Organic & Biomolecular Chemistry

Cite this: *Org. Biomol. Chem.*, 2012, **10**, 7078

www.rsc.org/obc

PAPER

O^4-Alkyl-2′-deoxythymidine cross-linked DNA to probe recognition and repair by O^6-alkylguanine DNA alkyltransferases†‡

Francis P. McManus, Derek K. O'Flaherty, Anne M. Noronha and Christopher J. Wilds*

Received 10th April 2012, Accepted 3rd July 2012
DOI: 10.1039/c2ob25705j

DNA duplexes containing a directly opposed O^4-2′-deoxythymidine-alkyl-O^4-2′-deoxythymidine (O^4-dT-alkyl-O^4-dT) interstrand cross-link (ICL) have been prepared by the synthesis of cross-linked nucleoside dimers which were converted to phosphoramidites to produce site specific ICL. ICL duplexes containing alkyl chains of four and seven methylene groups were prepared and characterized by mass spectrometry and nuclease digests. Thermal denaturation experiments revealed four and seven methylene containing ICL increased the T_m of the duplex with respect to the non-cross-linked control with an observed decrease in enthalpy based on thermodynamic analysis of the denaturation curves. Circular dichroism experiments on the ICL duplexes indicated minimal difference from B-form DNA structure. These ICL were used for DNA repair studies with O^6-alkylguanine DNA alkyltransferase (AGT) proteins from human (hAGT) and *E. coli* (Ada-C and OGT), whose purpose is to remove O^6-alkylguanine and in some cases O^4-alkylthymine lesions. It has been previously shown that hAGT can repair O^6-2′-deoxyguanosine-alkyl-O^6-2′-deoxyguanosine ICL. The O^4-dT-alkyl-O^4-dT ICL prepared in this study were found to evade repair by hAGT, OGT and Ada-C. Electromobility shift assay (EMSA) results indicated that the absence of any repair by hAGT was not a result of binding. OGT was the only AGT to show activity in the repair of oligonucleotides containing the mono-adducts O^4-butyl-4-ol-2′-deoxythymidine and O^4-heptyl-7-ol-2′-deoxythymidine. Binding experiments conducted with hAGT demonstrated that the protein bound O^4-alkylthymine lesions with similar affinities to O^6-methylguanine, which hAGT repairs efficiently, suggesting the lack of O^4-alkylthymine repair by hAGT is not a function of recognition.

Introduction

Interference by ICLs of critical cell events involving DNA unwinding is exploited in cancer chemotherapy regimens employing bis-alkylating agents.[1] The potency of these agents may be diminished by the ability of cancer cells to repair the very lesions they induce. Numerous DNA repair pathways including direct, base- and nucleotide-excision repair (NER), homologous recombination (HR), non-homologous end joining and DNA-mismatch repair remove various DNA lesions.[2] Some pathways, such as NER, are complex with broad substrate specificities whereas direct repair by AGT, which involves one protein, has a narrow range of damage that it repairs. ICL damage is challenging to repair as information on both DNA strands is affected. Removal of the ICL on one strand of DNA leaves behind damage on the opposing strand, complicating error-free repair from a template strand. Repair pathways such as NER, HR and translesion DNA synthesis (TLS) have all been implicated in ICL repair in mammalian cells, however, there is an increasing realization that ICL processing may depend on the nature of the lesion making it challenging to generalize how specific ICLs are repaired.[3]

An approach pursued by a number of groups to enhance our understanding of these processes involves the preparation of ICL DNA substrates by solution and solid-phase synthesis for repair studies.[4] A number of elegant examples have been reported including the preparation of various phosphoramidites of nucleosides and other molecules to introduce site-specific lesions in DNA for various experiments.[5] Repair and binding experiments can be conducted by incorporating the oligonucleotides containing lesions into specific plasmids.[6]

Due to the instability of certain ICL formed in DNA as a consequence of DNA treatment with bifunctional alkylating agents it is at times necessary to modify the structure of the ICL by producing a mimic similar in structure but exhibiting improved stability to enable biophysical and repair studies. For example,

Department of Chemistry and Biochemistry Concordia University, 7141 Sherbrooke Street West, Montreal, Canada.
E-mail: cwilds@alcor.concordia.ca; Fax: +1-514-848-2424 ext.5798

† In memory of Professor Har Gobind Khorana (1922–2011), acknowledging his legacy to the scientific community.

‡ Electronic supplementary information (ESI) available: ^{1}H and ^{31}P NMR Spectra, ESI-MS Spectra of DNA, C-18 RP-HPLC traces of phosphodiesterase digests of DNA, T_m and CD data of mono-adduct DNA, DNA Repair gels and results by AGTs and EMSA binding data. See DOI: 10.1039/c2ob25705j

the alkyl sulfamate hepsulfam and the nitrogen mustard mechlorethamine both form similar ICL between two N7 atoms of guanine in DNA containing a 5′-GNC.[7] Alkylation at the N7 atom of guanine destabilizes the nucleoside. For example, the half-life of the mechlorethamine induced ICL is of the order of hours, which can limit experiments that may be conducted.[8]

Our lab recently reported the preparation of a chemically stable O^6-2′-deoxyguanosine-alkyl-O^6-2′-deoxyguanosine (O^6-dG-alkyl-O^6-dG) ICL *via* a versatile synthesis strategy that enables introduction of an alkyl lesion of any desired length and orientation.[9] Repair studies of the O^6-dG-alkyl-O^6-dG ICL with AGT proteins were conducted. AGT proteins, which are found in all kingdoms of life, are responsible predominantly for the repair of O^6-methyl-2′-deoxyguanosine (O^6MeG) lesions, which are renowned for causing point mutations by disrupting normal Watson–Crick base pairing proving that this protein is integral in maintaining genomic integrity.[10] hAGT, the most thoroughly characterized AGT protein, repairs alkylated DNA by flipping the damaged base out of the helix and into its active site where the alkyl group is transferred from the point of lesion to the active site Cys145 residue.[11] Once alkylated, this protein is degraded by the ubiquitin pathway.[12]

It was found that hAGT can remove an O^6-dG-alkyl-O^6-dG ICL in mismatched[13] and 5′-GNC sequence motif DNA, the latter designed to mimic the lesion formed by hepsulfam.[14] We reported that hAGT reduces hepsulfam cytotoxicity in experiments with CHO cells, suggesting that this drug may form O^6-dG-alkyl-O^6-dG ICL given that hAGT repairs O^6-alkylated-guanine. Interestingly, neither AGT homologues from *E. coli* (Ada-C or OGT) were capable of repairing these ICL.

DNA mismatches can arise from heteroduplexes formed during homologous recombination, errors in DNA replication, deamination and base damage by alkylating agents.[15] The exocyclic atoms of mispaired nucleobases are more susceptible to alkylation as a result of altered hydrogen bonding patterns that expose these atoms to the external environment, as observed by the enhanced reactivity of mispaired thymidine residues to osmium tetroxide.[16] Moreover, formation of mutagenic alkylated nucleobases, such as O^4MeT, can cause mispairs when encountered by DNA polymerases to form, for example, an O^4MeT:dG mismatch.[17]

Formation of pyrimidine–pyrimidine ICL introduced by bifunctional alkylating agents have been reported. DNA containing mispaired dC residues treated with mechlorethamine has been shown to form ICL linking N^3 atoms.[18] This recent discovery of ICL formation involving two mispaired pyrimidines is unique and unprecedented. Not much is known about the properties of these ICL because of their recent appearance and is an area of research that requires further investigation.

Based on a straightforward synthetic methodology and the observation that directly opposed O^6-dG-alkyl-O^6-dG ICL could be repaired by hAGT, we set out to prepare O^4-2′-deoxythymidine-alkyl-O^4-2′-deoxythymidine (O^4-dT-alkyl-O^4-dT) ICL in a directly opposed motif as well as their respective mono-adducts (Fig. 1) by adapting methods described by Swann's group requiring the preparation of convertible nucleosides.[19] These ICL were prepared by solid-phase oligonucleotide synthesis and the influence of various alkyl linker lengths on duplex stability and structure, as well as recognition and repair by AGTs from human and *E. coli*, were explored.

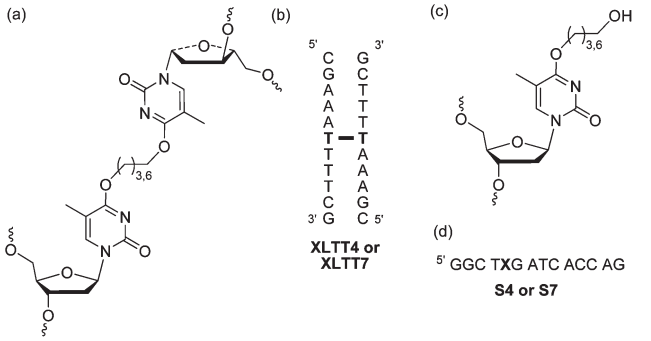


Fig. 1 Structures of (a) O^4-2′-deoxythymidine-alkyl-O^4-2′-deoxythymidine interstrand cross-link, (b) duplex where T–T is the ICL, (c) O^4-alkyl-2′-deoxythymidine mono-adducts and (d) DNA sequence where X contains the mono-adduct.

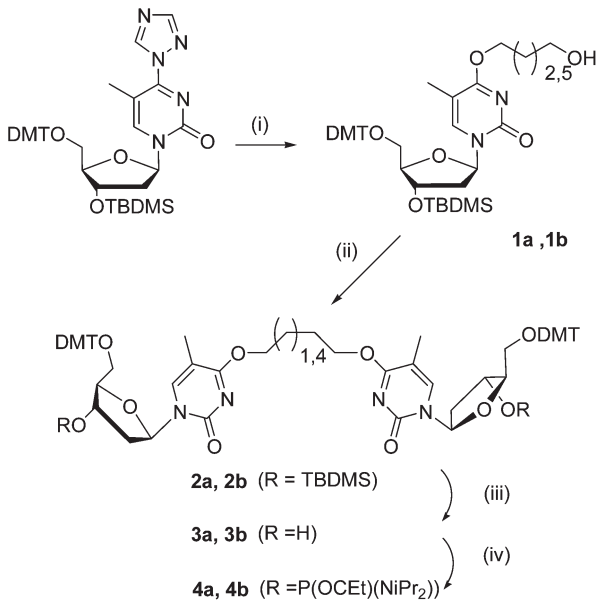
The formation of O^4-methyl-2′-deoxythymidine (O^4MeT) in DNA is highly mutagenic due to the preferential incorporation of dG opposite this alkylated base during replication by DNA polymerases, as indicated previously.[17] O^4MeT is a minor alkylation product with respect to its mutagenic O^6MeG counterpart. DNA treated with *N*-methyl-*N*-nitrosourea produces 126 times more O^6MeG than O^4MeT.[20] Though found in relatively low abundance this lesion is important as it has been observed in the DNA of smokers and is partly responsible for the adverse effects of tobacco.[21]

In mammalian cells, O^4MeT is more toxic than O^6MeG both in normal and repair deficient systems.[22] Unlike O^6MeG, O^4MeT is not affected by the expression or suppression of hAGT activity nor by mismatch repair (MMR) explaining its added toxicity. NER deficient cells exhibit increased sensitivity to O^4MeT indicating that this lesion is possibly repaired *via* the NER pathway in mammalian cells.[23] *E. coli* has evolved two proteins for the direct repair pathway involving the repair of O^4MeT lesions, OGT and the C-terminal of the adaptive response protein Ada, highlighting the importance of repairing this lesion *in vivo*, underscoring an interesting difference in substrate preference between the human and *E. coli* proteins.[24]

Results and discussion

Chemical synthesis of modified nucleosides

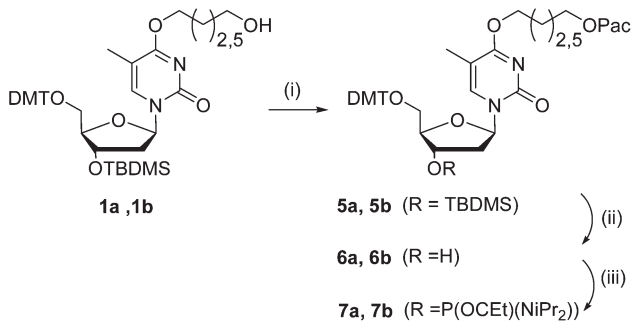
The structures of the O^4-dT-alkyl-O^4-dT cross-links and the mono-adducts are shown in Fig. 1 in addition to the sequences of the respective cross-links (**XLTT4** and **XLTT7**) and single stranded oligonucleotides (**S4** and **S7**). The synthesis of the O^4-dT-alkyl-O^4-dT cross-linked bis-phosphoramidites **4a** and **4b**, containing a four or seven carbon linker, is shown in Scheme 1. The convertible nucleoside 3′-*O*-(*tert*-butyldimethylsilyl)-5′-*O*-(4,4′-dimethoxytrityl)-4-triazolyl-2′-deoxythymidine was transformed into adducts **1a** or **1b** with the sodium salt of 1,4-butanediol or 1,7-heptanediol, respectively.[19] Reacting either **1a** or **1b** with the convertible nucleoside yielded the fully protected dimers **2a** and **2b**. The 3′-OH groups of these compounds, protected as silyl ethers, were liberated by treatment with tetrabutylammonium fluoride to produce compounds **3a** and **3b**,



Scheme 1 Reagents and conditions: (i) sodium salt of 1,4-butanediol (1a) or 1,7-heptanediol (1b), dioxane, (ii) 3′-*O*-(*tert*-butyldimethylsilyl)-5′-*O*-(4,4′-dimethoxytrityl)-4-triazolyl-2′-thymidine, 1,8-diazabicyclo-[5.4.0]undec-7-ene, pyridine, (iii) tetrabutylammonium fluoride (1 M in THF), THF, (iv) *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride, diisopropylethylamine, THF.

which were then phosphitylated with *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride in the presence of diisopropylethylamine to produce bis-phosphoramidites **4a** and **4b**. ^{31}P NMR signals at 148.07 and 148.28 ppm for **4a** and 148.07 and 148.28 ppm for **4b** confirmed the presence of the phosphoramidite functionality in these molecules (see ESI‡).

