CHAPTER I

INTRODUCTION

DNA Structure

Every living organism is created with the ability to store and transmit hereditary information through biomolecules of deoxyribonucleic acid (DNA) [1]. DNA was first recognized as genetic information by Avery and coworkers in 1944 [2]. Almost a decade later, the structure of DNA was reported in the famous 1953 *Nature* paper by Watson and Crick [3], and from this structure we were provided insight into how genetic information is both replicated during cell division and expressed into other biological macromolecules like ribonucleic acid (RNA) and proteins.

DNA is comprised of two long polymeric chains, coiled around the same axis and following a right-handed helix. The monomers which make up DNA are specifically adenine (A), guanine (G), cytosine (C), and thymine (T) [1]. Each monomer contains either a purine (A and G) or a pyrimidine (C and T) base and a deoxyribose sugar connected through a β-N-glycosidic bond. Together the sugar and base comprise what is referred to as a nucleoside. The chemical structures of the four DNA nucleosides are shown in Figure 1-1.

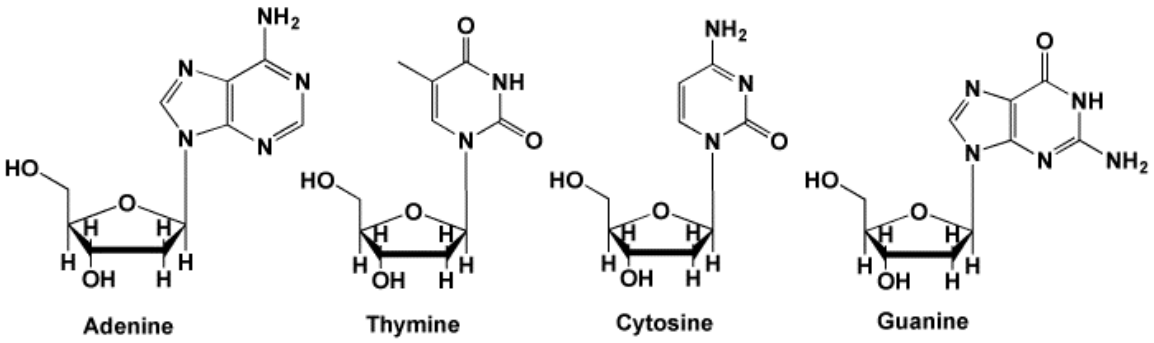


Figure 1-1: The chemical structures of DNA nucleosides.

The sugars of multiple nucleosides are connected by a phosphate group via the 5′ and 3′-hydroxyl groups, creating an oligonucleotide. These are formed with the phosphate groups on the outside of the helix and the bases extending horizontally beyond the backbone into the center of the strand. The nucleotides have directionality with one end of the strand being labeled as 5′ and the opposite end of the same strand being labeled as 3′. This nomenclature is determined by the ribose carbon to which the phosphate group is attached.

The individual strands of duplex DNA are antiparallel and are, by convention, read from the 5′-end to the 3′-end. Each residue in a single strand is separated by a distance of 3.4 Å and the strand turns every 10 residues due to a 36° angle between residues [3]. Duplex DNA contains a major groove and a minor groove. The depth of each groove is similar yet the major groove of the helix is wide and the minor groove is narrow [4]. The chemical structure of DNA and its components are shown in Figure 1-2.

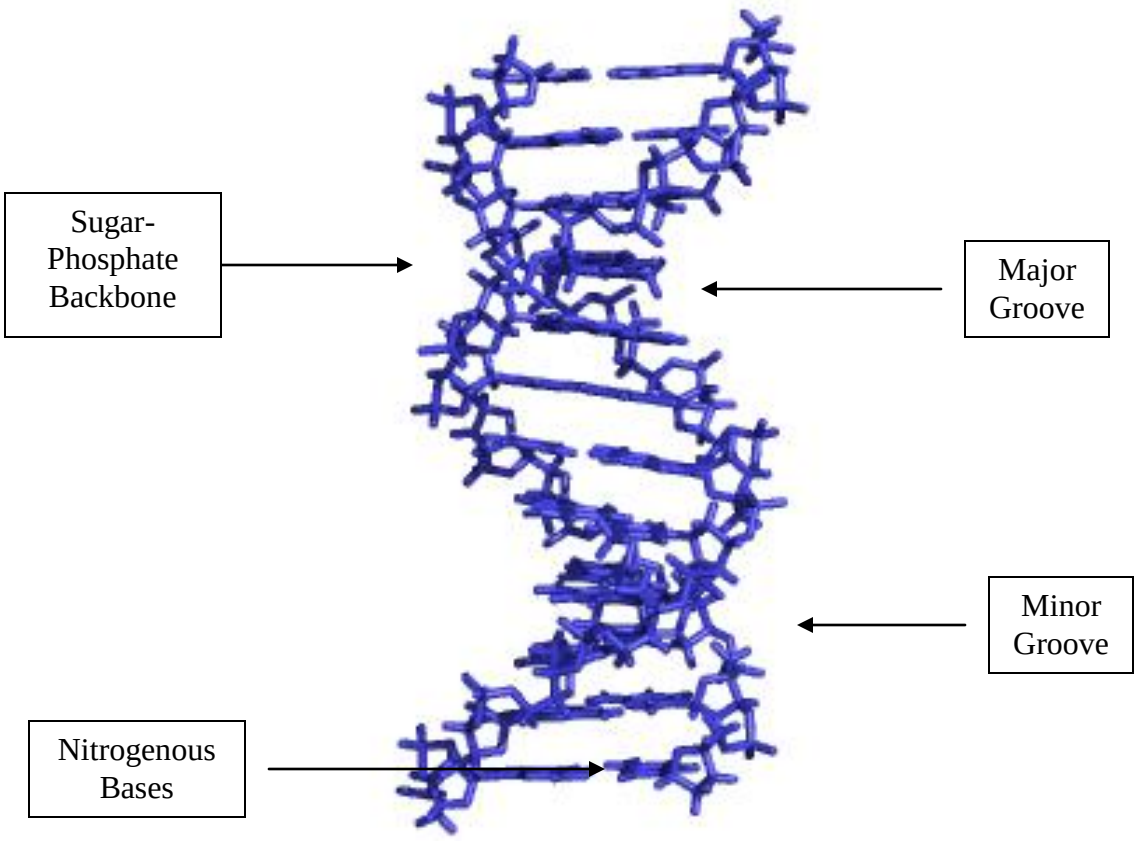


Figure 1-2: Chemical structure and components of B-DNA**.**

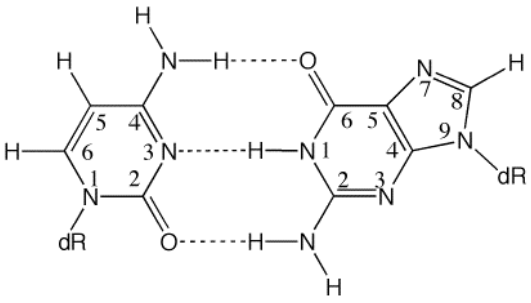
The structure described above represents B-DNA; however, other forms of DNA exist in nature and have structural differences. This change in structure is dependent on the DNA sequence as well as its environment [5]. The three most common forms of DNA are A, B, and Z, and examples of all three forms of DNA exist in nature. Like B-DNA, the A form follows a right-handed helix. A-DNA, however, is shorter than B-DNA with the same number of bases, and the bases are tilted toward the helix. The major and minor grooves of A-DNA differ in depth with the minor groove being more shallow [4]. This conformation typically exists at high salt concentrations [6].

Z-DNA follows a left-handed helix. This form of DNA is longer and thinner than B-DNA, and, in contrast to A-form DNA, the minor grove is deeper than the major

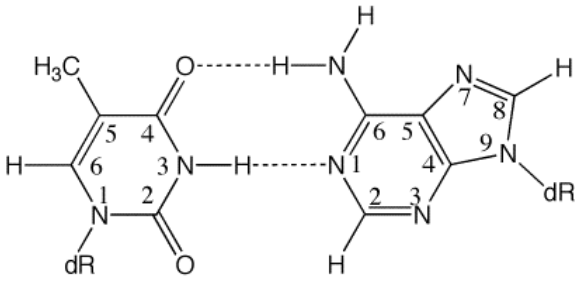
groove. Z-DNA is typically a sequence of alternating guanines and cytosines in high salt concentrations [7-9]. This conformation does not exist as a stable feature of duplex DNA. Instead, it is a short-lived structure that can be induced by biological activity before quickly disappearing [10].

In duplex DNA, the bases extend from the strand perpendicularly to the axis and are hydrogen-bonded to complementary bases in the opposite strand, making the double-helix model. Hydrogen bonding is a chemical mechanism involving electrostatic interactions between hydrogen atoms and oxygen, nitrogen, and fluorine atoms [11]. In the canonical Watson-Crick base-pairing model, hydrogen bonding occurs between specific purines and pyrimidines, for example, adenines bind to thymines and guanines bind to cytosines.

Hydrogen bonding between two purine bases is not energetically preferred because the close proximity of the bases would result in repulsion due to overlapping atoms. Pyrimidine:pyrimidine hydrogen bonding would not be energetically favorable either because the atoms would not be close enough for the electrostatic interactions to take place. Other hydrogen bonding motifs exist in nature including Hoogsteen and Wobble base-pairing; however, these are more common in RNA [12]. The Watson-Crick hydrogen bonding patterns are shown in Figure 1-3.



Cytosine:Guanine



Thymine:Adenine

Figure 1-3: Watson-Crick model of hydrogen bonding.

The primary source of stability in an oligonucleotide comes from the base-stacking interactions involving the pi orbitals of the aromatic rings of DNA bases [13]. Hydrogen bonding adds to the stability of duplex DNA. Therefore DNA sequences with a high G:C base pair content are considered to be more stable; however, the contribution of the electrostatic interactions to the stability of the complex is not substantial. The weak hydrogen bonds between the bases in duplex DNA allow separation of the strands to occur readily without disrupting the backbone so that replication can take place.

DNA is duplicated throughout living organisms by means of templated polymerization [1]. A pre-existing strand of DNA becomes the template on which DNA is synthesized. Because of the strict rules of base-pairing in DNA, the selection of which monomer will be added to the nascent strand is dictated by the bases protruding from the

template strand of DNA. Therefore, from a single molecule of duplex DNA, two identical strands are created.

DNA Damage

DNA is under constant attack from various exogenous and endogenous sources, including alterations by normal cellular processes [14]. The result of these attacks leads to an assortment of DNA injuries including oxidation, alkylation, and hydrolysis. The rate of DNA damage can be anywhere from 1,000 to 1,000,000 lesions per cell per day [15].

Multiple DNA repair mechanisms exist to remove and repair DNA lesions, including direct repair, base excision repair, nucleotide excision repair, double-strand break repair, and cross-link repair [16-18]. Therefore, most occurrences of DNA damage are readily corrected by these repair pathways. If the damage is left unrepaired, during replication the geometry of the modified bases is unrecognizable to replication polymerases. This may result in blocked replication and ultimately cell death or progression by error-prone translesion DNA polymerases, resulting in mutagenesis, and ultimately a plethora of deleterious effects including cancer [19-21]. The occurrence of DNA damage can be attributed to DNA replication errors, the chemical instability of DNA in solution, and covalent modification of DNA bases by endogenous or exogenous carcinogens [22-25].

DNA Replication Errors

The precise diffusion of genetic information from one cell to its daughters is necessary for organism survival. This is dependent on many things, not excluding extreme accuracy in the replication of DNA. While DNA replication takes place, double-strand breaks (DSB) may happen from stalled replication forks [26]. Another error would be the misincorporation of a nucleotide in the growing DNA strand. DNA damage occurring from replication errors is not frequent, however, due to the high fidelity of DNA replication and specific proofreading processes [27].

Chemical Instability of DNA

DNA is not stable under physiological conditions [28]. In fact, the primary structure of DNA is dynamic and prone to constant change. The N-glycosyl bond in DNA is labile, making it susceptible to attack by water, resulting in hydrolysis [22]. Depurination is the most frequent covalent change inducing abasic sites [29]. DNA bases, particularly cytosine, guanine, adenine, and 5-methylcytosine, are also vulnerable to spontaneous hydrolytic deamination at specific temperatures and pH values. The loss of the exocyclic amino group in this reaction results in uracil, xanthine, hypoxanthine and thymine, respectively [22]. Figure 1-4 illustrates those sites most susceptible to depurination and hydrolytic deamination.

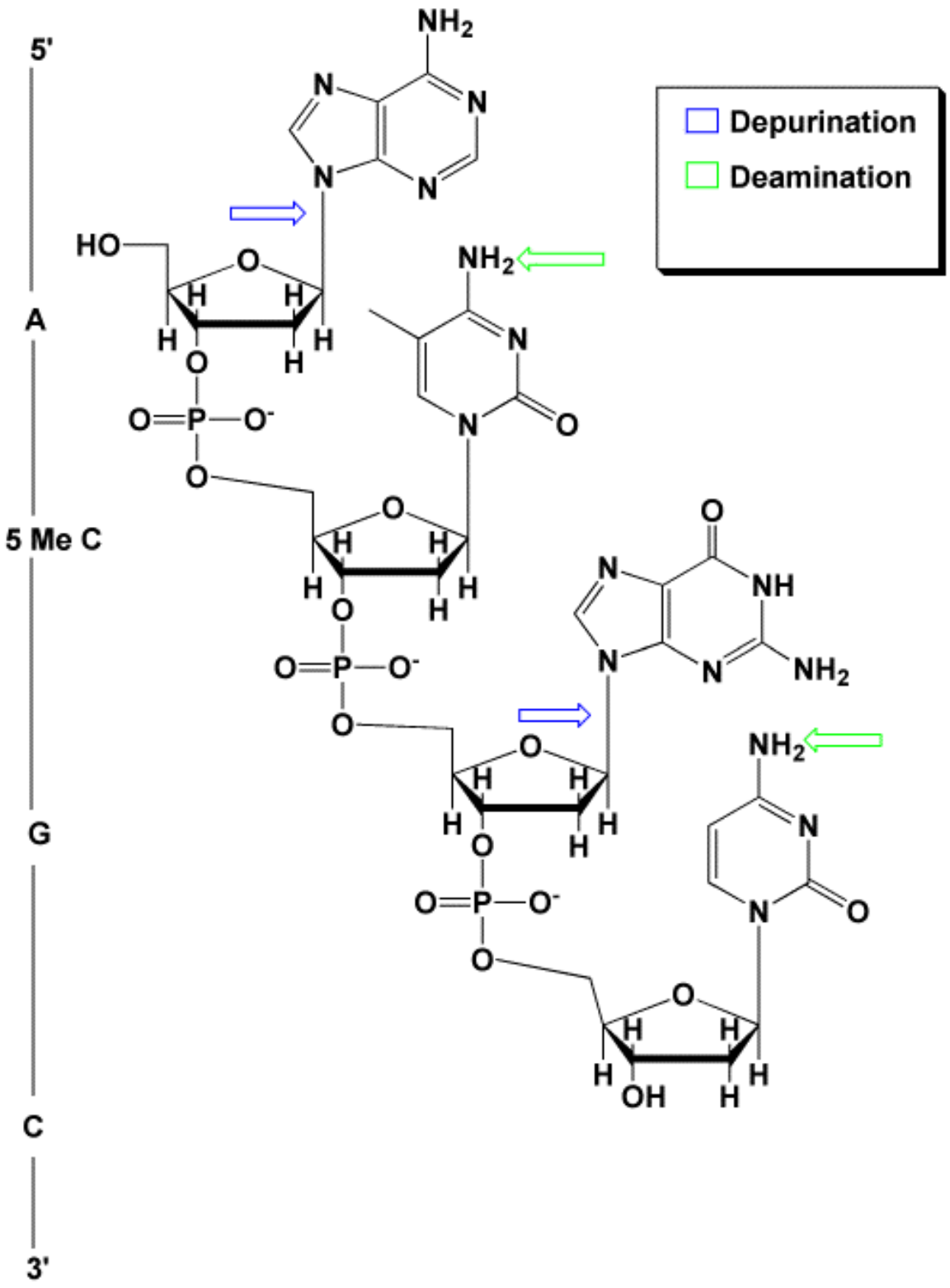


Figure1-4: DNA sites susceptible to depurination and hydrolytic deamination.

Covalent Modification of DNA Bases

The reality that the rate of cancer occurrence increases exponentially with increasing age [30, 31] suggests that insults to DNA occur in normal cellular metabolism through endogenous sources. One of the most common sources is oxygen free radicals resulting from respiration, phagocytosis, inflammation, and cell injury [14]. Endogenous DNA damage by oxygen free radicals occurs at a rate similar to that of depurination [32]. Every day an estimate of 20,000 alterations induced by oxygen free radicals occur in DNA per human cell [33, 34]. Reactive oxygen species including superoxide anions, hydrogen peroxide and hydroxyl radicals lead to oxidative modifications. A common modification resulting from oxidative damage is the 8-hydroxy-deoxyguanine adduct [22]. Endogenous cellular metabolites also covalently modify DNA. An example of this is S-adenosylmethionine, which results in DNA methylation. Other processes including chlorination, glycosylation, nitrosylation, and covalent attack by lipid peroxidation products can lead to DNA injury and can be carcinogenic [28, 35, 36].

DNA is also bombarded with exogenous assaults including several physical agents and chemical compounds. Of those physical agents, ionizing radiation and UV radiation are the most common. Ionizing radiation comes from X-rays and gamma rays and has been a source of DNA damage to living organisms leading to an assortment of DNA lesions [37-39]. Gamma-irradiation is most likely to cause damage to thymine bases, where the C5-C6 double bond is attacked by ·OH, producing thymine glycol, among other derivatives. Other examples of base damage include 8-hydroxyguanine, 5-hydroxymethyluracil, 4,5-diamino-5-formamidopyrimidine, and urea [40]. Single-strand and double-strand breaks also result from γ-irradiation at low levels [41].

The absorption of photons by DNA bases is due to UV radiation. UV radiation, broken into three subgroups based on wavelength, is capable of producing cyclobutane pyrimidine dimers, pyrimidine-pyrimidine (6-4) photoproducts, purine lesions, and thymine glycol. In addition to nucleotide base damage, UV radiation can result in DNA cross-links as well as single- and double-strand breaks [40].

Several sources of exogenous chemical carcinogens exist. Exposure to these chemical agents can come from food and drink consumption, drug intake, and exposure to environmental contaminants. The interaction of DNA with potential chemotherapy targets was one of the first studies of drug interactions with DNA. It was thought that, as a result of DNA modification by these drugs, DNA replication would be impeded, bringing about replication arrest of cancer cells [40]. Since then, numerous carcinogenic and mutagenic agents have been studied.

Several known carcinogens are electrophilic species that will readily attack nucleophilic sites on DNA bases. The nucleophilic sites susceptible to covalent modification by alkylating species are shown in Figure 1-5. There are also many other carcinogens which require metabolic activation by Cytochrome P450 in order to attack DNA [42-44]. An example is 2-Acetylaminofluorene (AAF) which is metabolized to the active form N-sulfoxy-AAF before alkylating guanine bases [45, 46]. Likewise, butadiene is a known carcinogen which requires metabolic activation to mono- and diepoxides in order to alkylate DNA bases [47]. The covalent modification of DNA bases by alkylating agents can cause structural perturbations to DNA which can result in blocked DNA replication or mutagenesis. In addition, when bases are modified by

alkylating species, the N-glycosylic bond is weakened. Therefore, as a result of alkylation, depurination or depyrimidination is likely to take place [48].

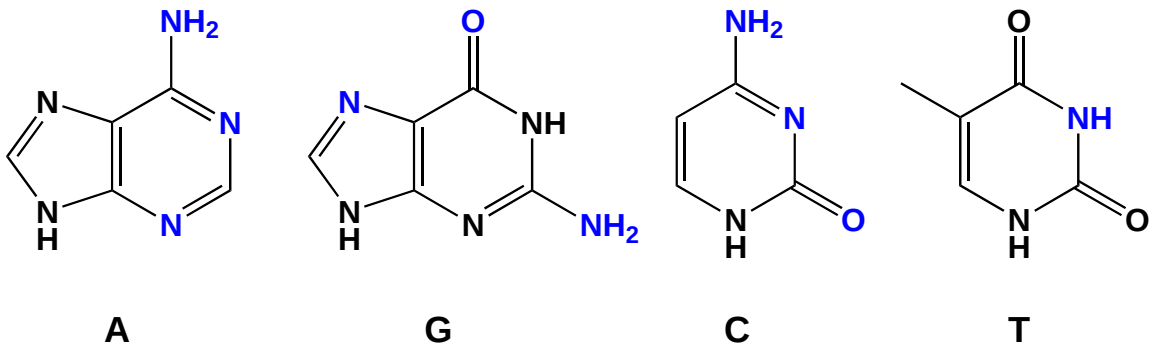


Figure 1-5: Nucleophilic sites on DNA nucleosides susceptible to covalent modification.

Reactions with bisfunctional alkylating species can result in DNA cross-links [49-51]. If the electrophile attacks two nucleophilic sites on the same strand, it is referred to as an intrastrand cross-link. However, if the sites are on opposite strands it is referred to as an interstrand cross-link. Examples of these species are mitomycin [52, 53] and *cis*-platinum(II) diaminodichloride [54], which are used specifically in chemotherapy.

Mutagenesis

The insertion of non-complementary nucleotides by DNA polymerases opposite a lesion site is defined as a mutation. Mutations occur at a rate of 10^{-10} mutations per nucleotide per cell per generation in normal cells [14, 28]. The accretion of mutations in key genes, such as oncogenes and tumor suppressor genes results in carcinogenesis [55].

Multiple studies have shown a direct relationship between DNA damage and mutagenesis. Schaaper and Dunn showed that the presence of damaged nucleotides in template strands was a contributing factor to mutations occurring in both *E. coli* and

mammalian cells [58]. Mutagenicity can also result from depurination, as is evidenced by the increased misincorporation opposite abasic sites by DNA polymerases [59-61]. Mutations caused by oxygen free radicals have been observed in eukaryotic cells after exposure to high concentrations of oxygen [62]. The major mutagenic base lesion produced by oxidative damage is 8-hydroxy-guanine [22]. Instead of base-pairing with cytosine, the modified guanine will base pair with adenine, inducing mutations upon replication.

Carcinogenesis

The exposure of normal cells to physical or chemical carcinogens represents the first stage in the creation of cancer, or carcinogenesis [63]. DNA, which we know to be genetic material through the research of Avery *et al.* [2], is the cellular target for chemical carcinogens. When the extent of DNA damage increases beyond the ability of the DNA repair mechanisms, mutations accumulate. This leads to the promotion of genetic changes in cells, which can lead to abnormal cellular generation in comparison with surrounding unaltered cells [16].

The history of chemical carcinogenesis goes back to the 18th century when Dr. P. Pott discovered an epidemiologic connection between chimney sweeps and scrotal cancer [64]. This observation was confirmed experimentally by exposing rabbits to coal tar, and consequently, inducing cancer [65]. Polycyclic aromatic hydrocarbons, specifically benzo[*a*]pyrene [66], were identified as the carcinogenic components in coal tar [67]. Later it was reported that, in order for DNA to be modified by chemical carcinogens, activation to an electrophilic species must take place [68].

With the determination of the DNA structure, the mechanisms of chemical-mediated carcinogenesis have been a key focus of scientific investigation. Of particular interest in this laboratory is the examination of the mechanism of butadiene-mediated carcinogenesis.

Butadiene

1, 3-Butadiene (BD) is a colorless gas with a mildly aromatic odor [69]. BD has a melting point of -4.4 °C and a boiling point of -108 °C [70]. BD is slightly water soluble, but predominantly soluble in alcohol, ether, acetone, and benzene [70, 71]. The four-carbon small molecule with a molecular weight of 54.4 g/mol is generated as a by-product of ethylene production and is widely used industrially to make synthetic rubber and plastics including polystyrene, butadiene rubber, adiponitrile, and styrene-butadiene latex [72]. The industrial prevalence of BD increased during World War II as the synthetic rubber industry emerged [73]. By the early 1990's, BD was ranked among the top twenty synthetic organic chemicals produced in the United States with a production of over three billion pounds yearly. BD is also a widely spread environmental contaminant as traces have been found in side stream tobacco smoke, gasoline vapor, and automotive exhaust [74, 75]. Due to widespread human BD exposure, the mechanisms of BD-induced mutagenicity and carcinogenicity are of substantial interest.

Butadiene-Induced Carcinogenicity

Butadiene exposure comes from industrial and environmental sources and occurs primarily through inhalation. Inhalation studies have shown that butadiene induces

neoplastic lesions at multiple sites in B6C3F1 mice and Sprague-Dawley rats, with mice being more sensitive than rats [76, 77]. Upon exposure to BD, development of lymphomas, brain tumors, and mammary adenocarcinomas in mice has been observed [78, 79].

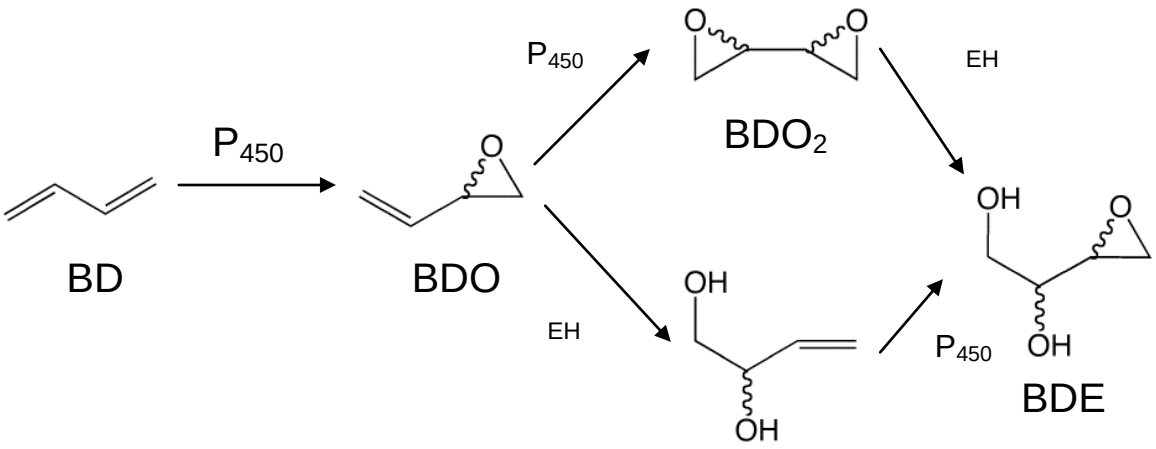
According to the National Toxicology Program, BD is a known human carcinogen [80]. The U.S. Environmental Protection Agency has classified BD as carcinogenic to humans by inhalation [81]. Results from epidemiology studies have shown a substantial amount of evidence linking occupational BD exposure with a higher risk of developing leukemia [82-84]. Although BD is known to be a probable human carcinogen, the exact means of BD-mediated carcinogenesis is not well known. In the mid-to-late 1980's, a considerable amount of literature was published describing BD carcinogenicity in laboratory animals [77, 85, 86]. Since then, a significant effort has been put forth to gaining a better understanding of the mechanisms of BD-induced carcinogenicity.

Metabolites of Butadiene

The two double bonds in BD which make it ideal for polymerization reactions also make it prone to epoxidation at two sites. It is believed that the epoxide butadiene metabolites, butadiene monoepoxide (BDO) and butadiene diepoxide (BDO_2), play a significant role in the carcinogenicity of BD by forming covalent adducts at nucleophilic sites in DNA bases. Evidence for this arises from data which suggests that *in vitro* mutagenesis studies involving BD require metabolic activation [87]. BDO and BDO_2 are also considered to be mutagenic in bacteria and carcinogenic in mice and rats [73].

The epoxide metabolites of BD were first observed in liver microsomes of rats upon exposure to butadiene [88]. 1,3-Butadiene is metabolized *in vivo* by specific isoforms of cytochrome P450 [89]. Results from Elfarra and coworkers indicated that 2E1 and 2A6 were the major isoforms responsible for BD oxidation in human liver microsomes [90]. BDO can be further metabolized by isoforms of cytochrome P450 including 2E1 and 3A4 to form BDO_2 [88, 91-93]. 1,2-Dihydroxy-3-butene is formed after hydrolysis of BDO by epoxide hydrolase [92, 94]. This can be further metabolized by cytochrome P450 to hydroxymethylvinyl ketone (HMVK) [95]. These secondary products can be further metabolized to 1,2-dihydroxy-3,4-epoxybutane (BDE) [92, 96]. A schematic of BD metabolism is shown in Scheme 1-1.

These epoxide metabolites of BD are highly reactive electrophiles capable of alkylating DNA bases and producing covalent modifications [97-99]. If left unrepaired in DNA, these modifications result in DNA base changes and deletions, which can lead to an overabundance of deleterious consequences including sister chromatid exchanges, micronuclei, and chromosomal aberrations [100-102]. The epoxide metabolites of BD were shown to be directly mutagenic in *Salmonella*, *Saccharomyces*, and *Drosophila* [103].



Scheme 1-1: The metabolism of butadiene.

BDO is the primary metabolite of butadiene formed both *in vivo* and *in vitro* [104-108]. Both the *R* and *S* stereoisomers of BDO have been shown to form in rabbits and humans and at similar rates in rat liver microsomes [104, 109]. BDO is known to be mutagenic and carcinogenic and is capable of inducing small deletions and point mutations at A:T and G:C base pairs [97,110]. BDO is also genotoxic in mice, producing chromosomal aberrations and sister chromatid exchanges *in vivo* [111].

BDO_2 is formed less *in vivo* than BDO, but is more genotoxic [112]. This is most likely due to the bisfunctional nature of BDO_2 and its ability to form cross-links with DNA and proteins. BDO_2 also induces point mutations, small deletions, and large scale deletions [113]. Data was reported showing that BDO_2 causes DNA cross-links between guanine bases [114]. The primary DNA cross-link involving BDO_2 occurs at the N7-dG site [115]. Studies have shown that BDO_2 is hundred-fold more mutagenic than BDO [97]. In mice, BDO_2 is produced in significantly higher amounts, which may be a cause for the higher sensitivity to BD mutagenicity and carcinogenicity in mice versus rats [72]. BDO_2 is also mutagenic in bacteria, yeast, and *Drosophila* germ cells [116-119].

Adducts Arising from Butadiene Metabolites

Adducts arising from BDO have been observed on all four DNA bases [120]. Under physiological conditions, the main sites of alkylation are N7-dG, N1-dA, N3-dA, and N3-dC [63]. Alkylation at dG is preferred [103]. The most reactive positions in DNA are the N7 site in deoxyguanine and the N3 and N^6 of adenine [121-123]. N7-dG adducts of BDO are also the major reaction products of BDO and BDO_2 with deoxyguanine in double-stranded calf thymus (CT) DNA [120, 124]. N^2-dG and N1-dG adducts have both been isolated *in vitro* [125, 126]. N3, N^6, N7, and N1 dA adducts have also been observed [103, 127]. The reaction of deoxycytidine with BDO was shown to result in several adducts including a diastereomeric pair of N3-dC, a diastereomeric pair of O2-dC, as well as the deamination products of the N3-dC adducts, diastereomeric pairs of N3-dU adducts [128]. N3-dT adducts were also observed *in vitro* [128].

The main adduct *in vivo* is the N7-dG adduct [63]. *In vivo* studies have shown N7-THB-dG, as well as other dG adducts arising from BDO, in the liver of mice exposed to BD [129]. Trace amounts of BDO- derived N7-dG adducts were found in liver DNA of rats [130]. N7-THB-dA and N6-THB-dA adducts have also been observed in rodents [127, 131, 132]. Currently, the only BD-derived adducts to have been isolated from humans is the N1-(2,3,4-trihydroxybutyl)-adenine adduct [133].

The N3-(2-Hydroxy-3-buten-1-yl)-2′-deoxyuridine Adducts

Alkylation of deoxycytosine has occurred with several electrophilic species and has resulted in deamination to deoxyuridine [134-137]. The electrophilic BDO attacks the N3 position of dC probably via an S_N2 reaction to form *R*- and *S*-N3-(2-hydroxy-3-

buten-1-yl)-2′-deoxycytosine adduct [138]. Selzer *et al.* were the first to identify the stereoisomeric N3-dC adducts of BDO which are generated upon exposure of nucleosides and calf thymus (CT) DNA to BD [120, 128]. These adducts are extremely unstable and readily convert to stereoisomeric N3-deoxyuridine (N3-dU) adducts via hydrolytic deamination [128].

Alkylation of dC by simple hydroxyalkyl epoxides, followed by hydrolytic deamination to dU, has been studied by Solomon and coworkers [134, 138]. N3-dU adducts of other reactive epoxides, including ethylene oxide and propylene oxide, are also formed via hydrolytic deamination of N3-dC adducts [138]. Upon alkylation at the N3 position, the pKa of dC increases from 4.3 to 8.8. Therefore, at a physiological pH, the imino group on the C4 carbon is protonated. This positive charge is delocalized with N3, creating a positive charge at the C4 position. A cyclic intermediate catalyzed by the hydroxyl group on the adduct is formed, and ammonia is released. Finally, the transient ring is hydrolytically reopened leaving a stable dU adduct [134, 138].

The N3-dU adduct is potentially mutagenic because of its stability. The half-life of the N3-dU adducts is 168 hours [128]. Therefore, the N3-dU adducts have the potential to remain stable in DNA long enough to affect not only the base pairing of the parent nucleoside, but to also potentially block DNA replication.

Recently, site-specific mutagenesis studies were carried out on a racemic mixture of the N3-dU diastereomeric adducts in the COS-7 mammalian cells [139]. The N3-dU adduct gave an overall mutation yield of 97% with 32.5% C to A, 53.4% C to T, and 11% C to G mutations. These results indicate that the stereoisomeric N3-dU adducts are some of the most mutagenic adducts arising from butadiene monoepoxide. Following

mutagenesis studies, DNA lesion bypass experiments were performed. The N3-dU adducts were toxic to replication polymerases δ, ε, ι and κ. A more significant finding was polymerase η was reported to misincorporate dNTPs at the lesion site, primarily dATP and dGTP. Pol η was not an efficient extender of the primer strand. Pol ζ, which was not able to insert dNTPs opposite the lesion site, was able to extend the primer to yield full-length extension products.

Studies of mutations in the *hprt* of rodents and humans, as well as mutations in *lac*I and *lac*Z mice have shown that G:C base pairs play a significant role in BD-induced mutagenesis [140]. Therefore, although BDO-induced N3-dU adducts represent minor adducts, their presence in cellular DNA may play a role in BD-mediated carcinogenesis. Several methods exist for the elucidation of information which might lead to a better understanding of butadiene-mediated carcinogenicity. These processes would fall into three categories: biological, biochemical, and structural.

Biological and Biochemical Studies of Carcinogen-Bound DNA

Site-Specific Mutagenesis Studies

Advances in organic synthesis have been beneficial to the progress of the site- and stereospecific production of reactive metabolites. This has afforded a greater investigation of their roles in mutagenic processes, including replication bypass of DNA lesions. These studies have the potential to identify individual DNA adducts which may be responsible for BD-induced mutagenesis.

The process of elucidating mutations caused by DNA lesions was conceived in the late 1970's [141]. The basic method involves site-specifically synthesizing a modified oligonucleotide primer. The next step requires hybridization with single-stranded DNA containing the specific gene of interest. Then the single-stranded fragment is extended by a DNA polymerase, and the emergent DNA is introduced into a host cell and cloned. Finally, the mutants are determined.

DNA Lesion Bypass Assays

The bypass of DNA lesions is a normal cellular process involving error-free and error-prone DNA polymerases [142]. The error-free mechanism does not result in mutations opposite the lesion site; however, the error-prone mechanism frequently introduces mutations. Following mutagenesis studies, *in vitro* lesion bypass assays provide further information on the DNA polymerases that are involved in bypassing DNA damage. There may be a direct correlation between the accretion of mutations and carcinogenesis. Therefore, understanding the mechanism of lesion bypass by specific DNA polymerases is a significant key to the understanding of butadiene-mediated carcinogenicity.

To identify lesion bypass DNA polymerases, certain polymerases are purified and their response to DNA damage is examined. The reactions contain a controlled amount of dNTPs, a ^{32}P-labeled primer, and one or multiple DNA polymerases. The components are incubated in a reaction buffer for a specified time and then quenched using a stop-solution, typically comprised of EDTA. The reaction products are then resolved on a polyacrylamide gel and visualized by audioradiography.

Structural Studies of Carcinogen-Bound DNA

Before the mid-1970's very few, if any, biomolecular structures were available. By the late 1980's that number began to increase, and the amount of available structures has grown to over 30, 000 [143]. Many of the structures that are available were determined by X-ray crystallography. The large size of proteins and protein complexes make X-ray crystallography a necessary approach to structural elucidation. Nuclear magnetic resonance (NMR) spectroscopy has also proven to be a powerful method for the three-dimensional structure determination of smaller biomolecules, like DNA. NMR spectroscopy provides a means for biomolecules to be determined in solution. This is a beneficial supplement to X-ray crystallography in that all biomolecules are naturally in aqueous environments and many of these molecules are difficult to crystallize.

NMR Spectroscopy

NMR methods make use of many phenomena. Two of those are through-bond interactions and through-space interactions. Through-bond interactions are scalar couplings through bonding electrons [144], while through-space interactions involve dipolar interactions between two different nuclei within approximately 5 Å of each other [145]. The two experiments providing through-bond and through-space information are correlated spectroscopy (COSY) and nuclear Overhauser effect spectroscopy (NOESY) NMR experiments, respectively. For NMR structural refinements, COSY data gives rise to information about coupling constants between various atoms. These coupling constants, through the Karplus equation, provide torsion angle restraints for the back-bone and sugars of the oligonucleotide [144]. NOESY interactions provide distance

restraints for both non-exchangeable and exchangeable protons [146]. Together, these experiments enable structural elucidation by restraining molecular dynamics calculations.

The NOE is dependent on many factors including molecular tumbling and internuclear distance. For a two-spin, rigid system, the NOE, at short mixing times, is proportional to r^{-6}, where r is the distance between two protons [145]. It is this relationship that makes NOESY data particularly useful in the structural refinement of biological macromolecules. The cross-peaks on a two-dimensional NOESY spectrum, arising from two spins, are inversely related to the distance between those spins.

NOE signals arise when the populations in the energy levels of two almost identical nuclei (spin = ½) are perturbed from equilibrium and then relaxed back to their equilibrium position during a specific mixing time [145], causing a change in intensity of one resonance. This perturbation of equilibrium occurs as a result of longitudinal spin relaxation, T_1, after a stimulus is applied to induce this transition. For NMR spectroscopy, the stimulus is a magnetic field that must be oscillating at the exact frequency of the spins [144].

This population transfer can be explained by looking at the effect of one spin, I, when irradiating a second spin, S. For a two-spin system, there are four energy levels corresponding to the α- and β- states of the individual spins. When the nuclei equilibrate, they do so via a specific pathway identified by a relaxation rate constant, W. The two cross-relaxation pathways are the transitions that are responsible for the NOE [144].

The rate of change for the difference between the intensity of the spin in the presence of the NOE, once it has been perturbed from equilibrium, I, and the intensity at

equilibrium, I_0, is proportional to the cross-relaxation rate and can be written as the Solomon equation [147]:

$$\frac{d(I_z(t) - I)_i}{dt} = -\rho_i(I_z(t)_I - I_{0i}) - \sum_{i \neq j} \sigma_{ij}(I_z(t)_j - I_{0j}),$$

where I(t) is the signal intensity at time t after perturbation has occurred and I_0 is the signal intensity at equilibrium. The Solomon equation describes the relaxation rate of the NMR signals which is proportional to the longitudinal relaxation rate, ρ, and the cross-relaxation rate, σ. The longitudinal relaxation rate and the cross-relaxation rate are both proportional to the spectral density function, J(ω). The spectral density function is a Fourier transform of the time-correlation function, which describes the random rotational diffusion of the spins. The spectral density function is inversely proportional to r^6. Assuming that the molecule is rigid and the only movement to be isotropic rotational tumbling, the spectral density function can be described by the following equation [145]:

$$J_{ij}(\omega) = \frac{1}{4\pi} \frac{1}{r^6} \left(\frac{\tau_c}{1 + \omega^2 \tau_c^2} \right).$$

From this equation we can see that the spectral density function is directly proportional to r^{-6} [145]. Because the cross-relaxation rate is directly related to the spectral density function and the signal intensities are proportional to the cross-relaxation rates, we can

say that the signal intensities are directly proportional to the inverse of r^6, establishing a distance dependence. Through this relationship, NMR refinement of duplex DNA is possible, in part, by establishing distance constraints between two spins whose cross-peaks can be assigned on a two-dimensional NOESY spectrum.

NMR Refinement of Duplex DNA

The NMR refinement process is the combination of experimental NMR restraints and force-field simulations to elucidate a macromolecular structure. This is typically done through molecular dynamic simulations [145]. Figure 1-6 lays out the necessary steps involved in NMR structural refinement.

The structural determination and refinement of modified duplex DNA begins with the acquisition and assignment of two-dimensional NMR spectra. DNA contains an overwhelming amount of hydrogen atoms that are observable in NMR spectra. In order to reduce the complexity of assigning the many protons, the proton resonances are spread out in a two-dimensional spectrum. Another means of spectrum simplification involves the use of deuterated solvent, most commonly D_2O. Labile protons covalently bonded to nitrogen and oxygen atoms in DNA rapidly exchange with deuterium, thereby removing many of the protons in DNA. The remaining non-exchangeable protons are more easily assigned under these circumstances.

NOESY experiments are primarily used for the assignment of non-exchangeable and exchangeable protons. The protons have fingerprint chemical shift values and they are located in the diagonal of the spectrum. The chemical shift values of specific oligonucleotide protons are separated into specific regions in the NOESY spectrum. To

each side of the diagonal, the cross-peaks represent a dipolar interaction between two protons. Once each proton is assigned, the chemical shift values can be used to identify the internuclear interactions producing the cross-peaks [146]. These regions are illustrated in the NOESY spectrum shown in Figure 1-7.

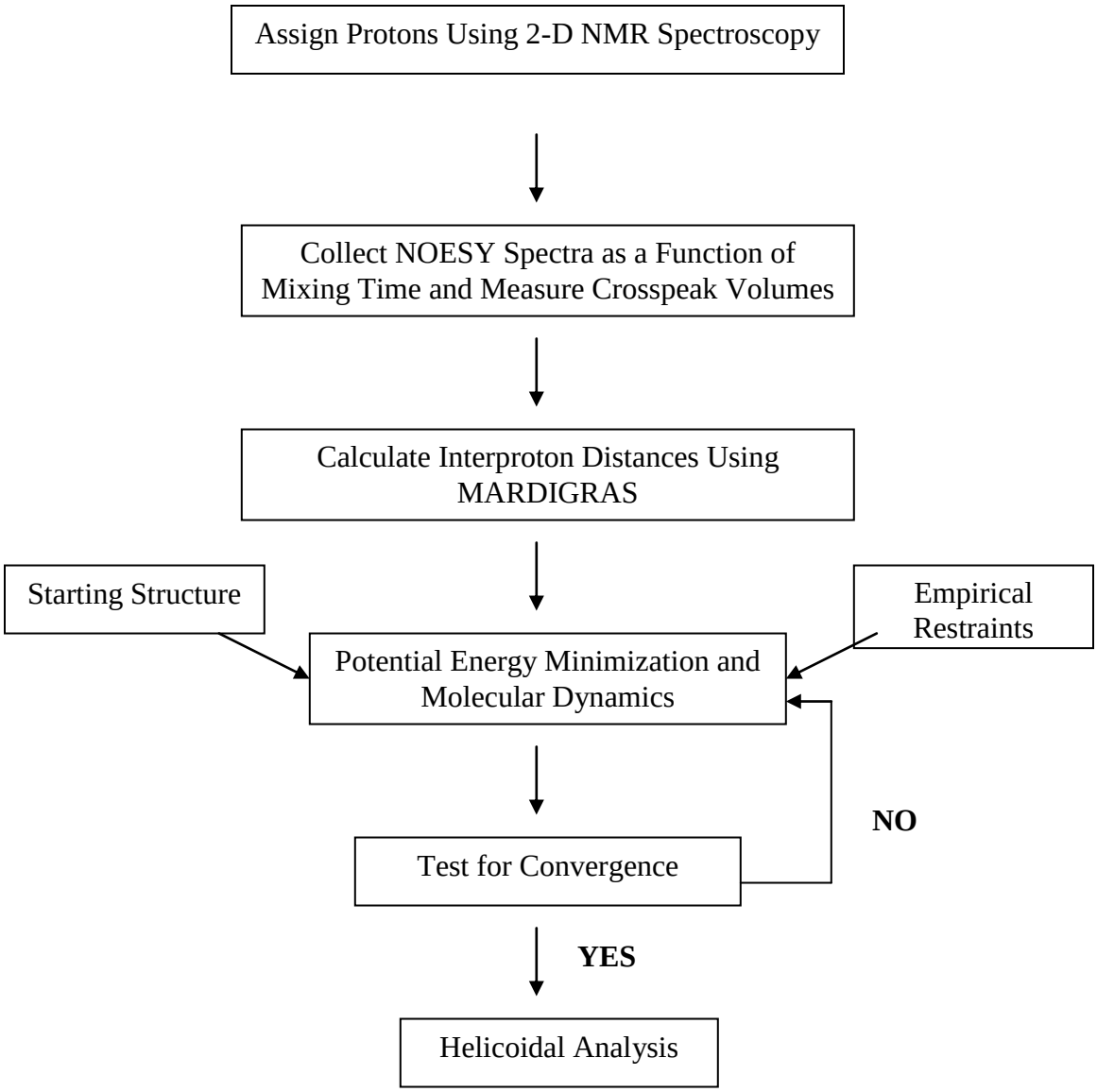


Figure 1-6: Strategy for NMR-derived solution structure determination.

