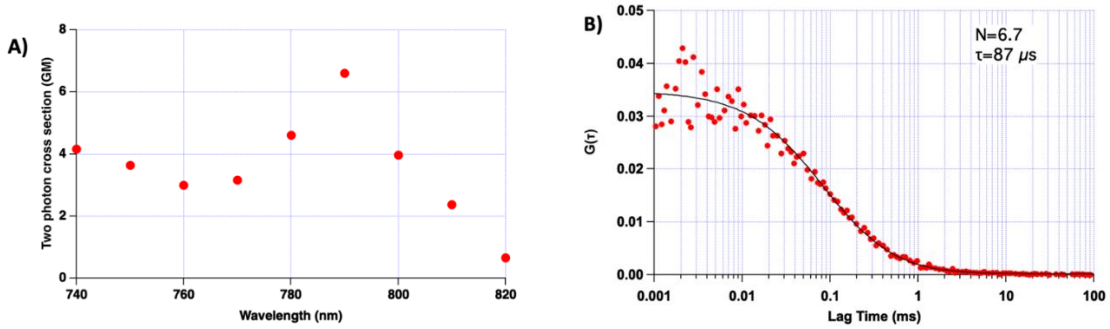

Figure 40. A) Two-photon absorption cross section of **pA** in ethanol. B) Fluorescence correlation spectroscopy for a **pA**-labeled oligonucleotide. The correlation curve is shown in red. The fit (black curve) gave an average of 6.7 molecules in the excitation volume with a diffusion time of 87 µs. The measured brightness was about 500 Hz /molecule. Adapted from ref. [351]

4. Conclusions- and outlooks

Sparked by a few pioneering literature reports on **2AP** and other isomorphic nucleosides decades ago, the field of research on fluorescent nucleoside analogues has grown immensely. Though the isomorphic nucleoside analogue **2AP** has been for a long time the gold standard probe and is still largely used for studying DNA and RNA, a large number of other FNAs have been proposed, having improved spectroscopic properties and environment sensitivity. The systematic search for new fluorescent analogues of the virtually non-fluorescent natural nucleosides has delivered a toolset of structurally diverse isomorphic and expanded FNAs, which have been utilized as probes of conformational polymorphism, dynamics and interactions of nucleic acids mainly in solution at ambient conditions. Selection of a FNA for a given application is necessarily a case-by-case study, which depends on the definition of the most important properties needed for that application and those that can be sacrificed. A first level of selection is based on the need to have or not FNAs structurally as close as possible to the natural nucleosides. If both the size and the Watson Crick base pairing of the natural nucleosides are critical, then isomorphic FNAs described in chapter 2 will be a first choice. Expanded FNAs described in chapter 3 are larger than the natural nucleosides, but many of them preserve Watson-Crick base pairing. Therefore, they generally preserve the secondary structure of the labelled nucleic acids, but can hinder the interactions with proteins and ligands. After having selected the FNA type (isomorphic or expanded), the second level of selection will be based on the spectroscopic properties needed for the application. A quick preliminary selection can be made based on the data in Table 1 for the isomorphic FNAs and Table 3 for the expanded FNAs. Parameters such as the position of the absorption and emission spectra, the brightness (given by the product of ε and ϕ_F), the sensitivity to the environment and sequence context (underlined by the differences in ϕ_F between the free nucleoside and the ssDNA and dsDNA, as well as through the range of ϕ_F values in the ssDNA or dsDNAs) and the fluorescence lifetime values will be used individually or combined to guide the selection.

Although a number of FNAs has been developed, a point of saturation has not been reached. An influx of new ideas and methodological developments are needed to create next generations of FNAs with improved nucleobase-mimicking and photophysical properties to conquer the deficiencies of the existing ones. We describe here several research avenues that could be promising in this regard. While reports on the design and synthesis of new FNAs appear regularly in the literature, very few of them go beyond the most basic steady-state spectroscopic characterization. Including time-resolved

fluorescence measurements in combination with QM calculations into the standard characterization pipelines for new FNAs would be highly beneficial for understanding more deeply their photophysics and photochemistry. This understanding is instrumental for laying the foundations for future applications of FNAs and, in particular, for rational design of appropriately labeled nucleic acid sequences to exhibit optimal spectroscopic changes in the presence of a given protein or other external effectors and ligands. An in depth knowledge on FNA photophysics included in ONs is also critical to properly extract and interpret the full set of information that results from a given experiment. Active involvement of ultrafast time-resolved spectroscopy and theoretical approaches of quantum chemistry will definitively catalyze the development of improved FNAs.

In many FNAs discussed in this review, installation of bulky substituents or modification of the nucleobase core improved the spectroscopic properties, but also affected the *syn-anti* equilibrium of the nucleoside and caused destabilization and perturbations of the conformational landscape of the labeled nucleic acids. Isomorphic and isofunctional nucleoside analogues such as the **thN** and **^{tz}N** nucleoside families (sections 2.2 and 2.3) present the key advantage of being perfect isosteric mimics of the natural purines. We envision that more isosteric and isofunctional FNAs could be synthesized using the atomic mutagenesis and heteroatom shuffling approaches.

The vast majority of FNAs described in the literature has been used for probing DNA. At the same time, natural RNAs possess a much broader structural, conformational, and functional diversity linked to the multitude of roles they play in the cell. More efforts are needed to adapt the most promising FNAs for labeling of RNAs using either the solid-phase phosphoramidite method, in vitro transcription with engineered RNA polymerases, or alternative next-generation enzymatic labeling methods, such as PLOR[358,359]. The main caveats of the classical methods are either the limitation in the size of labeled RNAs or the lack of position selectivity of the labeling, which should be addressed with the next-generation methods to site-selectively incorporate FNAs into long functional biologically relevant RNAs.

The development of the single-molecule FRET technology has resulted into a deeper understanding of conformational heterogeneity of biological molecules in bulk samples. A similar leap could be made with FNAs, if they could be adapted for single-molecule spectroscopy experiments. Their UV excitation and limited photostability are fundamental limitations in this direction. Despite that, proof-of-concept single-molecule detection has been notably reported for **2AP**, **pC**, and **pA**.[350,351,360] Development of FNAs with high brightness as well as high two-photon absorption (2PA) cross-section and increased photostability might pave the way towards routine single-molecule applications.

Last but not least, the vast majority of the reported experiments with FNAs have been performed using reconstituted systems in buffers. Performing experiments with FNAs in the context of living cells is an exceptional challenge, requiring *in cellulo* labelling strategies or delivery methods as well as means to increase the signal-to-noise ratio by suppressing the autofluorescence of the cell.

We believe that the upcoming decade will bring us many exciting reports on spectroscopic probing of nucleic acids with FNAs. An integrated approach, combining advanced spectroscopy, quantum chemistry and molecular dynamics simulations with structural biology information will be the key to reveal new fundamental biological mechanisms involving DNA and RNA.

5. Conflicts of interest

There are no conflicts of interest to declare.

6. Acknowledgements

This work was supported by the Agence Nationale de la Recherche (ANR blanc SMFLUONA), the Labex NIE, the Centre National pour la Recherche Scientifique (CNRS) and the European Research Council through the Advanced Grant 2014 (hybridFRET, number 671208) to CAMS. Y.M. is grateful to the Institut Universitaire de France (IUF) for support and providing additional time to be dedicated to research. D.D. was supported by the CNRS and a post-doctoral fellowship from the Fonds Régional de coopération pour la recherche of the Région Grand Est (EpiARN project). S.C. was supported by a PhD fellowship from Région Grand Est and A.B. by a PostDoc position within the ERC project hybridFRET. We thank the anonymous reviewers for their helpful comments and constructive feedback. We apologize to authors whose research articles could not be cited due to space constraints.