The mono-adduct lesions were introduced at the O^4 atom of dT as shown in Scheme 2. Starting with compound **1a** or **1b**, the free terminal alcohol functionality of the adduct was protected with phenoxyacetyl chloride to yield **5a** or **5b**. The 3′-OH groups of these nucleosides were deprotected prior to the phosphitylation step by treatment with tetrabutylammonium fluoride to yield **6a** or **6b**. They were then converted to phosphoramidites as confirmed by the presence of two diagnostic signals in ^{31}P NMR at 148.08 and 148.28 ppm for **7a** and 148.08 and 148.28 ppm for **7b** (see ESI‡).



Scheme 2 Reagents and conditions: (i) phenoxyacetyl chloride, triethylamine, THF, (ii) tetrabutylammonium fluoride (1 M in THF), THF, (iii) *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride, diisopropylethylamine, THF.

Synthesis of modified and ICL DNA

The solid-phase syntheses of cross-linked duplexes **XLTT4** and **XLTT7** were performed on a 1 μmol scale using either bis-phosphoramidite **4a** or **4b** to prepare the oligomers containing the four or seven carbon cross-link, respectively. Given the lability of the cross-link, 'fast-deprotecting' 3′-*O*-2′-deoxyphosphoramidites were employed (0.1 M in acetonitrile) with phenoxyacetic anhydride as the capping agent to prevent an undesired *N*-acetylation reaction that could occur if acetic anhydride was used.[25] Moreover, the directly opposed motif allows for ICL duplex synthesis without the need to remove additional protecting groups around the cross-linked site which could compromise the ICL or the use of 5′-*O*-2′-deoxyphosphoramidites. The concentration of 3′-*O*-bis-phosphoramidite **4a** or **4b** in acetonitrile was 0.05 M, and each were allowed to react for an extended period (30 min) to ensure that both phosphoramidite functionalities would couple to two oligonucleotide chains on the solid-support, thus enabling ICL formation.[26] Solid-phase assembly of the duplex was completed by extension with 3′-*O*-2′-deoxyphosphoramidites at the cross-linked site.

A variation to the standard oligonucleotide deprotection protocol was required due to the labile nature of O^4-alkylthymine adducts. These milder deprotection conditions utilized a solution of 10% 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) in anhydrous ethanol for 48 h at room temperature to reduce the likelihood of ICL cleavage linking the O^4 atoms upon deprotection.[19] The crude deprotection mixture containing duplexes of **XLTT4** and **XLTT7** were purified by polyacrylamide gel electrophoresis, rather than SAX HPLC, due to difficulties in removing excess DBU used in the deprotection step. The purified oligonucleotides were eluted from the gels and desalted. ESI mass spectrometry analysis of **XLTT4** and **XLTT7** revealed they had molecular weights of 6715.6 and 6758.3 Da (expected 6716.5 and 6758.6 Da), consistent with the expected values (see ESI‡). The composition of the cross-linked oligomers **XLTT4** and **XLTT7** was further confirmed by digestion to their constituent nucleosides with a mixture of snake venom phosphodiesterase and calf intestinal phosphatase followed by C-18 reversed phase HPLC analysis (see ESI‡). The presence of cross-linked nucleosides was established by the appearance of one additional peak, other than the four standard nucleosides, that eluted from the HPLC with retention times of 21.7 for the four carbon and 29.9 min for the seven carbon O^4-dT-alkyl-O^4-dT cross-link. The later elution of the seven carbon cross-link can be attributed to the greater hydrophobicity of the longer alkyl linker. The ratios extracted from the chromatogram of the component 2′-deoxynucleosides and cross-linked nucleosides were in good agreement with the theoretical compositions (see ESI‡).

For the mono-adduct lesions, O^4-butyl-4-ol-dT and O^4-heptyl-7-ol-dT, introduced in sequences **S4** and **S7** (see Fig. 1), similar solid-phase synthesis conditions as described for the ICL substrates were employed. For these oligonucleotides, cleavage and deprotection conditions involved incubation of the CPG (controlled pore glass) with a 10% solution of DBU in 1,4-butanediol

(for **S4**) or 1,7-heptanediol (for **S7**) for 5 days at room temperature to ensure that the adducts were not compromised.[19] The DBU was neutralized with acetic acid before the CPG was rinsed with acetonitrile (50% in water) to solubilize the fully deprotected product as a result of the poor solubility of DNA in the diols. These oligonucleotides were also characterized by ESI mass spectrometry and nuclease digestion to ensure their composition (see ESI‡).

UV thermal denaturation studies of DNA duplexes

The effect of the directly opposed O^4-dT-alkyl-O^4-dT ICL on duplex stability was assessed through UV thermal denaturation experiments. The melting profiles of **XLTT4** and **XLTT7**, shown in Fig. 2, exhibit a sigmoidal denaturation and hyperchromicities comparable to the control duplex containing a T–A base pair instead of the cross-link. The melting temperatures observed for the cross-linked oligonucleotides **XLTT4** and **XLTT7** were 66 and 50 °C, respectively, higher than the control duplex (with a T_m of 44 °C).

This melting temperature trend was observed earlier by our group for duplexes containing a directly opposed O^6-2′-deoxyguanosine-alkyl-O^6-2′-deoxyguanosine mismatch present in an 11 bp duplex with increases in T_m of 23 and 33 °C for the heptyl and butyl linkages, respectively, over the control duplex.[27] This same tendency has also been reported when ICL of various lengths are introduced between the N^3 atoms of thymidine where an increase in T_m of 16 and 31 °C were observed for heptyl and butyl linkages over the non-alkylated control.[28]

Thermodynamic parameters of the ICL and control duplexes were studied to investigate the origin of the increased thermal stability observed for the ICL DNA (Table 1). Varying duplex concentration did not affect the T_m observed for the ICL duplexes but did influence that of the control, as expected. The thermodynamic parameters indicate a reduction in the enthalpy for the ICL containing duplex, which was more pronounced for the heptyl *versus* butyl linker. One possible reason may be due to the disruption of the base pairs around the cross-linked site as a result of steric effects of the larger linker.

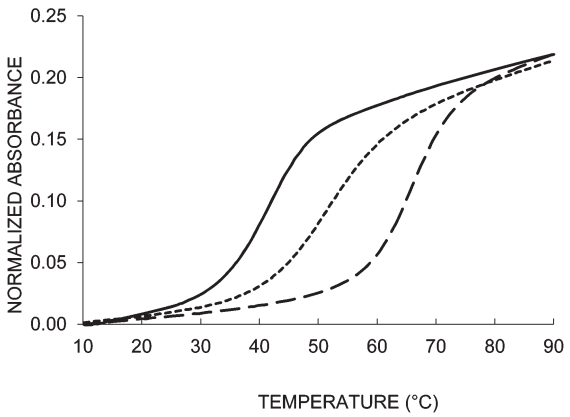


Fig. 2 Absorbance (A_{260}) *versus* temperature profiles of cross-linked duplexes **XLTT4** (— —), **XLTT7** (- - -) and non-cross-linked DNA (—).

Table 1 Thermodynamic parameters of DNA duplexes

DNA	$\Delta H°$ (kJ mol^{-1})	$\Delta S°$ (J mol^{-1} K^{-1})
Control (T–A)	−298 ± 23	−823 ± 71
XLTT4	−249 ± 7	−735 ± 20
XLTT7	−179 ± 4	−551 ± 12

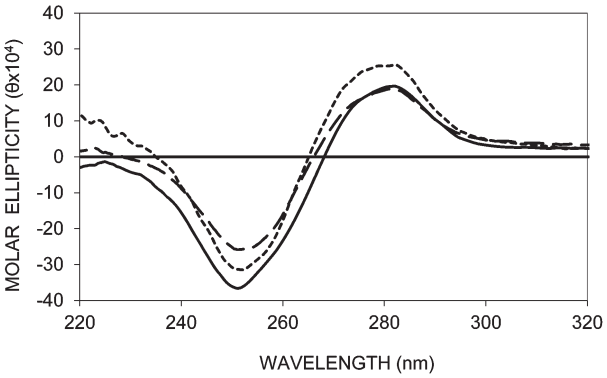


Fig. 3 Far-UV circular dichroism spectra of cross-linked duplexes **XLTT4** (— —), **XLTT7** (- - -) and non-cross-linked DNA (—).

The presence of the ICL caused an increase in the entropy of the duplex due to the preorganized nature of the unimolecular system thus favouring thermal stability of the DNA. The contribution of the increased entropy outweighed the negative enthalpic effect in the ICL system, consistent with what has been previously observed with ICL formed by *trans*-diamminedichloroplatinum(II).[29]

The influence of alkyl chain length of the mono-adducts at the O^4 atom of 2′-deoxythymidine on stability was also assessed. There was a 10 °C reduction in duplex stability for the duplexes containing O^4MeT, O^4-butyl-4-ol-dT (**S4**) or O^4-heptyl-7-ol-dT (**S7**) (48 °C) compared to the non-alkylated control (58 °C) demonstrating that for these adducts the length of the alkyl chain attached at the O^4 atom of 2′-deoxythymidine had no substantial effect on duplex stability (see ESI‡). These results are supported by previous studies demonstrating that O^4-ethyl-2′-deoxythmidine/2′-deoxyadenosine base pair within a duplex decreased the T_m by 14 °C with respect to the native control.[30]

Circular dichroism spectroscopy of DNA duplexes

CD spectra of the cross-linked duplexes **XLTT4** and **XLTT7** exhibited signatures characteristic of B-form DNA with a positive signal centered around 275 nm, a crossover around 260 nm and a negative signal around 250 nm, as shown in Fig. 3.[31] These results indicate that the O^4-dT-alkyl-O^4-dT ICL caused minimal distortion to the global B-form structure.

Similarly, duplexes containing O^4MeT, O^4-butyl-4-ol-dT (**S4**) or O^4-heptyl-7-ol-dT (**S7**), showed little deviation in their CD signature relative to the unmodified control duplex demonstrating minimal distortion from the global B-form structure. These findings are in agreement with NMR reports on O^4MeT containing duplexes that suggest such lesions do not cause major structural deformation of the duplex.[32]

Repair and binding of AGTs with cross-linked and mono-alkylated substrates

Repair studies conducted on **XLTT4** and **XLTT7** cross-links revealed that none of the AGT proteins from human or *E. coli* were able to completely eliminate the cross-links, as observed in Fig. 4. Moreover, none of the proteins were able to create the intermediate product, which consists of one repair step leading to an AGT–ssDNA species.[13,14]

EMSA were conducted with **XLTT4**, **XLTT7** and a non-cross-linked control duplex, which has a T–A match where the T–T cross-link is located, to determine if the lack of activity of hAGT towards these lesions was inherent to the binding and detection of the lesion or if the lack of repair was a function of catalysis.

Results of the EMSA indicate that the hAGT stoichiometry for the various DNA was the same, with two proteins for every 11 bp duplex (see ESI‡). These results are consistent with the findings of others, which stipulates that hAGT binds every 4 nucleotides.[33] This stoichiometry was the same for the cross-linked DNA and for the control demonstrating that the presence of the cross-link does not affect how hAGT interact together when bound to the DNA. The dissociation constant of the control duplex was much higher than the cross-links, which was unanticipated.

