A COMPUTATIONAL INVESTIGATION OF THE FORMATION AND STRUCTURE OF

DNA INTRASTRAND CROSS-LINKS INITIATED BY THE URACIL RADICAL

Cassandra D.M. Churchill
B.Sc., University of Lethbridge, 2009

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfillment of the
Requirements for the Degree

MASTER OF SCIENCE
in Chemistry

Department of Chemistry and Biochemistry
University of Lethbridge
Lethbridge, Alberta, Canada

© Cassandra D.M. Churchill, 2011

For my family.

Abstract

Using computational methods, the formation pathways and structures of four experimentally-observed DNA intrastrand cross-links are determined. These lesions originate from the uracil radical and are of particular importance due to their potential role in the activity of the 5-halouracils as radiosensitizing agents in anti-tumour treatments. The formation pathways are studied with density functional theory under conditions relevant to both UV and ionizing radiation. Results reveal these intrastrand cross-links are likely to form under therapeutic conditions and provide an explanation for their observed formation preferences. The structures of the cross-links in DNA are determined with molecular dynamics simulations to reveal the distortions these lesions induce to the helix. This has provided information about the potential biological implications of these lesions, where results indicate intrastrand cross-links are likely both mutagenic and cytotoxic. Therefore, these lesions may contribute to tumour cell death in the therapeutic use of halouracils.

Acknowledgements

I would like to thank my supervisor, Dr. Stacey Wetmore, who first introduced me to research and computational chemistry. She has provided me with this excellent opportunity, as well as guidance and support throughout the course of both my undergraduate and graduate careers. I would also like to thank our collaborator, Dr. Leif Eriksson, for his contributions to this research, as well as for the opportunity to visit him at the National University of Ireland – Galway.

Thank you to the members of my supervisory committee, Drs. René Boeré and Ute Wieden-Kothe, for their time, support and insights into my research project. I would also like to thank my external examiner, Dr. Tom Woo, for his willingness to share his expertise and interest in this thesis.

I would like to acknowledge the Natural Science and Engineering Research Council (NSERC), Alberta Innovates-Technology Futures, Alberta Scholarship Programs and the University of Lethbridge for student scholarships.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

8-oxoG	7,8-dihydro-8-oxoguanine
A	adenine
$A^{\bullet+}$	adenine radical cation
^{Br}U	5-bromouracil
$^{Br}U^{\bullet-}$	5-bromouracil radical anion
C	cytosine
^{Cl}U	5-chlorouracil
CPD	cyclobutane pyrimidine dimer
DEA	dissociative electron attachment
DFT	density functional theory
DNA	deoxyribonucleic acid
DSB	double-strand break
dsDNA	double-stranded DNA
^{F}U	5-fluorouracil
G	guanine
$G^{\bullet+}$	guanine radical cation
GAFF	general AMBER forcefield
GGA	generalized gradient approximation
Gy	gray, SI unit of radiation
IC	internal conversion
IEF-PCM	integral equation formalism variant of the polarizable continuum model; implicit solvation model
ISC	intersystem crossing
^{I}U	5-iodouracil
mC	5-methylcytosine
MD	molecular dynamics
MM	molecular mechanics
NER	nucleotide excision repair
ONIOM	our own N-layered integrated molecular orbital + molecular mechanics; hybrid method
P	pseudorotational phase angle
PSET	photoinduced single electron transfer

QM	quantum mechanics
QM/MM	quantum mechanics/molecular mechanics; hybrid method
RNA	ribononucleic acid
SCF	self-consistent field
SSB	single-strand break
ssDNA	single-stranded DNA
T	thymine
$T^{\bullet-}$	thymine radical anion
TLS	translesion synthesis
U	uracil
$U^{\bullet}$	uracil-5-yl radical
^{X}U	5-halouracils (including ^{Cl}U, ^{Br}U and ^{I}U)
ZPVC	zero-point vibrational correction

Chapter 1. Introduction

1.1.   General Background

Free radicals are commonly described as highly reactive compounds that contain one or more unpaired electrons (a non-zero spin).[1] However, an alternative definition also describes radicals as "any transient species (atom, molecule or ion) [with] a short lifetime in the gaseous phase under ordinary laboratory conditions", which includes reactive species with zero spin (such as singlet oxygen and singlet carbene) and excludes stable species with non-zero spin (such as triplet oxygen and nitric oxide).[1] These chemically-reactive species have important implications in many fields such as atmospheric chemistry, combustion and cellular biology.

Within our cells, radicals can be advantageous and are necessary in the activation of certain enzymes, signal transduction and oxidative stress responses.[2] However, radicals are also prone to initiating unwanted side reactions. Deoxyribononucleic acid, commonly known as DNA, is susceptible to a variety of detrimental reactions with radicals that form as a result of cellular processes (for example, as by-products of metabolism) or external sources (for example, radiation). These reactions have been shown to form many different lesions in DNA that contribute to cell death, aging, and disease.[2,3]

Although radicals have been implicated to cause cancer, the controlled creation of radicals can be used to kill unwanted cells. This is the approach employed in the therapeutic use of radiation to minimize or eliminate tumours. A recent strategy of this type involves the direct creation of DNA radicals, which react with other DNA components to damage DNA and thereby induce cell death. This thesis uses molecular modeling to examine the formation mechanism and structure of damaging DNA intrastrand cross-links that are

generated by the uracil-5-yl radical, as well as their potential to contribute to the death of tumour cells.

1.2. DNA Overview

DNA contains the information necessary for life. DNA is universal to all living organisms and stores information in genes through the sequencing of its components. From these genes, ribonucleic acid (RNA) and proteins are constructed to carry out cell functions.

DNA is a biopolymer comprised of repeating monomeric units called nucleotides (Figure 1–1(a)).[4,5] Each nucleotide consists of a phosphate group and a nucleoside, where a nucleoside is composed of a 2′-deoxyribose sugar and a nucleobase (Figure 1–1(a, b)). Each monomeric unit contains a glycosidic linkage (C1′–N9 or C1′–N1) between the 2′-deoxyribose sugar and one of the four nitrogenous DNA nucleobases: adenine (A), thymine (T), guanine (G) and cytosine (C) (Figure 1–1). The uracil (U) nucleobase is naturally found in RNA and is structurally similar to T, but differs in the presence of hydrogen at C5 rather than a methyl group. The purine nucleobases, A and G, are comprised of two fused rings, while the pyrimidines, C and T (U), consist of one ring. The 5′ end of each nucleotide is linked to the 3′ end of the previous nucleotide via a phosphodiester bond to create a DNA strand involving a deoxyribose-phosphate backbone with a 5′ to 3′ directionality (Figure 1–1(c)).

DNA adopts a double-stranded helical structure in which the two strands run in anti-parallel directions. Within this complex, each nucleobase is engaged in Watson-Crick hydrogen bonds with its complementary base in the opposite strand to form G···C and A···T (A···U) base pairs (Figure 1–1(a, b)). Our genes are encoded within the sequencing of these

2
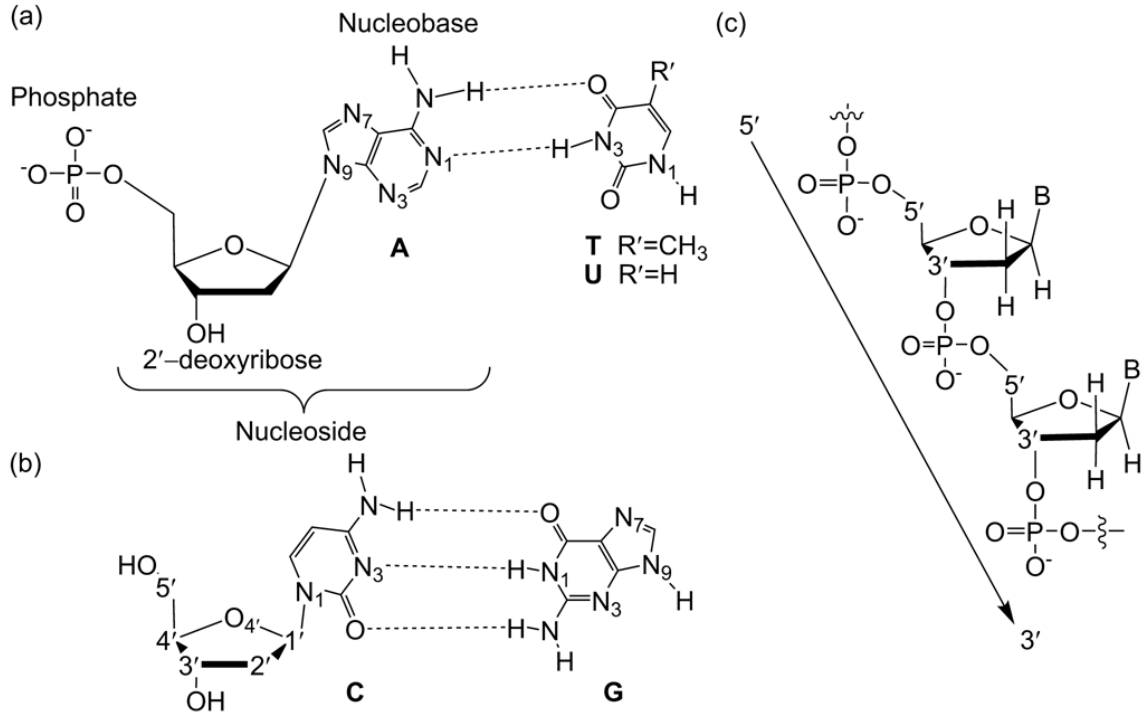
Figure 1–1: Illustration of the structure and numbering of DNA components. (a) A nucleotide unit (phosphate, 2′-deoxyribose and nucleobase) showing the Watson-Crick hydrogen bonding between A and T/U. (b) A nucleoside unit showing the Watson-Crick hydrogen bonding between C and G. (c) An illustration of the polymeric nature of DNA showing the deoxyribose-phosphate backbone, nucleobases (B) and strand direction.

base pairs, with specific hydrogen bonds ensuring the integrity of the genetic code. Hydrogen bonds, in conjunction with inter- and intrastrand π–π stacking interactions between the nucleobases, contribute to the stability of the helix. These non-covalent interactions have the benefit of being easily reversible, which allows the DNA strands to separate and unwind during biological processes, such as replication and transcription.

Due to the number of degrees of freedom in double-stranded DNA (dsDNA), this macromolecule can adopt many different conformations. Under biological conditions, DNA adopts the B-conformation (B-DNA), which is also observed in X-ray crystal structures.[4,5] This average conformation consists of a right-handed double helix. The backbone of each

nucleotide is described by six torsion angles (α, β, γ, δ, ε and ζ, Figure 1–2(a)), the average values of which are well-established for B-DNA,[6] and the deoxyribose moiety adopts C2′–*endo* puckering (Figure 1–2(a)). Each nucleobase is in the *anti* orientation ($\chi = 180 \pm 90°$, Figure 1–2(a))[4] about the glycosidic bond ensuring the Watson-Crick face is available for hydrogen bonding with the complementary nucleobase in the opposite strand. Alternatively, if the *syn* ($\chi = 0 \pm 90°$)[4] orientation is present in DNA, Watson-Crick hydrogen bonding cannot occur. Instead the Hoogsteen face is exposed to the complementary strand, which may create a mispair upon replication and induce a mutation. Grooves, designated as major (width = 11.7 Å, depth = 8.5 Å) and minor (width = 5.7 Å, depth = 7.5 Å), are present between the strands (Figure 1–2(b)),[5] and play an important role in initiating DNA–protein interactions.[7]
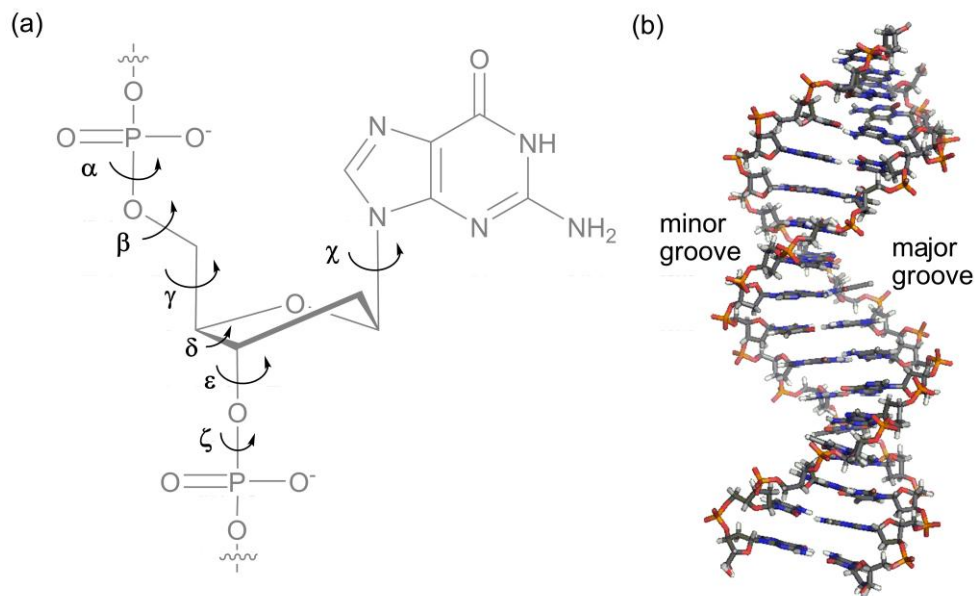
Figure 1–2: Illustration of select structural parameters used to define conformations of DNA, including (a) important torsion angles within a nucleotide and (b) the major and minor grooves of a DNA helix.

1.3. Modified DNA

1.3.1. Common Forms of DNA Damage

The structure of DNA is prone to damage from both natural processes (for example, nucleobase deamination and hydrolysis) and environmental agents (for example, UV sunlight, X-rays).[8-12] As the carrier of genetic material, modifications to DNA components have important biological repercussions. Although mechanisms have evolved to repair many forms of DNA damage,[13] some repair pathways are prone to errors and lesions are not always eliminated.[14] While some mutations are responsible for genetic variation and evolution,[14] others result in disease and/or cause cell death.

The following sections will discuss the effects of UV and ionizing radiation on DNA, and the resulting lesions. Radiation is of particular interest due to its ability to generate radicals and excited states that result in high yields of DNA damage.[15,16] Furthermore, these forms of radiation are responsible for the production of the uracil radical that leads to a variety of DNA lesions, including intrastrand cross-links, which have anti-tumour capabilities.

1.3.1.1. Exposure of DNA to UV radiation

Absorption of a UV photon by a molecule results in the excitation of an electron from the ground state. When an electron within DNA is excited, π-orbital overlap between the nucleobases allows for charge delocalization and charge transfer states to occur.[17] Specifically, absorption of a UV photon by DNA results in the creation of a delocalized excitation over several stacked nucleobases within a strand.[18] After a few picoseconds, this excited state localizes to a charge-separated dimer state, which creates a π–radical cation (hole) and a π–radical anion (electron).[19] Although the extent of charge separation depends on the identity and redox potential of the two bases,[19] the longest-lived and most stable

charge separation occurs between a guanine radical cation (G$^{\bullet+}$) due to the low ionization potential of G,[20] and a thymine radical anion (T$^{\bullet-}$) due to the high electron affinity of T.[21]

To avoid DNA damage from excited states, the charge-separated dimer undergoes electron-hole recombination.[22] This arises through an internal conversion (IC)[i] process in which a transition occurs from the first singlet excited state to the singlet ground state through conical intersections (potential energy surface degeneracies) that are easily accessible from the excited state through structural distortions.[22] When transferred to the ground state, the molecule is in an excited vibrational state, where the bases return to equilibrium by vibrationally transferring energy to surrounding solvent molecules.[23] In this way, DNA resists UV damage.

Despite the intrinsic IC relaxation mechanism of DNA, UV radiation is known to damage DNA. For example, alternate states may become populated through intersystem crossings (ISC),[ii] which create long-lived triplet states that are precursors for certain DNA lesions.[24] Alternatively, damage may also originate directly from an IC. The most commonly-formed UV-lesion is the cyclobutane pyrimidine dimer (CPD) involving thymine, which forms from the cycloaddition of the C5–C6 double bonds of adjacent Ts to create a cyclobutane ring (Figure 1–3(a)). This product originates from the first singlet excited state, where an IC enables ultrafast dimer formation.[23] However, since dimerization is faster than base motion, the two adjacent Ts must be in precisely the right conformation upon UV irradiation to form CPDs.[23,25] Otherwise, the IC will return DNA to its undamaged, pre-irradiation state. Enzymes have evolved to repair CPDs. For example, the photolyase enzyme is present in many organisms and utilizes visible light to photocleave the

---

[i] Internal conversion: transition from a higher electronic state to a lower electronic state in which the spin state remains the same (example: from S_1 to S_0).
[ii] Intersystem crossing: transition from a higher electronic state to a lower electronic state in which the spin state changes (example: from T_1 to S_0).

interpyrimidine bonds.[26,27] Alternatively, CPDs are removed from human DNA by nucleotide excision repair (NER) pathways through cleavage the DNA backbone.[26] Nevertheless, this lesion is not always repaired and has been directly linked to skin cancer.[27]

As discussed above, the most stable charge-separated states involve $G^{\bullet+}$ and $T^{\bullet-}$. If charge recombination does not occur, these species may cause DNA damage through subsequent reactions. Indeed, these radical ions are precursors to a variety of DNA lesions. For example, $G^{\bullet+}$ can react with water or oxygen to form mutagenic 7,8-dihydro-8-oxoguanine (8-oxoG, Figure 1–3(b)).[28,29] Generation of $T^{\bullet-}$ can cleave the glycosidic bond (initiate base release) or the C3′–O3′ or C5′–O5′ bonds (initiate strand breaks).[30]
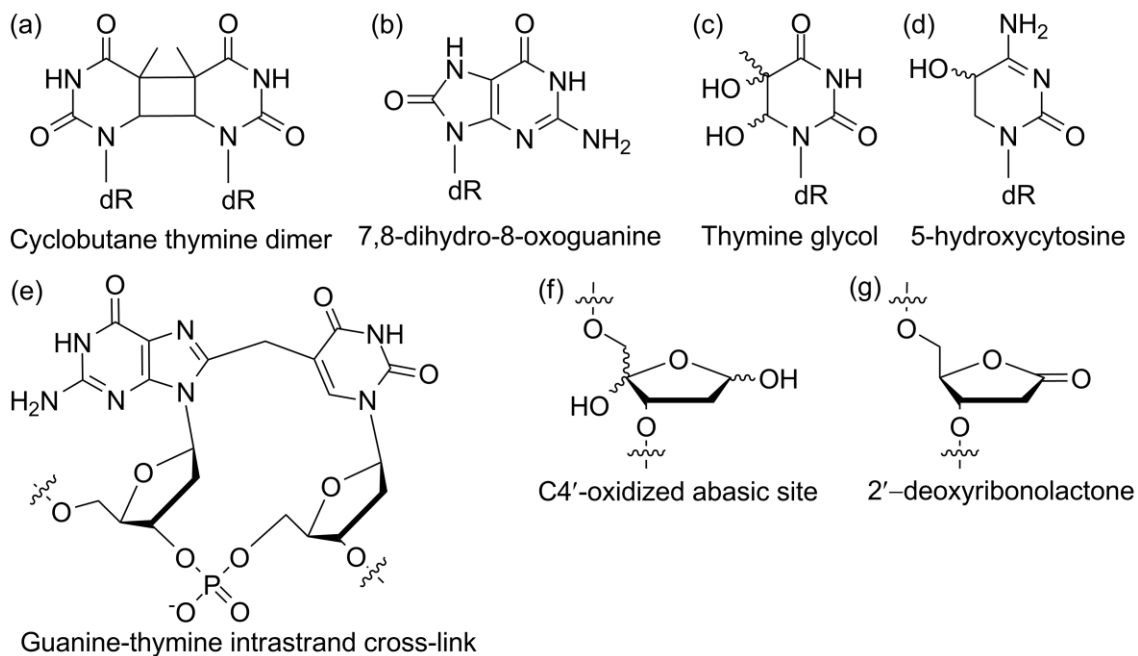
Figure 1–3: Common DNA lesions resulting from exposure of DNA to UV and/or ionizing radiation.

1.3.1.2. Exposure of DNA to Ionizing Radiation

A photon of ionizing radiation (for example, γ- or X-rays) contains enough energy to remove an electron from (ionize) a molecule. Since ionizing radiation is absorbed by the components of a mixture in proportion to their weight contributions, the majority of energy is absorbed by water upon cellular exposure.[31] Therefore, the resulting damage to DNA is indirect and caused by the radicals produced from the radiolysis of water (hydroxyl radicals (•OH), solvated electrons ($e^-_{(aq)}$) and hydrogen atoms (H•), Figure 1–4).

$$H_2O \xrightarrow{\text{ionizing radiation}} H_2O^{•+} + e^-$$

$$H_2O \xrightarrow{\text{ionizing radiation}} H_2O^*$$

$$H_2O^{•+} \longrightarrow {}^•OH + H^+$$

$$e^- + n\,H_2O \longrightarrow e^-_{(aq)}$$

$$H_2O^* \longrightarrow {}^•OH + H^•$$
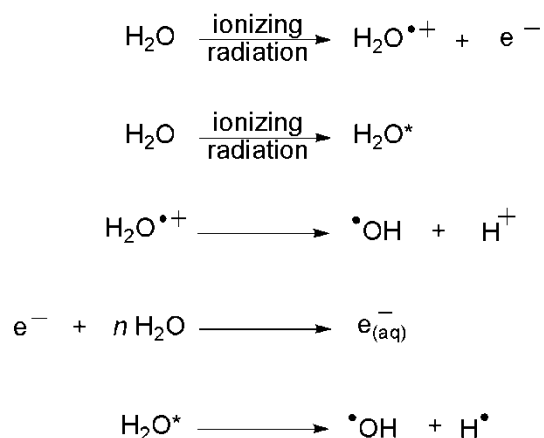
Figure 1–4: Reactions resulting from the exposure of water to ionizing radiation. See Ref. 31

Cells are protected from the harmful effects of these radicals by antioxidants, where •OH cannot travel more than 2 Å before being scavenged.[31] Nevertheless, •OH can cause DNA damage. For example, •OH easily undergoes addition to double bonds. This can occur at C5 of a pyrimidine to form a C6-centered radical, which subsequently leads to a pyrimidine glycol (Figure 1–3(c)) and 5-hydroxypyrimidine (Figure 1–3(d)).[10] Alternatively, •OH can add to C8 of G to create an N7-centred radical, which forms 8-oxoG (Figure 1–3(b)).[10,12] The hydroxyl radical also participates in hydrogen abstraction reactions, with pyrimidine nucleobases being particularly susceptible. For example, hydrogen abstraction at the C5

8

methyl group of T may produce intrastrand cross-links (Figure 1–3(e)).[10] The •OH species may also attack the 2′-deoxyribose moiety to create abasic sites (Figure 1–3(d)) and 2′-deoxyribonolactone (Figure 1–3(g)), which can subsequently form detrimental strand breaks.[12]

The solvated electron ($e^-_{(aq)}$) is the most powerful reducing agent and has a substantial lifetime compared to other (•OH, H•) radical species, which allows it to persist in a cell and produce a variety of lesions.[31] This electron efficiently attaches to DNA to form π–radical anions localized on pyrimidines, most commonly thymine (T•−), which subsequently breaks the glycosidic or a C–O backbone bond.[30] Like •OH, the hydrogen atom (H•) prefers to add to the C5–C6 double bond of pyrimidines, but hydrogen abstraction by H• is unlikely.[31]

1.3.2. *Intentional DNA Damage*

While DNA damage is commonly considered undesirable and much research has focused on the repair of detrimental lesions,[13,26,32] DNA damage can be harnessed as a tool to kill unwanted tumour cells. Many current anti-cancer treatments employ this strategy in an attempt to selectively induce apoptosis in tumour cells.[33-35] While repair mechanisms have evolved to overcome common forms of DNA damage, complex lesions such as double-strand breaks, tandem lesions (adjacent damaged nucleotides in the same strand) and clustered lesions (multiple damage sites within 20 base pairs) have shown particular promise in anti-tumour therapies due to their disruption of DNA structure and difficulty to repair.[15,16,31,33] When coupled with the disruption of the normal functioning of DNA repair pathways in tumour cells, this makes complex DNA lesions ideal anti-cancer agents. [15,16,36] However, the search for new and efficient cancer treatments is never-ending, and is fuelled

by both the side-effects and resistances that accompany many drugs. The following sections examine well-known and efficient anti-cancer agents.

1.3.2.1. Cisplatin

Cisplatin, the common name for *cis*-diaminodichloroplatinum(II) ($PtCl_2(NH_3)_2$), is a widely-used anti-tumour therapy that lethally damages DNA and stops fundamental cellular processes. Cisplatin consists of a platinum(II) centre coordinated to two Cl^- ions and two NH_3 groups in a square planar geometry. Under cellular conditions, cisplatin undergoes aquation to produce the reactive dication, $Pt(H_2O)_2(NH_3)_2^{2+}$ (Figure 1–5(a)).[34,37,38] This species attacks N7 of purine nucleobases in DNA to form intrastrand cross-links with the major product (65%) being the 1,2–GG lesion (Figure 1–5(b)), followed by 1,2–AG (25%) and 1,3–GG (5-10%), and a small portion of interstrand cross-links also being formed.[38,39] The major GG lesion is known to induce a bend of 40° towards the major groove.[39,40] The recognition of this DNA damage involves over 20 different proteins, which ultimately ends in cell apoptosis (programmed cell death).[34]
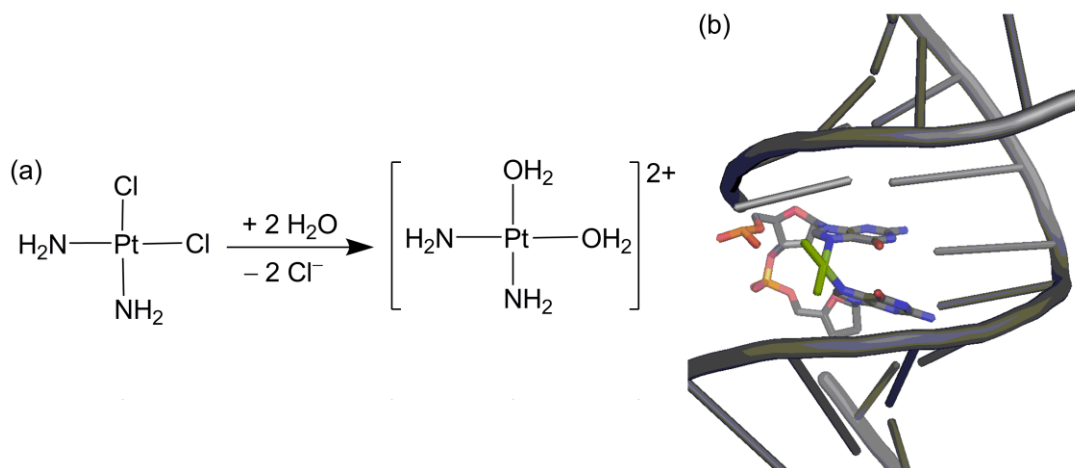
Figure 1–5: (a) The cellular activation of cisplatin to the reactive aqua species and (b) the structure of the major 1,2–GG intrastrand cross-link with cisplatin (green) bonded to N7 of adjacent guanine nucleobases. Crystal structure geometry (PDB: 3LPV) from Ref. 39

The therapeutic use of cisplatin has been faced with challenges. Side effects occur due to cell death in normal tissues at drug concentrations required to maintain effectiveness in cancerous tissues. Second-generation platinum drugs like carboplatin improve upon cisplatin in this respect by minimizing side effects,[31,32] but tumours are prone to develop resistances after prolonged exposure to carboplatin.[37] Alternatively, oxaliplatin has provided increased sensitivity in resistant cells and minimized toxicity in normal tissues.[34] However, despite these apparent successes, platinum-based compounds are only prescribed for a small range of tumours in humans.[37] Therefore, the search is ongoing for improved therapeutic strategies.

1.3.2.2. Radiation

Radiation has been used as an anti-tumour treatment since the late 19th century.[41] The ability to control radiated areas, coupled with the small travel radius of the resulting radicals, allows for the preferential targeting and killing of tumour cells. Furthermore, the tendency for radiation to produce clustered lesions makes this treatment particularly effective.[15,16] The radiation products discussed above (Section 1.4.1.2.) induce cell death through complex signalling cascades, with DNA double strand breaks (DSBs) implicated as the primary contributor.[42] A 1-Gy dose results in 10^5 ionizations per cell, which produces up to 2000 single strand breaks (SSBs) and 40 DSBs.[43]

Radiation therapy has also been used synergistically with chemotherapeutic agents. For example, cisplatin exposure in conjunction with radiation increases the uptake of the platinum compound into the cell.[41] Additionally, the effectiveness of ionizing radiation can be improved by administering radiosensitizing agents, which amplify DNA damage. Such compounds will be discussed in the following section.

1.4.   5-Halouracils and Their Use in Anti-Tumour Therapies

1.4.1. Fluorouracil

5-fluorouracil ([F]U) is currently used as a chemotherapeutic agent and exerts cytotoxic effects in a variety of ways. The van der Waals radius of F (1.35 Å, Figure 1–6(a)) is similar to that of H (1.2 Å), making [F]U a U analogue.[44] Through cellular exposure to [F]U and subsequent conversion to active metabolites, [F]U is incorporated into the U position of RNA.[45] [F]U disrupts normal RNA functioning, such as the assembly of RNA-protein complexes and post-translational modifications by inhibiting the conversion of uridine to pseudouridine.[45] 5-fluoro-2′-deoxyuridine monophosphate may also act as a competitive inhibitor of thymidylate synthase, an enzyme that catalyzes the *de novo* synthesis of thymidine monophosphate from 2′-deoxyuridine monophosphate.[45] This decreases the concentration of thymidine in the nucleotide pool, which results in the incorporation of [F]U and U into DNA.[45] Although repair mechanisms exist to remove U and [F]U from DNA,[46] these bases will continue to be incorporated into DNA due to low concentrations of thymidine, eventually leading to cancer cell death.[45]
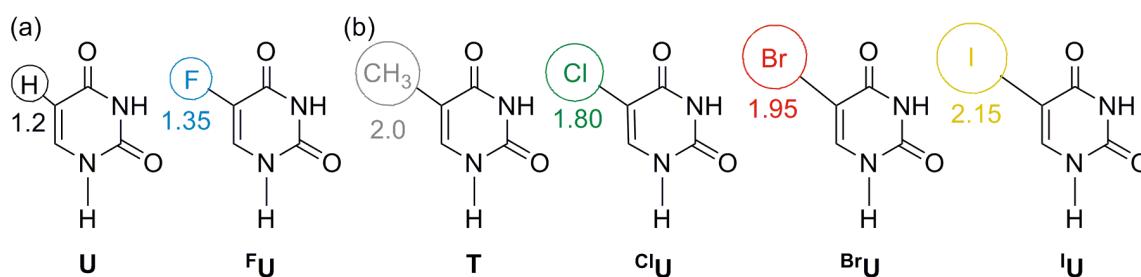
Figure 1–6: The structures and van der Waals radii (Å) of the C5 substituents[44] for (a) uracil and its analogue, 5-fluorouracil, as well as (b) thymine and its 5-halouracil analogues.

1.4.2. *Chloro-, Bromo- and Iodouracils*

The use of 5-chlorouracil ([Cl]U), 5-bromouracil ([Br]U) and 5-iodouracil ([I]U) in anti-tumour therapies differs substantially from that of [F]U. These three 5-halouracils ([X]U) are analogues of the T nucleobase due to the similarities in van der Waals radii (Figure 1–6(b)) of Cl (1.80 Å), Br (1.95 Å) and I (2.15 Å) to the C5–methyl (2.0 Å) of T.[44] Therefore, cellular exposure to these 5-halo-2′-deoxyuridines results in [X]U incorporation into the T position of DNA upon replication. Furthermore, [X]U are radiosensitizing agents that enhance the effects of radiation therapy on tumour cells.[44] In this regard, they must first be "activated" by radiation, which causes dissociation of the halogen and production of a radical, before exerting their cytotoxic effects.

The radiosensitizing ability of these halouracils was fortuitously discovered in the 1950's when examining the mutagenic effects of radiation on cells while employing [Br]U and [I]U as density labels in DNA.[44,47] Shortly thereafter, the potential for these compounds to act as anti-cancer agents was recognized, which led to numerous *in vitro* and *in vivo* studies, as well as clinical trials employing ionizing radiation.[44,48,49] Several factors made the administration of these 5-halo-2′-deoxyuridines with radiation appealing. Halouracils showed relatively low toxicity in non-radiated tissues[50] and toxicity could be minimized by administering 5-bromo-2′-deoxyuridine rather than 5-iodo-2′-deoxyuridine.[44] DNA damage could also be controlled though both the preferential incorporation of [X]U into rapidly dividing tumour cells and limiting the radiated area.[44,48] Furthermore, the radiation product of [X]U was reactive under the hypoxic conditions found in tumour cells, compared to other radicals that require reaction with O_2 prior to initiating DNA damage.[51]

Despite these apparent benefits, clinical trials failed to reveal a survival advantage for patients.[44] One major obstacle was achieving adequate and preferential incorporation

into tumour DNA without overexposing normal tissues.[44,52] Developments in recent years concerning preferential [X]U incorporation into tumour DNA have renewed interest in their use as anti-tumour therapies. Specifically, administration of 5-chloro-2′-deoxycytidine relies on the elevated levels of deoxycytidine kinase and cytosine deaminase in tumour cells for preferential incorporation of [Cl]U into tumour DNA over adjacent normal tissues where these enzymes are in lower concentrations.[53,54] Once 5-chloro-2′-deoxycytidine enters a tumour cell, it is metabolized by deoxycytidine kinase to 5-chloro-2′-deoxycytidine monophosphate and then deaminated by cytosine deaminase to produce 5-chloro-2′-deoxyuridine monophosphate. Additional phosphorylation reactions by higher kinases yield 5-chloro-2′-deoxyuridine triphosphate, from which [Cl]U is incorporated into the T position of DNA during replication. Interestingly, 5-chloro-2′-deoxyuridine monophosphate, like 5-fluoro-2′-deoxyuridine monophosphate, can also inhibit thymidylate synthase to increase the incorporation of [Cl]U over T in tumour DNA. Complementing the above, a recent study considered mutations to human deoxycytidine kinase that increase its specificity for 5-halo-2′-deoxycytidines.[55]

1.4.2.1. The Uracil-5-yl Radical

Following the discovery of the radiosensitizing ability of the halouracils, the observed cytotoxicity was attributed to DNA-damaging reactions initiated by the uracil-5-yl radical (U•), which forms following exposure of [X]U-containing DNA to either UV or ionizing radiation.[56] U• is a σ-radical, which is more reactive than π-carbon radicals and allows U• to damage DNA without first reacting with O_2.[51] Products of U• commonly occur in the presence of an adjacent 5′–purine nucleobase (see Section 1.5) and literature examining reactions of U• typically use [Br]U since it is the closest analogue of T among all [X]U (Figure 1–6). Therefore, generation of U• will be discussed below for exposure of the 5′–G[Br]U–3′ and

5′–A[Br]U–3′ sequences to UV and ionizing radiation. While identical mechanisms for radical generation occur for [Cl]U, differences exist for [I]U, which will also be outlined.

1.4.2.2. UV Irradiation of the Halouracils

Due to the enhanced reactivity of purine–[Br]U sequences in DNA and the low ionization potential of the purine nucleobases,[20] it was long believed that creation of U• in dsDNA was initiated by a photoinduced single electron transfer (PSET) from the purine adjacent to [Br]U. In the proposed process, the purine base absorbs a UV photon, which excites an electron and oxidizes the purine to a purine radical cation, while [Br]U is reduced to a radical anion ([Br]U•−).[57-59] In natural DNA, the excess electron on T•− is located in a π* orbital and can undergo radiationless charge recombination via ICs. However, the excess electron in [Br]U•− occupies the σ* orbital localized on the C5–Br bond,[60] which results in the dissociation of a bromide anion (Br−) and the creation of U• adjacent to a 5′–purine radical cation (for example, Figure 1–7(a)).

With the discovery of diverse products of U• and differences in sequence preferences, the commonly-accepted mechanism of U• generation presented above required re-evaluation. Since G has the lowest ionization potential among the natural DNA nucleobases,[20] PSET should be preferred in the 5′–G[Br]U–3′ sequence, which should result in more DNA damage in this sequence. However, certain DNA lesions selectively form in the 5′–A[Br]U–3′ sequence (Section 1.5.1.).[57-59,61-65] To explain this discrepancy, fast back-electron transfer in the 5′–G[Br]U–3′ sequence was proposed to prevent the loss of Br− and generation of U•.[57,61] However, other DNA lesions preferentially form in the 5′–G[Br]U–3′ sequence (Section 1.5.3.).[66-68]
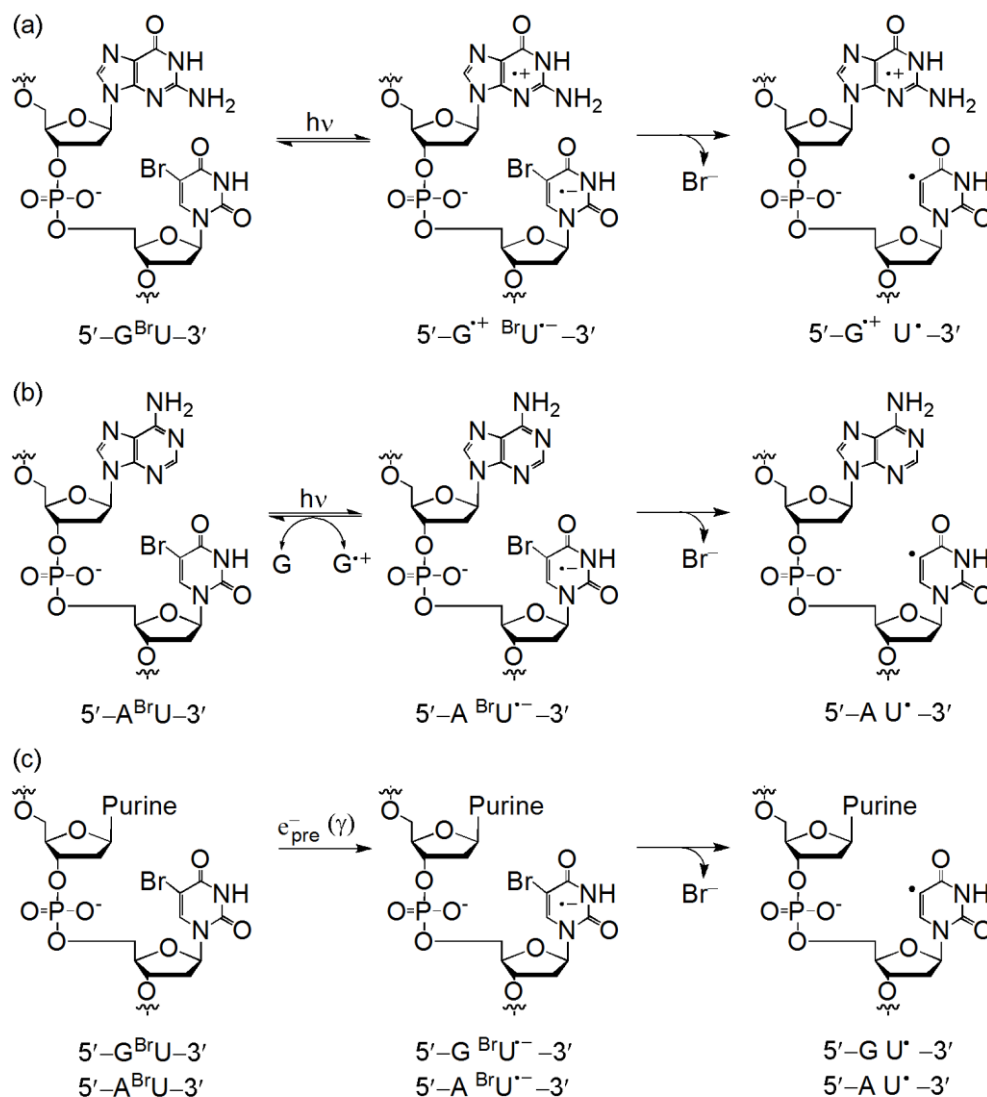
Figure 1–7: The generation of the uracil-5-yl radical following exposure of [Br]U-containing DNA to (a, b) UV or (c) ionizing radiation in the presence of a 5′–G or 5′–A.

More recently, these inconsistencies were addressed using the 5′–GAA[Br]UT–3′ sequence, which was identified as a "hotspot" for U• generation and DNA damage.[62,64] This sequence is rendered inactive if G is substituted for a nucleobase with a higher ionization potential,[64] which indicates that G, not A, is oxidized and donates an electron to [Br]U and that electron transfer can occur over a longer range than initially expected (Figure 1–7(b)).

Furthermore, reactivity was maximized with two intervening A bases between G and [Br]U, which verifies the role of A in slowing back-electron transfer from [Br]U.[64] This indicates long-range sequence effects are important in generating U•, which may also affect the preferences for the products formed.

In light of the discovery that G is the electron donor, a recent computational study examined possible excitations occurring when [Br]U-containing DNA is exposed to UV radiation.[69] Excitations were studied in stacked 5′–A[Br]U–3′ and 5′–G[Br]U–3′ nucleobase dimers in averaged crystal-structure geometries using CASPT2/CASSCF and the 6-311G(d) basis set. In both sequences, UV radiation at 300 nm results in a local excitation on the [Br]U chromophore ([Br]U*), rather than a direct electron (charge) transfer as previously proposed.[57-59] In the first excited state, [Br]U* is highly oxidative and capable of oxidizing a nearby G to form the charge-separated species. Subsequently, Br⁻ dissociates to generate U•. The direct creation of charge transfer states (A•⁺ [Br]U•⁻ or G•⁺ [Br]U•⁻) from UV photons does not occur due to the high energies of these states and the low oscillator strengths.

It is now clear that UV irradiation of [Br]U-containing dsDNA results in a local excitation of [Br]U, followed by oxidation of a nearby G to create either the 5′–G•⁺U•–3′ or 5′–AU•–3′ species in DNA (Figure 1–7 (a, b)). While a 5′–A•⁺U•–3′ reactant is of little relevance to dsDNA, it is applicable to UV irradiation of 5′–A[Br]U–3′ dinucleoside monophosphates used in experiment since no G is present[66-68] and may be relevant to single-stranded DNA (ssDNA) where long-range electron transfer from a nearby G is less efficient.

Generation of U• in dsDNA containing [I]U has a smaller sequence selectivity compared with [Br]U.[59,70] This was proposed to occur for one of two reasons:[59] (1) dehalogenation occurs faster for I⁻ than Br⁻, which means slowing back-electron transfer

with adjacent adenines is not as important; or (2) C5–I bond homolysis occurs. Support has been found for both scenarios. Using the previously-determined "hotspot" sequence (5′–GA$_n$^{X}UT–3′), maximum reactivity was found for ^{I}U when n=2, as was the case for ^{Br}U, which indicates heterolysis from an electron transfer processes.[70] However, the amount of DNA damage observed was higher for ^{I}U than ^{Br}U,[64,70] which supports I$^-$ dissociation being faster than Br$^-$ dissociation. For n=0, 6 or 7, no reactivity was observed for ^{Br}U, while ^{I}U produces DNA damage, which indicates C5–I homolysis occurs.[70] Therefore, homolysis and heterolysis of ^{I}U is highly dependent on the DNA sequence.[70] In the instance ^{I}U undergoes homolysis upon UV exposure, DNA damaging reactions must originate from the 5′–GU$^\bullet$–3′ and 5′–AU$^\bullet$–3′ species.

1.4.2.3. Ionizing Irradiation of the Halouracils

As discussed in Section 1.3.1.2., most reactions following exposure of cellular DNA to ionizing radiation result from the radiolysis of water. In regards to U$^\bullet$ generation in ^{X}U-containing DNA, the free electron (e^-) is the most important species. For many years it was believed $e^-_{(aq)}$ was responsible for the creation of U$^\bullet$, where the high electron affinity of ^{Br}U caused the base to act as an efficient electron sink.[71] Upon attachment of the solvated electron to ^{Br}U, the electron occupies the σ* orbital localized on the C5–Br bond of ^{Br}U$^{\bullet-}$, which causes dissociation of Br$^-$ and generation of U$^\bullet$. However, Br$^-$ dissociation occurs slowly and is competitive with electron transfer.[72]

A more efficient mechanism for the generation of U$^\bullet$ from free-electrons has recently been proposed.[61,62] Prior to solvation, a free electron exists in a precursor state ($e^-_{(pre)}$), which has an energy close to the dissociative electron attachment (DEA) energies of halogen-containing compounds.[71] However, reactions with $e^-_{(pre)}$ have been difficult to study

due to its femtosecond (fs) lifetime. Recent advances in laser spectroscopy have provided resolution on the fs-level, which allows for the observation of the transition state ($^{Br}U^{\bullet-}{}^*$) for electron attachment by monitoring transient absorption. Experiments examining 5-bromo-2′-deoxyuridine in solution have determined that $e^-_{(pre)}$ is responsible for DEA in ^{Br}U.[72,73] Specifically, attachment of $e^-_{(pre)}$ to ^{Br}U (0.2 ps) creates the vibrationally-excited, dissociative $^{Br}U^{\bullet-}{}^*$ species.[73] Subsequently, ultrafast (1.5 ps) Br$^-$ dissociation occurs before electron transfer is possible.[72,73]

In terms of the series of ^XUs, dissociative electron attachment increases as ^{Cl}U < ^{Br}U << ^IU, and can be attributed to ^IU having two vibrationally-excited $e^-_{(pre)}$ states, while ^{Cl}U and ^{Br}U have one.[73] Therefore, it can be conclusively determined that DNA damage from the exposure of ^XU-containing dsDNA to ionizing radiation, originates from U$^\bullet$ within the 5′–GU$^\bullet$–3′ and 5′–AU$^\bullet$–3′ sequences.

1.5. Reactions of the Uracil-5-yl Radical

A range of lesions are produced as a result of reactions initiated by U$^\bullet$. These have been detected following the exposure of ^XU-containing DNA to UV or ionizing radiation. The following sections will discuss these products.