7. References

1 F. Spitz and E. E. M. Furlong, *Nat Rev Genet*, 2012, **13**, 613–626.
2 T. Glisovic, J. L. Bachorik, J. Yong and G. Dreyfuss, *FEBS Letters*, 2008, **582**, 1977–1986.
3 M. W. Hentze, A. Castello, T. Schwarzl and T. Preiss, *Nat Rev Mol Cell Biol*, 2018, **19**, 327–341.
4 E. Lerner, T. Cordes, A. Ingargiola, Y. Alhadid, S. Chung, X. Michalet and S. Weiss, *Science*, 2018, **359**, eaan1133.
5 M. Dimura, T.-O. Peulen, H. Sanabria, D. Rodnin, K. Hemmen, C. A. Hanke, C. A. M. Seidel and H. Gohlke, *Nat Commun*, 2020, **11**, 5394.
6 H. Sanabria, D. Rodnin, K. Hemmen, T.-O. Peulen, S. Felekyan, M. R. Fleissner, M. Dimura, F. Koberling, R. Kühnemuth, W. Hubbell, H. Gohlke and C. A. M. Seidel, *Nat Commun*, 2020, **11**, 1231.
7 L. Bacic, A. Sabantsev and S. Deindl, *Current Opinion in Structural Biology*, 2020, **65**, 61–68.
8 M. I. J. Stich, L. H. Fischer and O. S. Wolfbeis, *Chem. Soc. Rev.*, 2010, **39**, 3102.
9 A. P. DEMCHENKO, *INTRODUCTION TO FLUORESCENCE SENSING: materials and devices.*, SPRINGER, S.l., 2021.
10 J. Widengren, V. Kudryavtsev, M. Antonik, S. Berger, M. Gerken and C. A. M. Seidel, *Anal. Chem.*, 2006, **78**, 2039–2050.
11 C. E. Crespo-Hernández, B. Cohen, P. M. Hare and B. Kohler, *Chem. Rev.*, 2004, **104**, 1977–2020.
12 L. J. Kricka and P. Fortina, *Clinical Chemistry*, 2009, **55**, 670–683.
13 N. Klöcker, F. P. Weissenboeck and A. Rentmeister, *Chem. Soc. Rev.*, 2020, 10.1039.D0CS00600A.
14 B. Y. Michel, D. Dziuba, R. Benhida, A. P. Demchenko and A. Burger, *Front. Chem.*, 2020, **8**, 112.
15 K. Wang, D. Wang and B. Wang, *Current Organic Chemistry*, 2015, **19**, 1011–1020.
16 S. H. Weisbrod and A. Marx, *Chem. Commun.*, 2008, 5675.
17 M. Hocek, *J. Org. Chem.*, 2014, **79**, 9914–9921.
18 M. Hocek, *Acc. Chem. Res.*, 2019, **52**, 1730–1737.
19 C. Holzhauser and H.-A. Wagenknecht, *J. Org. Chem.*, 2013, **78**, 7373–7379.
20 Y. N. Teo and E. T. Kool, *Chem. Rev.*, 2012, **112**, 4221–4245.
21 C. Wojczewski, K. Stolze and J. W. Engels, *Synlett*, 1999, **1999**, 1667–1678.

22 M. Merkel, K. Peewasan, S. Arndt, D. Ploschik and H.-A. Wagenknecht, *ChemBioChem*, 2015, **16**, 1541–1553.
23 F. D. Steffen, R. Börner, E. Freisinger and R. K. O. Sigel, *chimia (aarau)*, 2019, **73**, 257–261.
24 N. Dai and E. T. Kool, *Chem. Soc. Rev.*, 2011, **40**, 5756.
25 D. L. Wilson and E. T. Kool, *ACS Chem. Biol.*, 2018, **13**, 1721–1733.
26 G. Hwang, *Molecules*, 2018, **23**, 124.
27 F. Hövelmann and O. Seitz, *Acc. Chem. Res.*, 2016, **49**, 714–723.
28 H. Ji and K. Astakhova, in *RNA and DNA Diagnostics*, eds. V. A. Erdmann, S. Jurga and J. Barciszewski, Springer International Publishing, Cham, 2015, pp. 161–181.
29 H. Hwang and S. Myong, *Chem. Soc. Rev.*, 2014, **43**, 1221–1229.
30 J. Huang, X. Yang, X. He, K. Wang, J. Liu, H. Shi, Q. Wang, Q. Guo and D. He, *TrAC Trends in Analytical Chemistry*, 2014, **53**, 11–20.
31 H. Zhang, F. Li, B. Dever, X.-F. Li and X. C. Le, *Chem. Rev.*, 2013, **113**, 2812–2841.
32 C.-H. Leung, H.-J. Zhong, H.-Z. He, L. Lu, D. S.-H. Chan and D.-L. Ma, *Chem. Sci.*, 2013, **4**, 3781.
33 A. Boutorine, D. Novopashina, O. Krasheninina, K. Nozeret and A. Venyaminova, *Molecules*, 2013, **18**, 15357–15397.
34 S. Preus and L. M. Wilhelmsson, *ChemBioChem*, 2012, **13**, 1990–2001.
35 C.-H. Leung, D. S.-H. Chan, H.-Z. He, Z. Cheng, H. Yang and D.-L. Ma, *Nucleic Acids Research*, 2012, **40**, 941–955.
36 T. Ihara and Y. Kitamura, *Journal of Photochemistry and Photobiology C: Photochemistry Reviews*, 2012, **13**, 148–167.
37 R. T. Ranasinghe and T. Brown, *Chem. Commun.*, 2011, **47**, 3717.
38 M. E. Østergaard and P. J. Hrdlicka, *Chem. Soc. Rev.*, 2011, **40**, 5771.
39 B. Juskowiak, *Anal Bioanal Chem*, 2011, **399**, 3157–3176.
40 J. M. Pagano, C. C. Clingman and S. P. Ryder, *RNA*, 2011, **17**, 14–20.
41 K. Wang, Z. Tang, C. J. Yang, Y. Kim, X. Fang, W. Li, Y. Wu, C. D. Medley, Z. Cao, J. Li, P. Colon, H. Lin and W. Tan, *Angew. Chem. Int. Ed.*, 2009, **48**, 856–870.
42 N. Venkatesan, Y. Jun Seo and B. Hyean Kim, *Chem. Soc. Rev.*, 2008, **37**, 648.
43 A. V. Orden and J. Jung, *Biopolymers*, 2008, **89**, 1–16.
44 S. A. E. Marras, S. Tyagi and F. R. Kramer, *Clinica Chimica Acta*, 2006, **363**, 48–60.
45 A. P. Silverman and E. T. Kool, *Trends in Biotechnology*, 2005, **23**, 225–230.
46 A. Okamoto, Y. Saito and I. Saito, *Journal of Photochemistry and Photobiology C: Photochemistry Reviews*, 2005, **6**, 108–122.
47 J. Zheng, R. Yang, M. Shi, C. Wu, X. Fang, Y. Li, J. Li and W. Tan, *Chem. Soc. Rev.*, 2015, **44**, 3036–3055.
48 B. Faltin, R. Zengerle and F. von Stetten, *Clinical Chemistry*, 2013, **59**, 1567–1582.
49 D. O. Wang and A. Okamoto, *Journal of Photochemistry and Photobiology C: Photochemistry Reviews*, 2012, **13**, 112–123.
50 A. Okamoto, *Chem. Soc. Rev.*, 2011, **40**, 5815.
51 J. Guo, J. Ju and N. J. Turro, *Anal Bioanal Chem*, 2012, **402**, 3115–3125.
52 G. Bao, W. J. Rhee and A. Tsourkas, *Annu. Rev. Biomed. Eng.*, 2009, **11**, 25–47.
53 M. J. Davies, A. Shah and I. J. Bruce, *Chem. Soc. Rev.*, 2000, **29**, 97–107.
54 S. Verma and F. Eckstein, *Annu. Rev. Biochem.*, 1998, **67**, 99–134.
55 N. M. Bell and J. Micklefield, *ChemBioChem*, 2009, **10**, 2691–2703.
56 V. J. LiCata and A. J. Wowor, in *Methods in Cell Biology*, Elsevier, 2008, vol. 84, pp. 243–262.
57 H. Beltz, C. Clauss, E. Piémont, D. Ficheux, R. J. Gorelick, B. Roques, C. Gabus, J.-L. Darlix, H. de Rocquigny and Y. Mély, *Journal of Molecular Biology*, 2005, **348**, 1113–1126.
58 A. A. Deniz, M. Dahan, J. R. Grunwell, T. Ha, A. E. Faulhaber, D. S. Chemla, S. Weiss and P. G. Schultz, *Proceedings of the National Academy of Sciences*, 1999, **96**, 3670–3675.