To address the low binding affinity of hAGT for the control duplex a set of oligos were designed, where the central T in the control sequence (5′ CGA AAT TTT CG) was replaced by the other 3 natural 2′-deoxynucleotides, and assessed for binding by hAGT. Binding studies were conducted over a large range of protein to obtain a rough estimate of the K_d for the hAGT–DNA complexes. Results indicate that hAGT preferentially binds to DNA with a dG nucleotide that is further than 2 nucleotides from the ends (see ESI‡). This same hAGT binding trend has been observed for long A–T tracts.[34] The lack of a central dG explains the low binding affinity observed for the control sequence. The increased binding affinity observed for **XLTT4** and **XLTT7** over the control may be attributed to the large hydrophobic adduct at the O^4 atoms of the central dT, due to the presence of the linkers, and not as a result of the thermodynamic properties of the resulting DNA duplexes. This hydrophobic interaction between alkyl adduct and hAGT has been proposed as the basis of substrate discrimination by the protein for O^6MeG and O^6-benzyl-dG over dG.[35]

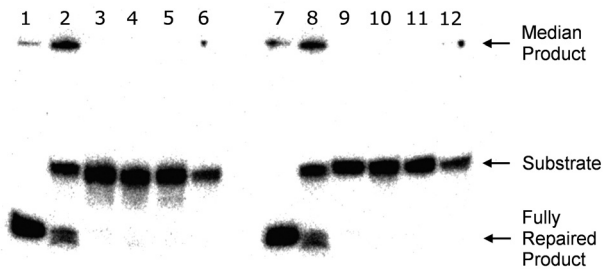


Fig. 4 Denaturing PAGE to demonstrate lack of repair of **XLTT4** and **XLTT7** by hAGT, Ada-C and OGT involving 2 pmol of DNA with 60 pmol of AGT. Lane 1, Control DNA; lane 2, **XLGG7** + hAGT; lane 3, **XLTT4**; lane 4, **XLTT4** + hAGT; lane 5, **XLTT4** + Ada-C; lane 6, **XLTT4** + OGT; lane 7, Control DNA; lane 8, **XLGG7** + hAGT; lane 9, **XLTT7**; lane 10, **XLTT7** + hAGT; lane 11, **XLTT7** + Ada-C; lane 12, **XLTT7** + OGT.

Experiments probing the repair of larger alkyl adducts at the O^4 atom of dT show a different trend than the ones observed for methylation (see ESI‡). For the larger lesions, O^4-butyl-4-ol-dT and O^4-heptyl-7-ol-dT, only OGT showed any ability to remove these alkylations to a reasonable level, indicating only OGT out of all 3 proteins is capable of repairing large lesions at the O^4 atom of thymine.

To assess the lack of repair of the adducts at the O^4 atom of dT by hAGT, binding assays were conducted. A general trend from the EMSA could be observed for the binding affinity of hAGT to the various mono-adduct DNA (see ESI‡). First, hAGT had very similar binding affinities to all mono-adduct DNA, which was roughly 1.8 times better than the non-alkylated control. Second, hAGT bound O^4-alkyl-dT lesions with very similar affinity as it did O^6MeG and the size of the lesion at the O^4 atom of dT had no substantial effect on binding for the alkylation sizes tested.

The observed increase in binding affinity for the adduct or ICL containing DNA and hAGT over their unalkylated controls suggests hAGT is capable of flipping out the alkylated thymidine and placing the lesion into its active site as required for repair. These results are consistent with our previous findings which demonstrate that hAGT is capable of flipping out cross-linked nucleotides as demonstrated through the repair of O^6-dG-alkyl-O^6-dG ICL.[13,14]

Modelling of O^4MeT and O^4-butyl-4-ol-dT in hAGT active site

To understand the underlying substrate discrimination observed for AGT with respect to the mono-adducts, molecular modelling was performed (coordinates for these models are provided in the ESI‡). hAGT is the only AGT that has been co-crystallized with a DNA substrate providing valuable insight into its active conformation. The substrate differences observed between various AGT homologues suggest substantial variations in the active sites amongst proteins. Attempts to model methyl and butyl-4-ol adducts with a homology model of OGT based on hAGT (generated from PDB ID: 1T38) were performed. However, the minimized structures did not correlate with the experimental evidence. Our modelling results suggest the methyl adduct fits in the active site of hAGT without steric clashes (Fig. 5). The

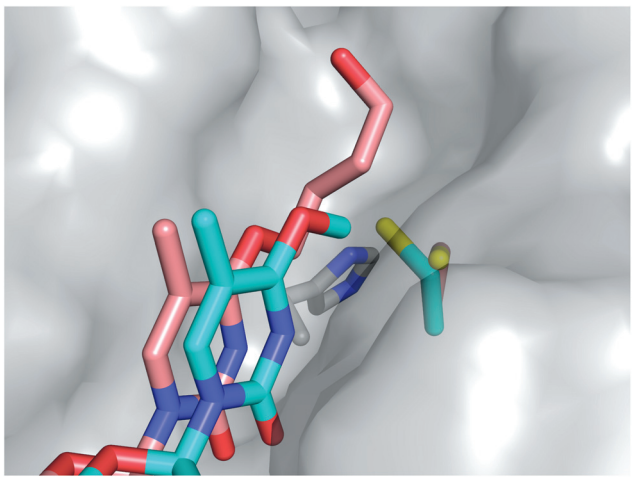


Fig. 5 Overlay of O^4MeT (cyan) and O^4-butyl-4-ol-dT (pink) in the active site of hAGT (white). The adopted rotamers of Cys145 are shown for O^4MeT (cyan) and O^4-butyl-4-ol-dT (pink) and His146 (grey).

model of the butyl-4-ol adduct proposes a steric clash in the active site of the protein between the γ-carbon of the lesion and the protein due to the adoption of the *gauche* conformation at the β-carbon of the adduct. This interaction causes Cys145 to adopt the incorrect rotamer for the alkyl transfer reaction to take place when the butyl-4-ol adduct is present. In the mechanism of hAGT repair, His146 is responsible for proton abstraction of a nearby water molecule from Cys145 allowing the formation of the thiolate anion. In the case where Cys145 is rotated away from His146 (as observed for the butyl-4-ol lesion) the thiolate anion will not be produced and the protein will be inactive.

Experimental

General methods and materials

5′-*O*-Dimethoxytrityl-N^2-phenoxyacetyl-2′-deoxyguanosine, 3′-*O*-dimethoxytrityl-2′-deoxyribonucleoside-5′-*O*-(β-cyanoethyl-*N*,*N*′-diisopropyl)phosphoramidites and *N*,*N*-diisopropylamino-cyanoethylphosphonamidic chloride were purchased from ChemGenes Inc. (Wilmington, MA). 5′-*O*-Dimethoxytrityl-2′-deoxyribonucleoside-3′-*O*-(β-cyanoethyl-*N*,*N*′-diisopropyl)phosphoramidites and protected 2′-deoxyribonucleoside–CPG supports were purchased from Glen Research (Sterling, Virginia). All other chemicals and solvents were purchased from the Aldrich Chemical Company (Milwaukee, WI) or EMD Chemicals Inc. (Gibbstown, NJ). Flash column chromatography was performed using silica gel 60 (230–400 mesh) obtained from Silicycle (Quebec City, QC). Thin layer chromatography (TLC) was performed using precoated TLC plates (Merck, Kieselgel 60 F_{254}, 0.25 mm) purchased from EMD Chemicals Inc. (Gibbstown, NJ). NMR spectra were recorded on a Varian 500 MHz NMR spectrometer at room temperature. ^{1}H NMR spectra were recorded at a frequency of 500.0 MHz and chemical shifts were reported in parts per million downfield from tetramethylsilane. ^{13}C NMR spectra (^{1}H decoupled) were recorded at a frequency of 125.7 MHz with tetramethylsilane as a reference. ^{31}P NMR spectra (^{1}H decoupled) were recorded at a frequency of 202.3 MHz with H_3PO_4 used as an external standard. High resolution mass spectrometry of modified nucleosides were obtained using an LTQ OrbitrapVelos – ETD (Thermo Scientific) at the Concordia University Centre for Biological Applications of Mass Spectrometry. The mass spectrometer was operated in full scan, positive ion detection mode. Ampicillin, isopropyl β-D-thiogalactopyranoside (IPTG), and most other biochemical reagents as well as polyacrylamide gel materials were purchased from Bioshop Canada Inc (Burlington, ON). Ni-NTA Superflow Resin was purchased from Qiagen (Mississauga, ON). Complete, Mini, EDTA-free Protease Inhibitor Cocktail Tablets were obtained from Roche (Laval, QC) Nitrocellulose filters (0.20 μm) were obtained from Millipore. XL-10 Gold and Bl21(DE3) *E. coli* cells were obtained from Stratagene (Cedar Creek, TX). DpnI, T4 polynucleotide kinase (PNK), Unstained Protein Molecular Weight Marker and restriction enzymes EcoRI and KpnI were obtained from Fermentas (Burlington, ON). [γ-^{32}P]ATP was purchased from Amersham Canada Ltd (Oakville, ON). Phusion Polymerase was obtained from New England Biolabs (Ipswich, MA). DNA primers for site directed mutagenesis and cloning were purchased from Biocorp (Montreal, QC). Oligonucleotides containing O^4MeT required for repair studies were a kind gift from TriLink Biotechnologies (San Diego, CA).

Chemical synthesis of nucleosides

The synthesis of phosphoramidites **4a**, **4b**, **7a** and **7b** are shown in Schemes 1 and 2.

3′-*O*-(*tert*-Butyldimethylsilyl)-5′-*O*-(4,4′-dimethoxytrityl)-O^4-(hydroxybutyl)-thymidine (1a)

1,4-Butanediol (1.60 g, 17.8 mmol) was dissolved in THF (15 mL). Sodium metal (2.35 g, 102 mmol) was added. After 5 h, the salt was extracted to be immediately used in the subsequent reaction. 3′-*O*-(*tert*-Butyldimethylsilyl)-5′-*O*-(4,4′-dimethoxytrityl)-4-triazolyl-thymidine (2.5 g, 3.52 mmol) was dissolved in dioxane (10 mL).[16] Sodium 4-hydroxybutan-1-olate was added dropwise. After 15 h, the solvent was evaporated *in vacuo*, the crude product was taken up in dichloromethane (50 mL) and the solution was washed with two portions of sodium bicarbonate (2 × 50 mL). The organic layer was dried over sodium sulfate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a hexanes–ethyl acetate solvent system (2 : 3) to afford 2.05 g (79.8%) of product as a colorless foam. R_f (SiO_2 TLC): 0.24 hexanes–ethyl acetate (2 : 3). $\lambda_{max(ACN)}$ = 282 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 7.98 (s, 1H, H6), 7.30–7.45 (m, 10H, Ar), 6.85–6.88 (m, 3H, Ar), 6.36 (dd, 1H, H1′, *J* = 6 Hz), 4.53 (m, 1H, H3′), 4.47 (m, 2H, $ArOCH_2$), 4.02 (m, 1H, H4′), 3.85 (s, 6H, OCH_3), 3.75 (t, 2H, CH_2), 3.56 (dd, 1H, H5′), 3.31 (dd, 1H, H5′), 2.56 (m, 1H, H2′), 2.27 (m, 1H, H2′), 1.90 (tt, 2H, CH_2), 1.74 (tt, 2H, CH_2), 1.64 (s, 1H, OH), 1.57 (s, 3H, $ArCH_3$), 0.87 (s, 9H, $SiC(CH_3)_3$), 0.08 (s, 6H, $Si(CH_3)_2$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.4, 158.7, 156.0 144.4, 139.5, 135.53, 135.51, 130.11, 130.10, 128.2, 127.9, 127.1, 113.24, 113.21, 104.6, 86.7, 86.5, 86.3, 71.0, 67.1, 62.4, 62.2, 42.2, 29.1, 25.7, 25.1, 17.9, 11.7, −4.62, −4.94. IR (thin film); ν_{max} (cm^{-1}) = 3414, 3061, 2953, 1669, 1509, 1252, 1178, 1109, 1035, 835, 782. HRMS (ESI-MS) *m*/*z* calculated for $C_{41}H_{54}N_2O_8SiNa^+$ 753.3546: found 753.3555 $[M + Na]^+$.