1.5.1. Single- and Double-Strand Breaks and the 2′-Deoxyribonolactone Intermediate

In the 1970's, strand breaks were identified as the primary product of U$^\bullet$.[74] These lesions occur following exposure of ^{Br}U-containing DNA to either UV[49,57-59,61,62,64,74-77] or ionizing[49,63,65,78-80] radiation, despite differences in the formation mechanism. However, more SSBs are induced from UV[77] than ionizing[63] radiation. Both SSBs and DSBs may form, and involve the creation of a nick or gap in the backbone of one or both strands, respectively.[49,57-59,61,63,65,75-80] While a DSB commonly results from the concurrent formation

of SSBs on opposite strands in close proximity, they may also form from a one-photon process.[74,75]

Although the exact mechanism for strand break formation is not fully understood, certain aspects of the reaction pathway are known (Figure 1–8).[81-87] Many intermediates exist and the strand break can adopt many structural forms. Regardless of whether strand breaks are formed from UV or ionizing radiation, the reaction is initiated by U• abstracting a hydrogen atom from the 5′–adjacent deoxyribose moiety to form uracil and a sugar radical. This particular step has been examined in the 5′–AU•–3′ sequence with molecular modeling, which found abstraction was preferred from C2′ of the deoxyribose rather than C1′.[88,89]

Following the formation of the sugar radical, electron transfer to a nearby G•⁺ regenerates G and creates a cationic sugar, which can undergo a hydride shift (Figure 1–8). From these sugar-cation intermediates, several reactions may take place. A nick can form between the 5′–phosphate of U-containing nucleotide and the adjacent sugar,[58,83,84] which results in the formation of 3′-keto-2′-deoxynucleotide or furanylnucleotide (Figure 1–8 (1) and (2), respectively).[59,83] Alternatively, a nucleobase may be released with the formation of the alkali-labile 2′-deoxyribonolactone intermediate (Figure 1–8 (3)).[61,81,82,85-87] This lesion is mutagenic by preferentially pairing with G[86,87] and cytotoxic by inducing DNA–protein cross-links.[90] From this lactone residue, successive elimination reactions release the sugar as a methylenefuranone to generate a gap in the DNA backbone (Figure 1–8 (4)).[87]

The formation of strand breaks with UV radiation is highly sequence dependent in dsDNA.[57,59,61] Specifically, strand breaks are formed 8 times more efficiently with a 5′–flanking A than G in synthetic oligonucleotides.[57] Enhancements in SSB formation occur in the "hotspot" 5′–GAA^BrU^BrU–3′ sequence, where intervening adenines play a role in
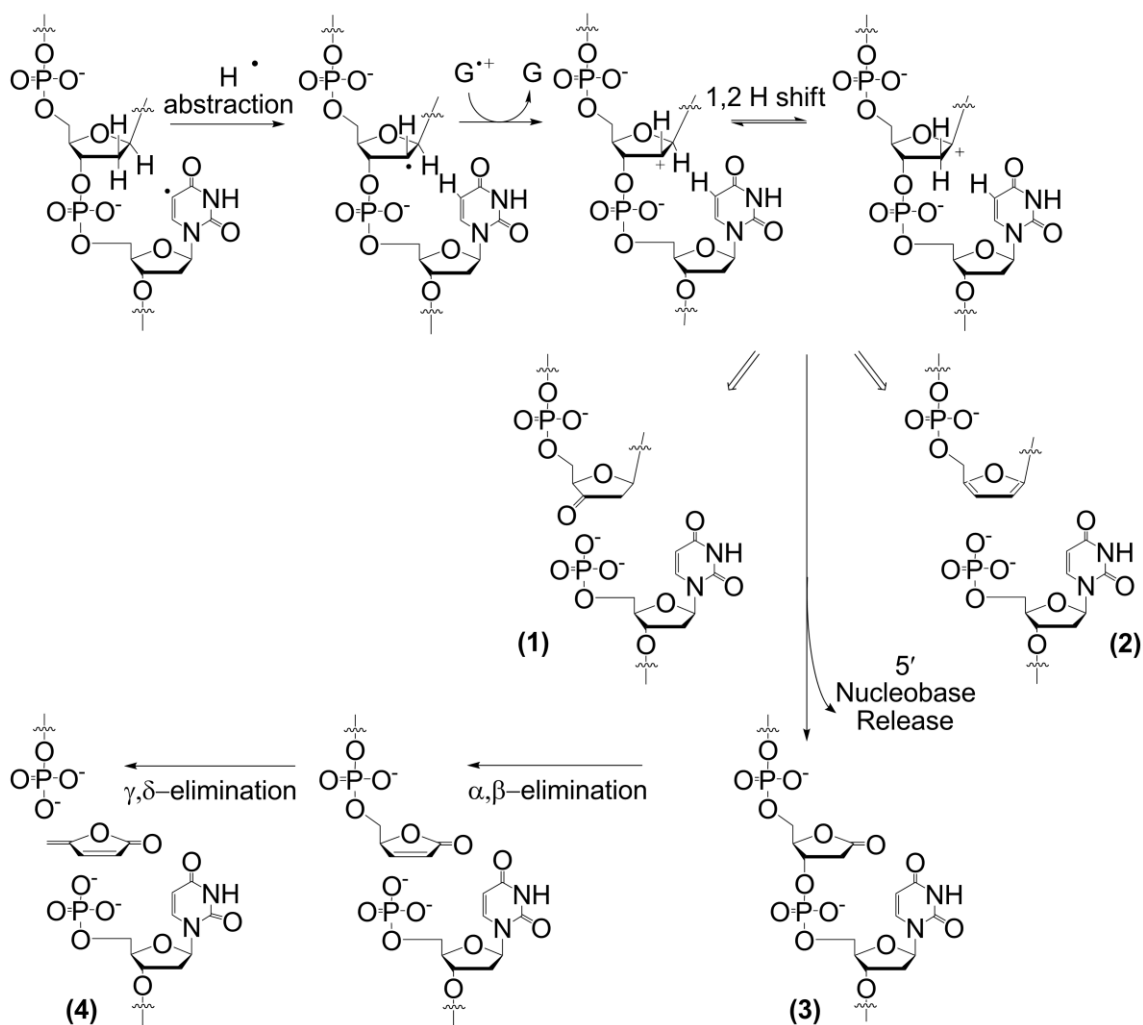
Figure 1–8: Outline of the reaction pathway for strand break formation resulting from the uracil-5-yl radical.

slowing back electron transfer to G[.+].[62] Interestingly, this sequence specificity is not observed in ssDNA where π–orbital overlap is disrupted,[58] which provides evidence for the role of π–mediated intrastrand electron transfer in the formation of U[.]. Strand breaks resulting from ionizing radiation of [Br]U-containing DNA have also shown no sequence preferences in ssDNA and dsDNA.[65]

Strand breaks have been implicated as the primary initiator of cell death resulting from U•.[50] In particular, DSBs are considered among the most cytotoxic lesions since their repair is complex and may cause deletions or insertions.[42] However, SSBs and DSBs do not completely account for the cell death observed in survival curves, which suggests additional forms of damage or cellular processing of the initial damage to a DSB.[49] Indeed, other [Br]U photoproducts were detected as early as 1973, but were not identified.[74] More recent studies have clarified the identity of these products, which are discussed below.

1.5.2. *Interstrand Cross-links*

Interstrand cross-links resulting from U•, which involve covalent bonds between nucleobases on opposite strands, were first reported in 2005.[50,63,65,77,80] An initial study exposed synthetic 25-base, [Br]U-containing duplex oligonucleotides to γ-rays.[63] Fully complementary dsDNA was considered in addition to semi-complementary dsDNA containing a central 5-base mismatch (AA[Br]UAA//CCCCC). These sequences were irradiated (^{60}Co, 1.18 MeV and 1.33 MeV) at a dose of 3.91 Gy min^{-1}, and interstrand cross-links detected with denaturing PAGE gel electrophoresis. However, using this detection method, the composition and exact connectivity of these cross-links cannot be determined. Interstrand cross-links were found to form exclusively in the mismatched region, which may either adopt a bulged or a zipper-like structure.[63,91] Regardless of the exact conformation of the mismatched region, it presents an environment unique from ssDNA and dsDNA that favours interstrand cross-link formation.

Interstrand cross-links were also detected following UV irradiation of the same sequences at 252 nm, 280 nm, 290 nm, 302 nm and 313 nm, with maximum damage detected at 280 nm.[77] Again, interstrand cross-links were exclusively observed in the mismatched region. Strand breaks were also detected, but formed preferentially in

complementary DNA. Therefore, the authors proposed that *in vivo*, UV irradiation of dsDNA containing [Br]U will primarily result in SSBs, while mismatched regions will form interstrand cross-links.[77]

Additional studies have examined the conformation and sequence dependence of interstrand cross-link formation following [Br]U activation by γ-rays.[65,80] While interstrand cross-links form in semi-complementary B-DNA, they are not produced in complementary or semi-complementary A-DNA.[80] This conformation-dependence indicates the mismatched region adopts a zipper-like structure rather than a bulge, which would be less affected by the differences between A- and B-DNA.[80] Mismatched DNA in the zipper arrangement naturally occurs in human centromeres, which may be a suitable target for the use of [X]U in anti-tumour therapies.[50] Interstrand cross-link formation was also found to depend on sequence. Twelve variations in the 5-base mismatch region were considered, with 10–12 distinct cross-links detected overall.[65] The presence of pyrimidines in the mismatch region promotes the formation of interstrand cross-links and SSBs, while G does not.[65] Since the most abundant interstrand cross-links occur when C is present on the opposite strand of the mismatch, the C5–C6 double bond of C was proposed to be the preferred site of radical attack.

Interstrand cross-links are particularly detrimental DNA lesions due to the inability of the helix to separate during fundamental cellular processes like replication or transcription.[92] Furthermore, interstrand cross-link repair involves the creation of a DSB.[93] Indeed, previous studies that detected DSBs resulting from U$^\bullet$ in cellular DNA may have actually detected the attempted enzymatic repair of an interstrand cross-link. This supports previous proposals that cellular processing of the initial damage to a DSB may explain why DBSs cannot fully account for cell survival curves.[49]

1.5.3. Intrastrand Cross-links

In addition to the strand breaks and interstrand cross-links produced by U[•], intrastrand cross-links have been identified.[66-68] These lesions involve a covalent bond between uracil and an adjacent purine nucleobase.[66-68] The sequence-dependent formation of intrastrand cross-links following UVB irradiation of [Br]U-containing synthetic DNA was first reported in 2006.[66] The intrastrand cross-links were identified in dinucleoside monophosphates following UVB-irradiation of the 5′–G[Br]U–3′, 5′–[Br]UG–3′ and 5′–A[Br]U–3′ sequences. Following UV irradiation of a solution of each dinucleoside monophosphate at 290 nm, photoproducts were separated using high-performance liquid chromatography (HPLC) and a reverse-phase column. Each fraction was subjected to electrospray ionization-mass spectrometry (ESI-MS) and tandem spectrometry (MS/MS) to determine the composition of the cross-link and its characteristic fragment ions. To establish the connectivity of each cross-link, ^{1}H-NMR was used. These cross-linked dinucleoside monophosphates were used as standards to identify the intrastrand cross-links formed in two synthetic, complementary dsDNA oligonucleotides with the 5′–d(ATGGC**G[Br]UG**CTAT)–3′ and 5′–d(ATGGC**A[Br]UG**CTAT)–3′ sequences. After irradiating the oligonucleotides at 290 nm, enzymatic digestion produced free nucleotides and cross-linked dinucleoside monophosphates. The cross-links were identified using LC-MS/MS, and comparing to the retention times and characteristic fragment ions of the standards.

Six different intrastrand cross-links were identified in dinucleoside monophosphates (Figure 1–9). Each cross-link involves the C5 site of U and either C8 (C8) or N2 (C2) of G (A). For example, a cross-link forms between C8 of a 5′–G and C5 of a 3′–U, denoted as 5′–G[8–5]U–3′ (Figure 1–9(a)). In synthetic dsDNA, 5′–U[5–N2]G–3′ (Figure
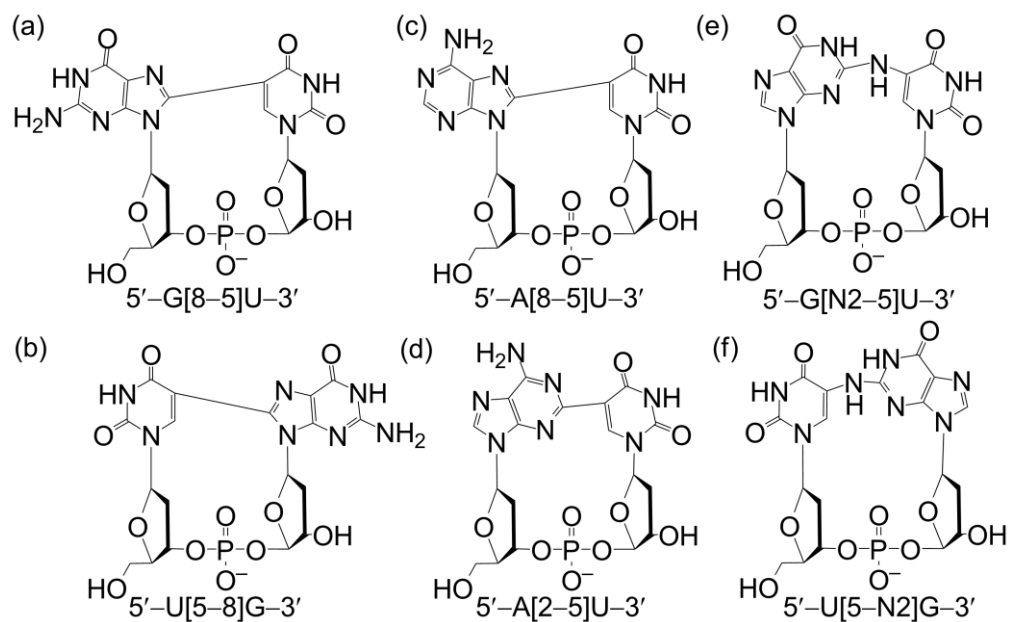
Figure 1–9: Intrastrand cross-links resulting from the uracil-5-yl radical.

1–9(f)) was not detected. As a general trend, cross-link yields increase as 5′–A[2–5]U–3′ < 5′–A[8–5]U–3′ < 5′–U[5–8]G–3′ < 5′–G[8–5]U–3′ and with irradiation time (Table 1–1). This is in contrast to the sequence specificity observed for strand break formation, where a 5′–A is preferred over a 5′–G. 5′–G[N2–5]U–3′ was omitted from the above trend since it is only observed at large radiation doses with a questionable (0.2 ± 0.2 %) yield. Interestingly, 5′–U[5–8]G–3′ is not detected in the 5′–A[Br]UG–3′ sequence despite its high yield in 5′–G[Br]UG–3′ sequences (Table 1–1). This is indicative of the importance of the local sequence in U• generation and in the identity of the products formed.

In 2007, intrastrand cross-links produced by U• were identified in cellular DNA.[67] MCF-7 human breast cancer cells were cultured in 5-bromo-2′-deoxyuridine and subsequently exposed to UVB radiation at 290 nm, enzymatically digested and subjected to

25

Table 1–1: Yields of the intrastrand cross-link products resulting from UVB-irradiated [Br]U-containing oligonucleotides See Ref. 66.[a]

	5'–G[8–5]U–3'	5'–U[5–8]G–3'	5'–A[8–5]U–3'	5'–A[2–5]U–3'	5'–G[N2–5]U–3'
10 min irradiation					
5'–G[Br]UG–3'	0.5 ± 0.1	0.16 ± 0.05	N/A	N/A	–[b]
5'–A[Br]UG–3'	N/A	–[b]	0.09 ± 0.03	0.03 ± 0.01	N/A
90 min irradiation					
5'–G[Br]UG–3'	3 ± 1	1.1 ± 0.2	N/A	N/A	0.2 ± 0.2
5'–A[Br]UG–3'	N/A	–[b]	0.10 ± 0.05	0.05 ± 0.02	N/A

[a]Percent conversion of the parent duplex averaged from 3-6 measurements. [b]Not detected.

the same LC-MS/MS identification procedure described above. Four intrastrand cross-links (Figure 1–9 (a–d)) were detected and their yields quantified. The relative abundances of the lesions (Table 1–2) parallels that presented above. Interestingly, the yield of the most and least abundant cross-links decrease with increasing irradiation time, which may reflect a product decomposition or preferred repair of these cross-links.[67]

Table 1–2: Yields of the intrastrand cross-link products resulting from UVB-irradiated [Br]U-containing cellular DNA. See Ref. 67.[a]

	5'–G[8–5]U–3'	5'–U[5–8]G–3'	5'–A[8–5]U–3'	5'–A[2–5]U–3'
9 min irradiation	130 ± 20	57 ± 10	5 ± 1	1.1 ± 0.1
70 min irradiation	89 ± 10	71 ± 10	7 ± 1	–[b]

[a]Expressed as lesions/10^6 nucleosides. [b]Not detected.

Since $G^{\bullet+}$ is a well-known precursor to 8-oxoG, detection of 8-oxoG can confirm the presence of $G^{\bullet+}$. Therefore, the yields of 8-oxoG in irradiated DNA were quantified to provide insight into the mechanism of $U^{\bullet}$ generation. Indeed, the yield of 8-oxoG in UV-irradiated [Br]U-substituted DNA was greater than natural DNA, which verifies the role of G oxidation in UV-induced $U^{\bullet}$ generation.

The yields of [Br]U- and U-containing nucleotides were also measured following irradiation and digestion of DNA to shed light on the efficiency of [Br]U activation.[67] Prior to UV exposure, [Br]U replaced 30.6% of T in DNA. Following 9 minutes of irradiation, 14.5% of T

remained as [Br]U while 10.8% T was converted to U,[67] which may result from the first (hydrogen abstraction) step of strand break formation. The remaining 5.3% of T is converted to other products of U[•]. While these numbers support strand breaks as the dominant product of U[•], other lesions also form that might actually be more cytotoxic. Unfortunately, the digestion procedures employed in this study do not allow for the simultaneous detection of strand breaks and intrastrand cross-links.

To rationalize the sequence dependence of intrastrand cross-link formation, the authors stated intrastrand stacking was responsible. Specifically, it was proposed that stronger stacking facilitates better electron transfer and therefore higher cross-link yields. The authors reported that the stacking strength increases as 5′–A < 5′–G and 5′–pyrimidine–purine–3′ < 5′–purine–pyrimidine–3′.[94] However, recent computational investigations into the strength of nucleobase stacking interactions do not support these trends.[95-97] Therefore, it is unlikely that stacking strength plays a role in the observed cross-link formation preferences.

To explore the conformation-dependence of intrastrand cross-link formation, [Br]U-containing G-quadruplex DNA was also studied.[68] G-quadruplex structures, which are found in telomeres and promoter sequences, provide a unique environment to study intrastrand cross-link formation due to the presence of various loop regions that differ in the proximity of adjacent bases and strength of stacking interactions. DNA was UVA-irradiated at 365 nm and the 5′–G[8–5]U–3′ intrastrand cross-link was isolated and detected using the techniques described above. Indeed, the yield of this cross-link is greater in loops where G and [Br]U are in closer proximity, which the authors again attribute to stronger stacking interactions. However, the closeness of the nucleobases also shortens the distance between C8 of G and C5 of U, which has been previously noted in the literature and proposed to

influence the preferences of intrastrand cross-link formation.[98-102] Interestingly, the overall yield of 5′–G[8–5]U–3′ was 2-fold higher in the G-quadruplex than in synthetic dsDNA. A closer examination of the quadruplex sequence suggests this may be attributed to a new "hotspot" sequence repeated throughout the motif, 5′–GGG[Br]UTA–3′, which favours intrastrand cross-link formation. Specifically, G stretches have a lower ionization potential than a single G and adjacent pyrimidines provide a more efficient electron sink,[103] which may enhance the generation of 5′–G•+U•–3′.

The above experimental work has conclusively identified the composition and connectivity of four intrastrand cross-link that form following UV irradiation of [Br]U-containing cellular DNA. However, the exact structure of these lesions and distortions to the helix are unknown. Furthermore, there is no information on the reaction by which these lesions form. Although a reaction pathway has been proposed (Figure 1–10),[66] A-containing cross-links are erroneously generated by an A•+-containing reactant (Figure 1–10(b)). Furthermore, the origin of the observed sequence selectivity is unknown. This thesis sets out to fill these gaps in our knowledge of U•-mediated intrastrand cross-link formation. However, first, the information known about related intrastrand cross-links is presented.

1.6. Related Purine–Pyrimidine Intrastrand Cross-links

Given the complexity of cellular responses to DNA damage, insight into the implications of U-containing intrastrand cross-links may be provided through knowledge of the biological processing of comparatively well-studied lesions with similar structures. Indeed, related intrastrand cross-links have been examined using experimental and computational methods, and significantly more is known about the formation and structure of these lesions, as well as their implications in DNA repair and replication. Furthermore,
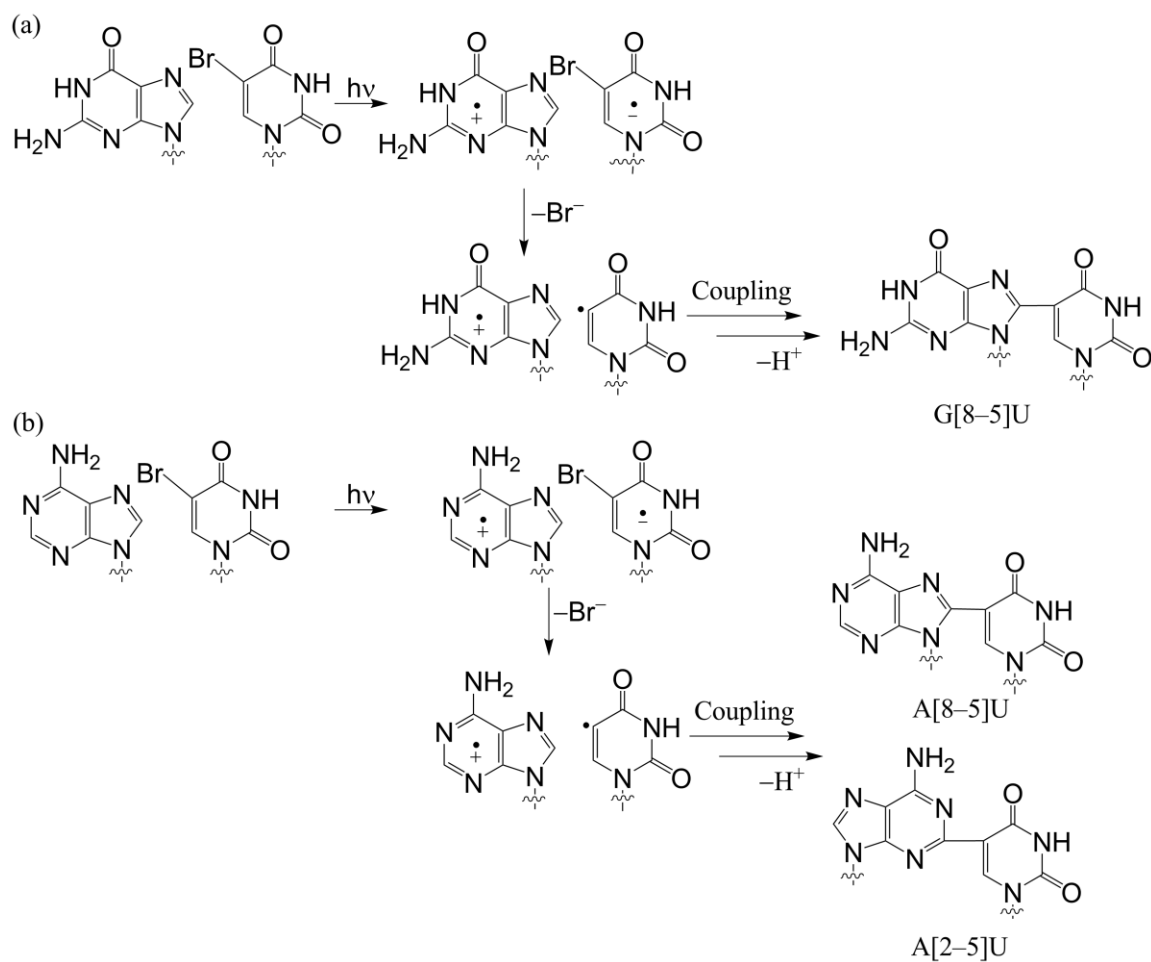
Figure 1–10: Proposed formation mechanism of intrastrand cross-links from UV irradiation of [Br]U-containing cellular DNA in the presence of (a) G or (b) A. See Ref. 66.

the biological processing of related lesions can provide clues as to the potential contributions of purine–U intrastrand cross-links to the observed cell death when [X]U are used in anti-tumour therapies. These studies are discussed below.

1.6.1. *Experimental Studies*

Intrastrand cross-links have been detected in natural DNA following exposure to ionizing radiation and originate from reactions with •OH that create a C5- (or C5-methyl-) pyrimidine radical.[98,102,104-106] This results in purine–pyrimidine cross-links with a [8–5] (or

[8–5-Me]) linkage (for example, Figure 1–3(e)), where preferences are observed for G over A, and for the purine in the 5′–position.[98,102,104] For example, experiments isolated four intrastrand cross-links in γ-irradiated natural DNA, which varied in their yields as 5′–G[8–5-Me]T–3′ > 5′–T[5-Me –8]G–3′ > 5′–A[8–5-Me]T–3′ > 5′–T[5-Me –8]A–3′.[98] Therefore, the formation of these intrastrand cross-links is similar to those produced by U$^\bullet$ formed upon exposure of ^{Br}U-containing DNA to ionizing radiation.

The effects of related purine–pyrimidine cross-links on repair and replication have been studied.[104-111] In general, lesions that cause substantial disruptions to the DNA helix are likely to be repaired, where bulky lesions, including intrastrand cross-links, are removed via a nucleotide excision repair (NER) pathway. The excision of the 5′–G[8–5]C–3′,[109] 5′–G[8–5-Me]mC–3′ (mC, 5-methylcytosine)[109] and 5′–G[8–5-Me]T–3′[108,109] intrastrand cross-links by UvrABC (*E. coli*) Nuclease was studied *in vitro* and a correlation between duplex destabilization and repair efficiency was found, with 5′–G[8–5-Me]mC–3′ repair being least efficient.[109] Additionally, 5′–G[8–5-Me]T–3′ is repaired less efficiently than other lesions (CPDs, bulky damage sites).[108] This is attributed to the lack of recognition of the intrastrand cross-link, despite the detection of the resulting DNA distortion and binding of the repair enzyme.[108] These results indicate G–pyrimidine cross-links may escape NER pathways and persist in cells. Furthermore, compromised repair functions in tumours may also contribute to the perseverance of these lesions in cancer cells.[36,54]

Intrastrand cross-links that are not repaired are detrimental to DNA replication. Studies of the 5′–G[8–5]C–3′[104,107,110] and 5′–G[8–5-Me]T–3′[105,106,112] cross-links and various non-human DNA polymerases found these intrastrand cross-links can be cytotoxic and mutagenic. Specifically, certain polymerases stall at the lesion,[105-107,110] while others incorporate the correct nucleotide across the cross-linked pyrimidine prior to

stalling.[105,107,110] This is due to a lack of an appropriate hydrogen bonding partner for G. If a stalled polymerase is not resolved, the complex can collapse into a DSB[113] or completely block replication, leading to cell death. [105,106] Alternatively, replication past the cross-link can occur with translesion synthesis (TLS). TLS by yeast polymerase η[98,100] or *E.coli* polymerase V[110] erroneously incorporates a purine nucleotide opposite the cross-linked G, likely due to steric considerations.[104,106,110] In contrast, *in vitro* TLS of 5′–G[8–5-Me]T–3′ by human polymerase η was found to insert the correct nucleotides opposite the cross-link.[111] However, *in vivo* replication of 5′–G[8–5-Me]T–3′ and 5′–T[5-Me–8]G–3′ in mammalian cells caused cellular mutations (transversions, transitions, and deletions) with the later displaying greater mutational frequency, where these errors were attributed to TLS by other polymerases or the involvement of accessory proteins.[111]

1.6.2. *Computational Studies*

To elucidate the mutagenesis mechanism of the above cross-links, calculations were performed on 5′–G[8–5-Me]T–3′ and 5′–G[8–5]C–3′. Using the semi-empirical PM3 method and a dinucleoside monophosphate model, it was found that the Watson-Crick hydrogen-bonding face of the cross-linked pyrimidine is unperturbed, while the cross-linked G adopts a *syn* orientation to expose the Hoogsteen face.[104,106] Although this provides an explanation for stalling and error-prone TLS by some polymerases, it cannot explain the correct replication of 5′–G[8–5-Me]T–3′ by human polymerase η, or the mutations observed afterwards.

In an attempt to elucidate the observed formation preferences of T-containing intrastrand cross-links,[98,102,104] molecular modeling has been used to study the mechanism of formation of GT- and AT-containing intrastrand cross-links.[114,115] Using a dinucleoside monophosphate model and B3LYP/6-31G(d,p), cross-link formation pathways were

determined. Overall, the reactions are initiated by bond formation and followed by the removal of H8• of the purine. For the four cross-links considered, the barrier for breaking the C8–H8 bond is similar in magnitude. Therefore, the selectivity for G over A occurs during bond formation, where the barrier is smaller for G than A. However, the preference for sequences with a 5′–purine, rather than a 3′–purine is proposed to occur in the second step since H8 is more accessible for removal from the cross-linked intermediate in 5′–purine sequences.

The above illustrates the strength of a computational approach in the examination of both the structures and formation mechanisms of radical-initiated lesion formation. Computational methods will be used in this thesis to study U-containing intrastrand cross-links to provide insight into the formation and structure of these lesions, as well as their biological implications.

1.7.   Thesis Overview

This thesis uses molecular modeling to examine the formation mechanism and structure of the four purine–U intrastrand cross-links that occur in cellular DNA from U•. Although experimental studies have identified these cross-links, their structures and mechanisms for formation are unknown. Furthermore, the effects of these lesions on the DNA helix and the biological fate of cross-linked DNA have not been studied. Therefore, the goal of this thesis is to determine the formation mechanism of the four U-containing intrastrand cross-links under conditions relevant to UV and ionizing radiation. Additionally, the structure of these lesions, as well as the distortions induced to the helix, will be established.

Using the proposed reaction pathways (Figure 1–10) as a basis, the formation of these lesions will be studied under the conditions relevant to UV (from 5′–G•+U•–3′ and

5′–AU•–3′) and ionizing (from 5′–GU•–3′ and 5′–AU•–3′) radiation. Although these lesions have not been detected in studies employing ionizing radiation, indirect evidence suggests they may contribute to the cytotoxicity of [Br]U in tumours. For example, strand breaks and certain interstrand cross-links are produced regardless of the form of irradiation used to create U•.[49,57-59,61-65,74-80] Furthermore, purine–pyrimidine intrastrand cross-links have been detected following the exposure of natural DNA to ionizing radiation.[98,102,104-106] Indirect evidence also exists for the formation of intrastrand cross-links following exposure of [Br]U-substituted DNA to ionizing radiation, where an observed decrease in the number of SSBs in the 5′–G[Br]U–3′ sequence was proposed to be due to preferential formation of intrastrand cross-links.[65] Additionally, it is possible these intrastrand cross-links have been formed in studies employing ionizing radiation, but were not detected due to the digestion and detection procedures commonly used, which complicates the simultaneous identification of SSBs, DSBs and inter- and intrastrand cross-links.[50,65,116] Therefore, it is important to examine the formation of these lesions by ionizing radiation and thereby determine the likelihood of their formation during the use of [X]U in conjunction with (ionizing) radiation as an anti-tumour therapy.

A high sequence dependence has been reported for intrastrand cross-link formation resulting from U•. Although experimental reports postulate this is attributed to enhancement of intramolecular electron transfer due to the strength of intrastrand stacking interactions,[66-68] recent computational investigations of nucleobase stacking do not support this proposal.[95-97] Furthermore, intrastrand cross-links formed following the exposure of natural DNA to ionizing radiation report the same sequence dependence,[98,102,104] but their formation does not involve an intramolecular electron transfer, which indicates a more general selection rule is responsible. Interestingly, experimental studies have noted a trend between cross-link yields and the distance between the two sites coupling to form the

bond,[98-102] though the origin of this correlation is unknown. Studies of purine[8–5-Me]T formation by the Barone group have proposed the barrier for the bond formation step controls the G > A preference, while the accessibility of the hydrogen atom for removal dictates the 5′–purine > 3′–purine preference.[114,115] By determining the formation pathways of the four U•-induced intrastrand cross-links in this thesis, the correlations noted experimentally and the factors proposed by Barone *et al.* can be tested and new ideas can be put forth. Therefore, an additional goal of this thesis is to provide an enhanced understanding of the factors that dictate intrastrand cross-link formation.

In addition to the above contributions, this thesis will determine the structures of these U-containing lesions within DNA and the resulting distortions to the helix. By comparing these results to the structures of related purine–pyrimidine cross-links and their effect on fundamental biological processes (repair and replication), as well as to DNA distortions induced by other successful anti-tumour therapies, insight will be provided into the biological fate of the purine–U intrastrand cross-links in tumour cells, and their potential to contribute to the observed cytotoxicity of [Br]U.

Before the formation of these intrastrand cross-links can be studied, an accurate computational approach must be determined. Chapter 2 details the selection of a computational model that is used to study intrastrand cross-link formation in the remainder of this thesis. Chapter 3 describes the formation of the most abundant cross-link, 5′–G[8–5]U–3′, under conditions relevant to both UV and ionizing radiation, as well as its resulting structure in the DNA helix and its biological implications. Chapter 4 similarly describes the formation and structure of 5′–U[5–8]G–3′, and begins to establish the preferences of intrastrand cross-link formation based on comparison to the 5′–G[8–5]U–3′ cross-link. Chapter 5 investigates both 5′–A[8–5]U–3′ and 5′–A[2–5]U–3′, and tests the ideas

proposed in Chapter 4, as well as provides additional insight into the factors governing cross-link formation. Finally, Chapter 6 provides an overview of this thesis and outlines possible future directions of this work.

Chapter 2. Determining an Accurate Computational Approach to Model Reactions Between Adjacent DNA Components[iii]

2.1.   Introduction

One of the main goals of this thesis is to use molecular modeling to examine the formation mechanisms of U-containing intrastrand cross-links. Unfortunately, a variety of computational models (DNA truncations, phosphate groups, environments) and methods have been used to gain information about radical-initiated DNA damage involving adjacent nucleic acid components,[88,89,114,115,117-120] which makes logically selecting the best computational approach for subsequent studies nearly impossible. Therefore, this chapter will identify an accurate computational approach that can be used to study the structures and formation pathways of intrastrand cross-links.

Although the complex features of DNA have been well characterized experimentally, modern computational methods cannot always accurately reproduce the structure of this biomolecule.[95,121-123] Adding to this computational challenge is the relatively large size of even short DNA oligomers, which necessitates structural truncations for improved computational efficiency. The structural model investigated in the present study must be large enough to describe key features of a DNA helix, yet small enough that highly accurate (density functional theory (DFT)) methods suitable for thermodynamic and kinetic studies of DNA damage can be employed. While it is possible to model base–base intrastrand reactions in the absence of the sugar-phosphate backbone,[119] computational constraints are required to obtain planar arrangements of the bases, which are not suitable when studying modified DNA since the exact structure is often unknown. In contrast, inclusion of the DNA

backbone naturally restricts the geometric freedom available to the nucleobases. Furthermore, the torsional degrees of freedom in the backbone may play an important role in determining the structure or reaction barriers. The inclusion of the DNA backbone will also allow the optimum computational approach identified in this chapter to be used in future studies of reactions between other nucleic acid components (for example, base–sugar reactions). Therefore, this chapter focuses on the ability of a dinucleoside monophosphate (Figure 2–1) to replicate the structure of this unit in a single strand within B-DNA.
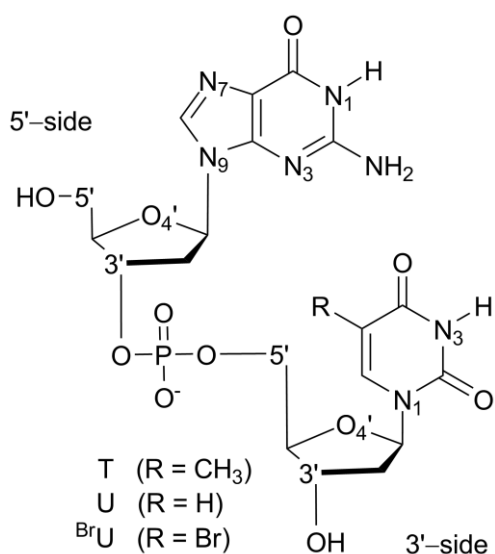
Figure 2–1: The (anionic) dinucleoside monophosphate and sequences considered in this chapter, as well as the structure and numbering of G and pyrimidine (T, U, BrU) nucleobases.

There are many questions surrounding how to correctly describe the DNA phosphate residues in small computational models, which has been depicted in the literature as an anionic group,[120,124-137] a neutral group resulting from full protonation[88,89,114,115,120,127-129,132,137] or a neutral group resulting from inclusion of a counterion[10,95,121-123,133,137-146] (typically Na+).[133,137,141,143,145,146] In addition to issues surrounding the atomic composition of the computational model, the level of theory must

37

be carefully considered. In the present work, DFT is used due to its computational efficiency and resistance to spin contamination compared to commonly-used *ab initio* techniques,[147] which make it an ideal choice for studies of radical reactions. Unfortunately, it is unlikely that all DFT methods will reproduce the fundamental features of a DNA double helix, since the ability to describe stacking interactions between DNA nucleobases is a common obstacle for many widely-used DFT functionals.[95,121,148] Nevertheless, the presence of a physical link between the nucleobases via the deoxyribose-phosphate backbone may sufficiently restrain base motion to permit characterization of stacked arrangements. Therefore, careful testing of phosphate models and methods must be performed to identify a combination that accurately mimics the conformation of dsDNA. Different environmental conditions must also be investigated since the role of the surrounding medium on the calculated structure and reactivity of DNA is complex.[149]

In this chapter, the ability of a dinucleoside monophosphate (Figure 2–1) to predict B-DNA structures is tested for three sequences (5′–GT–3′, 5′–GU–3′ and 5′–G^{Br}U–3′). By including these three base variations, a given computational approach will be tested for consistency in the reproduction of dsDNA structures when a purine is in the 5′ position relative to a natural (T) or modified (U, ^{Br}U) pyrimidine. ^{Br}U is considered since it is the dominant ^{X}U used in experimental studies examining DNA damage resulting from U•[50,57-59,61,63,65-68,74-77,80,83,84,150] and was used to generate the intrastrand cross-links of interest in this thesis.[66-68] U, a form of DNA damage that results from deamination of C, is considered due to its structural similarities to T, while the corresponding T sequence was chosen to allow for a direct comparison to experimental data for natural DNA. All pyrimidines were placed adjacent to 5′–G due the experimentally-observed preference for intrastrand cross-link formation in the 5′–G^{Br}U–3′ sequence.[66-68] The three sequences will be considered using three DFT functionals (B3LYP, MPWB1K and M06-2X), which were chosen

based on their high use in the literature, availability in common software packages, ability to describe non-covalent interactions and/or general success for chemical reactions.[121,148,151-156] All sequence and method combinations will be considered with three phosphate models (neutral, anionic and counterion) and two surrounding environments (gas phase and water), which collectively represent the extremes of the environmental conditions relevant to DNA.[157,158]

The overall suitability of a specific computational approach (phosphate model, environment and method combination) will be judged based on its ability to reproduce the structural features of the natural (T-containing) DNA double helix. Although modifications to DNA nucleobases can drastically affect the structure of the double helix, comparisons to the structure of natural DNA are expected to be valid for all bases considered in the present work due to the structural similarities of U and [Br]U to T. Only one study has attempted to develop an accurate computational approach for modeling B-DNA using a dinucleoside monophosphate system.[133] Although this previous contribution was valuable in formulating the work in this chapter, the present study considers a greater number of phosphate models, computational methods and environmental conditions. Furthermore, this approach is based on the intention to use the best computational model and method combination to study radical-induced intrastrand cross-link formation, which imposes additional criteria on the computational methodology. In addition to the rather specific applications of this model, the approach developed here can be applied to a broader range of DNA modeling, such as the prediction of structures involving natural or modified DNA components and the investigation of a variety of radical-initiated reactions between adjacent (intrastrand) nucleic acid components.[10,12,26,98,102,104-106,142,159-162]

2.2.   Computational Details

2.2.1.   *Generation of the Dinucleoside Monophosphate*

The dinucleoside monophosphate models used in the present work (Figure 2–1) were initially built to resemble a natural B-DNA double helix (described in Section 1.2.2.) since the primary goal is to retain a dsDNA structure rather than analyze the complex conformational space of a dinucleoside monophosphate. In general, two starting points are typically used to generate DNA fragments: (1) a crystal structure, or (2) a common software package that employs averaged geometries from natural DNA crystal structures. While the former approach may be preferred, experimental structures do not exist for the 5′–G^{Br}U–3′ sequence in DNA, which is common for many forms of modified DNA. In contrast, the latter approach provides a universal and consistent way to build structures regardless of the modification under consideration.

In the present study, the B-conformation of a natural 5′–GT–3′ dinucleoside monophosphate with an anionic phosphate group (Figure 2–2) was assembled using the Nucleic Acid Database built in the HyperChem 5.1 program suite.[163] Although HyperChem inserts a deoxyribose moiety in the C3′–*exo* puckering, the sugar puckering following full optimization has been shown to be independent of the initial input.[133] Similarly, the present work shows that structures generally adopt the C2′–*endo* puckering (defined by a pseudorotational phase angle (P) between 144 – 190°,[4] Figure 2–3) of B-DNA following optimization. Additionally, to obtain a structure representative of a DNA double helix, the terminal C5′–hydroxyl group was constrained such that the ∠(HO5′C5′C4′) torsion angle (Figure 2–1) equals 180°, which directs this group away from G and thereby prevents hydrogen bonds that are non-native to natural B-DNA.
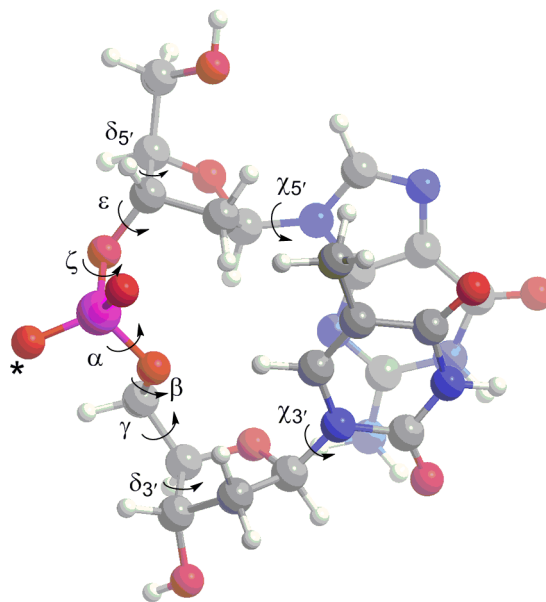
Figure 2–2: Important torsion angles in the backbone of a natural 5'–GT–3' dinucleoside monophosphate with an anionic phosphate group. The phosphate oxygen protonated to obtain the neutral phosphate model is indicated with an asterisk.

(a)

$$\tan P = \frac{(\nu_4 + \nu_1) - (\nu_3 + \nu_0)}{2 \times \nu_2 \,(\sin 36° + \sin 72°)}$$

(b)
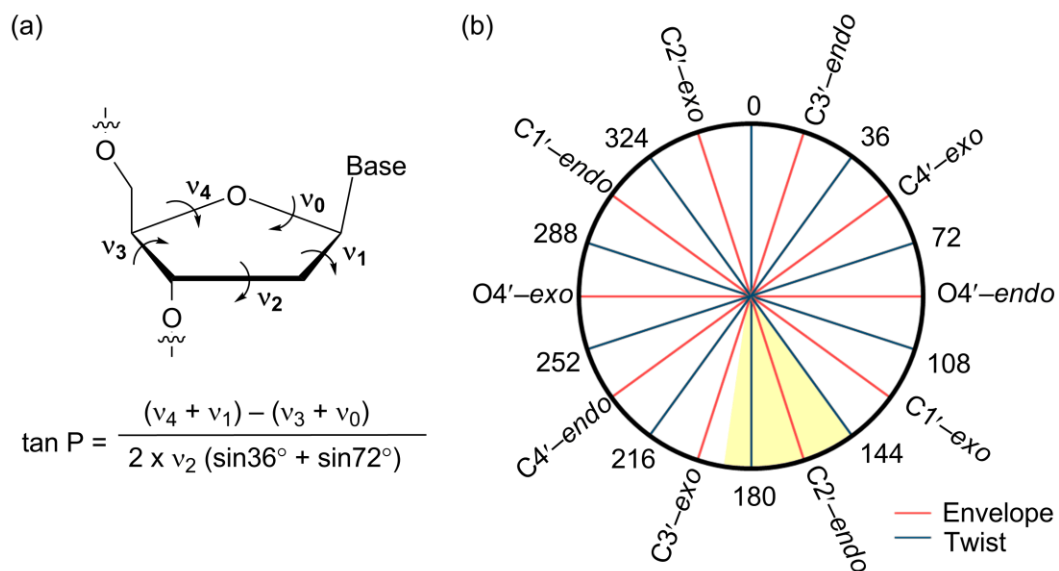
Figure 2–3: Definition of the (a) pseudorotation phase angle (P, °) and illustration of the (b) pseudorotation cycle and corresponding sugar puckering classifications describing the furanose ring in nucleic acids. Classifications belonging to the envelope and twist forms are indicated, and the range of phase angles found in B-DNA (P = 144 – 190°)[4] is highlighted.

2.2.2. The Consideration of Different Computational Approaches

Although the phosphate unit within the DNA backbone contains a monoanionic charge and several computational studies of DNA have used an anionic model (see, for example, Refs. 120, 124-137), a charge-neutralized model may better mimic the true biological system by accounting for stabilizing forces in the surroundings. Indeed, under biological conditions, the presence of cation–water networks near the phosphate backbone are known to provide stability to cellular DNA.[4] Furthermore, computational studies have found that charge-neutralized systems better represent the structure and conformational preferences of natural DNA.[127-129,131,133,137,164] Many groups have employed neutralized phosphate models obtained by coordinating a counterion,[130,132,133,137-141,143,145,146] commonly Na+,[133,137,141,143,145,146] to the phosphate residue. In this work, the counterion phosphate model was generated by inserting a sodium counterion (Na+) equidistant (approximately 2.2 Å) from the two phosphate oxygens. Alternatively, some groups have used fully protonated phosphate groups in DNA models,[88,89,114,115,120,127-129,132,137] which neutralize the charge while avoiding the addition of a computationally-demanding heavy atom. In the present study, the phosphate oxygen furthest from the DNA nucleobases (Figure 2–2, denoted with an asterisk) was protonated to generate the neutral model, which prevents interactions with this acidic proton that are non-native to B-DNA. For each phosphate model, the methyl group at the C5 position of T was subsequently changed to hydrogen (U) or bromine ([Br]U).

For all phosphate model and sequence combinations, optimizations were performed in both the gas phase and water. Full (water) solvation was accounted for implicitly using the IEF-PCM formalism[165] and a dielectric of 78.4. Although an isolated dinucleoside monophosphate is fully solvated, the conditions are more complex for a dinucleoside unit within double-stranded DNA, which likely experiences a range of environments

(polarities).[166,167] Indeed, a series of dielectric constants have been found to describe the regions within, and surrounding, DNA.[168-172] Due to this complexity, computational studies have used a variety of dielectric constants ranging between the gas phase and water extremes to mimic a DNA environment (see, for example Refs. 88, 89, 114, 117, 119, 120, 131, 137, 141, 149, 158 and 173). Therefore, a dielectric of 78.4 was chosen in order to evaluate the structural differences at the two (relevant) polarity extremes. The dsDNA environment will likely lie between the two (gas and water) environments considered.

Optimizations were performed with the B3LYP, MPWB1K and M06-2X functionals in combination with the 6-31G(d,p) basis set and frequency calculations characterized the resulting structures as minima. This basis set was selected for computational efficiency given the size of the dinucleoside monophosphate model, number of structures considered and intended future applications. As mentioned in the Section 2.1., the DFT functionals were chosen based on their popularity, availability, description of non-covalent interactions and/or success for thermochemical kinetic studies. [21,48,68-73] The choice of each functional is further justified in the Section 2.3.. Although PW91 has been previously used in conjunction with dinucleoside monophosphates to obtain B-like conformations of DNA,[133,145] this functional was not included in the present work since the ability of PW91 to treat radicals and reproduce reaction kinetics is unknown, which makes it an unreliable choice for studying the formation of U$^{\bullet}$-mediated intrastrand cross-links. While DFT-D methods are promising for describing dispersion-dominated systems,[174] they generally require the use of large basis sets, which are not suitable for a model of this size.

Using all three functionals, the nine phosphate model and sequence combinations were optimized in both the gas phase and water, which yields 54 final (optimized) structures. These optimized structures are carefully analyzed in Section 2.3. to identify the

computational model and method combination that most accurately reproduces the conformation of B-DNA. Although different input structures or initial orbital guesses could be used if a suitable structure is not obtained or convergence issues exist upon optimization of the HyperChem-generated structures, a systematic and reproducible approach has been used to build models of both natural and modified DNA. To avoid user bias and intervention, any model and method combination unable to directly yield optimized structures similar to dsDNA using this standard input was discarded. The model and method combination identified through this systematic analysis will therefore be useful for reliably studying systems with unknown structures, which includes intrastrand cross-links.

All calculations were performed using Gaussian 09.[175]

2.3. Results and Discussion

2.3.1. Judging the Accuracy of a Computational Approach

Combining the three phosphate models (anionic, neutral, counterion), three methods (B3LYP, MPWB1K, M06-2X), two environments (gas, water) and three sequences (5′–GT–3′, 5′–GU–3′, 5′–G^{Br}U–3′) yields 54 unique optimized structures. When identifying the most appropriate computational approach, a large emphasis will initially be placed on the ability to reproduce the base–base orientation found in natural B-DNA double helices since this is likely the most difficult geometric feature to reproduce with computational methods. To facilitate this analysis, the angle (designated as φ) between base planes (defined as the mean plane generated using endocyclic heavy atoms) was quantified using the Mercury software package.[176] The φ angle has been previously used to quantify base–base orientations in DNA[133,145,146] and combines the values of tilt and roll commonly used to describe nucleic acid structure.[4] To further aid structural characterization, a categorization scheme was developed to classify structures as: (1) distorted, (2) repelled, (3) tilted

($\varphi > 30°$) or (4) stacked ($\varphi < 30°$). An example of each classification is given in Figure 2–4. In distorted structures (Figure 2–4 (a)), the nucleobases do not interact with each other, but instead at least one base twists to form nucleobase–backbone interactions through hydrogen or halogen bonds. This orientation differs from the repelled configurations (Figure 2–4(b)) in which the nucleobases remain distant from each other, but do not interact with the backbone. In tilted structures (Figure 2–4(c)), π–π contacts are present between the bases, but the relative base–base orientation has substantial deviations from planarity. In contrast, stacked structures (Figure 2–4(d)) have near-parallel base–base arrangements due to attractive π–π interactions. While distorted and repelled base arrangements are not found in B-DNA, variations in base–base orientations commonly exist. The φ angle provides a means of quantifying these variations, where computational approaches yielding significant deviations from stacked structures (i.e., $\varphi > 30°$) will be discarded.
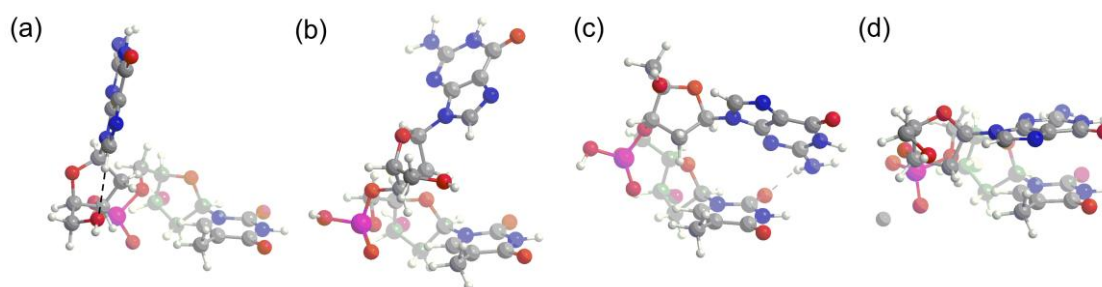
Figure 2–4: Representation of the four classifications of optimized base–base orientations: (a) distorted, (b) repelled, (c) tilted and (d) stacked.