59 A. Iqbal, S. Arslan, B. Okumus, T. J. Wilson, G. Giraud, D. G. Norman, T. Ha and D. M. J. Lilley, *Proceedings of the National Academy of Sciences*, 2008, **105**, 11176–11181.
60 R. M. Clegg, A. I. Murchie, A. Zechel and D. M. Lilley, *Proceedings of the National Academy of Sciences*, 1993, **90**, 2994–2998.
61 A. K. Wozniak, G. F. Schroder, H. Grubmuller, C. A. M. Seidel and F. Oesterhelt, *Proceedings of the National Academy of Sciences*, 2008, **105**, 18337–18342.
62 B. Hellenkamp, S. Schmid, O. Doroshenko, O. Opanasyuk, R. Kühnemuth, S. Rezaei Adariani, B. Ambrose, M. Aznauryan, A. Barth, V. Birkedal, M. E. Bowen, H. Chen, T. Cordes, T. Eilert, C. Fijen, C. Gebhardt, M. Götz, G. Gouridis, E. Gratton, T. Ha, P. Hao, C. A. Hanke, A. Hartmann, J. Hendrix, L. L. Hildebrandt, V. Hirschfeld, J. Hohlbein, B. Hua, C. G. Hübner, E. Kallis, A. N. Kapanidis, J.-Y. Kim, G. Krainer, D. C. Lamb, N. K. Lee, E. A. Lemke, B. Levesque, M. Levitus, J. J. McCann, N. Naredi-Rainer, D. Nettels, T. Ngo, R. Qiu, N. C. Robb, C. Röcker, H. Sanabria, M. Schlierf, T. Schröder, B. Schuler, H. Seidel, L. Streit, J. Thurn, P. Tinnefeld, S. Tyagi, N. Vandenberk, A. M. Vera, K. R. Weninger, B. Wünsch, I. S. Yanez-Orozco, J. Michaelis, C. A. M. Seidel, T. D. Craggs and T. Hugel, *Nat Methods*, 2018, **15**, 669–676.
63 S. Kilic, S. Felekyan, O. Doroshenko, I. Boichenko, M. Dimura, H. Vardanyan, L. C. Bryan, G. Arya, C. A. M. Seidel and B. Fierz, *Nat Commun*, 2018, **9**, 235.
64 A. Gansen, S. Felekyan, R. Kühnemuth, K. Lehmann, K. Tóth, C. A. M. Seidel and J. Langowski, *Nat Commun*, 2018, **9**, 4628.
65 T. D. Craggs, M. Sustarsic, A. Plochowietz, M. Mosayebi, H. Kaju, A. Cuthbert, J. Hohlbein, L. Domicevica, P. C. Biggin, J. P. K. Doye and A. N. Kapanidis, *Nucleic Acids Research*, 2019, **47**, 10788–10800.
66 S. Kosuri and G. M. Church, *Nat Methods*, 2014, **11**, 499–507.
67 M. Flamme, L. K. McKenzie, I. Sarac and M. Hollenstein, *Methods*, 2019, **161**, 64–82.
68 Y. Saito and R. H. E. Hudson, *Journal of Photochemistry and Photobiology C: Photochemistry Reviews*, 2018, **36**, 48–73.
69 M. Rist and J. Marino, *COC*, 2002, **6**, 775–793.
70 S. G. Srivatsan and A. A. Sawant, *Pure and Applied Chemistry*, 2010, **83**, 213–232.
71 A. A. Tanpure, M. G. Pawar and S. G. Srivatsan, *Isr. J. Chem.*, 2013, **53**, 366–378.
72 S. Manna and S. G. Srivatsan, *RSC Adv.*, 2018, **8**, 25673–25694.
73 W. Xu, K. M. Chan and E. T. Kool, *Nature Chem*, 2017, **9**, 1043–1055.
74 J. N. Wilson and E. T. Kool, *Org. Biomol. Chem.*, 2006, **4**, 4265.
75 R. W. Sinkeldam, N. J. Greco and Y. Tor, *Chem. Rev.*, 2010, **110**, 2579–2619.
76 A. Matarazzo and R. H. E. Hudson, *Tetrahedron*, 2015, **71**, 1627–1657.
77 D. W. Dodd and R. H. E. Hudson, *Mini-Reviews in Organic Chemistry*, 2009, **6**, 378–391.
78 L. M. Wilhelmsson, *Quart. Rev. Biophys.*, 2010, **43**, 159–183.
79 W. J. Schreier, P. Gilch and W. Zinth, *Annu. Rev. Phys. Chem.*, 2015, **66**, 497–519.
80 C. T. Middleton, K. de La Harpe, C. Su, Y. K. Law, C. E. Crespo-Hernández and B. Kohler, *Annu. Rev. Phys. Chem.*, 2009, **60**, 217–239.
81 K. Kleinermanns, D. Nachtigallová and M. S. de Vries, *International Reviews in Physical Chemistry*, 2013, **32**, 308–342.
82 J. Peon and A. H. Zewail, *Chemical Physics Letters*, 2001, **348**, 255–262.
83 M. Daniels and W. Hauswirth, *Science*, 1971, **171**, 675–677.
84 J.-M. L. Pecourt, J. Peon and B. Kohler, *J. Am. Chem. Soc.*, 2001, **123**, 10370–10378.
85 J.-M. L. Pecourt, J. Peon and B. Kohler, *J. Am. Chem. Soc.*, 2000, **122**, 9348–9349.
86 N. Ismail, L. Blancafort, M. Olivucci, B. Kohler and M. A. Robb, *J. Am. Chem. Soc.*, 2002, **124**, 6818–6819.
87 S. Perun, A. L. Sobolewski and W. Domcke, *J. Phys. Chem. A*, 2006, **110**, 13238–13244.
88 S. Perun, A. L. Sobolewski and W. Domcke, *J. Am. Chem. Soc.*, 2005, **127**, 6257–6265.

89 H. R. Hudock, B. G. Levine, A. L. Thompson, H. Satzger, D. Townsend, N. Gador, S. Ullrich, A. Stolow and T. J. Martínez, *J. Phys. Chem. A*, 2007, **111**, 8500–8508.
90 H. R. Hudock and T. J. Martínez, *Chemphyschem*, 2008, **9**, 2486–2490.
91 B. K. McFarland, J. P. Farrell, S. Miyabe, F. Tarantelli, A. Aguilar, N. Berrah, C. Bostedt, J. D. Bozek, P. H. Bucksbaum, J. C. Castagna, R. N. Coffee, J. P. Cryan, L. Fang, R. Feifel, K. J. Gaffney, J. M. Glownia, T. J. Martinez, M. Mucke, B. Murphy, A. Natan, T. Osipov, V. S. Petrović, S. Schorb, Th. Schultz, L. S. Spector, M. Swiggers, I. Tenney, S. Wang, J. L. White, W. White and M. Gühr, *Nat Commun*, 2014, **5**, 4235.
92 A. Banyasz, T. Douki, R. Improta, T. Gustavsson, D. Onidas, I. Vayá, M. Perron and D. Markovitsi, *J. Am. Chem. Soc.*, 2012, **134**, 14834–14845.
93 P. D. Wood and R. W. Redmond, *J. Am. Chem. Soc.*, 1996, **118**, 4256–4263.
94 R. González-Luque, T. Climent, I. González-Ramírez, M. Merchán and L. Serrano-Andrés, *J. Chem. Theory Comput.*, 2010, **6**, 2103–2114.
95 D. B. Bucher, B. M. Pilles, T. Carell and W. Zinth, *Proceedings of the National Academy of Sciences*, 2014, **111**, 4369–4374.
96 R. Improta, *J. Phys. Chem. B*, 2012, **116**, 14261–14274.
97 A. L. Sobolewski and W. Domcke, *Chemical Physics*, 2003, **294**, 73–83.
98 T. Schultz, E. Samoylova, W. Radloff, I. V. Hertel, A. L. Sobolewski and W. Domcke, *Science*, 2004, **306**, 1765–1768.
99 S. Perun, A. L. Sobolewski and W. Domcke, *J. Phys. Chem. A*, 2006, **110**, 9031–9038.
100 U. L. RajBhandary, S. H. Chang, A. Stuart, R. D. Faulkner, R. M. Hoskinson and H. G. Khorana, *Proceedings of the National Academy of Sciences*, 1967, **57**, 751–758.
101 A. Maelicke, F. Von Der Haar, M. Sprinzl and F. Cramer, *Biopolymers*, 1975, **14**, 155–171.
102 H. Kasai, M. Goto, K. Ikeda, M. Zama, Y. Mizuno, S. Takemura, S. Matsuura, T. Sugimoto and T. Goto, *Biochemistry*, 1976, **15**, 898–904.
103 D. C. Ward, E. Reich and L. Stryer, *J. Biol. Chem.*, 1969, **244**, 1228–1237.
104 A. C. Jones and R. K. Neely, *Quart. Rev. Biophys.*, 2015, **48**, 244–279.
105 T. M. Nordlund, S. Andersson, L. Nilsson, R. Rigler, A. Graeslund and L. W. McLaughlin, *Biochemistry*, 1989, **28**, 9095–9103.
106 S. M. Law, R. Eritja, M. F. Goodman and K. J. Breslauer, *Biochemistry*, 1996, **35**, 12329–12337.
107 O. O. Brovarets', H. Pérez-Sánchez and D. M. Hovorun, *RSC Adv.*, 2016, **6**, 99546–99557.
108 L. C. Sowers, G. V. Fazakerley, R. Eritja, B. E. Kaplan and M. F. Goodman, *Proceedings of the National Academy of Sciences*, 1986, **83**, 5434–5438.
109 P. A. Fagan, C. Fàbrega, R. Eritja, M. F. Goodman and D. E. Wemmer, *Biochemistry*, 1996, **35**, 4026–4033.
110 L. C. Sowers, Y. Boulard and G. V. Fazakerley, *Biochemistry*, 2000, **39**, 7613–7620.
111 G. V. Fazakerley, L. C. Sowers, R. Eritja, B. E. Kaplan and M. F. Goodman, *Biochemistry*, 1987, **26**, 5641–5646.
112 R. K. Neely, *Nucleic Acids Research*, 2005, **33**, 6953–6960.
113 R. Eritja, B. E. Kaplan, D. Mhaskar, L. C. Sowers, J. Petruska and M. F. Goodman, *Nucl Acids Res*, 1986, **14**, 5869–5884.
114 D. Xu, K. O. Evans and T. M. Nordlund, *Biochemistry*, 1994, **33**, 9592–9599.
115 S. P. Hancock, T. Ghane, D. Cascio, R. Rohs, R. Di Felice and R. C. Johnson, *Nucleic Acids Research*, 2013, **41**, 6750–6760.
116 A. Dallmann, L. Dehmel, T. Peters, C. Mügge, C. Griesinger, J. Tuma and N. P. Ernsting, *Angewandte Chemie International Edition*, 2010, **49**, 5989–5992.
117 M. Sholokh, R. Sharma, D. Shin, R. Das, O. A. Zaporozhets, Y. Tor and Y. Mély, *J. Am. Chem. Soc.*, 2015, **137**, 3185–3188.
118 K. Voltz, J. Léonard, P. T. Touceda, J. Conyard, Z. Chaker, A. Dejaegere, J. Godet, Y. Mély, S. Haacke and R. H. Stote, *Nucleic Acids Res*, 2016, **44**, 3408–3419.