3′-*O*-(*tert*-Butyldimethylsilyl)-5′-*O*-(4,4′-dimethoxytrityl)-O^4-(hydroxyheptyl)-thymidine (1b)

1,7-Heptanediol (1.7 g, 11.35 mmol) was dissolved in THF (10 mL). Sodium metal (1.7 g, 65.2 mmol) was added. After 5 h, the salt was extracted to be immediately used in the subsequent reaction. 3′-*O*-(*tert*-Butyldimethylsilyl)-5′-*O*-(4,4′-dimethoxytrityl)-4-triazolyl-thymidine (1.7 g, 1.13 mmol) was dissolved in dioxane (5 mL).[16] Sodium 7-hydroxyheptan-1-olate was added dropwise. After 15 h, the crude product was taken up in ethyl acetate (50 mL) and the solution was washed with two portions of sodium bicarbonate (2 × 50 mL). The organic layer was dried over sodium sulfate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a hexanes–ethyl acetate solvent system (6 : 4). Further purification of the product was achieved by flash

column chromatography using an acetonitrile–dichloromethane (1 : 9 → 3 : 25) solvent system to afford 0.594 g (42.1%) of product as a colorless foam. R_f (SiO_2 TLC): 0.43 hexanes–ethyl acetate (2 : 3). $\lambda_{max(ACN)}$ = 282 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 7.96 (s, 1H, H6), 7.29–7.34 (m, 10H, Ar), 6.86–6.88 (m, 3H, Ar), 6.37 (dd, 1H, H1′, J = 5.75 Hz), 4.53 (dd, 1H, H3′), 4.24 (t, 2H, $ArOCH_2$), 4.02 (m, 1H, H4′), 3.85 (s, 6H, OCH_3), 3.69 (m, 2H, CH_2OH), 3.56 (dd, 1H, H5′), 3.30 (dd, 1H, H5′), 2.55 (m, 1H, H2′), 2.25 (m, 1H, H2′), 1.80 (m, 2H, CH_2), 1.63 (m, 2H, CH_2), 1.57 (s, 3H, CH_3Ar), 1.47 (m, 6H, $(CH_2)_3$), 0.89 (s, 9H, $SiC(CH_3)_3$), 0.08 (s, 6H, $Si(CH_3)_2$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.5, 158.7, 156.0 144.4, 139.4, 135.56, 135.55, 130.12, 130.10, 128.2, 127.9, 127.1, 113.24, 113.21, 104.6, 86.7, 86.5, 86.3, 71.1, 67.3, 62.9, 62.4, 55.3, 42.2, 32.7, 29.8, 29.0, 28.4, 25.8, 25.7, 25.6, 17.9, 11.8, −4.62, −4.94. IR (thin film); ν_{max} (cm^{-1}) = 3420, 3060, 2930, 1668, 1509, 1252, 1177, 1109, 1035, 834, 782. HRMS (ESI-MS) *m/z* calculated for $C_{44}H_{60}N_2O_8SiNa^+$ 795.4017: found 795.4017 $[M + Na]^+$.

1-{O^4-[3′-*O*-(*tert*-Butyldimethylsilyl)-5′-*O*-(4,4′-dimethoxytrityl)-thymidyl]}-4-{O^4-[3′-*O*-(*tert*-butyldimethylsilyl)-5′-*O*-(4,4′-dimethoxytrityl)-thymidyl]}-butane (2a)

1a (0.154 g, 0.205 mmol) and 3′-*O*-(*tert*-butyldimethylsilyl)-5′-*O*-(4,4′-dimethoxytrityl)-4-triazolyl-thymidine (0.117 g, 0.164 mmol) was dissolved in pyridine (3 mL). DBU (0.062 g, 0.410 mmol) was added dropwise. After 3 d, the crude product was taken up in ethyl acetate (100 mL) and the solution was washed with four portions of sodium bicarbonate (4 × 50 mL). The organic layer was dried over sodium sulfate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a hexanes–ethyl acetate (1 : 1 → 2 : 3) solvent system to afford 0.151 g (67.7%) of product as a colorless foam. R_f (SiO_2 TLC): 0.22 hexanes–ethyl acetate (2 : 3). $\lambda_{max(ACN)}$ = 282 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 7.96 (s, 2H, H6), 7.30–7.46 (m, 20H, Ar), 6.86–6.88 (m, 6H, Ar), 6.37 (dd, 1H, H1′, J = 5.75 Hz), 4.53 (m, 2H, H3′), 4.47 (m, 4H, CH_2), 4.01 (m, 2H, H4′), 3.84 (s, 12H, OCH_3), 3.57 (dd, 2H, H5′), 3.31 (dd, 2H, H5′), 2.55 (m, 2H, H2′), 2.26 (m, 2H, H2′), 1.93 (m, 4H, CH_2), 1.57 (s, 6H, $ArCH_3$), 0.87 (s, 18H, $SiC(CH_3)_3$), 0.07 (s, 12H, $Si(CH_3)_2$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.4, 158.7, 155.9 144.4, 139.6, 135.51, 135.50, 130.11, 130.10, 128.2, 127.9, 127.1, 113.24, 113.21, 104.5, 86.7, 86.5, 86.3, 71.1, 66.6, 62.4, 55.2, 42.2, 25.7, 25.3, 17.9, 11.7, −4.62, −4.94. IR (thin film); ν_{max} (cm^{-1}) = 3057, 2954, 1671, 1608, 1509, 1463, 1329, 1252, 1178, 1035, 835, 736. HRMS (ESI-MS) *m/z* calculated for $C_{79}H_{98}N_4O_{14}Si_2Na^+$ 1393.6515: found 1393.6526 $[M + Na]^+$.

1-{O^4-[3′-*O*-(*tert*-Butyldimethylsilyl)-5′-*O*-(4,4′-dimethoxytrityl)-thymidyl]}-7-{O^4-[3′-*O*-(*tert*-butyldimethylsilyl)-5′-*O*-(4,4′-dimethoxytrityl)-thymidyl]}-heptane (2b)

1b (0.500 g, 0.647 mmol) and 3′-*O*-(*tert*-butyldimethylsilyl)-5′-*O*-(4,4′-dimethoxytrityl)-4-triazolyl-thymidine (0.368 g, 0.518 mmol) was dissolved in pyridine (10 mL). DBU (0.197 g, 1.29 mmol) was added dropwise. After 4.5 d, the crude product was taken up in ethyl acetate (150 mL) and the solution was washed with four portions of sodium bicarbonate (4 × 50 mL). The organic layer was dried over sodium sulfate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using an acetonitrile–dichloromethane (1 : 9 → 3 : 25) solvent system to afford 0.358 g (39.5%) of product as a colorless foam. R_f (SiO_2 TLC): 0.49 hexanes–ethyl acetate (2 : 3). $\lambda_{max(ACN)}$ = 282 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 7.99 (s, 2H, H6), 7.33–7.49 (m, 20H, Ar), 6.89–6.91 (m, 6H, Ar), 6.40 (dd, 2H, H1′, J = 12 Hz), 4.54 (dd, 2H, H3′), 4.44 (t, 4H, $ArOCH_2$), 4.03 (m, 2H, H4′), 3.86 (s, 12H, OCH_3), 3.58 (dd, 2H, H5′), 3.34 (dd, 2H, H5′), 2.58 (m, 2H, H2′), 2.26 (m, 2H, H2′), 1.81 (m, 4H, CH_2), 1.59 (s, 6H, CH_3Ar), 1.48 (m, 6H, $(CH_2)_3$), 0.88 (s, 18H, $SiC(CH_3)_3$), 0.07 (s, 12H, $Si(CH_3)_2$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.5, 158.7, 156.0 144.4, 139.4, 135.5, 130.13, 130.11, 128.2, 127.9, 127.1, 113.24, 113.21, 104.6, 86.7, 86.5, 86.3, 71.1, 67.3, 62.4, 55.3, 42.2, 29.0, 28.5, 25.9, 25.7, 17.9, 11.8, −4.61, −4.94. IR (thin film); ν_{max} (cm^{-1}) = 3058, 2952, 1671, 1533, 1508, 1327, 1252, 1177, 1035, 834. HRMS (ESI-MS) *m/z* calculated for $C_{81}H_{104}N_4O_{14}Si_2Na^+$ 1435.6985: found 1435.7037 $[M + Na]^+$.

1-{O^4-[5′-*O*-(4,4′-Dimethoxytrityl)-thymidyl]}-4-{O^4-[5′-*O*-(4,4′-dimethoxytrityl)-thymidyl]}-butane (3a)

2a (0.658 g, 0.485 mmol) was dissolved in THF (4.8 mL). TBAF (1 M in THF) (1.97 mL, 1.97 mmol) was added dropwise. After 30 min, the crude product was taken up in ethyl acetate (50 mL) and the solution was washed with two portions of sodium bicarbonate (2 × 50 mL). The organic layer was dried over sodium sulfate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a methanol–hexanes–ethyl acetate (1 : 5 : 4 → 1 : 4 : 5) solvent system to afford 0.514 g (93.0%) of product as a colorless foam. R_f (SiO_2 TLC): 0.21 methanol–hexanes–ethyl acetate (1 : 5 : 4). $\lambda_{max(ACN)}$ = 282 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 7.78 (s, 2H, H6), 7.23–7.37 (m, 20H, Ar), 6.78–6.81 (m, 6H, Ar), 6.34 (dd, 2H, H1′, J = 6 Hz), 4.52 (m, 2H, H3′), 4.40 (m, 4H, CH_2), 4.05 (m, 2H, H4′), 3.77 (s, 12H, OCH_3), 3.46 (dd, 2H, H5′), 3.35 (dd, 2H, H5′), 2.58 (m, 2H, H2′), 2.25 (m, 2H, H2′), 2.11 (s, 2H, OH),1.85 (m, 4H, CH_2), 1.56 (s, 6H, $ArCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.4, 158.7, 158.6, 144.5, 139.8, 135.6, 135.5, 130.1, 130.1, 128.1, 128.0, 127.0, 113.3, 105.0, 86.8, 86.5, 86.3, 66.8, 63.5, 60.4, 55.2 42.2, 25.2, 21.1, 14.2, 11.7. IR (thin film); ν_{max} (cm^{-1}) = 3353, 3057, 2953, 1661, 1509, 1431, 1329, 1251, 1177, 1095, 1035, 829, 735. HRMS (ESI-MS) *m/z* calculated for $C_{66}H_{70}N_4O_{14}Na^+$ 1165.4786: found 1165.4806 $[M + Na]^+$.

1-{O^4-[5′-*O*-(4,4′-Dimethoxytrityl)-thymidyl]}-7-{O^4-[5′-*O*-(4,4′-dimethoxytrityl)-thymidyl]}-heptane (3b)

2b (0.357 g, 0.263 mmol) was dissolved in THF (2 mL). TBAF (1 M in THF) (0.63 mL, 0.63 mmol) was added dropwise. After 25 min, the crude product was taken up in ethyl acetate (50 mL) and the solution was washed with three portions of sodium bicarbonate (3 × 50 mL). The organic layer was dried over sodium sulfate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a

methanol–hexanes–ethyl acetate (1 : 5 : 4 → 1 : 4 : 5) solvent system to afford 0.256 g (82.0%) of product as a colorless foam. R_f (SiO_2 TLC): 0.22 methanol–hexanes–ethyl acetate (1 : 11 : 8). $\lambda_{max(ACN)}$ = 282 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 7.77 (s, 2H, H6), 7.18–7.37 (m, 20H, Ar), 6.79–6.80 (m, 6H, Ar), 6.34 (dd, 2H, H1′, J = 12.5 Hz), 4.50 (m, 2H, H3′), 4.29–4.39 (m, 4H, CH_2), 4.05 (m, 2H, H4′), 3.78 (s, 12H, OCH_3), 3.46 (dd, 2H, H5′), 3.36 (dd, 2H, H5′), 2.58 (m, 2H, H2′), 2.25 (m, 2H, H2′), 1.70–1.74 (m, 4H, CH_2), 1.56 (s, 6H, $ArCH_3$), 1.39–1.41 (m, 6H, CH_2). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.6, 158.7, 156.2, 144.5, 139.5, 135.6, 135.5, 130.1, 128.1, 128.0, 127.1, 113.3, 105.0, 86.8, 86.4, 86.1, 71.9, 67.4, 63.4, 55.3, 42.1, 28.5, 25.8, 11.8. IR (thin film); ν_{max} (cm^{-1}) = 3350, 3058, 2932, 1654, 1532, 1508, 1458, 1328, 1251, 1177, 1035, 829, 735. HRMS (ESI-MS) *m/z* calculated for $C_{69}H_{76}N_4O_{14}Na^+$ 1207.5255: found 1207.5270 $[M + Na]^+$.