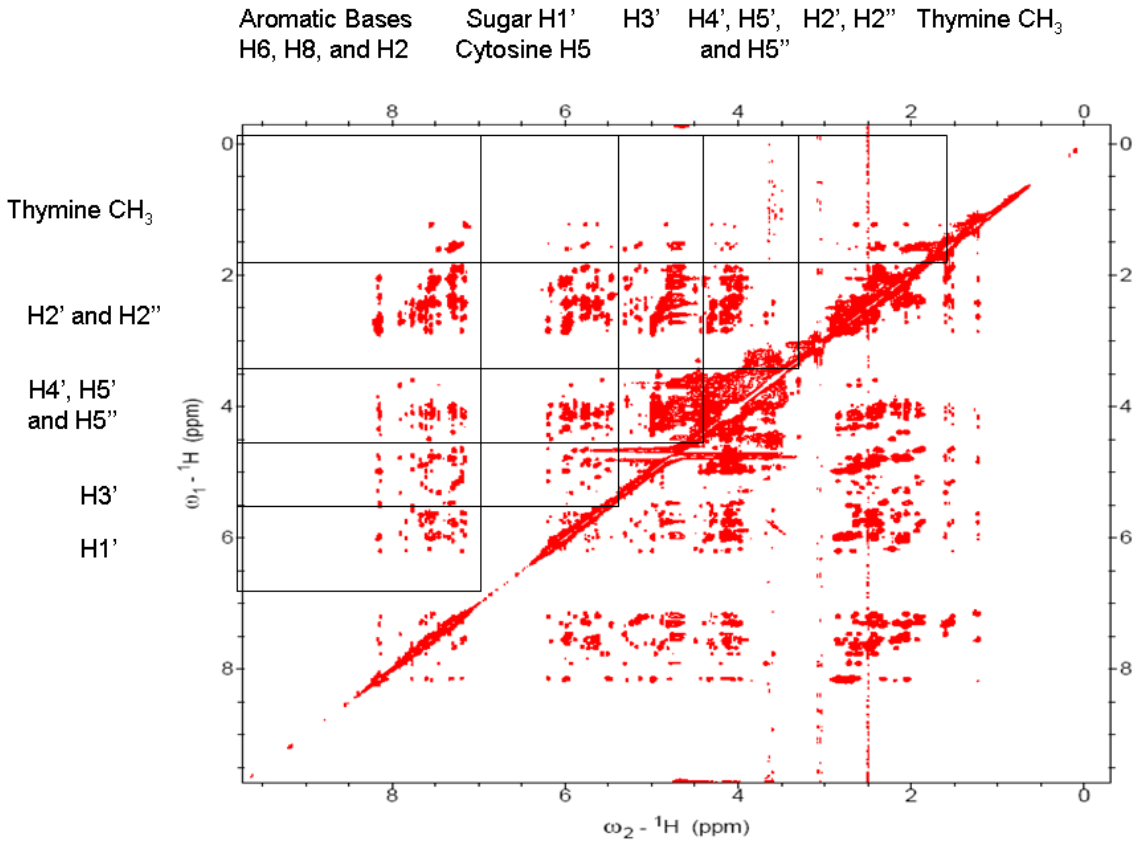


Figure 1-7: Chemical shift regions for non-exchangeable DNA protons in a two-dimensional NOESY spectrum.

Key to assigning NOESY spectra are the short (≤ 5 Å) distances between the aromatic proton of one nucleotide and the H1′, H2′, and H2″ protons of its 3′-neighboring base [148]. The interactions between these protons allow one to "walk" down the oligonucleotide beginning at the 5′-end and continuing through the entirety of the strand. The NOESY "walk" is illustrated in Figure 1-8. In addition to the NOESY data, several COSY experiments are used to complete the assignments.

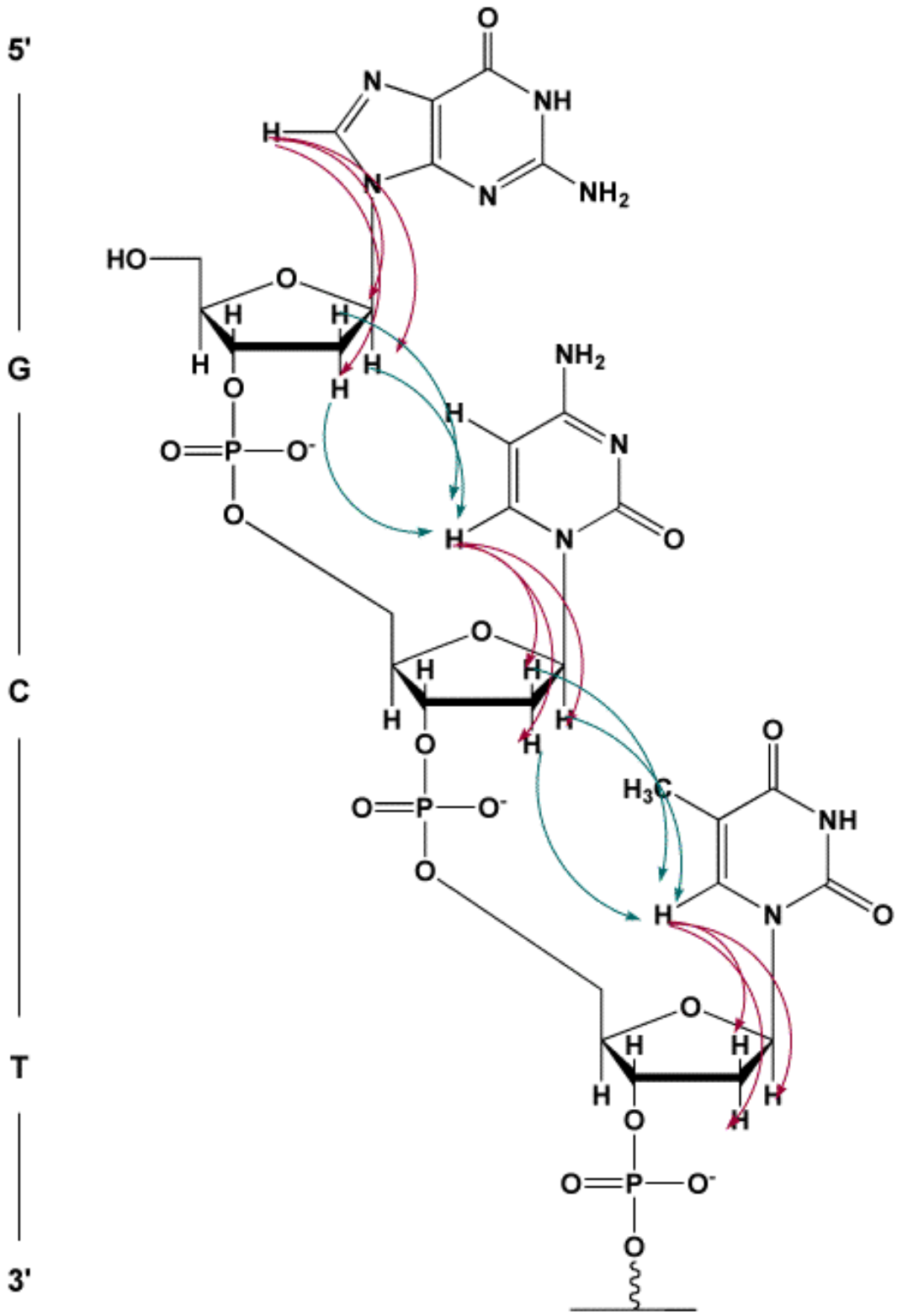


Figure 1-8: Dipolar interactions between aromatic and anomeric DNA protons observed in a NOESY spectrum. A NOESY spectrum can be assigned by "walking" from the 5′-end of the oligonucleotide to the 3′-end through intra-(red) and inter-nuclear (green) ^{1}H-^{1}H interactions.

In order to observe those protons that are exchangeable with deuterium, NOESY experiments are also acquired in a 90:10 H_2O:D_2O sample. This experiment provides information related to the hydrogen bonding of base-pairs. Because the stability provided by Watson-Crick hydrogen bonds decreases the rate of exchange of the imino protons with deuterium, imino protons stabilized by Watson-Crick hydrogen-bonding will produce a strong, sharp peak, observed on the diagonal of a two-dimensional spectrum. If Watson-Crick hydrogen bonding is disrupted, a weaker, broadened peak may be observed as a result of a faster rate of exchange.

Once all of the assignments are completed, the cross-peak volumes are measured using a Gaussian function and calculated into inter-proton distance restraints using the program, MARDIGRAS [149, 150]. These distance restraint files are combined with experimental and empirical torsion angle restraints to form a complete restraint file for MD calculations. Additional restraints, based on force-field models, can be used in order to refine the structure [145].

According to Beveridge and Ravishanker, molecular dynamics simulation is a necessary tool for the investigation of the structure of biological macromolecules and their motions [151]. To begin an MD simulation, an initial starting structure is necessary as a reference. This is typically the crystal structure of the system under investigation; however, it can be a built model. After a structure is obtained, it undergoes potential energy minimization (PEM) in order to relax the bonds and remove any unnecessary stress on the molecule [151]. The molecule then undergoes simulated annealing utilizing a defined force field. Force fields represents each atom in the system and provide the terms for bonds, bond angles, torsion angles, tetrahedral and planar geometry, hydrogen

bonding, and nonbonding interactions, including van der Waals and electrostatic forces [152]. The calculations described in this thesis use the AMBER 8 [153] force field.

During the simulated annealing protocol of MD calculations, the temperature is rapidly raised to 600K allowing the molecule to traverse high energy barriers in order to find a global minimum. After the temperatures have reached a high they are slowly brought down, usually by being tightly coupled to a heating bath. Once the temperature has been lowered, the simulation can sample the minima.

The purpose of running MD simulations is to find the global minimum region of a specific energy function, E_{tot}, which is given by the equation $E_{tot} = E_{cov} + E_{vdw} + E_{NMR}$, where the terms in the equation correspond to covalent interactions, non-bonded interactions and NMR experimental constraints, respectively [154]. The potential energy function, E_{NMR}, can be written as the differences between the distances present in the MD simulations and the distant constraints from the NMR experiments. To the potential energy function, upper and lower limits are given, and pre-determined force constants are applied, as a harmonic restraining function, with the intentions of keeping the distances within the upper and lower limits.

MD calculations identify a low energy state for the molecule, close to the initial reference structure, and in agreement with the experimental data [151]. The calculations are carried out on multiple starting structures. After rMD calculations, the emerging structure coordinate files are tested for convergence using pairwise root-mean-square-deviation (rmsd). The accuracy of the final structure is tested using complete relaxation matrix analysis, or CORMA [155]. Here, the theoretical intensities from the structure coordinate files are compared with the actual intensity files from NMR data. Finally,

helicoidal analysis is performed using the programs, 3DNA[156] and CURVES [157, 158]. The end result is a refined solution structure that can be used to further understand the chemistry of the biological macromolecule.

Dissertation Statement

This dissertation will discuss the elucidation of solution structures of stereoisomeric N3-(2-hydroxy-3-buten-1-yl)-2′-deoxyuridine (N3-dU) adducts site-specifically incorporated into three oligonucleotides. The adducts were chosen in order to better understand butadiene carcinogenicity at a molecular level. The N3-dU adducts have not been isolated *in vivo*, indicating that their presence in cellular DNA may only be at very low levels. They are of interest, however, because they are highly mutagenic in mammalian cells, yielding a majority of C to T transitions and C to A transversions. Structural comparisons of both diastereomers is relevant as the *R* diastereomer showed a 2:1 preference for C to T transitions versus C to A transversions, while the *S* diastereomer allowed equal numbers of both mutations. The three oligonucleotides contain the N3-dU adduct opposite adenine, thymine, and guanine, representing the C to T transition, the C to A transversion, and the non-mutated conditions, respectively. The goal of this dissertation is to discuss the solution structure of these adducts in a duplex DNA environment, relating the structures to our current understanding of their mutagenic potential.

The materials and methods used in this dissertation are detailed in Chapter II. Chapter III describes the structures obtained for the *R*- and *S*-BD-N3-dU modified oligonucleotide, 5'-d(GCTAGCXAGTCC)-3'•5'-d(GGACTAGCTAGC)-3', in which X=

R- or *S*-*BD*-N3-dU. These oligonucleotides model the misincorporation of dATP opposite the *R*- or *S*-N3-dU adducts, which if not recognized and repaired, would generate C to T transitions. Chapter IV discusses NMR structural studies of the stereoisomeric N3-dU adducts mismatched opposite thymine. These oligonucleotides model the misincorporation of dTTP opposite the *R*- or *S*-BD-N3-dU adducts, which if not recognized and repaired, would generate C to A transitions. Chapter V provides details about NMR solution structures obtained for the stereoisomeric butadiene N3-dU adducts placed opposite guanine. These model the incorporation of dGTP opposite the *R*- or *S*-BD-N3-dU adducts, which result in a non-mutagenic outcome. Finally, Chapter VI provides general conclusions and future directions.

CHAPTER II

MATERIALS AND METHODS

Oligodeoxynucleotide Synthesis

The unmodified oligodeoxynucleotides were synthesized by the Midland Certified Reagent Co. (Midland, TX) and purified by anion-exchange high performance liquid chromatography (AE-HPLC). The *R*- and *S*-BD-N3-dU modified 12-mer oligodeoxynucleotides 5'-d(GCTAGCXAGTCC)-3' were synthesized by Dr. Richard Hodge at the University of Texas Medical Branch. Dr. Hodge site-specifically incorporated N3-(3-butene-2-O-*tert*-butyldimethylsilyl)-2'-deoxyuridine at the seventh position of the 12-mer using phosphoramidite chemistry and standard oligonucleotide synthesis protocol as described [159].

Oligodeoxynucleotide Purification

The modified oligodeoxynucleotide 5'-d(GCTAGCXAGTCC)-3' was determined via enzyme hydrolysis and HPLC analysis to be a 60:40 mixture of *R* and *S* diastereomers, respectively. Separation of the oligodeoxynucleotides was conducted on a Beckman HPLC system (32 Karat software) with a diode array UV detector monitoring at 260 nm. A Waters YMC column was used with triethyl ammonium acetate and acetonitrile. A flow rate of 5 mL/min was used with the following gradient: 1-8% acetonitrile over 5 min, 8-12% acetonitrile over 25 min, 12-80% acetonitrile over 2 min, and then to 1% acetonitrile over 3 min. Because baseline separation of the two

oligonucleotides was not achieved, in order to attain maximum purity the samples were collected in three vials labeled “A”, “B”, and “C”, representing the first peak, the mixture, and the second peak, respectively. Peaks “A” and “C” were later identified as the *S* and *R* diastereomers, respectively, based on standard nucleosides. Figure 2-1 shows the HPLC trace of the enantiomeric separation.

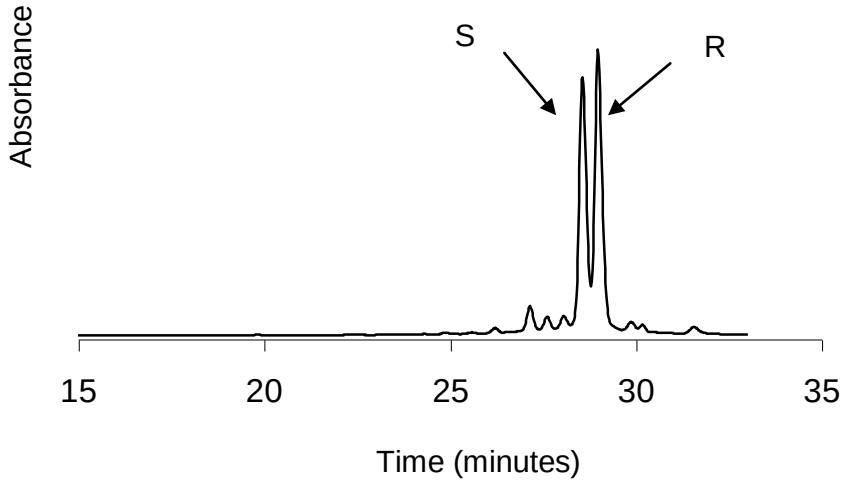


Figure 2-1: HPLC separation of the stereoisomeric BD-N3-dU modified oligodeoxynucleotides.

After elution from the HPLC, to each of the samples, 15 mL of 0.3M NaCl solution was added. The samples were incubated in the salt solution for 30 minutes at 4 °C before being lyophilized. The oligonucleotides were then resuspended in 1 mL of water and desalted using a G-25 Sephadex column in order to remove trace amounts of the triethyl ammonium acetate buffer system. Enzyme hydrolysis and HPLC analysis showed that both diastereomers were 96% pure. The HPLC traces of both diastereomers are shown in Figure 2-2.

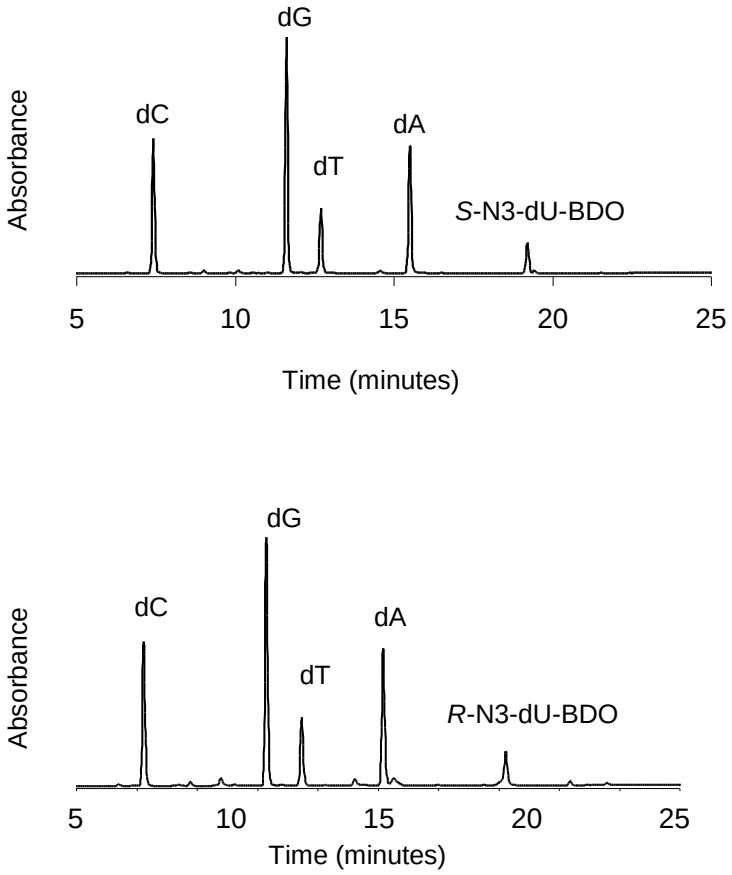


Figure 2-2: HPLC traces of nucleosides digested via enzyme hydrolysis reactions. (A) Nucleosides comprising the *S*-BD-N3-dU modified oligonucleotide, 5′-G[1]C[2]T[3]A[4]G[5]C[6]X[7]A[8]G[9]T[10]C[11]C[12]-3′. (B) Nucleosides comprising the *R*-BD-N3-dU modified oligonucleotide, 5′-G[1]C[2]T[3]A[4]G[5]C[6]X[7]A[8]G[9]T[10]C[11]C[12]-3′.

Sample Preparation

The concentration of the oligonucleotides was determined by calculating the optical density (OD) of the oligonucleotide strand at 254 nm using a Varian Cary 4E spectrophotometer. The ratio of nmol:OD was calculated to be 9.11 for the modified oligonucleotide. For the complementary strand, 5'-d(GGACTYGCTAGC)-3', the nmol:OD ratios were 8.51, 8.96, and 8.80, with Y being defined as adenine, thymine, and guanine, respectively.

For all six oligonucleotides, the *R* and *S* modified strands were each annealed to the appropriate complementary strands using a 1:1 molar ratio of the two single-stranded oligonucleotides. In a 1 mL Eppendorf tube, 0.5 μmol of each strand of oligonucleotide was added to 400 μL of phosphate buffer (10 mM NaH_2PO_4, 100 mM NaCl, 50 μM Na_2EDTA, pH 7.0). The solution sample was heated to 52 °C in a hot water bath for 15 minutes and slowly cooled to room temperature. A DNA Grade Biogel hydroxylapetite column (Bio-Rad Laboratories, Hercules, CA) was used to elute duplex DNA from a mixture of single-stranded and duplex DNA using a gradient from 10 to 100 mM NaH_2PO_4. The duplex DNA samples were lyophilized overnight, resuspended in 1 mL of H_2O and desalted on a G-25 Sephadex column. The purity of the samples was checked using MALDI-TOF mass spectrometry. The mass spectra were acquired in a negative reflector mode using a Voyager-DE (PerSeptive Biosystems, Inc.). The samples were fixed in a matrix containing 0.5 M 3-hydroxypicolinic acid and 0.1M ammonium citrate.

The modified oligodeoxynucleotide samples were lyophilized overnight and later dissolved in 500 μL of an NMR buffer (200 mM NaH_2PO_4, 100 mM NaCl, and 50 μM Na_2EDTA, pH 7.2) and 200 μL of internal standard (1 μM 3-(Trimethylsylil) propionic-$2,2,3,3_{d4}$-acid, TSP). For observing non-exchangeable protons, the samples were dissolved three times in 99.9% D_2O, followed by lyophilization. Finally, the samples were resuspended in 500 μL of 99.96% D_2O and transferred to a 500 mm NMR tube. For observing exchangeable protons, the samples were lyophilized and dissolved in a 9:1 ratio of H_2O and D_2O.

NMR Spectroscopy

For each of the studies, all of the NMR experiments were carried out at a frequency of 800.23 MHz and a temperature of 288K for non-exchangeable protons and 279K for exchangeable protons. Magnitude COSY spectra were acquired using 512 data points in the d1 dimension and 2048 data points in the d2 dimension. The indirect dimension was zero-filled causing the final matrix to be 2048 x 2048. A relaxation delay of 2.0 s was used and the sweep width was set to 10 ppm. Water suppression was accomplished using pre-saturation.

NOESY experiments for non-exchangeable protons were performed at mixing times of 250, 200, 150, and 80 ms sequentially without being removed from the spectrometer. At all four mixing times, the spectra were acquired with 512 data points in the d1 dimension and 2048 data points in the d2 dimension. The indirect dimension was zero-filled achieving a 2048 x 2048 overall matrix. The skewed sine-bell squared apodization was utilized with a 90° phase shift. States-TPPI quadrature detection was used. A relaxation delay of 2.0 s was used and the sweep width was set to 10 ppm. A square pre-saturation pulse was used for the purpose of water suppression. The chemical shift values were referenced to the water peak at 4.7 ppm. NOESY data for exchangeable protons were acquired at a mixing time of 250 ms. Spectra were recorded with a sweep width of 25 ppm and a relaxation delay of 1.5 s was used.

^{1}H-^{1}H TOCSY experiments were performed at mixing times of 80 and 120 ms. At both mixing times, the spectra were acquired with 512 data points in the d1 dimension and 2048 data points in the d2 dimension. The indirect dimension was zero-filled achieving a 2048 x 2048 overall matrix. The skewed sine-bell squared apodization was

utilized with a 90° phase shift. States-TPPI quadrature detection was used. A relaxation delay of 2.0 s was used and the sweep width was set to 10 ppm.

DQF-COSY spectra were obtained using 512 data points in the d1 dimension and 2048 data points in the d2 dimension. The indirect dimension was zero-filled achieving a 2048 x 2048 overall matrix. The skewed sine-bell squared apodization was utilized with a 180° phase shift. States-TPPI quadrature detection was used. A relaxation delay of 2.0 s was used and the sweep width was set to 10 ppm.

NMR data were processed and phased using XWINNMR (Bruker Biospin, Germany) and TopSpin (Bruker Biospin, Germany) on Silicon Graphics Octane workstations (Silicon Graphics, Inc., Mountain View, CA).

Structural Refinement

The chemical shift values of the oligonucleotide protons were assigned in the NOESY spectra for all four mixing times using the program SPARKY [160]. After each cross-peak was assigned, they were volume integrated using a Gaussian function in order to determine peak intensities. The values were saved in an intensity file for each mixing time. Along with the intensity files, PBD files of B-form DNA site-specifically modified with the stereoisomeric adducts were used to generate lower and upper bound distance restraints for molecular dynamics calculations using the program MARDIGRAS [149, 150]. MARDIGRAS calculations were run at isotropic correlation times of 2, 3, 4, 5, and 6 ns for both the base and sugar protons and at all four mixing times, producing twenty distance sets.

The output files were compared to canonical B-DNA internuclear distances and the ten best distance sets, determined by standard deviation, were averaged to give upper and lower bound distance restraints. These sets of distance restraints were divided into five classes based on the level of confidence in spectral assignment and integration. To the distance restraint file marked with the highest level of confidence, a force constant of 32 kcal mol^{-1} $Å^{-2}$ was assigned. To the subsequent distance restraint files, force constants of 29, 26, 23, and 20 kcal mol^{-1} $Å^{-2}$ were given in descending order. Empirical Watson-Crick distance restraints were used for all of the base pairs, excluding X^7:A^{18} and A^8:T^{17}, based on NOESY data suggesting a disruption in hydrogen-bonding at these sites.

Empirical restraints were used for backbone torsion angle and sugar pucker restraints, based on NOESY data suggesting B-form DNA with C2'-endo sugar geometries. Weak H2''-H3' and H3'-H4' cross-peaks in the COSY spectra, as well as weak H2''-H4' cross-peaks in the NOESY spectra indicated C2'-endo sugar conformations for each nucleotide [161, 162]. Based on this evidence, pseudorotational phase angles were determined to be in the range of 90° and 165°. The sugar torsion angles, ν_0-ν_4, were determined from these pseudorotational phase angle restraints. NOESY cross-peaks representing intranuclear interactions between base aromatic and sugar protons suggested that for all of the nucleotides, the glycosidic bond was in the *anti* conformation [155].

Restrained Molecular Dynamics

Twenty starting structures were created using the MKDNA program written by Dr. Jarrod Smith (Department of Biochemistry, Vanderbilt University) in the nucleic acid builder (NAB) suite [163]. The structures were created using differing values for the

helical twist, rise, inclination, and x-offset, falling within the boundaries of canonical A- and B-form DNA. The BD N3-dU adduct was built in xLEaP [153] and saved as the Object File Format file, DX.off. The atomic partial charges of the modified base were determined using GAUSSIAN 03 [164] and edited in the xLEaP file. The modified base was added to the unmodified starting structures using xLEaP [153]. Finally, the chirality of the Cβ carbon of the butadiene moiety in the modified base of each starting structure was distinguished using the program INSIGHT II (Acelrys, Inc.) The reference structures were energy minimized, restraining heavy atoms, with no added experimental restraints by the steepest descent algorithm for the first 10 steps followed by the conjugant gradients algorithm for a total of 200 iterations. The Sander module of AMBER 8.0 [153] was used for PEM.

The Sander module of the AMBER 8.0 force field [153] was used for molecular dynamics calculations, utilizing the simulated annealing protocol. The calculations were performed using the Generalized Born approach [165, 166] in which implicit solvent molecules are added to represent an aqueous environment. The potential energy function contained an NMR restraint component with terms used to describe the distance and angle restraints. Bond lengths involving hydrogen atoms were fixed using the SHAKE algorithm [167]. The cutoff radius for non-bonded interactions was 10 Å.

Calculations were initiated from twenty randomly seeded starting structures by coupling to a heating bath with a target temperature of 600 K. Simulated annealing calculations were carried out for 20 ps using NMR experimental distance restraints, empirical hydrogen-bond distance restraints, and empirical angle restraints. The temperature was rapidly brought to 600 K for the first 5000 steps, the temperature was

then slowly lowered to 100 K over the next 13,000 steps. For the final 2,000 steps, the molecule was cooled to a target temperature of 0 K.

Structure coordinates emerging from rMD calculations underwent PEM using the steepest descent algorithm for the first 10 steps followed by the conjugant gradients algorithm for a total of 200 iterations. CORMA [155] was used to perform back-calculations on the theoretical NMR intensities from the determined structures. Helicoidal analysis of the backbone was carried out using CURVES [157, 158].

CHAPTER III

SOLUTION STRUCTURES OF THE STEREOISOMERIC N3-(2-HYDROXY-3-BUTEN-1-YL)-2′-DEOXYURIDINE ADDUCTS ARISING FROM BUTADIENE MONOEPOXIDE OPPOSITE DEOXYADENINE IN A DODECAMER OLIGONUCLEOTIDE

Introduction

In the presence of DNA, the epoxide metabolites of butadiene are capable of alkylating nucleophilic sites on all four DNA bases forming several adducts [120]. These DNA lesions may play a role in butadiene-mediated carcinogenicity. Among several sites of possible covalent interaction, the electrophilic BDO has been shown to attack the N3 of dC probably via an S_N2 reaction to form stereoisomeric N3-(2-hydroxy-3-buten-1-yl)-2′-deoxycytidine adducts [128]. Selzer *et al.* were the first to report the N3-dC adduct of BDO upon exposure of nucleosides as well as single-stranded and double-stranded calf thymus (CT) DNA to BDO. The BD-N3-dC adducts have half-lives of less than three hours and will form the stereoisomeric N3-(2-hydroxy-3-buten-1-yl)-2′-deoxyuridine (N3-dU) adducts via hydrolytic deamination. Alternatively, the BD-N3-dU adducts are stable up to 168 hours [128]. The chemical structures and nomenclature of the *R*- and *S*-BD-N3-dU adducts are shown in Figure 3-1.

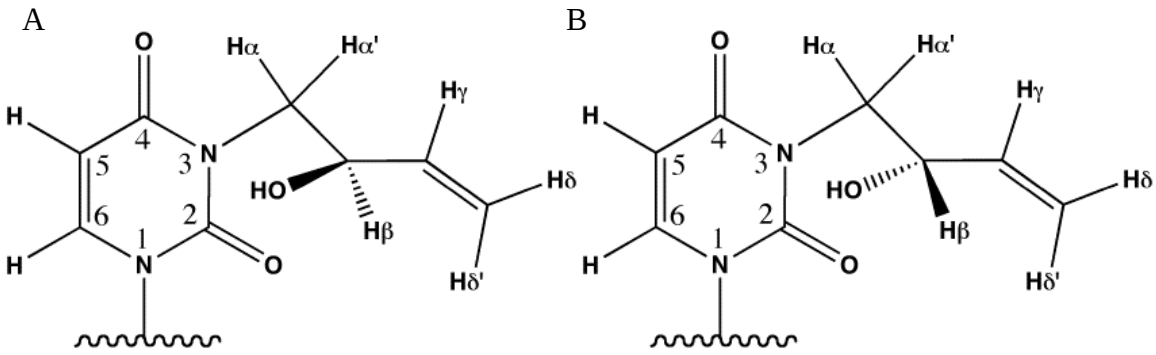


Figure 3-1: Chemical structure and nomenclature of the stereoisomeric N3-(2-hydroxy-3-buten-1-yl)-2′-deoxyuridine adducts. (A) The *R*-BD-N3-dU diastereomer. (B) The *S*-BD-N3-dU diastereomer. The pro-*S* and pro-*R* protona at the alpha carbon are designated as Hα and Hα′, respectively.

With a half-life on the order of seven days, it could be speculated that the BD-N3-dU adducts would remain in DNA long enough to induce mutations. Mutagenesis studies have been performed in COS-7 mammalian cells for both isomers [139]. These data confirm this suspicion, showing that the stereoisomeric adducts are highly mutagenic, producing a majority of C to T and C to A mutations. Stereoisomeric differences arise with respect to mutagenesis. While the *S*-BD-N3-dU adduct allows equal numbers of C to T and C to A mutations, the *R*-BD-N3-dU adduct has a 2:1 preference for the former [168]. The mutation frequencies of each stereoisomer are shown in Figure 3-2.

In light of the mutagenesis results, single nucleotide insertion and extension studies were performed on the mixture of adducts in order to classify the polymerases involved in bypassing the lesions [139]. The N3-dU adducts were toxic to replication polymerases δ, ε, and trans-lesion synthase polymerases, ι and κ. Polymerase bypass data indicated that polymerase η incorporates nucleotides opposite these adducts with the ability, along with polymerase ζ, to extend primers with mispaired nucleotides opposite the lesions. Pol η shows a preference for insertion of dGTP and dATP, which

corresponds with the mutagenesis data, suggesting that pol η may be one of the trans-lesion polymerases involved in replication past the stereoisomeric BD-N3-dU adducts [139].

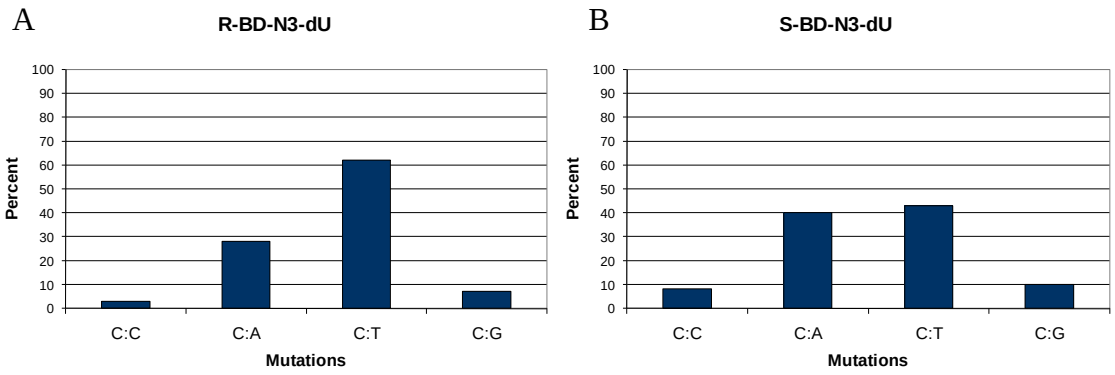


Figure 3-2: Mutagenic frequencies of the *R*- and *S*-N3-(2-hydroxy-3-buten-1-yl)-2′-deoxyuridine adducts in COS-7 mammalian cells. (A) Mutation percentages for the *R*-BD-N3-dU adduct. (B) Mutation percentages for the *S*-BD-N3-dU adduct. The adducts cause a majority of C:T and C:A mutations. The data suggests that the *R* isomer has a 2:1 preference for C:T transitions over C:A transversions, while the *S* isomer results in an equal number of both mutations.

Detection of the N3-dU adducts has not occurred yet *in vivo*, suggesting that these adducts may be present only at very low levels in cellular DNA. The adduct is found at a hydrogen-bonding site N3; therefore, steric hindrance would make it difficult for BDO to gain access to N3 in order to alkylate the deoxycytidine *in vivo* [63]. The N3-dU adducts are observed in both single-stranded and double-stranded CT DNA; however, there is a significant preference for single-stranded CT DNA, which would suggest that hydrogen-bonding does present a challenge to alkylation at that site [128]. Although only minor amounts of these adducts may be present in cells, they have the potential, if replicated, to be highly mutagenic.

Therefore, structural studies of the stereoisomeric BD-N3-dU adducts in duplex DNA have been carried out in order to better understand the role of these adducts in BD-

mediated mutagenicity and carcinogenicity. Stereospecific structural differences were observed by Merritt, *et al.*, between *R,R*- and *S,S*-N^6-THB-dA adducts, each having significantly different mutagenic frequencies in bacteria [89, 169]. The differential orientation of the butadiene lesions served to illustrate the role of adduct stereochemistry in biological processing. Other studies have corroborated the suggestion that stereochemistry plays a significant role in determining biological responses to damaged DNA bases [109, 170, 171].

Structural differences between the two enantiomers might explain the preferential insertion of dATP caused by the *R*-isomer in comparison to the *S*-isomer. Understanding the orientation of the butadiene moiety within DNA might provide an explanation for the misincorporation of nucleotides at the lesion site during replication. It is possible that during error-prone replication bypass by trans-lesion polymerases in COS-7 cells, the primer-template containing the *R*-BD-N3-dU adduct has a structural preference for the insertion of dATP versus dTTP, while the primer-template containing the *S*-BD-N3-dU adduct is oriented such that incorporation of either dATP or dTTP is permitted.

It was of interest, therefore, to determine the three-dimensional structure of the *R* and *S* isomers site-specifically incorporated into the dodecamer oligonucleotides, 5′-d($G^1C^2T^3A^4G^5C^6\underline{X}^7A^8G^9T^{10}C^{11}C^{12}$)-3′·5′-d($G^{13}G^{14}A^{15}C^{16}T^{17}\underline{A}^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$)-3′, where X is either the *R*- or *S*- BD-N3-dU DNA adduct. When choosing oligonucleotides for studies by NMR spectroscopy, it is important to select an oligonucleotide with a sequence that will produce quality NMR spectra. The dodecamer sequence above is one that has been studied with various elements of DNA damage in this laboratory and was chosen for this reason.

Two-dimensional NMR experiments were used to elucidate changes in the chemical environment of the *S* and *R* oligonucleotides as a result of modification at the lesion site, as well as provide experimental restraints for molecular dynamics calculations. When the lesions are placed opposite dA, the oligonucleotides represent the misincorporation of dATP, which if left unrepaired would result in a C to T transition.

NMR analysis reveals that for both adducts there is an increase in the rise between the adduct and its 3′- and 5′-neighboring bases. In both structures, the Watson-Crick hydrogen bonding is perturbed at the lesion site and at the 3′-neighboring base pair, A^{8}:T^{17}. For the structure containing the *R*-BD-N3-dU adduct, the butadiene moiety is oriented toward the minor groove in the 5′-direction. A hydrogen bond is predicted by MD calculations to exist between the hydroxyl group proton at Cβ and the O^{4} atom of the modified base. For the structure containing the *S*-BD-N3-dU adduct, the butadiene moiety is oriented in the minor groove toward the complementary strand. This forces the adenine opposite the lesion, A^{18}, into the major groove. The potential for hydrogen-bonding is predicted between the Cβ hydroxy group proton and the O^{2} atom of the adduct. The structural differences in the orientation of the butadiene moiety may explain the difference in nucleotide insertion preferences noted in the mutagenesis studies. This may indicate that the geometry of a lesion plays a role in the outcome of nucleotide insertion by DNA polymerases.

Results

R-Diastereomer

Sample Purity

The double-stranded *R*-BD-N3-dU modified oligonucleotide, 5′-d($G^1C^2T^3A^4G^5C^6\underline{X}^7A^8G^9T^{10}C^{11}C^{12}$)-3′·5′-d($G^{13}G^{14}A^{15}C^{16}T^{17}A^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$)-3′, was purified using the HPLC protocol described in Chapter II. Enzyme hydrolysis showed that the modified oligonucleotide was enantiomerically pure. The annealed duplex was separated from any remaining single-stranded DNA using a hydroxylapatite column. The presence of only duplex DNA in the solution sample was confirmed using capillary gel electrophoresis. The composition of the two single strands was verified using MALDI-TOF mass spectrometry and NMR spectroscopy. NMR experiments at a frequency of 800.23 MHz yielded spectra that provided valuable data.

Non-exchangeable protons

For the observation of non-exchangeable protons, the sample was prepared as discussed in Chapter II. Assignments of the purine and pyrimidine aromatic protons, the thymine methyl protons and the deoxyribose H1′, H2′, H2′′, and H3′ protons were successfully completed. Due to overlapping peaks and the effects of spin diffusion at higher mixing times, the H4′, H5′, and H5′′ protons were only partially assigned. The chemical shift values of the non-exchangeable protons are listed in Appendix A. The intensities of the cross-peaks in the NOESY spectrum indicate that the structure is similar to that of canonical B-DNA structure. Expanded plots of the NOESY spectrum for the *R*-diastereomer are shown in Figure 3-3. The expanded region of this spectrum shows cross-peaks resulting from intra- and internuclear dipolar interactions between purine and pyrimidine aromatic protons and the deoxyribose H1′ protons. In this region, nucleotides

can be sequentially assigned by beginning at the 5′ end of the strand and walking to the 3′-neighbor, as described in Chapter I.

In the modified strand of the *R* stereoisomer (Figure 3-3A), the NOE connectivity remained uninterrupted throughout the entire strand. The interaction between C^6 H1′ and X^7 H6, however, produced a weak cross-peak, suggesting that the distance between these two protons was increased as a result of the modification. We also observed a decreased intensity in the cross-peak for the interaction between X^7 H1′ and A^8 H8. In the complementary strand, sequential connectivity was completed without interruption. The interaction between T^{17} H1′ and A^{18} H8 produced a weak cross-peak (Figure 3-3B). A decrease in the intensity of the A^{18} H1′:G^{19} H8 cross-peak was also observed. These results were indicative of a localized perturbation at the lesion site, with an increase in the rise between the lesion site and its 5′- and 3′-neighboring base pairs.

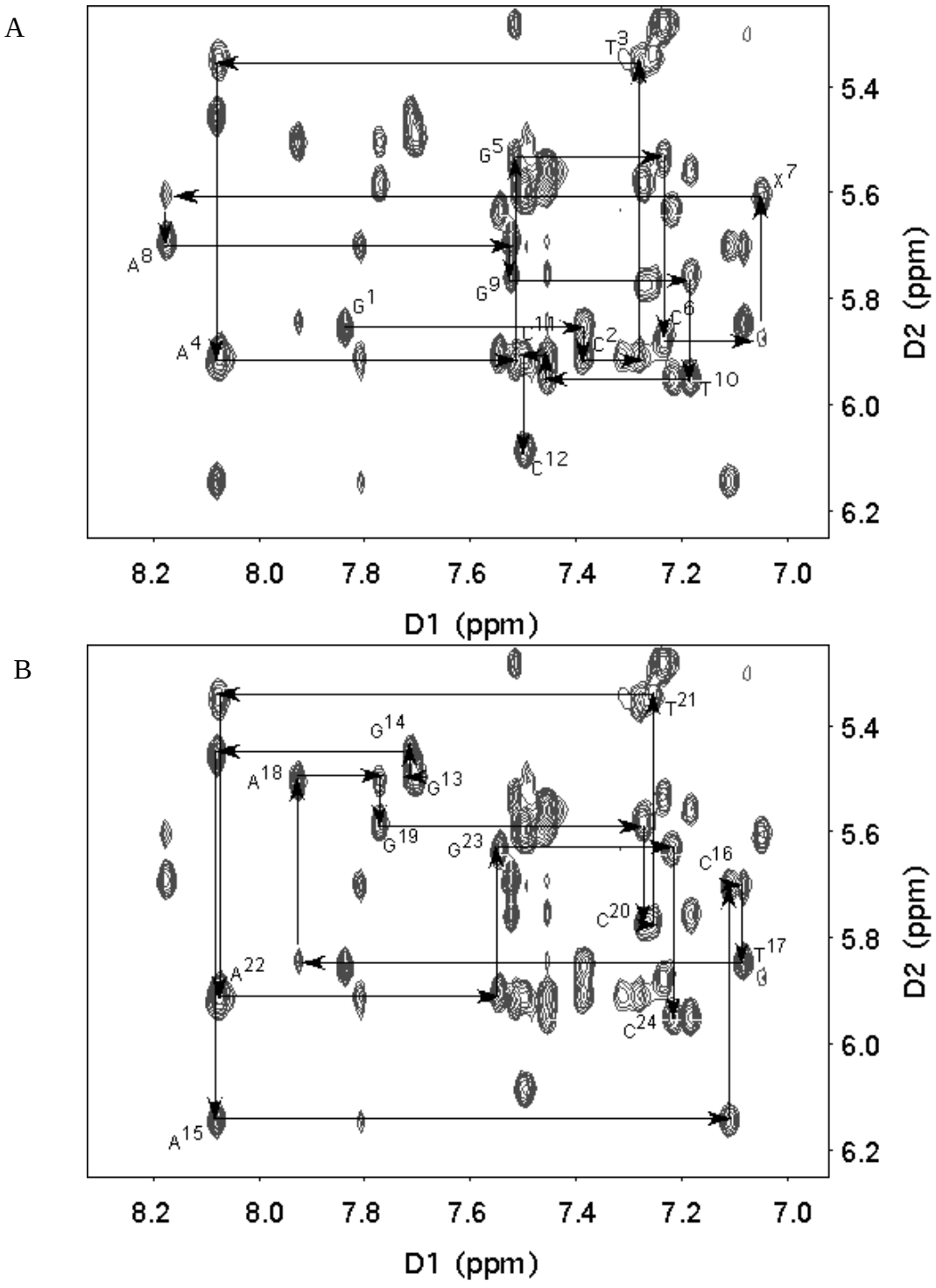


Figure 3-3: NOESY expansion plot of the anomeric proton to aromatic proton region for the *R*-BD-N3-dU modified sequence 5′-G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}X^{7}A^{8}G^{9}T^{10}C^{11}C^{12}-3′·5′-G^{13}G^{14}A^{15}C^{16}T^{17}A^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}-3′. (A) Nucleotides G^{1}-C^{12} of the modified oligonucleotide containing *R*-BD-N3-dU. (B) Nucleotides G^{13}-C^{24} of the complementary strand. NMR data were acquired at a mixing time of 250 ms and a frequency of 800.23 MHz.

