

# Chemical Science

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

## C-Linked 8-Aryl Guanine Nucleobase Adducts: Biological Outcomes and Utility as Fluorescent Probes

 Richard A. Manderville<sup>a</sup> and Stacey D. Wetmore<sup>b</sup>

 Received 00th January 20xx,  
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Aryl radical species derived from enzymatic transformations of aromatic mutagens preferentially react at the 8-site of the guanine (G) nucleobase to afford carbon-linked 8arylG adducts. The resulting lesions possess altered biophysical and genetic coding properties compared to the precursor G nucleoside in B-form DNA. Unlike other adducts, these lesions also possess useful fluorescent properties, since direct attachment of the 8aryl ring extends the purine  $\pi$ -system to afford G mimics with red-shifted excitation maxima and emission that can be sensitive to the microenvironment of the 8arylG base within nucleic acid structures. In B-form DNA, 8arylG adducts are disruptive to duplex formation because they prefer to adopt the *syn*-conformation about the bond connecting the nucleobase to the deoxyribose backbone, which perturbs Watson-Crick (WC) H-bonding with the opposing cytosine (C). Thus, in a B-form duplex, the emissive properties of 8arylG adducts can be employed as a tool to provide insight into adduct conformation, which can be related to their biological outcomes. However, since Gs preferentially adopt the *syn*-conformation in left-handed Z-DNA and antiparallel G-quadruplex (GQ) structures, 8arylG lesions can be inserted into *syn*-G positions without disrupting H-bonding interactions. In fact, 8arylG lesions can serve as ideal fluorescent probes in an antiparallel GQ because their emission is sensitive to GQ folding. This perspective outlines recent developments in the biological implications of 8arylG formation together with their utility as fluorescent G analogs for use in DNA-based diagnostic systems.

### Introduction

DNA provides a useful scaffold from which to design novel molecules with applications in diagnostics,<sup>1</sup> therapeutics,<sup>2</sup> nanoscience,<sup>3,4</sup> and genomics.<sup>5</sup> Chemical modification of the DNA nucleobases,<sup>5-11</sup> or the sugar-phosphate backbone,<sup>12,13</sup> can provide new desirable properties that expand the scope of DNA applications. Nucleobase modifications are of particular interest because these can change DNA base-pairing to expand the genetic alphabet,<sup>6</sup> afford new redox<sup>7,8</sup> or fluorescent<sup>9-11</sup> sensing bases, or affect the gene function in different tissues.<sup>5</sup>

In contrast to the design and synthesis of modified DNA bases for specific applications, useful nucleobase modifications have been identified by studying the cellular implications of DNA damage.<sup>14</sup> This is particularly true for DNA adducts (addition products), which stem from attack of the human genome by reactive chemical species. For example, 7,8-dihydro-8-oxo-2'-deoxyguanosine (8oxoG, Fig. 1) is a biomarker for oxidative DNA damage.<sup>8</sup> The biological impact of 8oxoG stems from the 8-keto conformation, which permits 8oxoG to form a stable Hoogsteen base pair with adenine.<sup>15</sup> Consequently, 8oxoG is mutagenic and unrepaired 8-oxoG lesions cause G→T transversions in cells.<sup>16</sup> Remarkably, 8oxoG possesses an even lower oxidation potential ( $E_{1/2} = 0.74$  V

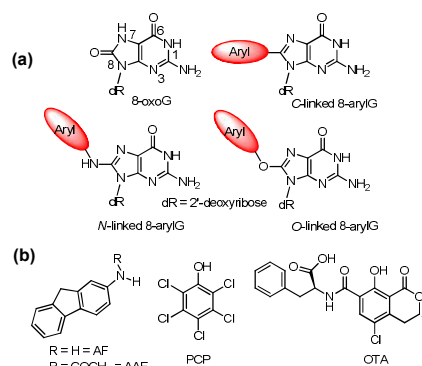


Fig. 1 (a) Structures of 8oxoG and 8arylG adducts with different linkages. (b) Structures of mutagens that produce the different 8arylG adducts.

versus NHE<sup>17</sup>) than the parent G ( $E_{1/2} = 1.29$  V versus NHE<sup>18</sup>), permitting its selective oxidation in DNA substrates. Thus, the 8oxoG lesion has become an effective redox tool for studying DNA electron transfer<sup>19,20</sup> and detecting single-base mismatches.<sup>21</sup>

Beyond the formation of 8oxoG, aryl groups can covalently attach to the 8-position of G following enzymatic transformations of different mutagens and carcinogens, and the resulting lesions have been classified as either nitrogen-, carbon- or oxygen-linked 8arylG adducts (Fig. 1).<sup>22-28</sup> Extensive efforts have focused on the biological impact of the N-linked derivatives because these lesions are produced by notorious chemical carcinogens that are present in, for example cigarette

<sup>a</sup> Department of Chemistry & Toxicology, University of Guelph, Guelph, ON, Canada N1G 2W1, E-mail: rmanderv@uoguelph.ca

<sup>b</sup> Department of Chemistry & Biochemistry, University of Lethbridge, Lethbridge, AB, Canada T1K 3M4, E-mail: Stacey.Wetmore@uleth.ca

smoke, fossil fuels and cooked meats.<sup>29</sup> The associated adducts have been detected in human cells,<sup>30,31</sup> including lesions arising from 2-aminofluorene (AFG) and N2-acetylaminofluorene (AAF<sub>2</sub>, Fig. 1b), which serve as prototype adducts to elucidate the molecular mechanisms of mutagenesis induced by arylamine carcinogens.<sup>32-34</sup> In contrast to the N-linked variants, relatively little is understood about the structure and biological impact of O- and C-linked 8arylG adducts. Nevertheless, these lesions also arise from our exposure to a variety of sources. For example, O-linked adducts have been connected with pentachlorophenol (PCP, Fig. 1b) found in pesticides, disinfectants and wood preservatives,<sup>22,23,25,27</sup> while the food mutagen ochratoxin A (OTA) arising from several species of (*Aspergillus* and *Penicillium*) fungi has been linked to the formation of a C-linked adduct (Fig. 1b).<sup>35-37</sup>

In terms of structure, the C-linked 8arylG adducts are unique because they lack the flexible tether that separates the aryl component from the G nucleobase in the N- and O-linked lesions. As a result, C-linked aryl moieties extend the purine  $\pi$ -system, which commonly results in fluorescent nucleobase analogues.<sup>9-11</sup> Emissive DNA bases enjoy a wide range of applications that include reporters of nucleic acid structure and function;<sup>38-43</sup> detectors of nucleobase damage, single nucleotide polymorphisms (SNPs) and mismatch dynamics;<sup>44-46</sup> probes for understanding protein-DNA interactions;<sup>47-50</sup> components of aptamers to provide a fluorescent signal upon target binding;<sup>51-53</sup> and oligonucleotide-based therapeutics.<sup>54</sup>

Ideal fluorescent probes provide emission switching properties while retaining the native behavior of the nucleic acid system being studied. Indeed, it is considered detrimental for probe incorporation to produce a major perturbation to duplex stability, H-bonding interactions, or folding characteristics. Modifications to the 8-position of G offer an attractive avenue to fluorescent probes because this site is not involved in canonical base-pairing interactions.<sup>38</sup> Nevertheless, fluorescent C-linked 8arylG adducts are detrimental to the stability of a B-DNA duplex (Fig. 2) because they prefer to adopt the *syn*-conformation.<sup>55</sup> In this scenario, the emissive properties of 8-arylG adducts can be exploited as a tool to provide insight into adduct conformation within the helix,<sup>24,28</sup> which can be related to biological outcome, as discussed above for 8oxoG.

Alternative oligonucleotide structures, such as Z-DNA<sup>56,57</sup> and antiparallel G-quadruplexes (GQs),<sup>58,59</sup> are also biologically relevant (Fig. 2). For example, Z-DNA plays a role in transcription and has been implicated in mutagenesis,<sup>57</sup> while GQs are active drug targets.<sup>59</sup> Both Z-DNA and GQs contain Gs that preferentially adopt the *syn*-conformation. In these DNA structures, 8arylGs can occupy *syn*-G positions without disturbing the overall stability and H-bonding interactions in the nucleic acid. In Z-DNA produced by alternating purine-pyrimidine sequences, the purines adopt the *syn*-conformation, while pyrimidines maintain the *anti*-

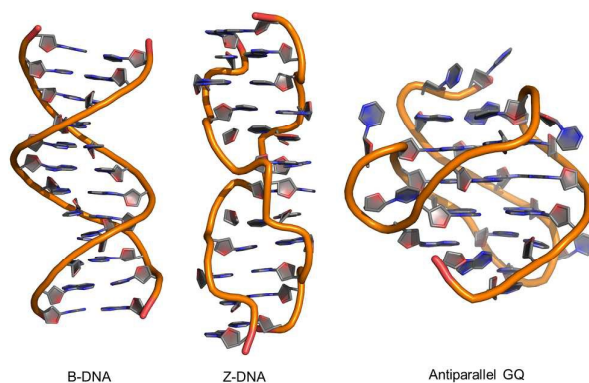


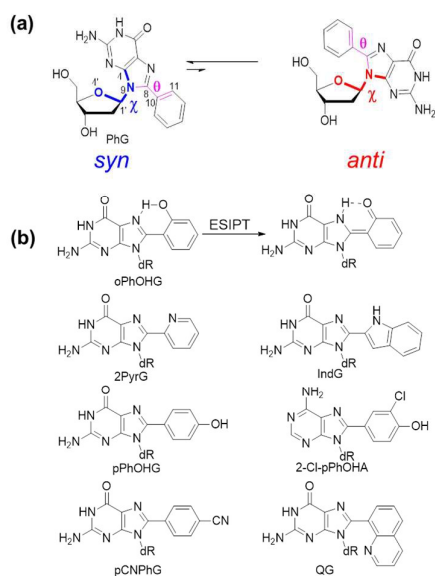
Fig. 2 Structures of B-form and Z-form DNA duplexes and an intramolecular antiparallel G-quadruplex (GQ).

