7098 *Biochemistry* **2009,** *48,* 7098–7109
DOI: 10.1021/bi9006429

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

NMR and Computational Studies of Stereoisomeric Equine Estrogen-Derived DNA Cytidine Adducts in Oligonucleotide Duplexes: Opposite Orientations of Diastereomeric Forms[†]

Na Zhang,[‡,#] Shuang Ding,[§,#] Alexander Kolbanovskiy,[∥] Anant Shastry,[∥] Vladimir A. Kuzmin,[∥] Judy L. Bolton,[⊥] Dinshaw J. Patel,[‡] Suse Broyde,[§] and Nicholas E. Geacintov*[,∥]

[‡]*Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065,* [§]*Department of Biology and* [∥]*Department of Chemistry, New York University, New York, New York 10003, and* [⊥]*Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, Chicago, Illinois 60612* [#]*N.Z. was the major contributor to the NMR experiments while S.D. assumed the major responsibilities for the computational and modeling aspects of this work*

Received April 15, 2009; Revised Manuscript Received June 9, 2009

ABSTRACT: The equine estrogens equilin (EQ) and equilenin (EN) are the active components in the widely prescribed hormone replacement therapy formulation Premarin. Metabolic activation of EQ and EN generates the catechol 4-hydroxyequilenin (4-OHEN) that autoxidizes to the reactive *o*-quinone form in aerated aqueous solutions. The *o*-quinones react predominantly with C, and to a lesser extent with A and G, to form premutagenic cyclic covalent DNA adducts in vitro and in vivo. To obtain insights into the structural properties of these biologically important DNA lesions, we have synthesized site-specifically modified oligonucleotides containing the stereoisomeric 1′*S*,2′*R*,3′*R*-4-OHEN−C3 and 1′*R*,2′*S*,3′*S*-4-OHEN−C4 adducts derived from the reaction of 4-OHEN with the C in the oligonucleotide 5′-GGTAG**C**GATGG in aqueous solution. A combined NMR and computational approach was utilized to determine the conformational characteristics of the two major 4-OHEN−C3 and 4-OHEN−C4 stereoisomeric adducts formed in this oligonucleotide hybridized with its complementary strand. In both cases, the modified C adopts an *anti* glycosidic bond conformation; the equilenin distal ring protrudes into the minor groove while its two proximal hydroxyl groups are exposed on the major groove side of the DNA duplex. The bulky 4-OHEN−C adduct distorts the duplex within the central GC*G portion, but Watson−Crick pairing is maintained adjacent to C* in both stereoisomeric adducts. For the 4-OHEN−C3 adduct, the equilenin rings are oriented toward the 5′-end of the modified strand, while in 4-OHEN−C4 the equilenin is 3′-directed. Correspondingly, the distortions of the double-helical structures are more pronounced on the 5′- or the 3′-side of the lesion, respectively. These differences in stereoisomeric adduct conformations may play a role in the processing of these lesions in cellular environments.

There is significant evidence from epidemiological studies that there is an increased risk of developing cancer of the breast and endometrium in women who are exposed to estrogens as a result of prolonged hormone replacement therapy (HRT) (*1*−*5*). Premarin is a widely used HRT formulation that contains the equine estrogens equilin (EQ) and equilenin (EN) as major components; they resemble human estrogens but contain one or two additional double bonds in the B-ring, respectively. Both are metabolized to 4-hydroxyequilin (4-OHEQ) and 4-hydroxyequilenin (4-OHEN),[1] respectively (*6*, *7*), and undergo reversible autoxidation to their *o*-quinone forms. However, the *o*-quinone derived from 4-OHEQ isomerizes to the 4-OHEN-derived *o*-quinone that yields 4-OHEN as a result of redox cycling (*7*). These quinoids react with DNA to form stereoisomeric cyclic covalent adducts with dC, dA, and dG (but not dT), yielding four stereoisomers in each case (*8*−*12*). These adducts are likely to contribute to the development of breast and endometrial cancers (*12*), as they have been detected in rat mammary tissues (*13*) and in human breast tissue of women (*14*).

[†]The experimental portion of this work was supported by National Institutes of Health and National Cancer Institute Grants CA112412 (N.E.G.) and CA-046533 (D.J.P.), and the computational aspects were supported by Grant CA-75449 (S.B.). Partial support for computational infrastructure and systems management was also provided by Grant CA-28038 (S.B.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

*Corresponding author. Tel: 212-998-8407. Fax: 212-998-8421. E-mail: ng1@nyu.edu.

[1]Abbreviations: 4-OHEN, 4-hydroxyequilenin; 4-OHEN−C, 4-hydroxyequilenin−cytosine; MD, molecular dynamics; rmsd, root-mean-square deviations; MM-PBSA, molecular mechanics Poisson−Boltzmann surface area.