Chemical shift analyses compare the chemical shift values of the DNA protons in the unmodified oligodeoxynucleotide sequence to those in the modified duplex. Changes in chemical shift values for the aromatic and anomeric H1′ protons of the modified and complementary strands are shown in Figure 3-4. The largest perturbation occurred at the modified nucleotide, X^7, where the H6 proton was shifted 0.3 ppm. Slight perturbations were also observed in the neighboring bases to the adduct. The chemical shifts of the H1′ proton of C^6 and A^8 were shifted 0.2 ppm. The H2′ and H2′′ protons of the adduct were shifted downfield 0.8 and 0.6 ppm, respectively. (Data not shown) In the complementary strand, chemical shift changes of 0.4 ppm were observed for the H1′ protons of T^{17} and A^{18} and a chemical shift change of 0.3 ppm was observed for the H8 proton of G^{19}. The observation of chemical shift changes primarily at the lesion site and its 3′- and 5′-neighboring base pairs suggests a localized perturbation to the structure of the oligonucleotide.

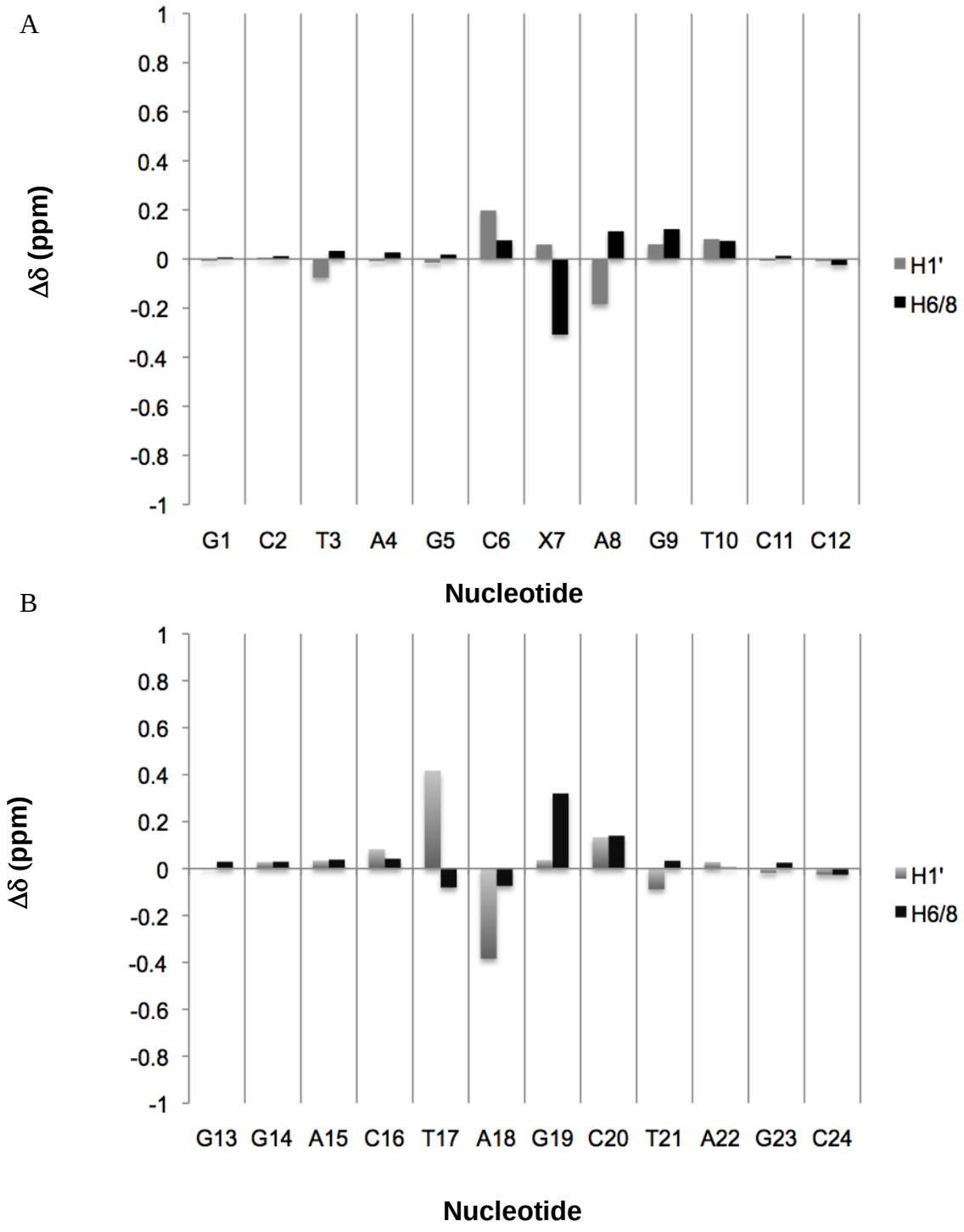


Figure 3-4: Chemical shift perturbations in the aromatic and anomeric protons of the R-BD-N3-dU modified sequence, 5′-G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}X^{7}A^{8}G^{9}T^{10}C^{11}C^{12}-3′·5′-G^{13}G^{14}A^{15}C^{16}T^{17}A^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}-3′, relative to the unmodified oligonucleotide. (A) Nucleotides G^{1}-C^{12} of the modified oligonucleotide containing R-BD-N3-dU. (B) Nucleotides G^{13}-C^{24} of the complementary strand.

Butadiene Protons

The six butadiene protons are labeled in Figure 3-1. Cross-peaks for the Hα, Hα′, Hβ, Hγ, Hδ, and Hδ′ protons were observed in a total correlation spectroscopy (TOCSY) spectrum. The TOCSY spectrum showing the scalar interactions between Hβ and the five other butadiene protons is shown in Figure 3-5. Hβ and Hγ were assigned using TOCSY and NOESY experiments. A DQF-COSY experiment was utilized for the assignment of Hδ and Hδ′. The Hδ′ proton was assigned based on the larger coupling constant. Unambiguous assignments of Hα and Hα′ were not achievable. However, the intensities in NOESY spectra at mixing times ranging from 80 to 250 ms between Hα and Hα′ and other protons were on the same order, suggesting that an ambiguous assignment of the protons would produce identical structures. The chemical shift values for the butadiene protons are listed in Appendix A.

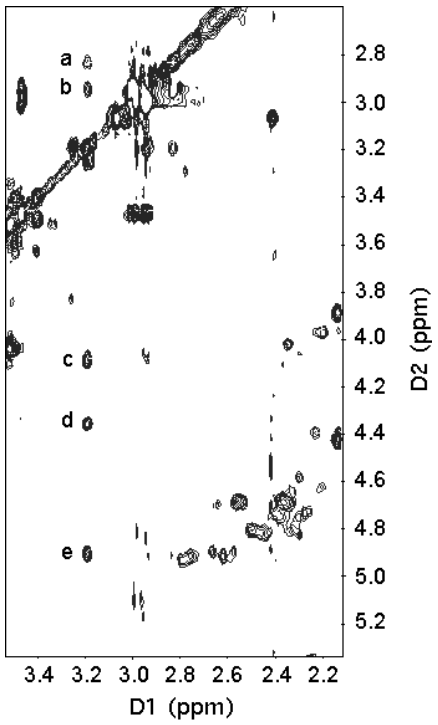


Figure 3-5: TOCSY spectrum of the R-BD-N3-dU butadiene protons. TOCSY crosspeaks between butadiene protons: a, Hα→ Hβ; b, Hα′→Hβ; c, Hγ→Hβ; d, Hδ→Hβ; e, Hδ′→Hβ. The spectrum was acquired at a mixing time of 80 ms and a frequency of 800.23 MHz.

Assignment of the butadiene protons was the first step in determining dipolar interactions between the butadiene protons and other oligonucleotide protons. The observation of cross-peaks resulting from these interactions provided information on the orientation of the butadiene moiety with respect to the oligonucleotide. The H2 and H1′ regions of the NOESY spectrum are shown in Figure 3-6. Strong interactions between Hα, Hα′, and Hβ of butadiene and A^{8} H2 were observed in the NOESY spectrum. Cross-peaks between the same butadiene protons and A^{18} H2 were also observed. An interaction between Hβ and T^{17} H1′ was observed in the anomeric region of the spectrum. Stronger interactions between the A^{8} H2 proton and the butadiene protons suggested that for the *R* diastereomer, these butadiene protons would be positioned closer to the 3′-neighboring base, A^{8}.

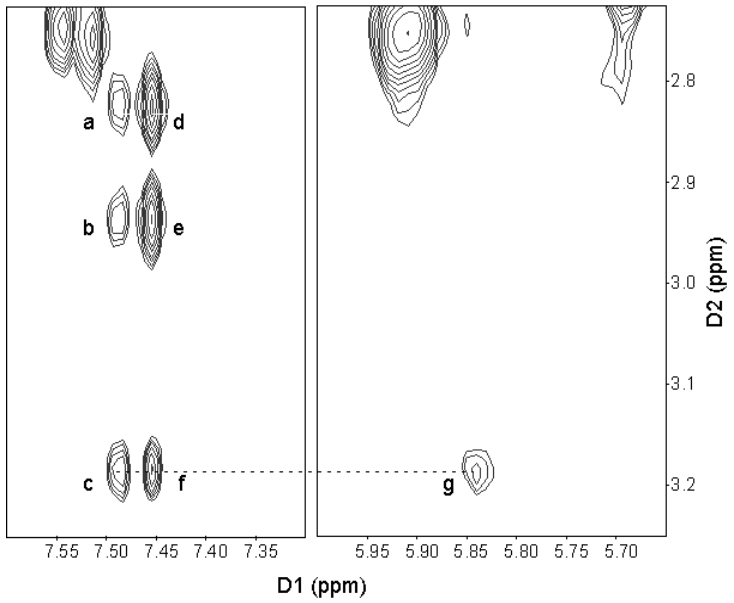


Figure 3-6: Dipolar proton:proton interactions between the *R*-BD moiety and other oligonucleotide protons. NOESY cross-peaks: a, Hα-A^{18} H2; b, Hα′-A^{18} H2; c, Hβ-A^{18} H2; d, Hα-A^{8} H2; e, Hα′-A^{8} H2; f, Hβ-A^{8} H2; g, Hβ-T^{17} H1′. NMR data were acquired at a mixing time of 250 ms and a frequency of 800.23 MHz.

Exchangeable Protons

The exchangeable imino protons were observed in a NOESY spectrum acquired according to the methods described in Chapter II. An expansion plot of the NOESY spectrum showing the imino proton region (12-15 ppm) is displayed in Figure 3-7. The imino protons of the terminal guanines, G^{1} and G^{13}, were not observable in the two-dimensional spectrum. This result is most likely due to fraying at the end of the helices, as well as a more rapid exchange of the terminal protons with solvent. The NMR experiment at 7 °C provided excellent data with well-resolved cross-peaks. Overlapping of cross-peaks did occur in the imino proton region, including those for T^{21} and T^{3} as well as G^{5} and G^{9}. In the two-dimensional spectrum, the T^{17} imino proton was significantly overlapped with that of T^{10}. However, the assignments for these cross-peaks were confirmed by the A^{15} H2:T^{10} N3H and A^{8} H2:T^{17} N3H interactions shown in Figure 3-7A.

The imino protons (Figure 3-7B) were sequentially connected beginning at the imino proton of G^{23} and walking to the closest imino proton, that of T^{3}. The T^{3} imino proton was "walked" to the T^{21} imino proton and the assignments continued without interruption to the G^{19} imino proton. The modification at the N3 position of X^{7} prevents Watson-Crick hydrogen bonding, causing NOE connectivity to be broken at G^{19}. The "walk" was expected to begin again at T^{17} and continue to the end. However, no interaction between T^{17} and G^{9} was observed (Figure 3-7B). The weak imino proton observed for T^{17} was indicative of a disrupted hydrogen bond at the A^{8}:T^{17} base pair. Evidence for a weakened hydrogen bond was also found in the H2 region of the

spectrum, where the interaction between T^{17} NH3 and A^{8} H2 was weakened (Figure 3-7A).

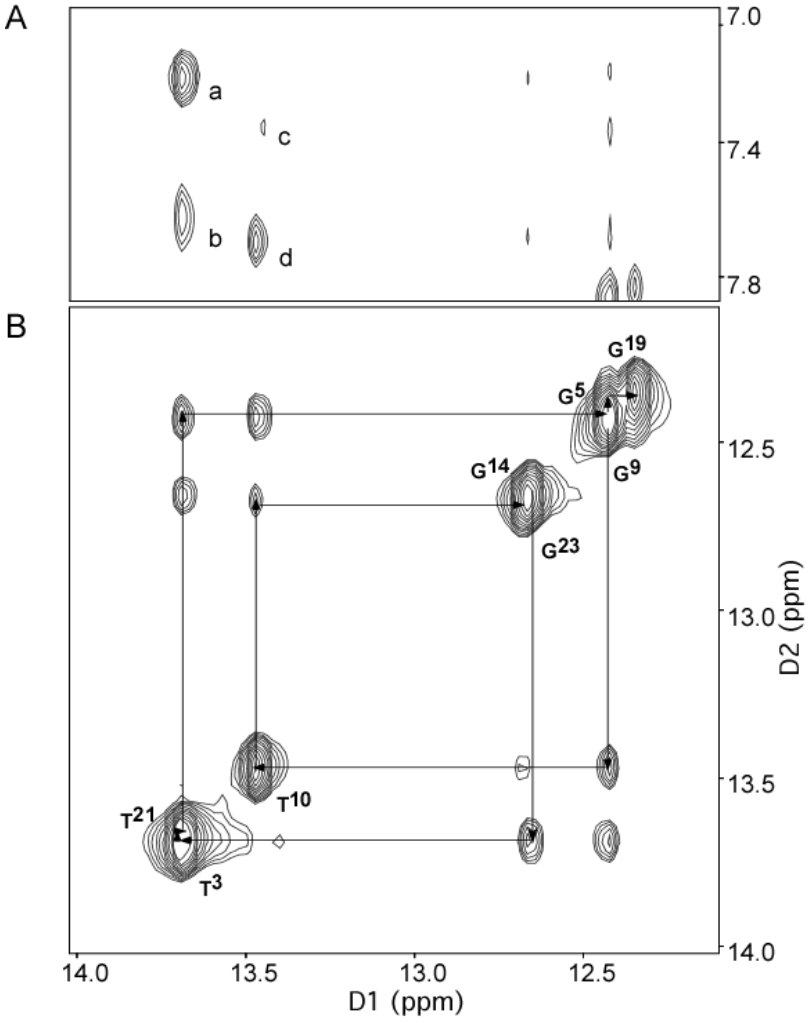


Figure 3-7: NOESY expansion plot of the imino proton region for the R-BD-N3-dU modified sequence, 5′-$G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}\underline{X}^{7}A^{8}G^{9}T^{10}C^{11}C^{12}$-3′·5′-$G^{13}G^{14}A^{15}C^{16}T^{17}A^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$-3′. (A) NOESY cross-peaks: a, A^{4} H2-T^{21} N3H; b, A^{22} H2-T^{3} N3H; c, A^{8} H2-T^{17} N3H; d, A^{15} H2-T^{10} N3H. (B) Sequential connectivity of imino protons starting at G^{23} and ending at G^{14}. Sequential connectivity is broken at G^{19} and resumed at G^{9}. NMR data were acquired at a temperature of 7 °C at a mixing time of 250 ms and a frequency of 800.23 MHz.

Structural Refinement

A total of twenty starting structures were used in the structural refinement of the duplex DNA containing the *R*-BD-N3-dU adduct opposite dA. These structures were built as described in Chapter II and energy minimized prior to rMD calculations. Restrained molecular dynamics calculations were carried out in implicit solvent using the

simulated annealing protocol. A total of 536 distance restraints, comprised of 303 intra-residue, 218 inter-residue, and 15 butadiene upper and lower bound restraints were used to restrain the MD calculations and are listed in Appendix B. The distance restraints were evenly distributed amongst the residues in the oligonucleotide. To restrain the backbone and sugar pucker, a total of 90 and 110 empirical restraints were used, respectively. In addition to the distance and angle restraints, 44 Watson-Crick hydrogen-bonding restraints were used, omitting hydrogen-bonding restraints for the X^{7}:A^{18} and A^{8}:T^{17} base pairs. These restraints can be found in Appendix C.

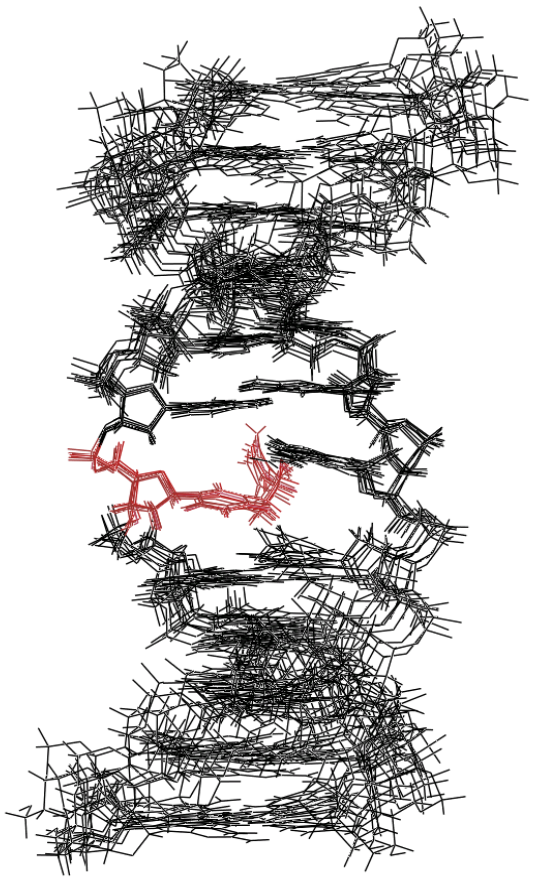


Figure 3-8: Overlay of ten solution structures emergent from simulated annealing calculations. The *R*-BD-N3-dU lesion is displayed in red. The rmsd of the ten internal base pairs is 1.18 Å.

Prior to the rMD calculations, the root mean square deviation (rmsd) of the twenty starting structures was 3.59 Å. Following 20 ps of calculations, the structural coordinates which emerged from the calculations were compared using pairwise rmsd. Ten of the twenty starting structures converged to give an rmsd value of 1.18 Å. An overlay of the structures is shown in Figure 3-8.

For the determination of structure accuracy, CORMA [155] calculations were run on the average structure from molecular dynamics calculations

performed in implicit solvent. The total R_1x value was $8.9x10^{-2}$, with an individual nucleotide range in R_1x values from $1.7x10^{-2}$ to $1.5x10^{-1}$. These values suggested that the average structure agrees well with NMR data. The results from the CORMA calculations with respect to the individual nucleotides are shown in Figure 3-9.

The average structure of the *R*-BD-N3-dU adduct opposite dA determined by rMD calculations in implicit solvent shows that the major perturbation resulting from modification at the X^7 position occurs at the lesion site. The lesion site of the average structure is shown in Figure 3-10. The butadiene moiety is oriented in the minor groove and in the 5′-direction. The position of the butadiene atoms forces the adenine opposite the lesion in the 3′-direction toward G^{19}. The position of the adduct generates a disruption in base pairing at both the lesion site and the 3′-neighboring base pair, A^8:T^{17}. Base stacking is also disrupted between the lesion and its 5′-neighboring base pair C^6.

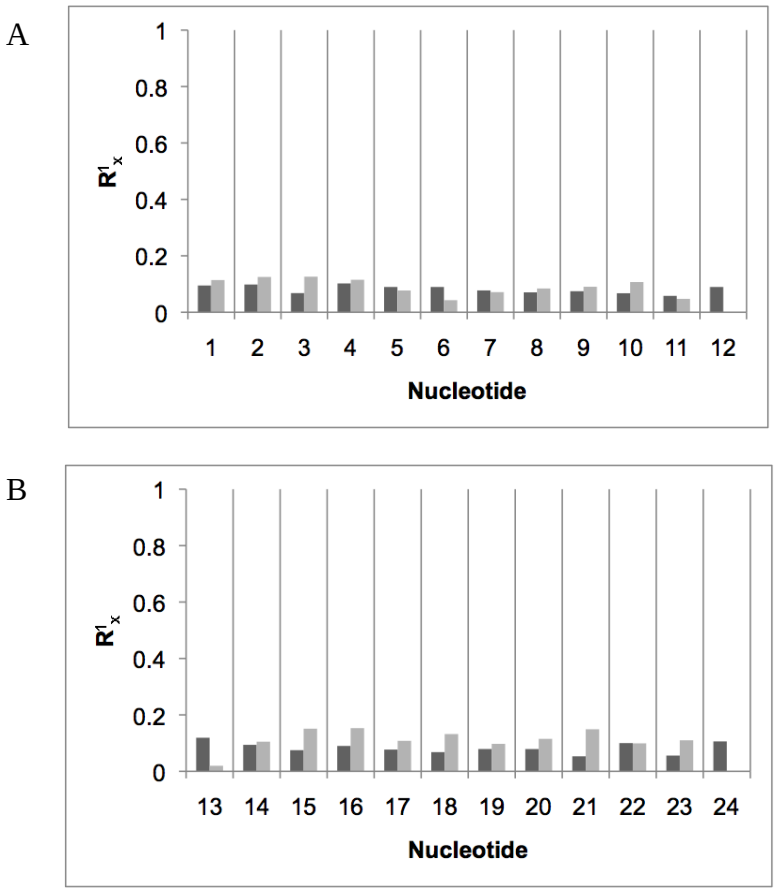


Figure 3-9: CORMA calculations on the average *R*-BD-N3-dU modified oligodeoxynucleotide calculated using a simulated annealing rMD protocol. (A) Nucleotides G^{1}-C^{12} of the modified oligonucleotide containing *S*-BD-N3-dU. (B) Nucleotides G^{13}-C^{24} of the complementary strand. The black bars represent intranuclear sixth root residuals and the gray bars represent internuclear sixth root residuals.

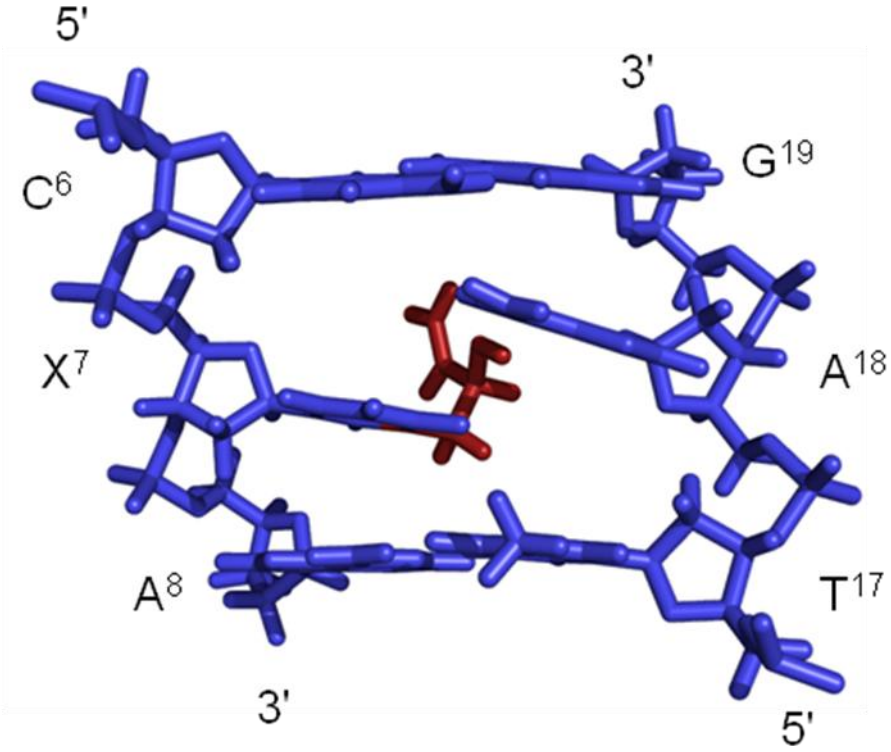


Figure 3-10: Lesion site of the average *R*-BD-N3-dU modified oligonucleotide. The BD moiety, illustrated in red, is oriented in the 5′-direction and displaces A^{18} in the 3′-direction.

Hydrogen Bonding Analysis

Initial calculations of the *R*-BD-N3-dU modified oligonucleotide indicated that the potential existed for a hydrogen bond between the hydroxyl proton and the O^4 atom of the modified nucleotide. To test this, the average structure extracted from molecular dynamics calculations in implicit solvent was placed in a truncated octahedral TIP3P solvent box after the addition of sodium counter ions. Prior to 5 ns of molecular dynamics calculations, the solvated molecule was energy minimized and equilibrated according to standard methods. Hydrogen bonding occupancies were extracted from molecular dynamics trajectories using PTRAJ (AMBER 8.0) [153]. The percent occupancy for the hydrogen bond between the hydroxyl proton and the O^4 atom on the

modified uracil was 72.9%. The distance between the hydroxyl proton and O^4 atom was 2.77 Å, and the O-H-O angle was determined to be 21.33 degrees. These data predicted hydrogen bonding interactions at this site. The modified base extracted from explicit solvent MD calculations is displayed in Figure 3-11.

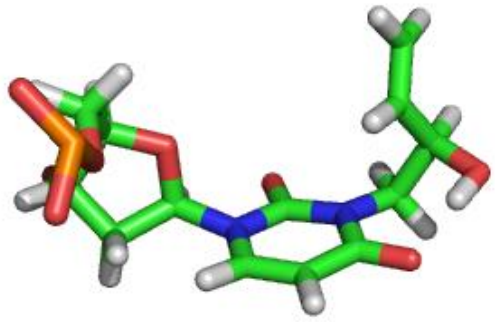


Figure 3-11: Structural representation of the *R*-BD-N3-dU DNA adduct extracted from 5 ns of MD calculations in explicit solvent. MD calculations predicted the potential for a hydrogen bond between the hydroxyl proton of the butadiene moiety and the O^4 atom of the modified base.

Helicoidal Analysis

A helicoidal analysis of the average solution structure was performed following rMD calculations. As a result of the modification at X^7, a 24º bend was present in the structure. The base pair parameters were normal except a slight displacement in the Y-axis at the lesion site (Figure 3-12). The parameters for base-base interactions showed that at the lesion site, the modification causes more than a 20° opening. Disruptions also occurred in the shear, the stretch, and the stagger, primarily at the lesion site (Figure 3-13). As was predicted by NOESY data, a 2 Å increase in rise was observed between X^7 and C^6 (Figure 3-14).

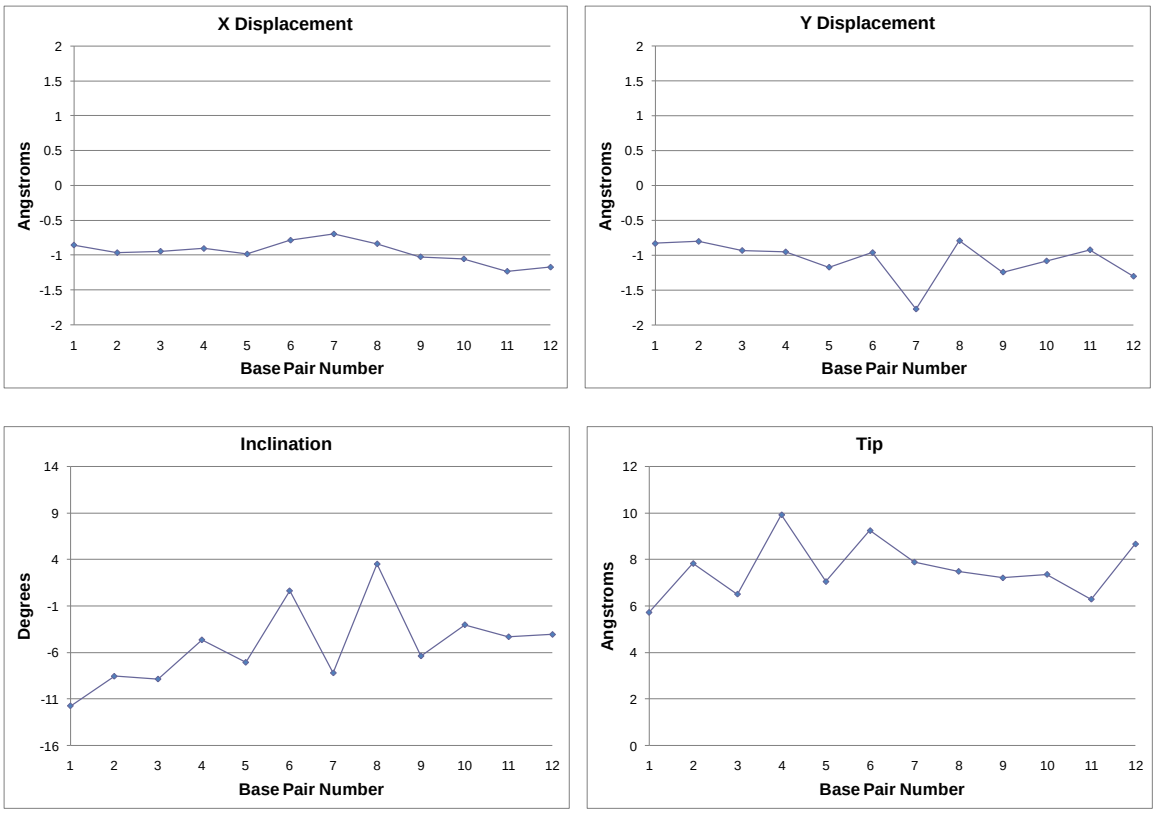


Figure 3-12: Global base pair helicoidal parameters. Helicoidal parameters for the *R*-BD-N3-dU modified oligonucleotide 5′-G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}<u>X</u>^{7}A^{8}G^{9}T^{10}C^{11}C^{12}-3′·5′-G^{13}G^{14}A^{15}C^{16}T^{17}A^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}-3′.

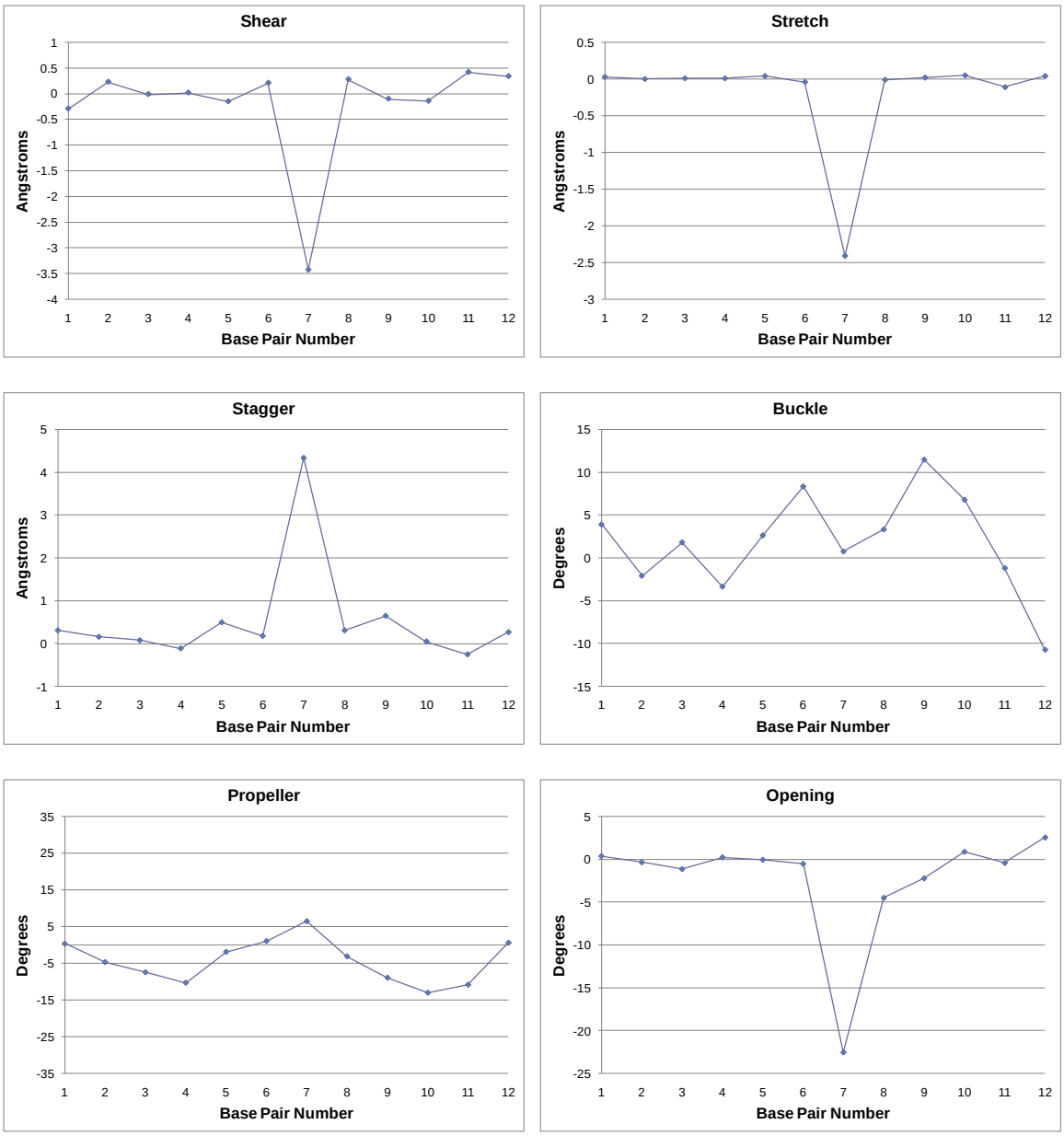


Figure 3-13: Global base-base helicoidal parameters. Helicoidal parameters for the *R*-BD-N3-dU modified oligonucleotide 5′-G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}<u>X</u>^{7}A^{8}G^{9}T^{10}C^{11}C^{12}-3′·5′-G^{13}G^{14}A^{15}C^{16}T^{17}A^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}-3′.

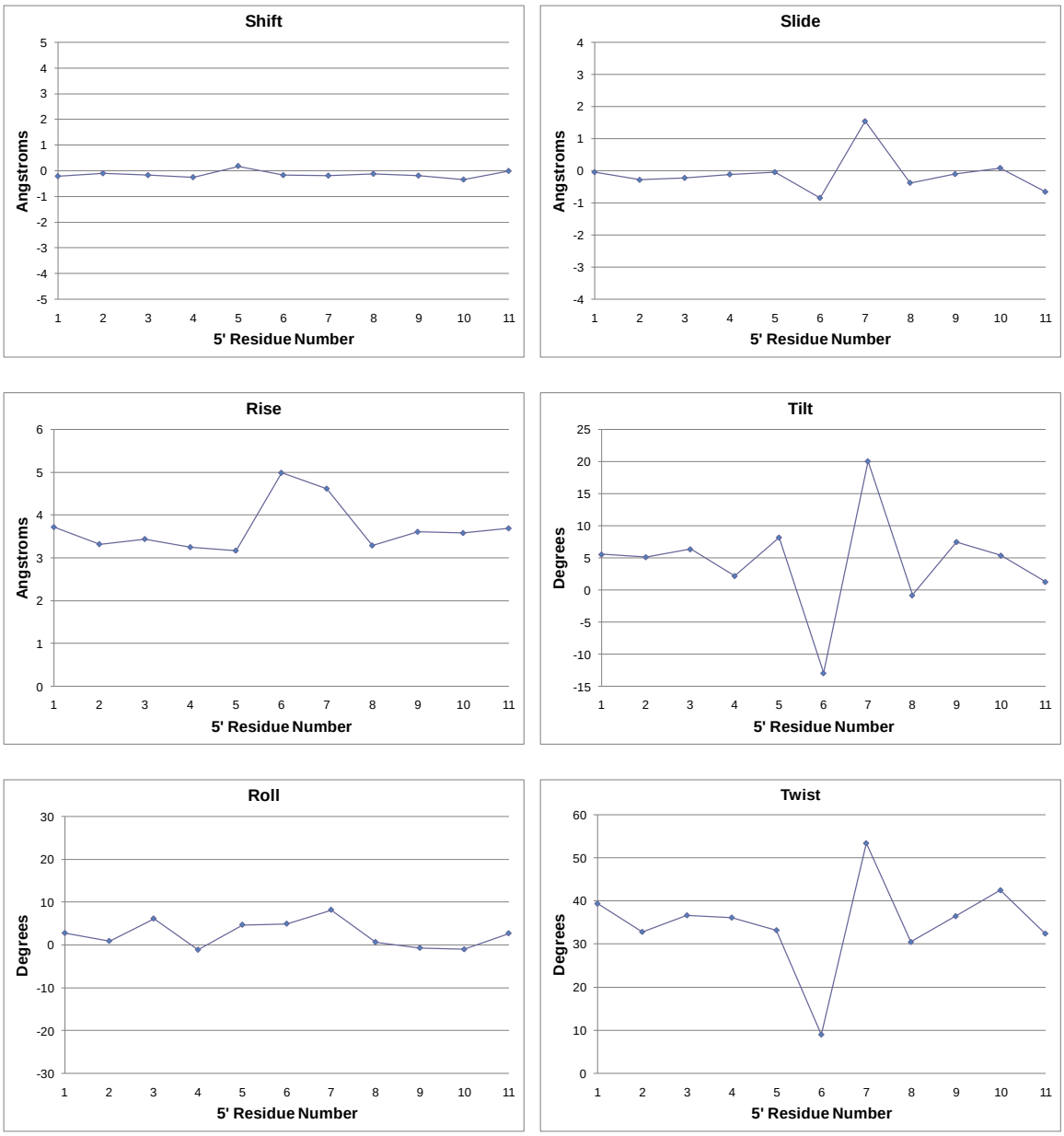


Figure 3-14: Global inter-base helicoidal parameters. Helicoidal parameters for the *R*-BD-N3-dU modified oligonucleotide 5′-$G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}\underline{X}^{7}A^{8}G^{9}T^{10}C^{11}C^{12}$-3′·5′-$G^{13}G^{14}A^{15}C^{16}T^{17}A^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$-3′.

S-Diastereomer

Sample Purity

The double-stranded *S*-BD-N3-dU modified oligonucleotide, 5′-d($G^1C^2T^3A^4G^5C^6\underline{X}^7A^8G^9T^{10}C^{11}C^{12}$)-3′·5′-d($G^{13}G^{14}A^{15}C^{16}T^{17}A^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$)-3′, was purified using the HPLC protocol described in Chapter II. Enzyme hydrolysis showed that the modified oligonucleotide was enantiomerically pure. The annealed duplex was separated from any remaining single-stranded DNA using a hydroxylapatite column. The presence of only duplex DNA in the solution sample was confirmed using capillary gel electrophoresis. The composition of the individual strands was verified using MALDI-TOF mass spectrometry and NMR spectroscopy. NMR experiments were carried out at a frequency of 800.23 MHz and yielded useful data.

Non-exchangeable Protons

For the observation of non-exchangeable protons, the sample was prepared as discussed in Chapter II. The purine and pyrimidine aromatic protons, thymine methyl protons and deoxyribose H1′, H2′, H2′′, and H3′ protons were assigned completely. Only partial assignment of the deoxyribose H4′, H5′, and H5′′ protons were completed. The chemical shift values of the non-exchangeable protons are listed in Appendix A. Expanded plots of the NOESY spectrum for the *S* diastereomer are shown in Figure 3-15.

Sequential connectivity throughout the modified strand was completed without interruption. A significant observation in the modified strand was a decrease in intensity for the interaction between C^6 H1′ and X^7 H6 (Figure 3-15A). A weak cross-peak was

also observed for the interaction between X^7 H1′ and A^8 H8. These weakened cross-peaks suggested a localized perturbation at the lesion site and its 5′- and 3′-neighbor base pairs, as well as an increase in the rise between these nucleotides resulting from the butadiene moiety. Sequential connectivity remained unbroken in the complementary strand as well. In comparison to the unmodified oligonucleotide, the intensities of the cross-peaks for the interactions between T^{17} H1′ and A^{18} H8, as well as A^{18} H1′ and G^{19} H8, were weakened (Figure 3-15B).

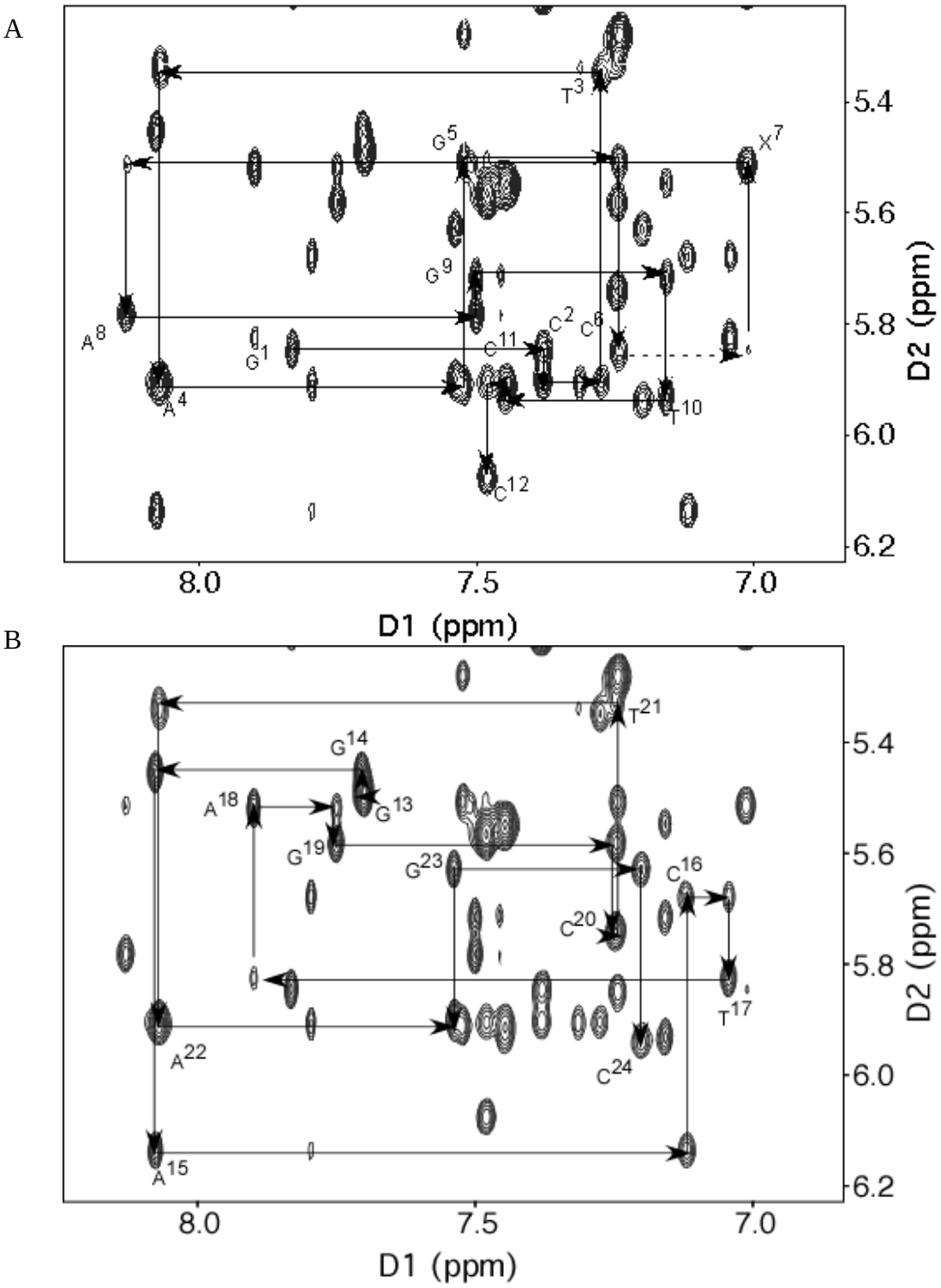


Figure 3-15: NOESY expansion plot of the anomeric proton to aromatic proton region for the *S*-BD-N3-dU modified sequence, 5′-$G^1C^2T^3A^4G^5C^6\underline{X}^7A^8G^9T^{10}C^{11}C^{12}$-3′·5′-$G^{13}G^{14}A^{15}C^{16}T^{17}A^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$-3′. (A) Nucleotides G^1-C^{12} of the modified oligonucleotide containing *S*-BD-N3-dU. (B) Nucleotides G^{13}-C^{24} of the complementary strand. NMR data were acquired at a mixing time of 250 ms and a frequency of 800.23 MHz.