Due to the large number of configurations considered, only structures of the optimized 5′–GT–3′ sequence, which are representative of all structures and can be directly compared to natural DNA, are provided in Figures 2–5 to 2–7. The remaining 5′–GU–3′ and 5′–G^{Br}U–3′ sequences are also discussed in the text, while the corresponding data is

included in Appendix A (Figures A–1 to A–6). In the discussion below, the relative base–

base orientation will first be considered to identify non-stacked structures. Since there is a

strong structural dependence on the functional used, the results for each functional will be

presented in turn. Second, important backbone torsion angles will be analyzed and

compared to experimental crystal structure data to select the most accurate computational

approach.

2.3.2. Analysis of the Base–Base Orientation

2.3.2.1. B3LYP

The popular (hybrid-GGA (generalized gradient approximation)) B3LYP functional

has dominated the computational chemistry literature since its development, as evidenced

by its use in 80% of all DFT studies between 1999 and 2006.[148] Successful applications of

this functional include the description of hydrogen-bonding interactions[177,178] and

radicals.[179] Despite the inability of B3LYP to describe stacking interactions,[148,153] it has been

applied to many biological systems including structural and conformational studies of

natural and modified DNA using models comprised of backbone

components,[124,125,130,134,136,143] nucleosides[126,137,180,181] and nucleotides,[127-129,132,135,137,144], as

well as larger and more complicated DNA systems.[182,183] In relation to future applications of

the present work, this functional has been used to study radical reactions between adjacent

nucleobases. For example, CPD formation between adjacent T nucleobases has been

examined.[117,119] B3LYP has also been used in conjunction with a dinucleoside

monophosphate and neutral (protonated) phosphate model to examine hydrogen

abstraction[88] resulting from nucleobase radicals, as well as the formation of T-containing

intrastrand cross-links that result following the exposure of natural DNA to ionizing

radiation.[114,115] While these studies observed larger inter-base distances than found in

natural DNA, B3LYP was deemed to provide sufficiently accurate structures by the authors.[114,115] Due to this rather surprising conclusion that B3LYP can describe dinucleoside monophosphates, coupled with the prominent use of this functional to examine the properties of DNA and the kinetics of the associated reactions, the ability of B3LYP to produce a conformation of a dinucleoside monophosphate characteristic of B-DNA is first considered.

Table 2–1: Classification of the dinucleoside monophosphate structures optimized with B3LYP, as well as the angle between the base planes (φ (°) in parentheses).[a–c]

B3LYP	Gas			Water		
	Anionic	Neutral	Counterion	Anionic	Neutral	Counterion
5′–GT–3′	distorted (80.2)	tilted (47.5)	tilted (56.8)	distorted (83.4)	repelled (38.4)	distorted (83.4)
5′–GU–3′	distorted (86.9)	tilted (48.5)	tilted (56.1)	distorted (87.6)	repelled (43.4)	distorted (78.6)
5′–G^{Br}U–3′	distorted (11.1)	tilted (37.7)	tilted (53.7)	distorted (67.9)	distorted (46.1)	distorted (33.0)

[a]For a representative illustration of the classifications, see Figure 2–4. A full description of the classifications is provided in Section 2.3.1.. [b]The corresponding optimized structures are provided in Figures 2–5, A–1 and A–4. [c]φ is the angle between the mean planes generated using endocyclic heavy atoms of each nucleobase.

All gas-phase B3LYP optimized structures (Table 2–1, Figures 2–5(a–c)) can be classified as either distorted or tilted regardless of the phosphate model used. Distorted structures are obtained for the anionic phosphate model, where at least one nucleobase twists to form close hydrogen contacts with the backbone (Figure 2–5(a)). In the case of the 5′–G^{Br}U–3′ sequence, a halogen bond (d(Br···O) = 3.241 Å) is formed between Br and the terminal 5′–hydroxyl oxygen. Significantly tilted structures are obtained with B3LYP for both the neutral (protonated) and counterion phosphate models (Figures 2–5(b)), where the angle between the nucleobase planes (φ) ranges from 37° to 57° (Table 2–1). Generally, these tilted arrangements are accompanied by pyramidalization of the amino group at C2 of

G, which has been noted in previous structural studies of dinucleoside monophosphates[28-30] and leads to an out-of-plane N–H⋯O hydrogen bond with O4′ of the 3′ sugar and/or the O4 carbonyl group of the pyrimidine (Figure 2–5(b, c)).

Figure 2–5: B3LYP structures of the 5′–GT–3′ dinucleoside monophosphate sequence optimized in the gas phase (a–c) and water (d–f) using the anionic (a, d), neutral (protonated) (b, e) and counterion (c, f) phosphate models.

In addition to the gas-phase results discussed above, optimizations in implicit solvent (water) were considered to determine the effect of the surrounding environment on the dinucleoside monophosphate structure. Solvent-phase optimizations do not improve the gas-phase results of the B3LYP functional (Table 2–1 and Figures 2–5(d–f)). In fact, the structures obtained in solvent are either similar to those obtained in the gas phase or deviate from B-DNA to a greater extent than the gas-phase geometries. For example, the

base orientations for the neutral model of the T and U-containing sequences are tilted in the gas phase, but repelled in water (Figures 2–5 (b) and 2–5 (e), which can likely be attributed to weakening of out-of-plane N–H···O contacts upon solvation. Similarly, the counterion model with all three sequences yields a tilted arrangement in the gas phase that contains an out-of-plane hydrogen bond between N2H of G and O4′ of the 3′ sugar, but a distorted arrangement in water that lacks this contact.

Thus, while previous studies deemed that gas-phase B3LYP calculations on a neutral dinucleoside monophosphate sufficiently recreate the structural features of DNA,[114,115] the above work clearly shows this approach is unacceptable since highly distorted geometries are obtained upon optimization. Therefore, even though the popular B3LYP functional has been successfully used in many biochemical applications,[137,177-179] its failure to describe stacking interactions, even in the presence of the physical restraint provided by the backbone, indicates that this functional should not be used to examine extended DNA systems. Indeed, previous B3LYP studies have required computational constraints between adjacent nucleobases to obtain a representative dsDNA geometry.[117,119] The following sections consider functionals that were designed to account for a range of non-covalent interactions, including those with large dispersion contributions.

2.3.2.2. MPWB1K

The (hybrid meta-GGA) MPWB1K functional was designed for thermochemistry, thermochemical kinetics and non-covalent interactions.[151] Computational benchmarks have illustrated that MPWB1K describes stacking interactions in biological systems.[121,184,185] Other successful biological applications of this functional include the description of C–H···π interactions involving aromatic amino acids,[186] ONIOM treatment of DNA–protein interactions,[187] and the reproduction of CCSD(T) barriers for phosphodiester bond

hydrolysis.[188] Most relevant to the present work, MPWB1K has been used in conjunction with a (neutral) dinucleoside monophosphate model to study hydrogen abstraction, and a stacked base arrangement was found.[89] These previous successes make MPWB1K a logical choice to consider in the present work.

Table 2–2: Classification of the dinucleoside monophosphate structures optimized with MPWB1K, as well as the angle between the base planes (φ (°) in parentheses).[a–c]

MPWB1K	Gas			Water		
	Anionic	Neutral	Counterion	Anionic	Neutral	Counterion
5′–GT–3′	tilted (38.1)	tilted (47.4)	stacked[d] (12.5)	tilted (33.9)	tilted (41.4)	stacked[d] (8.0)
5′–GU–3′	tilted (39.5)	stacked (14.5)	stacked[d] (14.6)	stacked (24.8)	stacked (8.9)	stacked (10.1)
5′–G^{Br}U–3′	tilted (37.3)	stacked (9.2)	stacked (9.3)	tilted (37.6)	distorted (52.8)	stacked[d] (7.8)

[a]For a representative illustration of the classifications, see Figure 2–4. A full description of the classifications is provided in Section 2.3.1.. [b]The corresponding optimized structures are provided in Figures 2–6, A–2 and A–5. [c]φ is the angle between the mean planes generated using endocyclic heavy atoms of each nucleobase. [d]Due to SCF convergence issues, structures were obtained using M06-2X-optimized structures as input rather than the standard HyperChem-generated structures.

MPWB1K gas-phase optimizations of the dinucleoside monophosphates considered in this study yield unsatisfactory results (Table 2–2, Figures 2–6(a–c)). Tilted structures are obtained for all sequences with the anionic phosphate model, where the angle between base planes (φ) ranges from 37° to 40° (Table 2–2). Optimization in water leads to similar structures with a slight decrease in φ, which ranges between 25° and 38°; however, technical issues create questions about the reproducibility of these results.[iv] When the neutral phosphate model is employed, stacked arrangements are obtained for sequences

---

[iv] Although optimizations of the anionic model in water starting from the same (HyperChem generated) input yield similar, tilted base–base orientations, small deviations exist between these optimized structures that are outside commonly-accepted variations for G09 optimizations.

containing U (φ = 14.5°) and [Br]U (φ = 9.3°). However, a significantly tilted orientation is found for the T sequence (φ = 47.4°), which suggests this method–phosphate model combination cannot uniformly predict dsDNA conformations. Optimizations in water offer no improvement, where the U sequence is stacked, but the T and [Br]U sequences are significantly tilted (φ = 41.4° and 52.8°, respectively).
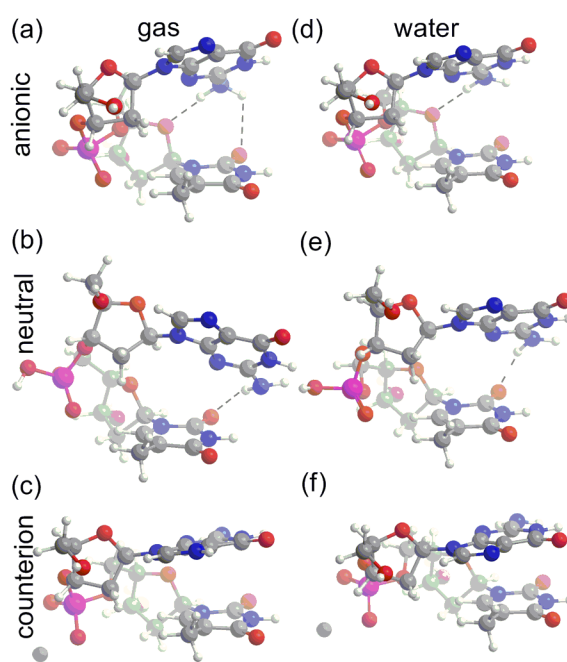
Figure 2–6: MPWB1K structures of the 5′–GT–3′ dinucleoside monophosphate sequence optimized in the gas phase (a–c) and water (d–f) using the anionic (a, d), neutral (protonated) (b, e) and counterion (c, f) phosphate models.

Calculations combining MPWB1K and the counterion phosphate model were plagued with computational challenges. Specifically, SCF convergence problems arose in both the gas phase and water. These issues most commonly occur when generating the molecular orbitals prior to the first optimization step and could not be fixed by providing

alternate orbital guesses. In an attempt to isolate and understand this problem, alternate input geometries were used. For the natural (5′–GT–3′) sequence, attempts to use experimental crystal structures as the input geometry were similarly unsuccessful, which eliminates the HyperChem-generated structures as the source of the problem. Interestingly, convergence could be obtained when geometries optimized with a different level of theory (M06-2X) were used as input,[v] which indicates MPWB1K is remarkably sensitive to the input geometry. Since these convergence problems only occurred for the counterion phosphate model, MPWB1K may have difficulty describing the orbitals of (Na^+) ions at certain geometries with respect to the deoxyribose-phosphate backbone.

While MPWB1K treats stacking interactions better than B3LYP and has been successfully applied in other biological studies,[89,121,184,186-189] it is not recommended when attempting to replicate the features of a dinucleoside monophosphate in dsDNA. Furthermore, while it is unknown whether the technical problems encountered in the present work will occur in other applications of MPWB1K, researchers interested in using this functional are cautioned to carefully test its performance for the particular problems and systems of interest.

2.3.2.3. M06-2X

Similar to MPWB1K, the (hybrid meta-GGA) M06-2X functional was parameterized to describe non-covalent interactions and kinetics.[155] Indeed, many computational benchmarks focusing on biological systems have concluded that M06-2X is a useful method

---

[v] Optimizations with MPWB1K and the counterion phosphate model, when using the HyperChem-generated inputs, were successful for the 5′–G^{Br}U–3′ sequence in the gas phase and the 5′–GU–3′ sequence in water. For the remaining sequences (5′–GT–3′ and 5′–GU–3′ in gas and 5′–GT–3′ and 5′–G^{Br}U–3′ in water), SCF convergence was too problematic when using the HyperChem-generated input structures and convergence could only be obtained when using M06-2X-optimized structures as input.

for studying π–π interactions.[154,156,185,189-194] For example, recent works on biomolecular binding motifs,[194] and DNA–protein π–π interactions,[185] as well as stacked A,[193] T[190] and U[190] dimers, found that M06-2X yields binding strengths comparable to CCSD(T). In fact, one study revealed that M06-2X more consistently recovers CCSD(T)/CBS π–π interaction energies than MPWB1K.[185] Furthermore, the successful examination of singlet-triplet transitions,[195] electronic excitation energies[196] and bond dissociation energies[197] with M06-2X is indicative of the general utility of this functional, as well as its likely success in studies of intrastrand cross-link formation and other reactions involving DNA radicals. Thus, M06-2X is a good candidate for investigating the structure of natural and modified DNA sequences. It has been reported that artificial oscillations exist on M06-2X potential energy surfaces describing weak, dispersion interactions, such as the benzene dimer, when using small integration grids.[198] However, no such oscillations were observed for DNA–protein π–π interactions with M06-2X and default integration grids.[185] Therefore, this artefact is not anticipated to affect the results of the present study.

Table 2–3: Classification of the dinucleoside monophosphate structures optimized with M06-2X, as well as the angle between the base planes (φ (°) in parentheses).[a–c]

M06-2X	Gas			Water		
	Anionic	Neutral	Counterion	Anionic	Neutral	Counterion
5′–GT–3′	stacked (28.8)	distorted (43.6)	stacked (12.9)	stacked (4.9)	distorted (36.2)	stacked (4.7)
5′–GU–3′	stacked (13.3)	distorted (41.4)	stacked (5.8)	stacked (3.8)	distorted (37.6)	stacked (5.1)
5′–G^{Br}U–3′	stacked (28.8)	stacked (7.6)	stacked (8.3)	stacked (5.6)	stacked (5.4)	stacked (5.0)

[a]For a representative illustration of the classifications, see Figure 2–4. A full description of the classifications is provided in Section 2.3.1.. [b]The corresponding optimized structures are provided in Figures 2–7, A–3 and A–6. [c]φ is the angle between the mean planes generated using endocyclic heavy atoms of each nucleobase.

As discussed for MPWB1K, a stacked nucleobase arrangement is not uniformly predicted for all sequences by M06-2X gas-phase optimizations with the neutral phosphate model (Table 2–3). Specifically, M06-2X yields a stacked 5′–G[Br]U–3′ sequence ($\varphi = 7.6°$), but distorted T- and U-containing sequences ($\varphi = 43.6°$ and $41.4°$, respectively). The corresponding optimizations in water remain inconsistent, where a stacked geometry is again only obtained for [Br]U. Since none of the functionals considered in the present work consistently predict stacked base–base arrangements for all nucleobase sequences when used with the neutral phosphate model, the problem must lie with the phosphate model rather than the functional employed. Despite previous applications of a neutral phosphate model in the literature,[88,89,114,115,120,127-129,132,137] this model cannot consistently reproduce the structure of dsDNA for the (5′–G–pyrimidine–3′) sequences considered and is eliminated as a viable phosphate model.
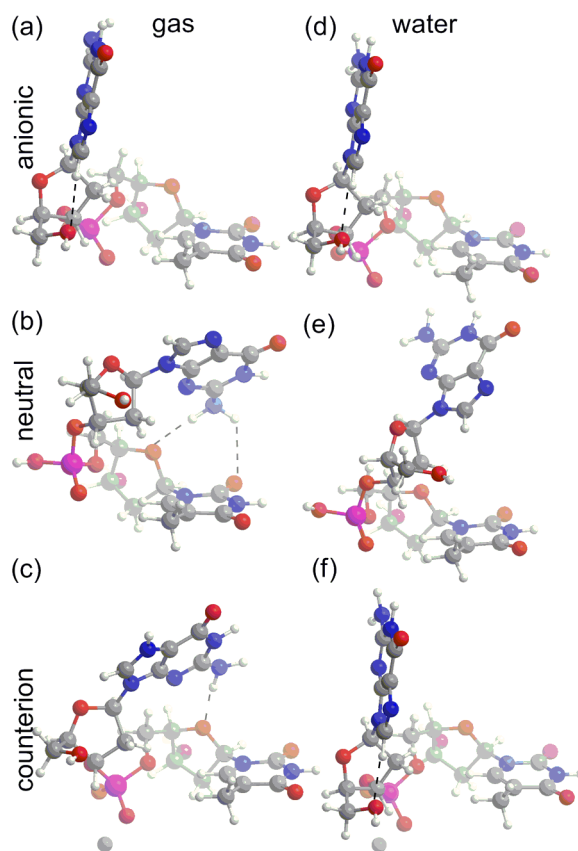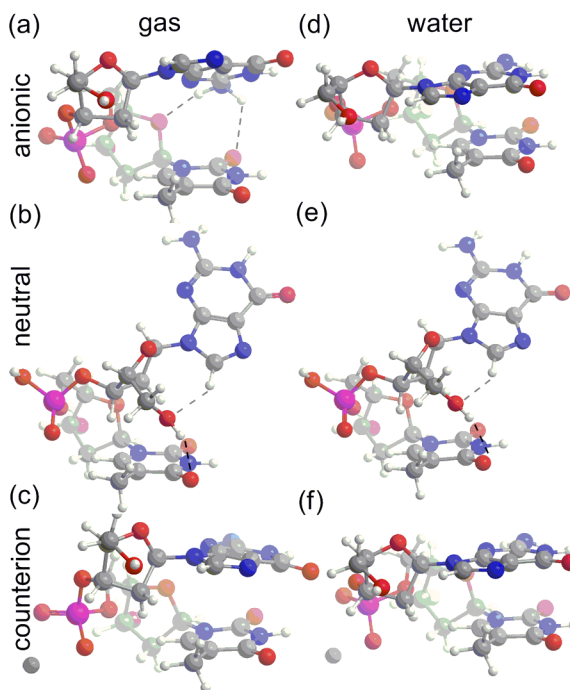

Figure 2–7: M06-2X structures of the 5′–GT–3′ dinucleoside monophosphate sequence optimized in the gas phase (a–c) and water (d–f) using the anionic (a, d), neutral (protonated) (b, e) and counterion (c, f) phosphate models.

Despite similar intended applications of the M06-2X and MPWB1K functionals,[151,155] the optimized structures obtained with the anionic dinucleoside monophosphate model are substantially different (Tables 2–2 and 2–3, Figures 2–6 and 2–7). Indeed, M06-2X consistently predicts stacked base arrangements in both the gas phase and water. Specifically, the angle between base planes (φ) in the gas-phase structures ranges between 13° and 29°, while the corresponding structures in water yield more parallel base–base arrangements ($4° < \varphi < 6°$). These results emphasize the superior performance of the anionic phosphate model over the protonated counterpart, and the ability of M06-2X to describe stacked arrangements in biomolecules, which is supported by recent literature.[154,156,185,189-194]

Like the anionic model, the counterion phosphate model yields stacked base–base arrangements for all sequences when optimized with M06-2X. However, the base planes are more parallel for the counterion model in the gas-phase ($6° < \varphi < 13°$) than the anionic model, whereas the inclusion of a counterion has a smaller effect when optimizations are performed in water ($\varphi < 5°$). For the T-containing sequence in the gas phase, the φ angle (12.9°) is similar to the value (8.7°) reported in a previous (gas-phase) study with the PW91 functional and a (Na^+) counterion phosphate model.[145] Based on this structural analysis, M06-2X optimizations with either the anionic or counterion phosphate model will provide stacked nucleobase arrangements using a dinucleoside monophosphate model of natural or modified DNA. In addition to predicting a stacked base arrangement, this functional reproduces other important structural features of dsDNA. For example, the nucleobases are in the DNA-preferred *anti* orientation ($\chi = 180 \pm 90°$, Figure 2–2)[4] about the glycosidic bond and close contacts (such as those between C8–H of G and the terminal O5′ hydroxyl group, or between C6–H of the pyrimidine and O5′)[4] are predicted. These features further testify to the ability of these combinations to mimic dsDNA structures.

Although optimizations in water lead to slightly more planar nucleobase arrangements than optimizations in the gas phase, the relative (stacked) base–base orientations obtained under the two environmental extremes are very similar. This indicates that π–π stacked nucleobase geometries are relatively unaffected by the polarity of the environment when using this computational approach, and optimizations in environments of intermediate polarity should also yield stacked nucleobase arrangements. Furthermore, the base–base orientation is more dependent on the chosen density functional and phosphate model than the environmental conditions imposed. Nevertheless, solvent may play a larger role in other aspects of DNA structure, such as the many torsional degrees of freedom in the DNA backbone. Therefore, to further assess the ability of the computational model and method combinations to reproduce dinucleoside monophosphate conformations relevant to B-DNA, the structure of the deoxyribose–phosphate backbone is next considered.

2.3.3. Analysis of the Deoxyribose-Phosphate Backbone Orientation

In the analysis of the backbone structure, important calculated torsion angles (Figure 2–2) will be compared to experimental averages[6] (Table 2–4) determined by a statistical analysis of structures in the Nucleic Acid Database with resolutions better than 1.9 Å that do not contain modifications, mismatches or molecules bound. Since an experimental crystal structure represents an average geometry, the flexible and dynamic nature of DNA, including a variety of possible backbone conformations, is not captured in this statistical analysis. Nevertheless, in cases where large differences exist between the optimized and (average) experimental backbone torsion angles, a comparison of the calculated and experimental structures will be used to rule out computational model and method combinations.

Although a comparison of the calculated and experimental backbone conformations is provided in Appendix A for all computational models, methods and sequences (Tables A–1 to A–3), the discussion within the text will focus on the computational method and model combinations that consistently yield stacked base–base orientations relevant to B-DNA (M06-2X with the anionic and counterion phosphate models in the gas phase and water). Table 2–4 provides a detailed comparison of the calculated and experimental backbone structures for the natural 5′–GT–3′ sequence.

Table 2–4: Backbone torsion angles (°) and deoxyribose pseudorotational phase angles (P, °) calculated with M06-2X for the 5′–GT–3′ sequence, as well as the average value (Exp.) and standard deviation (SD) obtained from experiment.[a,b]

∠[a]	Gas				Water				Exp.[c]	SD[c]
	Anionic		Counterion		Anionic		Counterion			
χ5′	210.3	[33.7]	236.8	[7.2]	227.2	[16.8]	230.0	[15.0]	258	14
δ5′	148.1	[7.1]	147.2	[6.2]	149.8	[8.8]	149.8	[8.8]	128	13
ε	161.9	[11.1]	170.6	[2.4]	158.0	[15.0]	155.8	[17.2]	184	11
ζ	279.0	[4.0]	264.9	[0.0]	276.8	[1.8]	277.2	[2.2]	265	10
α	271.3	[11.7]	295.1	[0.0]	274.6	[8.4]	274.4	[8.6]	298	15
β	250.5	[65.5]	197.2	[12.2]	240.7	[55.7]	236.8	[51.8]	176	9
γ	56.1	[0.0]	54.5	[0.0]	48.0	[0.0]	50.2	[0.0]	48	11
χ3′	277.9	[28.9]	257.9	[8.9]	265.0	[13.0]	263.8	[14.8]	241	8
δ3′	155.4	[14.4]	149.8	[8.8]	154.0	[16.0]	152.5	[11.5]	128	13
P5′	165.7		165.9		170.6		172.6		144–190[d]	
P3′	199.5		171.2		178.7		172.8		144–190[d]	

[a]See Figure 2–2 for definitions of backbone torsion angles. [b]Each value in square brackets represents the difference between the calculated angle and experimental average within one standard deviation, where a value of zero indicates that the calculated angle falls within the standard deviation of the experimental average. [c]See Ref. 6. [d]See Ref. 4.

For all computational model and method combinations that consistently yield a stacked base–base configuration, the smallest deviations from experiment occur in the ζ and γ torsion angles (Tables 2–4 and A–1 to A–3). The largest deviations occur in the β and χ angles. Deviations in χ, which controls the relative arrangement of the nucleobase with respect to the sugar moiety, are expected based on the known flexibility of DNA[4] and the

lack of stabilization from additional surrounding nucleobases. Since experiments reveal that a large correlation between the χ and β torsion angles is intrinsic to the dinucleoside monophosphate structure,[4] it is not surprising that large deviations also occur in the β angle in these models.

The backbone torsion angles and pseudorotation phase angles (sugar puckering) obtained using the counterion model are closer to the experimental values than those obtained using the anionic model (Table 2–4). Similarly, Polteva *et al.*[133] previously reported that charge-neutralized dinucleoside monophosphates better mimic characteristics of experimental crystal structures than anionic models. The effect of charge neutralization is more pronounced in the gas phase than water, where electrostatic interactions are weaker. Since charge neutralization also leads to more planar base arrangements (Table 2–4), the counterion phosphate model provides superior structural data and is the phosphate treatment of choice for use in conjunction with DFT methods to mimic large dsDNA systems.

Although the size of the basis set implemented in the present work was restricted to ensure that the best approach identified would be practical for studying U$^{\bullet}$-induced intrastrand cross-link formation, diffuse functions may affect the results obtained, especially for anionic systems. To verify that the poorer results obtained with the anionic phosphate model are not due to a lack of diffuse functions, optimizations of the 5′–GT–3′ sequence were performed using the 6-31+G(d,p) basis set and the anionic and counterion phosphate models in conjunction with M06-2X. Analysis of the optimized geometries (Table A–4) confirms that the charge-neutralized model provides superior results to the corresponding anionic model even upon the inclusion of diffuse functions. Therefore, the conclusions drawn above are valid.

The gas-phase backbone configurations obtained from optimizations of the counterion phosphate model are closer to experiment than those obtained in water. Nevertheless, inclusion of (implicit) water does not significantly affect the backbone torsional angles, and the global structures are qualitatively similar (Figures 2–5 to 2–7, A–1 to A–6). If other (kinetic) applications of the M06-2X counterion phosphate model require the inclusion of environmental effects, data indicates that (implicit) water can be incorporated into the calculation while maintaining structural accuracy. Since the structure of the dinucleoside monophosphate (relative base–base orientation, backbone torsional angles) is similar in both the gas phase and fully-solvated environments, comparable structures are expected for environments of intermediate polarity. Most importantly, it is clear from the results presented here that the choice of functional and phosphate model is more important for accurate reproduction of the conformation of a dinucleoside monophosphate in dsDNA than the environmental conditions. Taken together, these results indicate M06-2X in conjunction with the counterion model (in a variety of environments) should be used to study radical-induced intrastrand cross-link formation. This approach will also be applicable to studying structures and/or reactions involving adjacent natural or modified DNA components.

2.4.   Conclusions

In this chapter, 54 optimized structures of dinucleoside monophosphates were obtained by combining three methods (B3LYP, MPWB1K, M06-2X), three phosphate models (anionic, neutral, counterion), two environments (gas, water) and three sequences (5′–GT–3′, 5′–GU–3′, 5′–G^{Br}U–3′). These structures were compared to average structural parameters of natural (double-stranded) B-DNA to identify the optimum computational approach to study intrastrand cross-link formation. Despite the natural restraint provided by the phosphate backbone in the structural model, the B3LYP functional does not

accurately predict a characteristic B-DNA structure due to its inability to describe dispersion interactions. Furthermore, many technical difficulties were encountered with MPWB1K, which may limit the use of this functional. In contrast, depending on the phosphate model employed, stacked base–base orientations that are relevant to B-DNA could be obtained for all sequences with M06-2X.

Although a neutral (protonated) DNA phosphate model has been widely used in previous literature, this model cannot consistently predict stacked base–base alignments. In contrast, both the M06-2X anionic and counterion phosphate models yield stacked nucleobase arrangements. However, the counterion model yields slightly more planar relative base–base orientations and backbone torsion angles more comparable to experimental crystal structures of B-DNA. This result emphasizes the importance of charge neutralization in DNA models, and suggests that the counterion phosphate model should be used in conjunction with M06-2X/6-31G(d,p) to mimic larger dsDNA systems.

Comparison of two environmental extremes (gas phase and bulk water) shows that slightly more parallel base arrangements are calculated in water, but gas-phase backbone torsion angles are more representative of average crystal structure data. Nevertheless, since these differences are generally small and DNA is flexible, it is proposed that the global structure is relatively unaffected by the surrounding medium. Therefore, the use of M06-2X optimizations of a dinucleoside monophosphate with the counterion phosphate model in either the gas phase or water (or presumably an environment with intermediate polarity) are recommended to reliably mimic the structure of this unit within B-DNA.

The computational approach identified in this chapter will be useful for determining the local structure of natural or modified DNA for which experimental structural data is missing. This methodology will also be useful for studying a variety of reactions between

adjacent nucleic acid components, including reactions between two nucleobases or a nucleobase and a sugar moiety. Indeed, this model is used throughout the remaining chapters to study the structure and formation of U-containing intrastrand cross-links.

Chapter 3. The Formation and Structure of the 5′–G[8–5]U–3′ Intrastrand Cross-link[vi]

3.1.   Introduction

As discussed in Section 1.5.3., four DNA intrastrand cross-links have been detected following UV irradiation of [Br]U-containing cellular DNA. Although the connectivity of these intrastrand cross-link products has been determined,[66-68] the proposed reaction mechanism[66] has not been studied and no information exists on the orientation of the cross-links in DNA or their effects on the structure of the double helix. Therefore, this chapter begins by studying the most abundant intrastrand cross-link, 5′–G[8–5]U–3′ (Figure 3–1). Initially, the formation mechanism is determined under conditions relevant to UV radiation (5′–G[•+]U[•]–3′ reactant), which corresponds to the experimental conditions under which these lesions have been observed. Next, formation is studied from the 5′–GU[•]–3′ reactant, which is relevant to ionizing irradiation of [X]U-containing DNA and UV irradiation of [I]U since C5–I may undergo homolysis.[70] Since ionizing radiation is typically used in clinical applications, this work will establish whether 5′–G[8–5]U–3′ is likely to be generated under therapeutically-relevant conditions. Finally, the structures of the cross-link product, as well as cross-linked DNA, are analyzed and compared to that of natural DNA. Together, the calculations and discussion in this chapter will provide an enhanced understanding of the structure and formation mechanism of 5′–G[8–5]U–3′, and unveil the potential contribution of this lesion to the ability of [X]U to be effectively used in conjunction with radiation as an anti-tumour therapy.

[vi] Reprinted in part with permission from Churchill, C.D.M., Eriksson, L.A., Wetmore, S. D., Chem. Res. Toxicol., submitted for publication; manuscript ID: tx-2011-003239. Copyright 2011 American Chemical Society.
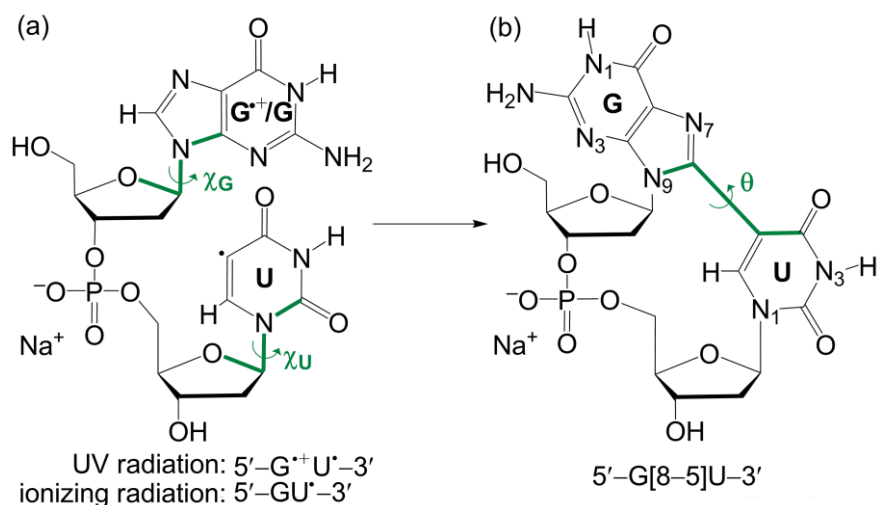
Figure 3–1: The structure and numbering of the (a) 5′–G$^{•+}$U$^{•}$–3′ and 5′–GU$^{•}$–3′ reactants and (b) the 5′–G[8–5]U–3′ product. The χ_G ($\angle$(O4′C1′N9C4)), χ_U ($\angle$(O4′C1′N1C2)) and θ ($\angle$(N9C8C5C4)) torsion angles describing the nucleobase orientations are indicated in bold.

3.2. Computational Details

3.2.1. DFT Calculations

The 5′–G[8–5]U–3′ lesion was studied using the computational approach recommended in Chapter 2. Specifically, a dinucleoside monophosphate model was complexed with a sodium counterion (Na$^+$) and optimizations were performed with IEF-PCM-M06-2X/6-31G(d,p) and a dielectric constant corresponding to water (ε = 78.4).[155,165] Although the results in Chapter 2 found no significant structural differences between gas- and solvent-phase optimizations, the effects of solvation were required for studying cross-link formation to naturally maintain the interaction between Na$^+$ and the phosphate group, whereas gas-phase optimizations resulted in the cation migrating away from the backbone as structural changes occurred along the reaction pathway. The unrestricted formalism was used for systems with unpaired electrons. Frequency calculations were performed to confirm the nature of all stationary points, and calculate scaled (0.9580)[199] zero-point

63

corrected energies (ΔE_{ZPVC}), scaled (0.9470)[199] thermal corrections to the enthalpy (ΔH) and unscaled thermal corrections to the Gibbs free energy (ΔG) under standard conditions (1 atm and 298.15 K). To obtain ΔG, the solvation energy (including the non-electrostatic component) was calculated with the SMD solvation model.[200] The reaction coordinates were confirmed by following the imaginary mode corresponding to the transition state in both directions to the neighbouring minima. Throughout the reaction pathway, the terminal C5′-hydroxyl group was constrained to eliminate interactions non-native to natural B-DNA ($\angle$(HO5′C5′C4′) = 180°). All energies are reported relative to the reactant. The relative stability of stationary points is discussed using the optimization energy (ΔE), unless otherwise noted. All DFT calculations were performed with Gaussian 09.[175]

3.2.2. *Molecular Dynamics Simulations*

To gain insight into the conformation(s) of the intrastrand cross-link within DNA, as well as helical distortion due to the formed cross-link, 55 ns molecular dynamics simulations of DNA containing 5′–G[8–5]U–3′ were performed using YASARA.[201] Double-stranded complementary B-DNA was generated with the 5′–d(GCATGGC**GT**GCTATGC)–3′ sequence using the *nucgen* module of AmberTools.[202] This sequence was chosen due to formation of 5′–G[8–5]U–3′ in the synthetic 5′–d(ATGGCG[Br]UGCTAT)–3′ oligonucleotide following UV exposure,[66] and include additional capping GC bases to reduce unwinding of the helix during the simulation.[203] Cross-linked DNA was generated by overlaying the backbone of the M06-2X-optimized dinucleoside monophosphate product onto the central 5′–GT–3′ sequence (bold). Simulations were performed using the AMBER03 forcefield[204] for the natural components and GAFF[205] (Table B–1) for the cross-linked dinucleotide (Figure B–2). For comparison, the corresponding natural strand was simulated for 28 ns. Simulation snapshots were saved every 25 ps and analysis was performed on all structures following a 3 ns equilibration. DNA strands were analyzed using the Curves+ program.[206] Further

64

details of the simulation (page B–5), including the atom types and charges used (Table B–1), can be found in Appendix B.

3.3. Results and Discussion

3.3.1. Cross-link Formation From UV Radiation

The 5′–G$^{\bullet+}$U$^{\bullet}$–3′ biradical reactant formed upon UV exposure of ^{X}U-containing DNA is initially in a triplet state (Figure 3–2(a)). The fully-optimized reactant (R$_T$, Figure 3–3) is characterized by an *anti* orientation of the nucleobases (χ_G = 255.4° and χ_U = 277.5°,[vii] Figure 3–1), a stacked base-base orientation and a slightly nonplanar U base (R$_T$, Figure 3–3). Both sugar moieties adopt the C2′–*endo* puckering characteristic of B-DNA. Mulliken spin densities reveal an unpaired electron localized at C5 of U (1.048 e$^-$) and an unpaired electron highly delocalized on G (C8 = 0.314 e$^-$, C5 = 0.312 e$^-$, N3 = 0.184 e$^-$, O6 = 0.132 e$^-$ and N2 = 0.104 e$^-$). Cross-link formation from UV radiation proceeds from this triplet reactant to a singlet product (5′–G[8–5]U–3′), which requires a spin flip. This can occur via one of two pathways (Figure 3–2(a)), which are discussed below and compared.

3.3.1.1. Pathway (1): Bond Formation on the Lowest-lying Triplet Surface

In the first pathway, bond formation occurs on the triplet surface (Pathway (1), Figure 3–2(a)). Prior to bond formation, a conformational change occurs (Figure 3–3), where U rotates about the glycosidic bond (χ_U) from the stacked R$_T$ (χ_U = 277.5°) to a T-shaped orientation in the intermediate (I1$_T$, χ_U = 179.7°). These minima are connected by

---

[vii] While the classical *anti* domain is defined as χ = 180 ± 90° (Section 1.2.), this particular orientation about the glycosidic bond is classified as high *anti* (270° < χ < 300°), which is common for purines and still allows the Watson-Crick hydrogen bonding to be maintained (See Ref. 6).
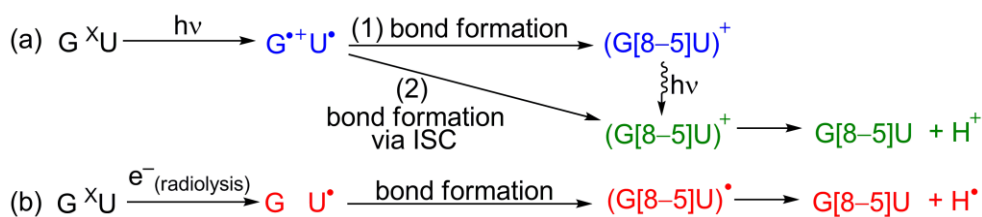
Figure 3–2: Reaction scheme for cross-link formation resulting from (a) UV and (b) ionizing radiation and important stationary points on the lowest-lying triplet state (blue), singlet ground state (green) and lowest-lying doublet surface (red).

a transition state (TS1$_T$) with a small barrier ($\Delta E^{\ddagger}$ = 21.0 kJ mol^{-1} (Figure 3–4(a)), $\Delta G^{\ddagger}$ = 4.3 kJ mol^{-1} (Table 3–1)). Subsequently, association occurs as the distance between C8 of G and C5 of U decreases from I1$_T$ (d(C8–C5) = 4.162 Å) through the transition state (TS2$_T$, d(C8–C5) =2.231 Å) to the cross-linked intermediate (I2$_T$, d(C8–C5) = 1.512 Å). During this process the C8–H8 bond lengthens, C8 becomes sp^3-hybridized and the G glycosidic bond (χ_G) changes from 243.7° (I1$_T$) to 282.7° (I2$_T$). Additionally, a twist (θ, Figure 3–1) of 207.2° occurs between the bonded nucleobases in I2$_T$. Mulliken spin densities for I2$_T$ show that the two unpaired electrons are delocalized over the G (N7 = 0.548 e$^-$, C4 = 0.116 e$^-$, O6 = 0.118 e$^-$ and N9 = 0.096 e$^-$) and U (C5 = 0.696 e$^-$, O4 = 0.151e$^-$ and N1 = 0.254 e$^-$) nucleobases. The barrier for association is 70.5 kJ mol^{-1} ($\Delta G^{\ddagger}$ = 61.7 kJ mol^{-1}) and bond formation releases energy (ΔE = –40.2 kJ mol^{-1}, Table 3–1).

Following cross-link formation, an adiabatic transition from I2$_T$ to the singlet cross-linked intermediate (I2$_S$) facilitates a spin flip, which is accompanied by the emission of phosphorescence (Figure 3–2(a)). These minima share similar DNA backbone conformations and only small rotations occur about the glycosidic bonds when going from I2$_T$ (χ_G = 282.7°, χ_U = 188.7°) to I2$_S$ (χ_G = 298.2°, χ_U = 190.8°). The I2$_S$ intermediate is
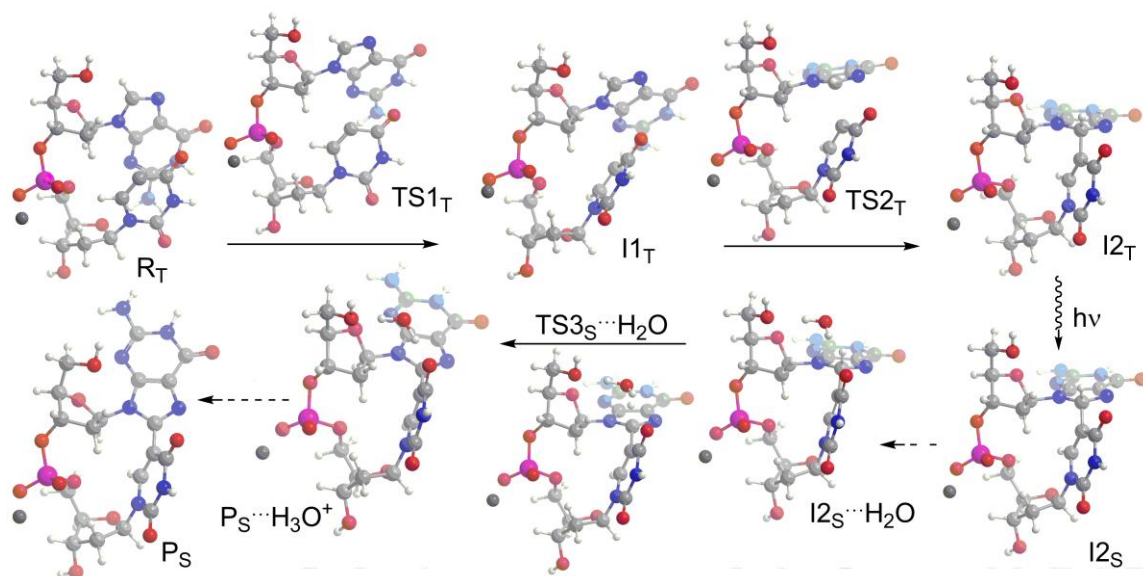
Figure 3–3: Stationary points for formation of 5′–G[8–5]U–3′ from 5′–G•+U•–3′.
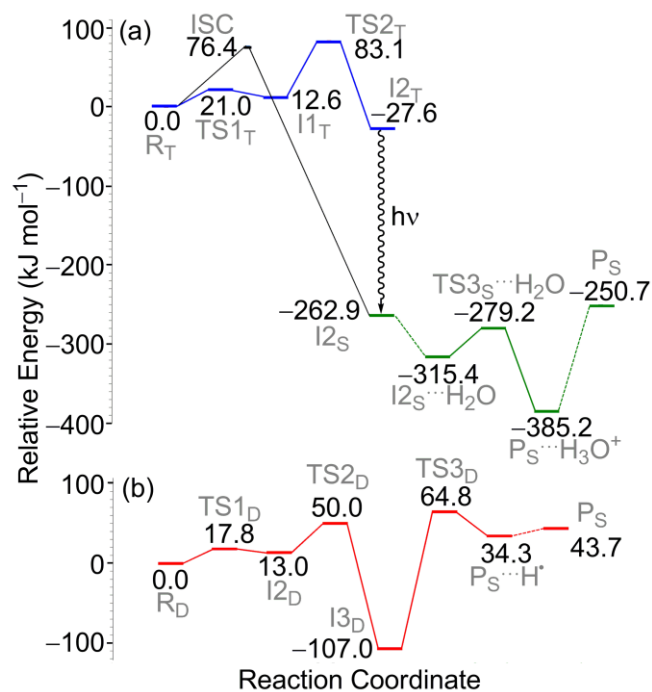
Figure 3–4: Cross-link formation resulting from (a) 5′–G•+U•–3′ on the triplet (blue) and singlet (green) surfaces or (b) 5′–GU•–3′ on the doublet surface (red). M06-2X energies (kJ mol⁻¹) are reported relative to the corresponding reactant.

Table 3–1: Relative energies and thermodynamic parameters (kJ mol^{-1}) for the formation of 5′–G[8–5]U–3′ from 5′–G$^{•+}$U$^{•}$–3′.[a]

Stationary Point[b]	$\Delta E^{c,d}$	$\Delta E_{ZPVC}^{d,e}$	$\Delta H^{d,f}$	ΔG^{g}	$T\Delta S^{h}$
R$_T$ + H$_2$O	0.0	0.0	0.0	0.0	0.0
TS1$_T$ + H$_2$O	21.0	19.3	18.5	4.3	14.2
I1$_T$ + H$_2$O	12.6	11.5	12.2	-5.2	7.0
TS2$_T$ + H$_2$O	83.1	78.5	77.9	56.5	21.4
I2$_T$ + H$_2$O	–27.6	–27.0	–27.4	–63.1	35.7
I2$_S$ + H$_2$O	–262.9	–255.8	–257.2	–303.3	46.1
I2$_S$···H$_2$O	–315.4	–294.7	–300.5	–268.9	–31.6
TS3$_S$···H$_2$O	–279.2	–270.7	–278.7	–237.1	–41.6
P$_S$···H$_3$O$^+$	–385.2	–369.5	–377.3	–344.1	–33.2
P$_S$ + H$_3$O$^+$	–250.7	–238.9	–241.0	–307.9	66.8

[a]Reactant generated from exposure of 5′–G^{x}U–3′ to UV radiation. [b]Refer to Figure 3–3 for the structures of corresponding stationary points. [c]Optimization energy. [d]Calculated with IEFPCM-M06-2X/6-31G(d,p). [e]Includes scaled (0.9580) zero-point vibrational energy correction. [f]Includes scaled (0.9480) thermal correction to the enthalpy. [g]Calculated using the SMD-M06-2X/6-31G(d,p) energy (including non-electrostatic component) and IEF-PCM-M06-2X/6-31G(d,p) thermal correction to the Gibbs Free Energy. [h]Calculated as TΔS = ΔH–ΔG.

characterized by d(C8–C5) = 1.505 Å and θ = 205.4°. The transition from I2$_T$ to I2$_S$ involves a large decrease in energy ($\Delta(\Delta E)$ = –235.3 kJ mol^{-1}). The vertical transition from I2$_T$ to the singlet surface releases a 178.9 kJ mol^{-1} photon, which corresponds to a wavelength of light (668.7 nm) in the visible region.

From I2$_S$, a proton is removed (Figure 3–2(a)) to obtain the experimentally-observed cross-linked product (P$_S$). To model this step, it was necessary to include a discrete water molecule, I2$_S$···H$_2$O. In the transition state for proton abstraction (TS3$_S$), the C8–H8 distance elongates (from 1.098 Å to 1.330 Å) and a barrier of 36.2 kJ mol^{-1} ($\Delta G^{\ddagger}$ = 31.8 kJ mol^{-1}, Table 3–1) is overcome to form a product complex (P$_S$···H$_3$O$^+$). In the final step of the reaction pathway, a decomplexation step yields the infinitely-separated products (P$_S$ + H$_3$O$^+$, Figure 3–3). Removal of the proton changes χ_G from 298.2° (I2$_S$) to 6.0° (P$_S$) as C8 becomes sp^2-hybridized, and negligibly affects χ_U.

3.3.1.2. Pathway (2): Bond Formation via an ISC

In contrast to the reaction pathway described above, bond formation may be coupled to an intersystem crossing (ISC, Figure 3–2(a)) between the triplet reactant (R_T) and the singlet cross-linked intermediate ($I2_S$). Since an ISC cannot be located using conventional optimization procedures, a coordinate-driven approach is used similar to that employed to study cyclobutane pyrimidine dimer formation.[119] Specifically, the lowest-lying triplet potential energy surface was examined as a function of the distance between the C8 atom of G and the C5 atom of U by constraining the C8–C5 bond and relaxing all other coordinates. The C8–C5 distance was systematically decreased from R_T in 0.2 Å increments between 4.2 Å and 1.6 Å. Subsequently, vertical excitations were performed from each optimized point to the singlet ground-state surface. The ISC is the intersection of these triplet and singlet surfaces, which was found to occur between 2.0 and 2.2 Å (Figure B–1).

Refined scans in 0.05 Å increments identified the ISC at approximately 2.09 Å at an energy of 76.4 kJ mol^{-1} relative to R_T. This energetic cost is attributed to the geometric distortion required to reach the ISC, which primarily changes χ_G, χ_U and the C8–C5 distance. Following an instantaneous spin flip at d(C8–C5) = 2.09 Å, the cross-linked intermediate undergoes vibrational relaxation to $I2_S$. During the bond-formation step, the C8–H8 bond length increases from 1.085 Å in R_T to 1.098 Å in $I2_S$. Following cross-link formation, proton removal from C8 occurs on the singlet surface in the same manner discussed above (from $I2_S$ to P_S, Figures 3–3 and 3–4(a)).