conformation, leading to a left-handed “zigzag” helix. Antiparallel GQs are composed of stacked G-quartets, which are stabilized by certain cations and contain alternating *syn*- and *anti*-Gs, and intervening sequences, which are extruded as single strand loops. The ability of 8arylG lesions to promote the formation of Z-DNA and GQs may provide a rationale for their biological activity.<sup>60-63</sup> In antiparallel GQs, 8arylGs also behave as ideal fluorescent probes, exhibiting emission that is sensitive to GQ folding.<sup>38,42,52,53</sup>

In this perspective, we discuss the synthesis, properties, biological impact and applications of 8arylG adducts. We start by summarizing the nucleoside structures and properties, and methods for their incorporation into oligonucleotide substrates. The structural and biological impact of 8arylG adducts within the B-DNA “NarI” recognition sequence (5'-G<sub>1</sub>G<sub>2</sub>CG<sub>3</sub>CC) is then presented, which highlights a comparison between the properties of the C-linked 8arylG adducts with the established biological properties of the N-linked variants. Novel carcinogenesis mechanisms for C-linked 8arylG adducts are then discussed based on their unique ability to promote Z-DNA and GQ formation. Finally, the utility of fluorescent 8arylG probes in duplex-GQ exchange systems is presented, which provides a signalling platform in biosensors. Together, the work highlighted within points toward a rich future for 8arylG probes in aptasensor development.

## Nucleoside synthesis, structure and properties

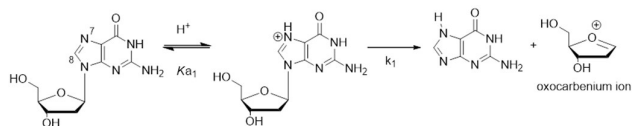
The C-linked 8arylG nucleosides are routinely synthesized from 8-BrG and the appropriate arylboronic acid using the Suzuki-Miyaura cross-coupling reaction.<sup>64</sup> Calculations (B3LYP/6-31G(d)) reveal that *anti*-structures are less stable than *syn*-structures, mainly since all *syn* minima contain an O5'-H...N3 hydrogen bond (1.80–1.96 Å).<sup>55,65-67</sup> The predicted *syn*-preference for the C-linked 8arylG nucleosides is supported by NMR spectra that exhibit a downfield shift of H2', C1', C3' and C4' and an upfield shift of C2' compared to the corresponding signals for native G.<sup>67,68</sup> Solid-state structures of C-linked 8arylG nucleosides, including an 8-quinolyl derivative (QG)<sup>24</sup> and an (*N,N*-dimethylaniline)guanosine analog,<sup>69</sup> are also *syn*. As a specific example, DFT calculations predict the *syn*



**Fig. 3** (a) *Syn* (left) and *anti* (right) structures of PhG. The dihedral angle  $\chi$  ( $\angle(O4'-C1'-N9-C4)$ ) defines the glycosidic bond orientation to be *anti* ( $\chi = 180 \pm 90^\circ$ ) or *syn* ( $\chi = 0 \pm 90^\circ$ ) and  $\theta$  ( $\angle(N9-C8-C10-C11)$ ) defines the degree of twist between the nucleobase and the 8-phenyl substituent. (b) Structures of 8-arylG probes.

orientation of the smallest (unsubstituted) 8-phenylG (PhG) derivative (Fig 3a) to be favoured over *anti* by  $25.1 \text{ kJ mol}^{-1}$ .<sup>24</sup> In a representative low-energy *syn*-structure,  $\chi = 66.8^\circ$  and the twist angle ( $\theta$ ) between the phenyl ring and the nucleobase is  $41.2^\circ$ .<sup>24</sup> The PhG nucleoside is also fluorescent in H<sub>2</sub>O ( $\lambda_{\text{em}} = 395 \text{ nm}$ ,  $\Phi_{\text{fl}} = 0.44$ <sup>24</sup>) following excitation at  $277 \text{ nm}$  ( $\log \varepsilon = 4.33$ <sup>70</sup>).

Given that PhG exhibits impressive emission intensity in H<sub>2</sub>O, it became apparent that manipulation of the phenyl ring with various substituents could generate new fluorescent nucleobase analogs with potentially useful emission switching properties (Fig. 3b). For example, the *ortho*-substituted phenolic nucleoside (oPhOHG) can produce a planar conformation in aprotic solvents due to an intramolecular H-bond between the phenolic OH and N<sup>7</sup> of the nucleobase.<sup>71</sup> The planar species absorbs at  $\sim 320 \text{ nm}$  and displays visible emission at  $476 \text{ nm}$  due to an excited-state intramolecular proton transfer (ESIPT) process to afford the keto tautomer (Fig. 3b).<sup>71</sup> In H<sub>2</sub>O, the intramolecular H-bond is disrupted and the nucleoside adopts a twisted structure ( $\lambda_{\text{vmax}} = 280 \text{ nm}$ ) that displays enol emission at  $\lambda_{\text{em}} = 395 \text{ nm}$  ( $\Phi_{\text{fl}} = 0.44$ ).<sup>71</sup> This differential fluorescence has proven useful for probing the local solvent environment of oPhOHG within duplex DNA.<sup>72</sup> The oPhOHG adduct also binds Cu(II) ( $\log K_{\text{a}} = 4.59$ ) and Ni(II) ( $\log K_{\text{a}} = 3.65$ ) effectively,<sup>73</sup> and the enol emission is quenched upon such metal coordination, illustrating the ability to use fluorescence to monitor metal ion binding by nucleic acids.<sup>73</sup> Likewise, the pyridyl derivative (2PyrG) has also been employed for metal ion binding by various DNA topologies.<sup>74</sup>



**Scheme 1** Acid-catalyzed hydrolysis of 2'-deoxyguanosine.

The attached pyridyl group in 2PyrG extends the Hoogsteen H-bonding face of the G nucleobase by an additional H-bonding acceptor. In contrast, the indole-linked derivative (IndG) extends the Hoogsteen H-bonding face by an H-bonding donor, and exhibits emission at  $395 \text{ nm}$  ( $\Phi_{\text{fl}} = 0.78$  in H<sub>2</sub>O,  $\lambda_{\text{ex}} = 321 \text{ nm}$ ) that is sensitive to WC (quenched emission) versus Hoogsteen H-bonding (enhanced emission intensity).<sup>75</sup> The *para*-substituted phenolic nucleoside (pPhOHG) exhibits pH-sensitive fluorescent properties.<sup>76</sup> The neutral adduct displays emission at  $390 \text{ nm}$  ( $\Phi_{\text{fl}} = 0.47$  in H<sub>2</sub>O) that is quenched upon phenolate formation ( $\text{p}K_{\text{a}} \sim 8.9$ ). This result prompted the synthesis of the 8-(2-chloro-phenol)adenosine analogue (2-Cl-pPhOHA) with a  $\text{p}K_{\text{a}}$  of 7.29 for fluorescent pH-sensing activity in the physiological pH range.<sup>76</sup>

The electron-rich nature of the G nucleobase also permits generation of visibly emissive derivatives by attaching electron-withdrawing 8-aryl groups to afford new derivatives with push-pull characteristics.<sup>24,42</sup> The pCNPhG nucleoside displays blue emission at  $468 \text{ nm}$  that is quenched in H<sub>2</sub>O ( $\Phi_{\text{fl}} = 0.04$ ), but lights-up in aprotic solvents ( $\Phi_{\text{fl}} = 0.43$  in CH<sub>3</sub>CN).<sup>42</sup> The 8-quinolyl derivative (QG) displays dual fluorescence in CH<sub>3</sub>CN at  $384$  and  $510 \text{ nm}$ . However, since red-shifted twisted intramolecular charge-transfer (TICT) states for QG are strongly quenched in H<sub>2</sub>O, the single peak for QG in H<sub>2</sub>O at  $407 \text{ nm}$  ( $\Phi_{\text{fl}} = 0.03$ ) was ascribed to locally excited (LE) emission.<sup>24</sup> The emission sensitivity to solvent polarity displayed by these donor-acceptor adducts has proven useful for predicting adduct conformation in duplex structures.<sup>24</sup>

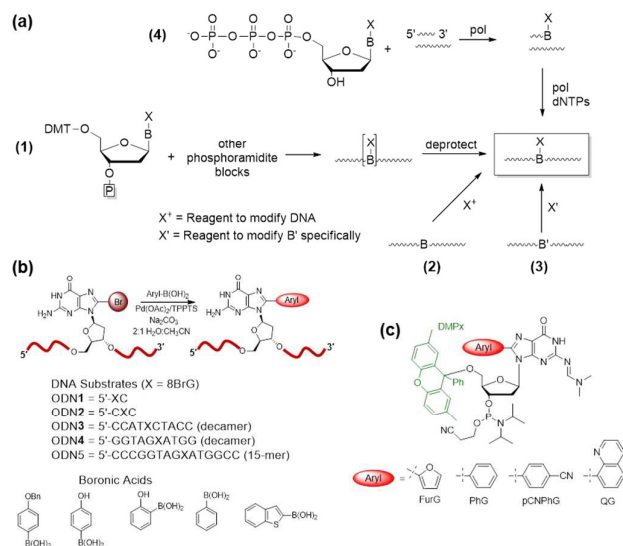
Modification of the 8-site of G with aryl groups also accelerates the rate of acid-catalyzed hydrolysis.<sup>70</sup> Hydrolysis of the unmodified base (Scheme 1) proceeds via a stepwise mechanism, involving initial protonation at N<sup>7</sup> ( $\text{p}K_{\text{a}}$  of the protonated species in  $2.34$ <sup>77</sup>) followed by rate-limiting unimolecular cleavage of the glycosidic bond ( $k_1$ ).<sup>22,70</sup> The oxocarbenium ion subsequently undergoes hydration to produce the 1'-hydroxylated-2'-deoxyribose sugar. Rates of hydrolysis for 8-arylG adducts are 5- to 45-fold greater than the unmodified base in  $0.1 \text{ M HCl}$ , which increases to 9- to 200-fold at pH 4.<sup>70</sup> The relative depurination efficiencies of C-linked 8-arylG adducts can be rationalized by the relative relief of steric strain in the twisted nucleoside (about the base-aryl moiety bond) upon deglycosylation, which yields a less twisted or even planar nucleobase.<sup>70</sup> Thus, the presence and chemical composition of the bulky moiety in C-linked 8-arylG lesions can have a significant effect on the structure and properties of the isolated nucleoside.