The chemical shift values of the anomeric and aromatic protons in the *S*-BD-N3-dU modified oligonucleotide were compared to those of the unmodified duplex. The changes in chemical shift values are shown in Figure 3-16. Like the *R*-diastereomer, the largest perturbation occurred at the modified nucleotide, X^7, where the H6 proton was shifted almost 0.4 ppm. In the modified strand, slight perturbations were observed in the 5′-neighboring base to the adduct. The chemical shift of the H1′ proton of C^6 was shifted 0.2 ppm. In the complementary strand, chemical shift changes of 0.4 ppm were observed for the H1′ protons of T^{17} and A^{18} and a chemical shift change of 0.3 ppm was observed for the H8 proton of G^{19}.

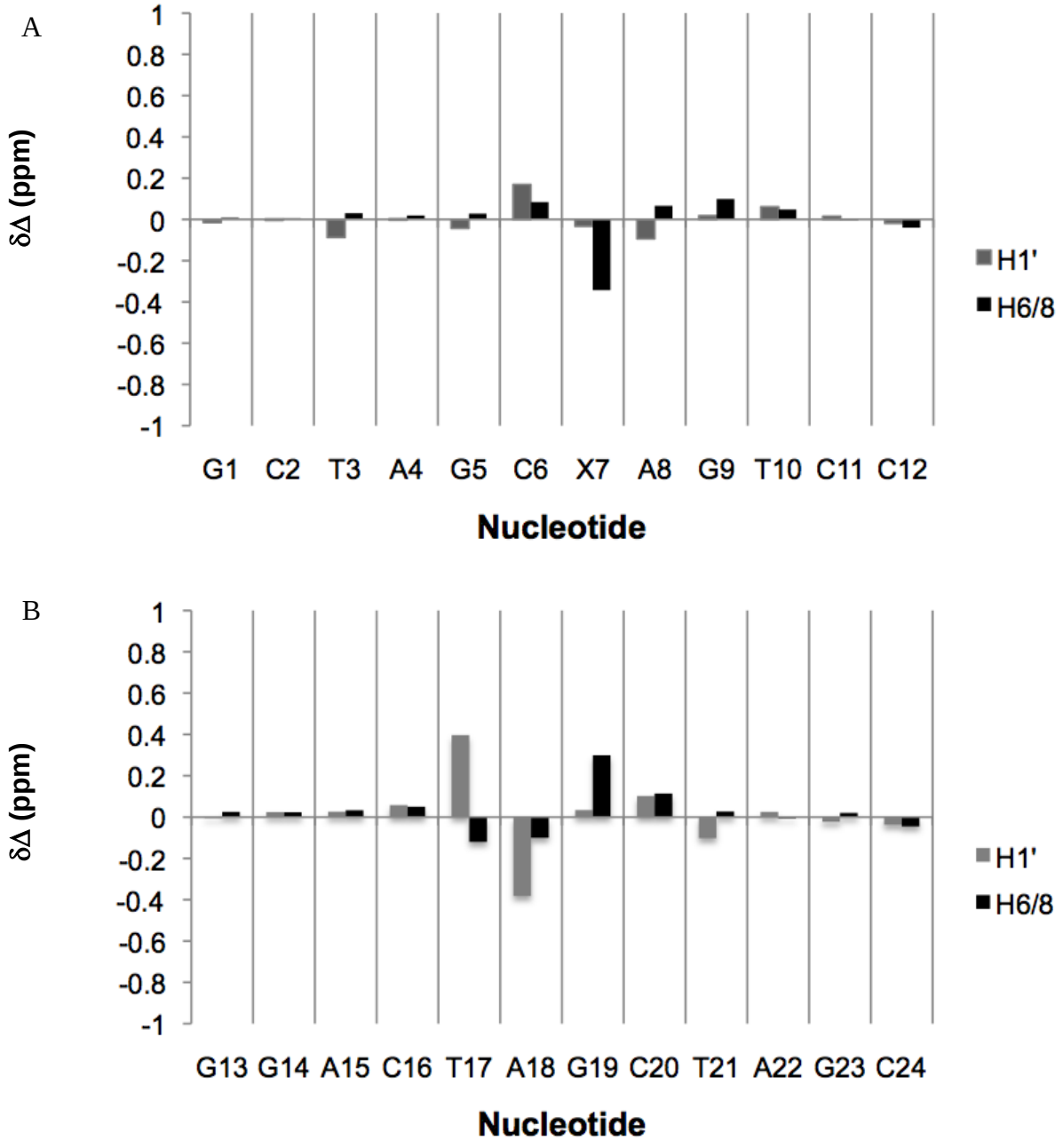


Figure 3-16: Chemical shift perturbations in the aromatic and anomeric protons of the *S*-BD-N3-dU modified sequence, 5′-G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}X^{7}A^{8}G^{9}T^{10}C^{11}C^{12}-3′·5′-G^{13}G^{14}A^{15}C^{16}T^{17}A^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}-3′, relative to the unmodified oligodeoxynucleotide. (A) Nucleotides G^{1}-C^{12} of the modified oligonucleotide containing *S*-BD-N3-dU. (B) Nucleotides G^{13}-C^{24} of the complementary strand.

Butadiene Protons

The Hα, Hα′, Hβ, Hγ, Hδ, and Hδ′protons of the S-diastereomer were also observed in a TOCSY spectrum. The TOCSY spectrum showing the scalar interactions between Hβ and the five other butadiene protons is shown in Figure 3-17. Hβ and Hγ were assigned using TOCSY and NOESY experiments. A DQF-COSY experiment was acquired in order to assign Hδ and Hδ′. Like those for the *R*-diastereomer, the protons on the alpha carbon, Hα and Hα′, were not able to be assigned unambiguously. However, the intensities in NOESY spectra at mixing times ranging from 80 to 250 ms between Hα and Hα′ and other protons were on the same order, suggesting that an ambiguous assignment of the protons would produce identical structures. The chemical shift values for the butadiene protons are listed in Appendix A.

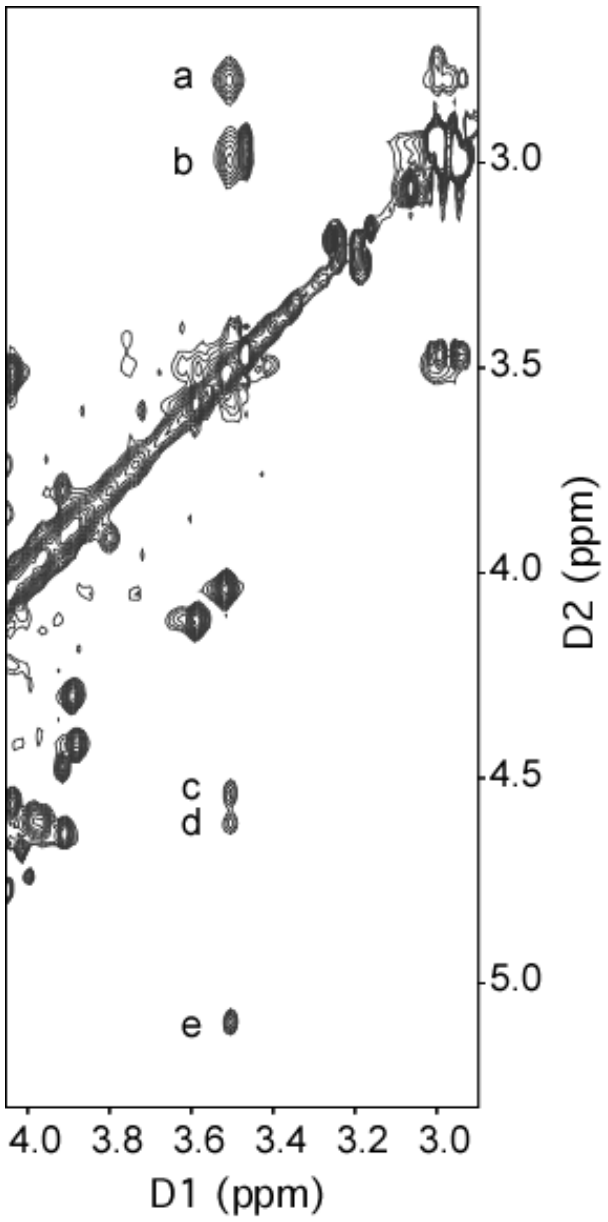


Figure 3-17: TOCSY spectrum of the *S*-BD-N3-dU butadiene protons. TOCSY cross-peaks between butadiene protons: a, Hα→ Hβ; b, Hα′→Hβ; c, Hγ→Hβ; d, Hδ→Hβ; e, Hδ′→Hβ. The spectrum was acquired at a mixing time of 80 ms and a frequency of .23 MHz.

Interactions between butadiene protons and other nucleotide protons were observed in the H1′ and H2 regions of the NOESY spectrum. Interactions were observed between A^{8} H2 and the Hα, Hα′ and Hβ protons (Figure 3-18). Cross-peaks for those interactions between A^{18} H2 and the Hα′ and Hβ protons were also observed. A cross-

peak for A^{18} H2 to Hα was observed but was not measurable due to overlapping peaks in the H2″ region of the NOESY spectrum. The presence of the butadiene moiety near A^{18} was validated by interactions between A^{18} H1′ and the butadiene protons Hα′ and Hβ. Finally, interactions between T^{17} H1′ and Hβ were observed in the NOESY spectrum. The proximity of these butadiene protons to A^{18} H2 and A^{8} H2 indicated that the butadiene moiety was oriented toward the minor groove of the duplex. While the interactions of Hα′ and Hβ with the H1′ protons of T^{17} and A^{18} suggested that the butadiene moiety was positioned toward the complementary strand.

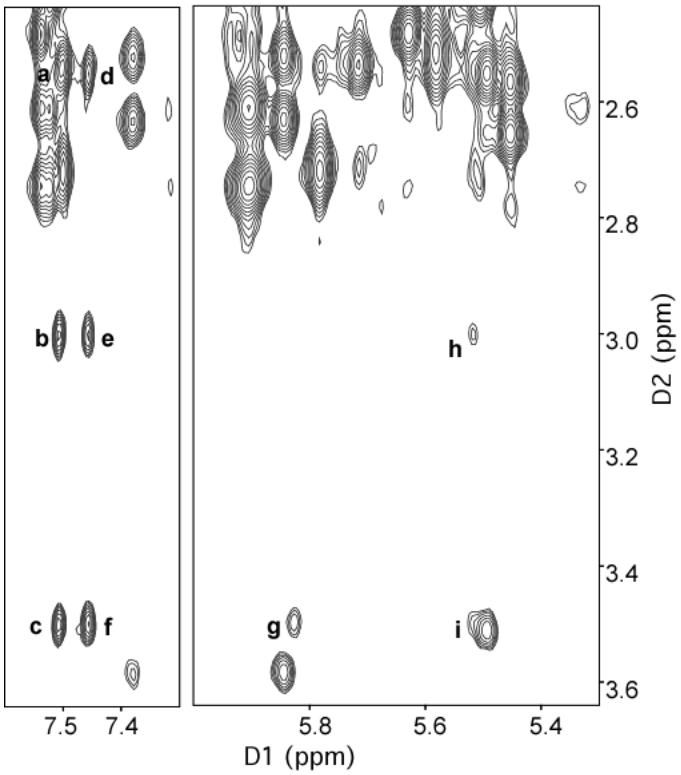


Figure 3-18: Dipolar proton to proton interactions between the S-BD-N3-dU moiety and other oligonucleotide protons. NOESY cross-peaks: a, Hα-A^{18} H2; b, Hα′-A^{18} H2; c, Hβ-A^{18} H2; d, Hα-A^{8} H2; e, Hα′-A^{8} H2; f, Hβ-A^{8} H2; g, Hβ-T^{17} H1′; h, Hα′-A^{18} H1′; i, Hβ-A^{18} H1′. NMR data were acquired at a mixing time of 250 ms and a frequency of 800.23 MHz.

Exchangeable Protons

The exchangeable imino protons were observed in a NOESY spectrum acquired according to methods described in Chapter II. An expansion plot showing the imino proton region (12-15 ppm) is shown in Figure 3-19. The imino protons of the terminal base guanines G^{1} and G^{13} were not observable in the two-dimensional spectrum. In the two-dimensional spectrum, the T^{17} imino proton of the *S* diastereomer was significantly overlapped with that of T^{10}. The assignments for these cross-peaks were confirmed by the A^{15} H2-T^{10} N3H and A^{8} H2-T^{17} N3H interactions shown in Figure 3-19A.

The imino protons (Figure 3-19B) were assigned sequentially beginning at the imino proton of G^{23} and "walking" to the T^{3} N3H. The T^{3} imino proton was "walked" to T^{21} and the assignments continued by connecting imino protons, "walking" toward the 3′ end of the modified strand. The modification at the N3 position of X^{7} prevents Watson-Crick hydrogen bonding, causing NOE connectivity to be broken at G^{19}. The imino proton for T^{17} was very weak and furthermore, there was no interaction between T^{17} and G^{9} (Figure 3-19B). These data suggest a disrupted hydrogen bond at T^{17}, which was corroborated by a broadened cross peak between T^{17} NH3 and A^{8} H2 in more upfield region of the spectrum (Figure 3-19A). Sequential connectivity was resumed at G^{9} and completed to the 3′-end of the modified strand.

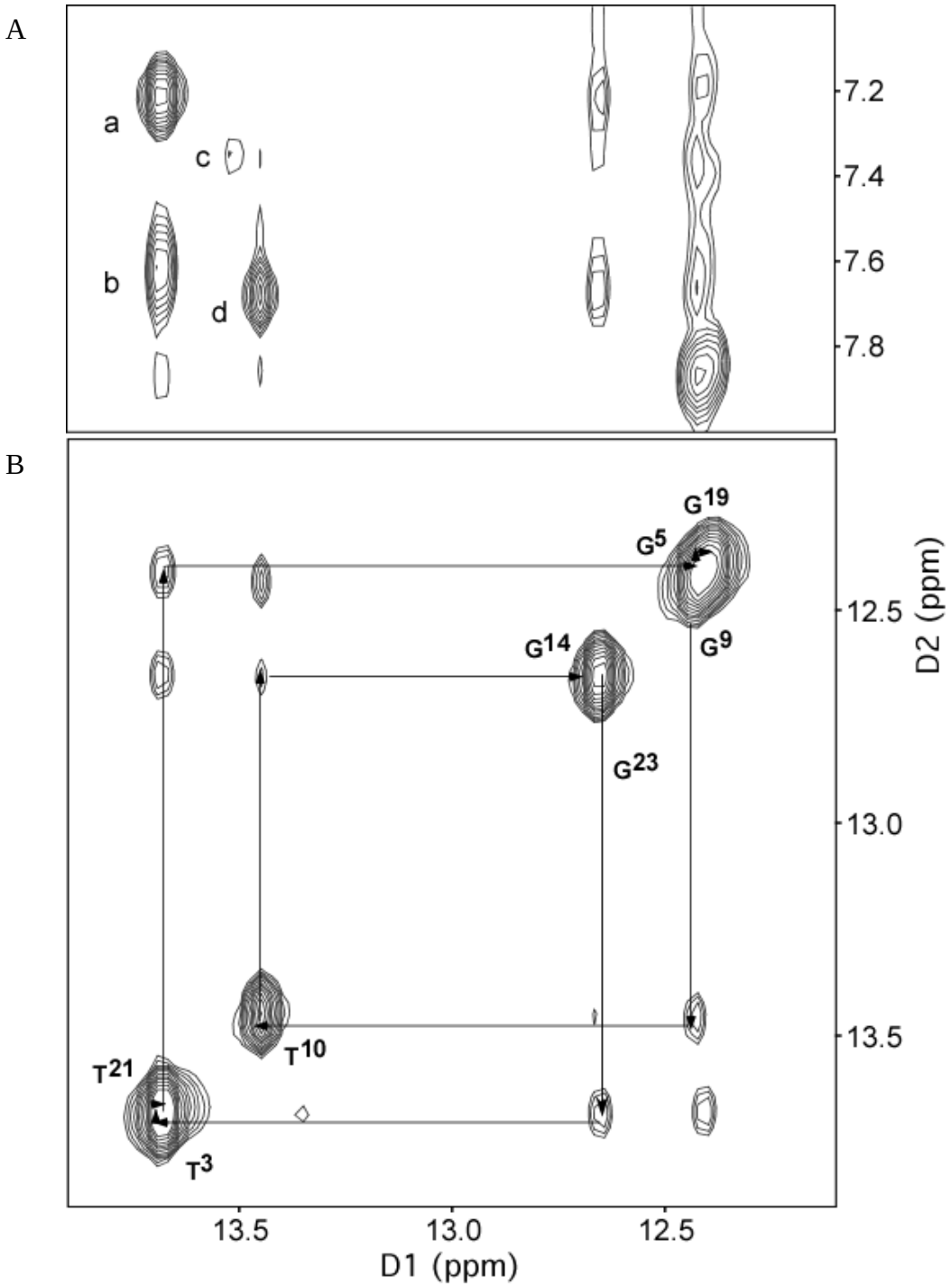


Figure 3-19: NOESY expansion plot of the imino proton region for the S-BD-N3-dU modified sequence 5′-$G^1C^2T^3A^4G^5C^6\underline{X}^7A^8G^9T^{10}C^{11}C^{12}$-3′·5′-$G^{13}G^{14}A^{15}C^{16}T^{17}A^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$-3′. (A) NOESY cross-peaks: a, A^4 H2-T^{21} N3H; b, A^{22} H2-T^3 N3H; c, A^8 H2-T^{17} N3H; d, A^{15} H2-T^{10} N3H. (B) Sequential connectivity of imino protons starting at G^{23} and ending at G^{14}. Sequential connectivity is disrupted at G^{19} and resumed at G^9. NMR data were acquired at 7 °C at a mixing time of 250 ms and a frequency of 800.23 MHz.

Structural Refinement

A total of twenty starting structures were used in the structural refinement of the duplex DNA containing the *S*-BD-N3-dU adduct. These structures were built as described in Chapter II and energy minimized prior to rMD calculations. The Generalized Born model was used to simulate a solvated environment. Twenty ps of rMD calculations using a simulated annealing protocol were used to generate an output coordinate file. A total of 529 distance restraints comprised of 323 intra-residue, 182 inter-residue, and 24 butadiene restraints, were used to restrain MD calculations. The distance restraints were evenly distributed amongst the residues in the oligonucleotide and are listed in Appendix B. To restrain the backbone and sugar pucker, a total of 90 and 110 empirical restraints were used, respectively. In addition to the distance and angle restraints, 44 Watson-Crick hydrogen-bonding restraints were used, omitting hydrogen-bond restraints for the X^{7}:A^{18} and A^{8}:T^{17} base pairs. These restraints are listed in Appendix C.

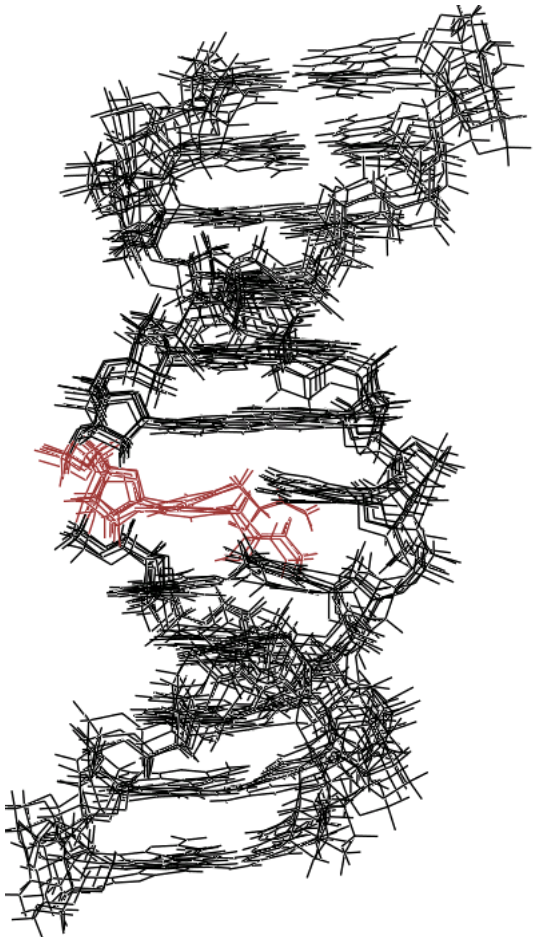


Figure 3-20: Overlay of six solution structures emergent from simulated annealing calculations. The S-BD-N3-dU lesion is displayed in red. The rmsd of the ten internal base pairs is 1.02 Å.

Prior to the rMD calculations, the rmsd of the twenty starting structures was 3.42 Å. Following 20 ps of calculations, the

structural coordinates which emerged from the calculations were compared using pairwise rmsd. Six of the twenty starting structures converged to give an rmsd value of 1.02 Å. An overlay of the structures is shown in Figure 3-20.

CORMA [155] calculations were run on the average structure emerging from rMD calculations. The total R_1x value was $8.4x10^{-2}$ with a range in individual nucleotide sixth root residuals from $1.7x10^{-2}$ to $1.6x10^{-1}$. CORMA data suggest that the average NMR solution structure agrees well with NMR data. The results from the CORMA calculations with respect to the individual nucleotides are shown in Figure 3-21.

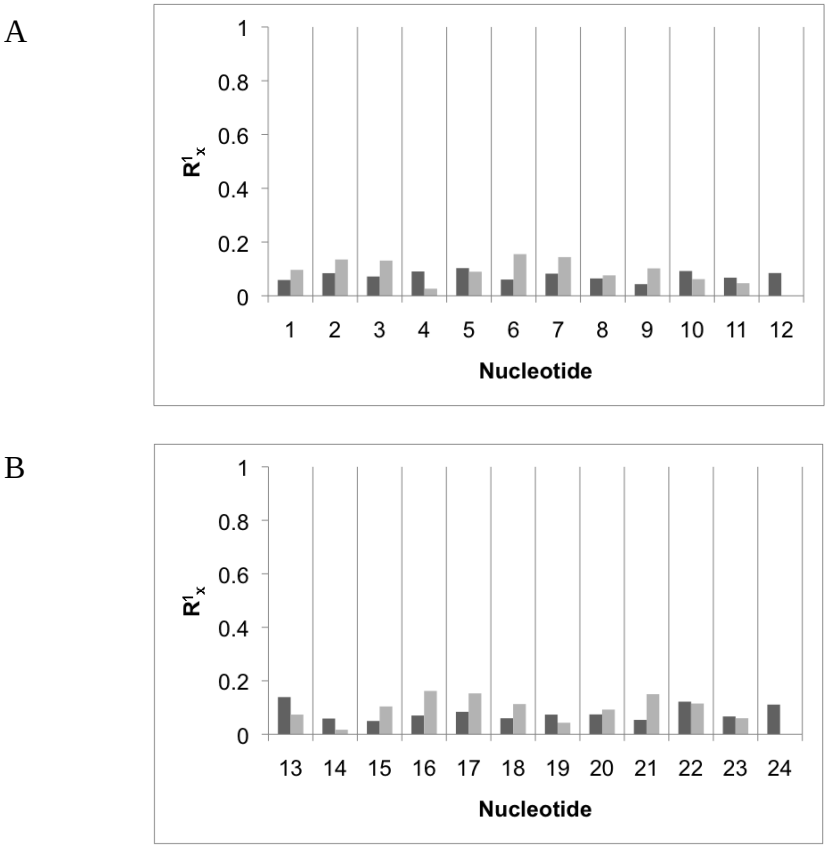


Figure 3-21: CORMA calculations on the average *S*-BD-N3-dU modified oligodeoxynucleotide calculated using a simulated annealing rMD protocol. (A) Nucleotides G^1-C^{12} of the modified oligonucleotide containing *S*-BD-N3-dU. (B) Nucleotides G^{13}-C^{24} of the complementary strand. The black bars represent intranuclear sixth root residuals and the gray bars represent internuclear sixth root residuals.

Structure of the S-BD-N3-dU Adduct Opposite dA

The average structure of the S-BD-N3-dU adduct opposite dA calculated using rMD calculations in implicit solvent shows that the major perturbation resulting from modification at the X^7 position occurs at the lesion site. The lesion site of the average structure is shown in Figure 3-22. The butadiene moiety is oriented in the minor groove toward the complementary strand. The position of the butadiene atoms forces the adenine opposite the lesion in the 3′-direction toward G^{19}. Base pairing is disrupted at both the lesion site and the 3′-neighboring base pair, A^8:T^{17} due to the position of the lesion. Base stacking is also disrupted at the lesion and its 5′-neighboring base C^6.

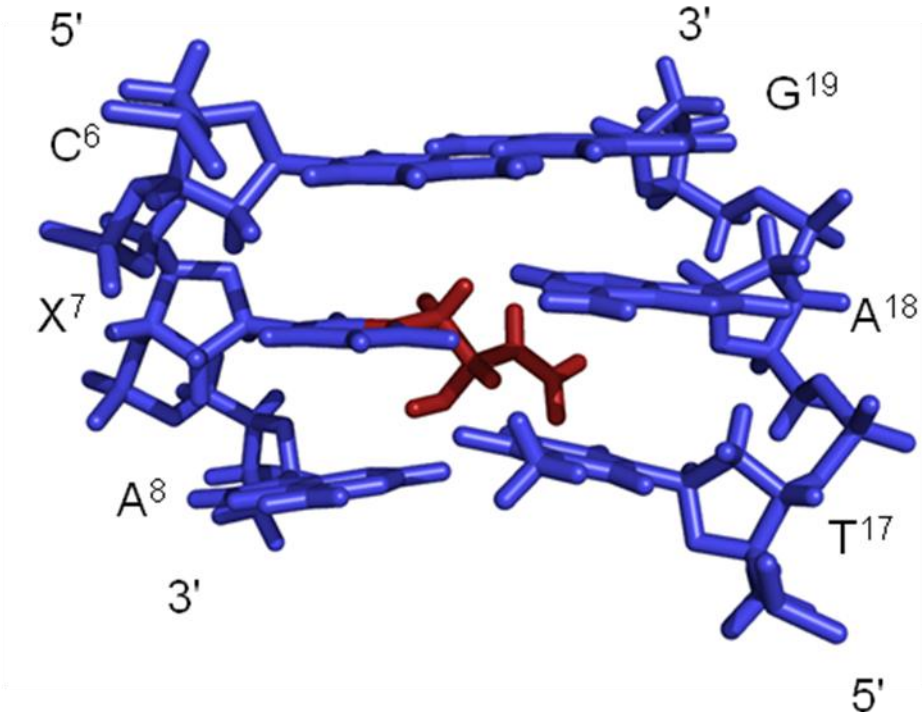


Figure 3-22: Lesion site of the average *S*-BD-N3-dU modified oligonucleotide. The BD moiety, illustrated in red, is oriented toward the complementary strand and displaces T^{17}, causing a disruption in base pairing at A^8:T^{17}.

Hydrogen Bonding Analysis

Like the *R*-diastereomer, initial calculations of the *S*-BD-N3-dU modified oligonucleotide indicated a potential for a hydrogen bond at the lesion site. For the *S*-diastereomer, however, this interaction would most likely occur between the hydroxyl proton and the O^2 atom of the modified uracil. To test this, the average structure extracted from MD calculations in implicit solvent was placed in a truncated octahedral TIP3P solvent box after the addition of sodium counter ions. Prior to 5 ns of MD calculations, the solvated molecule was energy minimized and equilibrated according to standard methods. Hydrogen bonding occupancies were extracted from MD trajectories using PTRAJ (AMBER 8.0) [153]. The percent occupancy for the hydrogen bond between the hydroxyl proton and the O^2 atom on the modified uracil was 85.8%. The distance between the hydroxyl proton and O^2 atom was 2.74 Å and the O-H-O angle was determined to be 19°. These data predicted a possible hydrogen bond at this location. The modified base extracted from explicit solvent molecular dynamics calculations is displayed in Figure 3-23.

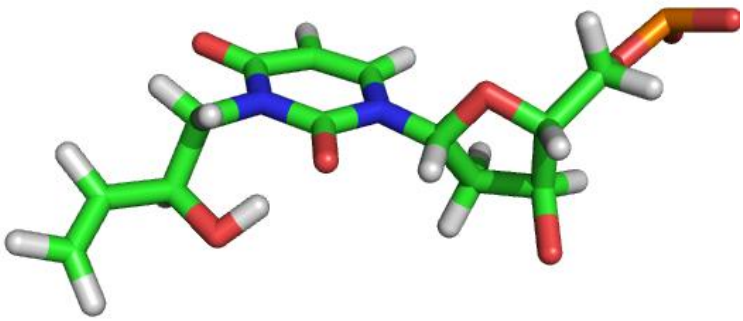


Figure 3-23: Structural representation of the modified *S*-BD-N3-dU extracted from molecular dynamics calculations in explicit solvent. MD calculations predict a hydrogen bond between the hydroxyl proton of the butadiene moiety and the O-atom of the modified base.

Helicoidal Analysis

An analysis of the helicoidal parameters of the average solution structure was performed following MD calculations. Similarly to the *R*-diastereomer, the *S*-BD-N3-dU adduct caused a 24º bend in the duplex at the lesion site. Helicoidal base pair parameters were normal (Figure 3-24). The parameters for base-base interactions showed that at the lesion site there was a disruption in the shear, the opening, and the stagger (Figure 3-25). Finally, helicoidal parameters for global inter-base interactions showed that there was disruption primarily at the lesion site in the slide, tilt, twist and rise (Figure 3-26). Unlike the *R*-diastereomer, an increase in the rise was not observed.

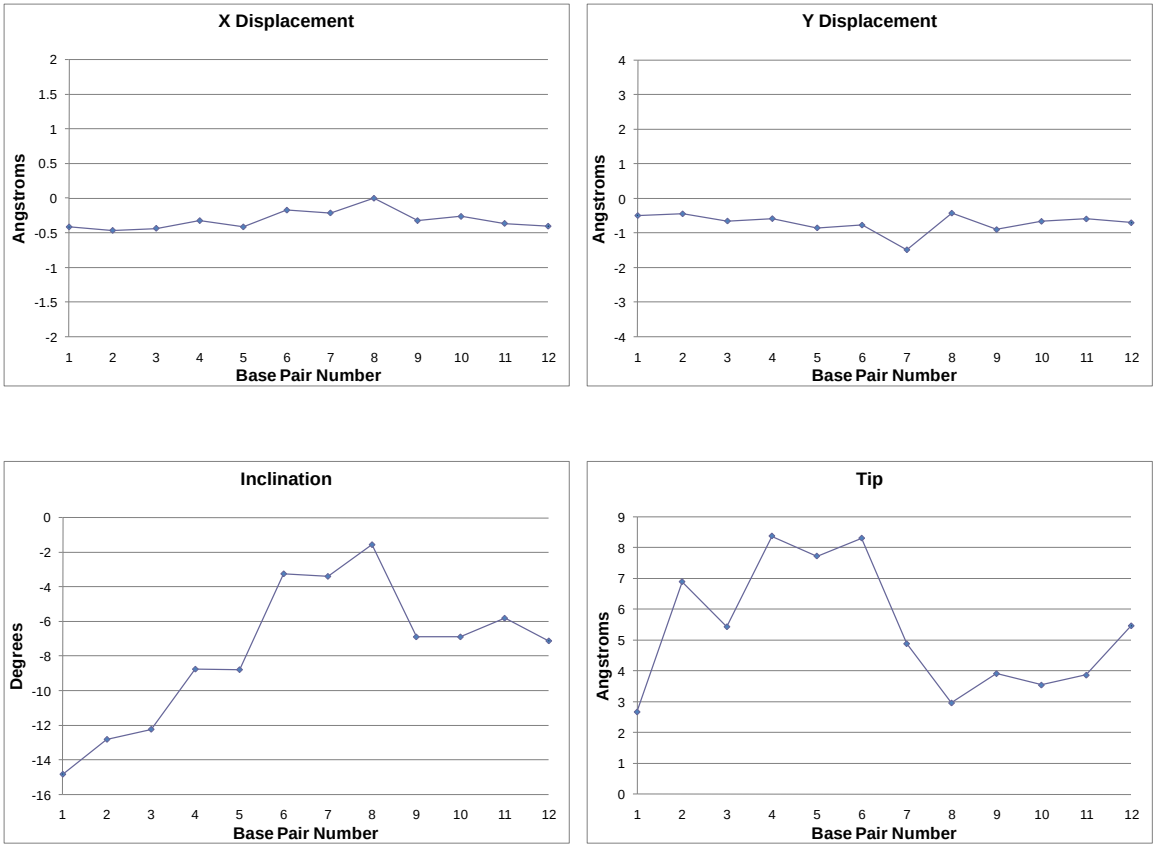


Figure 3-24: Global base pair helicoidal parameters. Helicoidal parameters for the *S*-BD-N3-dU modified oligonucleotide 5′-G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}<u>X</u>^{7}A^{8}G^{9}T^{10}C^{11}C^{12}-3′·5′-G^{13}G^{14}A^{15}C^{16}T^{17}A^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}-3′.

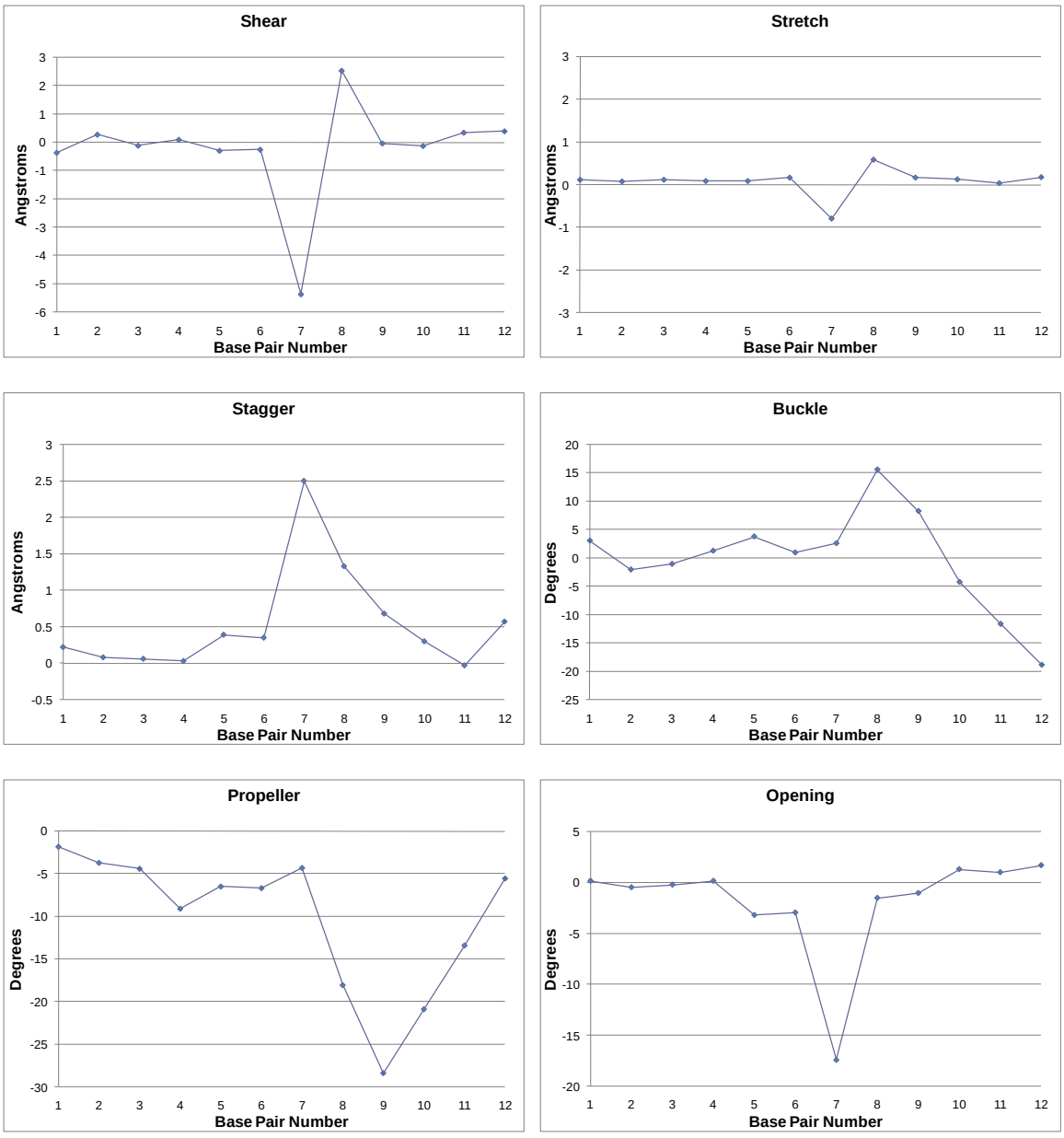


Figure 3-25: Global base-base helicoidal parameters. Helicoidal parameters for the *S*-BD-N3-dU modified oligonucleotide 5′-G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}<u>X</u>^{7}A^{8}G^{9}T^{10}C^{11}C^{12}-3′·5′-G^{13}G^{14}A^{15}C^{16}T^{17}A^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}-3′.

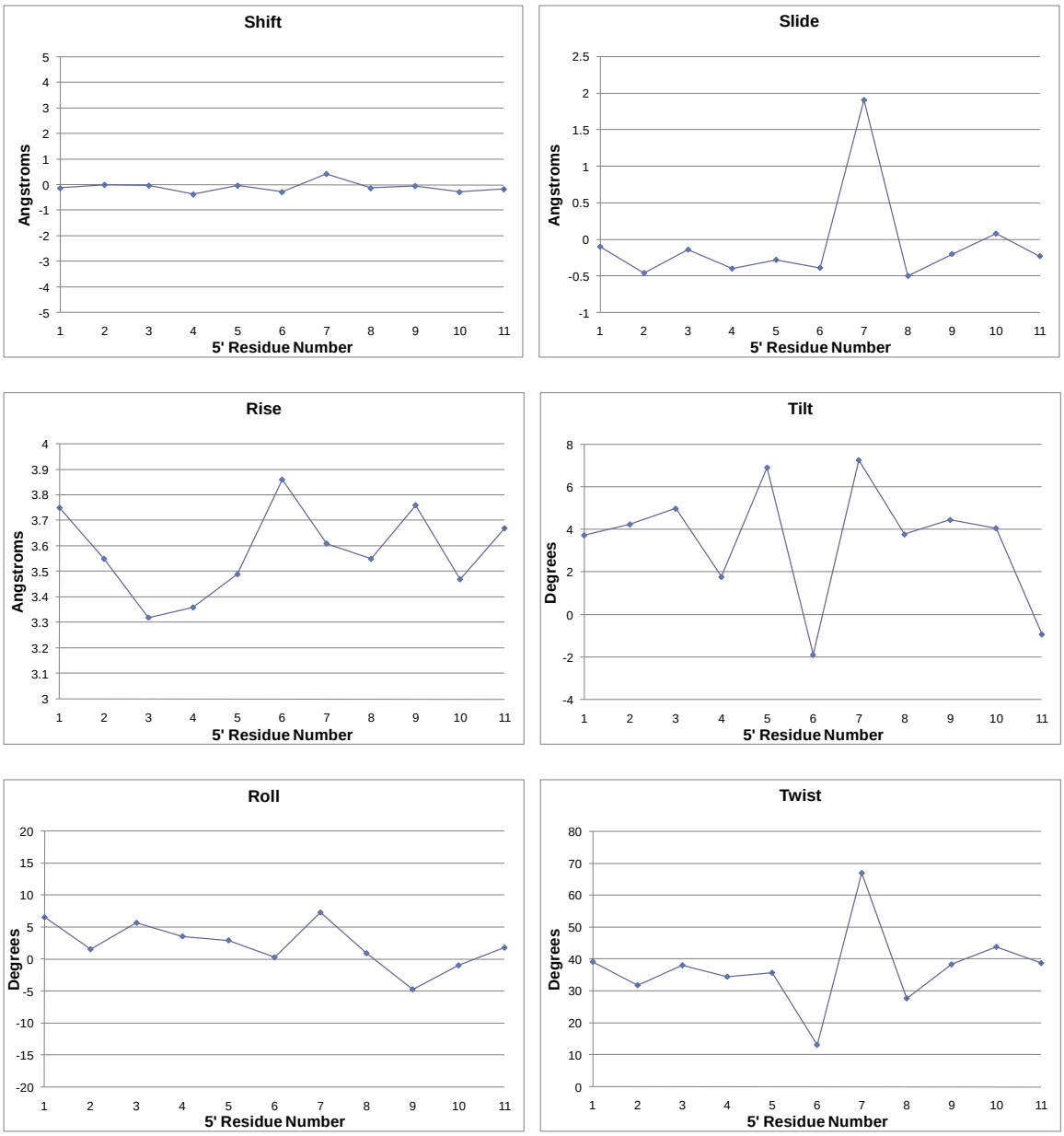


Figure 3-26: Global inter-base helicoidal parameters. Helicoidal parameters for the *S*-BD-N3-dU modified oligonucleotide 5′-G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}<u>X</u>^{7}A^{8}G^{9}T^{10}C^{11}C^{12}-3′·5′-G^{13}G^{14}A^{15}C^{16}T^{17}A^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}-3′.

Discussion

Mutagenesis studies in COS-7 mammalian cells indicated that the major mutations of the stereoisomeric N3-dU adducts arising from butadiene monoepoxide

were C to T transitions and C to A transversions. Stereoisomeric differences were observed with respect to these mutations as follows: while the *S*-diastereomer induced the insertion of adenine and thymine equally, the presence of the *R*-diastereomer yielded a 2:1 preference for the insertion of adenine. Therefore, it was hypothesized that the differences with respect to mutagenesis might be a result of structural differences in the orientation of the butadiene moiety when placed opposite different nucleotides. The first model for investigation was the dodecamer oligonucleotide, 5′-$G^1C^2T^3A^4G^5C^6\underline{X}^7A^8G^9T^{10}C^{11}C^{12}$-3′·5′-$G^{13}G^{14}A^{15}C^{16}T^{17}A^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$-3′, where both the *R*- and *S*-stereoisomers were site-specifically placed at the seventh position opposite dA.

Structural Analysis of the R-BD-N3-dU Adduct Opposite dA

NMR experiments provided valuable data, with several cross-peaks showing interactions between butadiene protons and other oligonucleotide protons. These data suggested that the two alpha protons would be closer in proximity to the 3′-neighbor, A^8, than the complementary base, A^{18}. Interactions with the H2 of both nucleotides placed the lesion in the minor groove. An interaction between the Hβ proton and the H1′ proton of T^{17} oriented Hβ toward the complementary strand. The Hγ, Hδ, and Hδ′ protons were oriented in the 5′-direction, producing an increase in the rise between X^7 and its 5′-neighbor C^6, as was observed in NOESY data. The modification at the seventh position created an opening in the helix. A disruption in hydrogen bonding was observed at both the lesion site and the 3′-neighboring base pair, A^8:T^{17}. This disruption in hydrogen bonding was supported by the absence of a dipolar interaction between T^{17} N3H and G^9

N3H as well as a broadened cross-peak for the interaction between A^8 H2 and T^{17} N3H. Finally, MD calculations predicted the possible formation of a hydrogen bond between the hydroxyl proton of the adduct and the O^4 atom of the modified base.