3.3.1.3. Comparison of UV Bond Formation Pathways

The UV formation pathways involve a large decrease in energy (ΔE_{Rxn} = −250.7 kJ mol^{-1}, Table 3–1). Both the associative and dissociative steps are spontaneous processes (Table 3–1) and the overall reaction is exothermic (ΔH_{Rxn} = −241.0 kJ mol^{-1}) and exergonic

(ΔG_{Rxn} = –307.9 kJ mol^{-1}). The majority of this stabilization is gained upon transition from the lowest-lying triplet state to the singlet ground state (Figure 3–4(a)), where energy is released vibrationally (ISC) or as a photon (phosphorescence). Regardless of the mechanism, the bond formation step has the largest barrier ($\Delta E^{\ddagger}$ = 70–76 kJ mol^{-1}, Figure 3–4(a)). Both pathways presented are viable mechanisms for cross-link formation from exposure to UV radiation and may occur experimentally. However, it should be possible to identify if bond formation occurs on the triplet surface due to the emission of phosphorescence (at approximately 668 nm; Pathway (1), Figure 3–2(a)). In theory, this signal should be distinguishable from the fluorescence (327 nm)[207] or phosphorescence (400-550 nm)[208] of natural DNA. However, the detection of phosphorescence will likely be complicated due to the low temperatures required for a measurable intensity[208] and low cross-link yields.[66,67]

3.3.2. *Cross-link Formation From Ionizing Radiation*

The optimized 5′–GU•–3′ doublet reactant (R_D, Figure 3–5) contains planar, stacked nucleobases in the high *anti* orientation.[vii] The 5′-deoxyribose adopts the sugar puckering of natural B-DNA, while the 3′-deoxyribose adopts a C3′–*exo* pucker. Mulliken spin densities show the unpaired electron is localized on C5 of U (1.054 e$^-$). The reaction proceeds from this reactant along the lowest-lying doublet surface to a cross-linked intermediate (Figure 3–2(b)). Cross-link formation is initiated by rotation of U from the stacked reactant structure (R_D, χ_G = 299.2° and χ_U = 274.9°) to a T-shaped intermediate ($I1_D$, χ_G = 239.1° and χ_U = 178.1°), which has a corresponding barrier of 17.8 kJ mol^{-1} ($\Delta G^{\ddagger}$ = 6.6 kJ mol^{-1}, Table 3–2). Subsequent cross-link formation occurs with a barrier of 37.0 kJ mol^{-1} ($\Delta G^{\ddagger}$ = 23.6 kJ mol^{-1}). This step is accompanied by a decrease in d(C8–C5) from 3.706 Å in $I1_D$ to 2.326 Å in $TS2_D$ and 1.517 Å in $I2_D$, as well as a substantial change in the G orientation (χ_G = 290.1°), C8 hybridization to sp^3 and a small change in U orientation (χ_U = 193.1°). In $I2_D$, the unpaired

electron is mainly delocalized on G (N7 = 0.545 e$^-$, C4 = 0.118 e$^-$ and O6 = 0.116 e$^-$). While U

rotation results in an increase in energy (ΔE = 13.0 kJ mol^{-1}), cross-link formation releases

energy, as I2$_D$ lies –107.0 kJ mol^{-1} below the reactant.
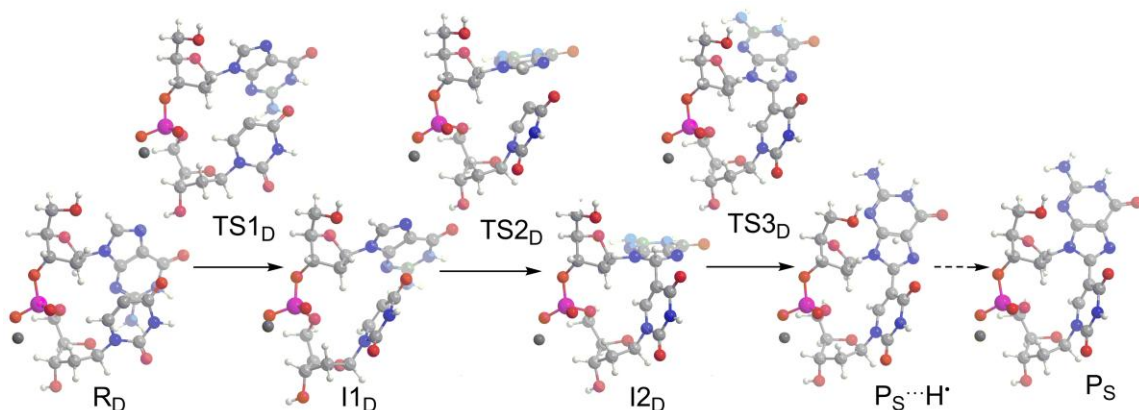
Figure 3–5: Stationary points for formation of 5′–G[8–5]U–3′ from 5′–GU•–3′.

Table 3–2: Relative energies and thermodynamic parameters (kJ mol^{-1}) for the formation of 5′–G[8–5]U–3′ from 5′–GU•–3′.[a]

Stationary Point[b]	ΔE[c,d]	ΔE$_{ZPVC}$[d,e]	ΔH[d,f]	ΔG[g]	TΔS[h]
R$_D$	0.0	0.0	0.0	0.0	0.0
TS1$_D$	17.8	13.3	13.2	6.6	6.6
I1$_D$	13.0	9.7	10.9	0.8	10.1
TS2$_D$	50.0	42.0	42.8	24.4	18.4
I2$_D$	–107.0	–109.1	–108.9	–136.7	27.8
TS3$_D$	64.8	44.0	43.8	28.6	15.5
P$_S$···H•	34.3	15.5	16.5	3.6	12.9
P$_S$ + H•	43.7	19.1	23.3	–14.5	37.8

[a]Reactant generated from exposure of 5′–G^{X}U–3′ to ionizing radiation. [b]Refer to Figure 3–5 for the structure of the corresponding stationary points. [c]Optimization energy. [d]Calculated with IEFPCM-M06-2X/6-31G(d,p). [e]Includes scaled (0.9580) zero-point vibrational energy correction. [f]Includes scaled (0.9480) thermal correction to the enthalpy. [g]Calculated using the SMD-M06-2X/6-31G(d,p)) energy (including non-electrostatic component) and IEF-PCM-M06-2X/6-31G(d,p) thermal correction to the Gibbs Free Energy. [h]Calculated as TΔS = ΔH−ΔG.

Following cross-link formation, a hydrogen atom is removed (Figure 3–2(b)) to

yield the same ground state singlet product (P$_S$) as the UV pathways. This transition has a

large barrier ($\Delta E^{\ddagger}$ = 171.8 kJ mol^{-1}, $\Delta G^{\ddagger}$ = 165.3 kJ mol^{-1}), which is associated with an increase in the C8–H8 distance from 1.099 Å (I2$_D$) to 1.837 Å (TS3$_D$) and a localization of the spin density on H8 (0.801 e$^-$). It is anticipated that a mediating molecule from the surroundings or within DNA will decrease this barrier. While attempts were made to include a mediating molecule (hydroxyl radical, water), stationary points could not be obtained. Regardless, the magnitude of the barrier for hydrogen-atom removal compared to the small barriers for the U rotation and bond formation steps suggest that hydrogen removal will be difficult.

The product complex (P$_S$⋯H•, Figure 3–5) connected to TS3 involves a hydrogen atom (0.964 e$^-$) bridged between the terminal O5′ of the backbone (d(O5′⋯H•) = 2.681 Å) and O4 of U (d(O4 ⋯H•) = 2.859 Å). A final decomplexation step (Figures 3–4(b) and 3–5) connects this complex (P$_S$⋯H•) to the infinitely separated (P$_S$ + H•) products. The overall reaction is endothermic (ΔH_{Rxn} = 23.3 kJ mol^{-1}) and slightly exergonic (ΔG_{Rxn} = –14.5 kJ mol^{-1}).

As mentioned in Chapter 1, 5′–G[8–5]U–3′ has not yet been experimentally detected following ionizing irradiation of ^{Br}U-containing DNA. However, the results presented in this chapter show that its formation is slightly thermodynamically favoured (ΔG_{Rxn} = –14.5 kJ mol^{-1}) under these conditions. In contrast, the (ionizing) radiation-induced formation of an experimentally-observed[98,105,209] intrastrand cross-link between C8 of a 5′–G and the C5–methyl radical in a 3′–T (5′–G[8–5-Me]T–3′) was calculated (B3LYP/6-31G(d,p)) to be highly unfavourable (ΔG_{Rxn} ≈ 167 kcal mol^{-1}).[115] The hydrogen atom removal step in the formation of 5′–G[8–5-Me]T–3′, modeled without a mediating molecule, was determined to have the largest barrier ($\Delta G^{\ddagger}$ ≈ 170 kJ mol^{-1}),[115] which is similar to that calculated in the present study for 5′–G[8–5]U–3′ (165.3 kJ mol^{-1}). Although the T-containing cross-link was

studied with B3LYP and the present study employs M06-2X, this comparison is expected to be valid since it is anticipated that the barrier for hydrogen-atom removal will be relatively unaffected by the choice in functional. Taken together, the facts that 5′–G[8–5]U–3′ formed from ionizing radiation is more kinetically and thermodynamically favoured than the experimentally-observed 5′–G[8-5-Me]T–3′ intrastrand cross-link suggests that 5′–G[8–5]U–3′ is a plausible product upon exposure of ^{X}U-containing DNA to ionizing radiation.

3.3.3. Comparison of Cross-link Formation From UV and Ionizing Radiation: Therapeutic Importance

There are substantial differences in the calculated reaction pathways for cross-link formation resulting from exposure of ^{X}U-containing DNA to UV or ionizing radiation (Figure 3–4). First, 5′–G[8–5]U–3′ formation from UV radiation is highly exergonic (ΔG_{Rxn} = –307.9 kJ mol^{-1}), while formation from ionizing radiation is only slightly exergonic (ΔG_{Rxn} = –14.5 kJ mol^{-1}). Second, the highest-energy step in the UV pathways involves bond formation and a small barrier ($\Delta E^{\ddagger}$ = 70–76 kJ mol^{-1}), while that for ionizing radiation involves hydrogen-atom removal and a comparatively large barrier ($\Delta E^{\ddagger}$ = 171.8 kJ mol^{-1}). This indicates 5′–G[8–5]U–3′ formation is more likely upon exposure to UV radiation. Although 5′–G[8–5]U–3′ has never been detected following ionizing radiation, this represents the therapeutically-relevant conditions. Nevertheless, as discussed in Section 1.7., indirect evidence suggests 5′–G[8–5]U–3′ will form due to ionizing radiation. Indeed, many of the same lesions (strand breaks, interstrand cross-links) can be formed by U$^{\bullet}$ generated from either UV or ionizing radiation.[49,50,57-59,61-65,74,75,77-80] Through comparisons to the experimentally-observed 5′–G[8-5-Me]T–3′, it is suggested 5′–G[8–5]U–3′ is likely to form in ^{X}U-substituted DNA following exposure to ionizing radiation. This includes DNA containing ^{Cl}U, which is currently the most promising ^{X}U in clinical applications.[53,54]

73

Therefore, future experimental work should both verify and quantify the formation of 5′–G[8–5]U–3′ in ˣU-substituted cellular DNA upon exposure to ionizing radiation.

3.3.4. The Cross-link Product: Structure and Biological Implications

While 5′–G[8-5]U–3′ has been experimentally identified as a product of UV radiation in a dinucleoside monophosphate,[66] synthetic oligonucleotides[66,68] and cellular human DNA,[67] its exact structure and orientation in DNA, as well as potential distortions inflicted upon the DNA helix, have not been determined. Given the evidence provided in this thesis that 5′–G[8–5]U–3′ is likely to form during the therapeutic use of ˣU, it is especially important to understand the structure of 5′–G[8–5]U–3′ and its effect on the DNA helix, which will help reveal the biological implications of this product including its potential to contribute to cell death during anti-cancer treatments. Comparison of the optimized geometries of 5′–G[8–5]U–3′ and the corresponding natural 5′–GT–3′ dinucleoside monophosphate obtained in Chapter 2 (Figure 3–6(a) and Table B–2) reveals only small changes in the backbone torsion angles and deoxyribose sugar puckering upon cross-link formation. The most substantial structural change to natural DNA is the orientation of the nucleobases. First, although U remains in the native *anti* orientation, G adopts a *syn* orientation, which has been similarly predicted for other purine–pyrimidine intrastrand cross-links with similar connectivity.[104,106,115,118] Second, the cross-link adopts a perpendicular orientation relative to the native stacked base arrangement, which suggests the local DNA environment will be significantly disrupted.
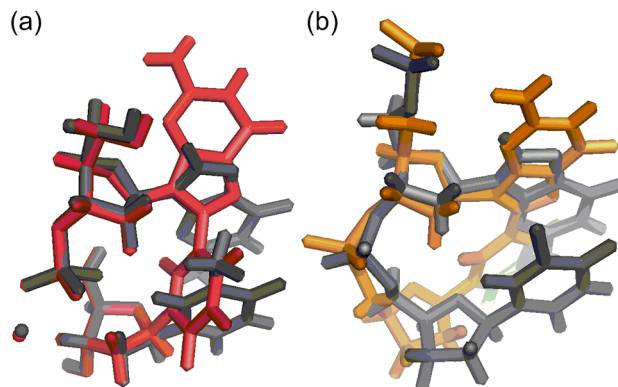
Figure 3–6: Overlay of (a) the natural 5′–GT–3′ sequence (grey) and 5′–G[8–5]U–3′ (red) obtained using the DFT dinucleoside monophosphate model, and (b) the natural (grey) and 5′–G[8–5]U–3′ (orange) dinucleotides obtained from the MD dsDNA model.

Molecular dynamics simulations were performed on dsDNA with the 5′–d(GCATGGCG[8–5]UGCTATGC)–3′ sequence to establish the cross-link structure within the double helix. The cross-link adopts the same conformation throughout the simulation (3 – 55 ns), where the DNA backbone becomes highly distorted and unwinds when the cross-link is present (Figure 3–6(b) and Table B–2). Although the small model predicts the cross-link to adopt a perpendicular base arrangement that would disrupt intrastrand base–base interactions and potentially cause significant distortions to neighbouring base pairs, changes to the backbone allow the cross-link to fall in the same plane as G in natural DNA (Figure 3–6). Therefore, the cross-link does not have a large impact on the flanking base pairs, which is supported by similar hydrogen-bonding occupancies in cross-linked (78–99 %) and natural (83–99 %) DNA. The twist between the G and U bases (θ, Table B–2) also decreases and occurs in the opposite direction when the cross-link is in DNA compared to a dinucleoside monophosphate. Regardless of the model considered, the Watson-Crick face of G is unavailable for hydrogen bonding with the opposing strand, which causes a loss of genetic information at the cross-linked G of 5′–G[8–5]U–3′ upon replication.

The conformation of the cross-link and flanking base pairs are unchanged throughout the course of the simulation, while the orientations of the nucleotides complementary to the cross-link change to give different interstrand base interactions. Specifically, three main conformations of cross-linked DNA (designated Conformers 1–3, see Appendix B, page B–5) were identified (Figure 3–7), each of which is stable for at least 8 consecutive nanoseconds over the course of the 55 ns simulation. Conformer 1 is characterized by hydrogen–bonding interactions between O4 in the Watson-Crick face of the cross-linked U and the amino groups in the Watson-Crick faces of the opposing C and A bases, which occur with occupancies of 89% and 34%, respectively. This creates an opening 5′ to the cross-link (Figure 3–7(a)) and local unwinding of the complementary strand. The helix width (interstrand C1′⋯C1′ distance between complementary bases) at the cross-link remains relatively unchanged at the G nucleotide (10.8 ± 0.6 Å) and decreases at the U nucleotide (9.8 ± 0.8 Å) compared to natural DNA (10.731 ± 0.189 Å and 10.614 ± 0.277 Å, respectively), and increases in flexibility.

In Conformer 2 (Figure 3–7(b)), both complementary bases form π–π interactions with the cross-link. Specifically, C adopts an *anti* orientation about the glycosidic bond (χ = 196.6 ± 8.8°) but directs the C5 and C6 atoms towards the cross-link-containing strand. This is accompanied by large distortions in the backbone of the C nucleotide as C intercalates between the cross-link and the 5′–flanking C⋯G base pair. A remains in the *anti* orientation and intercalates between the cross-link and the 3′–flanking G⋯C base pair. This creates a zipper-like conformation, which is known to naturally occur for mismatched DNA.[91] The helix narrows at both the G (8.622 ± 0.301 Å) and U (6.776 ± 0.333 Å) nucleotides, and there are significant distortions to the complementary A and C nucleotide backbones. Finally, in Conformer 3 (Figure 3–7(c)), the complementary A is intercalated 3′ of the cross-link in a
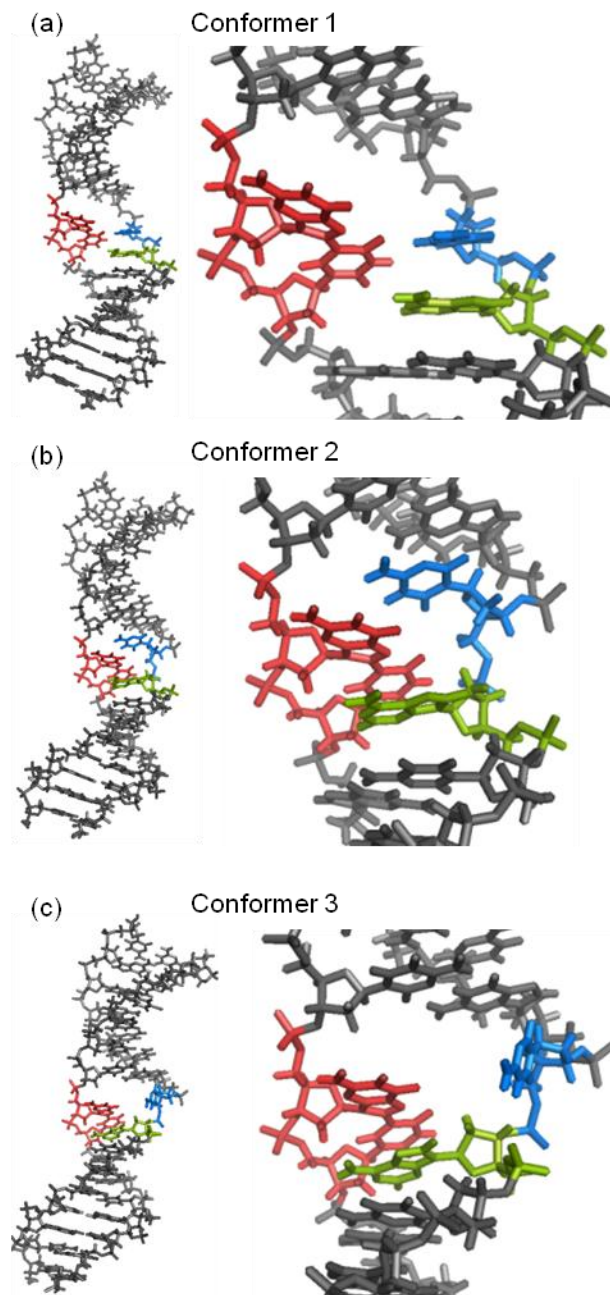
Figure 3–7: Conformers 1–3 (a–c, respectively) of 5'–G[8–5]U–3'-containing DNA obtained from molecular dynamics simulations, highlighting the cross-linked dinucleotide (red) and complementary C (blue) and A (green) nucleotides.

π–π interaction, while the C base is flipped out of the helix and displays significant flexibility in the extrahelical position (for example, see Figures 3–7(c) and B–5 (45ns)). C primarily

adopts the *anti* orientation (χ = 229.5 ± 28.1°) and the N4 amino group hydrogen bonds with the phosphate oxygen in the backbone of the A nucleotide (62.5% occupancy). However, more disordered states are also observed due to the flexibility in the backbone torsion angles. The interstrand distance remains compact at the cross-linked U nucleotide (7.927 ± 0.544 Å) whereas the extrahelical orientation of C causes the helix to widen substantially at the cross-linked G nucleotide (11.954 ± 0.841 Å).

Distortions to the helix caused by 5′–G[8–5]U–3′ might have implications in DNA repair.[108] For example, extrahelical bases, such as C in Conformer 3, can aid recognition of DNA mismatches.[210] Although distortions may enhance the excision of 5′–G[8–5]U–3′ *in vivo*, cell-cycle checkpoints and DNA repair are generally compromised in tumours and lesions often persist.[54] Therefore, the conformations of cross-linked DNA are likely to have a greater effect on DNA replication, a hyperactive process in rapidly-dividing tumour cells.[211] Indeed, related G–pyrimidine cross-links with [8–5] or [8–5-Me] connectivity were found to stall DNA polymerases,[105-107,110] and 5′–G[8–5]U–3′ may behave similarly. As discussed in Section 1.6.1., some DNA polymerases stall before the cross-link,[105,110] while others insert the correct nucleotide opposite the cross-linked pyrimidine and then stall at the cross-linked G due to the lack of an appropriate hydrogen-bonding partner.[105-107] A DSB may result if the stalled polymerase is not resolved.[203] This supports the observation that the cytotoxicity of [X]U following irradiation cannot be attributed to the direct formation of strand breaks from U•, but may instead involve cellular processing of the initial damage leading to a strand-break.[49,78] When translesion bypass synthesis is successful, steric considerations cause a purine nucleotide to be inserted opposite the cross-linked G,[104,106,110,112] which is mutagenic and may thereby hamper the functioning of tumour cells. Additionally, translesion synthesis is slower than high-fidelity polymerase replication, which may slow tumour growth.[111] The above discussion suggests 5′–G[8–5]U–3′ is both

mutagenic and cytotoxic. This indicates that DNA damage caused by U•, including intrastrand cross-links, may have broad biological implications through a variety of cellular pathways.

In addition to distortion localized to the region of the cross-link, a bend occurs towards the minor groove in all conformations (Figure 3–8 and Table B–3). This is significant since bending of DNA plays a role in regulating replication and transcription, and
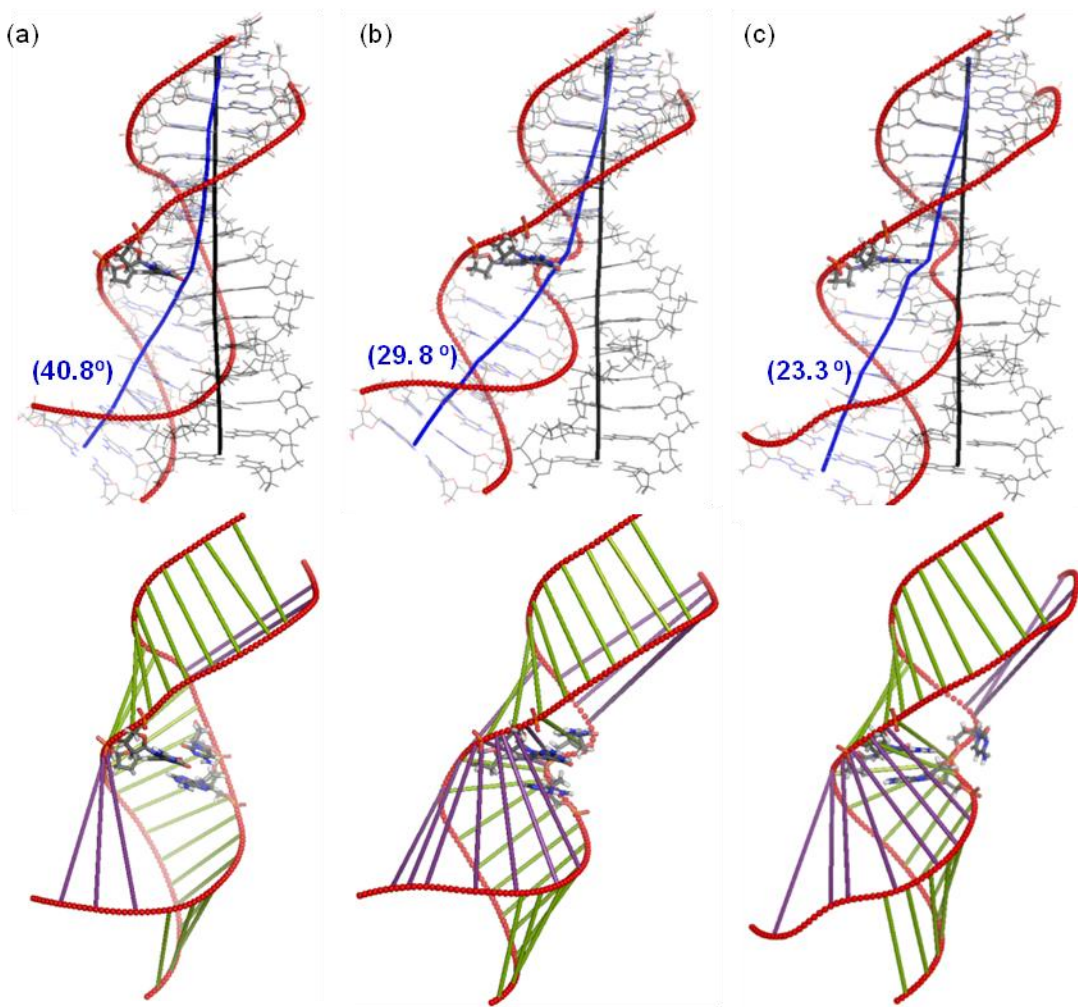
Figure 3–8: Illustrations of the global changes to the DNA helix when cross-linked DNA adopts Conformers 1–3 (a–c, respectively), where the helix axis (top, blue), backbone (red), as well as major (purple) and minor (green) grooves (bottom), are identified. The axis bend is provided in brackets and is compared to a value of 9.5° for natural DNA (black).

therefore alterations to this property are commonly observed in the use of many efficient anti-tumour therapies.[40] For example, cisplatin forms an intrastrand cross-link between adjacent G nucleotides that causes a ~35° bend towards the major groove.[33,40] Similarly, anticancer drugs like nitrogen mustard mechlorethamine form interstrand cross-links that cause up to a 17° bend in DNA,[212] while the intrastrand cross-links induced by mitomycin C cause a 15° bend.[213] The bend in 5′–G[8–5]U–3′-containing DNA has a substantial effect on the grooves of the helix (Figure 3–8 and Table B–3). At the cross-linked region, the minor groove broadens or even disappears (Conformer 2), while the major groove disappears in all conformations since the G moiety of the cross-link protrudes into this region. Disappearance of a groove is known to occur in highly deformed structures,[206] and may significantly affect the cell cycle since many DNA–protein interactions are initiated at the grooves.[7] Overall, alterations to both the bending and the grooves of natural DNA when the cross-link is present parallel conformational changes induced by binding of cisplatin,[33,40] and are therefore indicative of the potential of 5′–G[8–5]U–3′ to contribute to cell death through similar pathways.

3.4.   Conclusions

The formation pathways for 5′–G[8–5]U–3′ following exposure of ^{X}U-containing DNA to UV or ionizing radiation has been examined. Following UV irradiation and creation of the 5′–G$^{•+}$U$^{•}$–3′ reactant, the reaction may proceed through bond formation on the triplet surface or bond formation via an intersystem crossing (ISC), and is accompanied by a release of energy (ΔG_{Rxn} = −307.9 kJ mol^{-1}). While similarities in the barriers of the bond formation step ($\Delta E^{\ddagger}$ = 70–76 kJ mol^{-1}) preclude computational identification of the preferred pathway, bond formation on the triplet surface may be identified through the detection of phosphorescence. Alternatively, when the cross-link is formed from the 5′–GU$^{•}$–3′ reactant generated using ionizing radiation (or UV homolysis of ^{I}U), product

formation is less likely. Specifically, hydrogen-atom removal has the largest barrier ($\Delta E^{\ddagger}$ = 172 mol^{-1}) and the reaction is slightly exergonic (ΔG_{Rxn} = −14.5 kJ mol^{-1}). However, 5′–G[8–5]U–3′ formation upon exposure of ^{X}U-containing DNA to ionizing radiation is kinetically and thermodynamically more favoured than formation of related purine–pyrimidine cross-links that have been experimentally observed following exposure of natural DNA to ionizing radiation. It is therefore likely that 5′–G[8–5]U–3′ forms under therapeutically-relevant conditions.

The structure of the 5′–G[8–5]U–3′ intrastrand cross-link has been presented here for the first time. Regardless of the model used, the cross-linked G consistently adopts the *syn* orientation. This has important implications for DNA replication, where similar cross-links have been shown to stall polymerases or cause error-prone translesion synthesis. Distortions to the DNA helix when the cross-link is present are largely limited to the cross-linked dinucleotide and the opposing C and A nucleotides, with negligible effects on the flanking base pairs. Local distortions include unwinding of the cross-link backbone to accommodate the cross-linked nucleobases in the helix. Additionally, the nucleotides complementary to the cross-link adopt non-native configurations in response to the damage, which results in three main conformations of cross-linked DNA. Global changes to the DNA helix upon cross-link formation include a bend in the helix axis and alterations to both the major and minor grooves. These distortions parallel those induced by efficient anti-tumour therapies, such as cisplatin, which indicates 5′–G[8–5]U–3′ may contribute to the cytotoxicity of ^{X}U using similar mechanisms.

The following chapters will consider the formation and structure of the 5′–U[5–8]G–3′, 5′–A[8–5]U–3′, and 5′–A[2–5]U–3′ intrastrand cross-links. The primary goal is to determine whether similarities exist in the formation mechanism and distortion to the

DNA helices for these lesions and 5′–G[8–5]U–3′. Most importantly, attempts will be made to use calculations to establish the origin of the sequence selectivity of intrastrand cross-link formation and determine which cross-links are most likely to form under therapeutically-relevant conditions. Knowledge of the structures of these lesions in DNA will also provide insight into the biological implications of intrastrand cross-links and indicate how they may contribute to the cytotoxicity of ^{x}U in anti-tumour therapies.

Chapter 4. The Formation and Structure of 5′–U[5–8]G–3′

4.1. Introduction

Of the four intrastrand cross-links that form following UV irradiation of BrU-containing cellular DNA (Section 1.5.3.),[66,67] the formation pathway and structure of the most abundant cross-link (5′–G[8–5]U–3′) were determined in Chapter 3. The present chapter focuses on the other G-containing cross-link, 5′–U[5–8]G–3′ (Figure 1–8(b)), which is an isomer of 5′–G[8–5]U–3′ that differs in the nucleobase sequence. It is important to study this isomer for several reasons. Although 5′–U[5–8]G–3′ forms in lower quantities in cellular DNA with UV irradiation, yields do not necessarily dictate cytotoxicity. Additionally, isomeric sequences are likely to exhibit unique behaviours in terms of structure and formation mechanism. Furthermore, despite previous proposals, the magnitude of stacking interactions between nucleobases does not explain the sequence-selective formation of U-containing intrastrand cross-links (Section 1.5.3.).[66-68] Therefore, the present work will develop a new proposal to explain the formation preferences of these lesions and gain insight into the biological implications of U-containing intrastrand cross-links. Initially, the formation of 5′–U[5–8]G–3′ is studied from reactants relevant to UV (5′–U•G•+–3′) and ionizing (5′–U•G–3′) radiation, which will determine whether this cross-link, like 5′–G[8–5]U–3′, is likely to form under therapeutically-relevant conditions. Next, comparisons will be made between the calculated formation pathways of 5′–G[8–5]U–3′ and 5′–U[5–8]G–3′ to shed light on the observed sequence selectivity under experimental (UV light) conditions.[66,67] Finally, the structure of 5′–U[5–8]G–3′ within DNA and distortions caused to the helix will be presented and compared to those discussed for 5′–G[8–5]U–3′ in Chapter 3, as well as successful anti-tumour therapies.

4.2.   Computational Details

The formation of 5′–U[5–8]G–3′ is studied using the computational approach recommended in Chapter 2 and the same procedures described in detail in Section 3.2.. Molecular dynamics simulations were performed to determine the structure of this lesion within dsDNA with the 5′–d(GCATGGCG**TG**CTATGC)–3′ sequence. Cross-linked DNA was generated by overlaying the backbone of the M06-2X-optimized dinucleoside monophosphate product onto the central 5′–TG–3′ sequence (bold). The corresponding natural strand was also simulated for comparison. Details of the simulations are included in Appendix C, which are similar to that described for 5′–G[8–5]U–3′ in Appendix B.

4.3.   Results and Discussion

4.3.1.   Formation Mechanisms

4.3.1.1. Formation of 5′–U[5–8]G–3′ by UV Radiation

Initially, the 5′–U$^\bullet$ G$^{\bullet+}$–3′ reactant formed from UV radiation was optimized with both nucleobases in the *anti* orientation (χ_U = 214.4° and χ_G = 283.5°,[vii] R$_{T(U\text{-}anti\ G\text{-}anti)}$). However, cross-link formation cannot proceed directly from this reactant. Instead, both nucleobases must adopt the *syn* orientation about the glycosidic bond (χ_U = 67.9° and χ_G = 70.6°), where R$_{T(U\text{-}syn\ G\text{-}syn)}$ lies 11.5 kJ mol^{-1} (ΔG = 12.2 kJ mol^{-1}, Table 4–1) above R$_{T(U\text{-}anti\ G\text{-}anti)}$. Although *anti*-to-*syn* rotations are feasible in a nucleoside,[214] steric clashes may lead to higher energetic costs and thereby hinder this transition in dsDNA. However, experimental observation of 5′–U[5–8]G–3′ following UV exposure of ^{X}U in cellular DNA indicates such a large conformational change is possible in the helix. While initial attempts to model the conversion from R$_{T(U\text{-}anti\ G\text{-}anti)}$ to R$_{T(U\text{-}syn\ G\text{-}syn)}$ were unsuccessful, this was not investigated further since the exact effects of this important conformational change are likely not captured by the computational model used in the present work. Although these

steps likely play a role in the overall ease of formation of 5′–U[5–8]G–3′, the remainder of the formation pathway, including additional rearrangement and chemical steps, can still be examined with a dinucleoside monophosphate model.

In $R_{T(U\text{-}syn\ G\text{-}syn)}$, the nucleobases adopt a stacked arrangement, while the deoxyribose moieties adopt C1′–*exo* (P_U = 138.9°) and C2′–*endo* (P_G = 149.8°) puckering for U and G nucleotides, respectively. Mulliken spin densities show an unpaired electron localized on C5 of U (1.061 e⁻) and another delocalized on G (C5 = 0.317 e⁻, C8 = 0.258 e⁻, N3 = 0.247 e⁻, N2 = 0.137 e⁻ and O6 = 0.120 e⁻). The triplet cross-linked intermediate (I1$_T$) forms directly from $R_{T(U\text{-}syn\ G\text{-}syn)}$ through a transition state (TS1$_T$) with a barrier of 48.0 kJ mol⁻¹ (Figure 4–2, $\Delta G^{\ddagger}$ = 45.1 kJ mol⁻¹). I1$_T$ is characterized by negligible twist between the nucleobases (θ = 178.8°) and delocalization of the two unpaired electrons over both bases. Additionally, I1$_T$ contains the U base in the *syn* (χ_U = 8.9°) orientation, while G adopts an *anti* (χ_G = 162.6°) orientation. The formation of I1$_T$ is spontaneous (ΔG = –12.1 kJ mol⁻¹).

Following association on the triplet surface, a transition may occur from I1$_T$ to the singlet cross-linked intermediate (I1$_S$) through phosphorescence (Figure 4–1 and 4–2(a)). I1$_S$ is characterized by χ_U = 7.1°, χ_G = 158.9° and θ = 179.8°, as well as a similar backbone conformation as I1$_T$. The transition between these minima involves a decrease in energy (230.8 kJ mol⁻¹), the majority of which is released as a photon (175.5 kJ mol⁻¹, 681.5 nm) in the visible region of light. As discussed for formation of 5′–G[8–5]U–3′, bond formation may also proceed via an ISC between $R_{T(U\text{-}syn\ G\text{-}syn)}$ and I1$_S$. Using the procedure outlined in section 3.3.1.2., an ISC was estimated at d(C5–C8) = 2.04 Å with an energy of 47.6 kJ mol⁻¹ relative to $R_{T(U\text{-}syn\ G\text{-}syn)}$ (Figure 4–2(a), Figure C–1).
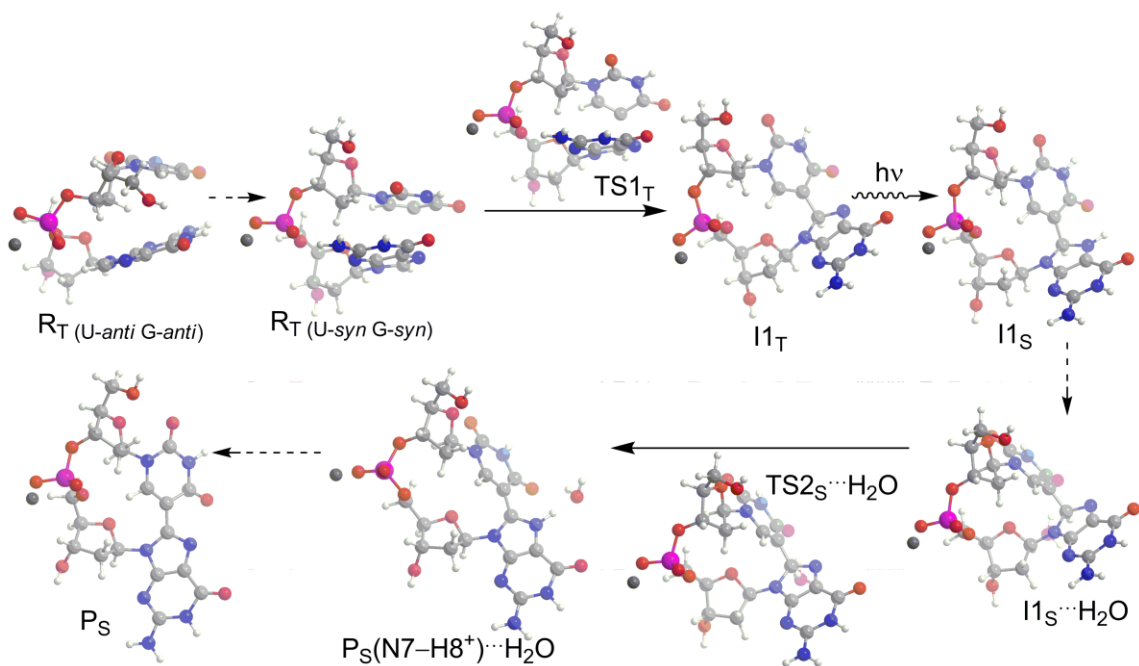
Figure 4–1: Stationary points for formation of 5′–U[5–8]G–3′ from 5′–U•G•+–3′.
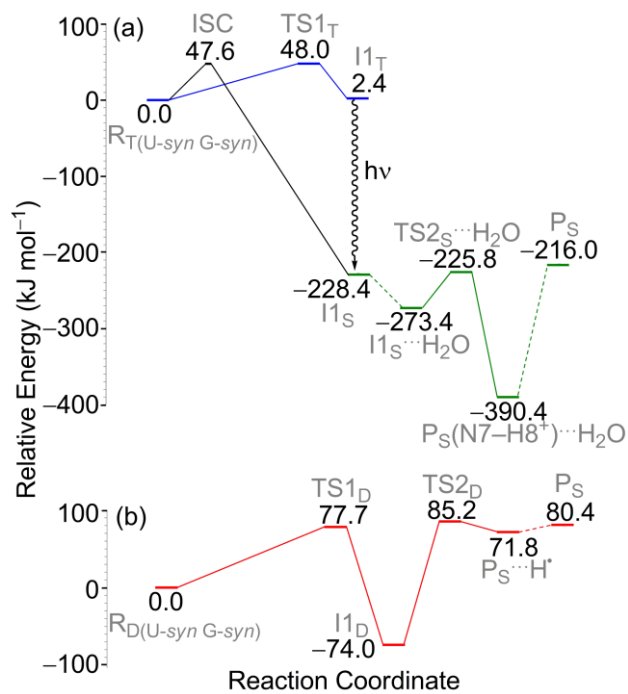
Figure 4–2: Cross-link formation resulting from (a) 5′–U•G•+–3′ on the triplet (blue) and singlet (green) surfaces or (b) 5′–UG•–3′ on the doublet surface (red). Energies (kJ mol⁻¹) are reported relative to the corresponding reactant.

Table 4–1: Relative energies and thermodynamic parameters (kJ mol^{-1}) for the formation of 5′–U[5–8]G–3′ from 5′–U•G•+–3′.[a]

Stationary Point[b]	ΔE_{Opt}[c,d]	ΔE_{ZPVC}[d,e]	ΔH[d,f]	ΔG[g]	$T\Delta S$[h]
R$_{T(U\text{-}anti\ G\text{-}anti)}$ + H$_2$O	−11.5	−9.1	−9.4	−12.2	2.8
R$_{T(U\text{-}syn\ G\text{-}syn)}$ + H$_2$O	0.0	0.0	0.0	0.0	0.0
TS1$_T$ + H$_2$O	48.0	46.3	44.1	45.1	−1.0
I1$_T$ + H$_2$O	2.4	1.1	−1.1	−12.1	11.0
I1$_S$ + H$_2$O	−228.4	−219.4	−221.0	−233.1	12.1
I1$_S$···H$_2$O	−273.4	−256.8	−259.6	−233.3	−26.3
TS2$_S$···H$_2$O	−225.8	−223.2	−228.1	−197.3	−30.8
P$_S$ (N7–H8$^+$)···H$_2$O[i]	−390.4	−373.2	−376.4	−328.6	−47.8
P$_S$ + H$_3$O$^+$	−216.0	−205.1	−206.7	−262.0	55.3

[a]Reactant generated from exposure of 5′–^{x}UG–3′ to UV radiation. [b]Refer to Figure 4–1 for the structures of the corresponding stationary points. [c]Optimization energy. [d]Calculated with IEF-PCM-M06-2X/6-31G(d,p). [e]Includes scaled (0.9580) zero-point vibrational energy correction. [f]Includes scaled (0.9480) thermal correction to the enthalpy. [g]Calculated using SMD-M06-2X/6-31G(d,p) energy (including non-electrostatic component) and IEF-PCM-M06-2X/6-31G(d,p) thermal correction to the Gibbs Free Energy. [h]Calculated as TΔS = ΔH–ΔG. [i]Cross-linked product protonated at N7 and hydrogen bonding with H$_2$O.

Regardless of the mechanism of bond formation, a proton is subsequently removed from C8 of G in I1$_S$. Water-mediated proton removal from I1$_S$···H$_2$O (d(C8–H8) = 1.099 Å) proceeds through TS2$_S$···H$_2$O (d(C8–H8)= 1.356 Å) with a barrier of 47.6 kJ mol^{-1} ($\Delta G^{\ddagger}$ = 36.0 kJ mol^{-1}). However, this step is coupled with water-mediated proton transfer to N7 of G to yield a P$_S$(N7–H8$^+$)···H$_2$O complex involving water bridged between N7 of G and O4 of U (Figure 4–1), and is accompanied by a release of energy ($\Delta(\Delta G)$ = −95.3 kJ mol^{-1}). Protonation of G is an artefact of using an isolated model since surrounding water networks will likely efficiently transfer the proton away from N7 in a DNA environment. In the final reaction step, the infinitely separated species (P$_S$ + H$_3$O$^+$, Figure 4–1) are obtained. The formation of 5′–U[5–8]G–3′ from 5′–U•G•+–3′ involves a large decrease in energy (ΔE_{Rxn} = 216.0 kJ mol^{-1}), the majority of which is lost upon transition from the triplet to the singlet surface as a photon (phosphorescence) or vibrationally dissipated in the form of heat (ISC).

The reaction is both exothermic (ΔH_{Rxn} = −206.7 kJ mol^{-1}) and exergonic (ΔG_{Rxn} = −262.0 kJ mol^{-1}), and involves spontaneous associative and dissociative steps (Table 4−1).

The barriers for both steps of the reaction are similar (46.7 − 48 kJ mol^{-1}), which makes it difficult to determine the relative ease of the chemical steps. However, within dsDNA, surrounding π–π and hydrogen-bonding interactions with nucleobases, as well as the energetic cost associated with backbone distortions, will hinder bond formation, while surrounding molecules (water, nucleobases) will assist H$^+$ removal. Therefore, the bond formation step is likely more difficult than H$^+$ removal in DNA. Since the barriers for bond formation via the triplet surface or coupled with an ISC are also very similar, it is difficult to computationally distinguish the preferred mechanism and both may occur under experimental conditions. Although the experimental detection of phosphorescence will verify that bond formation occurs along the lowest-lying triplet surface, this will not rule out the possibility that bond formation also occurs via an ISC. Additionally, if bond formation does occur on the triplet surface, the unique phosphorescence of 5′–G[8–5]U–3′ (668.7 nm) and 5′–U[5–8]G–3′ (681.5 nm) may provide an alternative, and more direct, means of detecting, quantifying or discriminating between the two cross-links. Both signals should also be unique from the fluorescence and phosphorescence of natural DNA even though their detection may be complicated by low cross-link yields and the low temperatures required for a measurable intensity.[66,67,207,208]

4.3.1.2. Formation of 5′–U[5–8]G–3′ by Ionizing Radiation

As found for the UV pathway, formation of 5′–U[5–8]G–3′ from ionizing radiation cannot proceed from 5′–U•G–3′ when both nucleobases adopt the *anti* (χ_U = 217.5° and χ_G = 280.9°[vii]) orientation ($R_{D(U\text{-}anti\ G\text{-}anti)}$, Figure 4–3). Rather, the reaction must proceed with both nucleobases in the *syn* (χ_U = 73.5° and χ_G = 69.1°) orientation ($R_{D(U\text{-}syn,\ G\text{-}syn)}$, Figure 4–3),

which lies 22.1 kJ mol^{-1} (ΔG = 17.0 kJ mol^{-1}, Table 4–2) above R$_{D(U\text{-}anti\ G\text{-}anti)}$. In R$_{D(U\text{-}syn,\ G\text{-}syn)}$, the sugar moieties adopt the C2'–*endo* puckering (P$_U$ = 147.4°, P$_G$ = 150.1°), and the Mulliken atomic spin density shows the unpaired electron localized on C5 of U (1.049 e$^-$).
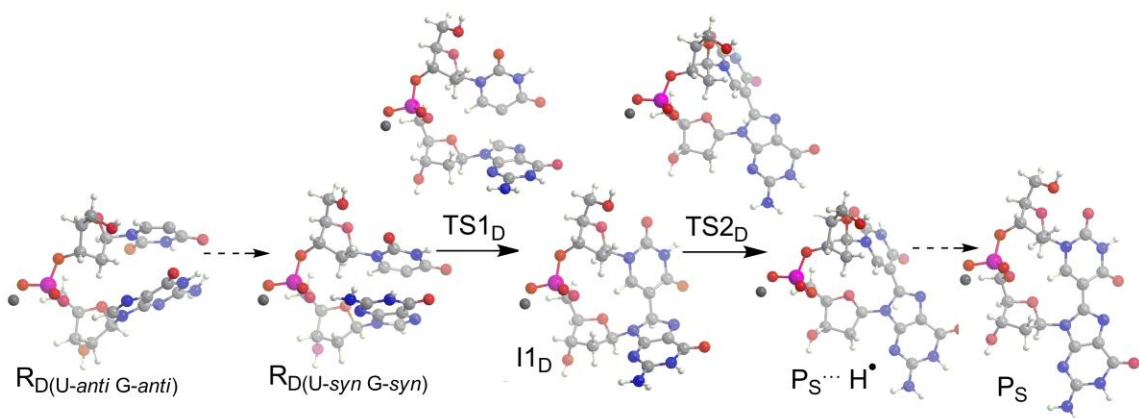
Figure: 4–3: Stationary points for formation of 5'–U[5–8]G–3' from 5'–U•G–3'.

Table 4–2: Relative energies and thermodynamic parameters (kJ mol^{-1}) for the formation of 5'–U[5–8]G–3' from 5'–U•G–3'.[a]

Stationary Point[b]	ΔE$_{Opt}$[c,d]	ΔE$_{ZPVC}$[d,e]	ΔH[d,f]	ΔG[g]	TΔS[h]
R$_{D(U\text{-}anti\ G\text{-}anti)}$	−22.1	−20.4	−20.8	-17.0	-3.8
R$_{D(U\text{-}syn\ G\text{-}syn)}$	0.0	0.0	0.0	0.0	0.0
TS1$_D$	77.7	71.0	71.5	57.9	13.5
I1$_D$	−74.0	−73.0	−73.5	-86.2	12.7
TS2$_D$	85.2	66.7	66.3	59.5	6.8
P$_S$···H•	71.8	53.5	55.4	40.9	14.5
P$_S$ + H•	80.4	56.8	61.0	31.1	29.9

[a]Reactant generated from exposure of 5'–^{X}UG–3' to ionizing radiation. [b]Refer to Figure 4–3 for the structures of the corresponding stationary points. [c]Optimization energy. [d]Calculated with IEF-PCM-M06-2X/6-31G(d,p). [e]Includes scaled (0.9580) zero-point vibrational energy correction. [f]Includes scaled (0.9480) thermal correction to the enthalpy. [g]Calculated using SMD-M06-2X/6-31G(d,p) energy (including non-electrostatic component) and IEF-PCM-M06-2X/6-31G(d,p) thermal correction to the Gibbs Free Energy. [h]Calculated as TΔS = ΔH–ΔG.

As observed for cross-link formation from UV radiation, the bond formation step occurs directly from $R_{D(U\text{-}syn\ G\text{-}syn)}$. The C5–C8 bond forms through a transition state (TS1$_D$) with a 77.7 kJ mol^{-1} ($\Delta G^{\ddagger}$ = 57.9 kJ mol^{-1}) barrier, and yields a cross-linked intermediate (I1$_D$) that lies 74.0 kJ mol^{-1} (ΔG = –86.2 kJ mol^{-1}) below the reactant. I1$_D$ is characterized by U in the *syn* orientation (χ_U = 6.3°), G in the *anti* orientation (χ_G = 164.7°), negligible twist between the bases (θ = 176.8°), d(C8–C5) = 1.515 Å, and an unpaired electron delocalized over the G moiety.