1-{O^4-[3′-*O*-(β-Cyanoethyl-*N*,*N*′-diisopropyl)-5′-*O*-(4,4′-dimethoxytrityl)-thymidyl]}-4-{O^4-[3′-*O*-(β-cyanoethyl-*N*,*N*′-diisopropyl)-5′-*O*-(4,4′-dimethoxytrityl)-thymidyl]}-butane (4a)

3a (0.200 g, 0.175 mmol) and diisopropylethylamine (0.068 g, 0.525 mmol) were dissolved in THF (2 mL), followed by *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride (0.104 g, 0.438 mmol) and the reaction was allowed to stir at room temperature for 30 min. Upon completion, the reaction was quenched by the addition of ethyl acetate (50 mL), and the solution was washed with sodium bicarbonate (3%, 2 × 25 mL) and brine (25 mL). The organic layer was dried over sodium sulfate, decanted, and evaporated. The product, a colorless powder, was precipitated from hexanes (0.152 g, 56.2%). R_f (SiO_2 TLC): 0.12, 0.18, 0.27 ethyl acetate. $\lambda_{max(ACN)}$ = 282 nm. ^{1}H NMR (500 MHz, d_6-acetone, ppm): 7.83–7.91 (s, 2H, H6), 7.47–7.50 (m, 4H, Ar), 7.21–7.39 (m, 14H, Ar), 6.87–6.91 (m, 8H, Ar), 6.29–6.34 (m, 2H, H1′), 4.70–4.74 (m, 2H, H3′), 4.37 (m, 4H, CH_2), 4.17–4.24 (m, 2H, H4′), 3.57–3.89 (m, 20H, CH_2OP, NCH, OCH_3), 3.34–3.51 (m, 4H, H5′), 2.74 (t, 2H, CH_2CN), 2.52–2.63 (m, 4H, H2′, CH_2CN), 2.34–2.41 (m, 2H, H2′), 1.89–1.91 (m, 4H, $(CH_2)_2$), 1.57–1.60 (m, 6H, $ArCH_3$) 1.15–1.20 (m, 18H, CH_3), 1.07–1.08 (m, 6H, CH_3). ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.28, 148.07. ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 170.0, 158.9, 158.8, 154.7, 144.97, 145.95, 140.03, 139.97, 135.7, 135.63, 135.60, 135.56, 130.17, 130.15, 128.2, 128.1, 127.87, 127.86, 126.91, 126.88, 118.1, 117.9, 113.1, 103.5, 103.4, 86.64, 86.61, 85.99, 85.9, 85.5, 85.3, 85.2, 73.6, 73.5, 73.1, 73.0, 66.2, 63.2, 63.0, 58.70, 58.69, 54.70, 54.68, 43.1, 43.0, 40.4, 40.3, 29.4, 25.1, 24.05, 24.03, 23.99, 23.93, 19.9, 19.8, 11.18, 11.17. IR (thin film); ν_{max} (cm^{-1}) = 3031, 2966, 1672, 1534, 1509, 1462, 1329, 1251, 1179, 1035. HRMS (ESI-MS) *m/z* calculated for $C_{84}H_{104}N_8O_{16}P_2Na^+$ 1565.6943: found 1565.6854 $[M + Na]^+$.

1-{O^4-[3′-*O*-(β-Cyanoethyl-*N*,*N*′-diisopropyl)-5′-*O*-(4,4′-dimethoxytrityl)thymidyl]}-7-{O^4-[3′-*O*-(β-cyanoethyl-*N*,*N*′-diisopropyl)-5′-*O*-(4,4′-dimethoxytrityl)-thymidyl]}-heptane (4b)

3b (0.170 g, 0.143 mmol) and diisopropylethylamine (0.056 g, 0.430 mmol) were dissolved in THF (2 mL), followed by *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride (0.085 g, 0.359 mmol) and the reaction was allowed to stir at room temperature for 30 min. Upon completion, the reaction was quenched by the addition of ethyl acetate (50 mL), and the solution was washed with sodium bicarbonate (3%, 2 × 25 mL) and brine (25 mL). The organic layer was dried over sodium sulfate, decanted, and evaporated. The product, a colorless powder, was precipitated from hexanes (0.146 g, 64.3%). R_f (SiO_2 TLC): 0.31, 0.46, 0.62 ethyl acetate. $\lambda_{max(ACN)}$ = 282 nm. ^{1}H NMR (500 MHz, d_6-acetone, ppm): 7.88–7.92 (s, 2H, H6), 7.48–7.51 (m, 4H, Ar), 7.23–7.40 (m, 14H, Ar), 6.87–6.93 (m, 8H, Ar), 6.30–6.35 (m, 2H, H1′), 4.70–4.76 (m, 2H, H3′), 4.30–4.32 (m, 4H, CH_2), 4.18–4.24 (m, 2H, H4′), 3.60–3.91 (m, 20H, CH_2OP, NCH, OCH_3), 3.39–3.52 (m, 4H, H5′), 2.77 (t, 2H, CH_2CN), 2.53–2.64 (m, 4H, H2′, CH_2CN), 2.37–2.42 (m, 2H, H2′), 1.74–1.78 (m, 4H, $(CH_2)_2$), 1.58–1.61 (m, 6H, $ArCH_3$) 1.44–1.49 (m, 6H, $(CH_2)_3$), 1.18–1.21 (m, 18H, CH_3), 1.09–1.11 (m, 6H, CH_3). ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.28, 148.07. ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 170.09, 170.08, 158.88, 158.86, 154.7, 145.0, 140.0, 139.9, 135.7, 135.63, 135.61, 135.57, 130.2, 130.1, 128.2, 128.1, 127.9, 127.8, 126.89, 126.86, 118.1, 117.9, 113.1, 103.5, 103.4, 86.62, 86.59, 86.0, 85.9, 85.5, 85.3, 85.2, 73.6, 73.5, 73.1, 73.0, 66.6, 63.2, 63.0, 58.7, 58.5, 54.68, 54.66, 43.1, 43.0, 40.4, 40.3, 29.4, 25.7, 24.04, 24.02, 23.97, 23.96, 23.9, 19.94, 19.89, 19.83, 19.77, 13.6, 11.17, 11.16. IR (thin film); ν_{max} (cm^{-1}) = 3057, 2966, 1672, 1510, 1463, 1329, 1252, 1179, 1036. HRMS (ESI-MS) *m/z* calculated for $C_{87}H_{110}N_8O_{16}P_2Na^+$ 1607.7412: found 1607.7285 $[M + Na]^+$.

3′-*O*-(*tert*-Butyldimethylsilyl)-5′-*O*-(4,4′-dimethoxytrityl)-O^4-(phenoxyacetyloxybutyl)-thymidine (5a)

1a (0.300 g, 0.410 mmol) was dissolved in THF (1.2 mL). Triethylamine (0.145 g, 1.43 mmol) followed by phenoxyacetyl chloride (0.122 g, 0.715 mmol) was added dropwise. The reaction was set in an ice bath for 25 min. After 1.5 h, the crude product was taken up in ethyl acetate (50 mL) and the solution was washed with three portions of sodium bicarbonate (3 × 50 mL). The organic layer was dried over sodium sulfate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a hexanes–ethyl acetate (3 : 7) solvent system to afford 0.266 g (75.1%) of product as a colorless foam. R_f (SiO_2 TLC): 0.39 hexanes–ethyl acetate (1 : 1). $\lambda_{max(ACN)}$ = 276 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 7.98 (s, 1H, H6), 7.26–7.50 (m, 13H, Ar), 6.85–7.09 (m, 5H, Ar), 6.37 (dd, 1H, H1′, J = 6 Hz), 4.69 (s, 2H, $PhOCH_2CO$), 4.53 (m, 1H, H3′), 4.42 (m, 2H, $ArOCH_2$), 4.30 (t, 2H, $COCH_2$),4.01 (m, 1H, H4′), 3.84 (s, 6H, OCH_3), 3.58 (dd, 1H, H5′), 3.31 (dd, 1H, H5′), 2.55 (m, 1H, H2′), 2.26 (m, 1H, H2′), 1.84 (m, 4H, CH_2), 1.57 (s, 3H, $ArCH_3$), 0.87 (s, 9H, $SiC(CH_3)_3$), 0.08 (s, 6H, $Si(CH_3)_2$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.3, 169.0, 158.7, 157.8, 155.6, 144.4, 139.7, 135.52, 135.50, 130.11, 130.09, 129.6, 128.2, 127.9, 127.1, 121.8, 114.6, 113.24, 113.21, 104.4, 86.7, 86.5, 86.3, 71.1, 66.5, 65.3, 64.8, 62.4, 55.2, 42.2, 25.7, 25.3, 25.1, 17.9, 11.7, −4.62, −4.94. IR (thin film); ν_{max} (cm^{-1}) = 3061, 2954, 1760, 1672, 1532, 1509, 1497, 1252, 1178, 835, 783. HRMS (ESI-MS) *m/z*

calculated for $C_{49}H_{60}N_2O_{10}SiNa^+$ 887.3914: found 887.3921 $[M + Na]^+$.

3′-*O*-(*tert*-Butyldimethylsilyl)-5′-*O*-(4,4′-dimethoxytrityl)-O^4-(phenoxyacetyloxyheptyl)-thymidine (5b)

1b (0.527 g, 0.682 mmol) was dissolved in THF (0.5 mL). Triethylamine (0.110 g, 1.09 mmol) was added followed by cooling of the reaction in an ice bath for 10 min. Phenoxyacetyl chloride (0.175 g, 1.02 mmol) was added dropwise. After 1.5 h, the solvent was evaporated *in vacuo*, the crude product was taken up in dichloromethane (50 mL) and the solution was washed with two portions of sodium bicarbonate (2 × 50 mL). The organic layer was dried over sodium sulfate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a hexanes–ethyl acetate (4 : 1 → 2 : 3) solvent system to afford 0.491 g (79.4%) of product as a colorless foam. R_f (SiO_2 TLC): 0.49 hexanes–ethyl acetate (1 : 1). $\lambda_{max(ACN)}$ = 276 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 8.00 (s, 1H, H6), 7.45 (dd, 2H, Ar), 7.25–7.34 (m, 9H, Ar), 6.85–7.04 (m, 7H, Ar), 6.37 (dd, 1H, H1′, *J* = 6.0 Hz), 4.68 (s, 2H, $PhOCH_2CO$), 4.53 (dd, 1H, H3′), 4.41 (t, 2H, $ArOCH_2$), 4.25 (t, 2H, CH_2OCO), 4.02 (m, 1H, H4′), 3.84 (s, 6H, OCH_3), 3.58 (dd, 1H, H5′), 3.30 (dd, 1H, H5′), 2.55 (m, 1H, H2′), 2.26 (m, 1H, H2′), 1.77 (quintet, 2H, CH_2), 1.71 (quintet, 2H, CH_2), 1.58 (s, 3H, CH_3Ar), 1.37–1.46 (m, 6H, $(CH_2)_3$), 0.88 (s, 9H, $SiC(CH_3)_3$), 0.07 (s, 6H, $Si(CH_3)_2$. ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.5, 169.1, 158.7, 157.8, 156.0, 144.4, 139.4, 135.5, 130.11, 130.09, 129.5, 128.2, 127.9, 127.1, 121.7, 114.6, 113.23, 113.21, 104.5, 86.7, 86.5, 86.3, 71.1, 67.2, 65.4, 65.3, 62.4, 55.2, 42.2, 28.8, 28.5, 28.4, 25.8, 25.71, 25.69, 17.9, 11.8, −4.62, −4.95. IR (thin film); ν_{max} (cm^{-1}) = 3062, 2928, 1760, 1675, 1510, 1464, 154, 1065, 832, 734. HRMS (ESI-MS) *m*/*z* calculated for $C_{52}H_{66}N_2O_{10}SiNa^+$ 929.4382: found 929.4384 $[M + Na]^+$.