119 P.-O. Lycksell, A. Gräslund, F. Claesens, L. W. McLaughlin, U. Larsson and R. Rigler, *Nucl Acids Res*, 1987, **15**, 9011–9025.
120 K. Evans, D. Xu, Y. Kim and T. M. Nordlund, *J Fluoresc*, 1992, **2**, 209–216.
121 A. Holmén, B. Nordén and B. Albinsson, *J. Am. Chem. Soc.*, 1997, **119**, 3114–3121.
122 R. K. Neely, S. W. Magennis, D. T. F. Dryden and A. C. Jones, *J. Phys. Chem. B*, 2004, **108**, 17606–17610.
123 S. Lobsiger, S. Blaser, R. K. Sinha, H.-M. Frey and S. Leutwyler, *Nature Chem*, 2014, **6**, 989–993.
124 E. L. Rachofsky, R. Osman and J. B. A. Ross, *Biochemistry*, 2001, **40**, 946–956.
125 S. Bharill, P. Sarkar, J. D. Ballin, I. Gryczynski, G. M. Wilson and Z. Gryczynski, *Analytical Biochemistry*, 2008, **377**, 141–149.
126 C. E. Crespo-Hernández, L. Martínez-Fernández, C. Rauer, C. Reichardt, S. Mai, M. Pollum, P. Marquetand, L. González and I. Corral, *J. Am. Chem. Soc.*, 2015, **137**, 4368–4381.
127 B. Cohen, P. M. Hare and B. Kohler, *J. Am. Chem. Soc.*, 2003, **125**, 13594–13601.
128 K. B. Hall, in *Methods in Enzymology*, Elsevier, 2009, vol. 469, pp. 269–285.
129 W. A. Wassam and E. C. Lim, *The Journal of Chemical Physics*, 1978, **68**, 433–454.
130 W. A. Wassam and E. C. Lim, *Journal of Molecular Structure*, 1978, **47**, 129–198.
131 E. C. Lim, *J. Phys. Chem.*, 1986, **90**, 6770–6777.
132 E. L. Rachofsky, J. B. A. Ross, M. Krauss and R. Osman, *J. Phys. Chem. A*, 2001, **105**, 190–197.
133 L. Serrano-Andres, M. Merchan and A. C. Borin, *Proceedings of the National Academy of Sciences*, 2006, **103**, 8691–8696.
134 L. Serrano-Andrés, M. Merchán and A. C. Borin, *Chem. Eur. J.*, 2006, **12**, 6559–6571.
135 S. Perun, A. L. Sobolewski and W. Domcke, *Molecular Physics*, 2006, **104**, 1113–1121.
136 V. Ludwig, M. S. do Amaral, Z. M. da Costa, A. C. Borin, S. Canuto and L. Serrano-Andrés, *Chemical Physics Letters*, 2008, **463**, 201–205.
137 K. Feng, G. Engler, K. Seefeld and K. Kleinermanns, *ChemPhysChem*, 2009, **10**, 886–889.
138 S. Lobsiger, R. K. Sinha, M. Trachsel and S. Leutwyler, *The Journal of Chemical Physics*, 2011, **134**, 114307.
139 M. Barbatti and H. Lischka, *Phys. Chem. Chem. Phys.*, 2015, **17**, 15452–15459.
140 C. Reichardt, C. Wen, R. A. Vogt and C. E. Crespo-Hernández, *Photochem. Photobiol. Sci.*, 2013, **12**, 1341.
141 H. Böhnke, K. Röttger, R. A. Ingle, H. J. B. Marroux, M. Bohnsack, A. J. Orr-Ewing and F. Temps, *Phys. Chem. Chem. Phys.*, 2018, **20**, 20033–20042.
142 L. W. McLaughlin, T. Leong, F. Benseler and N. Piel, *Nucl Acids Res*, 1988, **16**, 5631–5644.
143 I. Zagórowska and R. W. Adamiak, *Biochimie*, 1996, **78**, 123–130.
144 E. L. Rachofsky, E. Seibert, J. T. Stivers, R. Osman and J. B. A. Ross, *Biochemistry*, 2001, **40**, 957–967.
145 H. Böhnke, K. Röttger, R. A. Ingle, H. J. B. Marroux, M. Bohnsack, N. K. Schwalb, A. J. Orr-Ewing and F. Temps, *J. Phys. Chem. B*, 2019, **123**, 2904–2914.
146 M. A. O'Neill and J. K. Barton, *J. Am. Chem. Soc.*, 2004, **126**, 13234–13235.
147 M. A. O'Neil and J. K. Barton, *J. Am. Chem. Soc.*, 2002, **124**, 13053–13066.
148 J. M. Jean and K. B. Hall, *Biochemistry*, 2002, **41**, 13152–13161.
149 C. A. M. Seidel, A. Schulz and M. H. M. Sauer, *J. Phys. Chem.*, 1996, **100**, 5541–5553.
150 M. A. O'Neill, H.-C. Becker, C. Wan, J. K. Barton and A. H. Zewail, *Angew. Chem. Int. Ed.*, 2003, **42**, 5896–5900.
151 J. M. Jean and K. B. Hall, *Proceedings of the National Academy of Sciences*, 2001, **98**, 37–41.
152 J. Liang and S. Matsika, *J. Am. Chem. Soc.*, 2011, **133**, 6799–6808.
153 J. Liang, Q. L. Nguyen and S. Matsika, *Photochem. Photobiol. Sci.*, 2013, **12**, 1387.
154 C. R. Guest, R. A. Hochstrasser, L. C. Sowers and D. P. Millar, *Biochemistry*, 1991, **30**, 3271–3279.