## Incorporation into oligonucleotides

Four common approaches are utilized to incorporate modified DNA bases into oligonucleotides in a site-specific fashion (Fig. 4a).<sup>78</sup> The phosphoramidite approach (Fig. 4a (1)) requires the chemical synthesis of the modified phosphoramidite that can be incorporated into oligonucleotides using a DNA synthesizer. The modified phosphoramidite must be compatible with the DNA synthesis conditions (acidic, alkaline and oxidative) and deprotection (strongly alkaline) of the DNA strand.<sup>1</sup> Alternatively, postsynthetic strategies may involve treating an oligonucleotide with a reagent that will modify the desired nucleobase to generate the desired lesion (Fig. 4a (2)). An inherent problem with this approach is that multiple modifications may occur and positional isomers can be difficult to separate. To circumvent some of these challenges, an appropriately modified base may be incorporated site-specifically into an oligonucleotide and then chemically modified to generate the desired final product (Fig. 4a (3)). Finally, modified deoxynucleotide triphosphates may be incorporated enzymatically into a primer strand by the action of a DNA polymerase (Fig. 4a (4)).<sup>79</sup> Nevertheless, this approach is limited by the activity of polymerases toward a variety of bulky damaged products.

For site-specific incorporation of 8arylG bases into oligonucleotide substrates, we were concerned that their sensitivity to acids and oxidants may pose issues for their synthesis using the solid-phase approach with 8arylG phosphoramidites. The Hocek laboratory also reported that 8-Ph-dATP is too bulky to be a polymerase substrate,<sup>79</sup> which renders an enzymatic approach unfeasible. Thus, we developed a postsynthetic strategy utilizing the palladium-catalyzed (Suzuki-Miyaura) cross-coupling reaction (Fig. 4b).<sup>80</sup> In this strategy, the commercially available phosphoramidite of 8BrG is incorporated site-specifically into various oligonucleotide substrates using solid-phase DNA synthesis and then reacted with a range of arylboronic acids. This strategy avoids exposure of the 8arylG base to acids and oxidants, and can be employed to incorporate a single adduct into relatively short DNA substrates (3-15mers). However, the strategy has a number of limitations, including poor yields, the inability to incorporate multiple adducts into strands and limitations in generating the longer adducted DNA templates that are required to assess the biological impact of the lesion using DNA polymerases or repair enzymes.

A more efficient protocol for synthesis of DNA substrates containing 8arylG bases utilizes the 5'-O-2,7-dimethylpityl (DMPx) protecting group (Fig. 4c) in a solid-phase assisted synthesis strategy.<sup>81</sup> The DMPx group is more acid-labile than DMT (release of an aromatic carbocation versus a benzylic carbocation) and can be efficiently removed using 0.5% dichloroacetic acid (DCA) in dichloromethane rather than the 3% DCA required to remove the 5'-O-DMT protecting group.<sup>82</sup> Indeed, we successfully employed the DMPx group to incorporate a range of 8arylG bases (FurG, PhG, pCNPhG, QG,

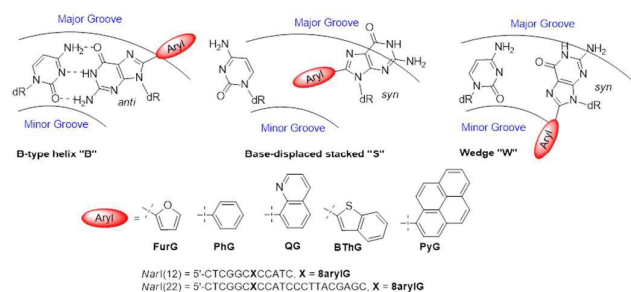


**Fig. 4** (a) Methods for synthesis of oligonucleotides containing modified DNA bases at defined positions. (b) Postsynthetic Suzuki-Miyaura cross-coupling of 8BrG-modified oligonucleotides for synthesis of 8arylG-modified DNA substrates. (c) Utility of the DMPx group for solid-phase synthesis of oligonucleotides containing acid-sensitive 8arylG adducts.

Fig. 4c) into DNA substrates using solid-phase synthesis. For incorporation of FurG, which exhibits a 55.2-fold increase in hydrolysis rate compared to dG in 0.1 M aqueous HCl, the 5'-O-DMPx group provided a 4-fold yield increase compared to use of the 5'-O-DMT group.<sup>1</sup> However, attempts to purify the modified DNA substrates with the final 5'-O-DMPx group attached using commercially available solid-phase extraction cartridges resulted in degradation of the oligonucleotide through exposure of the 8arylG base to both acid and water.<sup>81</sup> Therefore, although 8arylG bases can be incorporated into DNA substrates using solid-phase synthesis (DMPx or DMT protection), it is critical to remove the final 5'-OH protecting group on-column prior to cleavage of the oligonucleotide from the solid support using aqueous ammonium hydroxide.

## Structural and biological impact

As a first step toward understanding the biological impact of bulky DNA adducts, efforts have been made to relate adducted duplex structures to mutagenic outcomes. For N-linked 8arylG adducts, three distinct conformational themes have been characterized for the associated adducted DNA by NMR spectroscopy, which depend on the (*anti* or *syn*) glycosidic bond orientation and the location of the 8aryl group within the duplex (Fig. 5).<sup>29,34</sup> In the B-type conformation, the adduct adopts the *anti*-glycosidic orientation and maintains WC H-bonding with the opposing base C, which positions the 8aryl group in the solvent-exposed major groove. In the base-displaced stacked (*S*-type) conformation, the adduct adopts the *syn*-orientation and stacks the 8aryl ring between the flanking base pairs, which displaces the opposing C out of the helix. In the wedge (*W*-type) conformation, the 8arylG adduct also adopts the *syn*-conformation, but the 8aryl moiety is



**Fig. 5** Depictions of the three major conformations produced by 8arylG adducts, various 8aryl groups used to model C-linked 8arylG adducts and oligonucleotide sequences of *NarI*(12) and *NarI*(22).

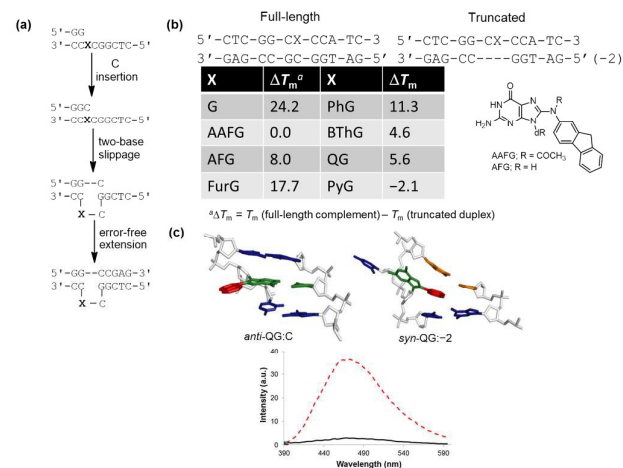
positioned in the minor groove and the opposing base remains in the helix. The distribution amongst the various conformations is dependent on the nature of the attached 8aryl moiety (planar-fused multiring systems versus single ring or nonplanar groups) and the neighbouring base sequence context.<sup>29,34</sup> For example, the N-linked 8arylG adduct produced by AAF induces all three conformational themes in normally paired DNA duplexes, while changes in the linker substitution results in the B and S-type conformations for the N-linked adduct produced by AF.<sup>29,32-34</sup> Thus, each adduct yields multiple conformations of adducted DNA, with each conformation leading to different mutational outcomes. In general, the B-type conformation will favour insertion of the correct base C opposite the lesion, but further extension can be both error free and error prone.<sup>32</sup> The mutagenic arrangement following C insertion opposite the N-linked adduct is induced by interactions between the polymerase and the bulky fluorenyl rings rather than by a base pair-stabilized misalignment.<sup>32</sup> The S-type conformation is strongly blocking because the fluorenyl rings occupy the position of the incoming dNTP within the polymerase; the S-type conformation is also associated with the production of deletion mutations.<sup>29</sup> The W-type conformation is strongly associated with misincorporation due to hydrogen bonding interactions between the incoming mismatched dNTPs with the Hoogsteen-face of the *syn*-adduct.<sup>29</sup>