5′-*O*-(4,4′-Dimethoxytrityl)-O^4-(phenoxyacetyloxybutyl)-thymidine (6a)

5a (0.200 g, 0.231 mmol) was dissolved in THF (2.4 mL). TBAF (1 M in THF) (0.277 mL, 0.277 mmol) was added dropwise. After 15 min, the crude product was taken up in ethyl acetate (50 mL) and the solution was washed with two portions of sodium bicarbonate (2 × 50 mL). The organic layer was dried over sodium sulfate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a methanol–hexanes–ethyl acetate (2 : 11 : 7) solvent system to afford 0.164 g (94.1%) of product as a colorless foam. R_f (SiO_2 TLC): 0.29 methanol–hexane–ethyl acetate (2 : 11 : 7). $\lambda_{max(ACN)}$ = 276 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 7.79 (s, 1H, H6), 7.18–7.37 (m, 13H, Ar), 6.79–6.96 (m, 5H, Ar), 6.34 (dd, 1H, H1′, *J* = 6.5 Hz), 4.61 (s, 2H, $PhOCH_2CO$), 4.50 (m, 1H, H3′), 4.36 (m, 2H, $ArOCH_2$), 4.23 (t, 2H, $COCH_2$),4.05 (m, 1H, H4′), 3.77 (s, 6H, OCH_3), 3.47 (dd, 1H, H5′), 3.36 (dd, 1H, H5′), 2.58 (m, 1H, H2′), 2.23 (m, 1H, H2′), 1.75–1.78 (m, 4H, CH_2), 2.04 (s, 1H, OH), 1.55 (s, 3H, $ArCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.4, 169.0, 158.7, 158.6, 157.8, 156.1, 144.5, 139.9, 135.6, 135.5, 130.08, 130.07, 129.6, 128.1, 128.0, 127.0, 121.8, 114.6, 113.3, 104.9, 86.8, 86.5, 86.4, 72.0, 66.7, 65.3, 64.8, 63.5, 55.2, 42.2, 25.3, 25.1, 11.7. IR (thin film); ν_{max} (cm^{-1}) = 3342, 3061, 2956, 1758, 1663, 1533, 1509, 1251, 1177, 1090, 755. HRMS (ESI-MS) *m*/*z* calculated for $C_{43}H_{46}N_2O_{10}Na^+$ 773.3049: found 773.3055 $[M + Na]^+$.

5′-*O*-(4,4′-Dimethoxytrityl)-O^4-(phenoxyacetyloxyheptyl)-thymidine (6b)

5b (0.491 g, 0.541 mmol) was dissolved in THF (4.5 mL). TBAF (1 M in THF) (0.65 mL, 0.65 mmol) was added dropwise. After 25 min, the solvent was evaporated *in vacuo*, the crude product was taken up in dichloromethane (50 mL) and the solution was washed with two portions of sodium bicarbonate (2 × 50 mL). The organic layer was dried over sodium sulfate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a methanol–hexanes–ethyl acetate (0 : 1 : 1 → 1 : 19 : 30) solvent system to afford 0.373 g (86.8%) of product as a colorless foam. $\lambda_{max(ACN)}$ = 276 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 8.00 (s, 1H, H6), 7.29–7.40 (m, 2H, Ar), 7.18–7.27 (m, 9H, Ar), 6.94–6.97 (m, 1H, Ar), 6.87–6.88 (m, 2H, Ar), 6.79–6.81 (m, 4H, Ar), 6.37 (dd, 1H, H1′, *J* = 6.25 Hz), 4.61 (s, 2H, $PhOCH_2CO$), 4.52 (m, 1H, H3′), 4.34 (t, 2H, $ArOCH_2$), 4.18 (t, 2H, CH_2OCO), 4.07 (m, 1H, H4′), 3.78 (s, 6H, OCH_3), 3.47 (dd, 1H, H5′), 3.36 (dd, 1H, H5′), 2.60 (m, 1H, H2′), 2.33 (s, 1H, OH), 2.25 (m, 1H, H2′), 1.72 (quintet, 2H, CH_2), 1.64 (quintet, 2H, CH_2), 1.56 (s, 3H, CH_3Ar), 1.33–1.37 (m, 6H, $(CH_2)_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.6, 169.1, 158.6, 157.8, 156.3, 144.5, 139.7, 135.6, 135.5, 130.09, 130.07, 129.5, 128.1, 127.9, 127.0, 121.7, 114.6, 113.2, 105.0, 86.8, 86.5, 86.3, 72.0, 67.3, 65.4, 65.3, 63.5, 55.2, 42.2, 28.8, 28.5, 28.4, 25.8, 25.7, 11.7. IR (thin film); ν_{max} (cm^{-1}) = 3346, 3060, 2934, 1757, 1663, 1532, 1509, 1251, 1177, 1091, 735. HRMS (ESI-MS) *m*/*z* calculated for $C_{46}H_{52}N_2O_{10}Na^+$ 815.3519: found 815.3517 $[M + Na]^+$.

3′-*O*-(β-Cyanoethyl-*N*,*N*′-diisopropyl)-5′-*O*-(4,4′-dimethoxytrityl)-O^4-(phenoxyacetyloxybutyl)-thymidine (7a)

6a (0.200 g, 0.266 mmol) and diisopropylethylamine (0.051 g, 0.398 mmol) were dissolved in THF (2.5 mL), followed by *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride (0.076 g, 0.319 mmol). The reaction was allowed to stir at room temperature for 30 min, where TLC revealed the presence of starting material. Diisopropylethylamine (0.006 g, 0.049 mmol) followed by *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride (0.011 g, 0.045 mmol) were added and the reaction was allowed to proceed for another 30 min. Upon completion, the reaction was quenched by the addition of ethyl acetate (50 mL), and the solution was washed with sodium bicarbonate (3%, 2 × 25 mL) and brine (25 mL). The organic layer was dried over sodium sulfate, decanted, and evaporated. The product, a colorless powder, was precipitated from hexanes (0.116 g, 46.2%). R_f (SiO_2 TLC): 0.49, 0.67 hexanes–ethyl acetate (2 : 8). $\lambda_{max(ACN)}$ = 276 nm. ^{1}H NMR (500 MHz, d_6-acetone, ppm): 7.90–7.94 (s, 1H, H6), 7.46–7.51 (m, 2H, Ar), 7.24–7.39 (m, 9H, Ar), 6.89–6.96 (m, 7H, Ar), 6.31–6.36 (m, 1H, H1′), 4.73–4.75

(m, 3H, H3′, $PhOCH_2CO$), 4.33 (m, 2H, $ArOCH_2$), 4.18–4.24 (m, 3H, H4′ & CH_2OCO), 3.60–3.89 (m, 10H, CH_2OP, NCH, OCH_3), 3.40–3.50 (m, 2H, H5′), 2.76 (t, 1H, CH_2CN), 2.54–2.63 (m, 3H, H2′, CH_2CN), 2.35–2.42 (m, 1H, H2′), 1.79–1.81 (m, 4H, $(CH_2)_2$), 1.57–1.61 (m, 3H, $ArCH_3$), 1.17–1.23 (m, 9H, CH_3), 1.09–1.10 (m, 3H, CH_3). ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.28, 148.08. ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 170.00, 169.99, 168.6, 158.9, 158.2, 154.7, 144.9, 140.1, 140.0, 135.7, 135.64, 135.62, 135.58, 130.18, 130.15, 129.4, 128.2, 128.1, 127.9, 127.0, 126.9, 121.1, 117.9, 114.5, 113.1, 103.5, 103.4, 86.64, 86.60, 86.0, 85.9, 85.5, 85.3, 85.2, 73.1, 66.1, 64.3, 63.2, 63.0, 58.7, 58.5, 54.70, 54.68, 43.1, 43.0, 40.4, 40.3, 25.1, 25.0, 24.05, 24.03, 23.99, 23.9, 19.9, 19.84, 19.79, 13.6, 11.17, 11.16. IR (thin film); ν_{max} (cm^{-1}) = 3060, 2966, 1758, 1671, 1607, 1508, 1330, 1251, 1180, 1034, 755. HRMS (ESI-MS) *m*/*z* calculated for $C_{52}H_{63}N_4O_{11}PNa^+$ 973.4128: found973.4131, $[M + Na]^+$.

3′-*O*-(β-Cyanoethyl-*N*,*N*′-diisopropyl)-5′-*O*-(4,4′-dimethoxytrityl)-O^4-(phenoxyacetyloxyheptyl)-thymidine (7b)

6b (0.205 g, 0.242 mmol) and diisopropylethylamine (0.050 g, 0.390 mmol) were dissolved in THF (1 mL), followed by *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride (0.073 g, 0.310 mmol). The reaction was allowed to stir at room temperature for 30 min, where TLC revealed the presence of starting material. Diisopropylethylamine (0.006 g, 0.049 mmol) followed by *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride (0.011 g, 0.045 mmol) were added and the reaction was allowed to react another 30 min. Upon completion, the reaction was quenched by the addition of ethyl acetate (50 mL), and the solution was washed with sodium bicarbonate (3%, 2 × 25 mL) and brine (25 mL). The organic layer was dried over sodium sulfate, decanted, and evaporated. The product, a colorless powder, was precipitated from hexanes (0.079 g, 32.2%). R_f (SiO_2 TLC): 0.70, 0.83 hexanes–ethyl acetate (2 : 8). $\lambda_{max(ACN)}$ = 276 nm. ^1H NMR (500 MHz, d_6-acetone, ppm): 7.90–7.93 (s, 1H, H6), 7.46–7.51 (m, 2H, Ar), 7.24–7.39 (m, 9H, Ar), 6.89–6.97 (m, 7H, Ar), 6.31–6.36 (m, 1H, H1′), 4.72–4.75 (m, 3H, H3′, $PhOCH_2CO$), 4.29 (m, 2H, $ArOCH_2$), 4.16–4.25 (m, 3H, H4′ & CH_2OCO), 3.60–3.90 (m, 10H, CH_2OP, NCH, OCH_3), 3.40–3.52 (m, 2H, H5′), 2.76 (t, 1H, CH_2CN), 2.53–2.63 (m, 3H, H2′, CH_2CN), 2.34–2.42 (m, 1H, H2′), 1.71–1.77 (m, 2H, CH_2), 1.57–1.68 (m, 5H, CH_2 & $ArCH_3$), 1.37–1.44 (m, 6H, $(CH_2)_3$), 1.16–1.23 (m, 9H, CH_3), 1.09–1.10 (m, 3H, CH_3). ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.28, 148.08. ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 170.09, 170.08, 168.6, 158.9, 158.2, 154.7, 145.0, 140.0, 139.9, 135.7, 135.64, 135.62, 135.58, 130.18, 130.16, 129.4, 128.20, 128.15, 127.87, 127.86, 126.91, 126.88, 121.2, 118.1, 117.9, 114.5, 113.2, 103.54, 103.46, 86.64, 86.61, 86.0, 85.9, 85.51, 85.47, 85.3, 73.6, 73.5, 73.1, 66.6, 64.8, 64.6, 63.2, 63.0, 59.7, 58.7, 58.5, 54.71, 54.69, 43.1, 43.0, 40.4, 40.3, 25.7, 25.5, 24.1, 24.04, 24.01, 23.9, 19.91, 19.85, 19.8, 13.7, 11.20, 11.18, 11.14. IR (thin film); ν_{max} (cm^{-1}) = 3060, 2965, 1758, 1671, 1607, 1509, 1252, 1180, 1086, 1035, 755. HRMS (ESI-MS) *m*/*z* calculated for $C_{55}H_{69}N_4O_{11}PNa^+$ 1015.4597: found 1015.4620 $[M + Na]^+$.

Solid-phase oligonucleotide synthesis and purification

DNA duplexes (**XLTT4** and **XLTT7**, whose sequences are shown in Fig. 1) and single stranded DNA (control strands, O^4-butyl-4-ol-dT (**S4**) and O^4-heptyl-7-ol-dT (**S7**)), were assembled on an Applied Biosystems Model 3400 synthesizer on a 1 μmol scale employing standard β-cyanoethylphosphoramidite cycles as indicated by the manufacturer with modifications to certain coupling times as indicated below.

Commercially available 3′-*O*-2′deoxynucleoside phosphoramidites, which were protected with fast deprotecting groups, were dissolved in anhydrous acetonitrile at a concentration of 0.1 M, the modified 3′-*O*-2′-deoxyphosphoramidites (**7a** and **7b**) at 0.15 M and the modified bis-3′-*O*-2′-deoxyphosphoramidites (**4a** and **4b**) at 0.05 M. Assembly of sequences began with detritylation (3% trichloroacetic acid (TCA) in CH_2Cl_2), followed by phosphoramidite coupling: commercial 3′-*O*-2′-deoxyphosphoramidites (2 min), modified phosphoramidites (**7a** and **7b**) (10 min) and modified bis-phosphoramidite (**4a** and **4b**) (30 min); capping was achieved with phenoxyacetic anhydride–pyridine–tetrahydrofuran (1 : 1 : 8, v/v/v; solution A, and 1-methyl-1*H*-imidazole–tetrahydrofuran 16 : 84 w/v; solution B) and oxidation (0.02 M iodine in tetrahydrofuran–water–pyridine 2.5 : 2 : 1) which followed every coupling. Removal of the 3′-terminal trityl group was carried out by the synthesizer to yield the final oligonucleotide on the solid support.