155 T. Ramreddy, Rao and G. Krishnamoorthy, *J. Phys. Chem. B*, 2007, **111**, 5757–5766.
156 J. M. Remington, A. M. Philip, M. Hariharan and B. Kohler, *The Journal of Chemical Physics*, 2016, **145**, 155101.
157 J. Godet, N. Ramalanjaona, K. K. Sharma, L. Richert, H. de Rocquigny, J.-L. Darlix, G. Duportail and Y. Mély, *Nucleic Acids Research*, 2011, **39**, 6633–6645.
158 J. Godet, C. Kenfack, F. Przybilla, L. Richert, G. Duportail and Y. Mély, *Nucleic Acids Research*, 2013, **41**, 5036–5048.
159 J. M. Remington, M. McCullagh and B. Kohler, *J. Phys. Chem. B*, 2019, **123**, 2291–2304.
160 D. A. Smith, L. F. Holroyd, T. van Mourik and A. C. Jones, *Phys. Chem. Chem. Phys.*, 2016, **18**, 14691–14700.
161 T. van Mourik and S. W. L. Hogan, *Struct Chem*, 2016, **27**, 145–158.
162 M. W. Frey, L. C. Sowers, D. P. Millar and S. J. Benkovic, *Biochemistry*, 1995, **34**, 9185–9192.
163 V. Purohit, N. D. F. Grindley and C. M. Joyce, *Biochemistry*, 2003, **42**, 10200–10211.
164 J. M. Beechem, M. R. Otto, L. B. Bloom, R. Eritja, L. J. Reha-Krantz and M. F. Goodman, *Biochemistry*, 1998, **37**, 10144–10155.
165 C. A. Dunlap and M.-D. Tsai, *Biochemistry*, 2002, **41**, 11226–11235.
166 A. M. Fedoriw, H. Liu, V. E. Anderson and P. L. deHaseth, *Biochemistry*, 1998, **37**, 11971–11979.
167 R. P. Bandwar and S. S. Patel, *J. Biol. Chem.*, 2001, **276**, 14075–14082.
168 P. G. Mitsis and J. G. Kwagh, *Nucleic Acids Research*, 1999, **27**, 3057–3063.
169 K. D. Raney, L. C. Sowers, D. P. Millar and S. J. Benkovic, *Proceedings of the National Academy of Sciences*, 1994, **91**, 6644–6648.
170 S. Gong, H. H. Yu, K. A. Johnson and D. W. Taylor, *Cell Reports*, 2018, **22**, 359–371.
171 S. Zhang, Q. Zhang, X. Hou, L. Guo, F. Wang, L. Bi, X. Zhang, H. Li, F. Wen, X. Xi, X. Huang, B. Shen and B. Sun, *EMBO Rep*, , DOI:10.15252/embr.202050184.
172 Q. Xu, Y. Zhang and C. Zhang, *Chem. Commun.*, 2015, **51**, 9121–9124.
173 S. Kirk, *Bioorganic & Medicinal Chemistry*, 2001, **9**, 2295–2301.
174 L. R. Ganser, M. L. Kelly, D. Herschlag and H. M. Al-Hashimi, *Nat Rev Mol Cell Biol*, 2019, **20**, 474–489.
175 C. K. Kwok, Y. Tang, S. M. Assmann and P. C. Bevilacqua, *Trends in Biochemical Sciences*, 2015, **40**, 221–232.
176 R. Schroeder, R. Grossberger, A. Pichler and C. Waldsich, *Current Opinion in Structural Biology*, 2002, **12**, 296–300.
177 Z. Lu and H. Y. Chang, *Current Opinion in Structural Biology*, 2016, **36**, 142–148.
178 S. A. Mortimer, M. A. Kidwell and J. A. Doudna, *Nat Rev Genet*, 2014, **15**, 469–479.
179 M. Menger, T. Tuschl, F. Eckstein and D. Porschke, *Biochemistry*, 1996, **35**, 14710–14716.
180 J. Ma, S. Haldar, M. A. Khan, S. D. Sharma, W. C. Merrick, E. C. Theil and D. J. Goss, *Proceedings of the National Academy of Sciences*, 2012, **109**, 8417–8422.
181 M. F. Soulière, A. Haller, R. Rieder and R. Micura, *J. Am. Chem. Soc.*, 2011, **133**, 16161–16167.
182 J. D. Ballin, S. Bharill, E. J. Fialcowitz-White, I. Gryczynski, Z. Gryczynski and G. M. Wilson, *Biochemistry*, 2007, **46**, 13948–13960.
183 J. Roca, N. Hori, S. Baral, Y. Velmurugu, R. Narayanan, P. Narayanan, D. Thirumalai and A. Ansari, *Proc Natl Acad Sci USA*, 2018, **115**, E7313–E7322.
184 R. D. Gray, L. Petraccone, J. O. Trent and J. B. Chaires, *Biochemistry*, 2010, **49**, 179–194.
185 Y. Xu, *Nucleic Acids Research*, 2006, **34**, 949–954.
186 T. Kimura, K. Kawai, M. Fujitsuka and T. Majima, *Tetrahedron*, 2007, **63**, 3585–3590.
187 R. C. Spitale, R. A. Flynn, E. A. Torre, E. T. Kool and H. Y. Chang, *WIREs RNA*, 2014, **5**, 867–881.
188 K. M. Weeks and D. M. Mauger, *Acc. Chem. Res.*, 2011, **44**, 1280–1291.

189 S. D. Gilbert, R. P. Rambo, D. Van Tyne and R. T. Batey, *Nat Struct Mol Biol*, 2008, **15**, 177–182.
190 C. M. Connelly, M. H. Moon and J. S. Schneekloth, *Cell Chemical Biology*, 2016, **23**, 1077–1090.
191 N. F. Rizvi and G. F. Smith, *Bioorganic & Medicinal Chemistry Letters*, 2017, **27**, 5083–5088.
192 M. D. Disney, I. Yildirim and J. L. Childs-Disney, *Org. Biomol. Chem.*, 2014, **12**, 1029–1039.
193 S. Shandrick, Q. Zhao, Q. Han, B. K. Ayida, M. Takahashi, G. C. Winters, K. B. Simonsen, D. Vourloumis and T. Hermann, *Angew. Chem. Int. Ed.*, 2004, **43**, 3177–3182.
194 K. A. Lacourciere, J. T. Stivers and J. P. Marino, *Biochemistry*, 2000, **39**, 5630–5641.
195 J. Parsons, M. P. Castaldi, S. Dutta, S. M. Dibrov, D. L. Wyles and T. Hermann, *Nat Chem Biol*, 2009, **5**, 823–825.
196 B. W. Allan and N. O. Reich, *Biochemistry*, 1996, **35**, 14757–14762.
197 B. Holz, *Nucleic Acids Research*, 1998, **26**, 1076–1083.
198 T. Lenz, E. Y. M. Bonnist, G. Pljevaljčić, R. K. Neely, D. T. F. Dryden, A. J. Scheidig, A. C. Jones and E. Weinhold, *J. Am. Chem. Soc.*, 2007, **129**, 6240–6248.
199 Y. Tor, S. Del Valle, D. Jaramillo, S. G. Srivatsan, A. Rios and H. Weizman, *Tetrahedron*, 2007, **63**, 3608–3614.
200 D. Shin, R. W. Sinkeldam and Y. Tor, *J. Am. Chem. Soc.*, 2011, **133**, 14912–14915.
201 A. R. Rovira, A. Fin and Y. Tor, *J. Am. Chem. Soc.*, 2015, **137**, 14602–14605.
202 S. Park, H. Otomo, L. Zheng and H. Sugiyama, *Chem. Commun.*, 2014, **50**, 1573.
203 T. N. V. Karsili, M. Thodika, L. Nguyen and S. Matsika, *Chemical Physics*, 2018, **515**, 434–440.
204 L. Martinez-Fernandez, K. Gavvala, R. Sharma, P. Didier, L. Richert, J. Segarra Martì, M. Mori, Y. Mely and R. Improta, *Chem. Eur. J.*, 2019, **25**, 7375–7386.
205 J. Kuchlyan, L. Martinez-Fernandez, M. Mori, K. Gavvala, S. Ciaco, C. Boudier, L. Richert, P. Didier, Y. Tor, R. Improta and Y. Mély, *J. Am. Chem. Soc.*, 2020, jacs.0c06165.
206 J. Feldmann, Y. Li and Y. Tor, *Chem. Eur. J.*, 2019, **25**, 4379–4389.
207 A. R. Rovira, A. Fin and Y. Tor, *Chem. Sci.*, 2017, **8**, 2983–2993.
208 N. Ben Gaied, N. Glasser, N. Ramalanjaona, H. Beltz, P. Wolff, R. Marquet, A. Burger and Y. Mély, *Nucleic Acids Res*, 2005, **33**, 1031–1039.
209 C. A. Kenfack, E. Piémont, N. Ben Gaied, A. Burger and Y. Mély, *J. Phys. Chem. B*, 2008, **112**, 9736–9745.
210 A. Nadler, J. Strohmeier and U. Diederichsen, *Angew. Chem. Int. Ed.*, 2011, **50**, 5392–5396.
211 L. Martinez-Fernandez, T. Gustavsson, U. Diederichsen and R. Improta, *Molecules*, 2020, **25**, 824.
212 M. E. Hawkins, W. Pfieiderer, A. Mazumder, Y. G. Pommier and F. M. Balis, *Nucl Acids Res*, 1995, **23**, 2872–2880.
213 M. E. Hawkins, *CBB*, 2001, **34**, 257–281.
214 M. E. Hawkins, W. Pfleiderer, F. M. Balis, D. Porter and J. R. Knutson, *Analytical Biochemistry*, 1997, **244**, 86–95.
215 M. E. Hawkins, W. Pfleiderer, O. Jungmann and F. M. Balis, *Analytical Biochemistry*, 2001, **298**, 231–240.
216 S. L. Driscoll, M. E. Hawkins, F. M. Balis, W. Pfleiderer and W. R. Laws, *Biophysical Journal*, 1997, **73**, 3277–3286.
217 K. Wojtuszewski Poulin, A. V. Smirnov, M. E. Hawkins, F. M. Balis and J. R. Knutson, *Biochemistry*, 2009, **48**, 8861–8868.
218 P. Wu, T. M. Nordlund, B. Gildea and L. W. McLaughlin, *Biochemistry*, 1990, **29**, 6508–6514.
219 B. A. Connolly and P. C. Newman, *Nucl Acids Res*, 1989, **17**, 4957–4974.
220 R. W. Sinkeldam, N. J. Greco and Y. Tor, *ChemBioChem*, 2008, **9**, 706–709.
221 N. J. Greco and Y. Tor, *J. Am. Chem. Soc.*, 2005, **127**, 10784–10785.
222 N. J. Greco and Y. Tor, *Tetrahedron*, 2007, **63**, 3515–3527.