C-Linked 8arylG adducts are much more rigid than the N-linked counterparts because they lack a flexible tether separating the 8aryl ring from the G nucleobase. They also tend to be highly twisted about the nucleobase–aryl moiety bond in order to reduce steric interactions that arise due to the closer proximity of the aryl group and sugar moiety, as well as decreased inherent flexibility within the bulky moiety, in the absence of the tether. As a result, C-linked 8arylG adducts likely exhibit a decreased tendency to produce the highly blocking/mutagenic S-type conformation compared to their N-linked counterparts. Nevertheless, the formation of C-linked 8arylG adducts has also been implicated in a variety of mutagenic outcomes. For example, arylhydrazines, which generate phenyl radicals, produce 8PhG adducts that are mutagenic in bacteria.<sup>60-62</sup> Benzo[a]pyrene (B[a]P) undergoes peroxidase-mediated oxidation to afford radical cations, where the potential involvement of an 8B[a]PG adduct produces

transversion (G → T and G → C) mutations in yeast.<sup>83</sup> The phenolic mycotoxin ochratoxin A (OTA) produces a bulky C-linked 8arylG adduct,<sup>35-37</sup> and increases the mutational frequency, as well as the induction of deletion mutations and double strand breaks, in the kidneys of male rats.<sup>84</sup>

To demonstrate relationships between 8aryl ring size, adduct conformation and *in vitro* mutagenicity for C-linked 8arylG adducts, five adducts with differing ring types (FurG, PhG, QG, BThG and PyG) were incorporated into the reiterated G<sub>3</sub>-position (X) of the *NarI* Type II restriction endonuclease recognition sequence (i.e. *NarI*(12) and *NarI*(22), Fig. 5).<sup>24,28</sup> The G<sub>3</sub>-site of *NarI* is part of a CpG dinucleotide repeat that is a hotspot for two-base deletion mutations induced by polycyclic N-linked 8arylG adducts in bacteria via a two-base slippage mechanism (Fig. 6a).<sup>85</sup> Optical experiments (ultraviolet (UV) thermal melting, circular dichroism (CD) and fluorescence) combined with molecular dynamics (MD) simulations were utilized to define the structural features of the adducted *NarI*(12) duplexes.<sup>24,28</sup> The adducted *NarI*(22) templates were also employed in primer elongation experiments to assess adduct impact on DNA replication *in vitro* by the polymerases-*Escherichia coli* DNA polymerase I Klenow fragment exo<sup>-</sup> (Kf), and DNA polymerase IV from *Sulfolobus solfataricus* P2 (Dpo4). High-fidelity polymerases, such as Kf, can fit one templating nucleotide in their active site, and favour accurate replication when correct WC base pairing is established with the incoming deoxynucleotide triphosphate (dNTP).<sup>86</sup> Bulky DNA adducts often stall or block DNA replication by high-fidelity polymerases, which *in vivo* is believed to be a signal for the recruitment of Y-family translesion polymerases. The Y-family polymerases have spacious solvent-exposed active sites that can accommodate bulky DNA lesions, while facilitating low-fidelity DNA replication.<sup>78,86</sup> Dpo4 is regarded as a prototypical Y-family polymerase that serves as an excellent model for investigating how structural features of adducts determine lesion bypass efficiency and fidelity.

In the *NarI*(12) duplex, C-linked 8arylG adducts paired opposite the correct base C strongly decrease duplex stability compared to the unmodified control due to their energetic preference for the *syn*-conformation.<sup>24,28</sup> More importantly, the ability of a lesion to stabilize the slippage product relative to the full-length duplex correlates with an ability to induce –2 frameshift mutagenesis in bacteria. As a result, at the G<sub>3</sub>-position of *NarI*, thermal melting parameters ( $T_m$  values) of the full-length complement duplex (with the adduct paired opposite the correct base C) have been compared to  $T_m$  values of *NarI*(12) hybridized to a truncated 10mer (two-base deletion) sequence (–2) (Fig. 6b).<sup>87</sup> The full-length complement and truncated duplexes containing the N-linked 8arylG adduct of AAF (AAFG) have the same  $T_m$  ( $\Delta T_m = 0.0$  °C, Fig. 6b). This highlights the ability of the AAFG lesion to stabilize the slippage product,<sup>87</sup> which correlates with reports that AAFG produces –2 deletions (91% mutational frequency) at the G<sub>3</sub>-position in *NarI* in bacteria.<sup>88</sup> In contrast, the AFG lesion stabilizes the slippage product to a lesser degree ( $\Delta T_m = 8.0$  °C)<sup>87</sup> and leads to only base-pair substitution mutations.<sup>88</sup> For the C-linked 8arylG adducts in *NarI*(12), full-length



**Fig. 6** (a) Proposed model for two-base slippage induced by N-linked 8arylG lesions in a CpG dinucleotide repeat sequence. (b) Thermal melting ( $T_m$ ) values for 8arylG lesions in full-length relative to truncated *Narl*(12) duplexes. (c) Most stable MD structures of QG (quinolyl group in red, G nucleobase in green) in *Narl*(12) paired opposite C (*anti*-QG:C), within the truncated duplex (*syn*-QG:-2) and emission spectra ( $\lambda_{ex} = 330$  nm) of QG within the full-length (solid black trace) and truncated (dashed red trace) duplexes.

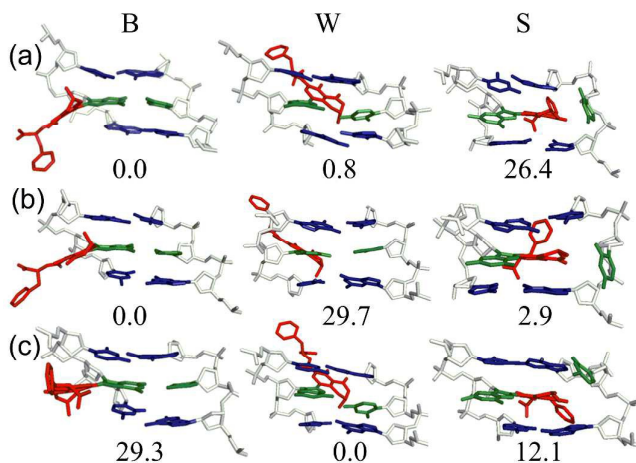
duplexes containing the single-ringed derivatives (FurG and PhG) are significantly more stable than the truncated duplex, suggesting an inability of these lesions to induce -2 deletions y stabilizing the slippage product. In contrast, the larger fused-ringed derivatives (BThG, QG and PyG) exhibit similar  $T_m$  values for the full-length and truncated duplexes, with the truncated duplex being more stable for the bulkiest PyG adduct ( $\Delta T_m = -2.1$  °C). These findings suggest that these lesions may induce deletion mutations when inserted into repeat sequences that are prone to slippage.<sup>24,28</sup>

MD simulations were carried out on the full-length and truncated duplexes, and adduct conformational preferences were ranked according to the calculated free energies.<sup>24,28</sup> In the full-length duplex with the lesion paired opposite C, C-linked 8arylG adducts favour the major groove (B-type, Fig. 5) conformation, although alternative *syn*-conformations are energetically accessible for FurG, PhG, BThG, QG (all W-type) and PyG (S-type). In the truncated duplexes, FurG, PhG and BThG favour the *anti*-conformation, while QG and PyG favour the *syn*-conformation. For the push-pull QG, its emissive response was consistent with the preferred conformation predicted by the MD simulations (Fig. 6c). Specifically, in the full-length duplex, the *anti*-conformation of QG is favoured over *syn*-structures by at least 25 kJ mol<sup>-1</sup>, which preferentially exposes the quinolyl moiety to the bulk aqueous solvent in the major groove (*anti*-QG:C, Fig. 6c), and quenches CT emission of QG (solid black emission trace, Fig. 6c). In contrast, the *syn*-conformation of QG in the truncated duplex sequesters the quinolyl moiety from the bulk aqueous solvent (*syn*-QG:-2, Fig. 6c) and results in a 12-fold increase in CT emission intensity (dashed red emission trace, Fig. 6c).<sup>24</sup>

In primer elongation assays using the *Narl*(22) templates, C-linked 8arylG adducts strongly block DNA replication by the high-fidelity DNA polymerase Kf<sup>-</sup> following insertion of a single base opposite the lesion, typically the correct base C.<sup>24,28</sup> This observation was consistent with their preference for the B-type structure, as predicted by MD simulations. In single-

nucleotide incorporation assays, the smallest lesion FurG produced the greatest levels of misincorporation (A > G >> T), which correlations with this lesion resulting in the smallest energy difference between the B and W-type adducted DNA conformations.<sup>24</sup> Using Dpo4 as a model translesion polymerase, C-linked 8arylG adducts were found to cause targeted (at the lesion site, i.e. base substitution) and semitargeted (in the vicinity of the lesion site, i.e. deletion) mutations.<sup>24,28</sup> The single-ringed derivatives (FurG and PhG) produced the greatest levels of misincorporation (A and G),<sup>24</sup> while the fused-ringed derivatives (QG, BThG and PyG) strongly blocked extension by Dpo4.<sup>24,28</sup>

Despite the reported reduced flexibility of the C-linked adducts discussed thus far, adducted DNA associated with the fungal carcinogen OTA has a much more complicated conformational profile. Although the genotoxicity of OTA has been debated in the literature, several studies have illustrated that OTA primarily results in a C-linked 8arylG adduct.<sup>35,36</sup> Using MD simulations, a dynamic conformational profile has been revealed for the C-linked OTAG lesion that involves at least two of the B, W and S-type conformations, with the energetically accessible orientations being highly dependent on the lesion site sequence context (Fig. 7), as well as the OTA ionization state.<sup>37</sup> This mixture of conformations correlates with the complicated mutagenic profile associated with OTA exposure, which includes base deletions and double strand breaks.<sup>84</sup> Furthermore, the adoption of the S-type conformation according to MD, coupled with observed double strand breaks, suggests that this bulkier C-linked adduct may inhibit replication.<sup>37</sup> Nevertheless, the genotoxic effects of OTA have been reported in only select tissues (outer stripe of the outer medulla rather than the entire kidney<sup>84</sup>), which suggests that certain conformations may dominate in particular cellular environments. This example highlights the complex interplay that can exist between the conformations adopted and biological outcomes for C-linked 8arylG adducts.