From I1$_D$, a hydrogen atom is removed as d(C8–H8) = 1.102 Å increases to 1.883 Å in the transition state (TS2$_D$) with a barrier of 159.2 kJ mol^{-1} ($\Delta G^{\ddagger}$ = 145.7 kJ mol^{-1}), and a localization of the unpaired electron on H8 (0.823 e$^-$). This substantial barrier will likely decrease in DNA since surrounding molecules may facilitate H$^{\bullet}$ removal. In the $P_S \cdots H^{\bullet}$ complex (Figure 4–3), the hydrogen atom (0.969 e$^-$) is bridged between N7 and O6 while interacting with the G π-system. The final reaction step (Figures 4–2(b) and 4–3) generates the infinitely separated P_S and H$^{\bullet}$, which is 80.4 kJ mol^{-1} (ΔG_{Rxn} = 31.1 kJ mol^{-1}) endothermic (endergonic).

Although 5′–U[5–8]G–3′ has not been experimentally reported following exposure of ^{X}U-containing DNA to ionizing radiation, the results of the calculated formation pathway can indicate whether the lesion is likely to form under these conditions. To verify that the substantial barrier for hydrogen atom removal in the pathway for 5′–U[5–8]G–3′ formation is plausible, comparisons were made to the calculated formation pathway[115] of an experimentally-observed[98,209] intrastrand cross-link between C5-methyl of 5′–T and C8 of 3′–G (5′–T[5-Me–8]G–3′) generated by exposure of natural DNA. Overall, the hydrogen-atom removal barrier and reaction thermodynamics are more favoured for 5′–U[5–8]G–3′ than the experimentally-observed 5′–T[5-Me–8]G–3′ cross-link (ΔG_{Rxn} = 97 kJ mol^{-1}, $\Delta G^{\ddagger}$ = 145.7

kJ mol^{-1} (hydrogen atom removal)). This indicates 5′–U[5–8]G–3′ can be formed by ionizing radiation, as well as from homolysis of C5–I in ^{I}U-containing DNA by UV radiation. Therefore, these results testify to the relevance of this cross-link to ^{X}U when used in anti-tumour therapies.

4.3.1.3. A Comparison of Cross-link Formation from UV and Ionizing Radiation

The formation of 5′–U[5–8]G–3′ from exposure of ^{X}U-substituted DNA to radiation requires both nucleobases to adopt the *syn* orientation regardless of the form of radiation used. However, the chemical steps of the reaction pathways differ considerably (Figure 4–2). Specifically, formation from UV radiation (ΔG_{Rxn} = –262.0 kJ mol^{-1}) is more thermodynamically favourable than from ionizing radiation (ΔG_{Rxn} = 31.1 kJ mol^{-1}). Additionally, the largest barrier in the UV pathways (bond formation, $\Delta E^{\ddagger}$ = ~48 kJ mol^{-1}) is relatively small compared to the chemical step with the largest barrier for ionizing radiation (hydrogen atom removal, $\Delta G^{\ddagger}$ = 145.7 kJ mol^{-1}). Therefore, 5′–U[5–8]G–3′ formation is more likely upon exposure to UV radiation. Nevertheless, the discussion in Section 4.3.1.2. suggests generation via ionizing radiation is also possible.

4.3.2. A Comparison of G-Containing Cross-links: Formation Preferences

As discussed in Section 1.5.3., the experimentally-observed formation of U$^{\bullet}$-mediated intrastrand cross-links in synthetic or cellular DNA by UV radiation measured a preference for 5′–G[8–5]U–3′ over 5′–U[5–8]G–3′.[66,67] For both cross-links, the bond formation step may take place on the triplet surface or is coupled with an ISC since the mechanisms occur with similar energetic penalties. Although the ISC and the transition state for the bond formation step occur at similar C5–C8 distances for both cross-links, the resulting energetic penalty in 5′–G[8–5]U–3′ ($\Delta E^{\ddagger}$ = 70 – 76 kJ mol^{-1}) is greater than 5′–U[5–8]G–3′ ($\Delta E^{\ddagger}$ = ~48 kJ mol^{-1}). Therefore, the dinucleoside monophosphate model

suggests the selectivity between G-containing cross-links cannot be explained by the barrier for this chemical step. However, bond formation results in a more stable cross-linked intermediate for 5′–G[8–5]U–3′ (ΔG = –303.3 kJ mol^{-1}, Table 3–1) than 5′–U[5–8]G–3′ (ΔG = –233.1 kJ mol^{-1}, Table 4–1), which may be an important factor for cross-link formation. The formation of both lesions by UV radiation is spontaneous, with 5′–G[8–5]U–3′ (ΔG_{Rxn} = –307.9 kJ mol^{-1}) more favoured than 5′–U[5–8]G–3′ (ΔG_{Rxn} = –262.0 kJ mol^{-1}). While these results suggest that cross-link formation preferences may be influenced by the stability of the cross-linked product, differences in the (product) structures predicted by a dinucleoside monophosphate model and within dsDNA (Sections 3.3.4. and 4.3.3.) may alter the thermodynamic preference calculated here. Since the barrier for the bond formation step has been eliminated as the cause of the sequence selectivity and thermodynamic preferences are unclear, other possibilities must be explored.

A previous computational study compared the formation pathways of the 5′–G[8–5-Me]T–3′ and 5′–T[5-Me–8]G–3′[115] cross-links produced following the exposure of natural DNA to ionizing radiation,[98,106,209] with the goal of understanding the sequence preference for a 5′–purine. Formation of the G[8–5-Me]T cross-links, as well as G[8–5]U, are initiated by bond formation.[115] However, while this is followed by hydrogen atom removal for 5′–G[8–5-Me]T–3′ and G[8–5]U, 5′–T[5-Me–8]G–3′ formation involves H• addition to C5 of T and subsequent H• removal from C5 of T.[115] The authors also found a competing reaction along the pathway for 5′–T[5-Me–8]G–3′ formation, where H• addition to C5 of T is followed by cleavage of the bond between C5 of T and 5-Me of T to produce U and an 8-methylated G with a radical at C8.[115] This additional pathway was proposed to partially account for the lower abundances of 5′–T[5-Me–8]G–3′ observed experimentally. However, the multiple H• transfers observed computationally are likely an artefact of an isolated, gas-phase model. Furthermore, no alternate reactions were found for the formation pathway of

5′–U[5–8]G–3′, and therefore another explanation is required for the sequence preferences for cross-links considered in this thesis.

For T-containing cross-link formation, the barriers for the H• removal or addition and thermodynamic factors indicate 5′–T[5-Me–8]G–3′ formation should be preferred[115] despite the higher yields of 5′–G[8–5-Me]T–3′ observed experimentally.[98,106,209] Therefore, in addition to the possibility of a side reaction discussed above, the sequence dependence was proposed to be influenced by the ease of H• removal from C8, which was judged based on the accessibility of the C8 site to other species.[114,115] Specifically, H• removal is more likely for 5′–purine–pyrimidine–3′ sequences since H8 occupies an external position rather than being located in a "molecular envelope" for 5′–pyrimidine–purine–3′ sequences.[114,115] Although this may be important for T-containing cross-links, calculations in the present thesis show this does not play a role in the selectivity of U-containing cross-links since the accessibility of C8 does not differ substantially in the cross-linked intermediates associated with 5′–G[8–5]U–3′ and 5′–U[5–8]G–3′. In fact, sp^3 hybridization of C8 in the U-containing cross-linked intermediates bends the nucleobases away from H8 and makes this position even more accessible for both sequences. Nevertheless, differences in the accessibility of C8 in T-containing cross-links may arise from the involvement of the C5-methyl of T, which lengthens and adds a bend to the link between the nucleobase rings. Regardless, the accessibility of C8 cannot be the factor dictating U-containing cross-link formation.

Interestingly, intrastrand cross-links between purines and pyrimidines in dsDNA exhibit a 5′–purine preference regardless of the method of generation (UV or ionizing radiation) or type (natural or modified, cellular or synthetic, dsDNA or dinucleoside monophosphate) of DNA.[66,67,98-102,104-108,209] This indicates that a more general selection rule dictates this preference. Experimental studies have noted a correlation between the

sequence preferences of purine–pyrimidine intrastrand cross-links and the proximity of the two sites forming the bond.[98-102] For example, it has been previously reported that the distance between the C8 of the purine and the C5-methyl group of T is 3.6 Å in a 5′–purine–T–3′ sequence, but 6.3 Å in the 5′–T–purine–3′ sequence.[98] In dsDNA, the preferences for GU cross-links are also supported by proximity, as evidenced by averaged fibre diffraction data[202] that find the C8 and C5 sites in the 5′–G sequence are 3.9 Å apart (4.1 – 4.3 Å in the corresponding calculated reactants), while this distance increases for 3′–G to 5.1 Å (5.8 – 6.3 Å in the corresponding calculated reactants), which is due to the natural twist of the DNA backbone. However, in the present study, the 5′–G[8–5]U–3′ formation pathways begin from reactants with nucleobases in the *anti* orientation, while the nucleobases must adopt the *syn* orientation for 5′–U[5–8]G–3′ formation. Once the bases adopt the *syn* conformation in the 5′–UG–3′ sequence, the C5–C8 distance decreases (~3.2 Å for the corresponding calculated reactants) and bond formation can occur. Therefore, proximity is an important factor in the preferences of intrastrand cross-link formation.

Although proximity can predict the formation preferences of intrastrand cross-links,[98] proximity cannot explain the sequence effect, despite being used as an argument in the literature.[98] Instead, this thesis emphasizes that proximity can only be used to predict these preferences since it is a simplistic property that plays a role in more complex factors. Specifically, proximity may be important due to the large energetic penalties associated with conformational changes prior to, or coupled with, the bond formation step when sites are far apart in dsDNA. Indeed, the large distances observed in (*anti*) pyrimidine–purine sequences require major conformational changes for cross-link formation to occur, such as the *anti*-to-*syn* transition necessary for 5′–U[5–8]G–3′. Alternatively, the distance between coupling atoms may be significant due to the high reactivity of nucleobase radicals compared to the time scale required for major conformational changes. While the

94

correlation between yield and proximity is undeniable, further insight into the exact role of proximity in intrastrand cross-link formation must be provided through studies of additional intrastrand cross-links.

4.3.3. The Structure of 5′–U[5–8]G–3′ and Helical Distortion

Although 5′–U[5–8]G–3′ forms in lower yields, it may be a more potent inducer of cell death than 5′–G[8–5]U–3′. To gain insight into the effect of the UG-product on DNA and tumour cells, its structure was studied. 5′–U[5–8]G–3′ obtained using the dinucleoside monophosphate model deviates from the corresponding natural sequence (Figure 4–4(a)). In the lesion, the 5′–sugar adopts C2′–*endo* puckering ($P_U = 153.8°$) and the 3′–sugar adopts a C1′–*exo* pucker ($P_G = 139.7°$). Differences are also observed in the backbone at ε_U and β_G (Table C–2), and 5′–U[5–8]G–3′ unwinds with respect to natural DNA. As found for 5′–G[8–5]U–3′, the cross-linked nucleobases adopt a perpendicular orientation relative to the stacked arrangement in natural DNA. However, results in Chapter 3 indicate different structures are obtained using more complete DNA models. Therefore, it is important to analyze the cross-link structure within dsDNA.

Molecular dynamics simulations were performed on complementary dsDNA containing 5′–d(GCATGGCGU[5–8]GCTATGC)–3′ to establish the structure of the cross-link within the helix, as well as the distortion to the helix after cross-link formation. Throughout the simulation (see Appendix C for details), cross-linked DNA adopts one of two conformations (designated Conformers 1 and 2), which differ in the orientation of the cross-linked dinucleotides in the helix, as well as the response of surroundings to the cross-link. The backbones of the cross-linked dinucleotides in Conformers 1 and 2 differ most
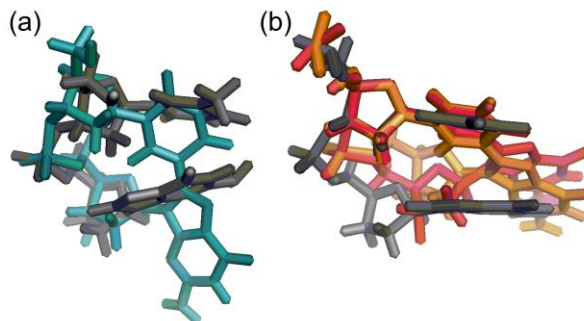
Figure 4–4: Overlay of (a) the natural 5′–TG–3′ sequence (grey) and 5′–U[5–8]G–3′ (blue) obtained using the dinucleoside monophosphate model, and (b) the natural (grey) and the of 5′–U[5–8]G–3′ (Conformers 1 (red) and 2 (orange)) dinucleotides obtained from the double stranded DNA model.

substantially at α_U, α_G and ζ_G (Table C–2). Deviations also exist in θ and χ, which results in a noticeable difference in the orientation of the cross-link in the two conformers (Figure 4–4(b)). Both conformers have C2′–*endo* puckering of the 5′–sugar and C1′–*exo* puckering of the 3′–sugar, as predicted by the dinucleoside monophosphate model.

The backbones of the cross-linked dinucleotides within Conformers 1 and 2 are unwound with respect to natural DNA (Figure 4–4(b)), which occurs to a greater extent than predicted by the dinucleoside monophosphate model Figure 4–4(b)). Furthermore, the cross-linked bases are largely in line with T in DNA, whereas the small model predicts a perpendicular arrangement between the natural and cross-linked nucleobases (Figure 4–4). These deviations between models are similar to those observed for 5′–G[8–5]U–3′ (Section 3.3.4.). Although U is in the *syn* orientation for both conformers (Table C–2), a widening of ∠(O4′C1′N1) also exposes O2, N3 and O4 of the Watson-Crick edge to the opposite strand, albeit in a distorted arrangement. Alternatively, the *anti* orientation of G (Table C–2) coupled with a widening of ∠(O4′C1′N9) exposes N7 and O6 of the Hoogsteen edge for base pairing. These distortions will affect the natural hydrogen-bonding patterns and duplex

stability. Therefore, genetic information is potentially lost upon cross-link formation, as found for 5′–G[8–5]U–3′. Indeed, the related 5′–T[5-Me–8]G–3′ cross-link is mutagenic *in vivo*.[111]

Although the conformers of 5′–G[8–5]U–3′-containing DNA differ solely in the orientation of the complementary nucleotides (Section 3.3.4), Conformers 1 and 2 of 5′–U[5–8]G–3′-containing DNA are defined by differences in the cross-linked dinucleotide, complementary nucleotides and flanking base pairs (Figure 4–5). In Conformer 1, the cross-link adopts an extrahelical position and the complementary A and C nucleotides do not interact with the cross-link. Instead, A adopts the *syn* orientation and moves further into the helix to form inter- and intrastrand stacking interactions with the base pair 3′ of it. Alternatively, C adopts the *anti* orientation and interacts with the backbone of the opposite strand. These significant distortions widen the helix at the cross-linked U nucleotide (d(C1′–C1′) = 13.272 ± 0.695 Å). The helix width is similar to, but more flexible than, natural DNA (10.672 ± 0.244 Å) at the cross-linked G nucleotide (10.785 ± 0.507 Å). Although the flanking G···C base pair 5′ of the cross-link is undisturbed (hydrogen bonding occupancy of 88 – 98 %), there are significant changes to the base pairs 3′ of the cross-link (Figure 4–5(a)). Most notably, G of the 3′-flanking C···G base pair adopts the high *anti*[vii] orientation (χ = 295.5 ± 14.4°), and interacts with the cross-link via π–π stacking interactions rather than hydrogen bonding to its complementary C, which adopts the *syn* orientation (χ = 306.7 ± 14.0°). These disruptions cause A of the next (3′) T···A base pair to be pushed out of the helix, which eliminates Watson-Crick hydrogen bonding in this base pair. Thus, significant distortions occur in the backbones of the cross-linked and complementary strand within in a 4 base-pair region that includes the cross-link and the two base pairs on the 3′ side.
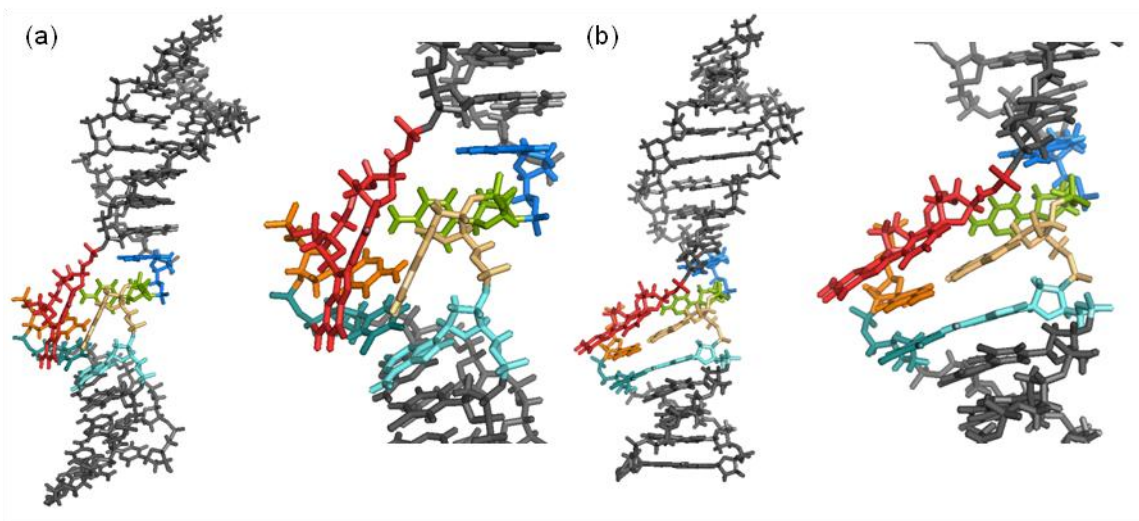
Figure 4–5: Conformers (a) 1 and (b) 2 of 5′–U[5–8]G–3′-containing DNA obtained from molecular dynamics simulations, highlighting the cross-linked dinucleotide (red) and complementary A (blue) and C (green) nucleotides, as well as the flanking C···G (orange) and T···A (teal) base pairs.

In Conformer 2, the cross-link remains extrahelical and does not interact with the complementary bases, which remain in similar orientations as described for Conformer 1(Figure 4–5(b)). However, the helix is wider at the cross-link compared to Conformer 1 (13.658 ± 0.524 Å and 12.161 ± 0.584 Å at U and G nucleotides, respectively). As with Conformer 1, the 5′-flanking G···C base pair is undisturbed and maintains Watson-Crick bonding (89 – 99%). Changes in the conformation of the cross-linked dinucleotides discussed above, as well as the backbones of the 3′-flanking C···G base pair result in the structural changes responsible for distinguishing the two conformers. These structural changes cause the 3′-flanking C···G base pair in Conformer 2 to engage in Watson-Crick hydrogen bonding (84 – 96 %), despite C in the *syn* (χ = 307.3 ± 11.7°) and G in the high *anti*[vii] (χ = 294.0 ± 27.2°) orientation. Furthermore, the entire C···G base pair interacts with the cross-link through π–π stacking interactions. Consequently, the next T···A base pair is
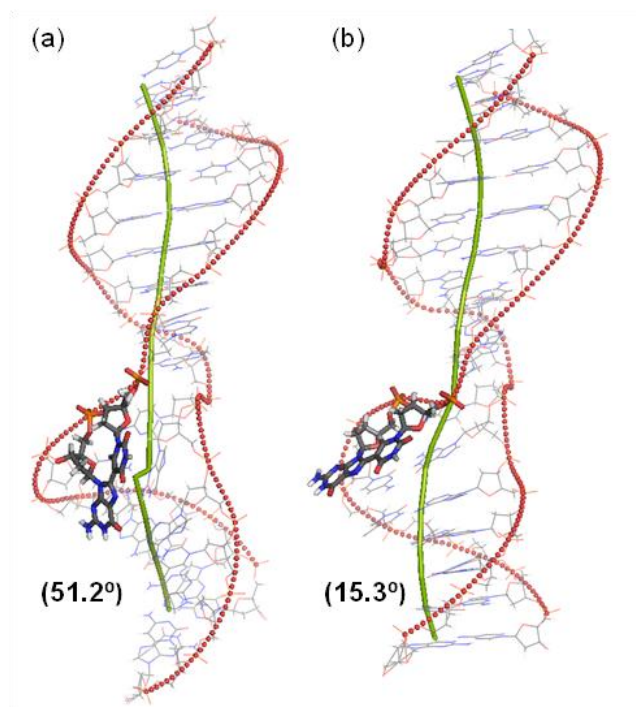
Figure 4–6: Illustration of the global structure of cross-linked DNA adopting Conformers (a) 1 and (b) 2, where the helix axis (green) and backbone (red) are identified. The axis bend is provided in brackets and is compared to a value of 9.5° for natural DNA.

able to engage in Watson-Crick hydrogen bonding (61 – 96 %). Nevertheless, significant distortions occur to the backbones of both strands at the cross-link and the two base pairs on its 3′ side.

The distortions in 5′–U[5–8]G–3′-containing DNA described above result in global changes to the helix. Specifically, significant alterations to the backbone occur and both conformers involve a bend in the helix axis towards the minor groove (Figure 4–6). This bend also alters the major and minor grooves. In Conformer 1, the major groove narrows substantially at the cross-link, while the minor groove becomes wider than the major groove. In fact, the major groove disappears in the region 3′ of the cross-link (Table C–3). Alternatively, in Conformer 2, the major groove disappears at the cross-linked dinucleotide,

while the minor groove narrows. As found for 5′–G[8–5]U–3′, these distortions to the helix axis are similar to those induced by successful anti-cancer agents, like cisplatin (35°),[33,40] mechlorethamine (17°)[212] and mitomycin C (15°).[213] However, the extent of this bend is greater for Conformer 1 of DNA containing 5′–U[5–8]G–3′ (51.2°) than any conformer of 5′–G[8–5]U–3′-containing DNA (23 – 41°). Alterations to the intrinsic bending of DNA may have implications in the regulation of fundamental biological processes, like repair, replication and/or transcription.[40] Alterations to the helix axis and grooves suggest 5′–U[5–8]G–3′, like 5′–G[8–5]U–3′, may contribute to the cytotoxicity of ^{X}U in anti-tumour therapies using mechanisms similar to successful anti-cancer drugs.

In the case of 5′–G[8–5]U–3′-containing DNA, the cross-linked G moiety protruded into the major groove, which changed the helix axis and grooves, but caused little disruption to the adjacent base pairs. In contrast, when the cross-link sequence is reversed, the cross-linked dinucleotide flips out of the helix and significantly disrupts the 3′-flanking base pairs, and may also induce a larger bend. These differences may also contribute to the preferences for 5′–G[8–5]U–3′ over 5′–U[5–8]G–3′. Due to the intricate pathways involved in cellular responses to DNA damage, which are further complicated in tumour cells, it is difficult to predict the biological implications of major distortions like extrahelical bases or bends in the helix axis. However, since 5′–U[5–8]G–3′ tends to inflict larger distortions to the helix than 5′–G[8–5]U–3′, there may be an increased likelihood of damage detection for the UG-lesion than the GU-counterpart despite compromised repair pathways in tumour cells.[54] Unfortunately, little is known about the repair of pyrimidine–purine intrastrand cross-links. Indeed, previous studies have examined the repair of 5′–G–pyrimidine–3′ cross-links,[106,108-110] but not the reverse sequence. Nevertheless, like other bulky lesions,[162] 5′–U[5–8]G–3′ is likely a substrate of NER, as proposed for 5′–G[8–5]U–3′ in Chapter 3.

If unrepaired, 5′–U[5–8]G–3′ is expected to affect the efficiency and fidelity of DNA replication, and may stall DNA polymerases.[104-107,110,111] If translesion synthesis occurs, the configuration of the cross-linked nucleobases may result in the insertion of purine nucleotides opposite both cross-linked nucleobases due to steric considerations,[5,9,22] which was similarly observed for 5′–G[8–5]U–3′ (Section 3.3.4.), and thereby cause transversion mutations. Therefore, it is likely that 5′–U[5–8]G–3′ is both mutagenic and cytotoxic, and will significantly contribute to the effectiveness of ^{X}U in anti-tumour therapies.

4.4. Conclusions

The formation pathway of 5′–U[5–8]G–3′ following UV or ionizing irradiation of ^{X}U-containing DNA has been characterized for the first time in this chapter. Under UV irradiation, the reaction proceeds from the 5′–U•G•+–3′ reactant with both bases in the *syn* orientation, and bond formation occurs through a transition state on the triplet surface or via an ISC. Similar barriers for both mechanisms (~48 kJ mol^{-1}, bond formation) make it difficult to identify the preferred formation pathway using computational means. Experiment can verify whether bond formation occurs on the triplet surface through the detection of phosphorescence (681.5 nm). These findings differ from the formation pathway for 5′–G[8–5]U–3′, where the bond formation step has a larger barrier (70 – 76 kJ mol^{-1}) and phosphorescence would occur at 668 nm. 5′–U[5–8]G–3′ formation from ionizing radiation, or homolysis of ^{I}U-containing DNA, also involves *syn* nucleobase orientations prior to base coupling, but is less likely than from UV radiation since hydrogen atom removal has a large barrier and the reaction is endergonic. However, formation of 5′–U[5–8]G–3′ via ionizing radiation has a lower barrier for hydrogen atom removal and is more thermodynamically favoured than the experimentally-observed 5′–T[5-Me–8]G–3′ cross-link that forms in natural DNA under the same irradiation conditions. This means

5′–U[5–8]G–3′, as well as its isomer 5′–G[8–5]U–3′, will likely form when ^{x}U is used in anti-tumour therapies.

Although both nucleobase stacking and the accessibility of C8 in the cross-linked intermediate have been previously proposed to dictate the preference for cross-links to form in 5′–purine than 3′–purine sequences, these factors cannot explain the experimentally-observed preference for 5′–G[8–5]U–3′ > 5′–U[5–8]G–3′. While the calculated barriers associated with the highest-energy chemical step (bond formation) also cannot explain this preference, the stability of the resulting cross-linked intermediate does agree with these yields and may therefore play a role. Additionally, experimental studies have noted a correlation in the distance between the sites forming the bond and the cross-link yield.[98-102] This chapter verifies that this proximity argument can be used to predict the relative yields of GU intrastrand cross-links. Furthermore, this chapter proposes the relationship between proximity and formation preference may be due to: (1) additional steps and/or larger barriers required to reorient sites that are not in close proximity; or (2) the fast time-scale of a radical reaction compared to necessary conformational changes. The first argument is supported by observations in this thesis that formation of 5′–U[5–8]G–3′ requires a major, and likely energetically-expensive, conformational change (*anti*-to-*syn* transition of both nucleobases), whereas 5′–G[8–5]U–3′ formation does not require *syn* orientations of the bases. This thesis also emphasizes that proximity may provide a universal predictor of the preferences for intrastrand cross-link formation, regardless of the method of generation or type of DNA considered.

The structure of 5′–U[5–8]G–3′ has been determined for the first time using a dinucleoside monophosphate model and dsDNA. 5′–U[5–8]G–3′-containing DNA adopts two conformers, where distortions are localized to the region of the cross-linked dinucleotide

and the two base pairs flanking the 3′ side. The distortion to dsDNA is significantly larger for 5′–U[5–8]G–3′- than 5′–G[8–5]U–3′. Therefore, 5′–U[5–8]G–3′ may be repaired, while 5′–G[8–5]U–3′ may persist in cells. Global changes to DNA containing 5′–U[5–8]G–3′ alter the grooves and create a bend in the helix axis that parallels distortions used by drugs, like cisplatin, to inflict tumour cell death. This indicates that, if left unrepaired, 5′–U[5–8]G–3′ may contribute to the cytotoxicity of ^{X}U in anti-tumour therapies using similar mechanisms.

This chapter has determined the formation mechanism for 5′–U[5–8]G–3′, and provided insight into its biological implications. Proximity has been shown to predict cross-link formation preferences for isomeric sequences. Furthermore, it is apparent that isomeric (5′–G[8–5]U–3′ and 5′–U[5–8]G–3′) intrastrand cross-links can inflict different distortions in the cross-link region, and yet both result in changes to the helix grooves and axis. The next chapter will examine the structure and formation of A-containing intrastrand cross-links to validate whether the proposed proximity argument can also predict the sequence selectivity of additional intrastrand cross-links and examine helical distortion to determine their potential cellular effects.

Chapter 5. Adenine–Uracil Intrastrand Cross-links

5.1.   Introduction

The A-containing cross-links are generated from a reactant with U$^{\bullet}$ adjacent to natural A, regardless of whether formed from the exposure of ^{X}U-containing cellular DNA to UV or ionizing radiation. The formation of the experimentally-observed UV-generated intrastrand cross-links exhibit a preference for 5′–A[8–5]U–3′ over 5′–A[2–5]U–3′ (Figure 5–1).[66,67] To provide a more complete picture of the factors governing cross-link formation, this chapter examines the formation mechanism of A-containing cross-links and complements the results presented in Chapters 3 and 4 for G-containing cross-links. Although a formation pathway has been previously proposed for A-containing cross-links (Section 1.5.3.),[66] the reaction within cellular DNA was incorrectly initiated from a reactant containing A$^{\bullet+}$ (Figure 1–10). Therefore, an updated, plausible formation mechanism (Figure 5–1) for these cross-links is investigated for the first time in this chapter. An attempt will be made to rationalize the product preferences within the A-containing cross-links using calculated kinetic and thermodynamic parameters, as well as the proximity argument proposed in Chapter 4. Comparisons will also be made between the formation mechanisms of G- and A-containing cross-links to provide insight into the overall preferences of intrastrand cross-link formation. Additionally, the structures of these cross-links in DNA, as well as the distortion inflicted to the helix when the cross-links are present, will be presented. Finally, the structures of the four cross-links in DNA in will be compared, and their biological implications discussed.
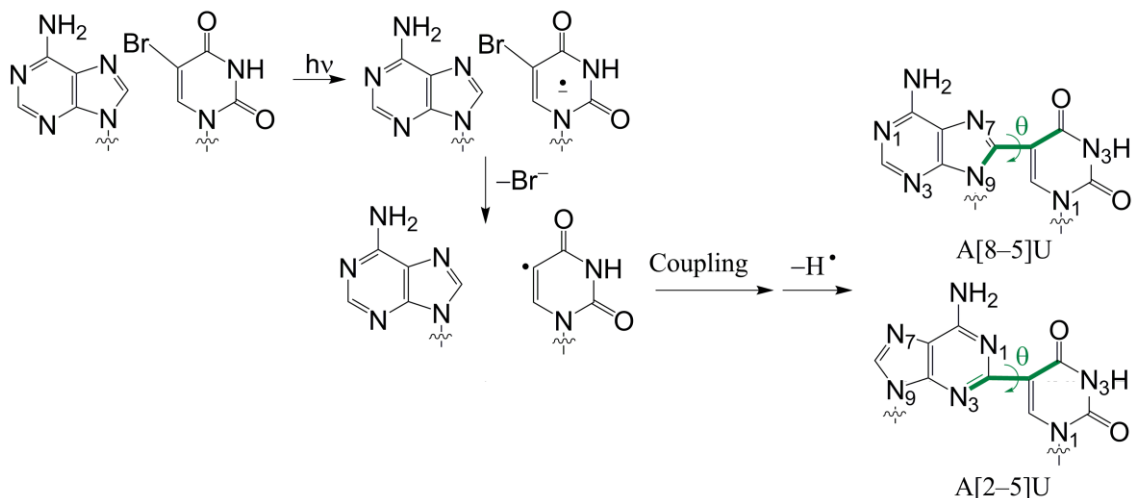
Figure 5–1: A new reaction scheme for the formation of A-containing cross-links from A and U•, corrected from that originally proposed (Ref. 66). Atomic numbering is shown and the torsion angles describing nucleobase orientation for 5′–A[8–5]U–3′ [θ(∠(N9C8C5C4))] and 5′–A[2–5]U–3′ [θ(∠(N2N3C5C4))] are indicated with bold.

5.2.   Computational Details

The formation of 5′–A[8–5]U–3′ and 5′–A[2–5]U–3′ are studied using the computational approach recommended in Chapter 2 and the same methodology described in Section 3.2.. As for the G-containing cross-links, molecular dynamics simulations were performed to determine the structure of these lesions within dsDNA with the 5′–d(GCATGGC**AT**GCTATGC)–3′ sequence. Due to substantial differences in the backbones of cross-linked and natural DNA (see Section 5.3.1.1.), cross-linked dsDNA was generated by overlaying the O5′ and C5′ atoms of both nucleosides in the M06-2X-optimized cross-links onto the corresponding atoms of the central 5′–AT–3′ (bold). The corresponding natural strand was also simulated for comparison. Details of the simulations are included in Appendix D.
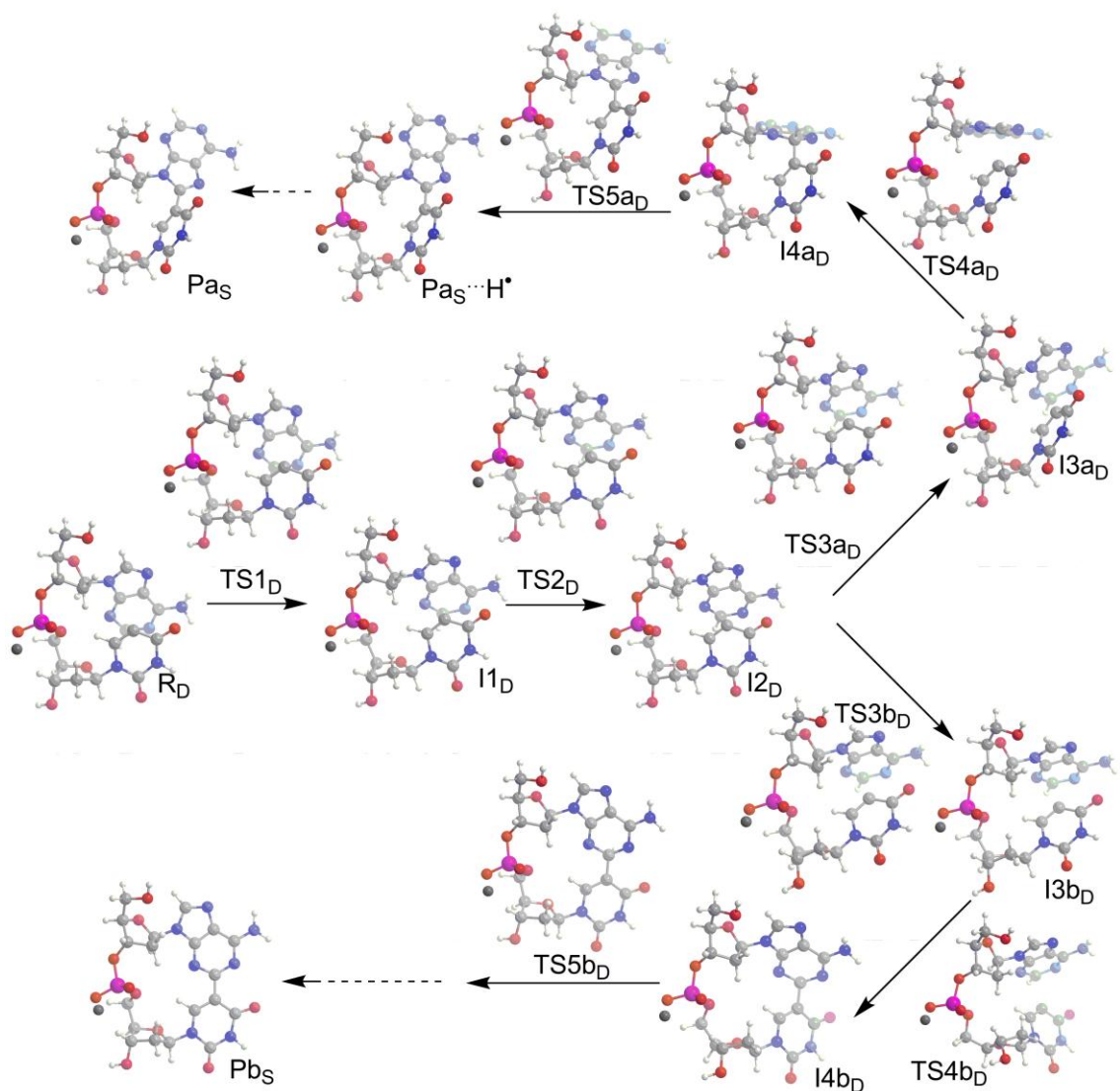
Figure 5–2: Stationary points for the formation of 5′–A[8–5]U–3′ (Pa$_D$) and 5′–A[2–5]U–3′ (Pb$_D$) from 5′–AU•–3′.

5.3.   Results and Discussion

5.3.1.   Formation Mechanisms

5.3.1.1. The Formation of 5′–A[8–5]U–3′ and 5′–A[2–5]U–3′

The 5′–AU•–3′ reactant (R$_D$, Figure 5–2) is characterized by a stacked configuration of *anti* nucleobases (χ_A = 225.4° and χ_U = 246.0°), and C2′–*endo* sugar puckering (P$_A$ = 169.4°,

$P_U = 174.0°$). Mulliken atomic spin densities show the unpaired electron is localized at C5 of U (1.044 e⁻). The production of both 5'–A[8–5]U–3' and 5'–A[2–5]U–3' proceeds from R_D through several low-energy bond rotations to $I2_D$. Overall, these rotations primarily occur in χ_A, χ_U and α, which yields an intermediate ($I2_D$) characterized by tilted (~42.2°) nucleobase planes and spin density localized at C5 of U (1.039 e⁻). Following $I2_D$, the pathways diverge to form each product (Figures 5–2 and 5–3).
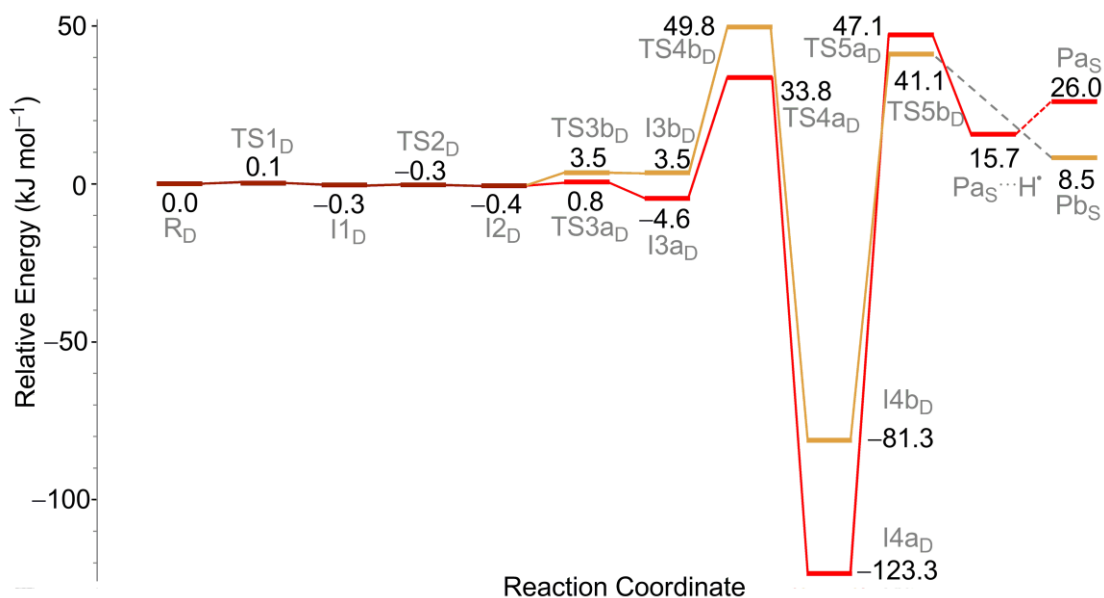
Figure 5–3: Cross-link formation on the doublet surface initiated by 5'–AU•–3', showing common steps along the pathway (dark red) leading to 5'–A[8–5]U–3' (pathway 'a', red) and 5'–A[2–5]U–3' (pathway 'b', orange). Energies (kJ mol⁻¹) are reported relative to the reactant.

During formation of 5'–A[8–5]U–3' (designated pathway "a", Figure 5–2), another rearrangement step is required after $I2_D$ prior to bond formation. Specifically, χ_U changes from 235.2° ($I2_D$) to 212.4° ($TS3a_D$) to 180.6° ($I3a_D$) through a 1.2 kJ mol⁻¹ ($\Delta G^{\ddagger} = 7.8$ kJ mol⁻¹) barrier. Subsequently, bond formation yields the cross-linked intermediate ($I4a_D$)

Table 5–1: Relative energies and thermodynamic parameters (kJ mol^{-1}) for the formation of 5′–A[8–5]U–3′ from 5′–AU•–3′.[a]

Stationary Point[b]	ΔE_{Opt}[c,d]	ΔE_{ZPVC}[d,e]	ΔH[d,f]	ΔG[g]	$T\Delta S$[h]
R$_D$	0.0	0.0	0.0	0.0	0.0
TS1$_D$	0.1	−0.3	−2.4	3.7	−6.1
I1$_D$	−0.3	−1.3	−0.9	−2.5	1.6
TS2$_D$	−0.3	−1.0	−3.1	5.1	−8.2
I2$_D$	−0.4	−1.0	−0.7	−1.3	0.6
TS3a$_D$	0.8	−1.3	−2.8	6.5	−9.3
I3a$_D$	−4.6	−6.1	−5.2	−5.6	0.4
TS4a$_D$	33.8	28.7	28.2	22.8	5.4
I4a$_D$	−123.3	−121.8	−123.2	−132.2	9.0
TS5a$_D$	47.1	30.0	28.5	31.6	−3.1
Pa$_S$···H•	15.7	−0.7	−0.9	2.9	−3.8
Pa$_S$ + H•	26.0	4.6	7.7	12.1	−4.4
TS3b$_D$	3.5	1.6	0.0	3.8	−3.8
I3b$_D$	3.5	2.3	2.9	−1.6	4.5
TS4b$_D$	49.8	43.2	43.4	37.5	5.9
I4b$_D$	−81.3	−79.1	−81.2	−84.1	2.9
TS5b$_D$	41.1	22.6	21.3	28.8	−7.5
Pb$_S$···H•	_[i]	_[i]	_[i]	_[i]	_[i]
Pb$_S$ + H•	8.5	−14.8	−11.0	−4.1	−6.9

[a]Reactant generated from exposure of 5′–A^{X}U–3′ to UV or ionizing radiation. [b]Refer to Figure 5–2 for the structures of the corresponding stationary points. [c]Optimization energy. [d]Calculated with IEF-PCM-M06-2X/6-31G(d,p). [e]Includes scaled (0.9580) zero-point vibrational energy correction. [f]Includes scaled (0.9480) thermal correction to the enthalpy. [g]Calculated using SMD-M06-2X/6-31G(d,p) energy (including non-electrostatic component) and IEF-PCM-M06-2X/6-31G(d,p) thermal correction to the Gibbs Free Energy. [h]Calculated as TΔS = ΔH−ΔG. [i]Stationary point could not be obtained.

with a decrease in the distance between C8 of A and C5 of U from 3.634 Å (I3a$_D$) to 2.274 Å (TS4a$_D$) to 1.515 Å (I4a$_D$) through a 38.3 kJ mol^{-1} ($\Delta G^{\ddagger}$ = 28.4 kJ mol^{-1}) barrier in a spontaneous process (Table 5–1). This step is accompanied by changes in χ_A and χ_U, a lengthening of d(C8–H8) as C8 becomes sp^3-hybridized, and delocalization of the unpaired electron on G (N7 = 0.493 e$^-$, N9 = 0.170 e$^-$, H8 = 0.178 e$^-$, C6 = 0.133 e$^-$ and C2 = 0.107 e$^-$). Subsequently, the H8 atom of A is removed with a 170.4 kJ mol^{-1} (TS5a$_D$, $\Delta G^{\ddagger}$ = 163.8 kJ mol^{-1}) barrier to yield the Pa$_S$···H• complex that contains a hydrogen atom (0.946 e$^-$) bridged

between O4 of U and the terminal O5′. A final decomplexation step (Figure 5–2) leads to infinitely separated species (Pa_S + H•). The 5′–A[8–5]U–3′ product is characterized by distortions in $α_U$ and $β_U$ (Figure 5–4(a) and Table D–2) compared to natural DNA. Product formation is accompanied by a 26.0 kJ mol^{-1} increase in energy, and is endothermic ($ΔH_{Rxn}$ = 7.7 kJ mol^{-1}) and endergonic ($ΔG_{Rxn}$ = 12.1 kJ mol^{-1}).
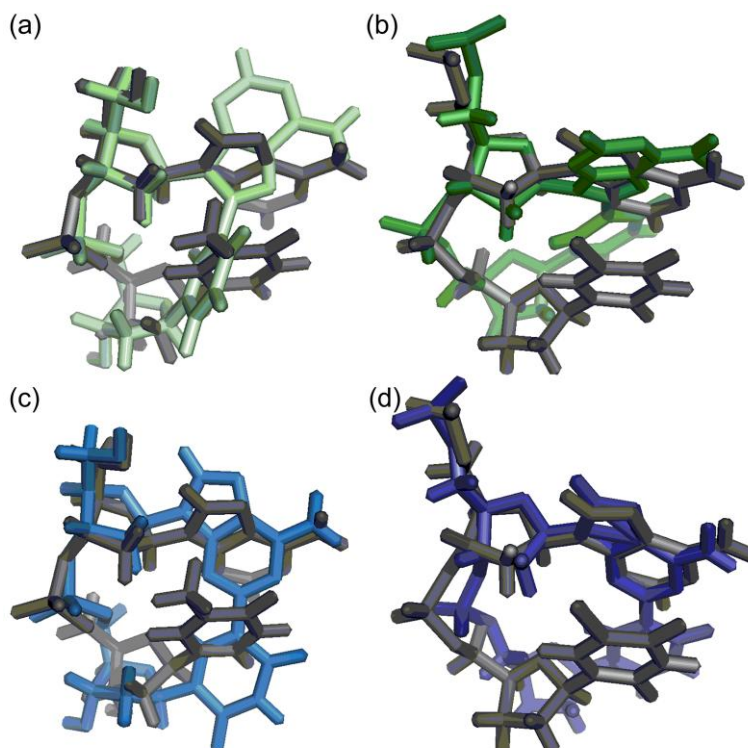
Figure 5–4: Overlay of (a, b) 5′–A[8–5]U–3′ (green) and (c, d) 5′–A[2–5]U–3′ (blue) with the natural 5′–AT–3′ sequence (grey) obtained from (a, c) the DFT dinucleoside monophosphate model (left, light) and the MD dinucleotides in the double stranded DNA model (right, dark).

Along the formation pathway of 5′–A[2–5]U–3′ (designated pathway "b", Figure 5–2), a rearrangement from $I2_D$ changes $χ_U$ and causes significant alterations to the backbone conformation (particularly $ε_A$ and $α_U$). This accommodates the movement of C5 of U towards C2 of A through a 3.9 kJ mol^{-1} ($ΔG^{‡}$ = 5.1 kJ mol^{-1}) barrier as d(C2–C5) decreases

from 3.562 Å (I2$_D$) to 3.386 Å (I3$_D$), and the unpaired electron remains localized on U (1.050 e$^-$). Next, bond formation occurs with a 46.3 kJ mol^{-1} (TS4a$_D$, G = 39.1 kJ mol^{-1}) barrier and changes in χ_A and χ_U, as well as the backbone conformation at ζ_A and γ_U, to give I4$_D$. In this structure, the backbone is highly distorted, the electron is delocalized over A (N3 = 0.395 e$^-$, N1 = 0.252 e$^-$, C5 = 0.235 e$^-$, H2 = 0.178 e$^-$ and C8 = 0.118 e$^-$) with negligible spin density on U (C5 = 0.018 e$^-$). Subsequently, the H2 atom is removed from A with a 122.4 kJ mol^{-1} ($\Delta G^{\ddagger}$ = 112.9 kJ mol^{-1}) barrier as d(C2–H2) increases from 1.105 Å (I4b$_D$) to 1.697 Å (TS5b$_D$) and the unpaired electron begins to localize on H2 (0.352 e$^-$) in TS5b$_D$. A complex between the cross-link product and a hydrogen atom could not be isolated along this pathway. However, the corresponding complexes were found for the three remaining cross-links due to bridging interactions between H$^{\bullet}$ and the cross-link, which are likely artefacts of using an isolated model. Interestingly, the barrier for H$^{\bullet}$ removal is lowest for 5′–A[2–5]U–3′ where this artefact is not present, which indicates hydrogen-atom removal will be easier under biological conditions as proposed in Chapters 3 and 4. Overall, the 5′–A[2–5]U–3′ structure is significantly distorted at α_U, β_U and γ_U (Figure 5–4(c) and Table D–2), and the formation of this cross-link is accompanied by an increase in energy of 8.5 kJ mol^{-1}, is exothermic (ΔH_{Rxn} = –11.0 kJ mol^{-1}) and slightly exergonic (ΔG_{Rxn} = –4.1 kJ mol^{-1}).