223 A. Nuthanakanti, I. Ahmed, S. Y. Khatik, K. Saikrishnan and S. G. Srivatsan, *Nucleic Acids Research*, 2019, **47**, 6059–6072.
224 R. W. Sinkeldam, P. A. Hopkins and Y. Tor, *ChemPhysChem*, 2012, **13**, 3350–3356.
225 M. Sholokh, R. Improta, M. Mori, R. Sharma, C. Kenfack, D. Shin, K. Voltz, R. H. Stote, O. A. Zaporozhets, M. Botta, Y. Tor and Y. Mély, *Angew. Chem. Int. Ed.*, 2016, **55**, 7974–7978.
226 P. Didier, J. Kuchlyan, L. Martinez-Fernandez, P. Gosset, J. Léonard, Y. Tor, R. Improta and Y. Mély, *Phys. Chem. Chem. Phys.*, 2020, **22**, 7381–7391.
227 Y. Araie, S. Ohtsuki, S. Park, M. Nagaoka, K. Umemura, H. Sugiyama, K. Kusamori, Y. Takahashi, Y. Takakura and M. Nishikawa, *Bioorganic & Medicinal Chemistry*, 2021, **29**, 115864.
228 L. S. McCoy, D. Shin and Y. Tor, *J. Am. Chem. Soc.*, 2014, **136**, 15176–15184.
229 Y. Li, A. Fin, L. McCoy and Y. Tor, *Angew. Chem. Int. Ed.*, 2017, **56**, 1303–1307.
230 S. Lyon and V. Gopalan, *ChemBioChem*, 2018, **19**, 142–146.
231 H. Otomo, S. Park, S. Yamamoto and H. Sugiyama, *RSC Adv.*, 2014, **4**, 31341–31344.
232 T. Schwartz, *Science*, 1999, **284**, 1841–1845.
233 M. Sholokh, R. Sharma, N. Grytsyk, L. Zaghzi, V. Y. Postupalenko, D. Dziuba, N. P. F. Barthes, B. Y. Michel, C. Boudier, O. A. Zaporozhets, Y. Tor, A. Burger and Y. Mély, *Chem. Eur. J.*, 2018, **24**, 13850–13861.
234 J.-L. Darlix, J. Godet, R. Ivanyi-Nagy, P. Fossé, O. Mauffret and Y. Mély, *Journal of Molecular Biology*, 2011, **410**, 565–581.
235 S. Bourbigot, N. Ramalanjaona, C. Boudier, G. F. J. Salgado, B. P. Roques, Y. Mély, S. Bouaziz and N. Morellet, *Journal of Molecular Biology*, 2008, **383**, 1112–1128.
236 V. Kilin, K. Gavvala, N. P. F. Barthes, B. Y. Michel, D. Shin, C. Boudier, O. Mauffret, V. Yashchuk, M. Mousli, M. Ruff, F. Granger, S. Eiler, C. Bronner, Y. Tor, A. Burger and Y. Mély, *J. Am. Chem. Soc.*, 2017, **139**, 2520–2528.
237 J. H. Han, S. Yamamoto, S. Park and H. Sugiyama, *Chem. Eur. J.*, 2017, **23**, 7607–7613.
238 J. H. Han, S. Park, F. Hashiya and H. Sugiyama, *Chem. Eur. J.*, 2018, **24**, 17091–17095.
239 R. W. Sinkeldam, L. S. McCoy, D. Shin and Y. Tor, *Angew. Chem. Int. Ed.*, 2013, **52**, 14026–14030.
240 R. A. Mizrahi, D. Shin, R. W. Sinkeldam, K. J. Phelps, A. Fin, D. J. Tantillo, Y. Tor and P. A. Beal, *Angew. Chem. Int. Ed.*, 2015, **54**, 8713–8716.
241 D. Valverde, A. Vasconcelos Sanches de Araujo, A. Carlos Borin and S. Canuto, *Phys. Chem. Chem. Phys.*, 2017, **19**, 29354–29363.
242 D. Valverde, A. V. Sanches de Araujo, S. Canuto and A. C. Borin, *ChemPhotoChem*, 2019, **3**, 916–924.
243 A. V. Sanches de Araújo, D. Valverde, S. Canuto and A. C. Borin, *J. Phys. Chem. A*, 2020, **124**, 6834–6844.
244 A. R. Rovira, A. Fin and Y. Tor, *J. Am. Chem. Soc.*, 2017, **139**, 15556–15559.
245 F. Hallé, A. Fin, A. R. Rovira and Y. Tor, *Angew. Chem.*, 2018, **130**, 1099–1102.
246 Y. Li, P. T. Ludford, A. Fin, A. R. Rovira and Y. Tor, *Chem. Eur. J.*, 2020, **26**, 6076–6084.
247 U. Jenal, A. Reinders and C. Lori, *Nat Rev Microbiol*, 2017, **15**, 271–284.
248 B. Holzberger, J. Strohmeier, V. Siegmund, U. Diederichsen and A. Marx, *Bioorganic & Medicinal Chemistry Letters*, 2012, **22**, 3136–3139.
249 C. A. Kenfack, A. Burger and Y. Mély, *J. Phys. Chem. B*, 2006, **110**, 26327–26336.
250 G. Kodali, K. A. Kistler, M. Narayanan, S. Matsika and R. J. Stanley, *J. Phys. Chem. A*, 2010, **114**, 256–267.
251 M. A. Kochman, M. Pola and R. J. D. Miller, *J. Phys. Chem. A*, 2016, **120**, 6200–6215.
252 A. Ambrus, D. Chen, J. Dai, T. Bialis, R. A. Jones and D. Yang, *Nucleic Acids Res*, 2006, **34**, 2723–2735.
253 J. Li, J. J. Correia, L. Wang, J. O. Trent and J. B. Chaires, *Nucleic Acids Res*, 2005, **33**, 4649–4659.