The oligomer-derivatized CPG support was removed from the column and placed into screw cap microfuge tubes fitted with Teflon lined caps. The oligonucleotides containing O^4dT modifications were removed from the support by incubation of the CPG in 500 μL of a 10% 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) solution in the respective alcohol (1,4-butanediol for O^4-butyl-4-ol-dT and 1,7-heptanediol for O^4-heptyl-7-ol-dT), or ethanol for cross-linked oligonucleotides, for 5 days in the dark at room temperature. The DBU was neutralized with acetic acid and the DNA solubilised in acetonitrile prior to being transferred into clean vials to eliminate the CPG. The DNA was dried down in a speed vacuum and then desalted using C-18 SEP PAK cartridges (Waters Inc.) prior to purification.

The cross-link and mono-adduct final products were purified from pre-terminated products by 20% 7 M urea denaturing polyacrylamide gels (19 : 1) using 1X TBE [89 mM Tris–HCl, 89 mM boric acid, 2 mM EDTA (pH 8.0)] as running buffer to eliminate any residual DBU, which could not be removed by SAX-HPLC. Prior to gel separation, 25 O.D. of each sample were dried *via* a speed-vac concentrator and resuspended in 100 μL of formamide. The bands corresponding to the desired products were excised from the gels and placed into 15 mL Falcon tubes and the gel slices submerged in 0.1 M sodium acetate. The tubes were allowed to shake overnight and the extracted oligomers desalted using C-18 SEP PAK cartridges.

Oligonucleotide characterization by ESI-MS and nuclease digestion

ESI mass spectra for oligonucleotides were obtained at the Concordia University Centre for Biological Applications of Mass Spectrometry using a Micromass Qtof2 mass spectrometer

(Waters) equipped with a nanospray ion source. The mass spectrometer was operated in full scan, negative ion detection mode.

The modified oligomers (0.1 A_{260} units) were analyzed by exonuclease digestion (snake venom phosphodiesterase: 0.28 units and calf intestinal phosphatase: 5 units, in 10 mM Tris, pH 8.1 and 2 mM magnesium chloride) for a minimum of 16 h at 37 °C. The resulting mixtures were analyzed by reversed phase HPLC performed on a Symmetry® C-18 5 μm column (0.46 × 15 cm) purchased from Waters Inc, Milford, MA. The column was eluted with a linear gradient of 0–60% buffer B over 30 min (buffer A, 50 mM sodium phosphate, pH 5.8, 2% acetonitrile and buffer B, 50 mM sodium phosphate, pH 5.8, 50% acetonitrile). The retention times of the eluted peaks (see ESI‡) were compared to the standard nucleotides which eluted at the following times: dC (4.4 min), dG (7.5 min), dT (8.2 min), dA (9.2 min), O^6-MeG (12.0 min), O^4-MeT (13.4 min), O^4-butyl-4-ol-dT (15.3 min), O^4-heptyl-7-ol-dT (25.6 min) and the cross-link dimers (21.7 for the four carbon and 29.9 min for the seven carbon cross-link), and the ratio of nucleosides was determined. The molecular weights of the modified oligomers were determined by ESI-MS, which correlated with the calculated mass.

UV thermal denaturation of DNA duplexes

The nearest-neighbour method was used to determine the molar extinction coefficients for all oligonucleotides (with units of (M^{-1} cm^{-1})). An equimolar amount of the complementary strands were mixed together and lyophilized to dryness for both the mono-adduct and the non-cross-linked duplex. The DNA was subsequently resuspended in 90 mM sodium chloride, 10 mM sodium phosphate, 1 mM EDTA buffer (pH 7.0) to give a final duplex concentration of 4 μM (OD_{260} = 0.5). To ensure proper duplex formation the solution was heated to 90 °C where it was held for 10 min, cooled slowly to room temperature and stored at 4 °C overnight prior to use. Degassing of the samples was achieved using a speed-vac concentrator for 2 min. Denaturation curves were obtained by monitoring absorbance at 260 nm at a rate of heating of 0.5 °C min^{-1}, using a Varian CARY Model 3E spectrophotometer fitted with a 6-sample thermostated cell block and a temperature controller. Processing of the data was performed using the method published by Puglisi and Tinoco.[36]

Thermodynamic parameters of the non-cross-linked control DNA were obtained by performing thermal denaturation studies in triplicate at concentrations of 0.6, 2.7, 12.6 and 58.5 μM duplex DNA in 1, 0.5, 0.2 and 0.1 cm cuvettes, respectively. A plot of $\ln(C_{tot})$ as a function of $1/T_m$, typically used for a non-self-complementary bimolecular systems, was employed to obtain $\Delta H°$ and $\Delta S°$ for the system as described by Puglisi and Tinoco.[36] A similar procedure was employed with the cross-linked DNA at concentrations of 0.6 and 2.7 μM. The T_m of the ICL were independent of oligonucleotide concentration due to the uni-molecular nature of the DNA and therefore the thermodynamic parameters were obtained by plotting $\ln(K)$ *versus* $1/T$ at a single concentration of 2.95 μM. The experiments were performed in triplicate (see ESI‡).

Circular dichroism (CD) spectroscopy of DNA duplexes

Circular dichroism experiments were carried out on a Jasco J-815 spectropolarimeter equipped with a Julaba F25 circulating bath as previously reported.[14] Briefly, the samples used for the UV thermal denaturation studies were placed at 4 °C overnight and subsequently used for circular dichroism. The spectra were an average of 5 scans at a rate of 20 nm min^{-1}, with a bandwidth of 1 nm, sampling wavelength of 0.2 nm in a fused quartz cells (Starna 29-Q-10). Scans were performed between 320 and 220 nm at 10 °C. The molar ellipticity [θ] was obtained from the relation $\theta = \varepsilon/Cl$, where ε is the relative ellipticity (mdeg), C is the molar concentration of the DNA duplex (moles L^{-1}), and l is the path length in cm.

AGT preparation and purification

All proteins were purified under native conditions under the promoter of the pQE30 vector as described elsewhere.[20] The procedures for site directed mutagenesis and transformations into XL-10 Gold *E. coli* cells were conducted as stated in the Stratagene manual. Cells containing a plasmid coding for wild-type or mutant Ada-C, hAGT and OGT were grown in 1 L of LB broth + 100 μg mL^{-1} ampicillin until an OD_{600} = 0.6 was reached. Induction of the protein was achieved by adding IPTG at a final concentration of 0.3 mM. The cells were incubated at 37 °C for 4 h with shaking at 225 rpm, harvested by centrifugation at 6000 × *g* at 4 °C for 20 min. 5 mL of resuspension buffer [20 mM Tris–HCl (pH 8.0), 250 mM NaCl, 20 mM β-mercaptoethanol supplemented with Complete, Mini, EDTA-free Protease Inhibitor Cocktail Tablets] was added for every gram of wet pellet, and the cells resuspended using a dounce homogenizer. Cell lyses was carried out using two passes on the French press followed by centrifuged at 17 000 × *g* for 45 min at 4 °C. The cell lysate was introduced on a Ni-NTA column pre-equilibrated with equilibration buffer (resuspension buffer lacking the Complete, Mini, EDTA-free Protease Inhibitor). After extensive washing of the column with the equilibration buffer supplemented with 20 mM imidazole the protein was eluted with equilibration buffer supplemented with 200 mM imidazole. Fractions having protein content were pooled and dialyzed against dialysis buffer [50 mM Tris–HCl (pH 7.6), 250 mM NaCl, 20 mM β-mercaptoethanol and 0.1 mM EDTA] using 8000 Da cutoff dialysis tubing. In general a yield of 8–10 mg of protein was obtained per L of culture.

Mono-adduct repair assay

A general oligonucleotide sequence was created with a BclI cut site to monitor the amount of repair of mono-adduct based on restriction digestion. The sequence of the damage containing strand was 5′ GGC TXG ATC ACC AG where: X represents dT (control sequence), O^4MeT, O^4-butyl-4-ol-dT (**S4**) or O^4-heptyl-7-ol-dT (**S7**). The sequence of the complement strand was 5′ CTG GT/G ATC AAG CC where the "/" represent the BclI cut site.

The damage containing strand was ^{32}P labelled at its 5′ end as previously described.[14] Briefly, a 20 μM solution of DNA was made in 1X PNK buffer along with 1 μL [γ-^{32}P]ATP

(10 μCi μL^{-1}) and 5 units of T4 PNK. The labelling reaction was conducted for 1 h at 37 °C after which the reaction was terminated by boiling the sample for 10 min.

40 pmol of labelled DNA was added to 50 pmol of the complement strand in a total volume of 20 μL of water making a 2 μM dsDNA solution with 20% excess of the non-damaged strand. The solution was boiled for 5 min, cooled slowly to room temperature and placed at 4 °C overnight to ensure proper duplex formation.

The repair reaction mixtures were constituted of 2 pmol of the duplex DNA and 10 pmol of AGT in a total volume of 15 μL in Activity Buffer [10 mM Tris–HCl (pH 7.6), 100 mM NaCl and 1 mM DTT] and allowed to react at 37 °C for 30 min, unless otherwise indicated. The reaction was terminated by boiling for 10 min. Prior to BclI digestion, $MgCl_2$ was added at a final concentration of 10 mM and the reaction mixture allowed to cool to room temperature. 7.5 units (0.5 μL) of BclI was added to the mixture and the solution incubated for 1 h at 37 °C. 24 μL of stop buffer [81 mM Tris–HCl, 81 mM boric acid, 1.8 mM EDTA and 1% SDS (sodium dodecyl sulfate) (pH 8.0) in 80% formamide] was added, boiled for 10 min and loaded on a 14 cm × 16 cm, 20% 7 M urea denaturing polyacrylamide gel (19 : 1) for separation. The gels were run using 1X TBE for 40 min at 400 V and the gels exposed to a storage phosphor screen. The image was captured on a Typhoon 9400 (GE Healthcare, Piscataway, NJ) and the autoradiography counts obtained by ImageQuant™ (Amersham Biosciences).

C145S hAGT variant binding studies with mono-adduct substrates

Binding reactions of the 14 bp mono-adduct DNA substrates consisted of 0.5 nM dsDNA (where an excess of 5% of undamaged strand was present to force duplex formation) and increasing C145S hAGT (ranging from 1 to 35.69 μM) in a total solution volume of 20 μl of binding buffer [10 mM Tris–HCl (pH 7.6), 100 mM NaCl, 1 mM DTT, 10 μg mL^{-1} BSA and 2.5% glycerol]. The samples were equilibrated for 30 min at room temperature and loaded on a 10% native polyacrylamide gel (75 : 1) and pre-ran with 10 mM Tris-Acetate (pH 7.6) and 100 mM NaCl. Electrophoresis was carried out at 21 °C at 100 V for 45 min.

Cross-link repair assay

Labelling the 5′ nucleosides of the cross-linked oligonucleotides **XLTT4** and **XLTT7** as well as the control DNA (non-cross-linked duplex) were performed as described above. Repair assays were performed with 60 pmol of AGT protein incubated with 2 pmol of labelled DNA in a total reaction volume of 15 μL made up of Activity Buffer. The samples were incubated at 37 °C for 16 h. The reaction was ended by adding 18.2 μL of stop reaction buffer and the mixture separated on a 17% 7 M urea denaturing polyacrylamide gel (19 : 1). The gels were run using 1X TBE for 1.5 h at 700 V and the gels exposed to a storage phosphor screen. The image was captured on a Typhoon 9400 (GE Healthcare, Piscataway, NJ) and the radioactive counts quantified by ImageQuant™ (Amersham Biosciences).

C145S hAGT variant binding studies with cross-linked substrates

Binding reactions were conducted as for the mono-adduct substrates with minor modification due to the different substrate size. Binding reactions consisted of 0.5 nM dsDNA, increasing C145S AGT ranging from 1 to 45.5 μM for the control duplex and from 1 to 6.5 μM for the cross-linked DNA in a total solution volume of 20 μl of binding buffer. The samples were equilibrated for 1 h at room temperature and loaded on a 15% native polyacrylamide gel (75 : 1) and ran with 0.5X TBE, which allowed the electrophoresis voltage to be increased without heating the system. Electrophoresis was carried out at 21 °C at 250 V for 25 min.

The monomeric dissociation constant (K_d) and stoichiometry (n) of AGT binding to DNA were obtained from the electrophoretic mobility shift assays as described previously.[14] The binding of n molecules of AGT protein [P] to 1 molecule of DNA [D] can be expressed by the equation: $n\mathrm{P} + \mathrm{D} \leftrightarrow \mathrm{P}n\mathrm{D}$. Taking the logarithm and rearranging the variables yields: $\log([\mathrm{P}n\mathrm{D}]/[\mathrm{D}]) = n \log[\mathrm{P}]\mathrm{free} + \log K_d$.