**Fig. 7** Representative MD structures and relative stability ( $\Delta G$ , kJ/mol) of the B (left), W (middle) and S (right) conformers of the monoanionic OTAG adduct incorporated into the (a) G<sub>1</sub>, (b) G<sub>2</sub> or (c) G<sub>3</sub> position in the *Narl*(12) duplex.

Alternative mechanisms for the toxicity of 8arylG adducts may stem from their ability to induce alternative DNA conformations. For example, the Gannett laboratory has examined the impact of 8PhG lesions on Z-DNA formation.<sup>60-62</sup> Formation of Z-DNA *in vivo* has been implicated in mutagenesis by stimulating large-scale gene deletions, translocations, and rearrangements.<sup>57</sup> A major structural requirement for Z-DNA formation is sequence (alternating purine-pyrimidine bases) and the addition of cationic species to screen electrostatic repulsion between adjacent phosphate groups in Z-DNA. It is also known that methylation of cytosine at the 5-carbon position in CpG runs can favour Z-DNA formation.<sup>89</sup> In an initial effort to gauge the impact of an 8PhG adduct on the B-/Z-DNA equilibrium, the simplest member 8PhG along with four *para*-substituted 8RPhG derivatives (R = CH<sub>3</sub>, CH<sub>2</sub>OCH<sub>3</sub>, CH<sub>2</sub>OH and CO<sub>2</sub><sup>-</sup>) were incorporated into d(CGCGCXCXCG)<sub>2</sub> (X = 8PhG or 8RPhG).<sup>60,61</sup> The effect of 8PhG on the B/Z equilibrium was determined using CD by monitoring the salt concentration required to generate a B/Z equilibrium equal to 1. The unmodified duplex required a salt concentration (NaCl) of ~ 3.2 M, while introduction of the 8PhG lesion lowered the salt concentration by a factor of ~3-25, depending on the nature of the *para*-substituent. These results suggest that 8PhG adducts change the position of the B/Z equilibrium by destabilizing the B-form rather than stabilizing the Z-form.<sup>60,61</sup>

Unfortunately, the duplex model previously used to determine the B/Z equilibrium was not ideal because it contained two 8PhG modifications that could potentially provide an additive or synergistic effect to favour the Z-form.<sup>62</sup> Furthermore, it would be highly unlikely that two 8PhG adducts would form on consecutive base steps *in vivo*. Thus, a hairpin sequence (d-(CG)<sub>5</sub>T<sub>4</sub>(CG)<sub>5</sub>) containing an intramolecular (CG)<sub>5</sub> duplex was synthesized in order to incorporate only one 8PhG modification.<sup>62</sup> Under physiological conditions (2 mM MgCl<sub>2</sub>, 10 mM NaCl, 140 mM KCl and 1 mM spermine, 37 °C, pH 7.4) the hairpin preferentially adopted the Z-form structure. This result strengthened the argument that 8PhG lesions favour Z-DNA through destabilization of the B-form, which may have biological significance. Specifically, sequences that can produce Z-DNA are most commonly found in gene promoter regions where Z-DNA can regulate transcription and nucleosome positioning.<sup>57</sup> Therefore, carcinogens that produce 8arylG adducts may overwhelm the normal cellular response to Z-DNA formation and lead to large-scale deletions in mammalian cells.<sup>62</sup>

Other unique DNA topologies that are more frequent in gene promoter regions include GQs.<sup>59</sup> Chromosome ends (telomeres) are capped with kilobase-long runs of the repeating 5'-(TTAGG)<sub>n</sub>-3' sequence.<sup>90</sup> Although most of the telomere resides as a duplex, the 3'-terminal 50-200 nucleotides are single-stranded and can fold into various GQ topologies. As described earlier, GQs are assembled through the sequential stacking of G-tetrads around a metal cation (Fig. 8), with the intervening sequences extruded as single-strand loops. The GQs topologies are classified depending on the orientation of the DNA strands. They can be parallel (all Gs in

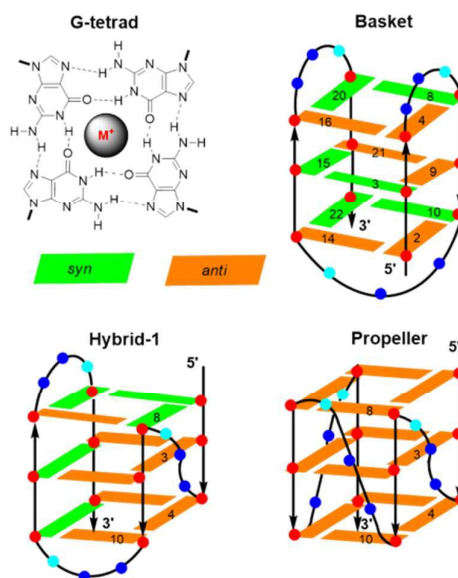


Fig. 8 Structures of a G-tetrad and the basket, hybrid and propeller folds of GQs produced by HTelo22. The 8arylG probe FurG was inserted into positions 3, 4, 8 or 10 of the GQs, *anti*-Gs are shown in brown, *syn*-Gs are shown in green.

the tetrad are in the *anti*-conformation), antiparallel (alternating *syn*- and *anti*-Gs) or hybrids thereof. Model studies utilizing a four-repeat section of the human telomeric DNA sequence (HTelo22, d[AG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>]) have demonstrated extensive GQ polymorphism (Fig. 8).<sup>63</sup> In Na<sup>+</sup> solution, HTelo22 produces a basket-type antiparallel GQ,<sup>58</sup> while in K<sup>+</sup> solution, mixed parallel/antiparallel (hybrid-1 or hybrid-2) GQ structures are the major conformations, with hybrid-1 being the major fold (Fig. 8).<sup>91</sup> In the crystalline state<sup>92</sup> containing K<sup>+</sup> or in K<sup>+</sup> solution containing certain additives (i.e., CH<sub>3</sub>CN, ethanol and polyethylene glycol (PEG)),<sup>93</sup> a propeller-type parallel-stranded GQ structure is favoured. Topologies of HTelo22 formed in K<sup>+</sup> (hybrid or parallel) are expected to be more biologically relevant than the antiparallel structure produced in Na<sup>+</sup>, because GQs have a stronger binding constant with K<sup>+</sup>, which is present in a higher cellular concentration (~ 140 mM [K<sup>+</sup>] versus ~ 10 mM [Na<sup>+</sup>]).<sup>94</sup>

The G-rich nature of the human telomere sequence makes it highly susceptible to electrophilic attack.<sup>94</sup> Hence, model studies have been conducted to determine the impact of the 8oxoG lesion on GQ formation by HTelo.<sup>94,95</sup> The lesion strongly perturbs GQ stability due to the inability of 8oxoG to form Hoogsteen base pairs with G. Placement of 8oxoG in an exterior tetrad forces an antiparallel topology in K<sup>+</sup> solution, while an unstable triplex-like topology is produced with 8oxoG in the middle tetrad.<sup>96</sup> Such studies prompted our laboratory to examine the structural impact of an 8arylG adduct on the GQ polymorphism exhibited by HTelo.<sup>63</sup> Unlike 8oxoG, 8arylG adducts can form Hoogsteen base pairs with G; however, their strong *syn*-preference may perturb or inhibit formation of certain GQ topologies that could impact telomeric function. As a representative 8arylG lesion, FurG was incorporated into various G-tetrad positions (3, 4, 8 or 10, see Fig. 8 for numbering) of HTelo22.<sup>63</sup> On the basis of CD



signatures,  $T_m$  analysis and fluorescence measurements, all FurG-modified HTelo22 sequences adopt the antiparallel fold in  $\text{Na}^+$  solution, with  $T_m$  values higher than the native HTelo22. However, in  $\text{K}^+$  solution, sequences with FurG at positions 3 and 4 exclusively form the antiparallel basket GQ that is produced in  $\text{Na}^+$  solution, while FurG at positions 8 and 10 produces the expected hybrid GQ that is favoured by the native HTelo22 sequence. At all positions examined, the FurG modification strongly impedes formation of the parallel fold that is produced by the native HTelo22 sequence in  $\text{K}^+$  solution in the presence of certain additives. These results demonstrate that production of 8arylG adducts within the human telomeric sequence may make it difficult to form certain GQ topologies, which could impact chromosome stability. Nevertheless, 8arylG adducts can stabilize GQ structures when placed in *syn*-G sites. The ability to stabilize GQ structures has important biological implications. For example, GQ formation can induce replication dependent double strand breaks.<sup>97</sup> Together, the above studies highlight the diverse structural and biological impacts of C-linked 8arylG adducts when incorporated into DNA structures.