Since both UV and ionizing irradiation of ^{x}U-containing DNA yield the same (5′–AU$^{\bullet}$–3′) reactant and A-containing cross-links have been experimentally observed using UV radiation, these lesions must also be produced by ionizing radiation. Additional support for the formation of 5′–A[8–5]U–3′ from ionizing radiation is provided by a comparison of its formation pathway to an experimentally-observed 5′–A[8–5–Me]T–3′ cross-link. Specifically, 5′–A[8–5–Me]T–3′ formation is highly endergonic (ΔG_{Rxn} = 162 kJ mol^{-1}) with bond formation ($\Delta G^{\ddagger}$ = 185 kJ mol^{-1}) and hydrogen-atom removal ($\Delta G^{\ddagger}$ = 154 kJ mol^{-1}) having substantial barriers. In contrast, 5′–A[8–5]U–3′ formation is more kinetically and

thermodynamically favoured with a smaller barrier for hydrogen-atom removal ($\Delta G^{\ddagger}$ = 163.8 kJ mol⁻¹) and overall slightly endergonic reaction (ΔG_{Rxn} = 12.1 kJ mol⁻¹) reaction.

5.3.1.2. Comparison of A-Containing Cross-links and Their Formation Preferences

The formation of both 5′–A[8–5]U–3′ and 5′–A[2–5]U–3′ is near thermoneutral (Table 5–1), with 5′–A[8–5]U–3′ formation endergonic (ΔG_{Rxn} = 12.1 kJ mol⁻¹) and 5′–A[2–5]U–3′ generation exergonic (ΔG_{Rxn} = –4.1 kJ mol⁻¹). These results suggest cross-link formation is not dictated by the thermodynamic stability of the product, although different cross-link structures are predicted by truncated and extended models (Section 5.3.1.), as discussed for GU cross-links (Sections 3.3.4. and 4.3.3.).

Since the formation of 5′–A[8–5]U–3′ and 5′–A[2–5]U–3′ share several stationary points at the beginning of the reaction, the selectivity between these two products must occur after these steps. For both cross-links, the step with the largest barrier, as calculated by the dinucleoside monophosphate model, is hydrogen-atom removal, which occurs with different barriers for 5′–A[8–5]U–3′ ($\Delta G^{\ddagger}$ = 163.8 kJ mol⁻¹) and 5′–A[2–5]U–3′ ($\Delta G^{\ddagger}$ = 112.9 kJ mol⁻¹). Therefore, the ease of this step cannot account for the 5′–A[8–5]U–3′ > 5′–A[2–5]U–3′ preference. Additionally, the barrier for this step in 5′–A[8–5]U–3′ is similar to that for the G-containing cross-links considered is this thesis, as well as T-containing cross-links calculated in another study[114,115] ($\Delta G^{\ddagger}$ = 150 – 170 kJ mol⁻¹). This indicates that the ease of hydrogen-atom removal does not dictate the formation preferences for any intrastrand cross-links.

Further comparisons of the formation pathways of 5′–A[8–5]U–3′ and 5′–A[2–5]U–3′ reveals the highest point on the surface for 5′–A[2–5]U–3′ involves the bond formation step (ΔG = 37.5 kJ mol⁻¹), which is higher on the reaction surface than the highest point for 5′–A[8–5]U–3′, which involves hydrogen-atom removal (ΔG = 31.6 kJ mol⁻¹). Closer

examination of the structures in the bond formation step reveals larger distortions occur for 5′–A[2–5]U–3′ to accommodate a bond between C2 of A and C5 of U compared to those observed for 5′–A[8–5]U–3′. This distortion is reflected (to some extent) in the bond-formation barrier, which is greater for 5′–A[2–5]U–3′ ($\Delta G^{\ddagger}$ = 39.1 kJ mol^{-1}) than 5′–A[8–5]U–3′ ($\Delta G^{\ddagger}$ = 28.4 kJ mol^{-1}), and causes this step to be the highest point on the surface for 5′–A[2–5]U–3′. Additionally, bond formation results in an intermediate that is $\Delta(\Delta G_{Step})$ = 33.9 kJ mol^{-1} more favourable for 5′–A[8–5]U–3′ than 5′–A[2–5]U–3′. Increased structural distortions are also present for 5′–A[2–5]U–3′ compared to 5′–A[8–5]U–3′ during the previous rearrangement step at TS3$_D$. However, the increased distortion for 5′–A[2–5]U–3′ is not reflected in the relative barriers at TS3$_D$, which is slightly lower ($\Delta(\Delta G)$ = 8.1 kJ mol^{-1}, Table 5–1)) for 5′–A[8–5]U–3′ than 5′–A[2–5]U–3′. This may suggest the energetic penalties associated with large conformational changes in DNA helices are not fully accounted for in a dinucleoside monophosphate model and require explicit consideration of the cross-link surroundings for an accurate description. Indeed, conformational changes might even be coupled with the bond-formation step in DNA. Given the large structural distortions observed in the rearrangement step for 5′–A[2–5]U–3′ and the importance of conformational changes established in Chapter 4, the rearrangement step cannot be eliminated as influencing cross-link formation preferences. The correlation between cross-link yields and stability of the cross-linked intermediate was also noted in Chapter 4, and is again observed for A-containing cross-links. Therefore, it is proposed that cross-link formation is governed by the barrier for the bond-formation step, the stability of the resulting cross-linked intermediate and the degree of structural distortion along the reaction pathway.

The above factors are not mutually exclusive from the proposal in Chapter 4 that the proximity of the sites forming the bond plays a role. Indeed, distances from averaged fibre

diffraction data relevant to the A-containing cross-links (d(C8–C5) = 3.9 Å and d(C2–C5) = 5.1 Å)[202] support the importance of proximity. Interestingly, before the formation pathways of the two A-containing lesions diverge at $I2_D$, C5 of U is nearly equidistant to C8 (3.550 Å) and C2 (3.562 Å) of A and ultimately the preference for 5′–A[8–5]U–3′ is dictated by the ease of subsequent backbone distortions and bond formation, as well as the stability of the cross-linked intermediate. This provides support for the proposal in Chapter 4 that the importance of proximity may manifest itself in the barriers involving conformational changes, which include backbone distortions. However, this chapter also establishes that the proximity argument is solely a predictive tool, and is most clear when distances are measured from B-DNA rather than points along a reaction pathway.

5.3.2.   *The Structures of 5′–A[8–5]U–3′ and 5′–A[2–5]U–3′ and Helical Distortion*

To assess the cross-link structure within dsDNA, as well as the distortions induced to the helix, molecular dynamics simulations were performed on 5′–A[8–5]U–3′- and 5′–A[2–5]U–3′-containing dsDNA in 5′–d(GCATGGCAUGCTATGC)–3′. Unlike the G-containing cross-links, each A-containing cross-linked helix adopts one conformer throughout the simulation (see Appendix D, page D–4 for details).

Within dsDNA, the backbone conformations of the 5′–A[8–5]U–3′ and 5′–A[2–5]U–3′ dinucleotides differ from natural DNA (Figure 5–4 (b) and (d), respectively), primarily due to unwinding of the backbone. However, larger deviations and unwinding are observed for 5′–A[2–5]U–3′ than 5′–A[8–5]U–3′(Table D–2), which correlate with the larger calculated bond-formation barrier (Table 5–1), and the larger distortions predicted by the dinucleoside monophosphate model along the formation pathway for 5′–A[2–5]U–3′. Within the helix, the deoxyribose moieties of 5′–A[8–5]U–3′ adopt C1′–*exo* puckering (P_A = 132.3°, P_U = 141.3°), while the 5′ sugar adopts a C1′–*exo* pucker (P_A = 138.4°) and the 3′ sugar

adopts a C2′–*endo* pucker (P_U = 159.6°) in 5′–A[2–5]U–3′. Additionally, the twist between the cross-linked bases in DNA is $\theta \approx 165^0$ for both lesions. Both cross-links are positioned in the same plane as A in natural DNA (Figure 5–4(b, d)). Within 5′–A[8–5]U–3′, A adopts the *syn* orientation about the glycosidic bond (χ_A = 42.0 ± 7.5°) to expose the Hoogsteen face to the complementary strand, while U adopts an *anti* (χ_U = 212.8 ± 8.7°) orientation and the Watson Crick face remains available for hydrogen bonding. Similar alterations to the hydrogen-bonding abilities of the cross-link were observed for 5′–G[8–5]U–3′ (Section 3.3.4.). In 5′–A[2–5]U–3′, both A and U adopt the *anti* orientation (χ_A = 203.4 ± 16.8°, χ_U = 201.7 ± 11.3°) and the Watson Crick faces are available for hydrogen bonding. In fact, the Watson-Crick faces of A in natural DNA and cross-linked A in 5′–A[2–5]U–3′ are similarly aligned (Figure 5–4(d)). Together, this indicates this lesion may be less mutagenic than 5′–A[8–5]U–3′.

Within the helices (Figure 5–5), the A-containing cross-links adopt extrahelical positions. Therefore, there is a lack of hydrogen bonding between the cross-links and complementary bases, and distortions are present in the backbone of the complementary T and A nucleotides. In 5′–A[8–5]U–3′, the helix widens at the cross-linked A (11.420 ± 0.878 Å) and U (12.499 ± 0.781 Å) nucleotides. Even greater widening is observed for 5′–A[2–5]U–3′ at the A (11.396 ± 1.277 Å) and U (12.732 ± 1.323 Å) nucleotides compared to natural DNA (10.716 ± 0.277 Å and 10.609 ± 0.288 Å for the A and T nucleotides, respectively). Although the complementary (T and A) bases remain in the helix, the above structural changes create a solvent-accessible cavity. The flanking base pairs are unperturbed in the presence of either 5′–A[8–5]U–3′ (hydrogen bonding occupancy of 79–98 %) or 5′–A[2–5]U–3′ (81–99 %) compared to natural DNA (80–98 %).
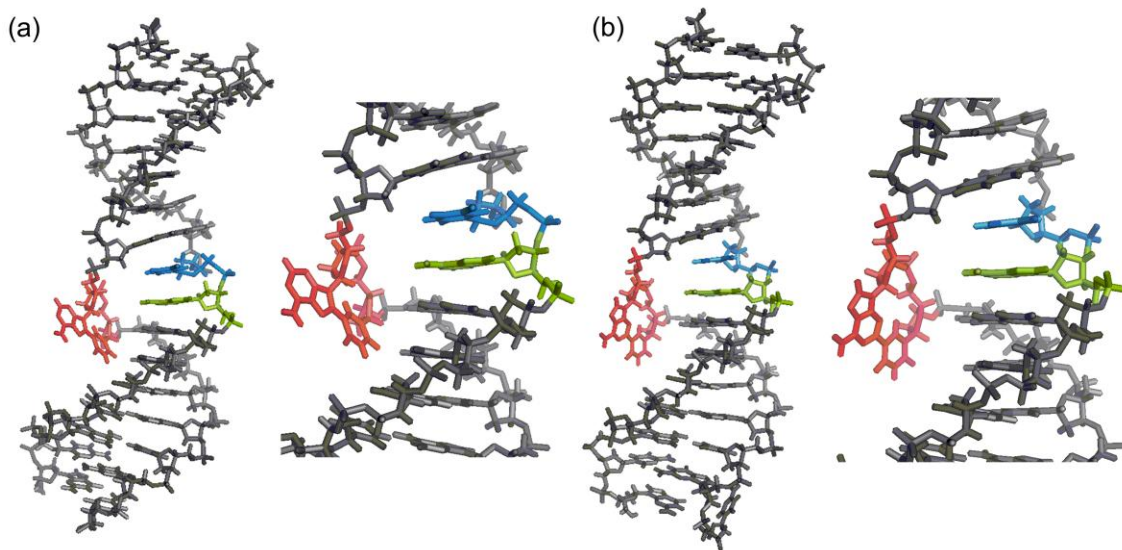
Figure 5–5: The MD conformations of (a) A[8–5]U–3′- and (b) 5′–A[2–5]U–3′-containing DNA, highlighting the cross-linked dinucleotide (red) and complementary T (blue) and A (green) nucleotides.

(26.4 °)   **(15.8 °)**
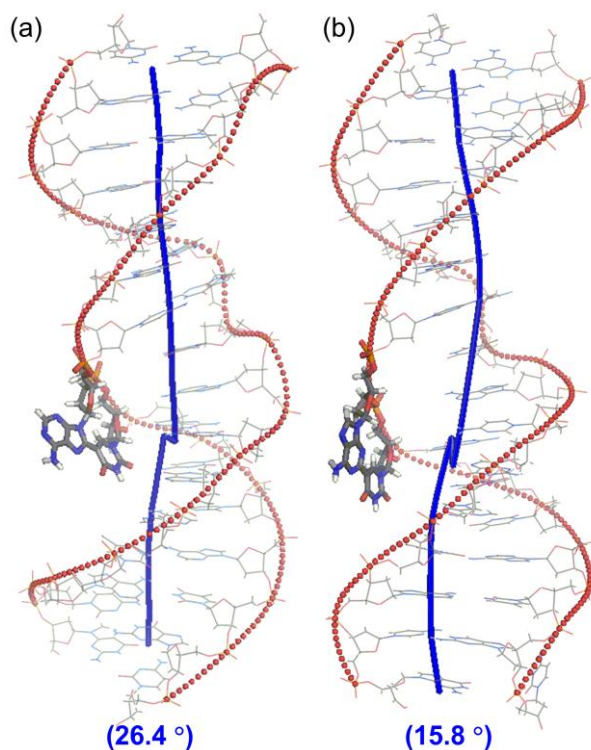
Figure 5–6: Illustration of the global structure of cross-linked DNA containing (a) 5′–A[8–5]U–3′ or (b) 5′–A[2–5]U–3′, where the helix axis (blue) and backbone (red) are shown. The axis bend is provided in brackets and is compared to a value of 5.6° for natural DNA.

Global distortions to the helix are inflicted by both 5′–A[8–5]U–3′ and 5′–A[2–5]U–3′. Specifically, the cross-links induce a bend in the helix (Figure 5–6), which is larger for 5′–A[8–5]U–3′ (26.4°) than 5′–A[2–5]U–3′ (15.5°) in comparison to the corresponding natural helix (5.6°). For both lesions, the major and minor grooves narrow at the cross-link and the major groove disappears at several regions along the strand (Table D–3). Substantial deviations in the backbones of the complementary nucleotides are also observed.

5.3.3. A Comparison of the Four Intrastrand Cross-links

5.3.3.1. Understanding the Sequence Selectivity

UV irradiation of [Br]U-containing cellular DNA produces four intrastrand cross-links with yields (5′–A[2–5]U–3′ < 5′–A[8–5]U–3′ < 5′–U[5–8]G–3′ < 5′–G[8–5]U–3′) demonstrating a sequence selectivity for G over A, and a 5′–purine over a 3′–purine. Due to differences in the mechanisms of UV-enabled U• generation in G- versus A-containing sequences (G•+U• versus AU• reactant, Section 1.4.2.2.), the formation of G- and A-containing cross-links under these conditions cannot be cross compared since the generation of U• may play a role in the observed preferences. However, the preferences within either the G- or A-containing cross-links can be rationalized. Cross-link yields can be predicted by the proximity of the two sites forming the bond, which was shown to agree with the yields of the four U-containing cross-links observed in cellular DNA (Section 4.3.2. and 5.3.1.2.). Proximity also accounts for 5′–G[N2–5]U–3′ and 5′–U[5–N2]G–3′ formation in (flexible) dinucleoside monophosphates, but not cellular DNA,[66,67] where fibre diffraction data measures d(N2–C5) = 5.3 Å for the 5′–G and d(C5–N2) = 7.8 Å for 3′–G sequences in B-DNA.[202] Despite the predictive power of proximity, its utility is limited to isomeric lesions and cannot be used to compare lesions of differing composition (G vs. A) since the proximity

between corresponding coupling sites is similar.[202] However, the importance of proximity due to the high reactivity of radicals cannot be ruled out. Interestingly, some support for the significance of proximity in DNA radical reactions is provided by the yields of UV-mediated cross-links. In particular, G-containing lesions originate from long-lived triplet states (G[•+]U[•]) where large conformational changes can occur, while A-containing cross-links originate from reactive doublet states (AU[•]) where strand break formation via hydrogen abstraction by U[•] from the nearby 5′–sugar is preferred. This may account for the preference for G over A in intrastrand cross-link formation from UV radiation.

This thesis proposes the structural distortions that occur during cross-link formation, the stability of the cross-linked intermediate and the ease of the bond-formation step work together to dictate cross-link preferences (Section 5.3.1.2.). In the case of G-containing cross-links, conformational changes and the stability of the cross-linked intermediate dictate the lesion preference. Specifically, upon formation of 5′–G[8–5]U–3′, little distortion is observed in the backbone and a very stable cross-link intermediate is formed with a 70 – 76 kJ mol^{-1} barrier for the bond-formation step. Alternatively, 5′–U[5–8]G–3′ formation involves a less-stable cross-linked intermediate and larger conformational changes since the nucleobases must adopt the *syn*, rather than *anti,* orientation, despite the lower barrier for the bond-formation step ($\Delta E^{\ddagger} = \sim 48.0$ kJ mol^{-1}). In the case of A-containing cross-links, the kinetics for the bond-formation step also play a role, and result in preferential formation of 5′–A[8–5]U–3′ over 5′–A[2–5]U–3′ (Section 5.3.1.2.).

Although these four intrastrand cross-links have not yet been observed following exposure of [X]U-containing DNA to ionizing radiation, the results presented in this thesis indicate all four lesions are likely to form when [X]U is used in conjunction with (ionizing)

radiation treatments as an anti-tumour therapy. However, the formation preferences of U-containing cross-links are unknown under these conditions. Because both G- and A-containing cross-links form from U• adjacent to the (undamaged) purine, comparisons can be made across all four cross-links. Since related purine–pyrimidine intrastrand cross-links[98,102,104] have demonstrated a sequence specificity, which is similar to the U-containing cross-links formed from UV radiation,[66,67] a sequence dependence is expected when U-containing intrastrand cross-links are formed from ionizing radiation. The barrier for the bond-formation step and the stability of the resulting cross-linked intermediate indicate a preference for 5′–G[8–5]U–3′ ($\Delta G^{\ddagger}$ = 23.6 kJ mol^{-1}, $\Delta(\Delta G_{Step})$ = –137.5 kJ mol^{-1}) followed by 5′–A[8–5]U–3′ ($\Delta G^{\ddagger}$ = 28.4 kJ mol^{-1}, $\Delta(\Delta G_{Step})$ = –126.6 kJ mol^{-1}), which also involves larger backbone distortions than its G-containing counterpart. As discussed in Section 5.3.1.2., the bond-formation step for 5′–A[2–5]U–3′ has a higher barrier and less stable cross-linked intermediate ($\Delta G^{\ddagger}$ = 39.1 kJ mol^{-1}, $\Delta(\Delta G_{step})$ = –82.5 kJ mol^{-1}), as well as even greater distortions to the DNA backbone along the formation pathway than 5′–A[8–5]U–3′. While the stability gained upon formation of the cross-link intermediate is similar for 5′–U[5–8]G–3′ (ΔG_{step} = –86.2 kJ mol^{-1}), it also involves the largest barrier for the bond-formation step ($\Delta G^{\ddagger}$ = 57.9 kJ mol^{-1}) and its formation requires the largest conformational change (both nucleobases transition from *anti* to *syn*). Although not explicitly examined in this thesis, this transition is high in energy and will further decrease the likelihood of this lesion forming in B-DNA. Thus, it is proposed that yields of cross-links generated upon exposure of [X]U-containing DNA to ionizing radiation will increase as 5′–U[5–8]G–3′ < 5′–A[2–5]U–3′ < 5′–A[8–5]U–3′ < 5′–G[8–5]U–3′. This proposal does not fully correlate with yields measured for T-containing cross-links observed in synthetic DNA (5′–T[5-Me–8]A–3′ < 5′–A[8–5-Me]T–3′ < 5′–T[5-Me–8]G–3′ < 5′–G[8–5-Me]T–3′).[98] However, the proposed yields for U-containing cross-links still have a preference for 5′–G[8–5]U–3′ over

5′–A[8–5]U–3′, which agrees with experimental observations for similar cross-links.[98-102,104-107,110,209]

The ease of the bond-formation step calculated in this thesis for U-containing cross-links does not support previous calculations stating the G > A preference of T-containing cross-link formation by ionizing radiation is explained by more facile bond formation for G regardless of (5′ or 3′) position.[114,115] However, these previous studies calculated sizeable barriers for bond formation in T-containing cross-links ($\Delta G^{\ddagger}$ = 88 – 100 kJ mol^{-1} for G and $\Delta G^{\ddagger}$ = 174 – 185 kJ mol^{-1} for A),[114,115] which were presumably poorly described by the B3LYP functional employed. Specifically, Chapter 2 found this computational approach to inadequately predict the structure of natural DNA, which suggests the reactant structures obtained in these studies are also inadequate and may involve repelled or distorted base arrangements. Furthermore, since this functional is well-known to poorly describe dispersion forces,[174] this will substantially affect the barriers. The results presented in this thesis provide a more accurate description of both the structures and barriers associated with intrastrand cross-link formation. Future work should re-examine T-containing intrastrand cross-link formation using the computational approach employed throughout this thesis (see Section 6.2.4.).

5.3.3.2. Cross-link Structures and Biological Implications

Since cross-link yield is not necessarily correlated with cytotoxicity, all four U-containing intrastrand cross-links may contribute to tumour cell death upon irradiation of ^{X}U. Additionally, the unique structure and composition of each cross-link may lead to different cellular implications. Therefore, the structures of the four U-containing intrastrand cross-links have been determined in this thesis for the first time. The structural differences observed for the isomeric 5′–G[8–5]U–3′ and 5′–U[5–8]G–3′ lesions, and the further

contrasting behaviour of the cross-linked helices containing 5′–G[8–5]U–3′ and 5′–A[8–5]U–3′ are both of notable interest. The intrahelical position of 5′–G[8–5]U–3′ may play a role in the formation preference for this cross-link compared to the extrahelical position of the three remaining cross-links, which form with lower yields. Indeed adopting an extrahelical position requires comparatively large backbone distortions, which provides additional support for the proposal that structural distortions play an important role in dictating formation preferences.

Regardless of the exact structure local to the cross-link, global distortions to the DNA helices parallel those induced by efficient anti-tumour therapies, like cisplatin,[33,40] for all cross-links. In particular, DNA helices with U-containing intrastrand cross-links exhibit a bend in the helix axis ranging from 15° to 51°. Additionally, alterations to both the major and minor grooves following cross-link formation mimic distortions caused by cisplatin, as well as other efficient inducers of apoptosis.[33,40,212,213] This indicates U-containing intrastrand cross-links may contribute to the cytotoxicity of ^{X}U in anti-tumour therapies using mechanisms similar to cisplatin. However, ^{X}U may actually prove to be a superior anti-cancer agent since strand breaks and interstrand cross-links may also form.[49,57-59,61,63,65,75-80] Specifically, ^{X}U-containing DNA may have the potential to create detrimental clustered lesions (2 or more damaged sites within 20 base pairs),[15,16] which may evade the resistances that accompany many drugs.

Experimental studies examining the synthesis of DNA containing G–pyrimidine intrastrand cross-links found that replication is highly dependent on the polymerase.[104-107,110] Commonly, DNA polymerases stall at the cross-linked G due to a lack of an appropriate hydrogen-bonding.[105,107,110] If a stalled polymerase is not resolved, the complex can collapse into a DSB[113] or completely block replication to cause cell death.[105,106]

This supports previous observations that direct formation of DSBs from U• cannot fully account for the observed cell death and that cellular processing of another lesion to a DSB is involved.[49] Alternatively, TLS may occur, which is commonly error-prone.[7,10,11] Although TLS of 5′–G[8–5-Me]T–3′ by human polymerase η inserts the correct nucleotides, mutations were still observed, which was attributed to the involvement of other polymerases or accessory proteins.[111] The structures presented in this thesis reveal that the Watson-Crick face of the purine in 5′–G[8–5]U–3′, 5′–U[5–8]G–3′ and 5′–A[8–5]U–3′ is unavailable for hydrogen-bonding and instead the Hoogsteen edge is exposed. Furthermore, the arrangement of the purine in these cross-links may cause DNA polymerases to favour the incorporation of a purine nucleotide opposite the G- or A-moieties due to steric considerations, which causes transversion mutations. In contrast, the Watson-Crick face of A is slightly altered in 5′–A[2–5]U–3′ due to the presence of the cross-link bond at C2 of A, which may affect the efficiency and fidelity of nucleotide incorporation upon replication. Furthermore, steric factors may generate a mispair by the DNA polymerase with a purine nucleotide inserted opposite A in 5′–A[2–5]U–3′ to create a transversion mutation. Additionally, the increased rigidity of the four cross-linked dinucleotides may stall the polymerase. Indeed, polymerases have been found to have difficulty bypassing lesions with locked orientations about the glycosidic bond (χ), even when the hydrogen-bonding face of the nucleobase is maintained, due to the inability of the lesion to undergo rotations necessary to fit into the enzyme active site.[215] Therefore, it will be important for future studies to consider the ability of human DNA polymerases to bypass these four cross-links. This will be particularly important for low-fidelity polymerases, which are over-expressed in tumour cells.[113] Inhibition of these polymerase may be an effective adjuvant to [x]U since DNA polymerases capable of TLS help cancer cells tolerate DNA damage.[113]

Given that many repair pathways are generally compromised in tumours[54] and tumour cell proliferation is highly dependent on the remaining, intact repair pathways,[36] lesion repair is an important indicator of the potential success of a given lesion to induce tumour cell death. Although the calculations presented in this thesis cannot explain how intrastrand cross-links are repaired, similarities can be drawn between U-containing and other intrastrand cross-links, as well as common bulky DNA lesions. NER pathways are the only repair mechanisms biologically available to remove intrastrand cross-links from DNA. Indeed, related G–pyrimidine intrastrand cross-links with [8–5] or [8–5-Me] connectivity are substrates for NER pathways involving UvrABC Nuclease (*E. coli*) and are removed with varying efficiencies.[108,109] Parallels have been drawn in the literature between G-containing intrastrand cross-links and addition products (adducts) resulting from addition of a bulky group to C8 of G since both structures involve a bulky (ring) substituent covalently attached to C8.[216] While both types of lesions undergo NER, recognition by UvrA is higher for 5′–G[8–5-Me]T–3′ than the C8-G adduct of N-acetyl-2-aminofluorene (AAF), which was proposed to be due to the larger alterations (bending, kinking and unwinding) to DNA induced by the cross-link.[108] Alternatively, incision by UvrC is less efficient for 5′–G[8–5-Me]T–3′ than the AAF adduct due to poorer recognition of the cross-link by UvrB.[108] This suggests intrastrand cross-links and C8-adducts are a distinct classes of lesions, and cross-links may resist repair to accumulate in cells more than other lesions, which may even be amplified in tumour cells due to hampered repair functions.[54] Indeed, tandem lesions, like intrastrand cross-links, are more challenging to repair than single-base lesions,[15,16,31,33] and therefore their formation is ideal for anti-cancer treatments.[15,16] Although the repair of U-containing intrastrand cross-links has not been studied, their synthesis in cellular DNA shows increased yields of 5′–U[5–8]G–3′ and 5′–A[8–5]U–3′ with irradiation time, but decreased yields of 5′–G[8–5]U–3′ and 5′–A[2–5]U–3′.[67] Although this

may be attributed to the preferential decomposition of 5′–G[8–5]U–3′ and 5′–A[2–5]U–3′, it may indicate that these lesions are more readily repaired *in vivo*.[67] Therefore, it will be of the utmost importance to determine the repair efficiencies of all U-containing intrastrand cross-links, and whether these lesions are excised by NER pathways in human tumour cells.

5.4. Conclusions

The formation of 5′–A[8–5]U–3′ and 5′–A[2–5]U–3′ share common initial steps that involve conformational changes. After these steps, an intermediate undergoes a unique conformational change, which is followed by bond formation and hydrogen-atom removal to form each product. The computational approach employed in this thesis finds the formation of both cross-links to be nearly thermoneutral and the largest barriers occur for hydrogen-atom removal, while bond formation is the highest-energy point for 5′–A[2–5]U–3′ and hydrogen-atom removal is the highest-energy point for 5′–A[8–5]U–3′. The ease of the bond-formation step and stability of the resulting cross-linked intermediate, as well as the degree of conformational changes along the pathway, dictate the preferences of intrastrand cross-link formation, which can be predicted by the proximity of the sites forming the bond. This proposal may be applicable to all intrastrand cross-links and thereby provide a universal explanation of the factors dictating the formation preferences regardless of the form of radiation used or type of DNA considered.

For the first time, the structures of four U-containing cross-links in DNA have been determined and compared using MD simulations. In both G- and A-containing cross-links, unwinding is observed in the cross-linked and complementary dinucleotides, as well as distortions to the grooves and a bend in the helix axis. Both A-containing cross-links and 5′–U[5–8]G–3′ adopt an extrahelical position, while 5′–G[8–5]U–3′ remains intrahelical. This major structural difference may contribute to the abundance of 5′–G[8–5]U–3′. All four

intrastrand cross-links are likely to be mutagenic upon replication due to changes in the exposed hydrogen-bonding face and/or the configuration of the cross-linked purine base. These lesions are also likely cytotoxic due to their potential ability to stall polymerases and difficulty to repair. Future work must be done to enhance our understanding of intrastrand cross-links, which is outlined in Chapter 6.

Chapter 6. Thesis Summary and Future Directions

6.1.   Summary

The goal of this thesis was to study four purine–U intrastrand cross-links (5′–G[8–5]U–3′, 5′–U[5–8]G–3′, 5′–A[8–5]U–3′ and 5′–A[2–5]U–3′) that form following UV irradiation of ^{Br}U-containing cellular DNA. This is of particular importance due to the applications of ^{X}U as a radiosensitizing agent in anti-tumour therapies. It was important to study the formation pathways of these lesions in order to determine whether U-containing intrastrand cross-links form under therapeutic conditions that employ ionizing radiation, as well as to provide insight into the observed sequence-selective formation of these lesions. Additionally, the structures of these intrastrand cross-links were previously unknown. These were important to determine since understanding the structures of these lesions in DNA, as well as the distortions they inflict on the helix, can be used to develop a sense of the biological implications of these lesions in (tumour) cells.

Various computational approaches were considered in Chapter 2 to determine their ability to reproduce the conformational features of DNA using a dinucleoside monophosphate model and high-level DFT calculations. Among the combinations considered, the M06-2X functional in conjunction with a (Na$^+$) counterion phosphate model was found to accurately describe the base–base orientation and backbone conformation of double-stranded B-DNA. This model was subsequently applied throughout the thesis to study the formation and structure of U-containing intrastrand cross-links. In addition to the specific applications in this thesis, this model is widely applicable. For example, it can be used to study non-covalent interactions and/or a variety of reactions involving adjacent nucleic acid components. An accurate DNA model of this size had not been previously

identified in the literature, where smaller truncations are more commonly used with high-level methods.

Using the computational approach recommended in Chapter 2, the formation pathways of four U-containing intrastrand cross-links were determined. The mechanism was first studied under conditions relevant to UV radiation, since this form of radiation was used to generate the lesions in experiments. G-containing cross-links, initiated from the $G^{•+}U^•$ reactant, were found to be generated with the largest barrier occurring for the formation step that may occur on the triplet surface (with the emission of phosphorescence) or be coupled with an intersystem crossing. Alternatively, the computational approach employed predicts A-containing cross-links, which originate from the $AU^•$ reactant, form with a substantial barrier occurring for the hydrogen atom removal step. However, this step will likely be more facile under biological conditions than found computationally. The calculated formation pathways of these lesions from ionizing radiation revealed all four intrastrand cross-links are likely to form under therapeutically-relevant conditions. Therefore, these lesions, in conjunction with interstrand cross-links and strand breaks, may contribute to the cytotoxicity of irradiated ^xU in tumour cells.

To shed light on their potential biological implications, the structures of these lesions in cells were determined for the first time using molecular dynamics simulations. Lesion formation commonly alters the available hydrogen-bonding faces of the cross-linked nucleobases, which may have implications in DNA replication by introducing mutations into the genome or stalling DNA polymerases if unrepaired. While the G-containing intrastrand cross-links adopt several different conformers throughout the simulation, only one conformer was found for each of the A-containing intrastrand cross-links. More importantly, with the exception of 5′–G[8–5]U–3′, all cross-linked dinucleotides adopted an

extrahelical position. This may be indicative of the larger backbone distortions and higher energetic costs required to form extrahelical lesions, and may also have implications in damage recognition and repair. All cross-linked helices contain bends in the helix axis ranging from 15–51˚, as well as changes to the grooves of the DNA helix and substantial distortions to the backbone in the region of the cross-linked nucleotides. These distortions may have implications in DNA repair since bending plays an important role regulating DNA–protein interactions.[40] However, given the repair efficiency of related intrastrand cross-links by NER, as well as the compromised repair functions in tumour cells, U-containing cross-links are likely to persist in cancer cells, where it is proposed they will be both mutagenic and cytotoxic.

In addition to determining the structures and formation pathways of the four intrastrand cross-links that result from U•, this thesis explains the sequence-specific formation of these lesions. In general, a simple comparison of thermodynamic and kinetic factors provided by the dinucleoside monophosphate model used in this thesis cannot provide a complete explanation for the preferences of intrastrand cross-link formation. Previous literature noted correlations between the cross-link preference and proximity of the two sites being coupled. This thesis proposes proximity can be used to predict cross-link yields and that the importance of proximity can be explained by the large conformational changes along the calculated reaction pathways required for cross-link formation when atoms are far apart, or also by the reactive nature of radicals compared to the time-scale required for large conformational changes. Indeed, the degree of structural distortion undergone upon U-containing cross-link formation, as well as the barrier of the bond formation step and stability of the resulting cross-linked intermediate, correlate with experimentally-observed cross-link yields,[66,67] which indicates these factors play a role in dictating the preferences of cross-link formation. However, the dinucleoside

monophosphate model used throughout this thesis is not without its limitations since conformational changes observed in a truncated model do not fully account for the distortions required for lesion formation in dsDNA. Therefore, to verify the predictions of this thesis, explicit consideration of the intrastrand cross-link surroundings is required. Nevertheless, the information gained offers insight into the formation preferences of U-containing intrastrand cross-links, and may also be useful for explaining the general sequence preferences of DNA intrastrand cross-link formation regardless of their method of generation or type of DNA used.

6.2. Future Research

6.2.1. A Large-Model Study of Intrastrand Cross-link Formation

This thesis characterized intrastrand cross-link formation pathways using a dinucleoside monophosphate model. Comparisons of the cross-link structures obtained using the dinucleoside monophosphate model to those obtained from MD simulations reveal structural differences, which indicate the DNA surroundings affect the conformation of the cross-link. While some structural distortions were captured by the dinucleoside monophosphate model, a more accurate depiction of cross-link formation would be afforded by a model capable of explicit inclusion of the cross-link surroundings. Under these conditions, it is expected distortions incurred along the formation pathway will be properly described and alterations to non-covalent interactions between nucleobases (both intra- and interstrand stacking interactions, as well as Watson-Crick hydrogen bonding) will be taken into account. Although some of these effects were mentioned in this thesis, it is difficult to predict how the surroundings will affect the calculated reaction pathways. For example, backbone flexibility is known to facilitate conformational changes within DNA through low transitional pathways involving many minima,[217] and therefore it is unclear

whether unwinding will affect the barriers for cross-link formation calculated in the present work. Alternatively, disruptions to base–base non-covalent interactions may be associated with large energetic costs. Overall, explicit inclusion of the DNA environment along the reaction pathway will provide further insight into the preferences dictating cross-link formation. In order to explicitly include the surrounding DNA helix, as well as solvent and ions, hybrid methods, such as QM/MM, must be used. This approach allows for large models to be partitioned such that the region in which the reaction is occurring can be examined using accurate, high-level methods, while low-level methods are applied to the surroundings. Indeed, this thesis has already established the foundation for such a study by identifying an accurate computational method to describe the high-level region (Chapter 2).

6.2.2. Other Products from the Uracil-5-yl Radical

Given the contrasting sequence preferences of UV-induced strand breaks (5′–A) and intrastrand cross-links (5′–G) generated from U•,[57,59,61,66-68] these may be competing reactions. The work presented in this thesis has fully determined the reaction pathways for intrastrand cross-link formation. Although the first step of strand break formation (hydrogen-abstraction) has been computationally studied in the 5′–A^{Br}U–3′ sequence,[88,89] the remainder of the reaction pathway has not been examined and there are many steps involved in the formation of these lesions (Figure 1–8). Furthermore, the hydrogen-abstraction step has not been examined for the 5′–G^{Br}U–3′ sequence. Therefore, the full reaction pathway for strand break formation should be determined and compared to that of the intrastrand cross-links to help understand the origin of these sequence specificities.

Interstrand cross-links resulting from U• have been detected in mismatched regions of synthetic DNA.[50,63,65,77,80] Despite the detection of several unique interstrand cross-links and high sequence selectivity, their composition and connectivity is currently unknown.

Interstrand cross-links are particularly detrimental to a cell due to the inability of the helix to separate during essential processes like replication and transcription. Furthermore, these lesions are complex to repair and require the coordination of multiple repair pathways.[218] Therefore, interstrand cross-links are particularly cytotoxic and a common DNA-damaging approach for many anti-tumour therapies.[33] More experimental work must be done to identify the composition and connectivity of these cross-links within synthetic DNA oligonucleotides, after which the reaction mechanism for formation may be examined computationally. Studying the formation of these lesions will be complicated by the experimentally-observed radical transfer from U$^{\bullet}$ to other nucleobases in the mismatch region, which means these reactions may not be directly initiated by U$^{\bullet}$. Furthermore, experimental studies should be extended to identifying these products in ^{X}U-containing cellular DNA. This is of particular interest for these lesions since formation is limited to mismatched DNA, likely in a zipper-like conformation, and therefore interstrand cross-links may not form in appreciable yields under therapeutic conditions. However, yield is not necessarily correlated to cytotoxicity and generation of only a few interstrand cross-links may be sufficient to induce cell death.

6.2.3. *Experimental Investigations of the Biological Implications of U-containing Intrastrand Cross-links*

U-containing intrastrand cross-links were first reported in 2006 following the exposure of ^{Br}U-containing DNA to UV radiation.[66-68] The results presented in thesis have offered the first suggestions that the same lesions form following the exposure of ^{X}U-containing DNA to ionizing radiation, which is relevant to the use of ^{X}U as radiosensitizing agents in anti-tumour therapies. Future experimental work must verify if U-containing intrastrand cross-links form under these conditions. This thesis has also shown that intrastrand cross-link lesions significantly affect the structure of the DNA helix.

Furthermore, related intrastrand cross-links were found to adversely affect DNA replication[104-107,110,111] and are a substrate for NER.[108,109] Therefore, subsequent experimental work should examine the structure of cross-linked DNA and the biological fate of these lesions. For example, it is important to determine how different DNA polymerases interact with these lesions in order to understand the mutagenic and cytotoxic potential of U-containing intrastrand cross-links. It is also important to understand which human polymerases are involved in TLS since this is related to the ability of a tumour cell to tolerate DNA damage and evade apoptosis.[113] Alternatively, the repair of these lesions through the NER pathway should be studied to determine whether these lesions will persist in (tumour) cells, or be repaired prior to exerting their mutagenic and/or cytotoxic effects.

6.2.4. *The Computational Investigation of Purine–Pyrimidine Intrastrand Cross-links*

By computationally studying the formation pathway of U•-initiated intrastrand cross-links, this thesis has proposed the experimentally-observed sequence selectivity is dictated by the magnitude of the barrier for the bond-formation step and the stability of the cross-linked intermediate, as well as the extent of conformational changes upon cross-link formation, which is related to the proximity of the two sites forming the bond. However, given the same formation preferences are observed for all reported purine–pyrimidine intrastrand cross-links regardless of their method of generation (UV or ionizing radiation) or type of DNA (natural or modified, cellular or synthetic, dsDNA or dinucleoside monophosphate), this proposal may be applicable to the formation of all intrastrand cross-links. To provide further support for this proposal, future computational work should examine the formation pathway of other, experimentally observed intrastrand cross-links that involve G or A and T, C or mC.[98,102,104-106] Although previous computational work investigated the formation pathways of purine–T intrastrand cross-links,[114,115] an inaccurate computational approach was used that calls into question the validity of the reported

131

results. Therefore, future investigations should employ the computational approach recommended in Chapter 2 and/or hybrid methods to study the formation mechanism of the intrastrand cross-links. Molecular dynamics simulations should also be performed in order to determine the structure of these lesions in DNA and relate these observations to the experimental studies of their biological implications (repair and replication).[104-111] This is particularly important since the work in this thesis has shown that relatively similar lesions can exhibit unique behaviours in the helix that may influence both their likelihood of formation and the biological responses they induce.

In summary, this thesis presents an extensive computational study of the experimentally-observed intrastrand cross-links initiated by U•. The first evidence has been provided to suggest U-containing intrastrand cross-links are likely to form under the (ionizing) radiation conditions used in therapeutic treatments with ^{X}U, and their resulting structures may exert both mutagenic and cytotoxic effects. In addition, a new explanation for the sequence-selective formation of intrastrand cross-links has been provided, while discrediting previous, incorrect proposals. The future work outlined above will provide further support for the proposals presented in this thesis and verify the computational findings, as well as provide an enhanced understanding of the function of therapeutics based on ^{X}U.

References

(1)   Herzberg, G. *The Spectra and Structures of Simple Free Radicals*; Dover Publications: Mineola, New York, 1971.

(2)   Droge, W. *Physiol. Rev* **2002**, *82*, 47.

(3)   Muller, F. L.; Lustgarten, M. S.; Jang, Y.; Richardson, A.; Van Remmen, H. *Free Radical Biol. Med.* **2007**, *43*, 477.

(4)   Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag New York Inc.: New York, NY, 1984.

(5)   Voet, D.; Voet, J. G. *Biochemistry*; 3 ed.; John Wiley & Sons: New York, 2004.

(6)   Schneider, B.; Neidle, S.; Berman, H. M. *Biopolymers* **1997**, *42*, 113.

(7)   Rohs, R.; Jin, X.; West, S. M.; Joshi, R.; Honig, B.; Mann, R. S. *Annu. Rev. Biochem* **2010**, *79*, 233.

(8)   Friedberg, E. C.; Walker, G. C.; Siefe, W. *DNA repair and mutagenesis*; American Society for Microbiology: Washington, DC, 1995.

(9)   Robertson, K. D.; Jones, P. A. *Carcinogenesis* **2000**, *21*, 461.

(10)   Gates, K. S. *Chem. Res. Toxicol.* **2009**, *22*, 1747.

(11)   Dahlmann, H. A.; Vaidyanathan, V. G.; Sturla, S. J. *Biochemistry* **2009**, *48*, 9347.

(12)   Cadet, J.; Douki, T.; Ravanat, J.-L. *Free Radical Biol. Med.* **2010**, *49*, 9.

(13)   Robertson, A.; Klungland, A.; Rognes, T.; Leiros, I. *Cell. Mol. Life Sci.* **2009**, *66*, 981.

(14)   Nilsen, H.; Krokan, H. E. *Carcinogenesis* **2001**, *22*, 987.

(15)   Eccles, L. J.; O'Neill, P.; Lomax, M. E. *Mutat. Res., Fundam. Mol. Mech. Mutagen.* **2011**, *711*, 134.

(16)   Sage, E.; Harrison, L. *Mutat. Res., Fundam. Mol. Mech. Mutagen.* **2011**, *711*, 123.

(17)   Markovitsi, D.; Gustavsson, T.; Banyasz, A. *Mutat. Res., Rev. Mutat. Res.* **2010**, *704*, 21.

(18)   Buchvarov, I.; Wang, Q.; Raytchev, M.; Trifonov, A.; Fiebig, T. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 4794.

(19)   Fiebig, T. *J. Phys. Chem. B* **2009**, *113*, 9348.

(20)   Wetmore, S. D.; Boyd, R. J.; Eriksson, L. A. *Chem. Phys. Lett.* **2000**, *322*, 129.

(21)   Duncan Lyngdoh, R. H.; Schaefer, H. F., III. *Acc. Chem. Res.* **2009**, *42*, 563.

(22)   Gustavsson, T.; Improta, R.; Markovitsi, D. *J. Phys. Chem. Lett.* **2010**, *1*, 2025.

(23)   Middleton, C. T.; de La Harpe, K.; Su, C.; Law, Y. K.; Crespo-Hernandez, C. E.; Kohler, B. *Annu. Rev. Phys. Chem.* **2009**, *60*, 217.

(24)   Blancafort, L.; Bertran, J.; Sodupe, M. *J. Am. Chem. Soc.* **2004**, *126*, 12770.

(25)   Hariharan, M.; McCullagh, M.; Schatz, G. C.; Lewis, F. D. *J. Am. Chem. Soc.* **2010**, *132*, 12856.

(26)   Sinha, R. P.; Hader, D. P. *Photochem. Photobiol. Sci.* **2002**, *1*, 225.

(27)   Pfeifer, G. P.; You, Y.-H.; Besaratinia, A. *Mutat. Res., Fundam. Mol. Mech. Mutagen.* **2005**, *571*, 19.

(28)   Ravanat, J.-L.; Douki, T.; Cadet, J. *J. Photochem. Photobiol., B* **2001**, *63*, 88.

(29)   Cadet, J.; Carell, T.; Cellai, L.; Chatgilialoglu, C.; Gimisis, T.; Miranda, M.; O'Neill, P.; Ravanat, J. L.; Robert, M. *Chimia* **2008**, *62*, 742.

(30)   Gu, J.; Xie, Y.; Schaefer, H. F., III. *Nucleic Acids Res.* **2007**, *35*, 5165.

(31)   von Sonntag, C. *Free-Radical-Induced DNA Damage and Its Repair: A Chemical Perspective*; Springer-Verlag: Berlin-Heidelberg, Germany, 2006.

(32)   Georgakilas, A. G. *Mutat. Res., Fundam. Mol. Mech. Mutagen.* **2011**, *711*, 1.

(33)   Rajski, S. R.; Williams, R. M. *Chem. Rev. (Washington, DC, U. S.)* **1998**, *98*, 2723.

(34)   Siddik, Z. H. *Oncogene* **2003**, *22*, 7265.

(35)   Roos, W. P.; Kaina, B. *Trends Mol. Med.* **2006**, *12*, 440.

(36)   Maga, G.; Hubscher, U. *Anti-Cancer Agents Med. Chem.* **2008**, *8*, 431.

(37)   Fuertes, M. A.; Castilla, J.; Alonso, C.; Perez, J. M. *Curr. Med. Chem.* **2003**, *10*, 257.

(38)   Jung, Y.; Lippard, S. J. *Chem. Rev. (Washington, DC, U. S.)* **2007**, *107*, 1387.

(39)   Todd, R. C.; Lippard, S. J. *J. Inorg. Biochem.* **2010**, *104*, 902.

(40)   Stehlikova, K.; Kostrhunova, H.; Kasparkova, J.; Brabec, V. *Nucleic Acids Res.* **2002**, *30*, 2894.

(41)   Connell, P. P.; Hellman, S. *Cancer Res.* **2009**, *69*, 383.

(42)   Radford, I. R. *Int. J. Radiat Biol.* **1985**, *48*, 45.

(43)   Hall, E. J.; Giaccia, A. J. *Radiobiology for the Radiobiologist*; 5th ed.; Lippincot Williams & Wilkins: Philadelphia, 2006.

(44)   Szybalski, W. *Cancer Chemother. Rep., Part 1* **1974**, *58*, 539.

(45)   Longley, D. B.; Harkin, D. P.; Johnston, P. G. *Nat. Rev. Cancer* **2003**, *3*, 330.

(46)   Grogan, B. C.; Parker, J. B.; Guminski, A. F.; Stivers, J. T. *Biochemistry* **2011**, *50*, 618.

(47)   Djordjevic, B.; Szybalski, W. *J. Exp. Med.* **1960**, *112*, 509.

(48)   Kinsella, T. J.; Mitchell, J. B.; Russo, A.; Morstyn, G.; Glatstein, E. *Int. J. Radiat. Oncol., Biol., Phys* **1984**, *10*, 1399.

(49)   Morris, S. M. *Mutat. Res.* **1991**, *258*, 161.

(50)   Dextraze, M.-E.; Gantchev, T.; Girouard, S.; Hunting, D. *Mutat. Res., Rev. Mutat. Res.* **2010**, *704*, 101.

(51)   Ding, H.; Greenberg, M. M. *J. Org. Chem.* **2010**, *75*, 535.

(52)   Li, Y.; Owusu, A.; Lehnert, S. *Int. J. Radiat. Oncol., Biol., Phys* **2004**, *58*, 519.

(53)   Greer, S.; Schwade, J.; Marion, H. S. *Int. J. Radiat. Oncol., Biol., Phys* **1995**, *32*, 1059.

(54)   Greer, S.; Alvarez, M.; Mas, M.; Wozniak, C.; Arnold, D.; Knapinska, A.; Norris, C.; Burk, R.; Aller, A.; Dauphinee, M. *Int. J. Radiat. Oncol., Biol., Phys* **2001**, *51*, 791.

(55)   Hazra, S.; Ort, S.; Konrad, M.; Lavie, A. *Biochemistry* **2010**, *49*, 6784.

(56)   Zimbrick, J. D.; Ward, J. F.; Myers, L. S. *Int. J. Radiat Biol.* **1969**, *16*, 525.

(57)   Cook, G. P.; Greenberg, M. M. *J. Am. Chem. Soc.* **1996**, *118*, 10025.

(58)   Cook, G. P.; Chen, T.; Koppisch, A. T.; Greenberg, M. M. *Chem. Biol. (Cambridge, MA, U.S.)* **1999**, *6*, 451.