Structural Analysis of the S-BD-N3-dU Adduct Opposite dA

The *S*-BD-N3-dU modified oligonucleotide also produced quality NMR data, with several cross-peaks showing interactions between butadiene protons and other oligonucleotide protons. Unlike the *R*-diastereomer, the two alpha protons were relatively equidistant from the H2 proton of A^8 and A^{18}. These interactions with the H2 of both nucleotides placed the moiety in the minor groove. An interaction between the Hβ proton and the H1′ proton of T^{17} and A^{18} oriented Hβ toward the complementary strand. The different NOEs observed in the NMR data were indicative of a different orientation for the *S*-diastereomer compared to the *R*-diastereomer. The final structure shows that the Hγ, Hδ, and Hδ′ protons are directed toward the complementary strand, disrupting base pairing at the lesion site and at the A^8:T^{17} base pair. Like the oligonucleotide containing the *R*-BD-N3-dU adduct, the modification at the seventh position created an opening in the helix, substantiated by weak NOESY interactions between several aromatic and anomeric protons at the lesion site as well as the neighboring base pairs. No substantial increase in rise was observed. The butadiene moiety causes a disruption in hydrogen bonding at both the lesion site and the 3′-neighboring base pair, A^8:T^{17}. This disruption in hydrogen bonding is supported by the absence of an interaction between T^{17} N3H and G^9 N3H as well as a broadened cross-peak for the interaction between A^8 H2 and $T^{`17}$ N3H. Finally, MD calculations also

predicted a hydrogen bond between the hydroxyl proton of the adduct and the O^2 atom of the modified base.

Biological Significance

Site-specific mutagenesis studies carried out in COS-7 mammalian cells indicated that the N3-dU adducts arising from butadiene monoepoxide were highly mutagenic [139]. These adducts which arise from BD-N3-dC adducts may be highly mutagenic due to their presence in DNA for extended periods of time [128]. Furthermore, N3-dU adducts are not expected to be repaired by uracil DNA glycosylase [63]. Analyses of mutations in *lac*I and *lac*Z mice indicate that mutations at G:C base pairs are noteworthy events in BD-induced mutagenicity [140]. Therefore, although these lesions have not yet been isolated *in vivo*, due to their high mutagenicity, they should not be overlooked.

Mutagenesis studies also showed that the insertion of dA opposite the lesion site occurs at a higher percentage in the presence of the *R*-BD-N3-dU adduct versus the *S*-BD-N3-dU adduct [168]. The propensity of trans-lesion polymerases to incorporate dA opposite the *R*-diastereomer twice as often as the *S*-diastereomer might be due to the differential orientations of the functional groups in the butadiene moiety. This hypothesis has been investigated using NMR spectroscopy as well as X-ray crystallography. Although X-ray crystallographic structures of Dpo4-DNA-dNTP complexes yielded excellent structures, the results of these studies indicated that the stereoisomeric BD-N3-dU adducts are not substrates for this polymerase [172]. Lesion bypass studies point to polymerases η and ζ as polymerases capable of bypassing the lesion site [139]. For that

reason, it will be of significant interest to investigate Pol η-DNA-dNTP complexes in the future.

Summary

Both the *R*- and *S*-BD-N3-dU adducts arising from butadiene monoepoxide caused major perturbations at the lesion site. However, differences were observed with respect to the orientation of the butadiene moieties. For the *R*-diastereomer, the butadiene moiety was located in the minor groove and oriented in the 5′-direction; whereas, the butadiene moiety of the *S*-diastereomer was oriented in the minor groove toward the complementary strand. A hydrogen bond was predicted between the proton of the hydroxyl group and the O^4 and O^2 atoms of the *R*- and *S*-BD-N3-dU modified bases, respectively.

CHAPTER IV

SOLUTION STRUCTURE OF THE STEREOISOMERIC N3-(2-HYDROXY-3-BUTEN-1-YL)-2'-DEOXYURIDINE ADDUCTS ARISING FROM BUTADIENE MONOEPOXIDE OPPOSITE DEOXYTHYMINE IN A DODECAMER OLIGONUCLEOTIDE

Introduction

The results of site-specific mutagenesis studies carried out in COS-7 mammalian cells using the stereoisomeric N3-dU BD adducts suggested that for the *R* and *S* adducts, more than thirty percent of the mutations produced by these lesions yielded C→A transversions [168]. The results from the mutagenesis studies also suggested that when inserting dNTPs opposite the lesion site, stereospecific differences would occur. For the *R* diastereomer, there was a 2:1 preference for the insertion of dA versus dT; whereas, for the *S* diastereomer, relatively equal amounts of both nucleotides were inserted.

It was hypothesized that a structural difference between oligonucleotides containing the *R* isomers annealed to differing complementary strands might explain the differences in nucleotide insertion. For the *R*-BD-N3-dU adduct placed opposite dT, a structural difference at the lesion site when compared to the same lesion opposite dA was expected. Whereas, for the *S*-BD-N3-dU adduct opposite dT, the same adduct orientation as the *S*-BD-N3-dU adduct opposite dA was expected. The model that was used for these studies was the dodecamer oligonucleotide 5'-d($G^1C^2T^3A^4G^5C^6\underline{X}^7A^8G^9T^{10}C^{11}C^{12}$)-3'•5'-d($G^{13}G^{14}A^{15}C^{16}T^{17}\underline{T}^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$)-3', where X is either the *R*- or *S*-BD-N3-dU adduct.

NMR analysis reveals that the overall structures of both modified oligonucleotides are similar. For both adducts there is an increase in the rise between the adduct and its 3'- and 5'-neighboring bases. Focusing on the lesion sites of both oligonucleotides, the orientation of the butadiene moiety is similar for both diastereomers, having no directionality, but rather being oriented toward the complementary stand and in the minor groove. The butadiene moiety lies between T^{18} and its 3'-neighbor T^{17}. The position of the lesion causes a change in the position of both T^{18} and T^{17}, leading to disruptions in hydrogen bonding between the X^7:T^{18} and A^8:T^{17} base pairs.

It was hypothesized that structural differences would be present when comparing the *R*-diastereomer in differing sequences. When the oligonucleotides placing the *R*-BD-N3-dU adduct opposite dA and dT are compared, structural differences at the lesion site are observed. For the *S*-diastereomer, the orientation of the butadiene moiety has is not related to its complementary base. These data might indicate that a specific orientation of the BD-N3-dU adduct is preferred during trans-lesion replication.

Results

R Diastereomer

Sample Purity

The double-stranded *R*-BD-N3-dU modified oligonucleotide, 5'-d($G^1C^2T^3A^4G^5C^6X^7A^8G^9T^{10}C^{11}C^{12}$)-3'·5'-d($G^{13}G^{14}A^{15}C^{16}T^{17}T^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$)-3',

was purified using the HPLC protocol described in Chapter II. Enzyme hydrolysis showed that the modified oligonucleotide was enantiomerically pure. A hydroxylapatite column was used for the separation of any remaining single-stranded DNA from the duplex sample. The presence of only duplex DNA in the solution sample was confirmed using capillary gel electrophoresis. MALDI-TOF mass spectrometry along with COSY and NOESY NMR experiments confirmed the composition of the two single strands. NMR experiments at a frequency of 800.23 MHz provided valuable data for the assignment and volume integration of oligonucleotide protons.

Non-exchangeable protons

For the observation of non-exchangeable protons in NMR spectra, the sample was prepared as discussed in Chapter II. Assignments of the purine and pyrimidine aromatic protons, the thymine methyl protons, and the deoxyribose H1', H2', H2", and H3' protons were successfully completed. Due to overlapping peaks and the effects of spin diffusion, the H4', H5', and H5" protons were only partially assigned. The chemical shift values of the non-exchangeable oligonucleotide protons are listed in Appendix A. Expansion plots of the NOESY spectrum for the *R* diastereomer are shown in Figure 4-1.

In the modified strand of the *R* stereoisomer (Figure 4-1A), the walk was completed without interruption. However, the interaction between C^6 H1' and X^7 H6 yielded a cross-peak with a weakened intensity. The change in the intensity of this cross-peak compared to the same cross-peak in the unmodified oligonucleotide suggested that the orientation of the butadiene lesion causes an increase in the distance between these protons. A decreased intensity was also observed in the cross-peak for the interaction

between X^7 H1' and A^8 H8. In the complementary strand, the "walk" also remained uninterrupted. The interaction between T^{17} H1' and T^{18} H6 produced a weak cross-peak (Figure 4-1B). A decrease in the intensity of the T^{18} H1':G^{19} H8 cross-peak was also observed. These results suggest a localized perturbation for the *R* stereoisomer, and predict a structure with an increase in the rise between the lesion site and its 5'- and 3'-neighboring base pairs.

A chemical shift analysis of the anomeric H1' and aromatic base protons was carried out in order to compare the chemical shift values of the DNA protons in the modified duplex to those in the unmodified oligodeoxynucleotide sequence. The chemical shift changes for these protons are shown in Figure 4-2. In the modified strand, only slight movement upfield was observed for the X^7 H6 and A^8 H1' protons. For both protons, the lesion produced a shift of 0.125 ppm and 0.105 ppm, respectively. In the complementary strand, a change in chemical shifts occurred at T^{17} H1', with a downfield shift of 0.109 ppm. This variation in chemical shift values was indicative of the lesion being proximal to the complementary strand, such that the chemical environment at these nucleotides was changed. The observation of chemical shift changes primarily at the lesion site suggested a localized perturbation to the structure of the oligonucleotide.

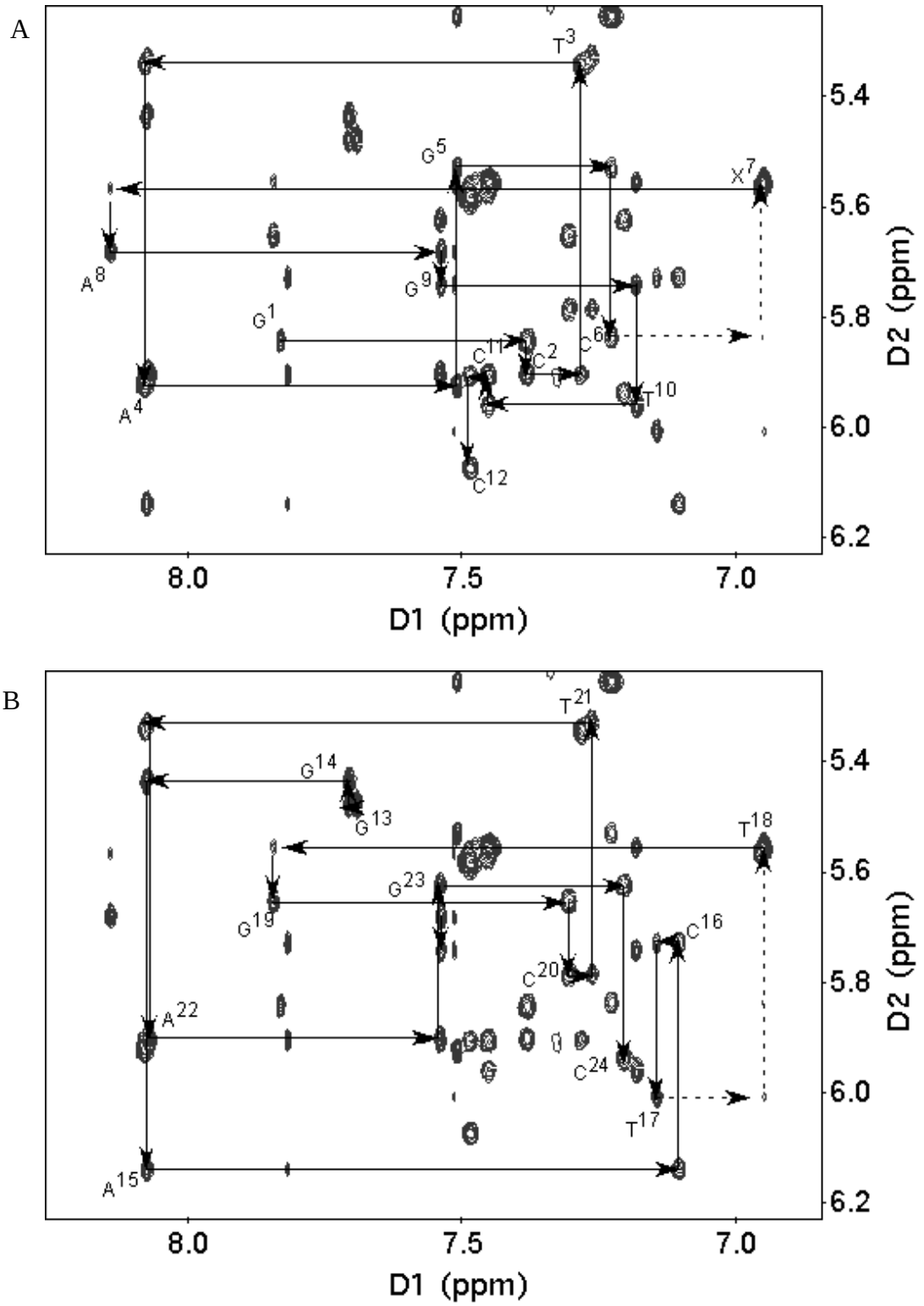


Figure 4-1: NOESY expansion plot of the anomeric proton to aromatic proton region for the R-BD-N3-dU modified sequence 5'-$G^1C^2T^3A^4G^5C^6\underline{X}^7A^8G^9T^{10}C^{11}C^{12}$-3'·5'-$G^{13}G^{14}A^{15}C^{16}T^{17}T^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$-3'. (A) Nucleotides G^1-C^{12} of the modified oligonucleotide containing *R*-BD-N3-dU. (B) Nucleotides G^{13}-C^{24} of the complementary strand. NMR data were acquired at a mixing time of 250 ms and a frequency of 800.23 MHz.

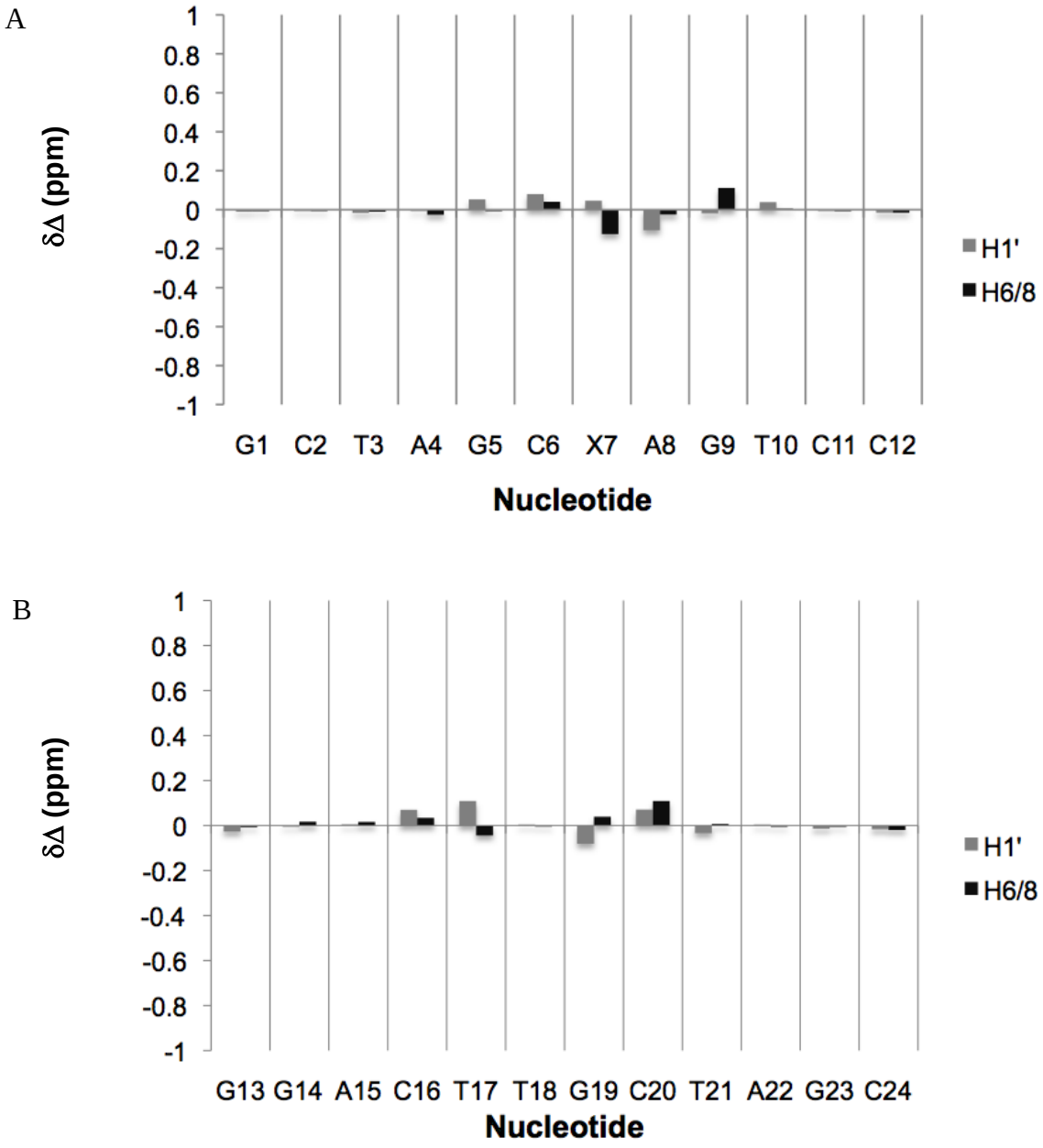


Figure 4-2: Chemical shift perturbations in the aromatic and anomeric protons of the R-BD-N3-dU modified sequence 5'-$G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}\underline{X}^{7}A^{8}G^{9}T^{10}C^{11}C^{12}$-3'·5'-$G^{13}G^{14}A^{15}C^{16}T^{17}T^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$-3' relative to the unmodified oligodeoxynucleotide. (A) Nucleotides G^{1}-C^{12} of the modified oligonucleotide containing *R*-BD-N3-dU. (B) Nucleotides G^{13}-C^{24} of the complementary strand.

Butadiene Protons

The *R*-BD-N3-dU modified deoxyuracil contains six butadiene protons, Hα, Hα′, Hβ, Hγ, Hδ, and Hδ′, which are labeled in Figure 3-1. Hβ and Hγ were assigned according to dipolar interactions in a NOESY experiment collected at a mixing time of 250 ms. Hδ and Hδ' were assigned using coupling constants from a DQF COSY experiment. Unambiguous assignments of the protons on the alpha carbon, Hα and Hα', was not achievable. However, the intensities in NOESY spectra between Hα and Hα' and other protons at all four mixing times are on the same order, suggesting that an ambiguous assignment of the protons would produce identical structures. The chemical shift values for the butadiene protons are listed in Appendix A.

Key evidence for the orientation of the butadiene moiety was found in a NOESY spectrum. NOESY data showed strong interactions between Hα, Hα', and Hβ of butadiene and A^8 H2 (Figure 4-3). A weak interaction between Hβ and T^{17} H1' was also observed in the anomeric region of the spectrum. These data suggested that the butadiene moiety would be oriented in the minor groove and the direction of the butadiene moiety would be toward the complementary strand.

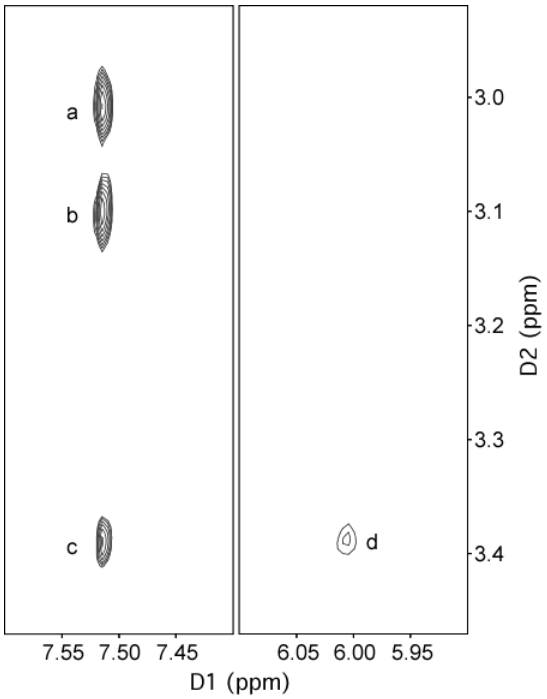


Figure 4-3: Dipolar interactions between the R-BD moiety and other oligonucleotide protons. NOESY cross-peaks: a, Hα-A^8 H2; b, Hα′-A^8 H2; c, Hβ-A^8 H2; d, Hβ-T^{17} H1′. NMR data were acquired at a mixing time of 250 ms and a frequency of 800.23 MHz.

Exchangeable Protons

The exchangeable imino protons were observed in a NOESY spectrum acquired according to methods described in Chapter II. NOESY data acquired at 7 °C produced quality spectra with well-resolved peaks. An expansion plot of the NOESY spectrum showing the imino proton region (12-15 ppm) is shown in Figure 4-4. The terminal guanine bases were not observed in the spectrum due to fraying at the end of the helices. In the imino region, the cross-peaks for G^{23} and G^{14} were overlapped; however, sequential assignment techniques allowed both cross-peaks to be assigned.

The imino protons (Figure 4-4B) from G^{23} to G^{14} were assigned according to the methods described in Chapter III. NOESY connectivity was completed without disruption from G^{23} to G^{19}. The modification at the N3 position of X^7 disrupts Watson-Crick hydrogen bonding, causing the NOE connectivity to be broken at G^{19}. The imino cross-peaks for T^{17} and T^{18} were not observed in this spectrum. Therefore, the "walk" was resumed at G^9 and continued without interruption to G^{14}.

In the H2 region of the spectrum (Figure 4-4A), the cross peak between T^{17} NH3 and A^8 H2 was not observable. These data agreed with the previous NOESY data, suggesting that the presence of butadiene causes a significant perturbation to T^{17}. The absence of imino cross-peaks for T^{17} and T^{18} as well as a dipolar interaction between the H2 proton of A^8 and the imino proton of T^{17} suggested that the protons were in rapid exchange with the solvent due to a weakened hydrogen bond and perturbed base-pairing interactions.

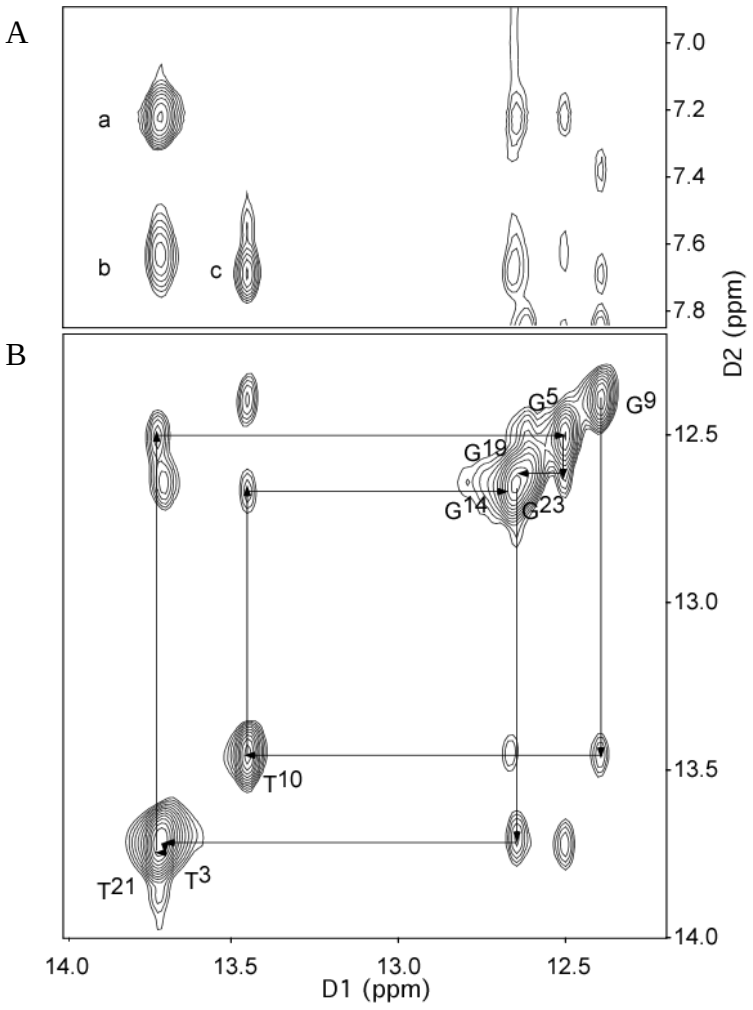


Figure 4-4: NOESY expansion plot of the imino proton region for the R-BD-N3-dU modified sequence 5′-$G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}\underline{X}^{7}A^{8}G^{9}T^{10}C^{11}C^{12}$-3′·5′-$G^{13}G^{14}A^{15}C^{16}T^{17}T^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$-3′. (A) NOESY cross-peaks: a, A^{4} H2-T^{21} N3H; b, A^{22} H2-T^{3} N3H; c, A^{8} H2-T^{17} N3H. (B) Sequential connectivity of imino protons starting at G^{23} and ending at G^{14}. Sequential connectivity is disrupted at G^{19} and resumed at G^{9}. NMR data were acquired at 7 °C at a mixing time of 250 ms and a frequency of 800.23 MHz.

Structural Refinement

Twenty initial structures were used as starting coordinate files for the structural refinement of the duplex DNA containing the *R*-BD-N3-dU adduct. These structures were built as described in Chapter II and energy minimized prior to rMD calculations. Restrained molecular dynamics calculations were carried out in implicit solvent using the simulated annealing protocol. A total of 545 distance restraints, comprised of 310 intra-residue, 204 inter-residue, and 31 butadiene restraints, were used to restrain MD

calculations. The distance restraints were evenly distributed amongst the residues in the oligonucleotide and are listed in Appendix B. To restrain the backbone and sugar pucker, a total of 90 and 110 empirical torsion angle restraints were used, respectively. In addition to the distance and angle restraints, 44 Watson-Crick hydrogen-bonding restraints were used, restraining the hydrogen bonds in all of the base pairs except X^{7}:T^{18} and A^{8}:T^{17}. These restraints are listed in Appendix C.

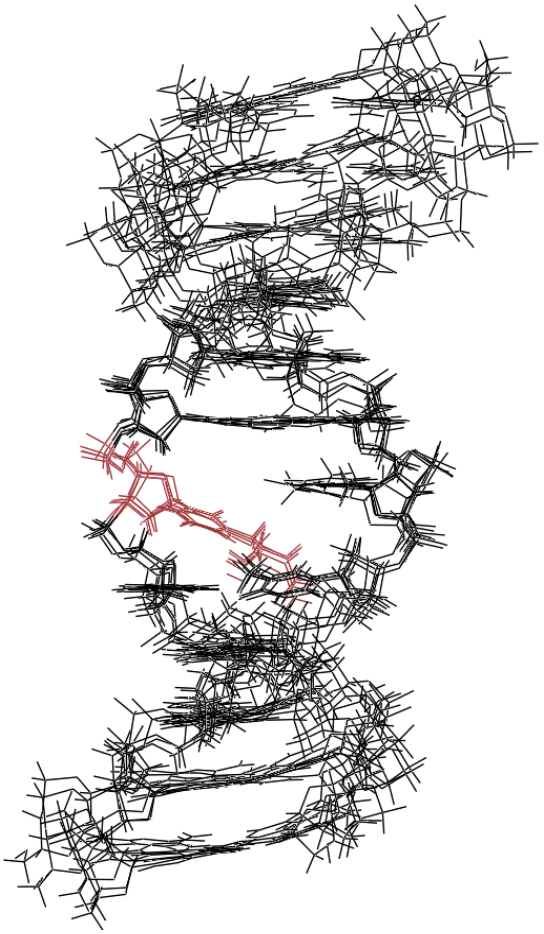


Figure 4-5: Overlay of the five convergent structures of the R-BD-N3-dU modified oligonucleotide calculated using a simulated annealing rMD protocol. The modified base is shown in red. The rmsd of the convergent structures was 1.47 Å.

Prior to the rMD calculations, the root mean square deviation (rmsd) of the twenty starting structures was 3.52 Å. Following 20 ps of MD calculations in implicit solvent, the structural coordinates which emerged from the calculations were compared using pairwise rmsd. Five of the twenty starting structures converged to give an rmsd value of 1.47 Å. An overlay of the five convergent structures is shown in Figure 4-5.

CORMA [155] calculations were run on the average structure in order to determine the accuracy of the structure with respect to the acquired NMR data. The total R_1x value was $8.0x10^{-2}$, with an individual nucleotide range from $2.5x10^{-2}$ to $1.8x10^{-1}$. These values suggested that

the average solution structure agrees well with the NMR data. The results from the CORMA calculations with respect to the individual nucleotides are shown in Figure 4-6.

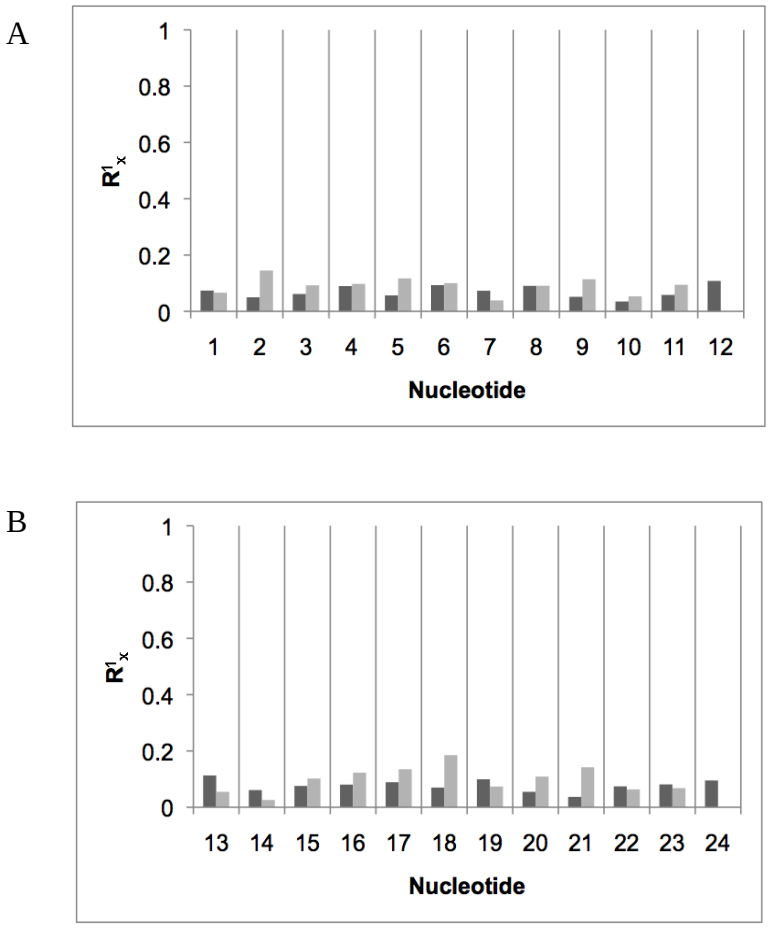


Figure 4-6: CORMA calculations on the average *R*-BD-N3-dU modified oligodeoxynucleotide calculated using a simulated annealing rMD protocol. (A) Nucleotides G^{1}-C^{12} of the modified oligonucleotide containing *R*-BD-N3-dU. (B) Nucleotides G^{13}-C^{24} of the complementary strand. The black bars represent intranuclear sixth root residuals and the gray bars represent internuclear sixth root residuals.

Structure of the R-BD-N3-dU Adduct Opposite dT

The five convergent structures were averaged to give a mean oligonucleotide structure. The average structure shows that the perturbation to the oligonucleotide remains localized at the lesion site. The lesion site of the average structure is shown in

Figure 4-7. The butadiene moiety possesses no directionality; rather it is fixed in the minor groove and oriented toward the complementary strand. The adduct positions itself between T^{17} and T^{18} in the complementary strand. The placement of the butadiene moiety causes a disruption in hydrogen bonding between X^7 and T^{18}, as well as A^8 and T^{17}. The butadiene protons, Hγ, Hδ, and Hδ′, are fixed such that T^{17} is pushed into the major groove. Base-stacking interactions are also perturbed between X^7 and C^6.

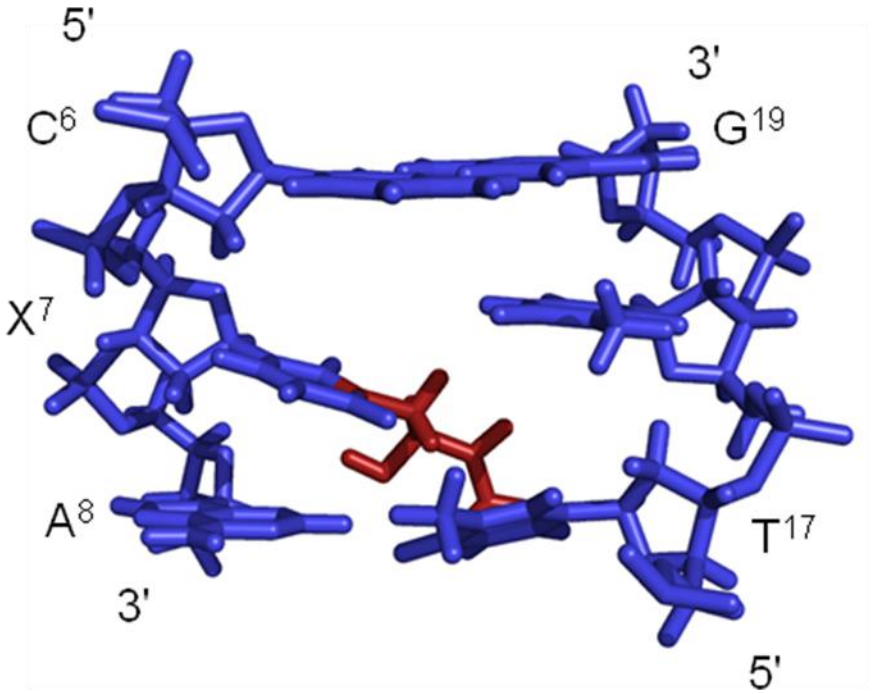


Figure 4-7: Lesion site of the average structure of the oligodeoxynucleotide containing R-BD-N3-dU opposite dT calculated using a simulated annealing rMD protocol. The butadiene moiety (red) is oriented in the minor groove between T^{18} and T^{17}.

Hydrogen Bonding Analysis

MD calculations in explicit solvent were performed on the average structure in order to investigate hydrogen bonding between the butadiene moiety and the modified

base. The average structure extracted from simulated annealing calculations in implicit solvent was placed in a truncated octahedral TIP3P solvent box after the addition of sodium counter ions. Prior to 5 ns of MD calculations, the solvated molecule was energy minimized and equilibrated according to standard methods. Hydrogen-bonding occupancies were extracted from MD trajectories using PTRAJ (AMBER 8.0) [153]. The percent occupancy for the hydrogen bond between the hydroxyl proton and the O^4 atom on the modified nucleotide was 6%. The distance between the hydroxyl proton and O^4 atom was 2.77 Å and the O-H-O angle was determined to be 21.33º. These data do not predict hydrogen bonding interactions at this site.

Helicoidal Analysis

An analysis of the helicoidal parameters of the average modified structure was performed following rMD calculations. In the sequence placing the modification opposite dT, the *R*-BD-N3-dU adduct caused a 16º bend in the duplex at the lesion site. Helicoidal base pair parameters were normal (Figure 4-8). The parameters for base-base interactions showed that at the lesion site there was a disruption in the shear and the stagger, and that the modification causes a 25º opening (Figure 4-9). Finally, helicoidal parameters for global inter-base interactions showed that there was a 1.5 Å increase in the rise at the lesion site (Figure 4-10).

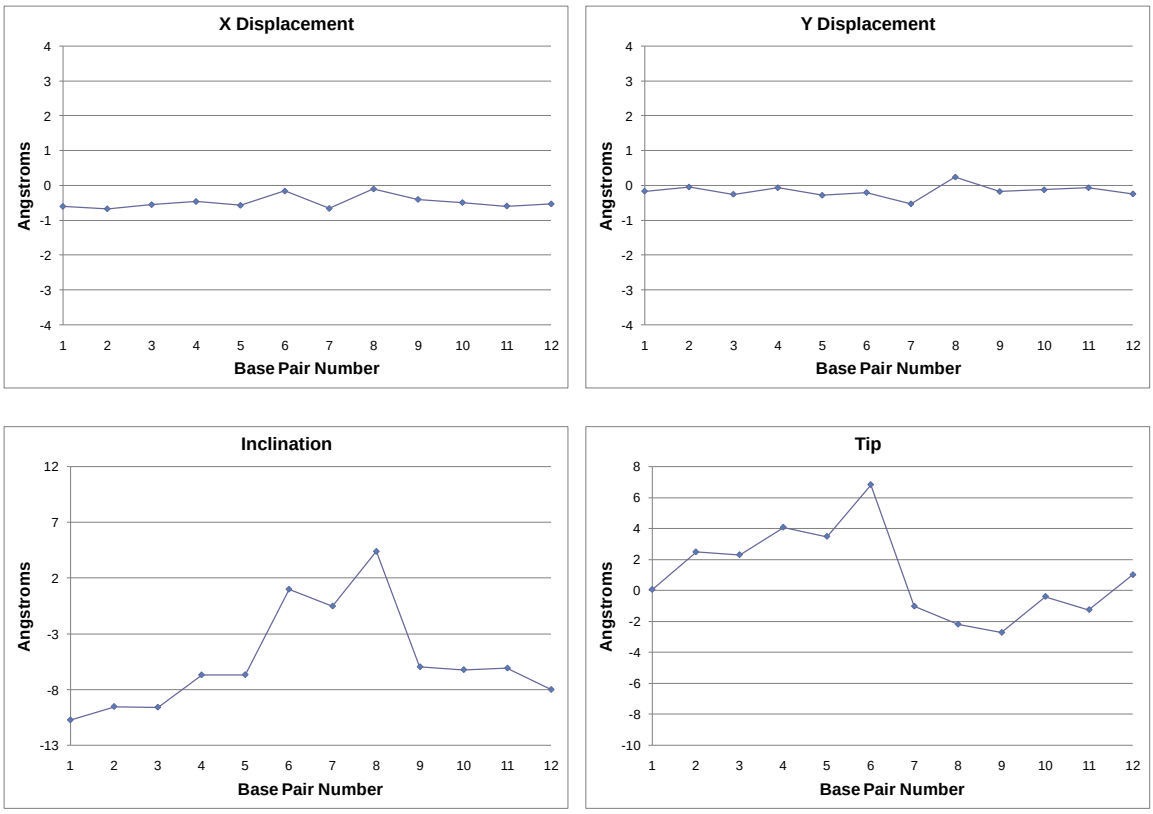


Figure 4-8: Global base pair helicoidal parameters. Helicoidal parameters for the *R*-BD-N3-dU modified oligonucleotide 5′-G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}X^{7}A^{8}G^{9}T^{10}C^{11}C^{12}-3′·5′-G^{13}G^{14}A^{15}C^{16}T^{17}T^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}-3′.

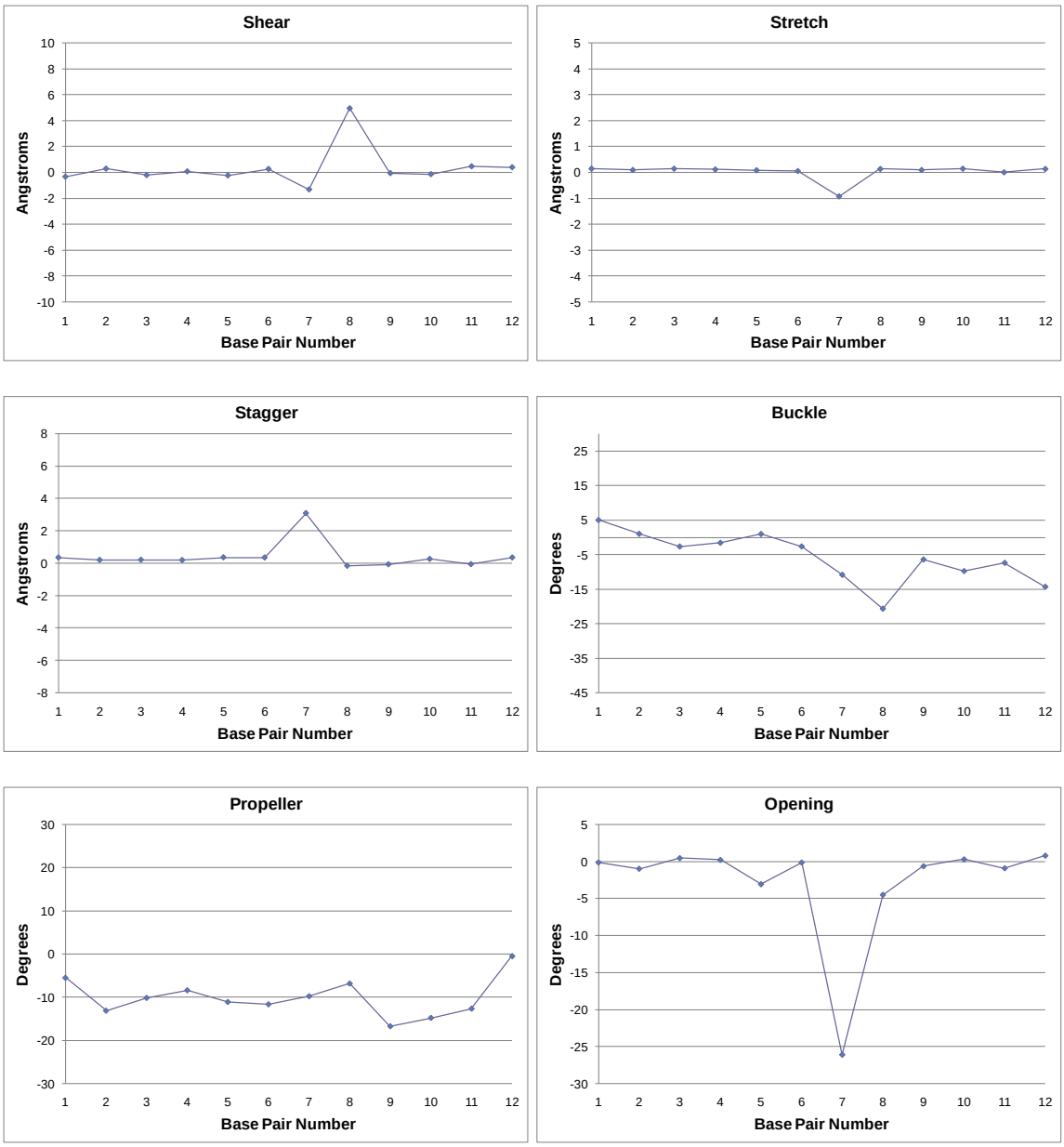


Figure 4-9: Global base-base helicoidal parameters. Helicoidal parameters for the *R*-BD-N3-dU modified oligonucleotide 5′-G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}X^{7}A^{8}G^{9}T^{10}C^{11}C^{12}-3′·5′-G^{13}G^{14}A^{15}C^{16}T^{17}T^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}-3′.