### Utility as fluorescent probes

Aptamers are single-stranded DNA, RNA or modified nucleic acids that are designed to fold into unique three-dimensional shapes, which are then recognized by a target molecule.<sup>98,99</sup> Given the stability and diversity of GQs, it is not surprising that many aptamers have been designed based on the GQ structural motif.<sup>100</sup> The antiparallel GQ scaffold provides an opportunity to employ fluorescent 8arylG probes for GQ detection strategies. Compared to other fluorescent guanine analogues, such as pteridines (3-MI and 6-MI)<sup>9</sup> and benzo-expanded guanine derivatives,<sup>11</sup> 8arylG probes are relatively easy to synthesize and have a *syn*-preference making them useful for defining the G conformation within the tetrad; the probe will generally stabilize the GQ when placed in a *syn*-G position, but destabilize the GQ when placed in an *anti*-G position. Furthermore, the G-G base-stacking, hydrogen bonding, and restricted motions within GQ structures can enhance the photoexcited lifetimes and energy transfer properties of a G residue.<sup>38</sup> These factors can amplify the emission of the 8arylG probe within the GQ structure compared to its emission in the duplex. Thus, 8arylG probes can be effective diagnostic tools in duplex-GQ exchange systems that are commonly employed in DNA-based detection strategies.<sup>42,52,53</sup> The level of precision offered by internal fluorescent nucleobase mimics can complement other approaches including end-labelled dyes and label-free dyes, because the internal probes may assist in establishing Gs in the tetrad versus loop positions, site(s) of ligand binding, and remove false negatives or positives that can occur with the use of external dyes that are not covalently attached to the aptamer.<sup>101</sup>

To test 8arylG probe performance in an antiparallel GQ aptamer, FurG and the donor-acceptor pCNPhG were inserted

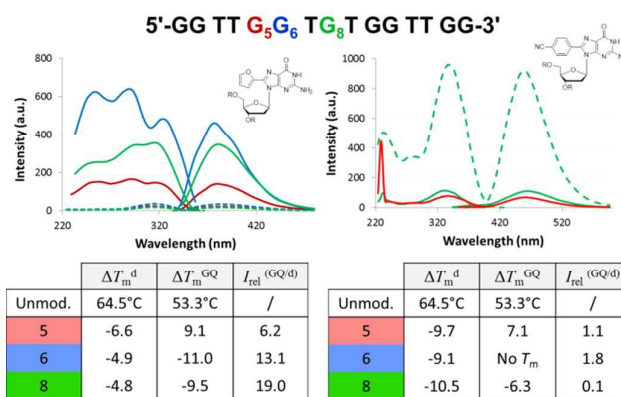
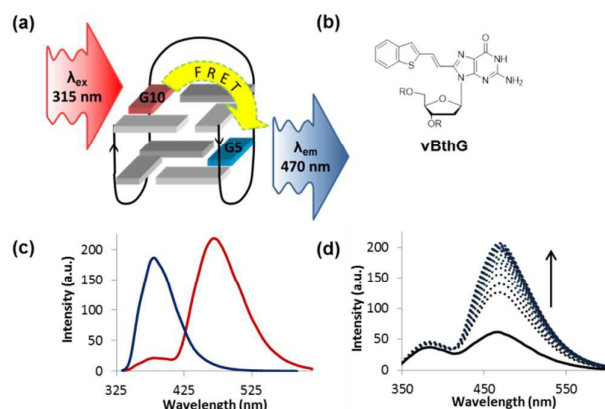


Fig. 9 TBA sequence with G<sub>5</sub>, G<sub>6</sub> and G<sub>8</sub> highlighted in red, blue and green, respectively. Excitation and emission spectra for mTBA with duplexes as dashed traces, GQs solid traces and different colours for the various positions of the 8arylG probe within mTBA. Thermal melting parameters for duplex (d) and GQ in  $\text{K}^+$  solution for mTBA compared to the unmodified sequence.

into various positions within the thrombin binding aptamer (TBA, Fig. 9).<sup>42</sup> The two 8arylG probes were inserted into the 5, 6 or 8 position that represent a *syn*-G position within a G-tetrad (G<sub>5</sub>), an *anti*-G position (G<sub>6</sub>) or a diagonal TGT loop position (G<sub>8</sub>). The duplex and GQ structures produced by the modified-TBA (mTBA) strands were compared to the corresponding structures adopted by native TBA using CD,  $T_m$  analysis, fluorescence and MD simulations.<sup>42</sup>

In the duplex structure, the emission of the FurG probe ( $\lambda_{em} = 384 \text{ nm}$ ,  $\Phi_{fl} = 0.49$ <sup>67</sup>) was strongly quenched (dashed traces, Fig. 9), while the emission exhibited a 6- to 19-fold increase in fluorescence intensity in the chair-like antiparallel GQ (Table, Fig. 9) compared to the duplex emission. The excitation spectra for FurG in the GQ also displays diagnostic energy-transfer bands at  $\sim 255$  and  $290 \text{ nm}$  that are absent in the duplex excitation spectra. The probe had a destabilizing influence on duplex stability ( $\Delta T_m \sim -5 \text{ }^\circ\text{C}$ ) and on GQ stability when placed at *anti*-G<sub>6</sub> and in the diagonal loop G<sub>8</sub>. However, at the *syn*-G<sub>5</sub> position, FurG increases GQ stability by  $\sim 9 \text{ }^\circ\text{C}$ . The bulkier pCNPhG probe had a stronger destabilizing influence on duplex stability and inhibited GQ formation when placed in the *anti*-G<sub>6</sub> position ( $T_m$  value could not be determined). Interestingly, the probe was not as destabilizing as FurG in the GQ when placed in the diagonal G<sub>8</sub> loop position (due to increased stacking with the G-tetrad); the pCNPhG probe also stabilized the GQ at *syn*-G<sub>5</sub> (by  $\sim 7 \text{ }^\circ\text{C}$ ). In contrast to the FurG probe, the push-pull pCNPhG exhibited quenched emission in the GQ structures at positions G<sub>5</sub> and G<sub>8</sub>, and at position G<sub>5</sub> in the duplex. However, in the duplex at position G<sub>8</sub>, where the probe is flanked by T residues, the probe was strongly emissive ( $\sim 10$ -fold increase compared to the GQ structure). Thus, FurG serves as an effective turn-on emissive probe in duplex-GQ exchange, while pCNPhG can be employed as a turn-off probe when inserted into the diagonal G<sub>8</sub> loop position of TBA.<sup>42</sup> Indeed, duplex-GQ exchange studies with FurG at *syn*-G<sub>5</sub> and pCNPhG at G<sub>8</sub> within mTBA have demonstrated the utility of these probes for  $\text{K}^+$  ion<sup>42</sup> and thrombin detection.<sup>52</sup>



**Fig. 10** (a) Schematic for FRET in TBA using FurG as the donor (D,  $\lambda_{\text{ex}} = 315$  nm) paired with an acceptor (A,  $\lambda_{\text{em}} = 470$  nm) in the G-tetrad. (b) Structure of the A probe vBthG. (c) Emission spectra ( $\lambda_{\text{ex}} = 315$  nm) of mTBA in  $\text{K}^+$ -solution with FurG (D, *syn*-G<sub>10</sub>) in the absence (blue trace) and presence (red trace) of vBthG (A, *syn*-G<sub>5</sub>) for a FRET efficiency of 88% in the antiparallel GQ. (d) Thrombin emission titration with mTBA (D/A, 10/5) with  $\lambda_{\text{ex}} = 315$  nm.

The turn-on emissive properties of FurG upon GQ folding suggested that it could serve as a donor (D) dye to be paired with an acceptor (A) G derivative for diagnostic fluorescence resonance energy transfer (FRET) signalling for GQ formation (Fig. 10a).<sup>53</sup> This prompted the synthesis of 8-vinylbenzo[*b*]thienyl-dG (vBthG, Fig. 8b), which has an absorbance maxima at  $\sim 380$  nm and will yield effective spectral overlap with the emission of FurG. The vBthG probe provides visible blue emission at 473 nm ( $\Phi_{\text{fl}} = 0.29$ ), but is not effective by itself for monitoring duplex-GQ exchange because its emission lacks sensitivity to the change in DNA topology.<sup>53,101</sup> However, when paired with FurG within *syn*-G positions of mTBA (i.e. D;A, 10;5), the D base can act as a switch upon GQ formation, turning on the visible fluorescence from the vBthG probe. The FRET efficiency of the D/A pair in the antiparallel GQ was 88% at positions 10;5 (Fig. 8c). The mTBA sample with the probes at *syn*-G<sub>5</sub> (A) and *syn*-G<sub>10</sub> (D) strongly decreased duplex stability (by 20 °C), but increased GQ stability (by  $\sim 9$  °C) and acted as an effective turn-on duplex-GQ exchange system (4-fold increase in emission intensity at  $\sim 470$  nm upon thrombin binding, Fig. 10d).<sup>53</sup> Overall, the ability of 8arylG bases to exhibit emission switching properties upon change in DNA topology (duplex-GQ exchange) provides a basis for their utility in DNA-based diagnostics.

## Conclusions and future prospects

The literature summarized above highlight the structural and *in vitro* mutagenic impact of C-linked 8arylG lesions together with their potential applications as fluorescent probes in duplex-GQ exchange systems. In terms of biological impact, until recently, few studies have focused on the C-linked 8arylG variety despite the fact that many chemical mutagens are known to produce such adducts. Our studies to date demonstrate that C-linked 8arylG lesions are stronger blocks of DNA synthesis than the corresponding N-linked derivatives, and can promote polymerase slippage and misincorporation. Future efforts should examine the conformational preferences

of C-linked 8arylG adducts in both duplex structures and in template:primers bound to DNA polymerases using high-resolution structural studies combined with MD simulations. In order for C-linked 8arylG adducts to induce mutagenicity, they must evade repair by nucleotide excision repair (NER) enzymes. NER of C-linked 8arylG adducts has yet to be examined. Here, it may be possible to take advantage of the fluorescent nature of C-linked 8arylG adducts to determine incision rates mediated by NER. Fluorescence-based assays for nucleotide selection preferences of Y-family polymerases opposite C-linked 8arylG adducts may also be highly insightful, especially in cases where the lesion induces  $-2$  base slippage; C-linked 8arylG adducts with push-pull character can exhibit enhanced emission when sequestered from the aqueous solvent (i.e. within a 2-base bulge). Finally, C-linked 8arylG adducts with defined structures have yet to be incorporated into DNA vectors for the determination of mutational frequency in cell-based assays. Such studies would better permit direct comparisons between the mutagenicity of C-linked and N-linked 8arylG lesions so that the biological impact of linkage type can be established.