254 A. T. Phan, V. Kuryavyi, K. N. Luu and D. J. Patel, *Nucleic Acids Res*, 2007, **35**, 6517–6525.
255 S. Müllar, J. Strohmeier and U. Diederichsen, *Org. Lett.*, 2012, **14**, 1382–1385.
256 M. E. Hawkins, in *Methods in Enzymology*, Elsevier, 2008, vol. 450, pp. 201–231.
257 M. E. Hawkins, in *Topics in Fluorescence Spectroscopy*, ed. J. R. Lakowicz, Kluwer Academic Publishers, Boston, 2003, vol. 7, pp. 151–175.
258 K. Wojtuszewski, M. E. Hawkins, J. L. Cole and I. Mukerji, *Biochemistry*, 2001, **40**, 2588–2598.
259 J. C. Myers, S. A. Moore and Y. Shamoo, *J. Biol. Chem.*, 2003, **278**, 42300–42306.
260 M. E. Hawkins, *Nucleic Acids Research*, 2004, **32**, e62–e62.
261 Y. Fu and C. He, *Current Opinion in Chemical Biology*, 2012, **16**, 516–524.
262 T. P. Hoernes and M. D. Erlacher, *WIREs RNA*, 2017, **8**, e1375.
263 T. Pan, *Cell Res*, 2018, **28**, 395–404.
264 K. E. Sloan, A. S. Warda, S. Sharma, K.-D. Entian, D. L. J. Lafontaine and M. T. Bohnsack, *RNA Biology*, 2017, **14**, 1138–1152.
265 S. Matsika, *J. Phys. Chem. A*, 2004, **108**, 7584–7590.
266 M. Z. Zgierski, S. Patchkovskii, T. Fujiwara and E. C. Lim, *J. Phys. Chem. A*, 2005, **109**, 9384–9387.
267 T. Gustavsson, Á. Bányász, E. Lazzarotto, D. Markovitsi, G. Scalmani, M. J. Frisch, V. Barone and R. Improta, *J. Am. Chem. Soc.*, 2006, **128**, 607–619.
268 S. Matsika, in *Photoinduced Phenomena in Nucleic Acids I*, eds. M. Barbatti, A. C. Borin and S. Ullrich, Springer International Publishing, Cham, 2014, vol. 355, pp. 209–243.
269 B. E. Billinghurst and G. R. Loppnow, *J. Phys. Chem. A*, 2006, **110**, 2353–2359.
270 M. Z. Zgierski, T. Fujiwara, W. G. Kofron and E. C. Lim, *Phys. Chem. Chem. Phys.*, 2007, **9**, 3206.
271 Á. Bányász, T. Gustavsson, E. Keszei, R. Improta and D. Markovitsi, *Photochem. Photobiol. Sci.*, 2008, **7**, 765.
272 X. Hua, L. Hua and X. Liu, *Phys. Chem. Chem. Phys.*, 2016, **18**, 13904–13911.
273 K. A. Kistler and S. Matsika, *J. Phys. Chem. A*, 2007, **111**, 8708–8716.
274 S. Laland and G. Serck-Hanssen, *Biochemical Journal*, 1964, **90**, 76–81.
275 S. F. Singleton, F. Shan, M. W. Kanan, C. M. McIntosh, C. J. Stearman, J. S. Helm and K. J. Webb, *Org. Lett.*, 2001, **3**, 3919–3922.
276 B. Gildea and L. W. McLaughlin, *Nucl Acids Res*, 1989, **17**, 2261–2281.
277 J. S. Driscoll, V. E. Marquez, J. Plowman, P. S. Liu, J. A. Kelley and J. J. Barchi, *J. Med. Chem.*, 1991, **34**, 3280–3284.
278 K. Ebenryter-Olbinska, J. Karolak-Wojciechowska and E. Sochacka, *Carbohydrate Research*, 2014, **392**, 7–15.
279 Y. Zhou and P. O. P. Ts'o, *Nucleic Acids Research*, 1996, **24**, 2652–2659.
280 C. J. Adams, J. B. Murray, J. R. P. Arnold and P. G. Stockley, *Tetrahedron Letters*, 1994, **35**, 1597–1600.
281 H. P. Rappaport, *Biochemical Journal*, 2004, **381**, 709–717.
282 D. N. Kaluzhny, S. N. Mikhailov, E. V. Efimtseva, O. F. Borisova, V. L. Florentiev, A. K. Shchyolkina and T. M. Jovin, *Nucleosides, Nucleotides and Nucleic Acids*, 2003, **22**, 1499–1503.
283 O. M. Subach, A. V. Khoroshaev, D. N. Gerasimov, V. B. Baskunov, A. K. Shchyolkina and E. S. Gromova, *Eur J Biochem*, 2004, **271**, 2391–2399.
284 E. Bignon, H. Gattuso, C. Morell, E. Dumont and A. Monari, *Chem. Eur. J.*, 2015, **21**, 11509–11516.
285 G. Ryseck, T. Schmierer, K. Haiser, W. Schreier, W. Zinth and P. Gilch, *ChemPhysChem*, 2011, **12**, 1880–1888.
286 K. Kistler and S. Matsika, *Photochem Photobiol*, 2007, **83**, 611–624.
287 K. A. Kistler and S. Matsika, *J. Phys. Chem. A*, 2007, **111**, 2650–2661.

288 S. Lobsiger, H.-M. Frey, S. Leutwyler, P. Morgan and D. Pratt, *J. Phys. Chem. A*, 2011, **115**, 13281–13290.
289 H. Yokoyama and R. Mizutani, *IJMS*, 2014, **15**, 20321–20338.
290 V. Vendrell-Criado, G. M. Rodríguez-Muñiz, M. C. Cuquerella, V. Lhiaubet-Vallet and M. A. Miranda, *Angew. Chem. Int. Ed.*, 2013, **52**, 6476–6479.
291 A. Fin, A. R. Rovira, P. A. Hopkins and Y. Tor, in *Modified Nucleic Acids*, eds. K. Nakatani and Y. Tor, Springer International Publishing, Cham, 2016, vol. 31, pp. 1–26.
292 N. J. Greco and Y. Tor, *Nat Protoc*, 2007, **2**, 305–316.
293 S. G. Srivatsan and Y. Tor, *J. Am. Chem. Soc.*, 2007, **129**, 2044–2053.
294 S. G. Srivatsan and Y. Tor, *Chem. Asian J.*, 2009, **4**, 419–427.
295 M. G. Pawar, A. Nuthanakanti and S. G. Srivatsan, *Bioconjugate Chem.*, 2013, **24**, 1367–1377.
296 Z. Wen, P. R. Tuttle, A. H. Howlader, A. Vasilyeva, L. Gonzalez, A. Tangar, R. Lei, E. E. Laverde, Y. Liu, J. Miksovska and S. F. Wnuk, *J. Org. Chem.*, 2019, **84**, 3624–3631.
297 P. K. Samanta and S. K. Pati, *Phys. Chem. Chem. Phys.*, 2015, **17**, 10053–10058.
298 R. W. Sinkeldam, A. J. Wheat, H. Boyaci and Y. Tor, *ChemPhysChem*, 2011, **12**, 567–570.
299 N. J. Greco, R. W. Sinkeldam and Y. Tor, *Org. Lett.*, 2009, **11**, 1115–1118.
300 T. Z. Cservenyi, A. J. Van Riesen, F. D. Berger, A. Desoky and R. A. Manderville, *ACS Chem. Biol.*, 2016, **11**, 2576–2582.
301 P. A. Hopkins, L. S. McCoy and Y. Tor, *Org. Biomol. Chem.*, 2017, **15**, 684–690.
302 H. Inoue, A. Imura and E. Ohtsuka, *NIPPON KAGAKU KAISHI*, 1987, 1214–1220.
303 C. Liu and C. T. Martin, *Journal of Molecular Biology*, 2001, **308**, 465–475.
304 S. J. O. Hardman, S. W. Botchway and K. C. Thompson, *Photochemistry and Photobiology*, 2008, **84**, 1473–1479.
305 Q. L. Nguyen, V. A. Spata and S. Matsika, *Phys. Chem. Chem. Phys.*, 2016, **18**, 20189–20198.
306 S. Taherpour and T. Lönnberg, *RSC Adv.*, 2015, **5**, 10837–10844.
307 P. Scharf and J. Müller, *ChemPlusChem*, 2013, **78**, 20–34.
308 M. S. Noé, A. C. Ríos and Y. Tor, *Org. Lett.*, 2012, **14**, 3150–3153.
309 K.-Y. Lin, R. J. Jones and M. Matteucci, *J. Am. Chem. Soc.*, 1995, **117**, 3873–3874.
310 L. M. Wilhelmsson, A. Holmén, P. Lincoln, P. E. Nielsen and B. Nordén, *J. Am. Chem. Soc.*, 2001, **123**, 2434–2435.
311 L. M. Wilhelmsson, P. Sandin, A. Holmén, B. Albinsson, P. Lincoln and B. Nordén, *J. Phys. Chem. B*, 2003, **107**, 9094–9101.
312 P. Sandin, K. Börjesson, H. Li, J. Mårtensson, T. Brown, L. M. Wilhelmsson and B. Albinsson, *Nucleic Acids Research*, 2008, **36**, 157–167.
313 K. Börjesson, P. Sandin and L. M. Wilhelmsson, *Biophysical Chemistry*, 2009, **139**, 24–28.
314 G. Stengel, M. Urban, B. W. Purse and R. D. Kuchta, *Anal. Chem.*, 2009, **81**, 9079–9085.
315 K. Börjesson, S. Preus, A. H. El-Sagheer, T. Brown, B. Albinsson and L. M. Wilhelmsson, *J. Am. Chem. Soc.*, 2009, **131**, 4288–4293.
316 D. D. Burns, K. L. Teppang, R. W. Lee, M. E. Lokensgard and B. W. Purse, *J. Am. Chem. Soc.*, 2017, **139**, 1372–1375.
317 A. Okamoto, K. Tainaka and I. Saito, *J. Am. Chem. Soc.*, 2003, **125**, 4972–4973.
318 A. Okamoto, Y. Saito and I. Saito, *Journal of Photochemistry and Photobiology C: Photochemistry Reviews*, 2005, **6**, 108–122.
319 A. Okamoto, K. Tainaka and I. Saito, *Tetrahedron Letters*, 2003, **44**, 6871–6874.
320 K. Miyata, R. Mineo, R. Tamamushi, M. Mizuta, A. Ohkubo, H. Taguchi, K. Seio, T. Santa and M. Sekine, *J. Org. Chem.*, 2007, **72**, 102–108.
321 M. Mizuta, K. Seio, A. Ohkubo and M. Sekine, *J. Phys. Chem. B*, 2009, **113**, 9562–9569.
322 T. Kanamori, Y. Masaki, M. Mizuta, H. Tsunoda, A. Ohkubo, M. Sekine and K. Seio, *Org. Biomol. Chem.*, 2012, **10**, 1007–1013.
323 K. Seio, T. Kanamori, M. Tokugawa, H. Ohzeki, Y. Masaki, H. Tsunoda, A. Ohkubo and M. Sekine, *Bioorganic & Medicinal Chemistry*, 2013, **21**, 3197–3201.