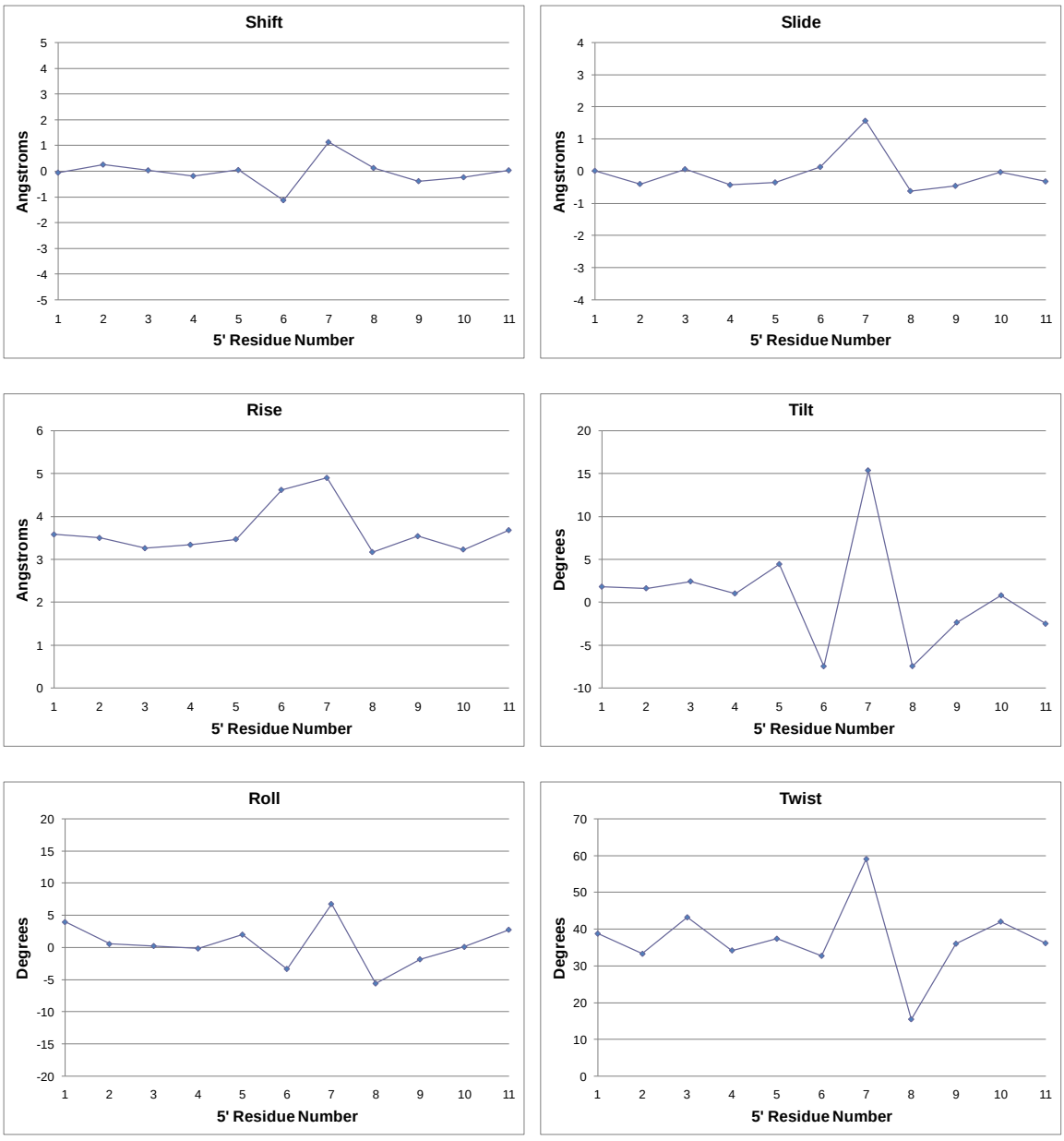


Figure 4-10: Global inter-base helicoidal parameters. Helicoidal parameters for the *R*-BD-N3-dU modified oligonucleotide 5′-$G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}\underline{X}^{7}A^{8}G^{9}T^{10}C^{11}C^{12}$-3′·5′-$G^{13}G^{14}A^{15}C^{16}T^{17}T^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$-3′.

S-Diastereomer

Sample Purity

The double-stranded *S*-BD-N3-dU modified oligonucleotide, 5'-d($G^1C^2T^3A^4G^5C^6\underline{X}^7A^8G^9T^{10}C^{11}C^{12}$)-3'·5'-d($G^{13}G^{14}A^{15}C^{16}T^{17}T^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$)-3', was purified using the HPLC protocol described in Chapter II. Enzyme hydrolysis showed that the oligonucleotide was enantiomerically pure. The composition of the two single strands was verified using MALDI-TOF mass spectrometry and COSY and NOESY NMR spectroscopy. Using a hydroxylapatite column, the annealed duplex was separated from any remaining single-stranded DNA. The presence of only duplex DNA in the solution sample was confirmed using capillary gel electrophoresis. NMR experiments at a frequency of 800.23 MHz provided valuable data.

Non-exchangeable Protons

For the observation of non-exchangeable protons in NMR data, the sample was prepared as discussed in Chapter II. The complete assignment of the purine and pyrimidine aromatic protons, the thymine methyl protons, and the deoxyribose H1′, H2', H2", and H3' protons was achieved. Due to overlapping peaks and spin diffusion at higher mixing times, only partial assignments were made for the deoxyribose H4', H5', and H5" protons. Expansion plots of the NOESY spectrum for the *S* diastereomer are shown in Figure 4-11.

Sequential connectivity was observed throughout the modified strand. A significant observation in the modified strand (Figure 4-11A) was a decrease in the

intensity between C^6 H1' and X^7 H6. A weak cross-peak for the interaction between X^7 H1' and A^8 H8 was also observed. These weakened cross-peaks were indicative of a localized perturbation around the lesion site and its 5'- and 3'-neighbor base pairs, as well as an increase in the rise between these nucleotides as a result of the butadiene modification. In the complementary strand, connectivity remained unbroken throughout the entirety of the strand. A cross peak having less intensity corresponding to the interaction between T^{17} H1' and T^{18} H6 was observed (Figure 4-11B). The interaction between T^{18} H1′ and G^{19} H8 also showed a weakened cross-peak. These data suggest, that like the *R*-diastereomer, an increase in the rise between the lesion and its neighboring base pairs will be observed.

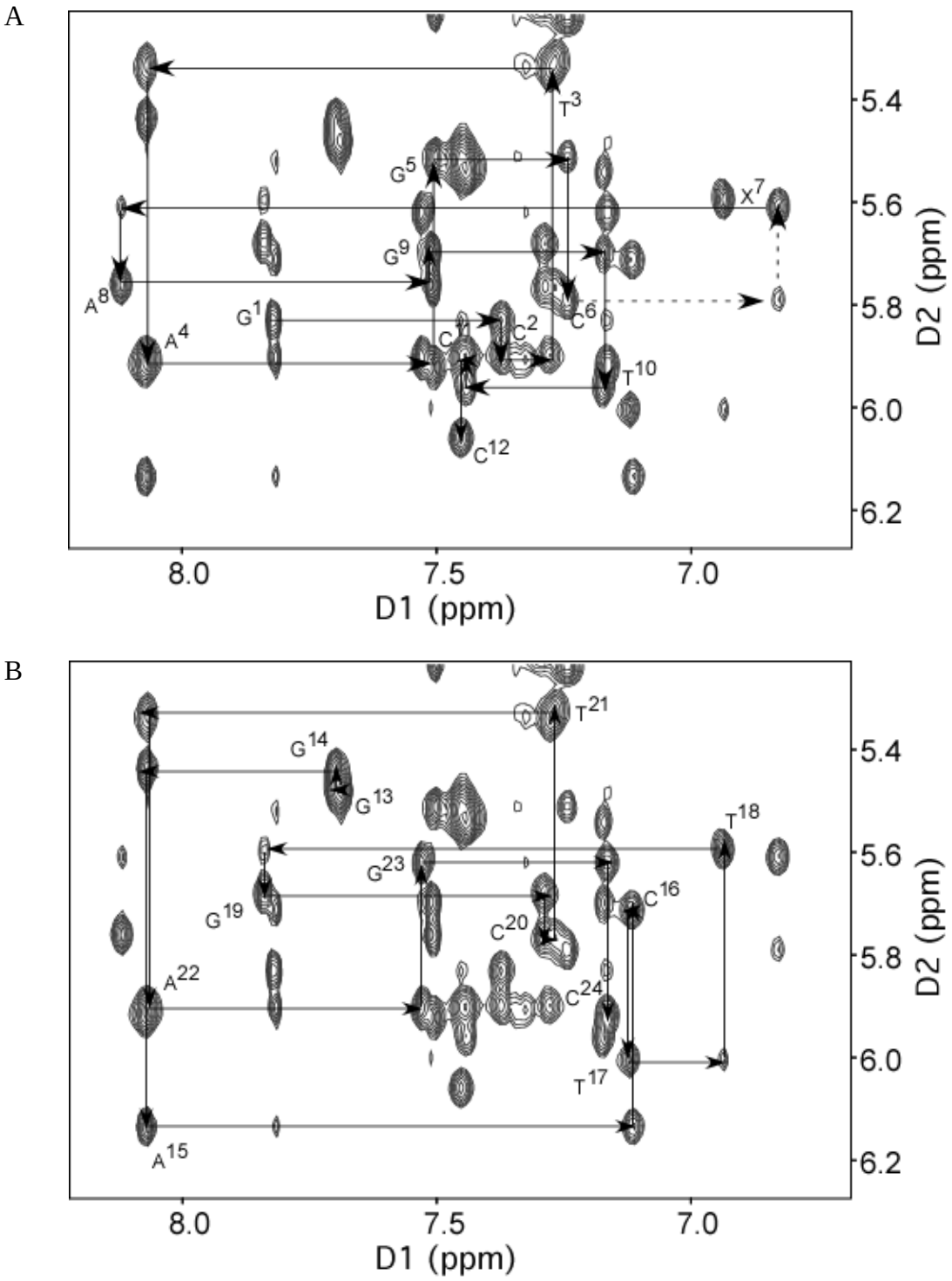


Figure 4-11: NOESY expansion plot of the anomeric proton to aromatic proton region for the S-BD-N3-dU modified sequence 5′-$G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}\underline{X}^{7}A^{8}G^{9}T^{10}C^{11}C^{12}$-3′·5′-$G^{13}G^{14}A^{15}C^{16}T^{17}T^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$-3′. (A) Nucleotides G^{1}-C^{12} of the modified oligonucleotide containing *S*-BD-N3-dU. (B) Nucleotides G^{13}-C^{24} of the complementary strand. NMR data were acquired at a mixing time of 250 ms and a frequency of 800.23 MHz.

A chemical shift analysis was performed on the anomeric H1′ protons and the purine and pyrimidine aromatic protons in the modified oligonucleotide, comparing the chemical shift values of the protons in the modified duplex to those in the unmodified duplex. The changes in chemical shift values are shown in Figure 4-12. In the modified strand, the largest perturbation occurred at the modified nucleotide, X^7, where the H6 proton was shifted upfield by 0.25 ppm. In the complementary strand, a chemical shift change was observed for the H1′ proton of T^{17}, with a downfield shift of 0.1 ppm.

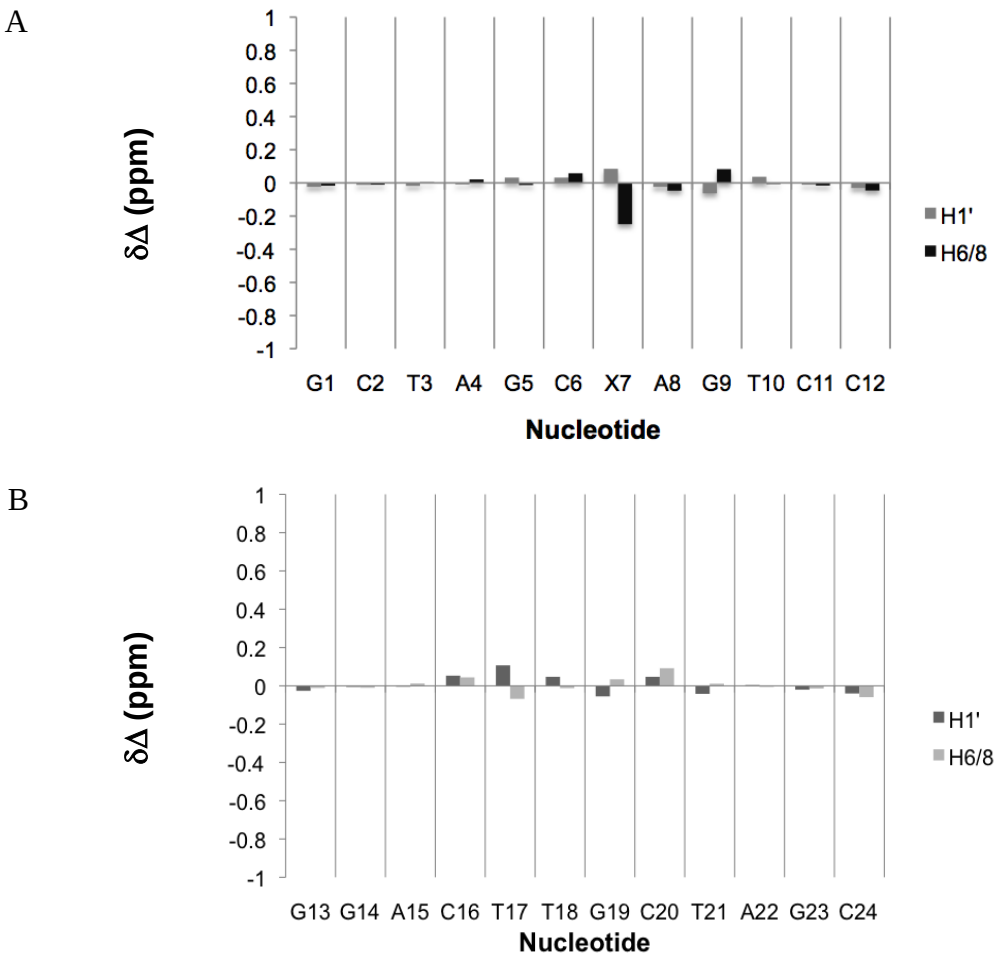


Figure 4-12: Chemical shift perturbations in the aromatic and anomeric protons of the S-BD-N3-dU modified sequence 5'-$G^1C^2T^3A^4G^5C^6\underline{X}^7A^8G^9T^{10}C^{11}C^{12}$-3'·5'-$G^{13}G^{14}A^{15}C^{16}T^{17}T^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$-3' relative to the unmodified oligodeoxynucleotide. (A) Nucleotides G^1-C^{12} of the modified oligonucleotide containing *R*-BD-N3-dU. (B) Nucleotides G^{13}-C^{24} of the complementary strand.

Butadiene Protons

The entire ^{1}H spin system of the butadiene moiety was observed in a TOCSY spectrum at a mixing time of 80 ms. Hβ and Hγ were assigned using TOCSY and NOESY experiments. A DQF-COSY experiment was used to assign Hδ and Hδ'. Like those on the *R*-BD-N3-dU adduct, the protons on the alpha carbon, Hα and Hα′, were not able to be unambiguously assigned. Therefore, ambiguous assignments were made for the alpha protons. The chemical shift values for the butadiene protons are listed in Appendix A.

Interactions between butadiene protons and other nucleotide protons were observed in the H1' and H2 regions of the NOESY spectrum. Those NOESY cross-peaks are shown in Figure 4-13. Interactions were observed between A^{8} H2 and the Hα, Hα′ and Hβ protons. An interaction between T^{17} H1' and Hβ was also observed in the NOESY spectrum. The proximity of these butadiene protons to A^{8} H2 indicates that the butadiene moiety will be oriented in the minor groove of the duplex.

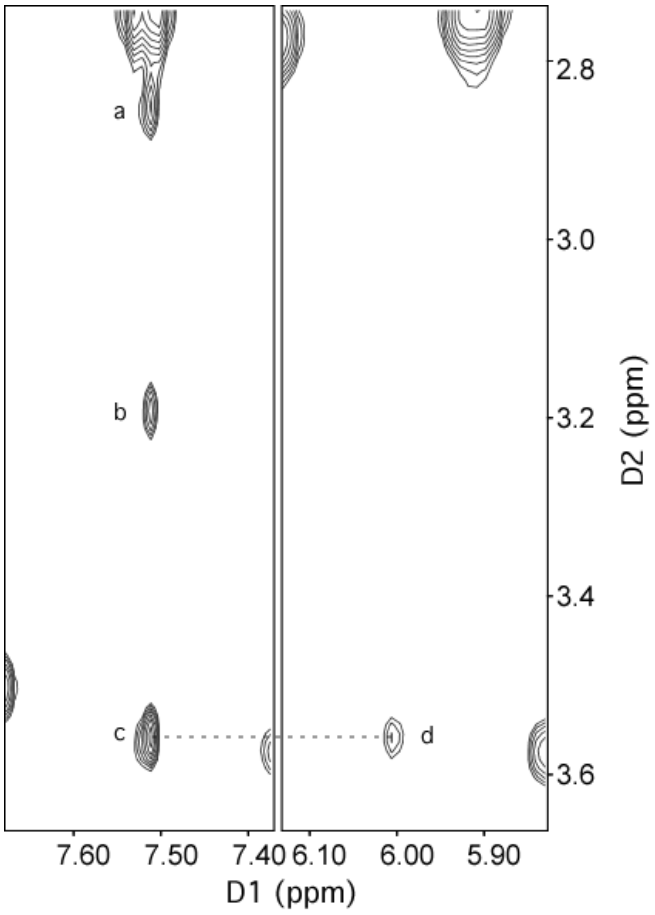


Figure 4-13: Dipolar interactions between the R-BD moiety and other oligonucleotide protons. NOESY cross-peaks: a, Hα-A^{8} H2; b, Hα′-A^{8} H2; c, Hβ-A^{8} H2; d, Hβ-T^{17} H1′. NMR data were acquired at a mixing time of 250 ms and a frequency of 800.23 MHz.

Exchangeable Protons

The exchangeable imino protons were observed in a NOESY spectrum acquired according to methods described in Chapter II. NOESY experiments were acquired at 7 °C and provided excellent spectra with well-resolved peaks. The imino cross-peaks of T^{3} and T^{21}, as well as G^{14} and G^{23} were slightly overlapped. An expansion plot displaying the imino proton region (12-15 ppm) is shown in Figure 4-14B. The imino protons of the terminal guanines, G^{1} and G^{13}, were not observable in the two-dimensional spectrum.

The imino protons were assigned by beginning at the imino proton of G^{23} and continuing to the T^{3} imino proton. The T^{3} imino proton was then "walked" to that of T^{21}, and the assignments were completed to G^{19}. The modification at the N3 position of X^{7} prevents Watson-Crick hydrogen bonding, causing NOE connectivity to be broken at G^{19}. Like the *R*-diastereomer, the imino cross-peaks for T^{17} and T^{18} were not observable in this spectrum. Therefore, sequential connectivity was resumed at G^{9} and completed at G^{14}.

In the H2 region of the spectrum (Figure 4-14A), a cross peak between T^{17} NH3 and A^{8} H2 was not observable. The absence of this cross-peak suggested that the protons had a greater access to solvent as a result of a weakened hydrogen bond and perturbed base pairing. This access to solvent leads to a more rapid exchange with solvent, such that no cross-peak is observable.

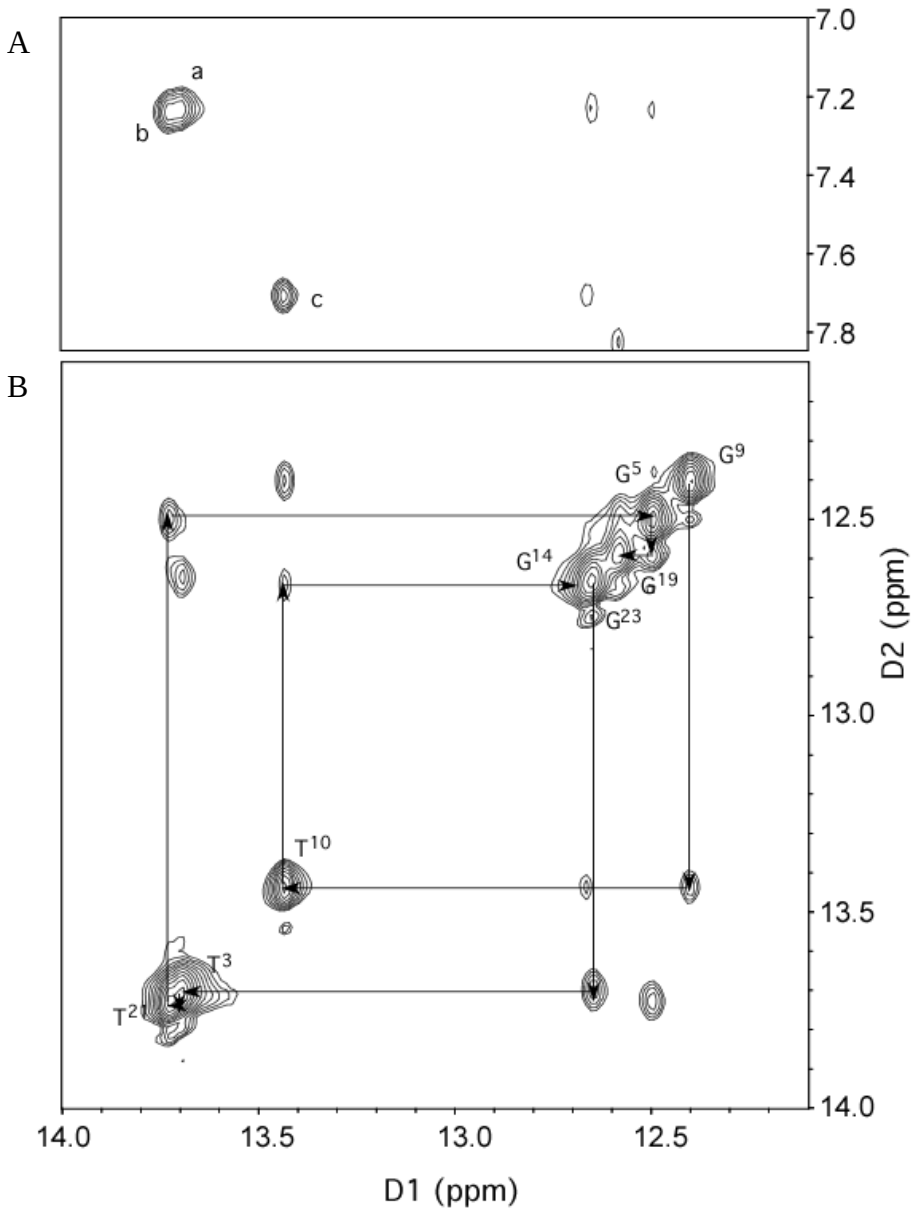


Figure 4-14: NOESY expansion plot of the imino proton region for the S-BD-N3-dU modified sequence 5'-$G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}\underline{X}^{7}A^{8}G^{9}T^{10}C^{11}C^{12}$-3'·5'-$G^{13}G^{14}A^{15}C^{16}T^{17}T^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$-3'. (A) NOESY cross-peaks: a, A^{22}H2-T^{3}N3H; b, A^{4}H2-T^{21}N3H; c, A^{15}H2-T^{10}N3H. (B) Sequential connectivity of imino protons starting at G^{23} and ending at G^{14}. Sequential connectivity is broken at G^{19} and continued at G^{9}. NMR data were acquired at 7 °C at a mixing time of 250 ms and a frequency of 800.23 MHz.

Structural Refinement

Restrained molecular dynamics calculations were carried out on a total of twenty initial coordinate files containing the *S*-BD-N3-dU adduct. These structures were built as described in Chapter 2 and energy minimized prior to rMD calculations. These calculations were carried out in implicit solvent using the simulated annealing protocol.

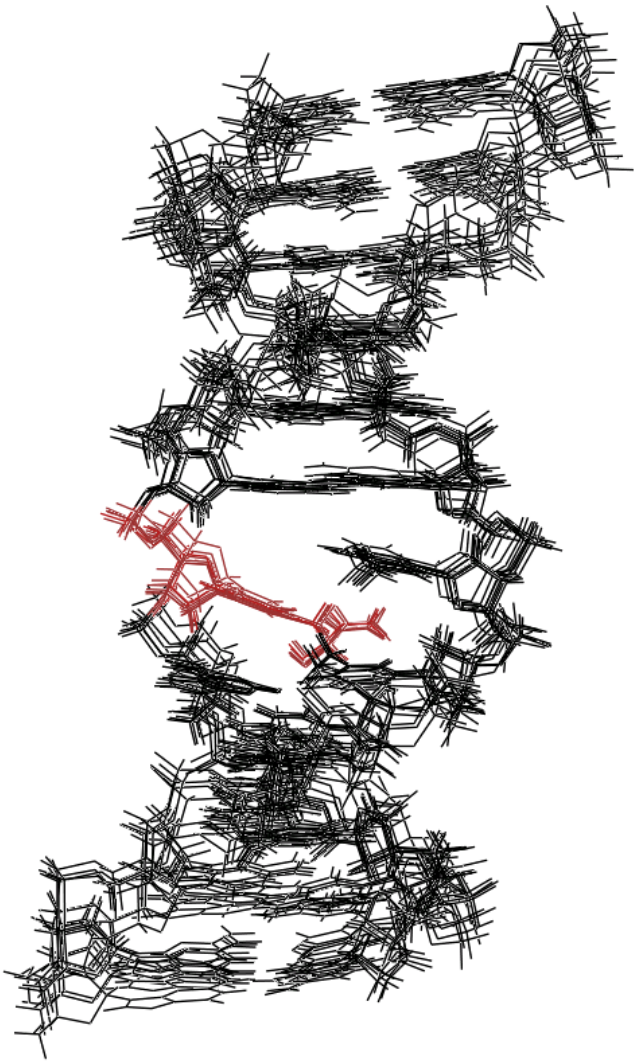


Figure 4-15: Overlay of the eight convergent structures of the S-BD-N3-dU modified oligonucleotide. The modified oligonucleotide is illustrated in red. The rmsd of the convergent structures is 1.02 Å.

A total of 408 distance restraints, comprised of 213 intra-residue, 178 inter-residue, and 17 butadiene restraints were used to restrain the molecular dynamics calculations. The distance restraints were evenly distributed amongst the residues in the oligonucleotide and are listed in Appendix B. To restrain the backbone and sugar pucker, a total of 90 and 110 torsion angle empirical restraints were used, respectively. In addition to the distance and angle restraints, 44 Watson-Crick hydrogen-bonding restraints were used, omitting hydrogen-bonding restraints for X^{7}:T^{18} and $A^{8}T^{17}$. These restraints

are listed in Appendix C.

The rmsd of the twenty starting structures proceeding rMD calculations was 2.73 Å. Following 20 ps of molecular dynamics calculations in implicit solvent, the structural coordinates which emerged from the calculations were compared using pairwise rmsd. Eight of the twenty starting structures converged to give an rmsd value of 1.02 Å. An overlay of these structures is shown in Figure 4-15. The coordinate files of these convergent structures were averaged to produce a mean structure.

CORMA [155] calculations were run on the average structure emerging from the simulated annealing calculations. The total R_1x value was $1.04x10^{-1}$, with individual nucleotide sixth root residuals ranging from $5.5x10^{-2}$ to $1.7x10^{-1}$. These data suggested that the average calculated structure agrees well with the NMR data. The results from the CORMA calculations with respect to the individual nucleotides are shown in Figure 4-16.

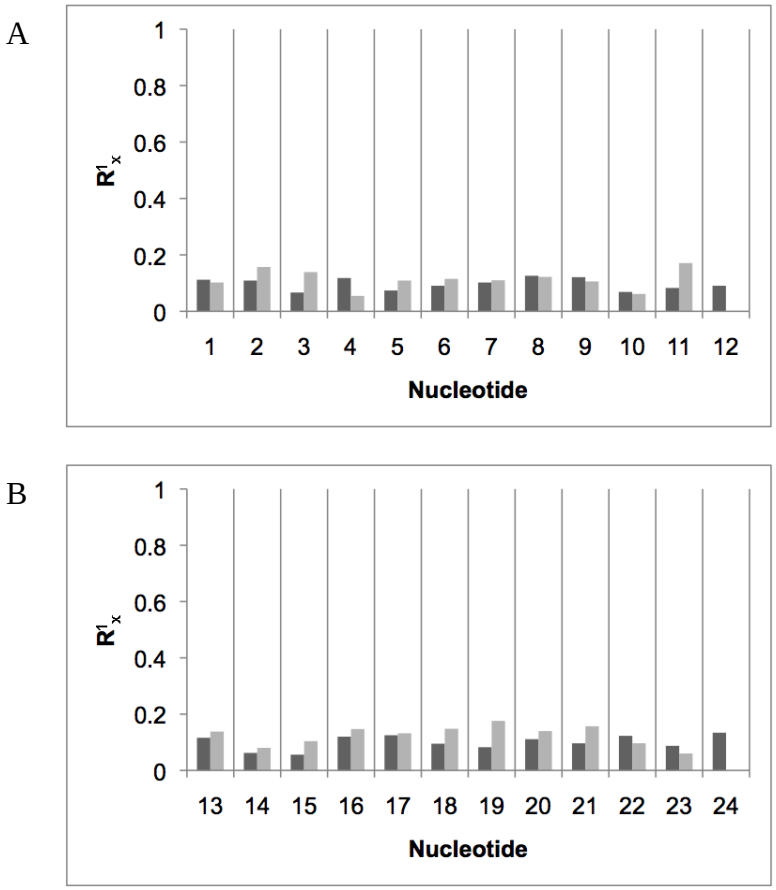


Figure 4-16: CORMA calculations on the average *S*-BD-N3-dU modified oligodeoxynucleotide calculated using a simulated annealing rMD protocol. (A) Nucleotides G^{1}-C^{12} of the modified oligonucleotide containing *S*-BD-N3-dU. (B) Nucleotides G^{13}-C^{24} of the complementary strand. The black bars represent intranuclear sixth root residuals and the gray bars represent internuclear sixth root residuals.

Structure of the S-BD-N3-dU Adduct Opposite dT

The average solution structure containing the *S*-BD-N3-dU adduct opposite dT has a localized perturbation at the lesion site. The lesion site of the average structure is shown in Figure 4-17. The butadiene moiety possesses no directionality, rather is fixed in the minor groove and oriented toward the complementary strand. The adduct is

positioned between T^{17} and T^{18} in the complementary strand. The placement of the butadiene moiety causes a disruption in hydrogen bonding between X^7 and T^{18}, as well as A^8 and T^{17}. A disruption in base stacking is also observed at the lesion site.

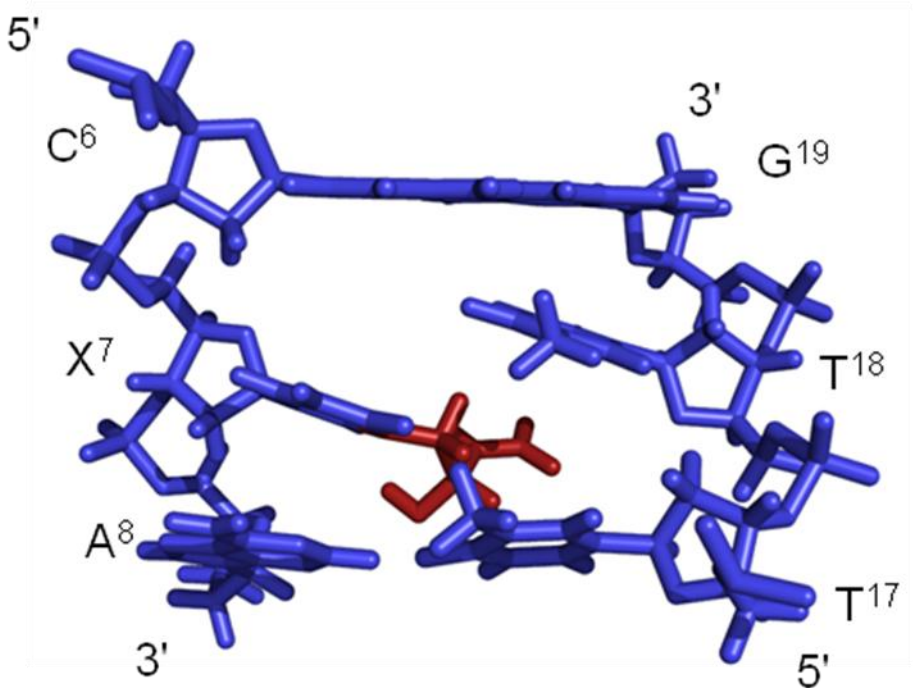


Figure 4-17: Lesion site of the average S-BD-N3-dU modified oligonucleotide structure calculated using a simulated annealing rMD protocol. The butadiene moiety (in gray) is oriented in the minor groove and located between T^{17} and T^{18}.

Hydrogen Bonding Analysis

A hydrogen bonding analysis was carried out on the average structure containing the *S*-BD-N3-dU adduct opposite dT in order to test for hydrogen bonding interactions between the hydroxyl proton of the butadiene moiety and the O^2 atom of the modified nucleotide. To test this, the average structure extracted from simulated annealing calculations in implicit solvent was placed in a truncated octahedral TIP3P solvent box after the addition of sodium counter ions. Prior to 5 ns of MD calculations, the solvated

molecule was energy minimized and equilibrated according to standard methods. Hydrogen bonding occupancies were extracted from molecular dynamics trajectories using PTRAJ (AMBER 8.0) [153]. The percent occupancy for the hydrogen bond between the hydroxyl proton and the O^2 atom on the modified uracil was 0.1%. Therefore, according to these data, no hydrogen bonding interactions at this position are predicted.

Helicoidal Analysis

An analysis of the helicoidal parameters of the average modified structure was performed following MD calculations. As a result of modification at the lesion site, the *S*-BD-N3-dU adduct caused a 25º bend in the duplex at the lesion site. Helicoidal base pair parameters were normal (Figure 4-18). The parameters for base-base interactions showed that at the lesion site there was a disruption in the shear and the stagger, as well as a 40° opening (Figure 4-19). Finally, helicoidal parameters for global inter-base interactions showed that there was a slight rise between X^7 and its two neighboring base pairs (Figure 4-20).

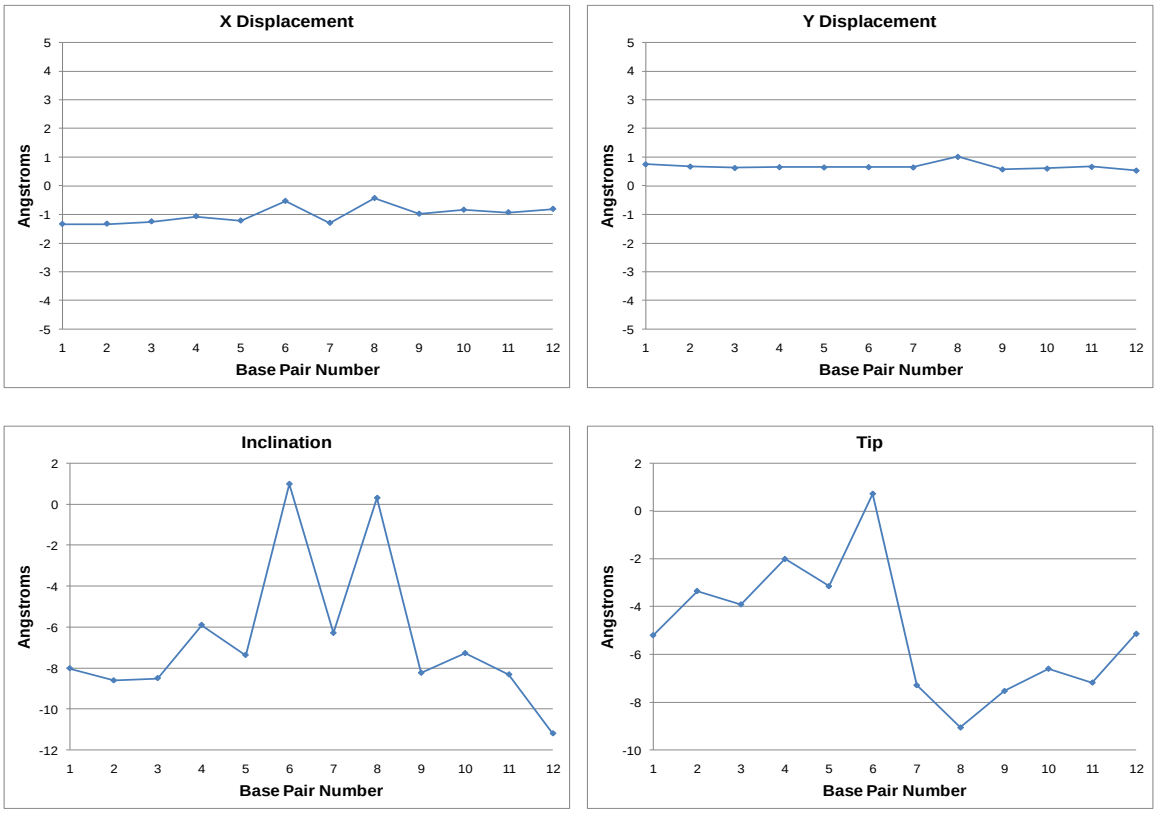


Figure 4-18: Global base pair helicoidal parameters. Helicoidal parameters for the *S*-BD-N3-dU modified oligonucleotide 5′-$G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}\underline{X}^{7}A^{8}G^{9}T^{10}C^{11}C^{12}$-3′·5′-$G^{13}G^{14}A^{15}C^{16}T^{17}T^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$-3′.

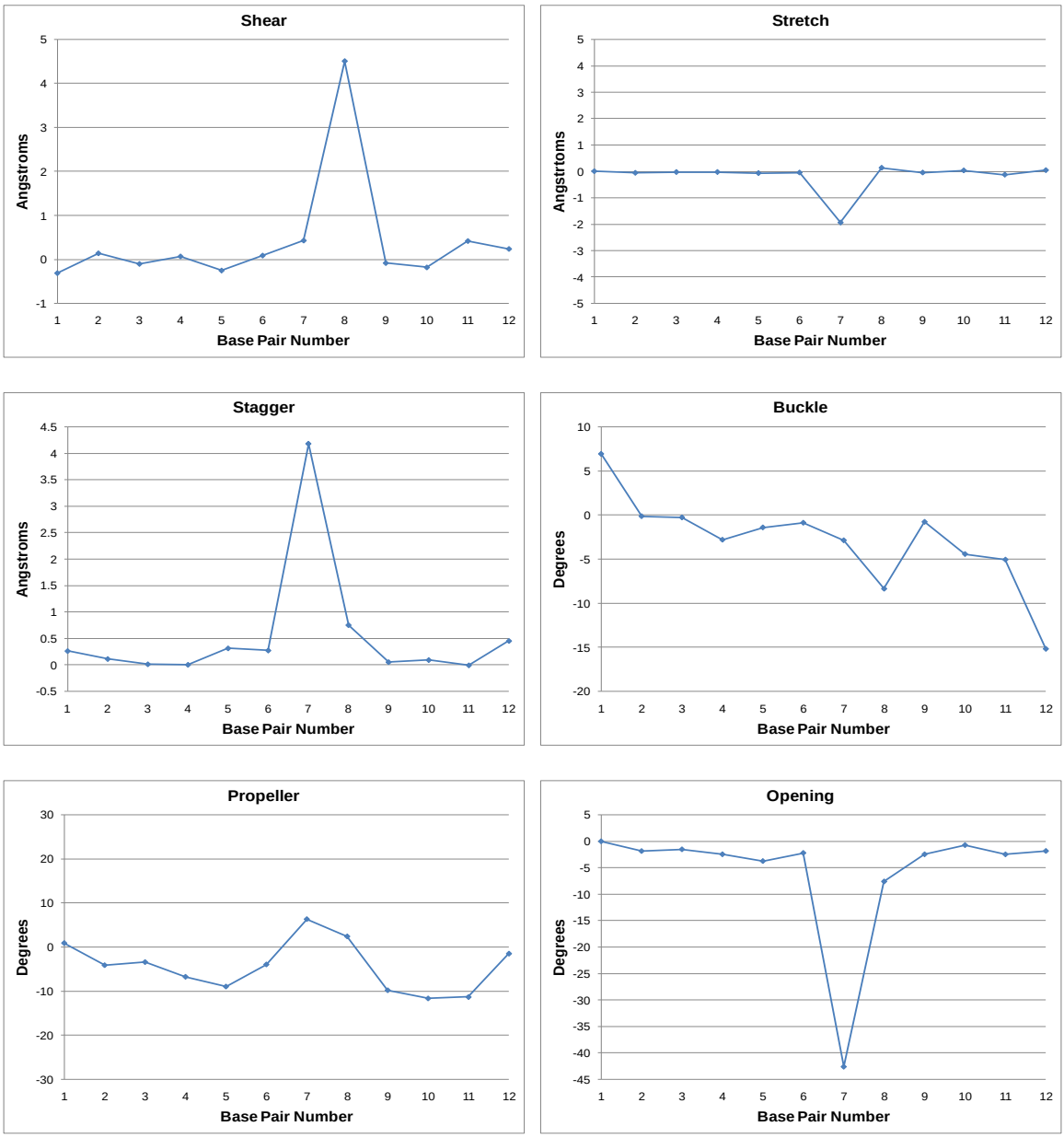


Figure 4-19: Global base-base helicoidal parameters. Helicoidal parameters for the *S*-BD-N3-dU modified oligonucleotide 5′-G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}<u>X</u>^{7}A^{8}G^{9}T^{10}C^{11}C^{12}-3′·5′-G^{13}G^{14}A^{15}C^{16}T^{17}T^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}-3′.

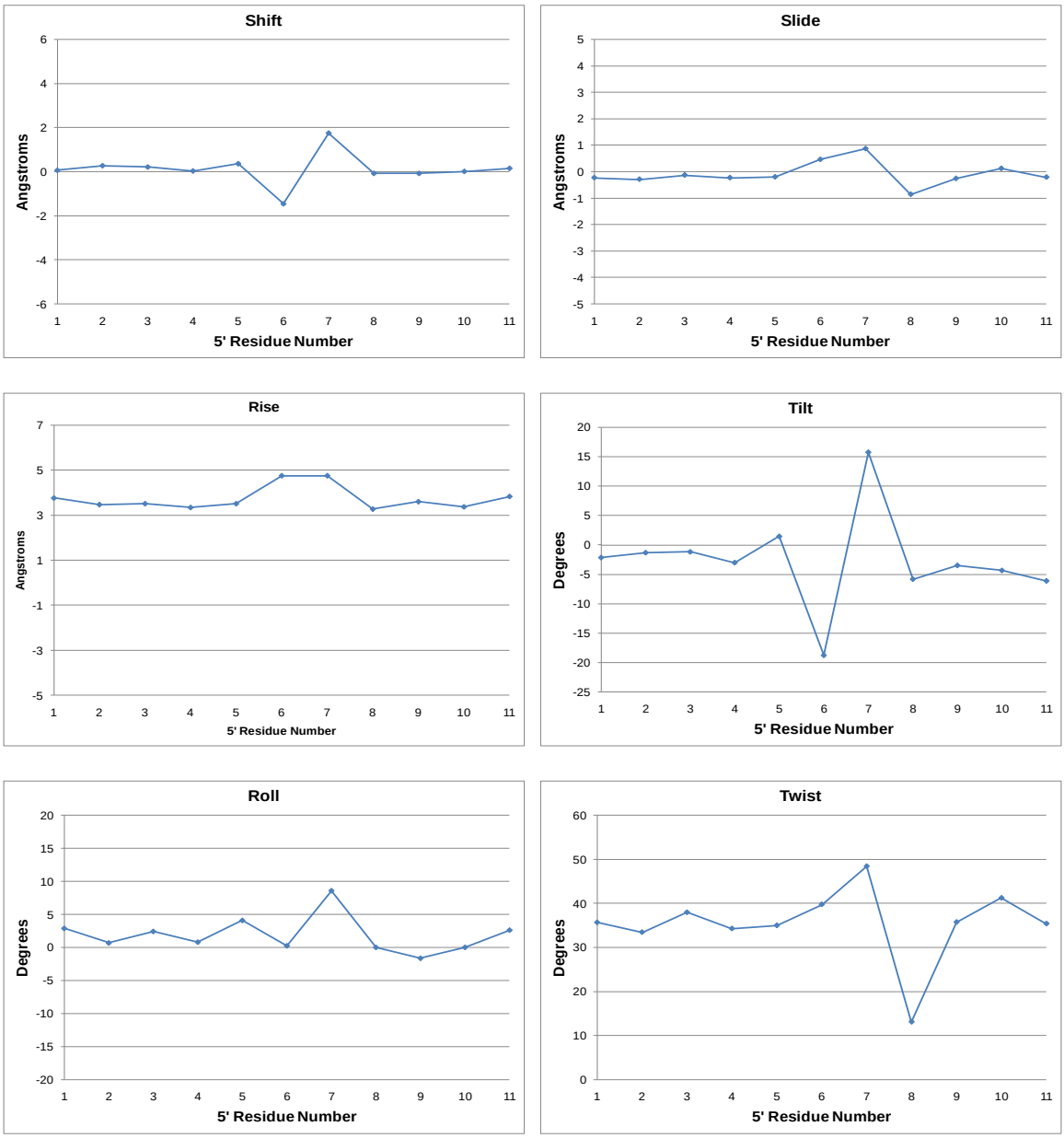


Figure 4-20: Global inter-base helicoidal parameters. Helicoidal parameters for the *S*-BD-N3-dU modified oligonucleotide 5′-$G^1C^2T^3A^4G^5C^6\underline{X}^7A^8G^9T^{10}C^{11}C^{12}$-3′·5′-$G^{13}G^{14}A^{15}C^{16}T^{17}T^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$-3′.