In terms of the utility of fluorescent 8arylG probes in duplex-GQ exchange systems, the 8-furyl derivative FurG exhibits favourable turn-on emission switching properties upon change in DNA topology due to energy-transfer from the unmodified Gs in the tetrad that is absent in the duplex structure. Specifically, the base can serve as a donor probe to be paired with an acceptor G for effective FRET signalling of GQ formation. So far, our studies have been limited to TBA, which forms an antiparallel GQ and can be used as “proof-of-principle” to test probe performance. However, many aptamers produce GQs upon target binding, including those for OTA, nucleolin, insulin, HIV-1 reverse transcriptase and ATP.<sup>100</sup> Many C-rich aptamers also exist that can be paired with GQ-producing complementary strands, such as aptamers for microcystins.<sup>102</sup> In this scenario, target binding to the aptamer releases the complementary strand containing the 8arylG probe, which is free to fold into a GQ to signal target binding. We expect 8arylG bases to have commercial applications in aptasensors over the next few years. One goal of our laboratories is to develop 8arylG probes that undergo excitation with visible light and can be readily employed at *anti*-G positions for applications within parallel GQ structures. This issue is particularly challenging because low energy emission is usually associated with large chromophores that may inhibit GQ folding or produce CT states that exhibit quenched emission in  $\text{H}_2\text{O}$  or a lack of emission sensitivity to changes in DNA topology. Nevertheless, the successes registered to date regarding the range of properties and applications of 8arylG bases help ensure that achieving these ambitious goals will push the current boundaries in many nucleic acid-based technologies.

## Acknowledgements

We gratefully acknowledge the Natural Sciences and Engineering Research Council of Canada [Discovery 311600-

2013 to RAM and 249598-07 to S.D.W.]; Canada Research Chair Program [950-228175 to S.D.W.] and the Canadian Foundation of Innovation [10679 to R.A.M. and 22770 to S.D.W.] for their generous support. Computational resources provided by Westgrid and Compute/Calcul Canada are greatly appreciated. Special thanks to Dr. Shana J. Sturla (ETH, Zurich) for helpful discussions and assistance with the biochemical assays. We also gratefully acknowledge the undergraduate, graduate students and postdoctoral research fellows that worked on these projects. Special thanks to Drs Michael Sproviero (UOG, now at New England BioLabs Inc., Ipswich, MA), Purshotam Sharma (UOL, now at Center for Computational Sciences, Central University of Punjab), Katherine M. Rankin (UOG, now at Sheridan College, Oakville, ON), Andrea L. Millen (UOL, now at Exova Canada Inc.), Aaron A. Witham (UOG, now at Environmental Bio-detection Products Inc., Mississauga, ON), Ms. Anne M. R. Verwey (UOG, now at Lab Services, UOG) and Mr. Darian Blanchard (UOG, now at Aterica Health Inc., Waterloo, ON).

## Notes and references

- P. M. E. Gramlich, C. T. Wirges, A. Manetto and T. Carell, *Angew. Chem. Int. Ed.*, 2008, **47**, 8350-8358.
- J. S. Waters, A. Webb, D. Cunningham, P. A. Clarke, F. Raynaud, F. di Stefano and F. E. Cotter, *J. Clin. Oncol.* 2000, **18**, 1812-1823.
- A. M. Mohammed and R. Schulman, *Nano Lett.*, 2013, **13**, 4006-4013.
- T. Wang, D. Schifffels, S. M. Cuesta, D. K. Fyngenson and N. C. Seeman, *J. Am. Chem. Soc.*, 2012, **134**, 1606-1616.
- M. Bachman, S. Uribe-Lewis, X. Yang, H. E. Burgess, M. Lurlaro, W. Reik, A. Murrell and S. Balasubramanian, *Nat. Chem. Biol.* 2015, **11**, 555-557.
- A. H. El-Sagheer and T. Brown, *Chem. Sci.* 2014, **5**, 253-259.
- J. L. Weishar, C. K. McLaughlin, M. Baker, W. Gabryelski and R. A. Manderville, *Org. Lett.* 2008, **10**, 1839-1842.
- A. M. Fleming, J. G. Muller, A. C. Dlouhy and C. J. Burrows, *J. Am. Chem. Soc.* 2012, **134**, 15091-15102.
- L. M. Wilhelmsson, *Q. Rev. Biophys.*, 2010, **43**, 159-183.
- R. W. Sinkeldam, N. J. Greco and Y. Tor, *Chem. Rev.*, 2010, **110**, 2579-2619.
- Y. N. Teo and E. T. Kool, *Chem. Rev.*, 2012, **112**, 4221-4245.
- H. Kaur, B. R. Babu and S. Maiti, *Chem. Rev.* 2007, **107**, 4672-4697.
- I. K. Astakhova and J. Wengel, *Acc. Chem. Res.*, 2014, **47**, 1768-1777.
- N. E. Geacintov and S. Broyde, in *The Chemical Biology of DNA Damage*, eds. N. E. Geacintov and S. Broyde, Wiley-VCH Verlag GmbH & Co., Weinheim, Germany, 2010, pp. 3-15.
- M. Kouchakdjian, V. Bodepudi, S. Shibutani, M. Eisenberg, F. Johnson, A. P. Grollman and D. J. Patel, *Biochemistry*, 1991, **30**, 1403-1412.
- S. Shibutani, M. Takeshita and A. P. Grollman, *Nature*, 1991, **349**, 431-434.
- S. Steenken, S. V. Jovanovic, M. Bietti and K. Bernhard, *J. Am. Chem. Soc.*, 2000, **122**, 2373-2374.
- S. Steenken and S. V. Jovanovic, *J. Am. Chem. Soc.*, 1997, **119**, 617-618.
- J. Joseph and G. B. Schuster, *J. Am. Chem. Soc.*, 2006, **128**, 6070-6074.
- P. A. Sontz, N. B. Muren and J. K. Barton, *Acc. Chem. Res.*, 2012, **45**, 1792-1800.
- H. H. Thorp and J. M. Sullivan, in *Radical and Radical Ion Reactivity in Nucleic Acid Chemistry*, ed. M. M. Greenberg, John Wiley & Sons Inc., Hoboken, New Jersey, 2009, pp. 357-387.
- M. S. Kuska, M. Majdi Yazdi, A. A. Witham, H. A. Dahlmann, S. J. Sturla, S. D. Wetmore and R. A. Manderville, *J. Org. Chem.*, 2013, **78**, 7176-7185.
- M. S. Kuska, A. A. Witham, M. Sproviero, R. A. Manderville, M. Majdi Yazdi, P. Sharma and S. D. Wetmore, *Chem. Res. Toxicol.*, 2013, **26**, 1397-1408.
- M. Sproviero, A. M. R. Verwey, K. M. Rankin, A. A. Witham, D. V. Soldatov, R. A. Manderville, M. I. Fekry, S. J. Sturla, P. Sharma and S. D. Wetmore, *Nucleic Acids Res.*, 2014, **42**, 13405-13421.
- A. A. Witham, P. Sharma, S. D. Wetmore, W. Gabryelski and R. A. Manderville, *J. Mass Spectrom.*, 2015, **50**, 81-87.
- P. Sharma, M. Majdi Yazdi, A. Merriman, R. A. Manderville and S. D. Wetmore, *Chem. Res. Toxicol.*, 2015, **28**, 782-796.
- A. A. Witham, A. M. R. Verwey, M. Sproviero, R. A. Manderville, P. Sharma and S. D. Wetmore, *Chem. Res. Toxicol.*, 2015, **28**, 1346-1318.
- M. Sproviero, A. M. R. Verwey, A. A. Witham, R. A. Manderville, P. Sharma and S. D. Wetmore, *Chem. Res. Toxicol.*, 2015, **28**, 1647-1658.
- B. Cho, in *The Chemical Biology of DNA Damage*, eds. N. E. Geacintov and S. Broyde, Wiley-VCH Verlag GmbH & Co., Weinheim, Germany, 2010, pp. 217-238.
- R. J. Turesky and L. L. Marchand, *Chem. Res. Toxicol.*, 2011, **24**, 1169-1214.
- G. Nauwelaërs, M. Bellamri, V. Fessard, R. J. Turesky and S. Langouët, *Chem. Res. Toxicol.*, 2013, **26**, 1367-1377.
- O. Rechkoblit, A. Kolbanovskiy, L. Malinina, N. E. Geacintov, S. Broyde, and D. J. Patel, *Nat. Struct. Biol.*, 2010, **17**, 379-389.
- H. Mu, K. Kropachev, L. Wang, L. Zhang, A. Kolbanovskiy, M. Kolbanovskiy, N. E. Geacintov, and S. Broyde, *Nucleic Acids Res.*, 2012, **40**, 9675-9690.
- V. G. Vaidyanathan, F. Liang, W. A. Beard, D. D. Shock, S. H. Wilson and B. P. Cho, *J. Biol. Chem.*, 2013, **288**, 23573-23585.
- J. Dai, M. W. Wright and R. A. Manderville, *J. Am. Chem. Soc.*, 2003, **125**, 3716-3717.
- P. G. Mantle, V. Faucet-Marquis, R. A. Manderville, B. Squillaci and A. Pfohl-Leszakowicz, *Chem. Res. Toxicol.*, 2010, **23**, 89-98.
- P. Sharma, R. A. Manderville and S. D. Wetmore, *Nucleic Acids Res.*, 2014, **42**, 11831-11845.
- A. Dumas and N. W. Luedtke, *J. Am. Chem. Soc.*, 2010, **132**, 18004-18007.
- K. Börhesson, S. Preus, A. El-Sagheer, T. Brown, B. Albinsson and L. M. Wilhelmsson, *J. Am. Chem. Soc.* 2009, **131**, 4288-4293.
- A. Nadler, J. Strohmeier and U. Diederichsen, *Angew. Chem. Int. Ed.*, 2011, **50**, 5392-5396.
- S. Park, H. Otomo, L. Zheng and H. Sugiyama, *Chem. Commun.*, 2014, **50**, 1573-1575.
- M. Sproviero, K. L. Fadock, A. A. Witham, R. A. Manderville, P. Sharma and S. D. Wetmore, *Chem. Sci.*, 2014, **5**, 788-796.
- G. Mata and N. W. Luedtke, *J. Am. Chem. Soc.* 2015, **137**, 699-707.
- P. Cekan and S. T. Sigurdsson, *Chem. Commun.*, 2008, 3393-3395.