(59)   Chen, T.; Cook, G. P.; Koppisch, A. T.; Greenberg, M. M. *J. Am. Chem. Soc.* **2000**, *122*, 3861.

(60)   Kobylecka, M.; Migani, A.; Asturiol, D.; Rak, J.; Blancafort, L. *J. Phys. Chem. A* **2009**, *113*, 5489.

(61)   Sugiyama, H.; Tsutsumi, Y.; Saito, I. *J. Am. Chem. Soc.* **1990**, *112*, 6720.

(62)   Watanabe, T.; Bando, T.; Xu, Y.; Tashiro, R.; Sugiyama, H. *J. Am. Chem. Soc.* **2004**, *127*, 44.

(63)   Cecchini, S.; Girouard, S.; Huels, M. A.; Sanche, L.; Hunting, D. J. *Biochemistry* **2005**, *44*, 1932.

(64) Watanabe, T.; Tashiro, R.; Sugiyama, H. *J. Am. Chem. Soc.* **2007**, *129*, 8163.

(65) Dextraze, M.-E.; Cecchini, S.; Bergeron, F.; Girouard, S.; Turcotte, K.; Wagner, J. R.; Hunting, D. J. *Biochemistry* **2009**, *48*, 2005.

(66) Zeng, Y.; Wang, Y. *Nucleic Acids Res.* **2006**, *34*, 6521.

(67) Zeng, Y.; Wang, Y. *Biochemistry* **2007**, *46*, 8189.

(68) Lin, G.; Zhang, J.; Zeng, Y.; Luo, H.; Wang, Y. *Biochemistry* **2010**, *49*, 2346.

(69) Storoniak, P.; Rak, J.; Polska, K.; Blancafort, L. *J. Phys. Chem. B* **2011**, *115*, 4532.

(70) Tashiro, R.; Nakamura, K.; Sugiyama, H. *Tetrahedron Lett.* **2008**, *49*, 428.

(71) Rivera, E.; Schuler, R. H. *J. Phys. Chem.* **1983**, *87*, 3966.

(72) Wang, C.-R.; Hu, A.; Lu, Q.-B. *J. Chem. Phys.* **2006**, *124*, 241102.

(73) Wang, C.-R.; Lu, Q.-B. *Angew. Chem. Int. Ed.* **2007**, *46*, 6316.

(74) Hutchinson, F. *Q. Rev. Biophys.* **1973**, *6*, 201.

(75) Krasin, F.; Hutchinson, F. *Biophys. J.* **1978**, *24*, 645.

(76) Krasin, F.; Hutchinson, F. *Biophys. J.* **1978**, *24*, 657.

(77) Cecchini, S.; Masson, C.; La Madeleine, C.; Huels, M. A.; Sanche, L.; Wagner, J. R.; Hunting, D. J. *Biochemistry* **2005**, *44*, 16957.

(78) Cornforth, M. N.; Bedford, J. S. *Radiat. Res.* **1987**, *111*, 385.

(79) Iliakis, G.; Kurtzman, S.; Pantelias, G.; Okayasu, R. *Radiat. Res.* **1989**, *119*, 286.

(80) Dextraze, M.-E.; Wagner, J. R.; Hunting, D. J. *Biochemistry* **2007**, *46*, 9089.

(81) Sugiyama, H.; Fujimoto, K.; Saito, I. *J. Am. Chem. Soc.* **1995**, *117*, 2945.

(82) Hwang, J. T.; Tallman, K. A.; Greenberg, M. M. *Nucleic Acids Res.* **1999**, *27*, 3805.

(83) Fujimoto, K.; Ikeda, Y.; Saito, I. *Tetrahedron Lett.* **2000**, *41*, 6455.

(84) Fujimoto, K.; Ikeda, Y.; Ishihara, S.; Saito, I. *Tetrahedron Lett.* **2002**, *43*, 2243.

(85) Roupioz, Y.; Lhomme, J.; Kotera, M. *J. Am. Chem. Soc.* **2002**, *124*, 9129.

(86) Kroeger, K. M.; Jiang, Y. L.; Kow, Y. W.; Goodman, M. F.; Greenberg, M. M. *Biochemistry* **2004**, *43*, 6723.

(87) Zheng, Y.; Sheppard, T. L. *Chem. Res. Toxicol.* **2004**, *17*, 197.

(88)   Schyman, P.; Zhang, R. B.; Eriksson, L. A.; Laaksonen, A. *Phys. Chem. Chem. Phys.* **2007**, *9*, 5975.

(89)   Schyman, P.; Eriksson, L. A.; Laaksonen, A. *J. Phys. Chem. B* **2009**, *113*, 6574.

(90)   Kroeger, K. M.; Hashimoto, M.; Kow, Y. W.; Greenberg, M. M. *Biochemistry* **2003**, *42*, 2449.

(91)   Chou, S.-H.; Chin, K.-H. *J. Mol. Biol.* **2001**, *312*, 769.

(92)   Grillari, J.; Katinger, H.; Voglauer, R. *Nucleic Acids Res.* **2007**, *35*, 7566.

(93)   Hinz, J. M. *Environ. Mol. Mutagen.* **2010**, *51*, 582.

(94)   Ornstein, R. L.; Rein, R.; Breen, D. L.; Macelroy, R. D. *Biopolymers* **1978**, *17*, 2341.

(95)   Sponer, J.; Riley, K. E.; Hobza, P. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2595.

(96)   Cysewski, P. *New J. Chem.* **2009**, *33*, 1909.

(97)   Svozil, D.; Hobza, P.; Sponer, J. *J. Phys. Chem. B* **2010**, *114*, 1191.

(98)   Bellon, S.; Ravanat, J.-L.; Gasparutto, D.; Cadet, J. *Chem. Res. Toxicol.* **2002**, *15*, 598.

(99)   Zhang, Q.; Wang, Y. *J. Am. Chem. Soc.* **2003**, *125*, 12795.

(100)   Zeng, Y.; Wang, Y. *J. Am. Chem. Soc.* **2004**, *126*, 6552.

(101)   Hong, H.; Wang, Y. *J. Am. Chem. Soc.* **2005**, *127*, 13969.

(102)   Zhang, Q. B.; Wang, Y. S. *Nucleic Acids Res.* **2005**, *33*, 1593.

(103)   Kanvah, S.; Joseph, J.; Schuster, G. B.; Barnett, R. N.; Cleveland, C. L.; Landman, U. *Acc. Chem. Res.* **2009**, *43*, 280.

(104)   Gu, C.; Wang, Y. *Biochemistry* **2004**, *43*, 6745.

(105)   Bellon, S.; Gasparutto, D.; Saint-Pierre, C.; Cadet, J. *Org. Biomol. Chem.* **2006**, *4*, 3831.

(106)   Jiang, Y.; Hong, H.; Cao, H.; Wang, Y. *Biochemistry* **2007**, *46*, 12757.

(107)   Gu, C.; Wang, Y. *Biochemistry* **2005**, *44*, 8883.

(108)   Yang, Z.; Colis, L. C.; Basu, A. K.; Zou, Y. *Chem. Res. Toxicol.* **2005**, *18*, 1339.

(109)   Gu, C.; Zhang, Q.; Yang, Z.; Wang, Y.; Zou, Y.; Wang, Y. *Biochemistry* **2006**, *45*, 10739.

(110)   Hong, H.; Cao, H.; Wang, Y. *Nucleic Acids Res.* **2007**, *35*, 7118.

(111)   Colis, L. C.; Raychaudhury, P.; Basu, A. K. *Biochemistry* **2008**, *47*, 8070.

(112) Raychaudhury, P.; Basu, A. K. *Biochemistry* **2011**, *50*, 2330.

(113) Lange, S. S.; Takata, K.; Wood, R. D. *Nat. Rev. Cancer* **2011**, *11*, 96.

(114) Xerri, B.; Morell, C.; Grand, A.; Cadet, J.; Cimino, P.; Barone, V. *Org. Biomol. Chem.* **2006**, *4*, 3986.

(115) Labet, V.; Morell, C.; Grand, A.; Cadet, J.; Cimino, P.; Barone, V. *Org. Biomol. Chem.* **2008**, *6*, 3300.

(116) Wojcik, A.; Bochenek, A.; Lankoff, A.; Lisowska, H.; Padjas, A.; Szumiel, I.; von Sonntag, C.; Obe, G. *J. Photochem. Photobiol., B* **2006**, *84*, 15.

(117) Durbeej, B.; Eriksson, L. A. *J. Photochem. Photobiol., A* **2002**, *152*, 95.

(118) Zhang, R. B.; Eriksson, L. A. *Chem. Phys. Lett.* **2006**, *417*, 303.

(119) Zhang, R. B.; Eriksson, L. A. *J. Phys. Chem. B* **2006**, *110*, 7556.

(120) Schyman, P.; Eriksson, L. A.; Zhang, R. B.; Laaksonen, A. *Chem. Phys. Lett.* **2008**, *458*, 186.

(121) Zhao, Y.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2701.

(122) Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2007**, *9*, 5291.

(123) Grimme, S.; Antony, J.; Schwabe, T.; Muck-Lichtenfeld, C. *Org. Biomol. Chem.* **2007**, *5*, 741.

(124) Foloppe, N.; MacKerell, A. D., Jr. *J. Phys. Chem. B* **1999**, *103*, 10955.

(125) Foloppe, N.; Nilsson, L.; MacKerell, A. D., Jr. *Biopolymers* **2001**, *61*, 61.

(126) Shishkin, O. V.; Pelmenschikov, A.; Hovorun, D. M.; Leszczynski, J. *J. Mol. Struct.* **2000**, *526*, 329.

(127) Shishkin, O. V.; Gorb, L.; Zhikol, O. A.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **2004**, *22*, 227.

(128) Shishkin, O. V.; Gorb, L.; Zhikol, O. A.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **2004**, *21*, 537.

(129) Gorb, L.; Shishkin, O.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **2005**, *22*, 441.

(130) Petrov, A. S.; Funseth-Smotzer, J.; Pack, G. R. *Int. J. Quantum Chem* **2005**, *102*, 645.

(131) Zhang, R. B.; Eriksson, L. A. *J. Phys. Chem. B* **2006**, *110*, 23583.

(132) Kosenkov, D.; Gorb, L.; Shishkin, O. V.; Sponer, J.; Leszczynski, J. *J. Phys. Chem. B* **2008**, *112*, 150.

(133) Poltev, V. I.; Anisimov, V. M.; Danilov, V. I.; Deriabina, A.; Gonzalez, E.; Jurkiewiez, A.; Les, A.; Polteva, N. *J. Biomol. Struct. Dyn.* **2008**, *25*, 563.

(134) Svozil, D.; Sponer, J. E.; Marchan, I.; Perez, A.; Cheatham, T. E., III; Forti, F.; Luque, F. J.; Orozco, M.; Sponer, J. *J. Phys. Chem. B* **2008**, *112*, 8188.

(135) Palamarchuk, G. V.; Shishkin, O. V.; Gorb, L.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **2009**, *26*, 653.

(136) MacKerell, A. D., Jr. *J. Phys. Chem. B* **2009**, *113*, 3235.

(137) Millen, A. L.; Manderville, R. A.; Wetmore, S. D. *J. Phys. Chem. B* **2010**, *114*, 4373.

(138) Leulliot, N.; Ghomi, M.; Scalmani, G.; Berthier, G. *J. Phys. Chem. A* **1999**, *103*, 8716.

(139) Rulisek, L.; Sponer, J. *J. Phys. Chem. B* **2003**, *107*, 1913.

(140) Gresh, N.; Sponer, J. E.; Spackova, N.; Leszczynski, J.; Sponer, J. *J. Phys. Chem. B* **2003**, *107*, 8669.

(141) Herbert, H. E.; Halls, M. D.; Hratchian, H. P.; Raghavachari, K. *J. Phys. Chem. B* **2006**, *110*, 3336.

(142) Delaney, J. C.; Essigmann, J. M. *Chem. Res. Toxicol.* **2008**, *21*, 232.

(143) Dhaouadi, Z.; Nsangou, M.; Hernandez, B.; Pfluger, F.; Liquier, J.; Ghomi, M. *Spectrochim. Acta, Part A* **2009**, *73*, 805.

(144) Tehrani, Z. A.; Fattahi, A.; Pourjavadi, A. *THEOCHEM* **2009**, *913*, 117.

(145) Poltev, V. I.; Anisimov, V. M.; Danilov, V. I.; Deriabina, A.; Gonzalez, E.; Garcia, D.; Rivas, F.; Jurkiewicz, A.; Les, A.; Polteva, N. *THEOCHEM* **2009**, *912*, 53.

(146) Poltev, V. I.; Anisimov, V. M.; Danilov, V. I.; Mourik, T. v.; Deriabina, A.; González, E.; Padua, M.; Garcia, D.; Rivas, F.; Polteva, N. *Int. J. Quantum Chem* **2010**, *110*, 2548.

(147) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*; 2nd ed.; Wiley-VCH: Toronto, 2001.

(148) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Phys. Chem. A* **2007**, *111*, 10439.

(149) Rayat, S.; Glaser, R. *J. Org. Chem.* **2003**, *68*, 9882.

(150) Murray, V.; Martin, R. F. *Nucleic Acids Res.* **1989**, *17*, 2675.

(151) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6908.

(152) Zhao, Y.; Tishchenko, O.; Truhlar, D. G. *J. Phys. Chem. B* **2005**, *109*, 19046.

(153) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157.

(154) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 1849.

(155) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.

(156) Hohenstein, E. G.; Chill, S. T.; Sherrill, C. D. *J. Chem. Theory Comput.* **2008**, *4*, 1996.

(157) Barrios, R.; Skurski, P.; Simons, J. *J. Phys. Chem. B* **2002**, *106*, 7991.

(158) Berdys, J.; Anusiewicz, I.; Skurski, P.; Simons, J. *J. Phys. Chem. A* **2004**, *108*, 2999.

(159) Box, H. C.; Dawidzik, J. B.; Budzinski, E. E. *Free Radical Biol. Med.* **2001**, *31*, 856.

(160) Cadet, J.; Douki, T.; Gasparutto, D.; Ravanat, J.-L. *Mutat. Res.-Fundam. Mol. Mech. Mutagen.* **2003**, *531*, 5.

(161) Cadet, J.; Douki, T.; Gasparutto, D.; Ravanat, J.-L. *Radiat. Phys. Chem.* **2005**, *72*, 293.

(162) Wang, Y. *Chem. Res. Toxicol.* **2008**, *21*, 276.

(163) HyperChem; version 5.1 ed.; Hypercube, Inc.: Gainesville, Florida 32601, USA, 1996.

(164) Sponer, J.; Leszczynski, J.; Hobza, P. *Biopolymers* **2001**, *61*, 3.

(165) Scalmani, G.; Frisch, M. J. *J. Chem. Phys.* **2010**, *132*, 114110(15).

(166) Arora, N.; Jayaram, B. *J. Phys. Chem. B* **1998**, *102*, 6139.

(167) Barbe, S.; Le Bret, M. *J. Phys. Chem. A* **2008**, *112*, 989.

(168) Young, M. A.; Jayaram, B.; Beveridge, D. L. *J. Phys. Chem. B* **1998**, *102*, 7666.

(169) Okamoto, A.; Tainaka, K.; Saito, I. *Bioconjugate Chem.* **2005**, *16*, 1105.

(170) Sinkeldam, R. W.; Tor, Y. *Org. Biomol. Chem.* **2007**, *5*, 2523.

(171) Sinkeldam, R. W.; Greco, N. J.; Tor, Y. *ChemBioChem* **2008**, *9*, 706.

(172) Pavanello, M.; Adamowicz, L.; Volobuyev, M.; Mennucci, B. *J. Phys. Chem. B* **2010**, *114*, 4416.

(173) Santoro, F.; Barone, V.; Improta, R. *J. Am. Chem. Soc.* **2009**, *131*, 15232.

(174) Johnson, E. R.; Mackie, I. D.; DiLabio, G. A. *J. Phys. Org. Chem.* **2009**, *22*, 1127.

(175)   Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Jr., J. A. M.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; Revision A.02 ed.; Gaussian, Inc.: Wallingford CT, 2009.

(176)   Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; van de Streek, J.; Wood, P. A. *J. Appl. Cryst* **2008**, *41*, 466.

(177)   Rabuck, A. D.; Scuseria, G. E. *Theor. Chem. Acc.* **2000**, *104*, 439.

(178)   Ban, F. Q.; Rankin, K. N.; Gauld, J. W.; Boyd, R. J. *Theor. Chem. Acc.* **2002**, *108*, 1.

(179)   Improta, R.; Barone, V. *Chem. Rev. (Washington, DC, U. S.)* **2004**, *104*, 1231.

(180)   Hocquet, A.; Leulliot, N.; Ghomi, M. *J. Phys. Chem. B* **2000**, *104*, 4560.

(181)   Yurenko, Y. P.; Zhurakivsky, R. O.; Ghomi, M.; Samijlenko, S. P.; Hovorun, D. M. *J. Phys. Chem. B* **2007**, *111*, 9655.

(182)   Gu, J.; Leszczynski, J. *Chem. Phys. Lett.* **2001**, *335*, 465.

(183)   Sharma, P.; Sponer, J. E.; Sponer, J.; Sharma, S.; Bhattacharyya, D.; Mitra, A. *J. Phys. Chem. B* **2010**, *114*, 3307.

(184)   Stepanian, S. G.; Karachevtsev, M. V.; Glamazda, A. Y.; Karachevtsev, V. A.; Adamowicz, L. *Chem. Phys. Lett.* **2008**, *459*, 153.

(185)   Rutledge, L. R.; Wetmore, S. D. *Can. J. Chem.* **2010**, *88*, 815.

(186)   Gil, A.; Branchadell, V.; Bertran, J.; Oliva, A. *J. Phys. Chem. B* **2007**, *111*, 9372.

(187)   Ruiz, R.; Garcia, B.; Ruisi, G.; Silvestri, A.; Barone, G. *THEOCHEM* **2009**, *915*, 86.

(188)   Ribeiro, A. J. M.; Ramos, M. J.; Fernandes, P. A. *J. Chem. Theory Comput.* **2010**, *6*, 2281.

(189)   Suresh, C. H.; Mohan, N.; Vijayalakshmi, K. P.; George, R.; Mathew, J. M. *J. Comput. Chem.* **2009**, *30*, 1392.

(190)   Gu, J.; Wang, J.; Leszczynski, J.; Xie, Y.; Schaefer, H. F., III. *Chem. Phys. Lett.* **2008**, *459*, 164.

(191) Pitonak, M.; Riley, K. E.; Neogrady, P.; Hobza, P. *ChemPhysChem* **2008**, *9*, 1636.

(192) Sherrill, C. D.; Takatani, T.; Hohenstein, E. G. *J. Phys. Chem. A* **2009**, *113*, 10146.

(193) Morgado, C. A.; Jurecka, P.; Svozil, D.; Hobza, P.; Sponer, J. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3522.

(194) Riley, K. E.; Pitonak, M.; Cerny, J.; Hobza, P. *J. Chem. Theory Comput.* **2010**, *6*, 66.

(195) Jacquemin, D.; Perpete, E. A.; Ciofini, I.; Adamo, C. *J. Chem. Theory Comput.* **2010**, *6*, 1532.

(196) Jacquemin, D.; Perpete, E. A.; Ciofini, I.; Adamo, C.; Valero, R.; Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2010**, *6*, 2071.

(197) Tognetti, V.; Cortona, P.; Adamo, C. *Int. J. Quantum Chem* **2010**, *110*, 2320.

(198) Johnson, E. R.; Becke, A. D.; Sherrill, C. D.; DiLabio, G. A. *J. Chem. Phys.* **2009**, *131*.

(199) Merrick, J. P.; Moran, D.; Radom, L. *J. Phys. Chem. A* **2007**, *111*, 11683.

(200) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 6378.

(201) Krieger, E.; Darden, T.; Nabuurs, S. B.; Finkelstein, A.; Vriend, G. *Proteins: Struct., Funct., Bioinf.* **2004**, *57*, 678.

(202) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvai, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A.; University of California, San Francisco: 2010.

(203) Cheatham, T. E., III; Case, D. A. In *Computational Studies of RNA and DNA*; Sponer, J., Lankas, F., Eds.; Springer: Dordrecht, The Netherlands, 2006; Vol. 2, p 45.

(204) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999.

(205) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157.

(206) Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K. *Nucleic Acids Res.* **2009**, *37*, 5917.

(207) Vaya, I.; Gustavsson, T.; Miannay, F.-A.; Douki, T.; Markovitsi, D. *J. Am. Chem. Soc.* **2010**, *132*, 11834.

(208) Yashchuk, V.; Kudrya, V.; Losytskyy, M.; Suga, H.; Ohul'chanskyy, T. *J. Mol. Liq.* **2006**, *127*, 79.

(209) Hong, H.; Cao, H.; Wang, Y.; Wang, Y. *Chem. Res. Toxicol.* **2006**, *19*, 614.

(210) Stivers, J. T. *Chem. Eur. J.* **2008**, *14*, 786.

(211) Berdis, A. J. *Biochemistry* **2008**, *47*, 8253.

(212) Rink, S. M.; Hopkins, P. B. *Biochemistry* **1995**, *34*, 1439.

(213) Rink, S. M.; Lipman, R.; Alley, S. C.; Hopkins, P. B.; Tomasz, M. *Chem. Res. Toxicol.* **1996**, *9*, 382.

(214) McConnell, T. L.; Wetmore, S. D. *J. Phys. Chem. B* **2007**, *111*, 2999.

(215) Jasti, V. P.; Das, R. S.; Hilton, B. A.; Weerasooriya, S.; Zou, Y.; Basu, A. K. *Biochemistry* **2011**, *50*, 3862.

(216) Munzel, M.; Szeibert, C.; Glas, A. F.; Globisch, D.; Carell, T. *J. Am. Chem. Soc.* **2011**, *133*, 5186.

(217) Sims, G. E.; Kim, S. H. *Nucleic Acids Res.* **2003**, *31*, 5607.

(218) Christmann, M.; Tomicic, M. T.; Roos, W. P.; Kaina, B. *Toxicology* **2003**, *193*, 3.

Appendix A

Appendix A is included in the electronic information.

Appendix B

Appendix B is included in the electronic information.

Appendix C

Appendix C is included in the electronic information.

Appendix D

Appendix D is included in the electronic information.

Appendix A

Table of Contents

	Page
Optimized Structures for Modified Sequences	

Figure A–1: B3LYP structures of the 5′–GU–3′ dinucleoside monophosphate sequence optimized in the gas phase (a–c) and water (d–f) using the anionic (a,d), neutral (b,e) and counterion (c,f) phosphate models.
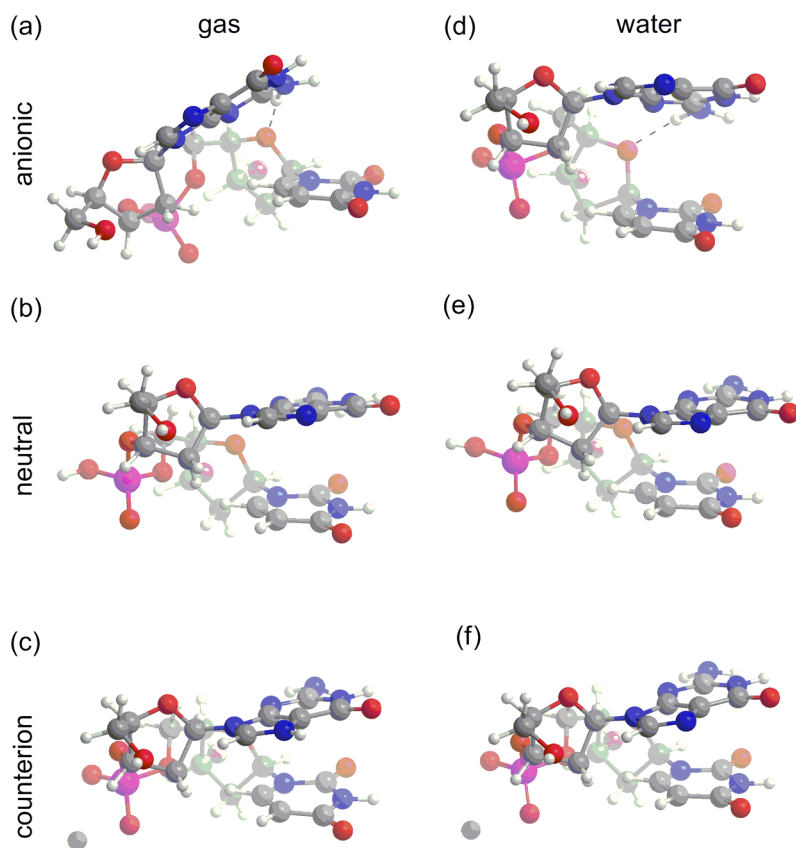
Figure A–2: MPWB1K structures of the 5′–GU–3′ dinucleoside monophosphate sequence optimized in the gas phase (a–c) and water (d–f) using the anionic (a,d), neutral (b,e) and counterion (c,f) phosphate models.
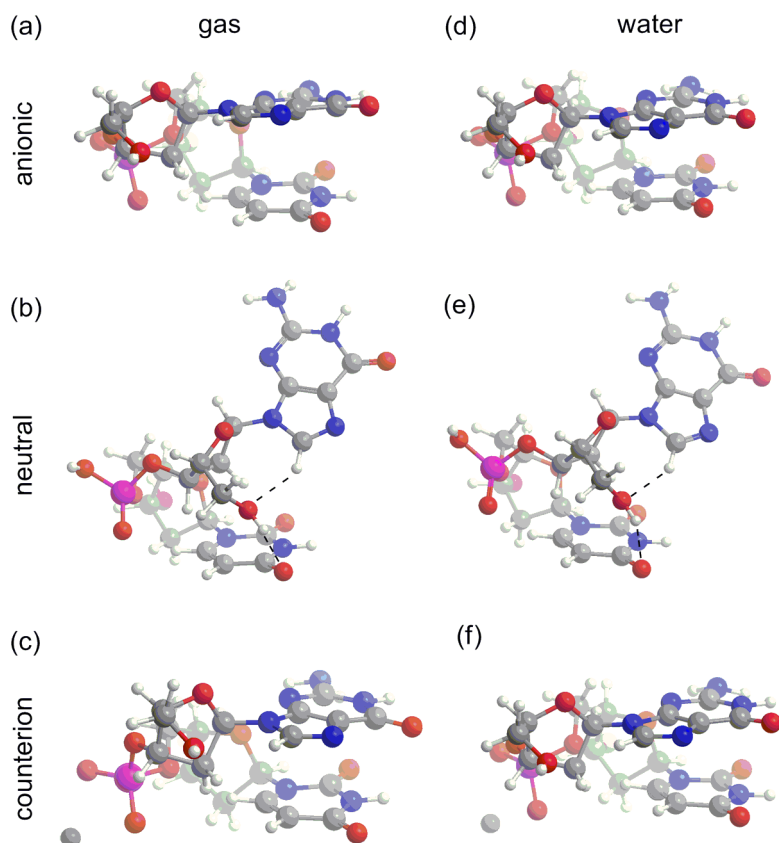
Figure A–3: M06-2X structures of the 5′–GU–3′ dinucleoside monophosphate sequence optimized in the gas phase (a–c) and water (d–f) using the anionic (a,d), neutral (b,e) and counterion (c,f) phosphate models.
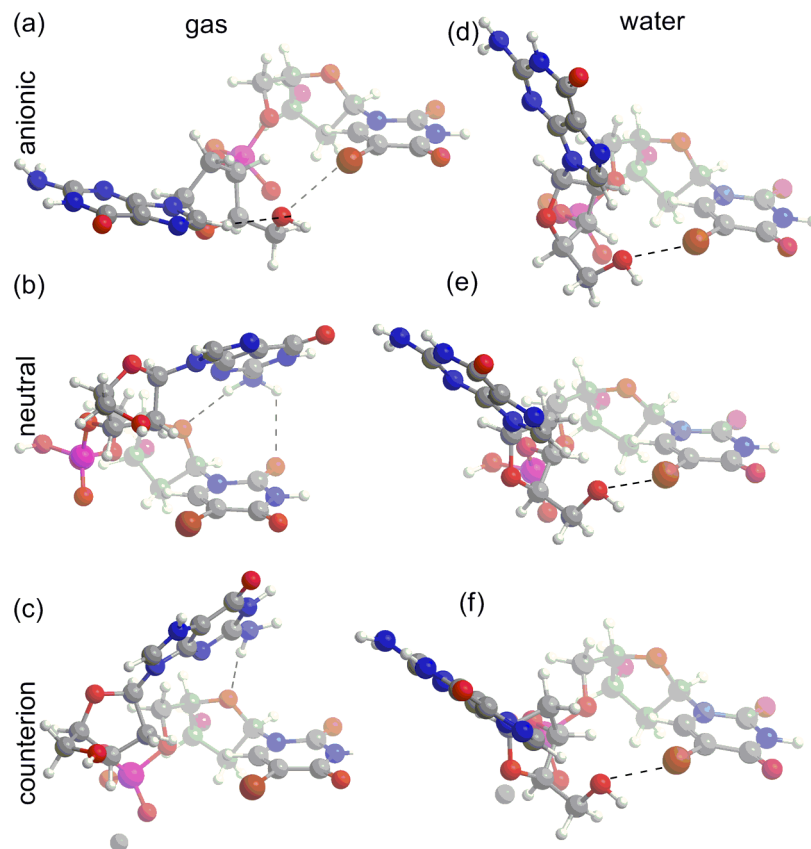
Figure A–4: B3LYP structures of the 5′–G^{Br}U–3′ dinucleoside monophosphate sequence optimized in the gas phase (a–c) and water (d–f) using the anionic (a,d), neutral (b,e) and counterion (c,f) phosphate models.
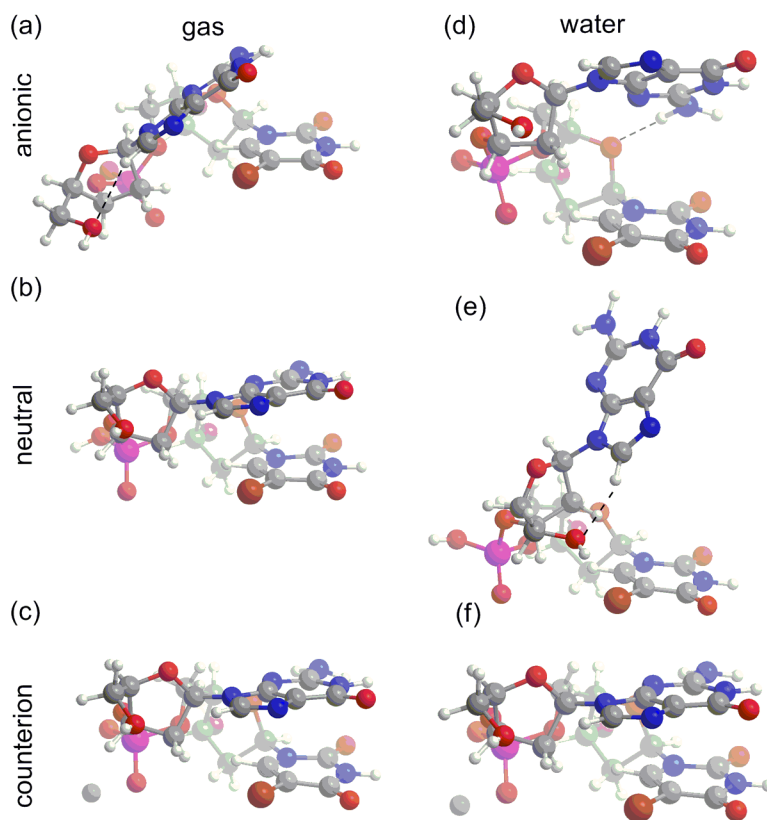
Figure A–5: MPWB1K structures of the 5′–G^{Br}U–3′ dinucleoside monophosphate sequence optimized in the gas phase (a–c) and water (d–f) using the anionic (a,d), neutral (b,e) and counterion (c,f) phosphate models.
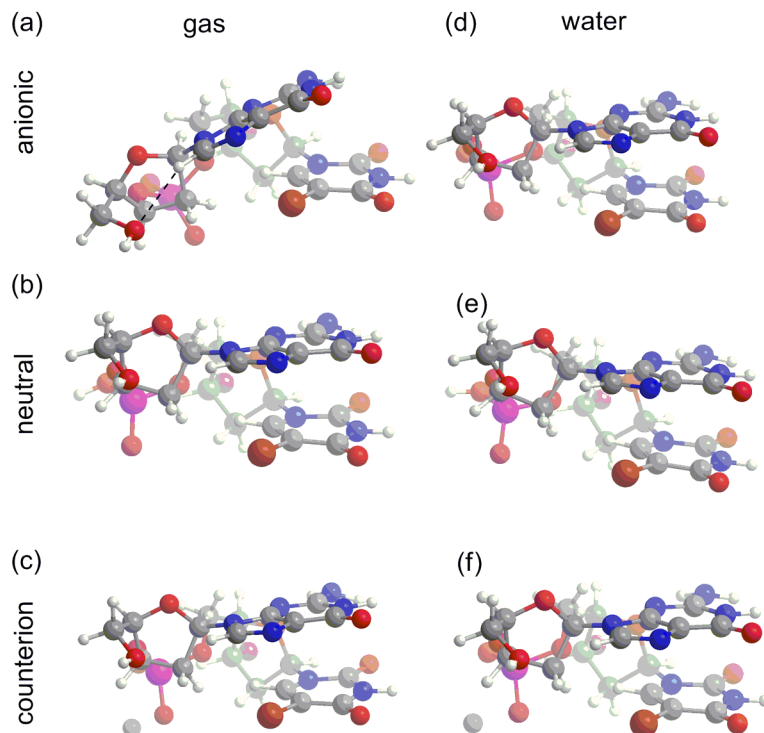
Figure A–6: M06-2X structures of the 5′–G^{Br}U–3′ dinucleoside monophosphate sequence optimized in the gas phase (a–c) and water (d–f) using the anionic (a,d), neutral (b,e) and counterion (c,f) phosphate models.

Table A–1: Backbone torsion angles (^{0}) and pseudorotation phase angles (P, ^{0}) of the 5′–GT–3′sequence calculated in the gas-phase and water, as well as the average value (Exp.) and standard deviation (SD) obtained from experiment.

Gas	Anionic			Neutral			Counterion				
∠[a]	B3LYP	MPWB1K	M06-2X	B3LYP	MPWB1K	M06-2X	B3LYP	MPWB1K[dc]	M06-2X	Exp.[b]	SD[b]
χ5′	232.0	201.3	210.3	195.3	237.9	244.9	200.7	236.7	236.8	258	14
δ5′	145.4	147.5	148.1	143.4	147.0	146.1	143.9	146.8	147.2	128	13
ε	201.8	167.5	161.9	193.7	197.6	224.1	182.2	160.1	170.6	184	11
ζ	277.6	278.3	279.0	261.7	173.9	275.7	278.8	272.7	264.9	265	10
α	261.0	268.6	271.3	314.8	23.2	310.4	295.4	285.5	295.1	298	15
β	260.0	247.5	250.5	185.8	84.5	152.6	229.3	221.6	197.2	176	9
γ	64.6	59.8	56.1	55.3	51.3	56.8	51.9	49.7	54.5	48	11
χ3′	251.3	273.7	277.9	262.4	250.6	268.5	240.3	255.0	257.9	241	8
δ3′	145.4	153.7	155.4	148.3	126.8	151.9	145.7	148.1	149.8	128	13
P5′	172.200	164.8	165.7	160.8	173.5	174.3	158.7	165.7	165.9	144–190[c]	
P3′	171.1	189.5	199.5	171.9	127.3	170.7	166.64	164.3	171.2	144–190[c]	

Water	Anionic			Neutral			Counterion				
∠[a]	B3LYP	MPWB1K	M06-2X	B3LYP	MPWB1K	M06-2X	B3LYP	MPWB1K[d]	M06-2X	Exp.[b]	SD[b]
χ5′	237.3	210.4	227.2	238.1	241.2	241.3	236.8	234.5	230.0	258	14
δ5′	147.0	150.1	149.8	146.2	147.2	145.7	147.2	150.8	149.8	128	13
ε	203.3	162.8	158.0	220.7	187.1	217.4	203.8	172.0	155.8	184	11
ζ	279.5	277.0	276.8	264.4	218.0	280.2	281.0	266.7	277.2	265	10
α	284.3	271.5	274.6	315.5	323.8	304.1	281.3	284.4	274.4	298	15
β	242.3	244.3	240.7	157.0	150.3	159.2	237.8	209.7	236.8	176	9
γ	59.8	60.5	48.0	58.2	60.3	57.4	54.7	50.6	50.2	48	11
χ3′	242.6	273.0	265.0	246.8	249.0	273.8	233.7	248.6	263.8	241	8
δ3′	146.8	155.1	154.0	143.7	147.0	154.5	147.1	146.6	152.5	128	13
P5′	177.8	170.1	170.6	177.6	176.8	174.8	178.4	170.8	172.6	144–190[c]	
P3′	170.1	184.1	178.7	159.1	158.3	179.9	167.1	157.7	172.8	144–190[c]	

[a]See Figures 2–2 and 2–3 for definitions of backbone torsion angles and pseudorotation phase angle, respectively. [b]See Ref. i. [c]See Ref. ii. [d]Due to SCF convergence issues, structures were obtained using M06-2X-optmized structures as input rather than the standard HyperChem-generated structures.

[i] Schneider, B.; Neidle, S.; Berman, H. M. *Biopolymers* **1997**, *42*, 113.

[ii] Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag New York Inc.: New York, NY, 1984.

Table A–2: Backbone torsion angles (^{0}) and pseudorotation phase angles (P, ^{0}) of the 5′–GU–3′ sequence calculated in the gas-phase and water, as well as the average value (Exp.) and standard deviation (SD) obtained from experiment.

Gas	Anionic			Neutral			Counterion				
∠[a]	B3LYP	MPWB1K	M06-2X	B3LYP	MPWB1K	M06-2X	B3LYP	MPWB1K[d]	M06-2X	Exp.[b]	SD[b]
χ5′	232.2	209.1	217.3	241.5	236.8	58.4	201.3	237.0	239.0	258	14
δ5′	146.0	149.6	147.7	147.3	149.0	144.0	144.2	147.8	149.3	128	13
ε	206.9	165.8	157.2	195.2	173.0	223.2	182.1	168.4	170.1	184	11
ζ	280.1	274.7	278.1	181.6	253.2	275.9	278.9	266.0	265.5	265	10
α	250.1	264.8	265.9	27.4	304.0	309.8	294.8	291.3	291.2	298	15
β	264.2	254.3	253.0	88.1	190.7	154.8	230.2	206.5	199.5	176	9
γ	64.2	54.2	53.7	51.2	53.5	57.2	51.3	51.8	52.5	48	11
χ3′	252.4	232.6	277.4	246.5	250.6	269.8	237.4	247.6	254.8	241	8
δ3′	146.6	151.8	155.6	119.1	147.7	152.3	146.0	149.2	149.7	128	13
P5′	172.7	164.5	164.4	180.2	168.6	173.1	159.1	169.0	169.2	140–190[c]	
P3′	170.8	170.3	196.6	121.2	165.4	172.7	166.5	168.0	172.5	144–190[c]	

Water	Anionic			Neutral			Counterion				
∠[a]	B3LYP	MPWB1K	M06-2X	B3LYP	MPWB1K	M06-2X	B3LYP	MPWB1K	M06-2X	Exp.[b]	SD[b]
χ5′	239.6	216.3	226.7	237.8	250.7	242.8	237.2	237.1	229.6	258	14
δ5′	147.7	150.8	149.9	212.8	149.9	145.9	146.7	151.3	149.6	128	13
ε	206.8	161.3	154.8	219.2	171.4	216.5	199.3	171.7	155.8	184	11
ζ	284.0	277.2	277.6	263.3	249.9	280.0	278.3	265.0	277.1	265	10
α	276.9	269.0	267.1	315.0	179.7	304.2	284.4	287.2	272.0	298	15
β	243.4	245.6	246.5	162.1	181.4	160.0	238.0	205.6	236.2	176	9
γ	55.2	57.7	52.8	58.0	55.1	57.2	54.0	53.3	51.7	48	11
χ3′	234.9	273.7	274.2	243.2	246.6	275.6	233.7	245.7	259.8	241	8
δ3′	148.3	155.4	155.8	144.3	147.6	154.9	146.8	149.5	152.5	128	13
P5′	178.7	171.9	171.7	178.0	180.0	174.0	178.1	173.5	172.8	144–190[c]	
P3′	170.2	187.9	189.2	160.0	160.1	180.7	166.9	167.9	176.8	144–190[c]	

[a]See Figures 2–2 and 2–3 for definitions of backbone torsion angles and pseudorotation phase angle, respectively. [b]See Ref. i. [c]See Ref. ii. [d]Due to SCF convergence issues, structures were obtained using M06-2X-optmized structures as input rather than the standard HyperChem-generated structures.

Table A–3: Backbone torsion angles (^{0}) and pseudorotation phase angles (P, ^{0}) of the 5′–G^{Br}U–3′sequence calculated in the gas-phase and water, as well as the average value (Exp.) and standard deviation (SD) obtained from experiment.

Gas	Anionic			Neutral			Counterion				
∠[a]	B3LYP	MPWB1K	M06-2X	B3LYP	MPWB1K	M06-2X	B3LYP	MPWB1K	M06-2X	Exp.[b]	SD[b]
χ5′	230.5	212.1	210.7	196.8	233.6	231.2	206.8	229.2	228.5	258	14
δ5′	144.5	150.7	150.6	144.2	149.2	149.6	145.3	146.4	147.0	128	13
ε	278.0	165.8	162.4	195.1	160.8	158.2	179.8	160.3	154.3	184	11
ζ	286.7	275.7	278.0	265.0	266.1	269.2	276.9	273.1	278.1	265	10
α	251.1	268.9	267.0	308.6	293.8	292.2	293.4	283.7	280.0	298	15
β	260.9	253.4	256.7	188.1	221.2	226.5	230.1	230.6	239.0	176	9
γ	57.6	52.4	52.1	52.1	45.3	44.4	51.8	47.1	47.0	48	11
χ3′	238.8	233.0	240.6	261.3	256.5	258.6	238.6	257.7	260.4	241	8
δ3′	148.5	152.5	154.7	148.3	150.6	151.7	147.1	151.1	153.1	128	13
P5′	169.5	166.6	167.8	162.3	167.4	168.4	160.9	162.9	165.9	144–190[c]	
P3′	177.3	170.3	173.9	172.7	168.3	171.7	172.2	171.9	175.3	144–190[c]	

Water	Anionic			Neutral			Counterion				
∠[a]	B3LYP	MPWB1K	M06-2X	B3LYP	MPWB1K	M06-2X	B3LYP	MPWB1K[d]	M06-2X	Exp.[b]	SD[b]
χ5′	238.0	202.1	227.3	237.5	236.7	232.9	237.5	229.8	227.7	258	14
δ5′	145.6	151.7	150.2	147.0	146.1	151.8	149.0	149.5	150.1	128	13
ε	219.0	166.2	202.8	276.1	215.3	161.1	275.6	162.5	157.2	184	11
ζ	298.4	279.6	276.6	270.8	273.0	265.7	286.8	271.8	276.6	265	10
α	253.4	268.6	276.4	301.6	304.5	295.1	277.6	280.4	277.6	298	15
β	249.5	245.4	242.3	192.5	176.2	217.4	222.2	227.3	238.1	176	9
γ	58.0	60.2	47.2	50.7	55.7	47.3	53.1	50.2	48.0	48	11
χ3′	238.7	272.7	260.3	225.6	237.1	253.1	231.9	252.4	258.5	241	8
δ3′	149.5	154.9	153.8	146.9	147.7	150.4	148.3	149.9	152.7	128	13
P5′	176.6	172.2	170.4	176.0	170.9	172.2	178.5	170.0	171.3	144–190[c]	
P3′	174.4	180.6	176.7	167.3	157.8	167.0	173.0	166.2	173.5	144–190[c]	

[a]See Figures 2–2 and 2–3 for definitions of backbone torsion angles and pseudorotation phase angle, respectively. [b]See Ref. i. [c]See Ref. ii. [d]Due to SCF convergence issues, structures were obtained using M06-2X-optmized structures as input rather than the standard HyperChem-generated structures.

Table A–4: Comparison of the 6-31G(d,p) and 6-31+G(d,p) backbone torsion angles (^{0}), pseudorotation phase angles (P, ^{0}) and the angle between nucleobase planes (φ, ^{0}) for the 5′–GT–3′ sequence calculated in the gas-phase and water with M06-2X for the anionic and counterion phosphate models, as well as the average value (Exp.) and standard deviation (SD) obtained from experiment.

| $\angle$[a] | Anionic | | | | Counterion | | | | | |
| | 6-31+G(d,p) | | 6-31G(d,p) | | 6-31+G(d,p) | | 6-31G(d,p) | | | |
	Gas	Water	Gas	Water	Gas	Water	Gas	Water	Exp.[b]	SD[b]
χ5′	201.7	233.9	210.7	227.3	238.3	260.5	228.5	227.7	258	14
δ5′	146.1	149.5	150.6	150.2	146.7	147.5	147.0	150.1	128	13
ε	167.2	162.6	162.4	202.8	154.4	160.3	154.3	157.2	184	11
ζ	279.4	272.6	278.0	276.6	276.7	268.3	278.1	276.6	265	10
α	272.1	283.5	267.0	276.4	283.1	300.1	280.0	277.6	298	15
β	249.4	223.5	256.7	242.3	230.7	188.3	239.0	238.1	176	9
γ	56.2	49.1	52.1	47.2	46.7	55.8	47.0	48.0	48	11
χ3′	279.1	254.3	240.6	260.3	257.9	251.5	260.4	258.5	241	8
δ3′	154.2	149.4	154.7	153.8	150.2	146.0	153.1	152.7	128	13
P5′	163.1	173.2	167.8	170.4	168.4	189.3	165.9	171.3	144–190[c]	
P3′	198.5	168.9	173.9	176.7	170.9	160.9	175.3	173.5	144–190[c]	
φ	32.1	6.1	28.8	4.9	11.6	6.8	12.9	4.7	–	

[a]See Figures 2–2 and 2–3 for definitions of backbone torsion angles and pseudorotation phase angle, respectively and Section 2.3.1. for a definition of φ. [b]See Ref. i. [c]See Ref. ii.

Appendix B

Table of Contents

Figure B–1: The lowest-lying triplet surface (blue, solid line) and singlet ground state surface (green, solid line) generated as a function of d(C8–C5), (Å) (described in Section 3.3.1.2), illustrating the intersystem crossing (degeneracy) between these states that connect R_T and $I2_S$ (see Figure 3–3 in text for structures).
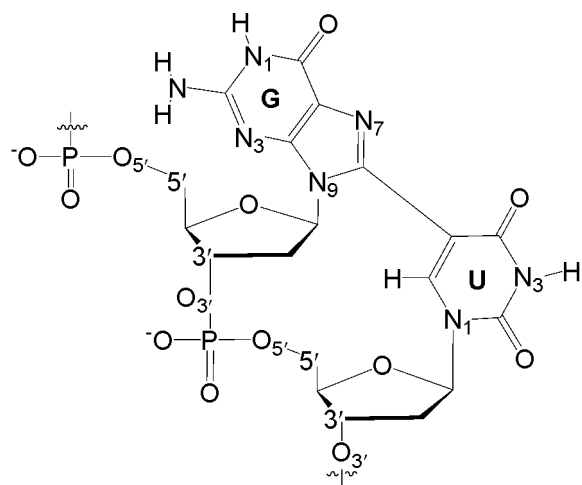
Figure B–2: Structure and chemical numbering of the cross-linked 5'–G[8–5]U–3' dinucleotide within double-stranded DNA.

Table B–1: GAFF atom types and charges assigned to the cross-linked dinucleotide. See Figure B–2 for chemical numbering.

	5'–G			3'–U	
Atom	Charges	Atom Type	Atom	Charges	Atom Type
N1	-0.4891	na	N1	-0.3546	na
H(N1)	0.3440	hn	C2	0.8077	c
C2	0.7810	cd	O2	-0.5970	o
N2	-0.9653	nn	N3	-0.4708	na
H(N2)	0.4360	hn	H(N3)	0.3376	hn
H(N2)	0.4360	hn	C4	0.7327	c
N3	-0.6569	nc	O4	-0.5652	o
C4	0.1390	cc	C5	-0.3722	cc
C5	0.1492	cd	C6	0.1287	cd
C6	0.5086	c	H(C6)	0.2093	h4
O6	-0.5722	o			
N7	-0.5879	nd			
C8	0.5206	cc			
N9	-0.1818	na			
C1'	0.0344	c3	C1'	0.0282	c3
H(C1')	0.1784	h2	H(C1')	0.1831	h2
C2'	-0.0891	c3	C2'	-0.0895	c3
H(C2')	0.0718	hc	H(C2')	0.0718	hc
H(C2')	0.0718	hc	H(C2')	0.0718	hc
C3'	0.2224	c3	C3'	0.0713	c3
O3'	-0.5211	os	O3'	-0.5286	os
H(C3')	0.0515	h1	H(C3')	0.0978	h1
C4'	0.1666	C3	C4'	0.1193	C3
H(C4')	0.0751	h1	H(C4')	0.1115	h1
O4'	-0.3718	os	O4'	-0.3733	os
C5'	0.0338	c3	C5'	0.0436	c3
H(C5')	0.0688	h1	H(C5')	0.0690	h1
H(C5')	0.0688	h1	H(C5')	0.0690	h1
O5'	-0.4974	os	O5'	-0.4051	os
O(P)	-0.7755	o	O(P)	-0.6992	o
O(P)	-0.7755	o	O(P)	-0.6992	o
P	1.1733	p5	P	0.9550	p5
Total	-0.9525			-1.0473	

Simulation Details

Simulations were performed using the AMBER03[1] forcefield and GAFF[2] parameters for the cross-linked dinucleotide (Figure B–2 and Table B–1). Cross-linked DNA was generated by overlaying the backbone of the M06-2X-optimized dinucleoside monophosphate product onto the central GT of the 5′–d(GCATGGC**GT**GCTATGC)–3′ oligonucleotide. Any close contacts or artificial overlaps resulting from the overlay process were eliminated. Subsequently, this structure was minimized in the gas-phase with the AMBER and GAFF forcefields, and the resulting structure used as input for the molecular dynamics simulation.