The data can be expressed with a plot of $\log [\mathrm{P}n\mathrm{D}]/[\mathrm{P}]$ as a function of $\log [\mathrm{P}]$. The slope of the graph represents the stoichiometry (n) and the observed K_d can be obtained by the relationship: $K_d = 10^{-(y-\mathrm{intercept})}$. The monomeric K_d shown in the tables was obtained by taking the nth root of the observed K_d.

Modelling of O^4MeT and O^4-butyl-4-ol-dT in hAGT active site

The alkylated DNA was placed in the active site of hAGT based on PDB entry 1t38 by mutating O^6MeG to O^4MeT. The placement of O^4MeT was verified with PDB entry 1yfh. PDB entry 1t38 was used as the template for hAGT where Ser145 was mutated back to Cys and the proper rotamer used as based on PDB entries: 1t39, 1eH6 and 1yfh. This primary model was used as the template for the addition of the butyl-4-ol adduct, which was added manually by PyMOL.

The preliminary models underwent conjugate gradient minimization, simulated annealing, and torsion angle dynamics using Crystallography & NMR System.[37] CHARMM (Chemistry at HARvard Macromolecular Mechanics) topology files were modified to use the altered nucleotides by adding restraints to keep the α-carbon of the lesion in a *syn* conformation with respect to the N^3 atom of the base, as observed in PDB entries: 1t38, 1t39, and 1yfh.[38]

The models were subjected to 2 sets of 100 steps of gradient minimization and 200 steps of simulated annealing. The first cycle was carried out at 500 K with slow cooling at a rate of 4.5 K per cycle of dynamics. The second cycle was carried out at 350 K with slow cooling at a rate of 2.5 K per cycle of dynamics. The minimized models were displayed using PyMol with the absence of a few amino acids for clarity of the active site.

Conclusions

DNA duplexes containing an opposed O^4-2′-deoxythymidine-alkyl-O^4-2′-deoxythymidine ICL were synthesized using a convertible nucleoside strategy to prepare a cross-linked nucleoside

dimer that was converted into a bis-phosphoramidite. Solid phase synthesis was employed to incorporate the cross-link into a specific orientation and position in a DNA duplex. This method is versatile and allows for various linker lengths to be introduced for a variety of studies including DNA repair. ICL duplexes containing a four and seven carbon linker evaded repair from various AGT proteins from human and *E. coli*.

Studies to investigate AGT proteins that can be engineered to repair these ICL and their repair by other systems are ongoing to enhance our understanding of how ICL lesions may be processed. Furthermore this may result in increasing the potency of bifunctional chemotherapeutic agents which create ICL lesions and hopefully limit their mutagenicity and other side effects such as the onset of resistance to these drugs.

Acknowledgements

We are grateful to Dr Anthony E. Pegg (Penn. State University) for the plasmids encoding the wild-type Ada-C, OGT and hAGT genes. This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canada Foundation for Innovation (CFI) and the Canada Research Chair (CRC) program. Both FPM and DKO were recipients of postgraduate fellowships from NSERC.

Notes and references

1 (*a*) N. Magana-Schwencke, J. A. Henriques and R. E. Moustachhi, *Proc. Natl. Acad. Sci. U. S. A.*, 1982, **79**, 1722–1726; (*b*) P. D. Lawley and D. H. Phillips, *Mutat. Res., Fundam. Mol. Mech. Mutagen.*, 1996, **355**, 13–40; (*c*) S. R. Rajski and R. M. Williams, *Chem. Rev.*, 1998, **98**, 2723–2796.
2 (*a*) M. L. Dronkert and R. Kanaar, *Mutat. Res.*, 2001, **486**, 217–247; (*b*) P. J. McHugh, V. J. Spanswick and J. A. Hartley, *Lancet Oncol.*, 2001, **2**, 483–490; (*c*) M. Christmann, M. T. Tomicic, W. P. Roos and B. Kaina, *Toxicology*, 2003, **193**, 3–34; (*d*) A. Sancar, L. A. Lindsey-Boltz, K. Unsal-Kacmaz and S. Linn, *Annu. Rev. Biochem.*, 2004, **73**, 39–85.
3 (*a*) D. M. Noll, T. M. Mason and P. S. Miller, *Chem. Rev.*, 2006, **106**, 277–301; (*b*) K. M. McCabe, S. B. Olson and R. E. Moses, *J. Cell. Physiol.*, 2009, **220**, 569–573; (*c*) T. V. Ho and O. D. Schärer, *Environ. Mol. Mutagen.*, 2010, **51**, 552–566.
4 A. Guainazzi and O. D. Schärer, *Cell. Mol. Life Sci.*, 2010, **67**, 3683–3697.
5 (*a*) K. Stevens, D. D. Claeys, S. Catak, S. Figaroli, M. Hocek, J. M. Tromp, S. Schürch, V. Van Speybroeck and A. Madder, *Chem.–Eur. J.*, 2011, **17**, 6940–6953; (*b*) K. Hattori, T. Hirohama, S. Imoto, S. Kusano and F. Nagatsugi, *Chem. Commun.*, 2009, 6463–6465; (*c*) T. Angelov, A. Guainazzi and O. D. Schärer, *Org. Lett.*, 2009, **11**, 661–664; (*d*) H. Huang, H. Y. Kim, I. D. Kozekov, Y. J. Cho, H. Wang and A. Kozekova, *et al.*, *J. Am. Chem. Soc.*, 2009, **131**, 8416–8424; (*e*) S. Hentschel, J. Alzeer, T. Angelov, O. D. Schärer and N. W. Luedtke, *Angew. Chem., Int. Ed.*, 2012, **51**, 1–5.
6 E. M. Hlavin, M. B. Smeaton, A. M. Noronha, C. J. Wilds and P. S. Miller, *Biochemistry*, 2010, **49**, 3977–3988.
7 (*a*) J. O. Ojwang, D. A. Grueneberg and E. L. Loechler, *Cancer Res.*, 1989, **49**, 6529–6537; (*b*) Q. Dong, D. Barsky, M. E. Colvin, C. F. Melius, S. M. Ludeman and J. F. Moravek, *et al.*, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 12170–12174; (*c*) R. T. Streeper, R. J. Cotter, M. E. Colvin, J. Hilton and O. M. Colvin, *Cancer Res.*, 1995, **55**, 1491–1498.
8 S. Kallama and K. Hemminki, *Chem.-Biol. Interact.*, 1986, **57**, 85–96.
9 C. J. Wilds, J. D. Booth and A. M. Noronha, *Tetrahedron Lett.*, 2006, **47**, 9125–9128.
10 A. E. Pegg, *Mutat. Res.*, 2000, **462**, 83–100.
11 D. S. Daniels, T. T. Woo, K. X. Luu, D. M. Noll, N. D. Clarke and A. E. Pegg, *et al.*, *Nat. Struct. Mol. Biol.*, 2004, **11**, 714–720.
12 K. S. Srivenugopal, X. H. Yuan, H. S. Friedman and F. Ali-Osman, *Biochemistry*, 1996, **35**, 1328–1334.
13 Q. Fang, A. M. Noronha, S. P. Murphy, C. J. Wilds, J. L. Tubbs, J. A. Tainer, G. Chowdhury, F. P. Guengerich and A. E. Pegg, *Biochemistry*, 2008, **47**, 10892–10903.
14 F. P. McManus, Q. Fang, J. D. Booth, A. M. Noronha, A. E. Pegg and C. J. Wilds, *Org. Biomol. Chem.*, 2010, **8**, 4414–4426.
15 (*a*) A. S. Waldman, *BioEssays*, 2008, **30**, 1163–1171; (*b*) M. F. Goodman, S. Creighton, L. B. Bloom and J. Petruska, *Crit. Rev. Biochem. Mol. Biol.*, 1993, **28**, 83–126; (*c*) T. A. Kunkel and K. Bebenek, *Annu. Rev. Biochem.*, 2000, **69**, 497–529; (*d*) T. Lindahl, *Nature*, 1993, **362**, 709–715.
16 A. Bhattacharyya and D. M. Lilley, *J. Mol. Biol.*, 1989, **209**, 583–597.
17 B. D. Preston, B. Singer and L. A. Loeb, *Proc. Natl. Acad. Sci. U. S. A.*, 1986, **83**, 8501–8505.
18 P. Rojsitthisak, N. Jongaroonngamsang, R. M. Romero and I. S. Haworth, *PLoS One*, 2011, **6**, e20745.
19 Y. Z. Xu and P. F. Swann, *Nucleic Acids Res.*, 1990, **18**, 4061–4065.
20 M. E. Dolan and A. E. Pegg, *Carcinogenesis*, 1985, **6**, 1611–1614.
21 S. S. Hecht, *Chem. Res. Toxicol.*, 1998, **11**, 559–603.
22 (*a*) G. T. Pauly and R. C. Moschel, *Chem. Res. Toxicol.*, 2001, **14**, 894–900; (*b*) K. B. Altshuler, C. S. Hodesand and J. M. Essigmann, *Chem. Res. Toxicol.*, 1996, **9**, 980–987.
23 J. C. Klein, M. J. Bleeker, H. C. Roelen, J. A. Rafferty, G. P. Margison and H. F. Brugghe, *et al.*, *J. Biol. Chem.*, 1994, **269**, 25521–25528.
24 (*a*) P. M. Potter, M. C. Wilkinson, J. Fitton, F. J. Carr, J. Brennand and D. P. Cooper, *et al.*, *Nucleic Acids Res.*, 1987, **15**, 9177–9193; (*b*) S. R. Paalman, C. Sung and N. D. Clarke, *Biochemistry*, 1997, **36**, 11118–11124; (*c*) M. C. Wilkinson, P. M. Potter, L. Cawkwell, P. Georgiadis, D. Patel and P. F. Swann, *et al.*, *Nucleic Acids Res.*, 1989, **17**, 8475–8484; (*d*) Q. Fang, S. Kanugula, J. L. Tubbs, J. A. Tainer and A. E. Pegg, *J. Biol. Chem.*, 2010, **285**, 8185–8195.
25 Q. Zhu, M. O. Delaney and M. M. Greenberg, *Bioorg. Med. Chem. Lett.*, 2001, **11**, 1105–1107.
26 (*a*) C. J. Wilds, F. Xu and A. M. Noronha, *Chem. Res. Toxicol.*, 2008, **21**, 686–695; (*b*) C. J. Wilds, E. Palus and A. M. Noronha, *Can. J. Chem.*, 2007, **85**, 249–256.
27 J. D. Booth, S. P. Murphy, A. M. Noronha and C. J. Wilds, *Nucleic Acids Symp. Ser.*, 2008, **52**, 431–432.
28 C. J. Wilds, A. M. Noronha, S. Robidoux and P. S. Miller, *J. Am. Chem. Soc.*, 2004, **126**, 9257–9265.
29 C. Hofr and V. Brabec, *Biopolymers*, 2005, **77**, 222–229.
30 C. Fabrega, R. Eritja, N. D. Sinha, M. K. Dosanjh and B. Singer, *Bioorg. Med. Chem.*, 1995, **3**, 101–108.
31 (*a*) W. C. Johnson Jr., *Methods Biochem. Anal.*, 1985, **31**, 61–163; (*b*) W. C. Johnson Jr., in *Circular Dichroism and the Conformational Analysis of Biomolecules*, ed. G. D. Fasman, Plenum Press, New York, 1996, p. 433.
32 M. W. Kalnik, M. Kouchakdjian, B. F. Li, P. F. Swann and D. J. Patel, *Biochemistry*, 1988, **27**, 100–108.
33 (*a*) C. A. Adams, M. Melikishvili, D. W. Rodgers, J. J. Rasimas, A. E. Pegg and M. G. Fried, *J. Mol. Biol.*, 2009, **389**, 248–263; (*b*) M. Melikishvili, J. J. Rasimas, A. E. Pegg and M. G. Fried, *Biochemistry*, 2008, **47**, 13754–13763.
34 M. G. Fried, S. Kanugula, J. L. Bromberg and A. E. Pegg, *Biochemistry*, 1996, **35**, 15295–15301.
35 D. S. Daniels, C. D. Mol, A. S. Arvai, S. Kanugula, A. E. Pegg and J. A. Tainer, *EMBO J.*, 2000, **19**, 1719–1730.
36 J. D. Puglisi and I. Tinoco Jr., *Methods Enzymol.*, 1989, **180**, 304–325.
37 A. T. Brunger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros and R. W. Grosse-Kunstleve, *et al.*, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 1998, **D52**, 30–42.
38 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187–217.