- 45 N. J. Greco, R. W. Sinkeldam and Y. Tor, *Org. Lett.*, 2009, **11**, 115-1118.
- 46 Y. Sato, M. Kudo, Y. Toriyabe, S. Kuchitsu, C.-X. Wang, S. Nishizawa and N. Teramae, *Chem. Commun.*, 2014, **50**, 515-517.
- 47 S. G. Srivatsan, N. J. Greco and Y. Tor, *Angew. Chem. Int. Ed.*, 2008, **47**, 6661-6665.
- 48 N. Dai and E. T. Kool, *Chem. Soc. Rev.*, 2011, **40**, 5756-5770.
- 49 C. Y. Lee, K. S. Park and H. G. Park, *Chem. Commun.*, 2015, **51**, 13744-13747.
- 50 D. Dziuba, R. Pohl and M. Hocek, *Chem. Commun.*, 2015, **51**, 4880-4882.
- 51 T. Li, R. Fu and H. G. Park, *Chem. Commun.*, 2010, **46**, 3271-3273.
- 52 M. Sproviero and R. A. Manderville, *Chem. Commun.*, 2014, **50**, 3097-3099.
- 53 D. J. M. Blanchard, T. Z. Cservenyi and R. A. Manderville, *Chem. Commun.*, 2015, **51**, 16829-16831.
- 54 G. F. Deleavey and M. J. Damha, *Chem. & Biol. Rev.*, 2012, **19**, 937-954.
- 55 A. L. Millen, C. K. McLaughlin, K. M. Sun, R. A. Manderville and S. D. Wetmore, *J. Phys. Chem. A*, 2008, **112**, 3742-3753.
- 56 A. H. J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel and A. Rich, *Nature*, 1979, **282**, 680-686.
- 57 G. Wang, L. A. Christensen and K. M. Vasquez, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, **103**, 2677-262.
- 58 Y. Wang and D. J. Patel, *Structure*, 1993, **1**, 263-282.
- 59 G. Biffi, D. Tannahill, J. McCafferty and S. Balasubramanian, *Nat. Chem.*, 2013, **5**, 182-186.
- 60 P. M. Gannett, S. Heavner, J. R. Daft, K. H. Shaughnessy, J. D. Epperson and N. L. Greenbaum, *Chem. Res. Toxicol.*, 2003, **16**, 1385-1394.
- 61 V. Vongsutilers, D. J. Phillips, B. C. Train, G. R. McKelvey, N. M. Thomsen, K. H. Shaughnessy, J. P. Lewis and P. M. Gannett, *Biophys. Chem.*, 2011, **154**, 41-48.
- 62 B. C. Train, S. A. Bilgesü, E. C. Despeaux, V. Vongsutilers and P. M. Gannett, *Chem. Res. Toxicol.* 2014, **27**, 1176-1186.
- 63 M. Sproviero, K. L. Fadock, A. A. Witham and R. A. Manderville, *ACS Chem. Biol.*, 2015, **10**, 1311-1318.
- 64 E. C. Western, J. R. Daft, E. M. Johnson II, P. M. Gannett and K. H. Shaughnessy, *J. Org. Chem.*, 2003, **68**, 6767-6774.
- 65 A. L. Millen, R. A. Manderville and S. D. Wetmore, *J. Phys. Chem. B*, 2010, **114**, 4373-4382.
- 66 A. L. Millen, C. D. M. Churchill, R. A. Manderville and S. D. Wetmore, *J. Phys. Chem. B*, 2010, **114**, 12995-13004.
- 67 K. M. Rankin, M. Sproviero, K. Rankin, P. Sharma, S. D. Wetmore and R. A. Manderville, *J. Org. Chem.*, 2012, **77**, 10498-10508.
- 68 A. Dumas and N. W. Luedtke, *Nucleic Acids Res.*, 2011, **39**, 6825-6834.
- 69 J. L. Sessler, M. Sathiosatham, K. Doerr, V. Lynch and K. A. Abboud, *Angew. Chem. Int. Ed.*, 2000, **39**, 1300-1303.
- 70 K. M. Schlitt, K. M. Sun, R. J. Paugh, A. L. Millen, L. Navarro-Whyte, S. D. Wetmore and R. A. Manderville, *J. Org. Chem.*, 2009, **74**, 5793-5802.
- 71 McLaughlin, D. R. Lantero and R. A. Manderville, *J. Phys. Chem. A*, 2006, **110**, 6224-6230.
- 72 A. Omumi, A. L. Millen, S. D. Wetmore and R. A. Manderville, *Chem. Res. Toxicol.*, 2011, **24**, 1694-1709.
- 73 A. Omumi, C. K. McLaughlin, D. Ben-Israel and R. A. Manderville, *J. Phys. Chem. B*, 2012, **116**, 6158-6165.
- 74 A. Dumas and N. W. Luedtke, *Chem.–Eur. J.*, 2012, **18**, 245-254.
- 75 K. M. Schlitt, A. L. Millen, S. D. Wetmore and R. A. Manderville, *Org. Biomol. Chem.*, 2011, **24**, 1694-1709.
- 76 K. M. Sun, C. K. McLaughlin, D. R. Lantero and R. A. Manderville, *J. Am. Chem. Soc.*, 2007, **129**, 1894-1895.
- 77 C. P. Da Costa and H. Sigel, *Inorg. Chem.*, 2003, **42**, 3475-3482.
- 78 F. P. Guengerich, *Chem Rev.*, 2006, **106**, 420-452.
- 79 P. Čapek, R. Pohl and M. Hocek, *Org. Biomol. Chem.*, 2006, **4**, 2278-2284.
- 80 A. Omumi, D. G. Beach, M. Baker, W. Gabryelski and R. A. Manderville, *J. Am. Chem. Soc.*, 2011, **133**, 42-50.
- 81 M. Sproviero, K. M. Rankin, A. A. Witham and R. A. Manderville, *J. Org. Chem.*, 2014, **79**, 692-699.
- 82 K. Tram, Y. S. Sanghvi and H. Yan, *Nucleosides, Nucleotides Nucleic Acids*, 2011, **30**, 12-19.
- 83 S. Sen, P. Bhojnagarwala, L. Francey, D. Lu, T. M. Penning and J. Field, *Chem. Res. Toxicol.*, 2012, **25**, 2117-2126.
- 84 D. Hibi, A. Kijima, Y. Suzuki, Y. Ishii, M. Jin, Y. Sugita-Konishi, T. Yanai, A. Nishikawa and T. Umemura, *Toxicology*, 2013, **304**, 92-99.
- 85 G. R. Hoffmann and R. P. P. Fuchs, *Chem. Res. Toxicol.*, 1997, **10**, 347-359.
- 86 S. Broyde, L. Wang, O. Rechkoblit, N. E. Geacintov and D. J. Patel, *Trends Biochem. Sci.*, 2008, **33**, 209-219.
- 87 C. E. Elmquist, J. S. Stover, Z. Wang and C. J. Rizzo, *J. Am. Chem. Soc.*, 2004, **126**, 11189-11201.
- 88 X. Tan, N. Suzuki, A. P. Grollman and S. Shibutani, *Biochemistry*, 2002, **41**, 14255-14262.
- 89 N. A. Temiz, D. E. Donohue, A. Bacolla, B. T. Luke and J. R. Collins, *PLoS one*, 2012, **7**, e35558.
- 90 R. E. Verdun and J. Karlseder, *Nature*, 2007, **447**, 924-931.
- 91 A. Ambrus, D. Chen, J. Dai, T. Bialis, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2006, **34**, 2723-2735.
- 92 G. N. Parkinson, M. P. Lee and S. Neidle, *Nature*, 2002, **417**, 876-880.
- 93 M. C. Miller, R. Buscaglia, J. B. Chaires, A. N. Lane and J. O. Trent, *J. Am. Chem. Soc.*, 2010, **132**, 17105-17107.
- 94 A. M. Fleming and C. J. Burrows, *Chem. Res. Toxicol.*, 2013, **26**, 593-607.
- 95 V. A. Szalai, M. J. Singer and H. H. Thorp, *J. Am. Chem. Soc.*, 2002, **124**, 1625-1631.
- 96 J. Zhou, A. M. Fleming, A. M. Averill, C. J. Burrows and S. S. Wallace, *Nucleic Acids Res.*, 2015, **43**, 4039-4054.
- 97 P. Murat and S. Balasubramanian, *Curr. Opin. Genet. & Develop.*, 2014, **25**, 22-29.
- 98 A. D. Ellington and J. W. Szostak, *Nature*, 1990, **346**, 818-822.
- 99 C. Tuerk and L. Gold, *Science*, 1990, **249**, 505-510.
- 100 W. O. Tucker, K. T. Shum and J. A. Tanner, *Curr. Pharm. Des.*, 2012, **18**, 2014-2026.
- 101 D. J. M. Blanchard, K. L. Fadock, M. Sproviero, P. Deore, T. Z. Cservenyi, R. A. Manderville, P. Sharma and S. D. Wetmore, *J. Mater. Chem. C*, 2016, DOI:10.1039/c5tc03354c.
- 102 A. Ng, R. Chinnappan, S. Eissa, H. Liu, C. Tlili and M. Zourob, *Environ. Sci. Technol.*, 2012, **46**, 10697-10703.