Simulations were performed using pre-programmed protocols (the macro md_run.mcr) in YASARA. Initializing this macro automates all procedures required to begin the simulation from the desired input structure through to the production phase, including cell neutralization, the addition of solvent and relaxation of solvent/solute. The simulation was performed in a solvent box with dimensions 21.0 Å from all atoms. Overall, the simulation cell for 5′–G[8–5]U–3′-containing DNA contained 12 932 water molecules, 54 Na^+, 24 Cl^- and the cross-linked helix with a charge of –30. For the corresponding natural dsDNA, the simulation cell contained 17 150 water molecules, 66 Na^+, 36 Cl^- and the oligonucleotides with a charge of –30. This results in neutral cells with a density of 0.997 g mL^{-1} and a physiologically-relevant NaCl concentration of 0.9% (mass percent). The simulations were performed at 298 K and 1 atm. Long-range Coulomb interactions were calculated using the particle-mesh Ewald (PME) summation and a 10.54 Å cutoff was used to calculate the van der Waals terms. A time step of 1.25 fs was used for intramolecular forces and 2.5 fs for intermolecular forces. Snapshots were saved every 10 000 simulation steps, or every 25 ps.

Based on the resulting RMSDs for the simulation of the cross-linked and natural strands (Figure B–3), analysis was performed on the simulation snapshots after 3.000 ns, which ensured an adequate equilibration phase. Upon analysis of the trajectory (Figure B–4 and B–5), it became evident that cross-linked DNA preferentially occupies certain conformations. A conformation was deemed of interest and further analyzed if present for a consecutive 8 ns, which resulted in 3 different conformations of cross-linked DNA (designated Conformers 1–3). These occurred from 3.000 ns to 11.700 ns, 26.800 ns to 37.725 ns and 42.625 ns to 55.000 ns. Average structures (Figure 3–7) were generated over these ranges. Select geometric parameters were averaged individually for each conformer using all structures within the corresponding range listed above, and standard deviations were calculated to indicate the variation and flexibility within a conformer. An average structure for natural DNA was generated from 3.000 ns to 28.000 ns. Curves+[3] analysis was performed on the averaged conformers to calculate the helical axis and groove parameters.

[1] Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem*. **2003**, *24*, 1999.

[2] Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem*. **2004**, *25*, 1157.

[3] Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K. *Nucleic Acids Res*. **2009**, *37*, 5917.
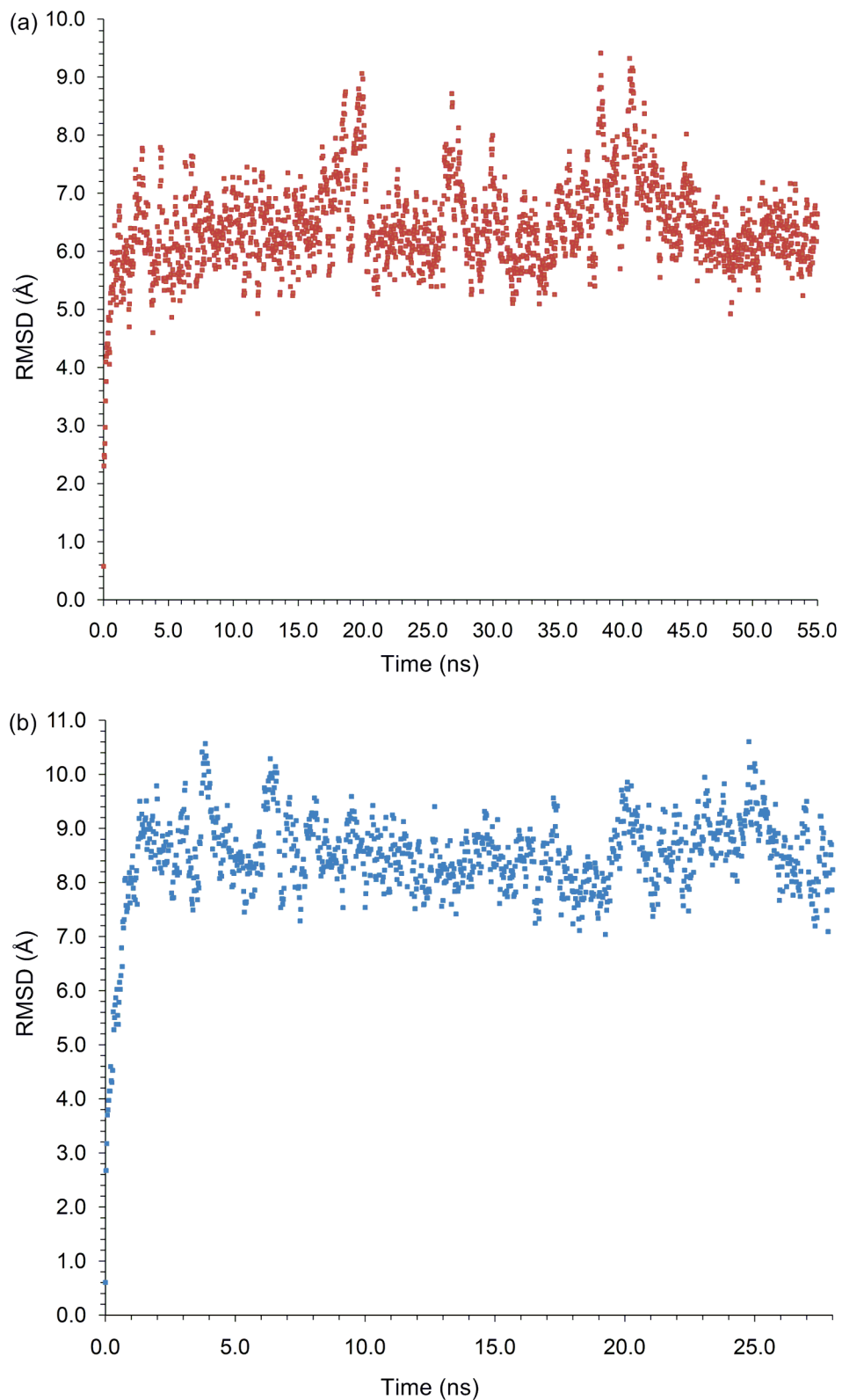
Figure B–3: Root mean square deviation (RMSD, Å) of the heavy atom coordinates in the DNA backbone during the simulation of (a) cross-linked and (b) natural DNA with respect to their input (gas-phase minimized) structure.

Table B–2: Backbone torsion angles of the dinucleoside monophosphates obtained from IEFPCM-M06-2X/6-31G(d,p) and dinucleotides obtained from molecular dynamics simulations for both natural and cross-linked DNA.

$\angle$[a]	M06-2X/6-31G(d,p)		Molecular Dynamics			
	5′–GT–3′	5′–G[8–5]U–3′	Natural[c]		Cross-linked[d]	
			Avg.	SD	Avg.	SD
α_G	–	–	290.5	10.4	133.5	26.6
β_G	–	–	162.2	19.4	181.1	19.2
γ_G	–	–	51.2	13.2	182.9	10.4
δ_G	149.8	156.3	122.6	21.7	142.2	9.3
ε_G	155.8	149.3	190.2	22.4	205.3	14.9
ζ_G	277.2	284.2	264.4	30.9	230.6	14.8
χ_G	230.0	6.0	239.8	19.3	43.0	7.4
α_U	274.4	266.8	293.6	9.1	140.8	57.0
β_U	236.8	240.6	168.1	15.8	188.0	11.2
γ_U	50.2	53.1	56.1	10.1	158.1	49.7
δ_U	152.5	155.2	111.0	20.8	134.3	12.7
ε_U	–	–	195.6	17.6	259.5	35.4
ζ_U	–	–	274.0	26.0	230.5	35.2
χ_U	263.8	182.7	235.6	16.5	217.5	12.3
θ[b]	–	243.8	–	–	164.0	7.9

[a]See Figure 1–2 for backbone torsion angles commonly used to describe DNA structure. [b]See Figure 3–1 for angle definition. [c]Values obtained as an average over all snapshots between 3.000 and 28.000 ns. [d]Values obtained as an average between 3.000 and 55.000 ns.
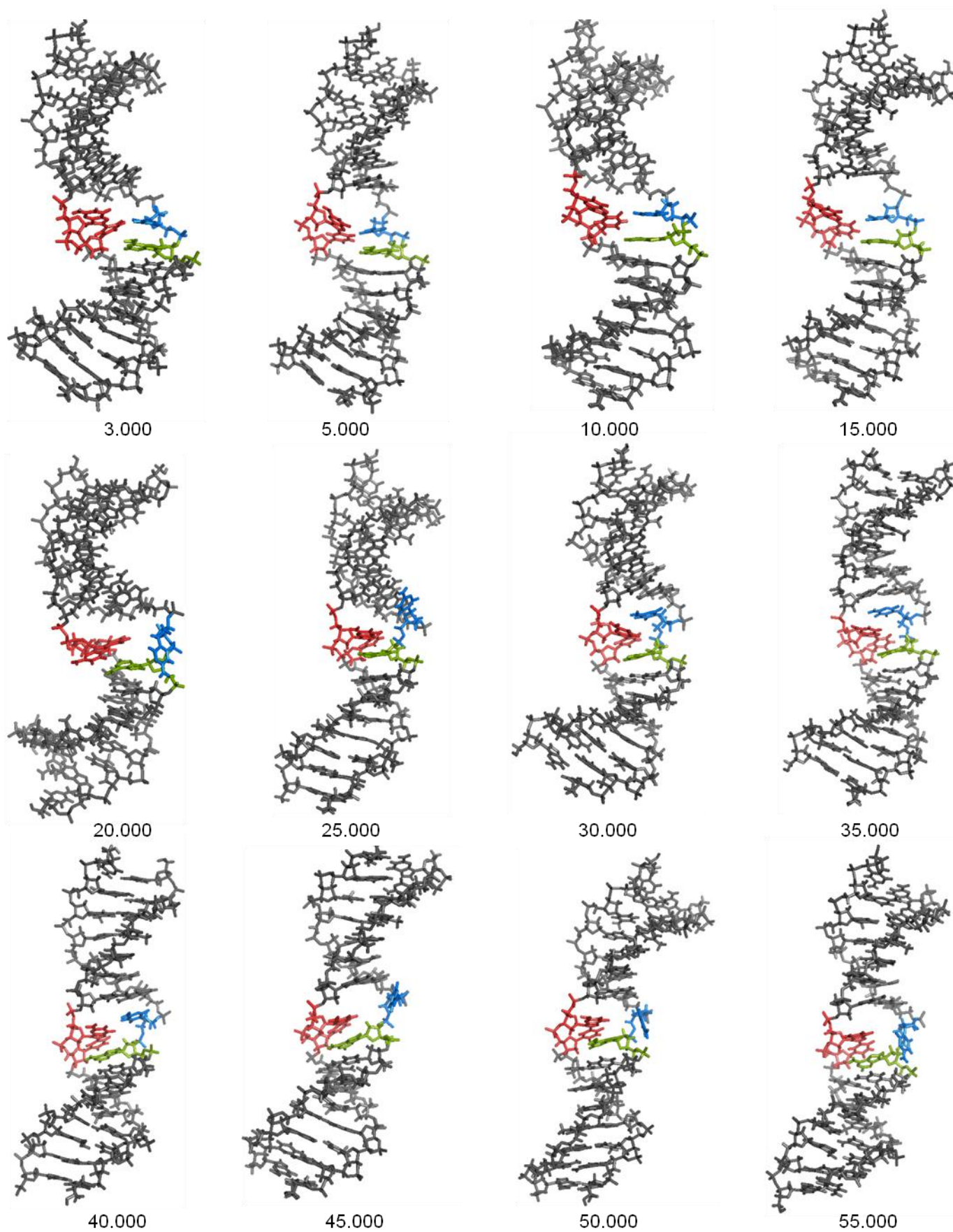
Figure B–4: Illustration of the trajectory of cross-linked DNA between 3.000 and 55.000 ns in 5 ns increments.
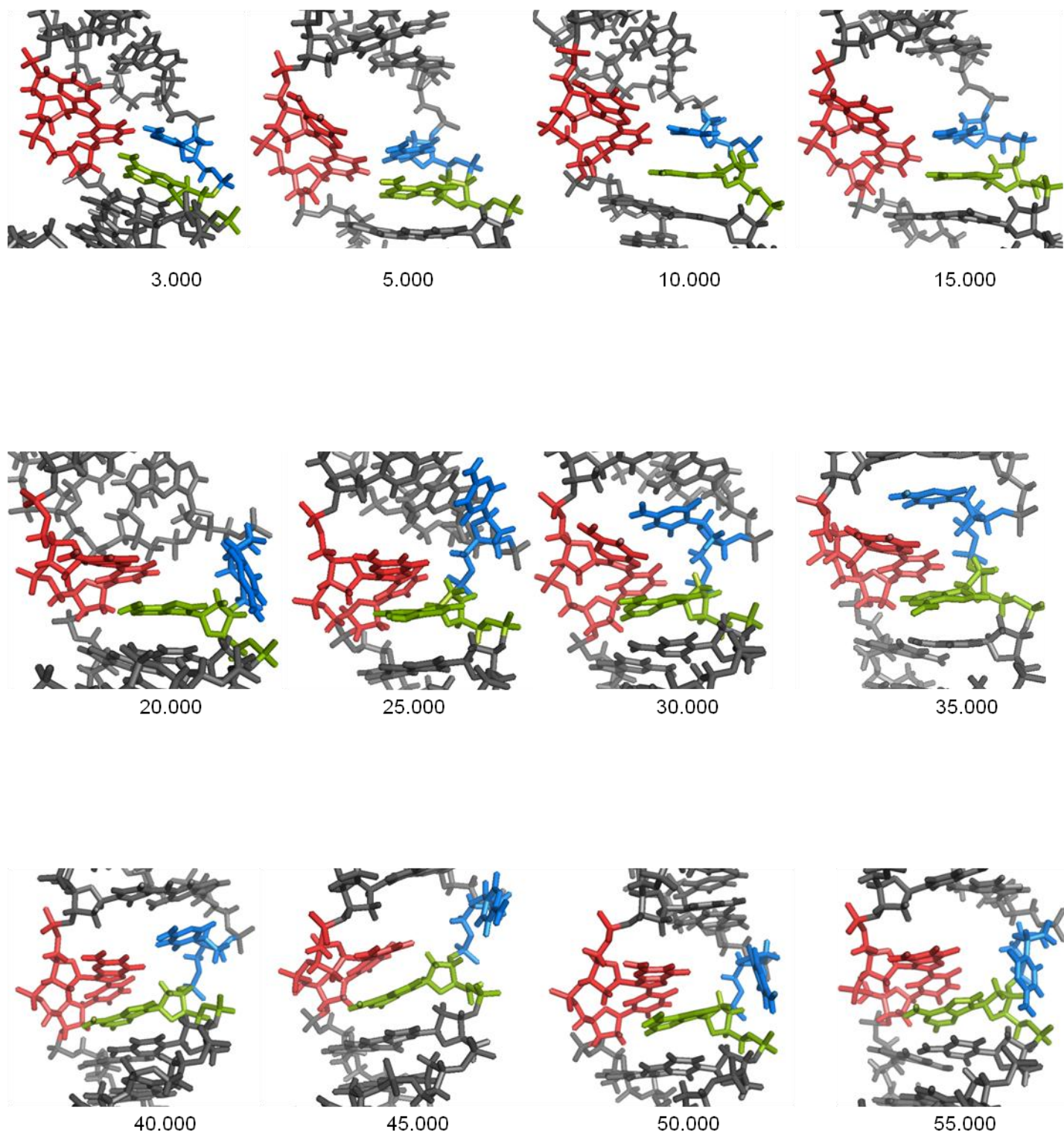
Figure B−5: Illustration of the cross-linked region along the trajectory between 3.000 and 55.000 ns in 5 ns increments.

Table B–3: Groove analysis for the Conformers 1–3 of cross-linked DNA and the corresponding natural strand.[a]

Level	Natural Minor W	Minor D	Major W	Major D	Conformer 1 Minor W	Minor D	Major W	Major D	Conformer 2 Minor W	Minor D	Major W	Major D	Conformer 3 Minor W	Minor D	Major W	Major D
3	7.0	4.0			6.8	4.2			6.8	3.9			7.1	3.9		
3.5	7.0	4.1			6.8	4.0			6.6	4.1			7.1	4.0		
4	7.5	3.7			6.9	4.0			6.7	4.1	14.4	5.7	7.5	3.8		
4.5	7.8	4.3			7.0	4.2	13.3	4.5	7.2	4.3	14.5	5.2	7.9	4.1	14.0	7.2
5	7.6	4.4	12.8	8.1	6.7	4.3	14.8	2.9	7.5	4.2	15	4	7.7	4.3	15.1	4.4
5.5	7.4	4.4	12.9	6.9	6.6	4.2	–[b]	–[b]	7.7	4.2	15.7	1.5	7.5	4.3	14.5	3.0
6	7.2	4.3	12.5	6.5	6.6	4.2	–[b]	–[b]	7.6	4.3	–[b]	–[b]	7.1	4.2	–[b]	–[b]
6.5	7.0	4.5	12.7	7.1	6.8	4.9	–[b]	–[b]	7.7	5.2	–[b]	–[b]	7.4	4.6	–[b]	–[b]
7	6.7	4.6	13.2	7.4	7.4	4.9	–[b]	–[b]	8.3	6.2	–[b]	–[b]	8.9	5.9	–[b]	–[b]
7.5[c]	6.8	4.9	12.9	5.8	8.5		–[b]		9.9		9.0		10.0		–[b]	
8[c]	6.8	4.7	12.4	6.8	9.2		–[b]		–[b]		7.6		8.6		–[b]	
8.5[c]	7.2	4.4	12.1	6.8	8.4		–[b]		–[b]		6.1		8.3		–[b]	
9[c]	7.7	4.2	12.8	6.6	8.0		–[b]		4.8		–[b]		8.2		8.7	
9.5[c]	7.4	4.8	13.4	5.6	7.8		–[b]		6.3		11.2		7.7		9.7	
10	6.8	4.9	13.6	6.8	7.1	3.7	–[b]	–[b]	6.9	3.1	11.7	3.4	7.0	4.1	10.9	2.3
10.5	6.3	4.7	13.4	6.9	6.6	4.2	–[b]	–[b]	7.5	3.7	12.3	4.7	7.5	4.4	11.9	3.5
11	5.9	4.4	14.2	6.2	6.5	4.4	13.4	5.8	7.6	4.1	12.7	5.3	7.9	4.3	12.1	4.4
11.5	5.9	4.7	15.1	4.3	7.2	4.3	12.4	6.3	7.7	4.2	12.9	5.9	8.0	4.2	13.5	3.8
12	6.1	4.8	15.0	1.4	7.5	4.0	11.8	6.8	7.6	3.9	13.8	5.5	7.8	4.0	15.2	6.0
12.5	6.3	4.9	15.4	4.3	7.7	3.9	–[b]	–[b]	7.2	4.1	14.1	5.4	7.3	4.1	–[b]	–[b]
13	6.2	4.7			7.5	3.7			6.6	4.1			6.6	4.0		
13.5	6.3	4.2			7.3	3.9			6.4	4.1			6.4	4.1		
14	7.3	3.8			7.2	3.8			6.6	4.1			6.7	4.0		

[a]Calculated using the Curves+[3] program. [b]Groove has disappeared, as calculated by Curves+. [c]Groove depths are not reported in the region of the intrastrand cross-link dinucleotides (levels 7.5 – 9.5) since this value is calculated with the vector defining the base pairs (constructed using C8 of purines and C6 of pyrimidines), which is highly distorted for the cross-link and its complementary bases.

Appendix C

Table of Contents

Figure C–1: The lowest-lying triplet surface (blue, solid line) and singlet ground state surface (green, solid line) generated as a function of d(C8–C5) (described in Section 4.3.1.1.), illustrating the intersystem crossing (degeneracy) between these states that connect R_T and $I2_S$ (see Figure 4–1 in text for structures).
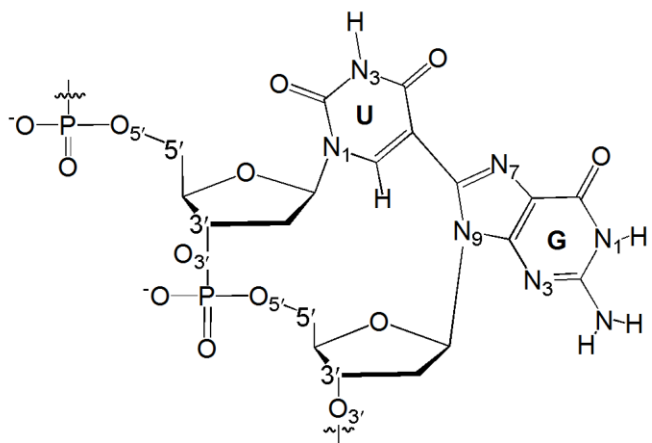
Figure C–2: Structure and chemical numbering of the cross-linked 5'–U[5–8]G–3' dinucleotide within double-stranded DNA.

Table C–1: GAFF atom types and charges assigned to the cross-linked dinucleotide. See Figure C–1 for chemical numbering.

	5'–U			3'–G	
Atom	Charges	Atom Type	Atom	Charges	Atom Type
N1	-0.3461	na	N1	-0.4891	na
C2	0.7979	c	H(N1)	0.3447	hn
O2	-0.5946	o	C2	0.7834	cd
N3	-0.4695	na	N2	-0.9653	nn
H(N3)	0.3376	hn	H(N2)	0.4362	hn
C4	0.7353	c	H(N2)	0.4362	hn
O4	-0.5648	o	N3	-0.6569	nc
C5	-0.3607	cc	C4	0.0890	cc
C6	0.1152	cd	C5	0.1492	cd
H(C6)	0.2103	h4	C6	0.5086	c
			O6	-0.5699	o
			N7	-0.5984	nd
			C8	0.5454	cc
			N9	-0.1861	na
C1'	0.0282	c3	C1'	0.0344	c3
H(C1')	0.1842	h2	H(C1')	0.1798	h2
C2'	-0.0883	c3	C2'	-0.0880	c3
H(C2')	0.0718	hc	H(C2')	0.0718	hc
H(C2')	0.0718	hc	H(C2')	0.0718	hc
C3'	0.2266	c3	C3'	0.0713	c3
O3'	-0.5200	os	O3'	-0.5268	os
H(C3')	0.0515	h1	H(C3')	0.0979	h1
C4'	0.1675	C3	C4'	0.1205	C3
H(C4')	0.0746	h1	H(C4')	0.1117	h1
O4'	-0.3706	os	O4'	-0.3719	os
C5'	0.0338	c3	C5'	0.0436	c3
H(C5')	0.0693	h1	H(C5')	0.0696	h1
H(C5')	0.0693	h1	H(C5')	0.0696	h1
O5'	-0.4962	os	O5'	-0.4016	os
O(P)	-0.7746	o	O(P)	-0.6969	o
O(P)	-0.7746	o	O(P)	-0.6969	o
P	1.1733	p5	P	0.9550	p5
Total	-0.9418		Total	-1.0581	

Simulation Details

Simulations were performed using the AMBER03[1] forcefield and GAFF[2] parameters for the cross-linked dinucleotide (Figure C–1 and Table C–1). Cross-linked DNA was generated by overlaying the backbone of the M06-2X-optimized dinucleoside monophosphate product onto the central TG of the 5′–d(GCATGGCG**TG**CTATGC)–3′ oligonucleotide. Any close contacts or artificial overlaps resulting from the overlay process were eliminated. Subsequently, this structure was minimized in the gas-phase with the AMBER and GAFF forcefields, and the resulting structure used as input for the molecular dynamics simulation.

Simulations were performed using pre-programmed protocols (the macro md_run.mcr) in YASARA. Initializing this macro automates all procedures required to begin the simulation from the desired input structure through to the production phase, including cell neutralization, the addition of solvent and relaxation of solvent/solute. The simulation was performed in a solvent box with dimensions 21.0 Å from all atoms. Overall, the simulation cell for 5′–U[5–8]G–3′-containing DNA contained 13 851 water molecules, 56 Na$^+$, 26 Cl$^-$ and the cross-linked helix with a charge of –30. For the corresponding natural dsDNA, the simulation cell contained 17 150 water molecules (density of 0.997 g mL^{-1}), 66 Na$^+$, 36 Cl$^-$ and the oligonucleotides with a charge of –30. This results in neutral cells with a density of 0.997 g mL^{-1} and a physiologically-relevant NaCl concentration of 0.9% (mass percent). The simulations were performed at 298 K and 1 atm. Long-range Coulomb interactions were calculated using the particle-mesh Ewald (PME) summation and a 10.54 Å cutoff was used to calculate the van der Waals terms. A time step of 1.25 fs was used for intramolecular forces and 2.5 fs for intermolecular forces. Snapshots were saved every 10 000 simulation steps, or every 25 ps.

Based on the resulting RMSDs for the simulation of the cross-linked and natural strands (Figure C–2), analysis was performed on the simulation snapshots after 4.000 ns, which ensured an adequate equilibration phase. Upon analysis of the trajectory (Figure C–3 and C–4), it became evident that cross-linked DNA preferentially occupies certain conformations. A conformation was deemed of interest and further analyzed if present for a consecutive 8 ns, which resulted in 2 different conformations of cross-linked DNA (designated Conformers 1 and 2). These occurred from 4.000 ns to 39.375 ns and 39.875 ns to 55.000 ns. Average structures (Figure 4–5) were generated over these ranges. Select geometric parameters were averaged individually for each conformation using all structures within the corresponding range listed above, and standard deviations were calculated to indicate the variation and flexibility within a conformation. An average structure for natural DNA was generated from 3.000 ns to 28.000 ns. Curves+[3] analysis was performed on the averaged conformations to calculate the helical axis and groove parameters.

[1] Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem*. **2003**, *24*, 1999.

[2] Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem*. **2004**, *25*, 1157.

[3] Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K. *Nucleic Acids Res*. **2009**, *37*, 5917.

Figure C–3: Root mean square deviation (RMSD, Å) of the heavy atom coordinates in the DNA backbone during the simulation of (a) cross-linked and (b) natural DNA with respect to their input (gas-phase minimized) structure.

Table C–2: Backbone torsion angles of the dinucleoside monophosphates obtained from IEFPCM-M06-2X/6-31G(d,p) and dinucleotides obtained from molecular dynamics simulations for both natural and cross-linked DNA.

$\angle$[a]	M06-2X/6-31G(d,p)		Molecular Dynamics					
	5'–TG–3'	5'–U[5–8]G–3'	Natural[c]		Conformer 1[d]		Conformer 2[e]	
			Avg.	SD	Avg.	SD	Avg.	SD
α_U	–	–	293.6	9.1	186.5	50.3	216.8	51.3
β_U	–	–	168.1	15.8	177.3	18.1	178.8	15.9
γ_U	–	–	56.1	10.1	214.3	58.1	230.9	71.6
δ_U	137.0	145.3	111.0	20.8	154.0	6.7	153.4	7.8
ε_U	200.0	155.8	195.6	17.6	182.9	6.8	176.6	6.6
ζ_U	278.1	276.2	274.0	26.0	267.8	6.4	263.7	5.7
χ_U	218.5	21.4	235.6	16.5	44.1	12.6	32.6	25.1
α_G	297.3	282.5	289.9	9.2	216.7	16.5	276.6	12.7
β_G	174.7	212.0	167.3	11.5	201.6	11.1	201.5	8.3
γ_G	58.5	51.2	49.4	12.3	51.7	14.2	39.0	9.2
δ_G	156.4	133.8	129.5	16.0	110.8	7.2	126.8	7.8
ε_G	–	–	233.1	43.7	289.1	11.1	282.8	19.5
ζ_G	–	–	199.4	57.2	90.0	14.0	115.6	23.2
χ_G	280.1	189.6			248.7	7.2	217.3	12.0
θ[b]	–	212.3	–	–	160.3	7.8	185.4	10.0

[a]See Figure 1–2 for backbone torsion angles commonly used to describe DNA structure. [b]See Figure 3–1 for angle definition and atom numbering. [c]Values obtained as an average over all snapshots between 3.000 and 28.000 ns. [d]Values obtained as an average between 4.000 and 39.375 ns. [e]Values obtained as an average between 39.875 and 55.000 ns.

Table C–3: Groove analysis for the two conformations of cross-linked DNA and the corresponding natural strand.[a]

Level	Natural Minor W	D	Major W	D	Conformation1 Minor W	D	Major W	D	Conformation 2 Minor W	D	Major W	D
3	7.0	4.0			6.9	3.9			6.8	4.0		
3.5	7.0	4.1			6.8	4.0			6.8	3.9		
4	7.5	3.7			6.9	4.0			7.1	3.8		
4.5	7.8	4.3			7.2	4.1	13.2	5.6	7.6	3.8	14.8	5.7
5	7.6	4.4	12.8	8.1	7.2	4.0	14.2	7	7.6	3.5	14.7	7.8
5.5	7.4	4.4	12.9	6.9	6.9	4.6	14.0	5.9	7.5	3.9	14.5	6.1
6	7.2	4.3	12.5	6.5	6.6	4.7	14.6	4.5	6.6	4.0	15.5	4.6
6.5	7.0	4.5	12.7	7.1	7.5	4.2	14.4	2.6	6.2	4.1	15.5	3.1
7	6.7	4.6	13.2	7.4	9.2	3.8	13.8	1.0	7.7	4.0	14.6	1.5
7.5	6.8	4.9	12.9	5.8	9.1	5.3	12.6	-1.3	6.3	4.8	12.9	−b
8	6.8	4.7	12.4	6.8	9.1	4.6	10.5	2.1	7.4	6.1	10.2	3.0
8.5 c	7.2	4.4	12.1	6.8	8.6		7.0		6.1		6.5	
9 c	7.7	4.2	12.8	6.6	−b		4.4		7.4		3.7	
9.5 c	7.4	4.8	13.4	5.6	13.8		1.2		0.8		−b	
10 c	6.8	4.9	13.6	6.8	11.6		2.4		2.1		−b	
10.5 c	6.3	4.7	13.4	6.9	9.8		5.7		4.6		−b	
11	5.9	4.4	14.2	6.2	8.7	9.2	8.9	-2.5	8.6		7.5	4.2
11.5	5.9	4.7	15.1	4.3	6.3	7.9	−b	−b	10.5	6.4	11.1	3.8
12	6.1	4.8	15.0	1.4	6.8	7.3	−b	−b	8.7	2.3	14.5	3.0
12.5	6.3	4.9	15.4	4.3	9.4	3.0	−b	−b	6.1	3.7	−b	−b
13	6.2	4.7			8.3	3.3	−b	−b	5.1	3.9		
13.5	6.3	4.2			7.1	3.8	14.4	2.7	5.8	4.1		
14	7.3	3.8			6.6	4.1			6.4	3.7		

[a]Calculated using the Curves+[3] program. [b]Groove has disappeared, as calculated by Curves+. [c]Groove depths are not reported in the region of the intrastrand cross-link dinucleotides (levels 8.5 – 10.5) since this value is calculated with the vector defining the base pairs (constructed using C8 of purines and C6 of pyrimidines), which is highly distorted for the cross-link and its complementary bases.
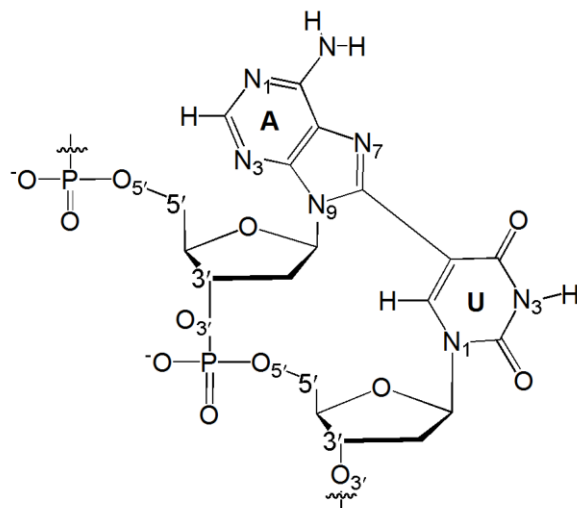
Appendix D

Table of Contents

Figure D–1: Structure and chemical numbering of the cross-linked 5'–A[8–5]U–3' dinucleotide within double-stranded DNA. The same numbering is used for 5'–A[2–5]U–3'

Table D–1: GAFF atom types and charges assigned to the cross-linked dinucleotides. See Figure D–1 for chemical numbering.

5′–A[8–5]U–3′						5′–A[2–5]U–3′					
5'–A			3'–U			5'–A			3'–U		
Atom	Charges	Atom Type	Atom	Charges	Atom Type	Atom	Charges	Atom Type	Atom	Charges	Atom Type
N1	-0.7667	nb	N1	-0.3544	na	N1	-0.7648	no	N1	-0.3416	na
C2	0.5846	ca	C2	0.8051	c	C2	0.7837	ca	C2	0.7903	c
H(C2)	0.0503	h5	O2	-0.5945	o	N3	-0.7397	nb	O2	-0.5942	o
N3	-0.7146	nb	N3	-0.4655	na	C4	0.3515	ca	N3	-0.4647	na
C4	0.4828	ca	H(N3)	0.3376	hn	C5	-0.2086	ca	H(N3)	0.3376	hn
C5	-0.2117	ca	C4	0.7338	c	C6	0.6990	ca	C4	0.7357	c
C6	0.6984	ca	O4	-0.5623	o	N6	-0.9041	nh	O4	-0.5618	o
N6	-0.9043	nh	C5	-0.3669	cd	H(N6)	0.4109	hn	C5	-0.3869	cd
H(N6)	0.4108	hn	C6	0.1399	cc	H(N6)	0.4109	hn	C6	0.1415	cc
H(N6)	0.4108	hn	H(C6)	0.2102	h4	N7	-0.6108	nd	H(C6)	0.2137	h4
N7	-0.5627	nc				C8	0.1983	cc			
C8	0.5251	cd				H(C8)	0.1877	h5			
N9	-0.2351	na				N9	-0.2583	na			
C1'	0.0344	c3	C1'	0.0282	c3	C1'	0.0706	c3	C1'	0.0588	c3
H(C1')	0.1830	h2	H(C1')	0.1877	h2	H(C1')	0.1838	h2	H(C1')	0.1884	h2
C2'	-0.0859	c3	C2'	-0.0861	c3	C2'	-0.1488	c3	C2'	-0.0854	c3
H(C2')	0.0718	hc	H(C2')	0.0718	hc	H(C2')	0.0342	hc	H(C2')	0.0762	hc
H(C2')	0.0718	hc	H(C2')	0.0718	hc	H(C2')	0.0342	hc	H(C2')	0.0762	hc
C3'	0.2165	c3	C3'	0.0713	c3	C3'	0.2120	c3	C3'	0.2429	c3
O3'	-0.5173	os	O3'	-0.5239	os	O3'	-0.5167	os	O3'	-0.4899	os
H(C3')	0.0515	h1	H(C3')	0.0984	h1	H(C3')	0.0546	h1	H(C3')	0.0563	h1
C4'	0.1692	c3	C4'	0.1207	C3	C4'	0.1695	C3	C4'	0.1210	C3
H(C4')	0.0753	h1	H(C4')	0.1133	h1	H(C4')	0.0754	h1	H(C4')	0.1136	h1
O4'	-0.3694	os	O4'	-0.3697	os	O4'	-0.3691	os	O4'	-0.3691	os
C5'	0.0338	c3	C5'	0.0436	c3	C5'	0.0490	c3	C5'	0.0598	c3
H(C5')	0.0705	h1	H(C5')	0.0704	h1	H(C5')	0.0708	h1	H(C5')	0.0706	h1
H(C5')	0.0705	h1	H(C5')	0.0704	h1	H(C5')	0.0708	h1	H(C5')	0.0706	h1
O5'	-0.4943	os	O5'	-0.3957	os	O5'	-0.4939	os	O5'	-0.3942	os
O(P)	-0.7730	o	O(P)	-0.6931	o	O(P)	-0.7726	o	O(P)	-0.6921	o
O(P)	-0.7730	o	O(P)	-0.6931	o	O(P)	-0.7726	o	O(P)	-0.6921	o
P	1.1733	p5	P	0.9550	p5	P	1.2008	p5	P	1.0113	p5
Total	-1.0236		Total	-0.9760		Total	-1.2923		Total	-0.7075	

Simulation Details

Simulations were performed using the AMBER03[1] forcefield and GAFF[2] parameters for the cross-linked dinucleotide (Figure D–1 and Table D–1). Cross-linked DNA strands were generated by overlaying the C5′ and O5′ atoms in the backbone of the M06-2X-optimized dinucleoside monophosphate product onto the central AT of the 5′–d(GCATGGC**AT**GCTATGC)–3′ oligonucleotide. Any close contacts or artificial overlaps resulting from the overlay process were eliminated. Subsequently, this structure was minimized in the gas-phase with the AMBER and GAFF forcefields, and the resulting structure used as input for the molecular dynamics simulation.

Simulations were performed using pre-programmed protocols (the macro md_run.mcr) in YASARA. Initializing this macro automates all procedures required to begin the simulation from the desired input structure through to the production phase, including cell neutralization, the addition of solvent and relaxation of solvent/solute. The simulation was performed in a solvent box with dimensions 21.0 Å from all atoms. For 5′–A[8–5]U–3′-containing DNA, the simulation cell contained 14 226 water molecules, 58 Na^+, 28 Cl^- and the cross-linked helix with a charge of –30. For 5′–A[2–5]U–3′-containing DNA, the simulation cell contained 12 435 water molecules, 53 Na^+, 23 Cl^- and the cross-linked helix with a charge of –30. For the corresponding natural dsDNA, the simulation cell contained 16 016 water molecules, 63 Na^+, 33 Cl^- and the oligonucleotides with a charge of –30. This results in neutral cells with a density of 0.997 g mL^{-1} and a physiologically-relevant NaCl concentration of 0.9% (mass percent).The simulations were performed at 298 K and 1 atm. Long-range Coulomb interactions were calculated using the particle-mesh Ewald (PME) summation and a 10.54 Å cutoff was used to calculate the van der Waals terms. A time step of 1.25 fs was used for intramolecular forces and 2.5 fs for intermolecular forces. Snapshots were saved every 10 000 simulation steps, or every 25 ps.

Based on the resulting RMSDs for the simulation of the cross-linked and natural strands (Figure D–2), analysis was performed on the simulation snapshots after 4.000 ns, which ensured an adequate equilibration phase. Upon analysis of the trajectory, it became evident 5′–A[8–5]U–3′- and 5′–A[2–5]U–3′-containing DNA each occupied only one conformation (designated a Conformer). A conformation was deemed of interest and further analyzed if present for a consecutive 8 ns. For DNA containing 5′–A[8–5]U–3′, a stable conformation occurred from 19.600 ns to 43.000 ns. For DNA containing 5′–A[2–5]U–3′, a stable conformation occurred from 4.000 ns to 55.000 ns. Average structures (Figure 5–5) were generated over these ranges. Select geometric parameters were averaged individually for each conformation using all structures within the corresponding range listed above, and standard deviations were calculated to indicate the variation and flexibility within a conformation. An average structure for natural DNA was generated from 4.000 ns to 29.000 ns. Curves+[3] analysis was performed on the averaged conformations to calculate the helical axis and groove parameters.

[1]Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem*. **2003**, *24*, 1999.
[2]Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem*. **2004**, *25*, 1157.
[3]Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K. *Nucleic Acids Res*. **2009**, *37*, 5917.
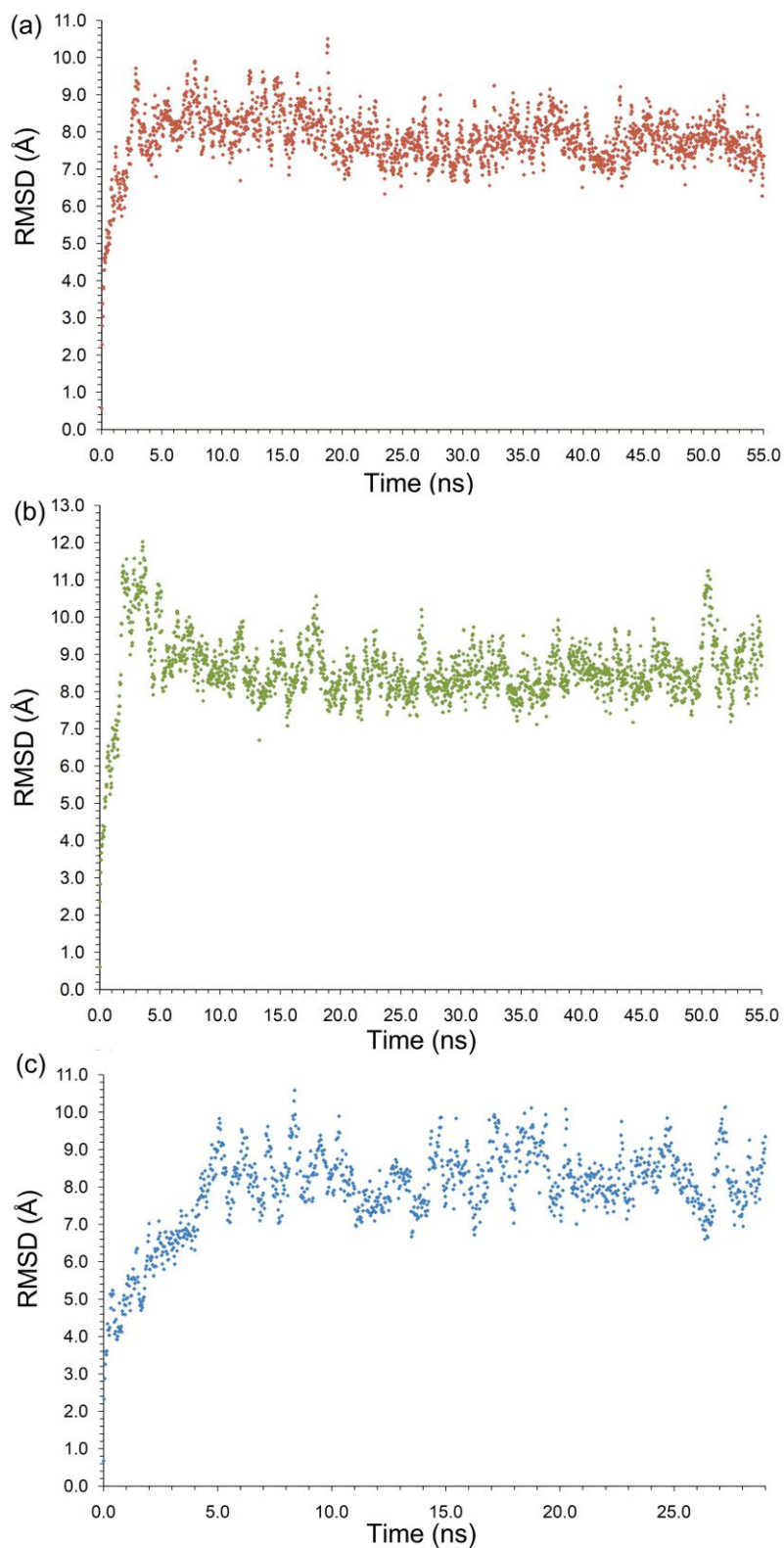
Figure D–2: Root mean square deviation (RMSD, Å) of the heavy atom coordinates in the DNA backbone during the simulation of (a) 5′–A[8–5]U–3′, (b) 5′–A[2–5]U–3′ and (c) natural DNA with respect to their input (gas-phase minimized) structure.

Table D–2: Backbone torsion angles of the dinucleoside monophosphates obtained from IEFPCM-M06-2X/6-31G(d,p) and dinucleotides obtained from molecular dynamics simulations for both natural and cross-linked DNA.

∠[a]	M06-2X/6-31G(d,p)			Molecular Dynamics					
	5'–AT–3'	5'–A[8–5]U–3'	5'–A[2–5]U–3'	Natural[c]		5'–A[8–5]U–3'[d]		5'–A[2–5]U–3'[e]	
				Avg.	SD	Avg.	SD	Avg.	SD
α_A	–	–	–	291.2	9.4	207.9	23.0	212.4	38.3
β_A	–	–	–	170.5	15.0	181.5	15.3	185.1	25.7
γ_A	–	–	–	53.4	11.8	195.7	14.3	190.5	21.0
δ_A	213.3	150.4	151.3	116.7	20.7	128.5	11.7	134.5	14.0
ϵ_A	169.1	155.6	168.2	192.1	25.2	199.4	12.3	208.0	20.8
ζ_A	264.1	279.8	267.9	263.3	34.4	226.5	13.3	221.2	20.2
χ_A	252.1	353.3	188.9	240.5	20.1	42.0	7.5	203.4	16.8
α_U	300.0	263.4	266.7	293.2	8.9	119.3	17.3	141.0	35.7
β_U	183.0	243.4	247.3	164.4	16.2	191.9	10.7	286.4	19.8
γ_U	57.1	54.7	43.1	56.3	10.3	184.2	9.5	187.1	29.6
δ_U	146.8	155.9	150.6	106.9	21.0	137.6	8.0	144.4	11.7
ϵ_U	–	–	–	197.0	22.4	281.9	10.0	263.9	30.3
ζ_U	–	–	–	266.5	35.7	130.6	19.5	132.5	30.5
χ_U	252.7	183.9	226.8	234.3	18.3	212.8	23.2	201.7	11.3
θ[b]	–	246.1	183.5	–	–	165.2	7.5	166.1	15.0

[a]See Figure 1–2 for backbone torsion angles commonly used to describe DNA structure. [b]See Figure 5–1 for angle definition. [c]Values obtained as an average over all snapshots between 4.000 and 29.000 ns. [d]Values obtained as an average between 19.600 ns to 43.000 ns. [e]Values obtained as an average between 3.000 ns to 55.000 ns.

Table D–3: Groove analysis for the two conformations of cross-linked DNA and the corresponding natural strand.[a]

	Natural				5′–A[8–5]U–3′				5′–A[2–5]U–3′			
	Minor		Major		Minor		Major		Minor		Major	
Level	W	D	W	D	W	D	W	D	W	D	W	D
3	6.4	4.3			7.5	3.7			6.6	4.0		
3.5	6.6	4.1			7.4	3.9			6.4	4.0		
4	7.4	3.8			7.5	3.6	–	–	6.4	4.0	13.5	5.9
4.5	7.7	4.2	12.7	7.2	7.5	4.1	–	–	6.8	4.2	14.1	3.8
5	7.4	4.3	12.8	7.5	7.2	4.4	–	–	6.9	4.3	16.6	5.6
5.5	7.1	4.4	12.8	7.5	7.1	4.3	16.7	3.9	6.9	4.3	15.3	5.0
6	6.8	4.4	13.2	8.0	7.2	4.4	14.2	3.4	7.0	4.2	15	4.8
6.5	6.8	4.5	13.6	8.1	6.8	5.5	13.5	1.6	7.3	4.6	14.5	3.6
7	7.0	4.4	13.5	8.1	6.9	6.1	–[b]	–[b]	7.9	5.2	–[b]	–[b]
7.5[c]	7.7	4.4	13.1		7.9	0.3	–[b]		7.9	1.2	12.1	
8[c]	7.9	3.9	12.1		6.2	3.1	11.7		5.8	2.8	10.8	
8.5[c]	7.7	4.0	11.3		6.6	4.5	10.6		5.6	4.3	10.1	
9[c]	7.6	3.9	11.8		5.8	5.5	9.5		4.8	5.0	9.6	
9.5[c]	7.5	4.3	12.7		7.2	4.5	10.0		5.4	4.8	9.8	
10	7.0	4.3	12.9	8.3	8.4	3.2	10.6	5.8	5.9	4.2	10.9	4.4
10.5	6.9	4.4	13.0	8.3	6.9	4.1	11.5	6.4	6.2	4.2	11.5	5.3
11	6.9	4.2	12.7	8.4	6.3	4.6	12.1	6.3	6.7	4.1	12.0	5.4
11.5	7.3	4.1	12.5	8.0	7.1	4.4	12.2	6.0	7.4	4.2	12.6	5.8
12	7.6	4.0	12.8	7.7	7.6	3.6	11.7	6.8	7.8	4.2	13.7	5.6
12.5	7.7	4.1	13.4	7.4	7.3	4	–[b]	–[b]	7.7	4.3	15.6	4.2
13	7.2	3.8			6.8	4.2			7.1	3.9		
13.5	6.8	4.1			6.6	4.1			6.9	4.0		
14	6.7	4.2			6.7	4			7.7	3.7		

[a]Calculated using the Curves+[3] program. [b]Groove has disappeared, as calculated by Curves+. [c]Groove depths are not reported in the region of the intrastrand cross-link dinucleotides (levels 7.5 – 9.5) since this value is calculated with the vector defining the base pairs (constructed using C8 of purines and C6 of pyrimidines), which is highly distorted for the cross-link and its complementary bases.