Discussion

The second model for investigation was the dodecamer oligonucleotide, 5′-$G^1C^2T^3A^4G^5C^6\underline{X}^7A^8G^9T^{10}C^{11}C^{12}$-3′·5′-$G^{13}G^{14}A^{15}C^{16}T^{17}T^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$-3′, where

both the *R*- and *S*-stereoisomers were site-specifically placed at the seventh position opposite dT. This model represents the insertion of dTTP opposite the lesion site, which if left unrepaired, results in a C to A transversion. This mutation was one of the major mutations caused by both the *R*- and *S*-BD-N3-dU adducts in COS-7 mammalian cells. It was reported that both stereoisomers induced the same percentage of C to A transversions, therefore, it was hypothesized that the orientation of the stereoisomeric lesions would be similar.

Structural Analysis of the R-BD-N3-dU Adduct Opposite dT

CORMA back calculations carried out on the average structure indicated that the theoretical NOE intensities of the structure agreed well with the actual NOE intensities at 250 ms. NMR data, showing interactions between butadiene protons and other oligonucleotide interactions between Hα, Hα′, and Hβ with the H2 of A^8 indicated that the butadiene moiety would reside in the minor groove. An interaction between the Hβ proton and the H1' proton of T^{17} oriented Hβ toward the complementary strand. Weak cross-peaks were also observed between the Hδ and Hδ' protons and A^8 H2, placing them in the minor groove, directly across from T^{17} in the complementary strand

The butadiene lesion created a 25° opening in the helix as well as a 1.5 Å rise between the modified nucleotide and its 3'- and 5'-neighbors, confirmed by weak NOESY interactions between several aromatic and anomeric protons at the lesion site and its neighboring base pairs. The intrusion of the butadiene moiety into the complementary strand shifted T^{17} slightly into the major groove, resulting in a 1 Å twist at the lesion site. Finally, disruptions in base pairing and disordered hydrogen bonding at both the lesion

site and the 3'-neighboring base pair, A^8:T^{17} were due to the orientation of the *R*-BD-N3-dU adduct. This disruption in hydrogen bonding was supported by the absence of a cross-peak for the interaction between A^8 H2 and T^{17} N3H.

Comparison of the Stereoisomeric BD-N3-dU Adducts Mismatched Opposite dT

CORMA back calculations carried out on the average structure containing the *S*-BD-N3-dU opposite thymine indicated that theoretical NOE intensities of the structure agreed well with the actual NOE intensities at 250 ms. Well-resolved NMR data showed several cross-peaks resulting from interactions between butadiene protons and other oligonucleotide protons, confirming the orientation of the butadiene moiety. Like the *R*-diastereomer, the interactions of Hα, Hα′, and Hβ with the H2 of A^8 placed the moiety in the minor groove. An interaction between the Hβ proton and the H1′ proton of T^{17} oriented this proton toward the complementary strand. The NOEs observed in the NMR data for the *S*-diastereomer were similar to those for the *R*-diastereomer and therefore suggested that the orientations of the lesion in both structures would be similar.

Like the oligonucleotide containing the *R*-BD-N3-dU adduct, for the *S*-diastereomer, the modification at the seventh position created an opening in the helix, as well as a 1 Å rise was also observed between the lesion its neighboring base pairs. Both lesions caused a disruption in hydrogen bonding at both the lesion site and the 3′-neighboring base pair, A^8:T^{17} along with base-stacking interference between X^7 and C^6.

Summary

Both the *R*- and *S*-BD-N3-dU adducts mismatched opposite deoxythymine caused disruptions in the complementary strand opposite the lesion site. Unlike the structures with the stereoisomeric lesions opposite deoxyadenine, the orientations of both butadiene moieties are similar. For both diastereomers, the butadiene moiety was located in the minor groove and oriented toward the complementary strand between T^{17} and T^{18}. Furthermore, the orientation of the *R*-BD-N3-dU is different depending on its complementary base. When the *R*-BD-N3-dU lesion is placed opposite dA, the butadiene moiety is oriented in the 5′-direction. Whereas, the same lesion opposite dT possesses no directionality. These findings correlate with the mutagenesis studies and may suggest that the orientation of the butadiene moiety in BD-N3-dU adducts plays a role in biological processing.

CHAPTER V

SOLUTION STRUCTURE OF THE STEREOISOMERIC N3-(2-HYDROXY-3-BUTEN-1-YL)-2'-DEOXYURIDINE ADDUCTS ARISING FROM BUTADIENE MONOEPOXIDE OPPOSITE DEOXYGUANINE IN A DODECAMER OLIGONUCLEOTIDE

Introduction

Site-specific mutagenesis studies in COS-7 mammalian cells suggested that the stereoisomeric N3-dU adducts arising from BDO are highly mutagenic and will rarely result in the insertion of dGTP during replication bypass [159]. It was hypothesized that the orientation of the BD moiety opposite dG in duplex DNA might elucidate why dGTP is seldom inserted. Therefore, the final model of structural investigation contained the stereoisomeric BD-N3-dU adducts opposite dG in the dodecamer oligonucleotide, 5'- d($G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}\underline{X}^{7}A^{8}G^{9}T^{10}C^{11}C^{12}$)-3'·5'-d($G^{13}G^{14}A^{15}C^{16}T^{17}\underline{G}^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$)-3'. When the lesions are placed opposite dG, the oligonucleotides represent the incorporation of dGTP, which would result in normal DNA replication.

Two-dimensional NMR experiments were used to elucidate changes in the chemical environment of the *R* and *S* oligonucleotides as a result of modification at the lesion site, as well as provide experimental restraints for rMD calculations. NMR analysis indicated that the major perturbation to the oligonucleotide structure occurs at the lesion site. For both adducts there is an increase in the rise between the adduct and its 3′- and 5′-neighboring bases. In both structures, Watson-Crick hydrogen bonding is

perturbed at the lesion site and at the 3′-neighboring base pair, A^{8}:T^{17}. For the structure containing the *R*-BD-N3-dU adduct, the butadiene moiety is oriented in the 5′-direction. For the structure containing the *S*-BD-N3-dU adduct, the butadiene moiety is oriented in the minor groove toward the complementary strand, forcing G^{18} into the major groove.

Results

R Diastereomer

Sample Purity

The double-stranded *R*-BD-N3-dU modified oligonucleotide, 5'-d($G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}\underline{X}^{7}A^{8}G^{9}T^{10}C^{11}C^{12}$)-3'·5'-d($G^{13}G^{14}A^{15}C^{16}T^{17}G^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$)-3', was purified using the HPLC protocol described in Chapter II. Enzyme hydrolysis showed that the oligonucleotide was enantiomerically pure. The composition of the two single strands was verified using MALDI-TOF mass spectrometry, as well as COSY and NOESY NMR experiments. The annealed duplex was separated from any remaining single-stranded DNA using a hydroxylapatite column. The presence of only duplex DNA in the solution sample was confirmed using capillary gel electrophoresis. NMR experiments at a frequency of 800.23 MHz provided valuable data for the assignment and volume integration of cross-peaks.

Non-exchangeable protons

NMR experiments were carried out in D_2O for the observation of non-exchangeable protons. The modified sample was prepared for these experiments as

discussed in Chapter II. Assignments of the purine and pyrimidine aromatic protons, the thymine methyl protons, and the deoxyribose H1', H2', H2", and H3' protons were successfully completed. Due to overlapping peaks, the H4', H5', and H5" protons were only partially assigned. The chemical shift values are listed in Appendix A.

Expansion plots of the NOESY spectrum showing the interactions between the aromatic and anomeric protons for the *R* diastereomer are shown in Figure 5-1. In the modified strand (Figure 5-1A), the "walk" was uninterrupted throughout the entirety of the sequence. The interaction between C^6 H1' and X^7 H6 produced a cross-peak that was overlapped with the T^{17} H1′-H6 cross-peak. However, the intensity of this cross-peak suggested that the interaction between C^6 H1' and X^7 H6 was weaker than other anomeric:aromatic internuclear interactions. We also observed a decreased intensity in the cross-peak for the interaction between X^7 H1' and A^8 H8. In the complementary strand, sequential connectivity was observed from the 5′- to 3′-terminus. It was observed in the complementary strand that the interaction between T^{17} H1' and G^{18} H8 produced a weak cross-peak (Figure 5-1B). A decrease in the intensity of the G^{18} H1':G^{19} H8 cross-peak is also observed. These results suggested a localized perturbation for the *R* stereoisomer, with an increase in the rise between the lesion site and its 5'- and 3'-neighboring base pairs.

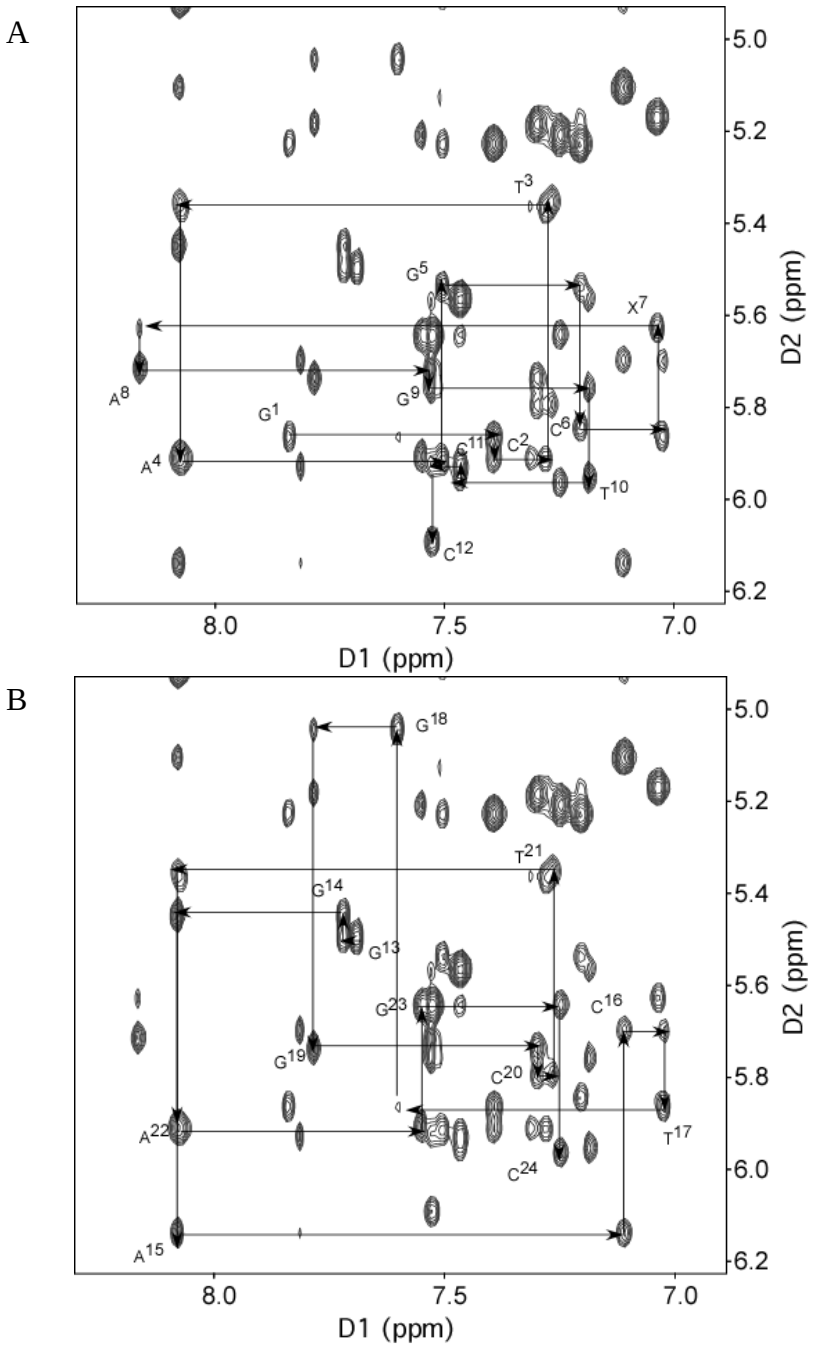


Figure 5-1: NOESY expansion plot of the anomeric proton to aromatic proton region for the *R*-BD-N3-dU modified sequence 5'-$G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}\underline{X}^{7}A^{8}G^{9}T^{10}C^{11}C^{12}$-3'·5'-$G^{13}G^{14}A^{15}C^{16}T^{17}G^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$-3'. (A) Nucleotides G^{1}-C^{12} of the modified oligonucleotide containing *R*-BD-N3-dU. (B) Nucleotides G^{13}-C^{24} of the complementary strand. NMR data were acquired at a mixing time of 250 ms and a frequency of 800.23 MHz.

A chemical shift analysis compared the chemical shifts of the aromatic and anomeric protons in the unmodified oligodeoxynucleotide sequence to those in the *R*-BD-N3-dU modified duplex. In the modified strand, the largest perturbation occurred at the modified nucleotide, X^7, where the H1' proton was shifted upfield by more than 0.4 ppm and the H6 proton was shifted downfield by 0.2 ppm. A significant change of 0.2 ppm was also observed for A^8 H1'. In the complementary strand, significant chemical shift changes of 0.4 and 0.5 ppm were observed for the H1' protons of T^{17} and G^{18}, respectively. A chemical shift change of 0.2 ppm was also observed for the H8 proton of G^{19}. The observation of chemical shift changes primarily at the lesion site and its 3'- and 5'-neighboring base pairs suggested a localized perturbation to the structure of the oligonucleotide.

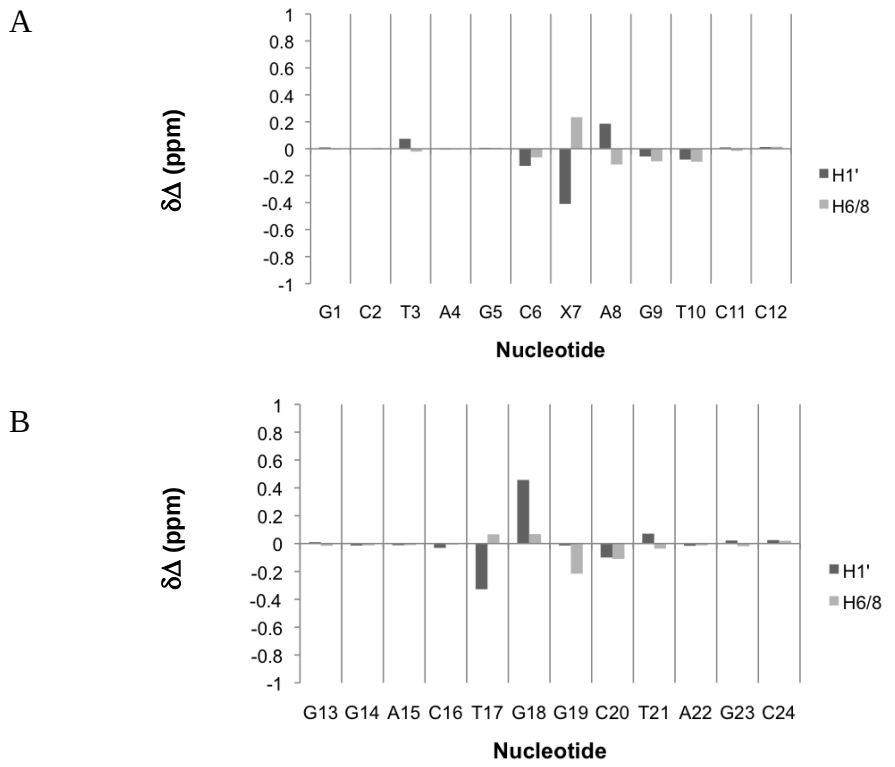


Figure 5-2: Chemical shift perturbations in the aromatic and anomeric protons of the *R*-BD-N3-dU modified sequence 5'-$G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}\underline{X}^{7}A^{8}G^{9}T^{10}C^{11}C^{12}$-3'·5'-$G^{13}G^{14}A^{15}C^{16}T^{17}G^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$-3' relative to the unmodified oligodeoxynucleotide. (A) Nucleotides G^{1}-C^{12} of the modified oligonucleotide containing *R*-BD-N3-dU. (B) Nucleotides G^{13}-C^{24} of the complementary strand.

Each of the six BD protons were observed in a TOCSY spectrum. Hβ and Hγ were assigned using TOCSY and NOESY experiments. Hδ and Hδ′ were assigned using a DQF-COSY experiment. It was not possible to discriminate between the cross-peaks produced by the protons on the alpha carbon, Hα and Hα′. Therefore, these protons were assigned ambiguously. The chemical shift values for the butadiene protons are listed in Appendix A.

NOESY experiments were used in order to observe interactions between the BD protons and other oligonucleotide protons. These interactions were used in order to determine the orientation of the BD moiety with respect to other protons in the

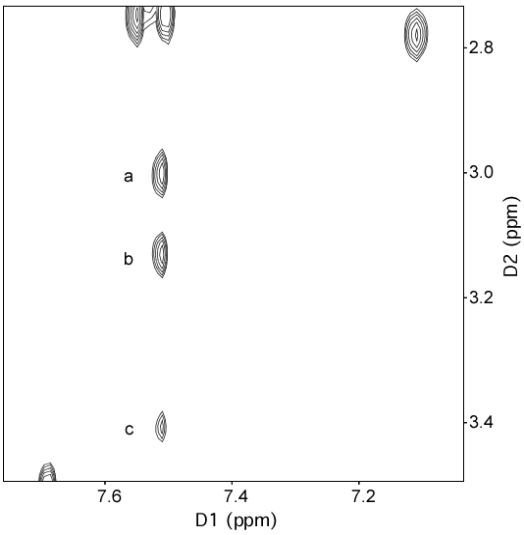


Figure 5-3: Dipolar proton:proton interactions between the R-BD moiety and other oligonucleotide protons. NOESY cross-peaks: a, Hα-A^8 H2; b, Hα′-A^8 H2; c, Hβ-A^8 H2. NMR data were acquired at a mixing time of 250 ms and a frequency of 800.23 MHz.

oligonucleotide. An expansion plot of the NOESY spectrum (Figure 5-3) shows strong interactions between Hα, Hα', and Hβ of butadiene and A^8 H2. Stronger interactions between the A^8 H2 proton and the butadiene protons, Hα and Hα', suggested that for the *R* diastereomer, these butadiene protons would be positioned closer to the 3'-neighboring base, A^8, whereas the Hβ proton would be oriented more so in the 5'-direction.

Exchangeable Protons

A NOESY spectrum acquired in H_2O was carried out in order to observe exchangeable imino protons. NOESY data were acquired at 7 °C at a frequency of 800 MHz in order to yield well-resolved imino cross-peaks. Nonetheless, the NOESY data yielded imino cross-peaks which were poorly resolved and overlapped.

The modification at the N3 position of X^7 prevents Watson-Crick hydrogen bonding between X^7 and G^{18}. Because hydrogen bonding had been disrupted at the A^8:T^{17} base pair in previous structures, we were interested in evaluating hydrogen-bonding interactions between A^8 and T^{17}. The H2 region (7-8 ppm) is shown in Figure 5-4 and provides insight to these interactions. The cross-peaks at 13.7 are for those interactions between the imino protons of T^3 and T^{21} and the H2 protons of A^{22} and A^4, respectively. These cross-peaks are integrated to one. The cross-peak at 13.5 ppm (a), is assigned to

T^{10} N3H: A^{15} H2 and is also integrated to one. Therefore, we do not observe a cross-peak between T^{17} N3H and A^{8} H2. The absence of this cross-peak indicates that hydrogen bonding was disrupted such that these protons had a greater access to solvent leading to a more rapid exchange with solvent.

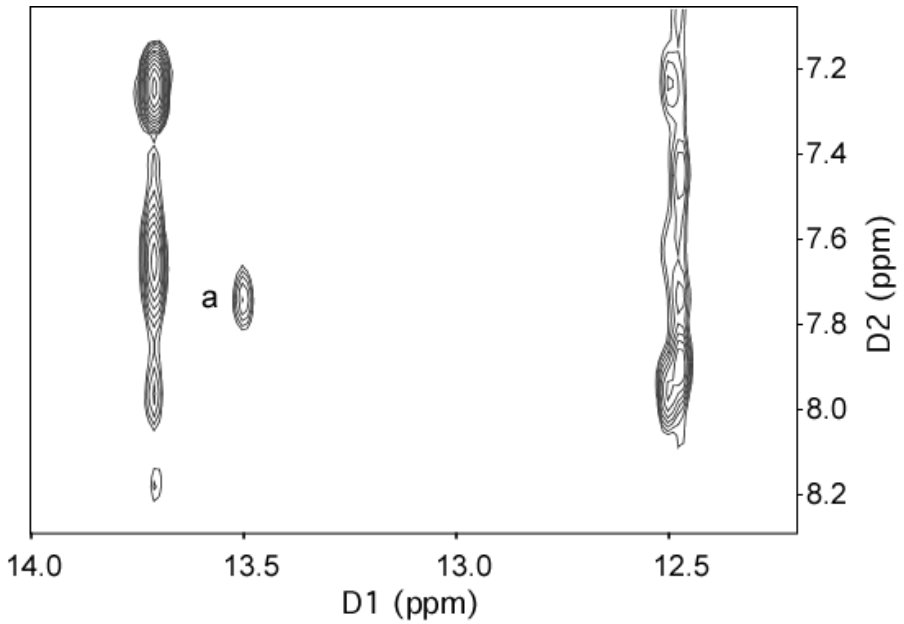


Figure 5-4: NOESY expansion plot of the imino and H2 proton region for the R-BD-N3-dU modified sequence 5'-$G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}\underline{X}^{7}A^{8}G^{9}T^{10}C^{11}C^{12}$-3'·5'-$G^{13}G^{14}A^{15}C^{16}T^{17}G^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$-3'. NOESY cross-peak: a, A^{15} H2 to T^{10} N3H. NMR data were acquired at 7 °C at a mixing time of 250 ms and a frequency of 800.23 MHz.

Structural Refinement

Twenty initial structures were used as coordinate files for structural refinement of the duplex DNA containing the *R*-BD-N3-dU adduct opposite dG. These structures were built as described in Chapter II and energy minimized prior to rMD calculations. Restrained molecular dynamics calculations were carried out on all twenty structures in implicit solvent utilizing the simulated annealing protocol. A total of 549 distance

Figure 5-5: Overlay of the five convergent R-BD-N3-dU modified oligonucleotide structures calculated using a simulated annealing rMD protocol. The modified nucleotide, X^7, is colored in red. The rmsd of the ten internal base pairs is 1.14 Å.

restraints, containing 331 intra-residue, 199 inter-residue and 19 butadiene restraints, were used to restrain the molecular dynamics calculations. The distance restraints, listed in Appendix B, were evenly distributed amongst the residues in the oligonucleotide. To restrain the backbone and sugar pucker, a total of 90 and 110 empirical torsion angle restraints were used, respectively. In addition to the distance and angle restraints, 44 Watson-Crick hydrogen-bonding restraints were used. These restraints are listed in Appendix C.

Prior to rMD calculations, the root mean square deviation (rmsd) of the twenty starting structures was 3.19 Å. Following 20 ps of calculations, the structural coordinates which emerged from the calculations were compared using pairwise rmsd. Five of the twenty starting structures converged to give an rmsd value of 1.14 Å. An overlay of the convergent structures is shown in Figure 5-5.

CORMA [155] calculations were run on the average structure taken from rMD calculations. The total R_1x value was $8.5x10^{-2}$, with the sixth root residual values for

individual nucleotides ranging from $2.8x10^{-2}$ to $1.4x10^{-1}$. These values suggested that the average calculated structure agreed well with the NMR data. The results from the CORMA calculations with respect to the individual nucleotides are shown in Figure 5-6.

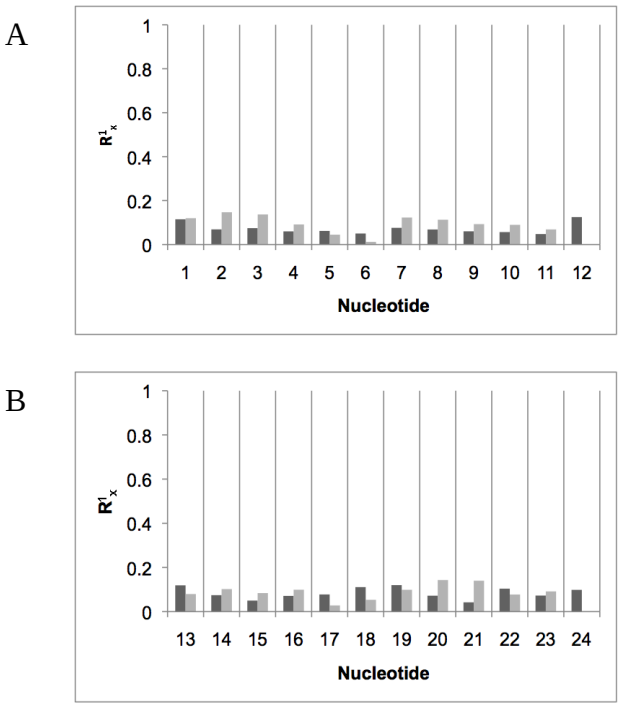


Figure 5-6: CORMA calculations on the average *R*-BD-N3-dU modified oligonucleotide calculated using a simulated annealing rMD protocol. (A) Nucleotides G^{1}-C^{12} of the modified oligonucleotide containing *R*-BD-N3-dU. (B) Nucleotides G^{13}-C^{24} of the complementary strand. The black bars represent intranuclear sixth root residuals and the gray bars represent internuclear sixth root residuals.

Structure of the R-BD-N3-dU Adduct Opposite dG

The average structure of the *R*-BD-N3-dU adduct opposite dG determined by rMD calculations in implicit solvent shows that the major perturbation resulting from modification at the X^{7} position occurs at the lesion site. The lesion site of the average structure is shown in Figure 5-7. The butadiene moiety is oriented toward the 5′-direction. The position of the butadiene atoms forces the guanine opposite the lesion in the 5′-direction toward T^{17}. The position of the adduct generates a disruption in hydrogen

bonding at both the lesion site and the 3′-neighboring base pair, A^8:T^{17}. Base stacking is also disrupted at the lesion site and its neighboring base pairs. In order to accommodate the butadiene moiety, a rise between X^7 and C^6, as well as G^{18} and G^{19} is also observed.

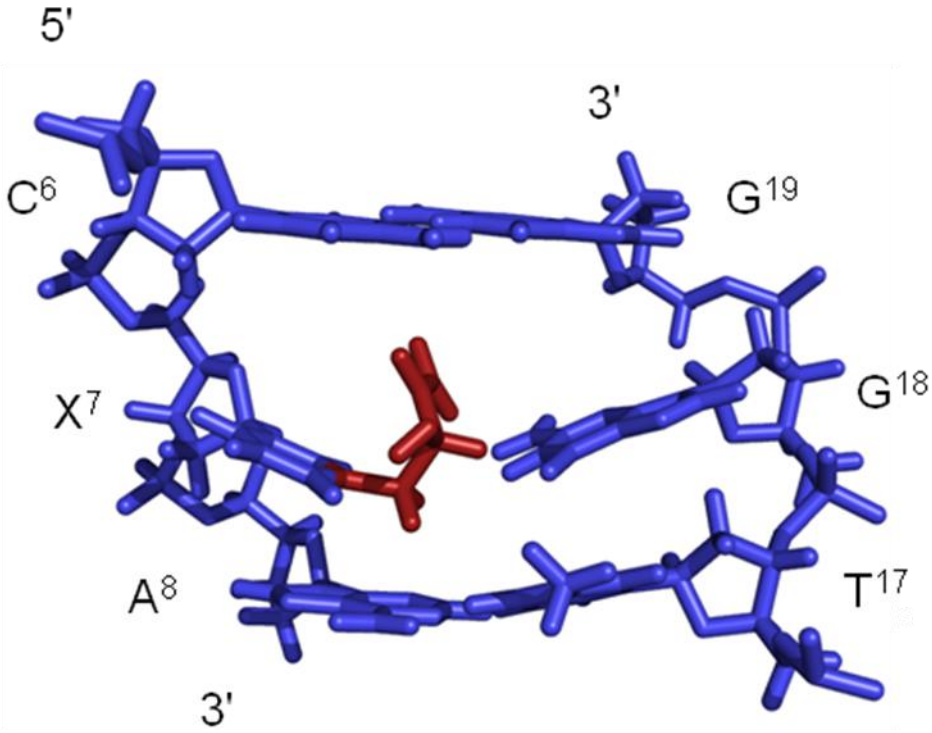


Figure 5-7: Lesion site of the average R-BD-N3-dU modified oligonucleotide structure. The butadiene moiety has a 5′-directionality and is disrupting to base pairing at the lesion site and its 3′-neighboring base pair.

Hydrogen Bonding Analysis

The average structure of the *R*-BD-N3-dU modified oligonucleotide determined from simulated annealing calculations using implicit solvent, indicated that the potential existed for a hydrogen bond between the hydroxyl proton and the O^4 atom of the modified nucleotide. To test this, the average structure was placed in a truncated

octahedral TIP3P solvent box after the addition of sodium counter ions. Prior to 5 ns of MD calculations, the solvated molecule was energy minimized and equilibrated according to standard methods. Hydrogen bonding occupancies were extracted from MD trajectories using PTRAJ (AMBER 8.0) [153]. Interestingly, the hydroxyl proton was shown to hydrogen bond to both O^4 and O^2. The percent occupancy for the hydrogen bond between the hydroxyl proton and the O^4 atom on the modified uracil was 17.2%, while the percent occupancy for the hydrogen bond between the hydroxyl proton and the O^2 atom was 20.3%. The distance between the hydroxyl proton and O^4 atom was 2.72 Å and the O-H-O angle was determined to be 17.5 degrees. The distance between the hydroxyl proton and O^2 atom was 2.76 Å and the O-H-O angle was determined to be 20.3 degrees. These data predict that the hydroxyl proton moves from one hydrogen-bonding interaction to another.

Helicoidal Analysis

An analysis of the helicoidal parameters of the average modified structure was performed following simulated annealing calculations. As a result of modification at the lesion site, the *R*-BD-N3-dU adduct caused a 17º bend in the duplex at the lesion site. Helicoidal base pair parameters were normal (Figure 5-8). The parameters for base-base interactions showed that at the lesion site there was a 2 Å increase in the shear, the stagger, and the stretch, as well as a 40° buckle (Figure 5-9). Finally, helicoidal parameters for global inter-base interactions showed that there was a significant rise between X^7 and its two neighboring base pairs (Figure 5-10).

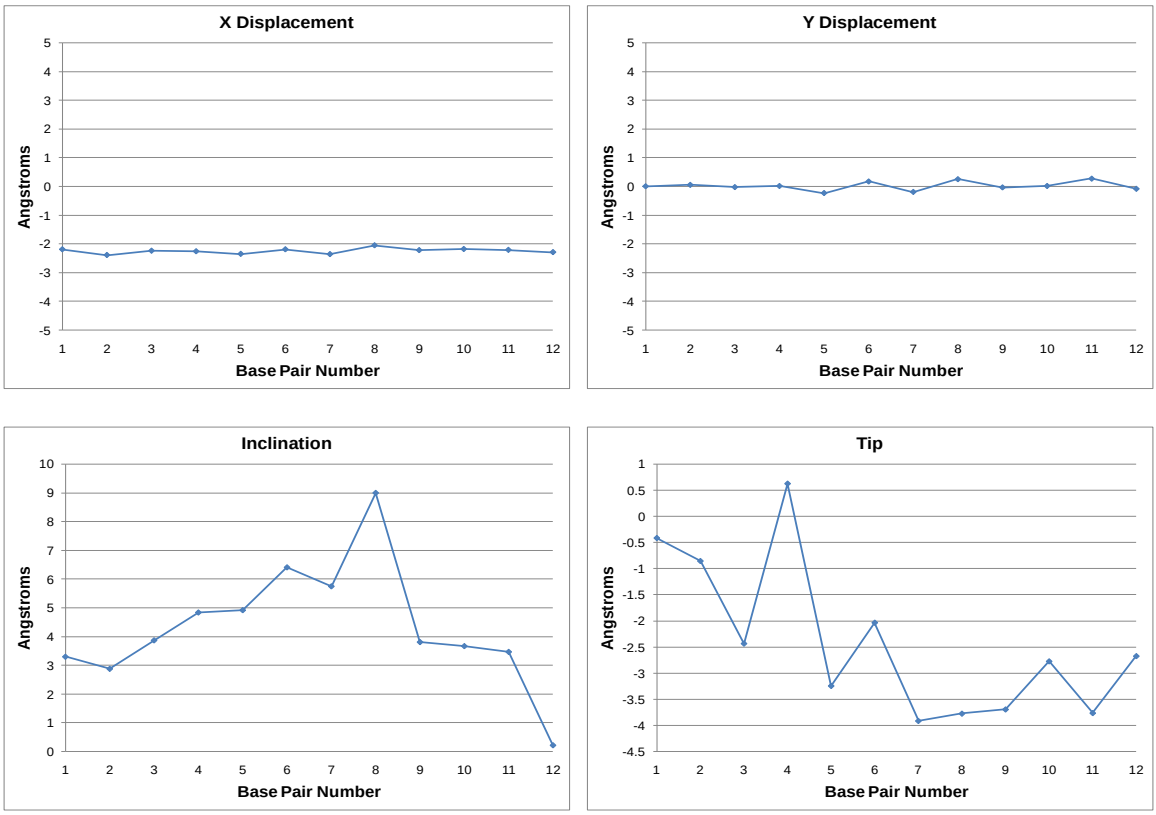


Figure 5-8: Global base pair helicoidal parameters. Helicoidal parameters for the *R*-BD-N3-dU modified oligonucleotide 5′-G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}<u>X</u>^{7}A^{8}G^{9}T^{10}C^{11}C^{12}-3′·5′-G^{13}G^{14}A^{15}C^{16}T^{17}G^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}-3′.

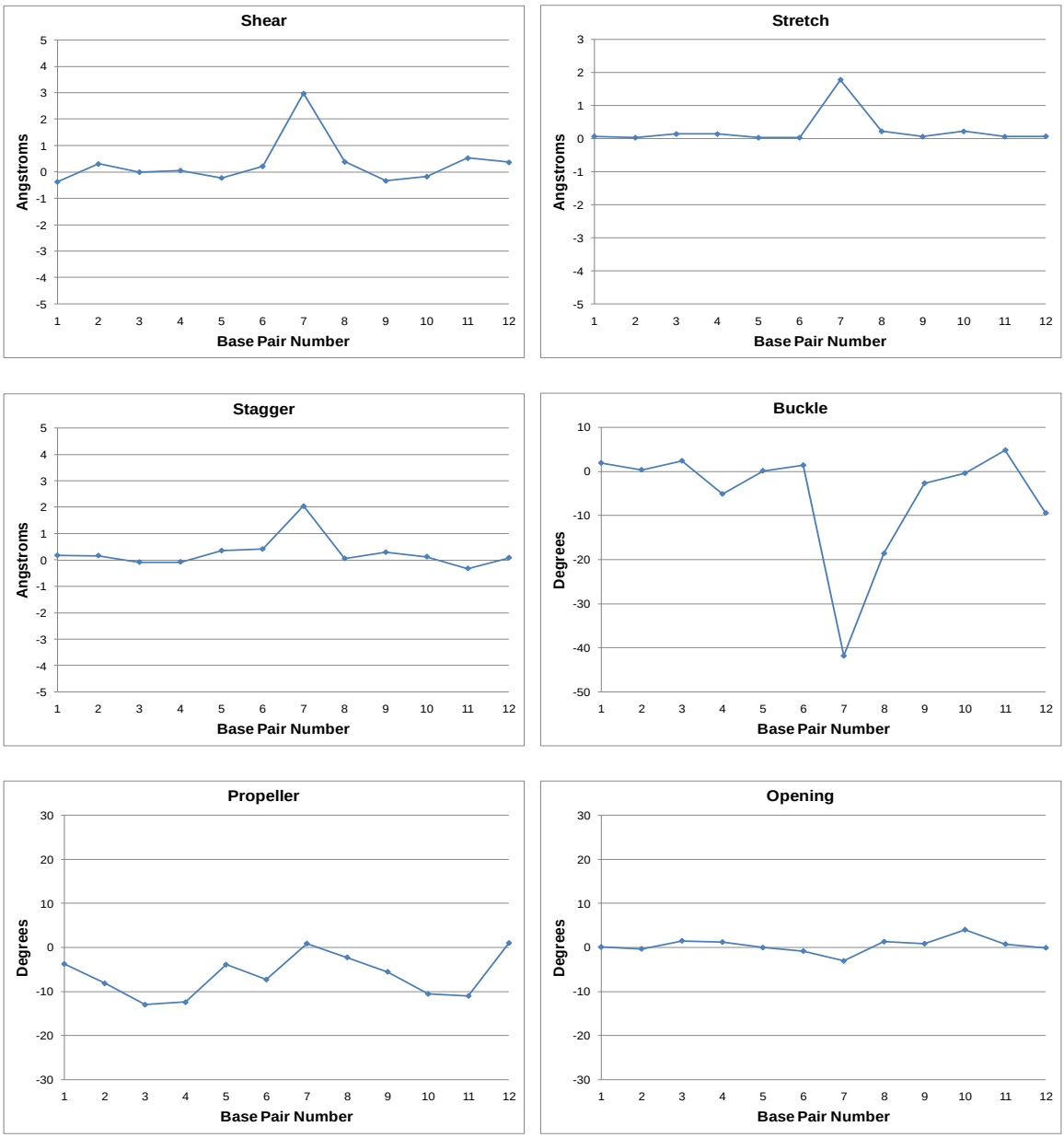


Figure 5-9: Global base-base helicoidal parameters. Helicoidal parameters for the *R*-BD-N3-dU modified oligonucleotide 5′-G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}X^{7}A^{8}G^{9}T^{10}C^{11}C^{12}-3′·5′-G^{13}G^{14}A^{15}C^{16}T^{17}G^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}-3′.

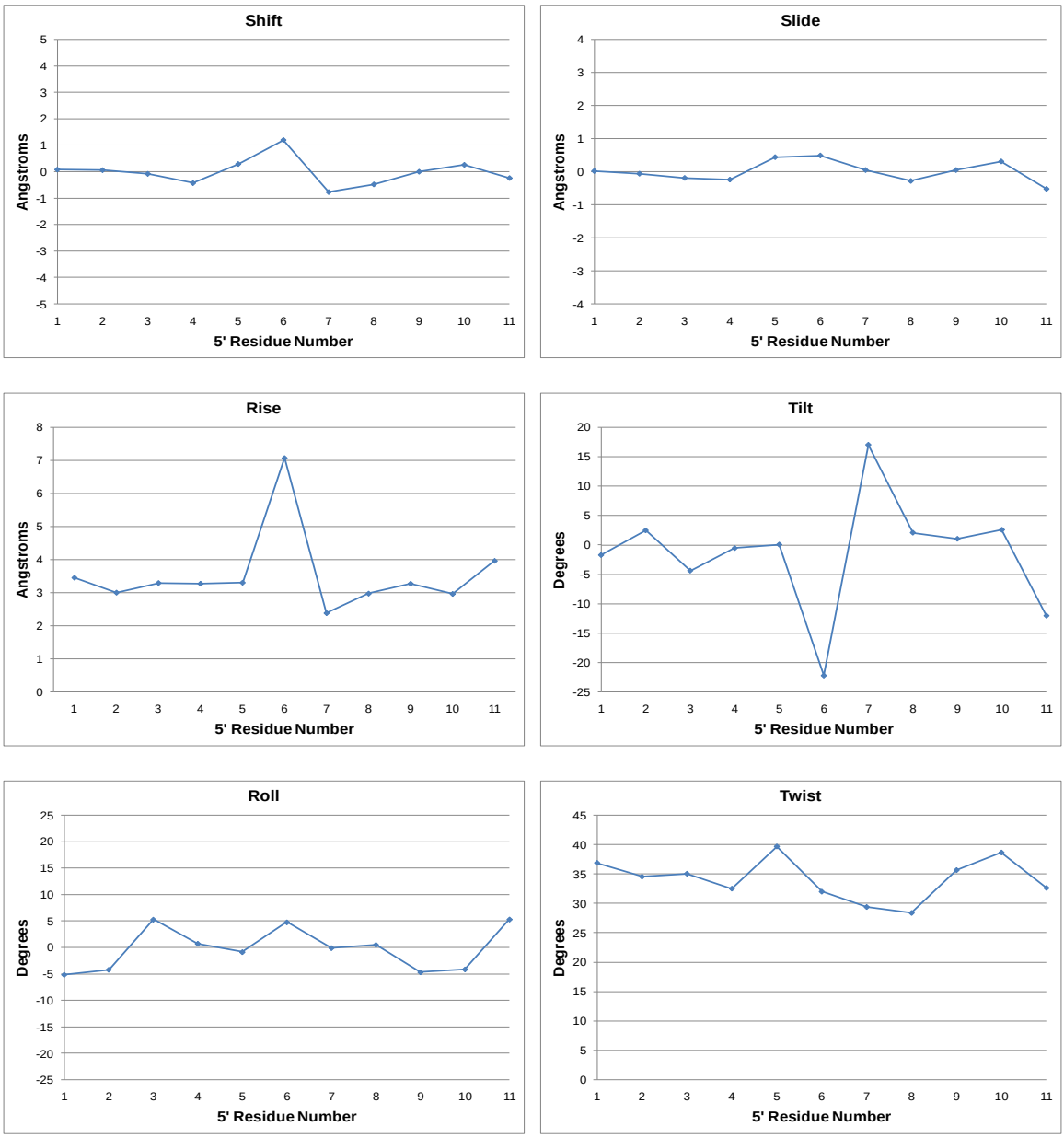


Figure 5-10: Global inter-base helicoidal parameters. Helicoidal parameters for the *R*-BD-N3-dU modified oligonucleotide 5′-G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}<u>X</u>^{7}A^{8}G^{9}T^{10}C^{11}C^{12}-3′·5′-G^{13}G^{14}A^{15}C^{16}T^{17}T^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}-3′.

S-Diastereomer

Sample Purity

The double-stranded *S*-BD-N3-dU modified oligonucleotide, 5'-d($G^1C^2T^3A^4G^5C^6\underline{X}^7A^8G^9T^{10}C^{11}C^{12}$)-3'·5'-d($G^{13}G^{14}A^{15}C^{16}T^{17}G^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$)-3', was purified using the HPLC protocol described in Chapter II. Enzyme hydrolysis showed that the oligonucleotide was enantiomerically pure. The composition of the two single strands was verified using MALDI-TOF mass spectrometry as well as COSY and NOESY NMR experiments. The annealed duplex was separated from any remaining single-stranded DNA using a hydroxylapatite column. The presence of only duplex DNA in the solution sample was confirmed using capillary gel electrophoresis. NMR experiments at a frequency of 800.23 MHz provided valuable data.

Non-exchangeable Protons

For the observation of non-exchangeable protons in NMR spectra, the sample was prepared as discussed in Chapter II. Complete assignments of the chemical shift values for the purine and pyrimidine aromatic protons, the thymine methyl protons, and the deoxyribose H1', H2', H2", and H3' protons was achievable. Due to overlapping peaks and spin diffusion at higher mixing times, only partial assignments were made for the deoxyribose H4', H5', and H5" protons. Expansion plots of the NOESY spectrum for the *S* diastereomer are shown in Figure 5-11.

Sequential connectivity was observed in the modified strand. A significant observation in the modified strand (Figure 5-11A) was a decrease in the intensity of the cross-peak for C^6 H1' and X^7 H6. A weak cross-peak for the interaction between X^7 H1'

and A^{8} H8 was also observed. These weakened cross-peaks suggested a localized perturbation around the lesion site and its 5'- and 3'-neighbor base pairs, as well as an increase in the rise between these nucleotides resulting from the butadiene moiety. In the complementary strand, NOE connectivity remained unbroken. A weak cross-peak was also observed between T^{17} H1' and G^{18} H8, as well as G^{18} H1' and G^{19} H8. These data suggested an increase in the distance between these protons (Figure 5-11B).