324 N. J. Leonard, M. A. Sprecker and A. G. Morrice, *J. Am. Chem. Soc.*, 1976, **98**, 3987–3994.
325 A. T. Krueger and E. T. Kool, *J. Am. Chem. Soc.*, 2008, **130**, 3989–3999.
326 M. Fuentes-Cabrera, B. G. Sumpter and J. C. Wells, *J. Phys. Chem. B*, 2005, **109**, 21135–21139.
327 S. K. Jarchow-Choy, A. T. Krueger, H. Liu, J. Gao and E. T. Kool, *Nucleic Acids Research*, 2011, **39**, 1586–1594.
328 H. Gardarsson and S. Th. Sigurdsson, *Bioorganic & Medicinal Chemistry*, 2010, **18**, 6121–6126.
329 P. Cekan and S. Th. Sigurdsson, *Chem. Commun.*, 2008, 3393.
330 F. Godde, J.-J. Toulmé and S. Moreau, *Biochemistry*, 1998, **37**, 13765–13775.
331 F. Godde, K. Aupeix, S. Moreau and J.-J. Toulmé, *Antisense and Nucleic Acid Drug Development*, 1998, **8**, 469–476.
332 H. Liu, J. Gao, L. Maynard, Y. D. Saito and E. T. Kool, *J. Am. Chem. Soc.*, 2004, **126**, 1102–1109.
333 J. Gao, H. Liu and E. T. Kool, *J. Am. Chem. Soc.*, 2004, **126**, 11826–11831.
334 G. Mata, O. P. Schmidt and N. W. Luedtke, *Chem. Commun.*, 2016, **52**, 4718–4721.
335 C. P. Lawson, A. F. Füchtbauer, M. S. Wranne, T. Giraud, T. Floyd, B. Dumat, N. K. Andersen, A. H. El-Sagheer, T. Brown, H. Gradén, L. M. Wilhelmsson and M. Grøtli, *Sci Rep*, 2018, **8**, 13970.
336 I. Okamura, S. Park, R. Hiraga, S. Yamamoto and H. Sugiyama, *Chem. Lett.*, 2017, **46**, 245–248.
337 S. Hirashima, J. H. Han, H. Tsuno, Y. Tanigaki, S. Park and H. Sugiyama, *Chem. Eur. J.*, 2019, **25**, 9913–9919.
338 J. H. Han, S. Hirashima, S. Park and H. Sugiyama, *Chem. Commun.*, 2019, **55**, 10245–10248.
339 S. Hirashima, H. Sugiyama and S. Park, *J. Phys. Chem. B*, 2020, **124**, 8794–8800.
340 G. N. Samaan, M. K. Wyllie, J. M. Cizmic, L.-M. Needham, D. Nobis, K. Ngo, S. Andersen, S. W. Magennis, S. F. Lee and B. W. Purse, *Chem. Sci.*, 2021, **12**, 2623–2628.
341 K. Miyata, R. Tamamushi, A. Ohkubo, H. Taguchi, K. Seio, T. Santa and M. Sekine, *Org. Lett.*, 2006, **8**, 1545–1548.
342 J. A. Secrist, J. R. Barrio and N. J. Leonard, *Science*, 1972, **175**, 646–647.
343 A. Okamoto, K. Tanaka, T. Fukuta and I. Saito, *J. Am. Chem. Soc.*, 2003, **125**, 9296–9297.
344 A. Okamoto, K. Tanaka, T. Fukuta and I. Saito, *ChemBioChem*, 2004, **5**, 958–963.
345 A. Dierckx, P. Dinér, A. H. El-Sagheer, J. D. Kumar, T. Brown, M. Grøtli and L. M. Wilhelmsson, *Nucleic Acids Research*, 2011, **39**, 4513–4524.
346 A. Dierckx, F.-A. Miannay, N. Ben Gaied, S. Preus, M. Björck, T. Brown and L. M. Wilhelmsson, *Chem. Eur. J.*, 2012, **18**, 5987–5997.
347 A. Wypijewska del Nogal, A. F. Füchtbauer, M. Bood, J. R. Nilsson, M. S. Wranne, S. Sarangamath, P. Pfeiffer, V. S. Rajan, A. H. El-Sagheer, A. Dahlén, T. Brown, M. Grøtli and L. M. Wilhelmsson, *Nucleic Acids Research*, 2020, **48**, 7640–7652.
348 B. Dumat, M. Bood, M. S. Wranne, C. P. Lawson, A. F. Larsen, S. Preus, J. Streling, H. Gradén, E. Wellner, M. Grøtli and L. M. Wilhelmsson, *Chem. Eur. J.*, 2015, **21**, 4039–4048.
349 A. F. Füchtbauer, M. S. Wranne, S. Sarangamath, M. Bood, A. H. El-Sagheer, T. Brown, H. Gradén, M. Grøtli and L. M. Wilhelmsson, *ChemPlusChem*, 2020, **85**, 319–326.
350 M. Bood, A. F. Füchtbauer, M. S. Wranne, J. J. Ro, S. Sarangamath, A. H. El-Sagheer, D. L. M. Rupert, R. S. Fisher, S. W. Magennis, A. C. Jones, F. Höök, T. Brown, B. H. Kim, A. Dahlén, L. M. Wilhelmsson and M. Grøtli, *Chem. Sci.*, 2018, **9**, 3494–3502.
351 R. S. Fisher, D. Nobis, A. F. Füchtbauer, M. Bood, M. Grøtli, L. M. Wilhelmsson, A. C. Jones and S. W. Magennis, *Phys. Chem. Chem. Phys.*, 2018, **20**, 28487–28498.
352 M. Kouchakdjian, M. Eisenberg, K. Yarema, A. Basu, J. Essigmann and D. J. Patel, *Biochemistry*, 1991, **30**, 1820–1828.
353 P. Sharma, L. A. Lait and S. D. Wetmore, *Phys. Chem. Chem. Phys.*, 2013, **15**, 15538.

354 T. Kent, T. D. Rusanov, T. M. Hoang, W. A. Velema, A. T. Krueger, W. C. Copeland, E. T. Kool and R. T. Pomerantz, *Nucleic Acids Res*, 2016, gkw721.
355 C. A. Buhr, M. D. Matteucci and B. C. Froehler, *Tetrahedron Letters*, 1999, **40**, 8969–8970.
356 A. Foller Larsen, B. Dumat, M. S. Wranne, C. P. Lawson, S. Preus, M. Bood, H. Gradén, L. Marcus Wilhelmsson and M. Grøtli, *Sci Rep*, 2015, **5**, 12653.
357 M. S. Wranne, A. F. Füchtbauer, B. Dumat, M. Bood, A. H. El-Sagheer, T. Brown, H. Gradén, M. Grøtli and L. M. Wilhelmsson, *J. Am. Chem. Soc.*, 2017, **139**, 9271–9280.
358 Y. Liu, E. Holmstrom, J. Zhang, P. Yu, J. Wang, M. A. Dyba, De Chen, J. Ying, S. Lockett, D. J. Nesbitt, A. R. Ferré-D'Amaré, R. Sousa, J. R. Stagno and Y.-X. Wang, *Nature*, 2015, **522**, 368–372.
359 Y. Liu, E. Holmstrom, P. Yu, K. Tan, X. Zuo, D. J. Nesbitt, R. Sousa, J. R. Stagno and Y.-X. Wang, *Nat Protoc*, 2018, **13**, 987–1005.
360 E. A. Alemán, C. de Silva, E. M. Patrick, K. Musier-Forsyth and D. Rueda, *J. Phys. Chem. Lett.*, 2014, **5**, 777–781.