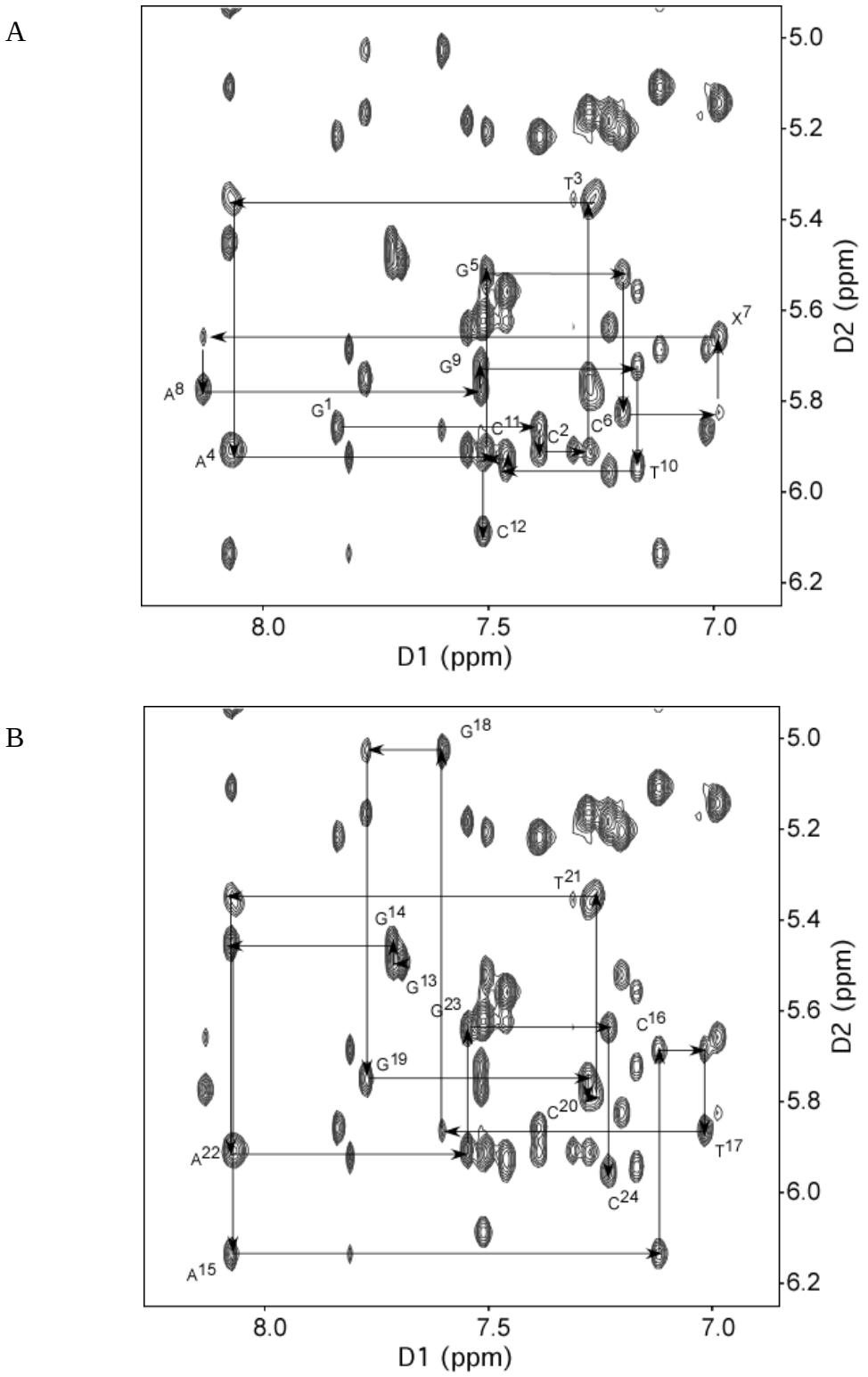


Figure 5-11: NOESY expansion plot of the anomeric proton to aromatic proton region for the *S*-BD-N3-dU modified sequence 5'-$G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}\underline{X}^{7}A^{8}G^{9}T^{10}C^{11}C^{12}$-3'·5'-$G^{13}G^{14}A^{15}C^{16}T^{17}G^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$-3'. (A) Nucleotides G^{1}-C^{12} of the modified oligonucleotide containing *S*-BD-N3-dU. (B) Nucleotides G^{13}-C^{24} of the complementary strand. NMR data were acquired at a mixing time of 250 ms and a frequency of 800.23 MHz.

A chemical shift analysis was performed on the anomeric and aromatic protons of the modified oligonucleotide, comparing the chemical shift values of the protons in the modified duplex to those in the unmodified duplex. In the modified strand, the largest perturbation occurred at the modified nucleotide, X^7, where the H1′ proton was shifted upfield more than 0.4 ppm and the H6 proton was shifted downfield by 0.3 ppm. In the complementary strand, chemical shift changes of almost 0.4 ppm were observed for the H1' protons of T^{17}, while the H1′ proton of G^{18} was shifted downfield by 0.5 ppm. A chemical shift change of 0.2 ppm was observed for the H8 proton of G^{19}. The changes in chemical shift, shown in Figure 5-12, are similar to those observed for the *R* diastereomer, and are indicative of a localized perturbation.

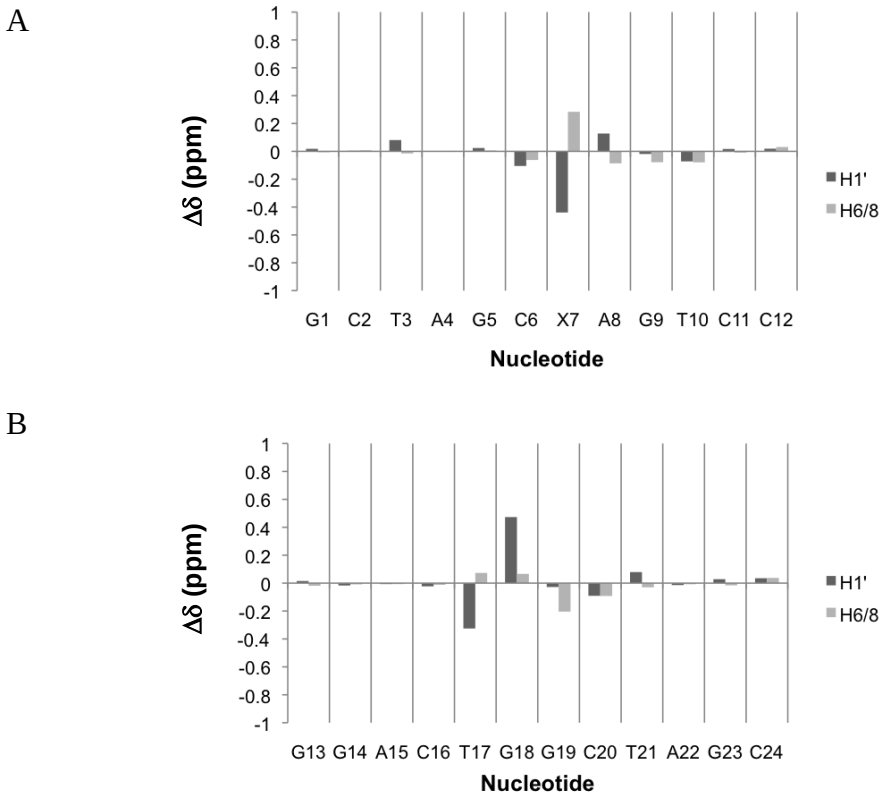


Figure 5-12: Chemical shift perturbations in the aromatic and anomeric protons of the S-BD-N3-dU modified sequence 5'-$G^1C^2T^3A^4G^5C^6\underline{X}^7A^8G^9T^{10}C^{11}C^{12}$-3'·5'-$G^{13}G^{14}A^{15}C^{16}T^{17}G^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$-3' relative to the unmodified oligodeoxynucleotide. (A) Nucleotides G^1-C^{12} of the modified oligonucleotide containing *S*-BD-N3-dU. (B) Nucleotides G^{13}-C^{24} of the complementary strand.

Butadiene Protons

The Hα, Hα', Hβ, Hγ, Hδ, and Hδ' protons were observed in a TOCSY spectrum. Hβ and Hγ were assigned using TOCSY and NOESY experiments. A DQF-COSY experiment was used to differentiate between Hδ and Hδ'. Like the R-diastereomer, the protons on the alpha carbon, Hα and Hα′, were ambiguously assigned. The chemical shift values for the butadiene protons are listed in Appendix A.

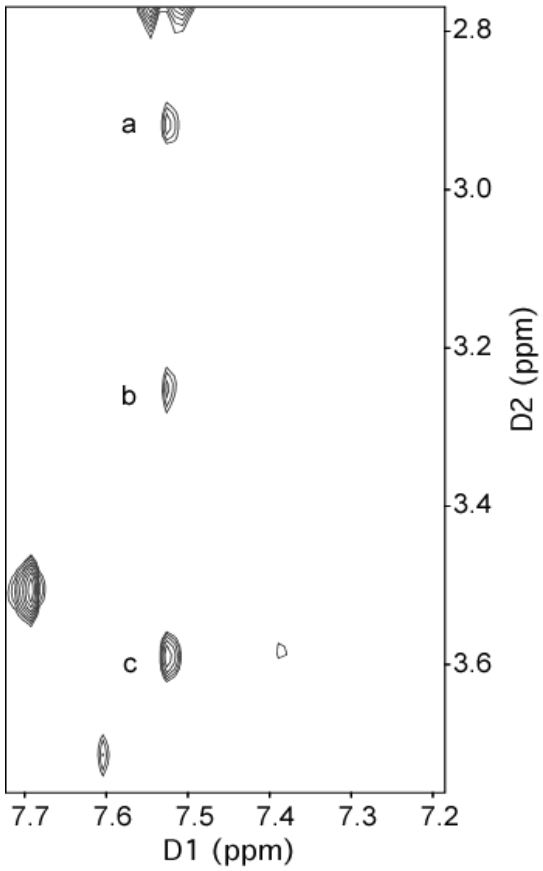


Figure 5-13: Dipolar proton:proton interactions between the S-BD moiety and the H2 proton of A8. NMR data were acquired at a mixing time of 250 ms and a frequency of 800.23 MHz.

Interactions between butadiene protons and other nucleotide protons were observed in the H2 region of the NOESY spectrum (Figure 5-13). Interactions were observed between A^8 H2 and the Hα, Hα′ and Hβ protons.

Exchangeable Protons

A NOESY spectrum acquired in H_2O was carried out in order to observe exchangeable imino protons. NOESY data were acquired at 7 °C at a frequency of 800 MHz in order to yield well-resolved imino cross-peaks. Nonetheless, the NOESY data yielded imino cross-peaks which were poorly resolved and overlapped.

The modification at the N3 position of X^7 prevents Watson-Crick hydrogen bonding between X^7 and G^{18}. Because hydrogen bonding had been disrupted at the A^8:T^{17} base pair in previous structures, we were interested in evaluating hydrogen-bonding interactions between A^8 and T^{17}. The H2 region (7-8 ppm) is shown in Figure 5-14 and provides insight to these interactions. The cross-peaks at 13.7 ppm are for those interactions between the imino protons of T^3 and T^{21} and the H2 protons of A^{22} and A^4, respectively. These cross-peaks are integrated to one. The cross-peak at 13.5 ppm (a), is assigned to T^{10} N3H: A^{15} H2 and is also integrated to one. Therefore, we do not observe a cross-peak between T^{17} N3H and A^8 H2. The absence of this cross-peak indicated that hydrogen-bonding interactions were disrupted such that the protons had a greater access to solvent leading to a more rapid exchange with solvent.

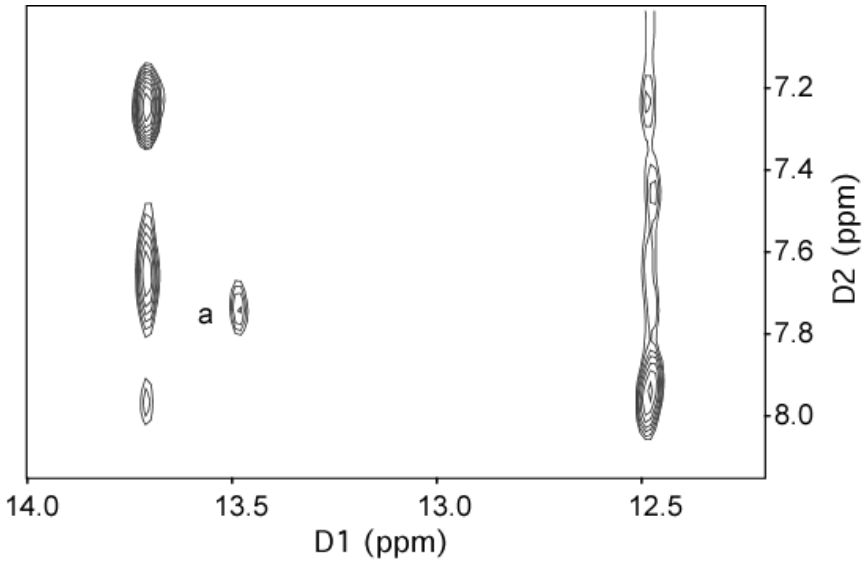


Figure 5-14: NOESY expansion plot of the imino and H2 proton region for the S-BD-N3-dU modified sequence 5'-$G^1C^2T^3A^4G^5C^6\underline{X}^7A^8G^9T^{10}C^{11}C^{12}$-3'·5'-$G^{13}G^{14}A^{15}C^{16}T^{17}G^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$-3'. NOESY cross-peak: a, A^{15} H2 to T^{10} N3H. NMR data were acquired at 7 °C at a mixing time of 250 ms and a frequency of 800.23 MHz.

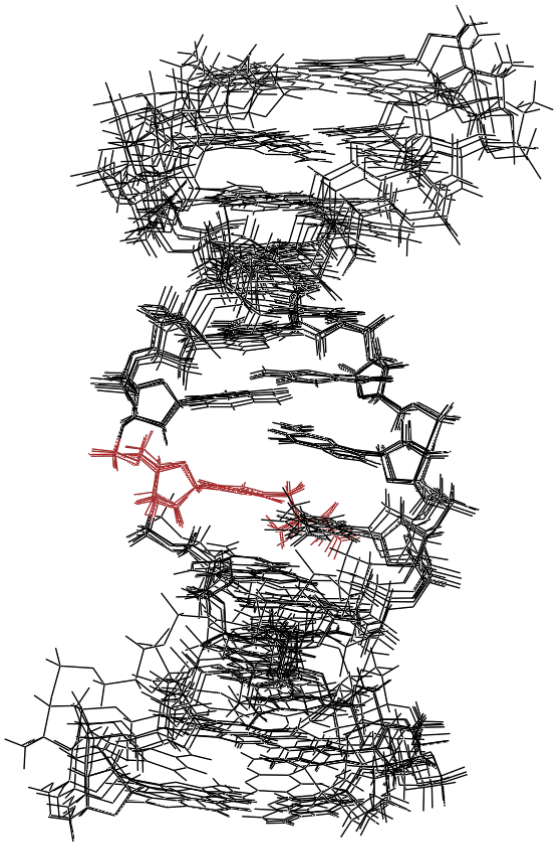


Figure 5-15: Overlay of the six S-BD-N3-dU modified oligonucleotide structures. The modified nucleotide is shown in red. The rmsd of the ten internal base pairs is 0.96 Å.

Structural Refinement

Twenty initial coordinate files were built as described in Chapter II and energy minimized prior to rMD calculations. Restrained molecular dynamics calculations were carried out on each of the twenty starting structures in implicit solvent using a simulated annealing protocol. An example of the MD input file is shown in Appendix D. A total of 592 distance restraints, evenly distributed amongst the residues in the oligonucleotide, were used to restrain MD calculations. These distance restraints included 354 intra-residue, 223 inter-residue, and 15 butadiene restraints and are listed in Appendix B. To restrain the backbone and sugar pucker, a total of 90 and 110 empirical torsion angle restraints were used, respectively. In addition to the distance and angle restraints, 44 Watson-Crick hydrogen-bonding restraints were used.

Prior to the rMD calculations, the root mean square deviation (rmsd) of the twenty starting structures was 3.58 Å. Following 20 ps of calculations, the structural coordinates which emerged from the calculations were compared using pairwise rmsd. Six of the

twenty starting structures converged to give an rmsd value of 0.96 Å. An overlay of the six structures is shown in Figure 5-15.

CORMA calculations were run on the average *S*-BD-N3-dU modified structure. The sixth root residual of the entire structure was determined to be 8.9×10^{-2}, indicating that the structure agreed well with NMR data. The individual nucleotide sixth root residuals ranged from 1.8×10^{-2} to 1.8×10^{-1}. The results from the CORMA calculations with respect to the individual nucleotides are shown in Figure 5-16.

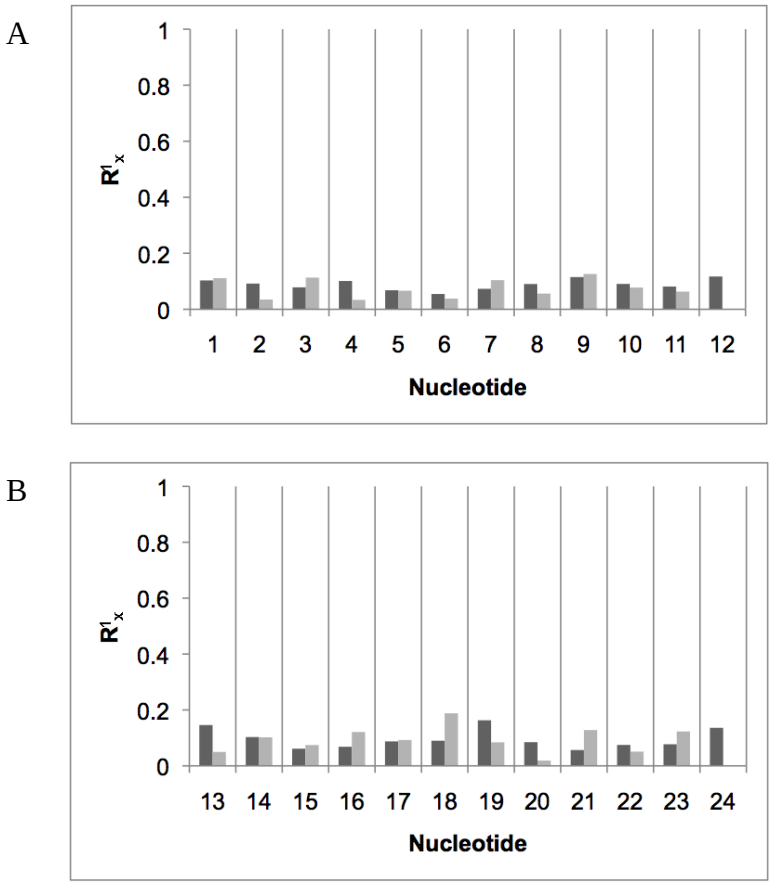


Figure 5-16: CORMA calculations on the average *S*-BD-N3-dU modified oligodeoxynucleotide calculated using a simulated annealing rMD protocol. (A) Nucleotides G^{1}-C^{12} of the modified oligonucleotide containing *S*-BD-N3-dU. (B) Nucleotides G^{13}-C^{24} of the complementary strand. The black bars represent intranuclear sixth root residuals and the gray bars represent internuclear sixth root residuals.

Structure of the S-BD-N3-dU Adduct Opposite dG

The average structure of the *S*-BD-N3-dU adduct opposite dG calculated using rMD calculations in implicit solvent shows that the major perturbation resulting from modification at the X^7 position occurs at the lesion site. The lesion site of the average structure is shown in Figure 5-17. The butadiene moiety is oriented in the minor groove toward the complementary strand. The position of the butadiene atoms forces T^{17} toward the major groove and displaces G^{18}, driving it in the 3′-direction, toward G^{19}. The position of the adduct causes the disruption of hydrogen bonding at both the lesion site and the 3′-neighboring base pair, A^8:T^{17}. We also observe that base stacking is disrupted at the lesion site, due to the disruption caused in the complementary strand by the butadiene lesion.

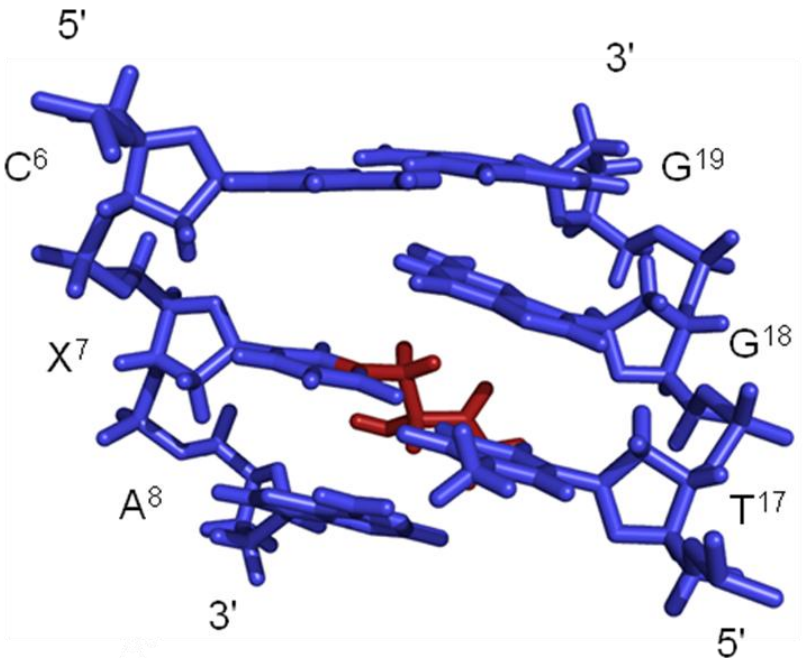


Figure 5-17: Lesion site of the average S-BD-N3-dU modified oligonucleotide calculated using a simulated annealing rMD protocol. The butadiene moiety (in gray) is oriented toward the complementary strand and displaces both G^{18} and T^{17}.

Hydrogen Bonding Analysis

Like the *R*-diastereomer, the initial calculations of the *S*-BD-N3-dU modified oligonucleotide produced an average structure which indicated that the potential existed for a hydrogen bond between the hydroxyl proton and one of the keto oxygen atoms on the modified nucleotide. For the *S*-diastereomer, it appeared that the hydrogen-bond would occur between the hydroxyl-proton and the O^2 atom. This was tested by extracting the average structure from simulated annealing calculations in implicit solvent and placing it in a truncated octahedral TIP3P solvent box after the addition of sodium counter ions. Prior to 5 ns of molecular dynamics calculations, the solvated molecule was energy minimized and equilibrated according to standard methods. Hydrogen-bonding occupancies were taken from MD trajectories using PTRAJ (AMBER 8.0) [153]. We observed that the S-diastereomer acted similarly to the R-diastereomer when placed opposite dG. Instead of hydrogen-bonding to one of the keto oxygen atoms, explicit solvent calculations predicted that the hydrogen-bond would occur at both the O^2 and O^4 atoms. The percent occupancy for the hydrogen bond between the hydroxyl proton and the O^4 atom on the modified uracil was 31.7%, while the percent occupancy for the hydrogen bond between the hydroxyl proton and the O^2 atom was 42.7%. The distance between the hydroxyl proton and O^4 atom was 2.75 Å and the O-H-O angle was determined to be 18.4 degrees. The distance between the hydroxyl proton and O^2 atom was 2.75 Å and the O-H-O angle was determined to be 20 degrees. These data predict that hydrogen bonding interactions occur at relatively similar percentages at both oxygen atoms on the modified deoxyuracil.

Helicoidal Analysis

An analysis of the helicoidal parameters of the average *S*-BD-N3-dU modified structure was performed following MD calculations. As a result of modification at the lesion site, the adduct caused a 25º bend in the duplex at the lesion site. Helicoidal base pair parameters were normal (Figure 5-18). The parameters for base-base interactions showed that at the lesion site there was a 2 Å change in the stretch and the stagger, as well as a 40° opening (Figure 5-19). Finally, helicoidal parameters for global inter-base interactions showed that there was a slight rise between X^7 and its two neighboring base pairs (Figure 5-20).

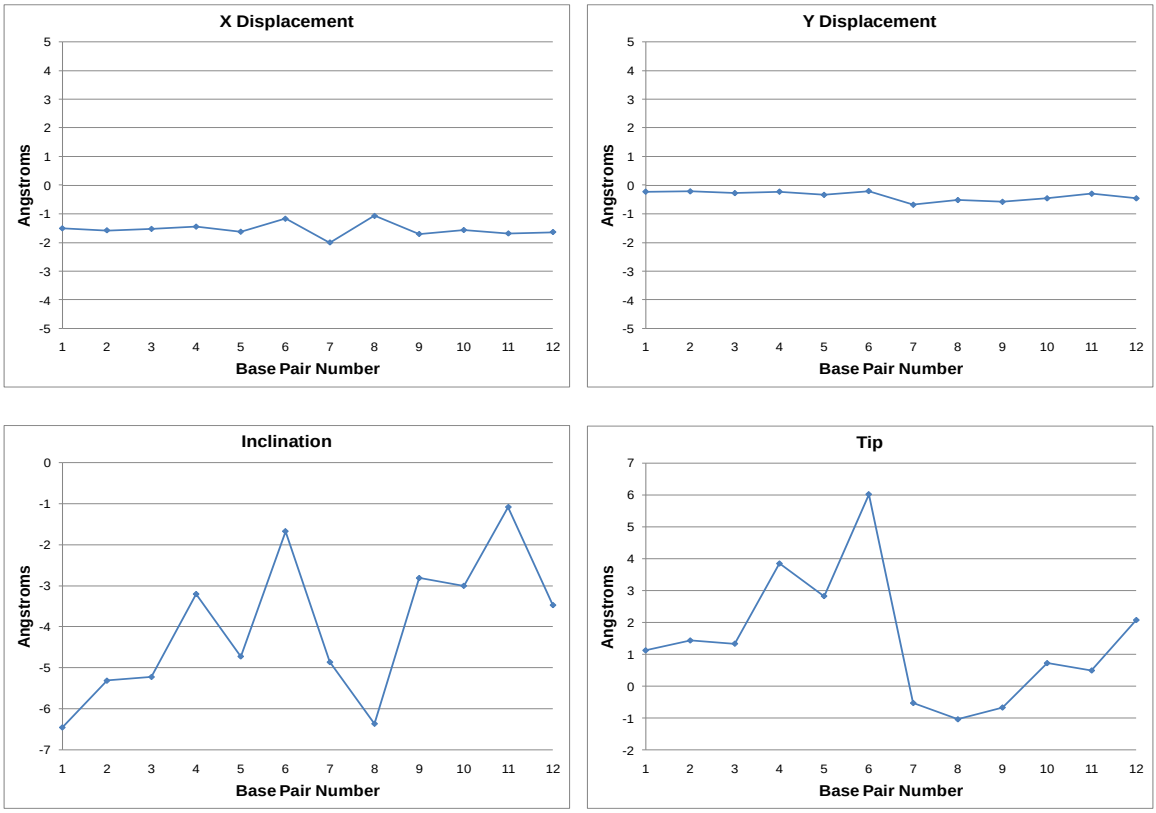


Figure 5-18: Global base pair helicoidal parameters. Helicoidal parameters for the *S*-BD-N3-dU modified oligonucleotide 5′-$G^1C^2T^3A^4G^5C^6\underline{X}^7A^8G^9T^{10}C^{11}C^{12}$-3′·5′-$G^{13}G^{14}A^{15}C^{16}T^{17}G^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}$-3′.

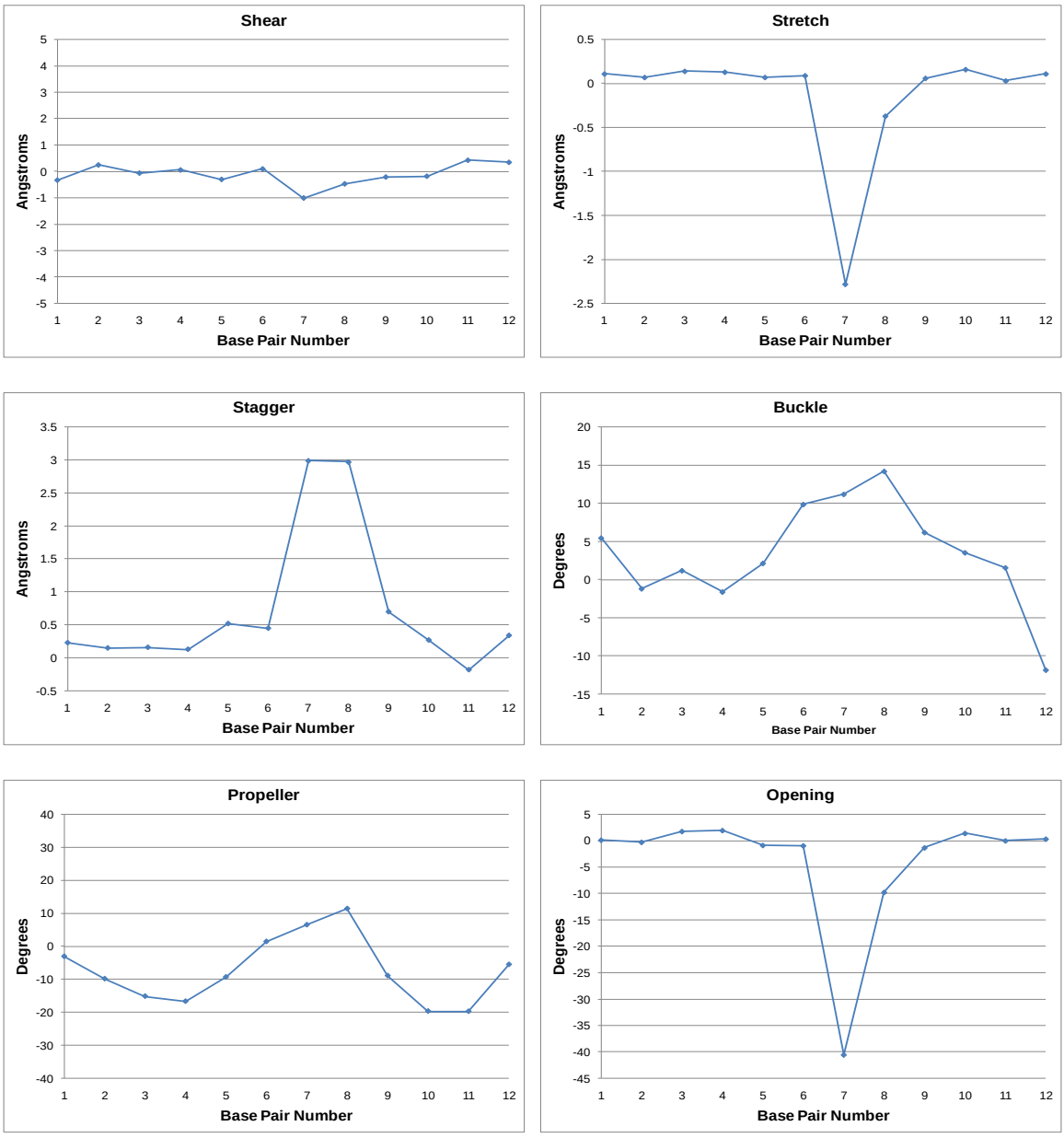


Figure 5-19: Global base-base helicoidal parameters. Helicoidal parameters for the *S*-BD-N3-dU modified oligonucleotide 5′-G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}X^{7}A^{8}G^{9}T^{10}C^{11}C^{12}-3′·5′-G^{13}G^{14}A^{15}C^{16}T^{17}G^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}-3′.

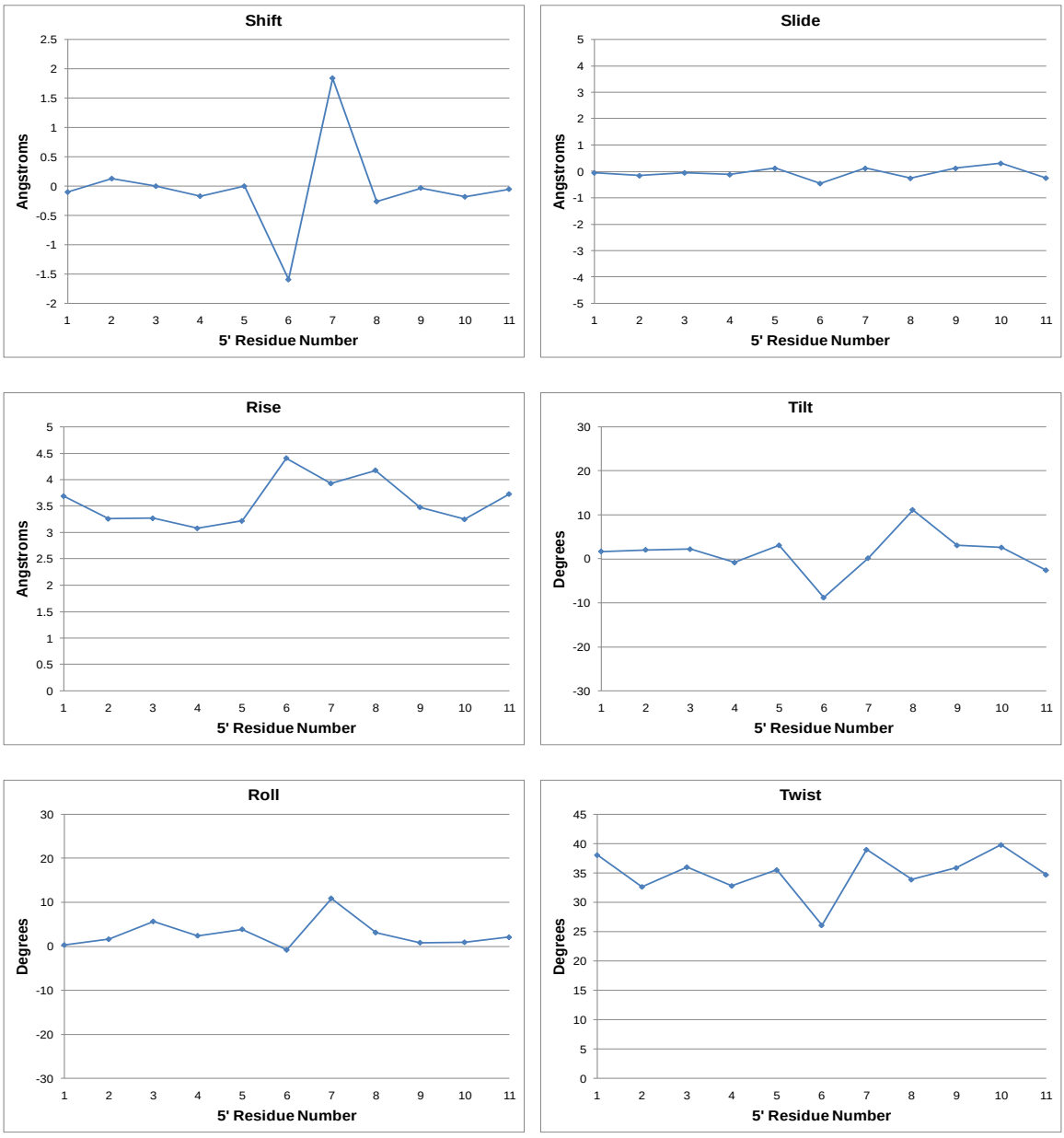


Figure 5-20: Global inter-base helicoidal parameters. Helicoidal parameters for the *S*-BD-N3-dU modified oligonucleotide 5′-G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}<u>X</u>^{7}A^{8}G^{9}T^{10}C^{11}C^{12}-3′·5′-G^{13}G^{14}A^{15}C^{16}T^{17}T^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}-3′.

Discussion

The final model of investigation was the dodecamer oligonucleotide, 5′-G^{1}C^{2}T^{3}A^{4}G^{5}C^{6}<u>X</u>^{7}A^{8}G^{9}T^{10}C^{11}C^{12}-3′·5′-G^{13}G^{14}A^{15}C^{16}T^{17}G^{18}G^{19}C^{20}T^{21}A^{22}G^{23}C^{24}-3′, where

both the *R*- and *S*-stereoisomers were site-specifically placed at the seventh position and annealed to a sequence that placed the butadiene lesion opposite dG. This model represents the insertion of dGTP opposite the lesion, which would result in non-mutagenic replication. The stereoisomeric BD-N3-dU adducts were shown to be more than 90% mutagenic in mammalian cells, with very little insertion of dG. It was of interest, therefore, to probe for differences in the structure of both isomers placed opposite dG when compared to the structures of the lesions opposite dA and dT.

Structural Analysis of the R-BD-N3-dU Adduct Opposite dG

The structure of the *R*-BD-N3-dU adduct opposite dG was determined using NMR experimental and empirical restraints. The average structure was determined to represent the NMR data accurately using CORMA back calculations. NMR data indicated that the overall structure would be similar to the canonical B-DNA structure, with a perturbation at the lesion site. The butadiene moiety is oriented in the 5′-direction with the two alpha protons closer in proximity to the 3′-neighbor, A^8. Interactions with the H2 of A^8 placed the moiety in the minor groove. Similarly to the R-diastereomer placed opposite dA, the Hγ, Hδ, and Hδ′ protons were oriented in the 5′-direction, producing a 4 Å increase in the rise between X^7 and its 5′-neighbor, C^6, as was observed in NOESY data. The butadiene moiety caused a disruption in hydrogen bonding at both the lesion site and the 3′-neighboring base pair, A^8:T^{17}. This disruption in hydrogen bonding was supported by the absence of a dipolar interaction between A^8 H2 and T^{17} N3H. Finally, MD calculations in explicit solvent predicted that the position of the BD

moiety forms a hydrogen bond between the hydroxyl proton of the adduct and the O^4 and O^2 atoms of the modified base.

Comparison of the Stereoisomeric BD-N3-dU Adducts Placed Opposite dG

The structure of the *S*-BD-N3-dU adduct opposite dG was determined using NMR experimental and empirical restraints. It was concluded that the final average structure represented the NMR data accurately according to CORMA back calculations. NMR data indicated that the major perturbation to the structure occurred at the lesion site. Several cross-peaks showing interactions between butadiene protons and other oligonucleotide protons were observed in NOESY data.

When placed opposite deoxyguanine, the stereoisomeric lesions differ with respect to the orientation of the butadiene moiety. While the *R*-diastereomer is oriented in the 5′-direction, the Hγ, Hδ, and Hδ′ protons of the *S*-diastereomer are directed toward the complementary strand, disrupting both hydrogen bonding and base stacking at the lesion site as well as the A^8:T^{17} base pair. For the *S*-diastereomer, the modification at the seventh position creates an opening in the helix. Like the *R*-diastereomer, MD calculations predict a hydrogen bond between the hydroxyl proton of the adduct and the O^4 and O^2 atoms of the modified base.

Summary

When placed opposite dG, the *R*- and *S*-BD-N3-dU adducts caused significant changes to the lesion site. Like the structures placing the modified nucleotide opposite dA, stereospecific differences were observed with respect to the orientation of the

butadiene moiety. For the *R*-diastereomer, the butadiene moiety was oriented in the 5′-direction; whereas, the butadiene moiety of the *S*-diastereomer was oriented in the minor groove toward the complementary strand. Both lesions resulted in disrupted base-stacking and hydrogen-bonding interactions at the lesion site.

Mutagenesis studies suggest that these adducts are highly mutagenic and that the insertion of dGTP opposite the lesion site only occurs at small percentages [168]. DNA lesion bypass studies, however, indicate that polymerase η inserts dGTP opposite the lesion site along with dATP [139]. The structures of the stereoisomeric BD-N3-dU adducts opposite dG are similar to those structures in which dA is complementary to the lesion site. This may indicate that the orientation of specific adducts plays a role in biological processing. X-ray crystal structures of Pol η-DNA-dNTP complexes would be a significant resource in gaining further insight into this role.

CHAPTER VI

CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions

Solution structures of the stereoisomeric N3-(2-hydroxy-3-buten-1-yl)-2'-dexoyuridine adducts site-specifically incorporated into dodecamer oligonucleotides opposite dA, dT, and dG have been elucidated using NMR spectroscopy and rMD calculations. The metabolites of butadiene, including butadiene monoepoxide and butadiene diepoxide, are thought to be involved in butadiene carcinogenicity, although, the exact mechanism of this is not clearly understood [63]. Mutagenesis studies on the *R*- and *S*-BD-N3-dU adducts, arising from butadiene monoepoxide, reveal that these lesions are highly mutagenic with the majority of mutations being C to T transitions and C to A transversions [159, 168]. A comparison of the mutagenic spectrum of both isomers suggests that there is a statistically significant preference for the insertion of dATP versus dTTP when the R-diastereomer is present in DNA. For the S-diastereomer, there is no partiality between dATP and dTTP, rather, its presence in DNA leads to the same percentage of insertion of both nucleotides [168].

NMR solution structures provide an illustration of both stereoisomers when placed opposite dA, dT, and dG in duplex DNA. In the case of each structure, the lesion creates major perturbations at the lesion site, producing an opening in the helix and causing a considerable rise between the modified nucleotide and it's 5′- and 3′-neighbors. It was observed that when the butadiene lesions were placed opposite purines, the adducts assumed different orientations, with the R-diastereomer being oriented in the 5′-direction

and the S-diastereomer being positioned more toward the complementary strand. However, when the lesions were inserted opposite the pyrimidine nucleotide, thymine, both isomers were oriented toward the complementary strand and fixed between the thymine opposite the lesion and its 5′-neighboring base. IIrrespective of the nucleotide placed opposite the lesion, the orientation of both diastereomers led to a disruption in both base stacking and hydrogen bonding at the X^7:Y^{18} and A^8:T^{17} base pairs.

The greater percentage of C to T mutations observed for the R diastereomer in site-specific mutagenesis studies may be partially explained by the alternate orientations of the stereoisomeric butadiene functional groups when placed opposite deoxyadenine. X-ray crystal structures of Dpo4-DNA-dNTP ternary complexes also suggested differences in the orientation of the butadiene moiety when placed opposite dA. There is evidence, however, that these lesions are not substrates for Dpo4, therefore, crystal structures of complexes containing pol η may paint a more accurate picture of the orientation of the butadiene moieties during biological processing.

In addition, NMR solution structures of the BD-N3-dU lesions opposite deoxyguanine reveal that the butadiene moieties assume similar orientations when placed opposite dG as when placed opposite dA. These structural similarities might explain why pol η inserts mainly dATP and dGTP at the lesion site. Again, crystal structures of complexes comprised of pol η will be useful for further understanding the structures of these adducts as they relate to BD-mediated mutagenicity.

Future Directions

X-ray Crystallographic Investigations of Pol η-DNA-dNTP Complexes

Ternary Dpo4-DNA-dNTP complexes, containing the *R*- and *S*-BD-N3-dU adducts, show that these DNA lesions are not substrates for Dpo4 and that the complexes are not catalytically active [172]. Evidence arising from DNA lesion bypass assays suggests that pol η and pol ζ are responsible for bypassing and extending past these lesions [139]. Pol η inserts dATP and dGTP at the lesion site. Pol ζ is then required to extend the oligonucleotide following insertion by pol η. These studies indicate that pol η might be responsible for the C to T transitions occurring in COS-7 mammalian cells. Therefore, X-ray crystal structures of a ternary complex containing pol η might provide structural information relevant to the processing of the BD-N3-dU adducts.

Continuing Investigations of Adducts Arising from Butadiene Metabolites

DNA adducts have been associated with carcinogenic processes in living organisms [63]. The active metabolites of BD are thought to be primarily responsible for the deleterious effects of butadiene, due to the formation of DNA adducts as well as DNA-DNA and DNA-protein cross-links. However, the exact involvement of these metabolites in the mechanism of butadiene-induced carcinogenicity is not known.

BD exposure in bacterial and mammalian cells induced primarily A to T and G to T transversions [89, 109]. The majority of mutations caused by the BD-N3-dU adducts are C to T and C to A. Therefore, although these adducts are significant due to their high mutagenicity, they do not represent the BD-induced adducts which lead to the

predominant mutations *in vivo*. Future research on adducts arising from BD metabolites remains to be carried out. The N7-dG adducts arising from BDO and the N7,N7-dG cross-links arising from BDO_2 represent the majority of BD-induced lesions [125, 173]. Although N7-dG-adducts are highly unstable due to the imidazole ring opening, their presence in DNA, even for shorter periods of time, might generate significant biological consequences [63]. In light of this, advances in the synthesis and structural elucidation of these adducts will be key in further investigations of the mechanisms of BD-induced carcinogenicity.