 Published on Web 06/15/2009

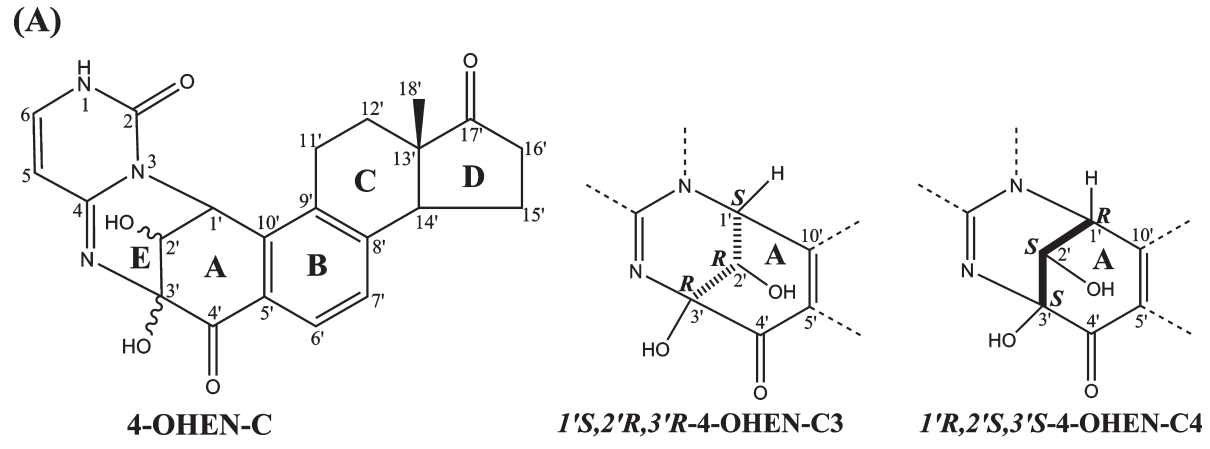

FIGURE 1: (A) Chemical structures and stereochemical characteristics of the 4-OHEN−C3 and 4-OHEN−C4 adducts. (B) Sequence of the 11mer B-DNA duplex. The starred bases denote the sites of the lesions studied in the singly modified duplexes as follows: the 4-OHEN−C3 lesion positioned at site C17* (C3(17*)) or 4-OHEN−C4 positioned at site C7* (C4(7*)) or C17* (C4(17*)).

Site-directed mutagenesis experiments in vitro and in vivo have shown that these DNA adducts impede the activities of polymerases and that translesion bypass occurs in an error-prone manner (*15*, *16*).

The chemical structures of the 4-OHEN adducts have been determined by Shen et al. (*8*, *9*). The Watson−Crick hydrogen-bonding edges of all the modified bases are obstructed by the formation of the cyclic adduct. Each 4-OHEN adduct has three chiral carbon atom centers within the region connecting the base to the 4-OHEN A-ring (Figure 1). However, the two covalent bonds formed between 4-OHEN and the base adopt only a *cis* configuration because the connecting ring would be highly strained if the configuration were *trans* (*10*). Thus, only four different stereoisomeric adducts are observed for each of the modified bases, with two of these formed in greater abundance in the case of the 4-OHEN−dC and 4-OHEN−dA adducts (*6*, *8*).

In the reaction of 4-OHEN with oligonucleotides in vitro, the cytosine base is the preferred target base. Guanine adducts are observed only in oligonucleotides that contain only G and the unreactive T residues (*11*). The absolute configurations of the 4-OHEN−C and 4-OHEN−A adducts have been recently determined by computed and experimentally measured optical rotatory dispersion, electronic circular dichroism, and mass spectrometry techniques (*17*). The four 4-OHEN−C stereoisomeric adducts form two pairs with inverse *R* and *S* configurations at the linkage site. The chirality of the C3′ atom determines the handedness of the ring system of the adduct, and the C2′ chirality governs the orientation of its OH group (Figure 1). The 4-OHEN−C1 and 4-OHEN−C2 stereoisomers are a pair with inverse *R* and *S* configurations, as are the 4-OHEN−C3 and 4-OHEN−C4 stereoisomers. These pairs are thus nearly mirror images of one another since only the absolute configuration of the OH group about the 2′ carbon atom is different. The symmetry is broken only by the equilenin D-ring with its C18′ methyl group that has *S* absolute configuration. In the 4-OHEN−C1/C2 pair, the C2′ OH group is oriented toward the cytosine base, while in the 4-OHEN−C3/C4 pair, this methyl group points toward the equilenin rings. In the reaction of 4-OHEN with cytosine or 2′-deoxycytidine, the reaction yields of C3/C4 (or dC3/dC4) are significantly larger than those of C1/C2 (or dC1/dC2) (*11*). It was hypothesized that the orientation of the bulky C2′ OH group toward the nucleobase target atoms in the case of C1/dC1 and C2/dC2 lowers the probability of formation of the cyclic products; however, in the case of the C3/dC3 and C4/dC4 adducts, this steric hindrance effect is absent, thus favoring a higher reaction yield of these latter products (*17*).

Using computational approaches, we have previously investigated the structures of the four stereoisomeric 4-OHEN−C, −A, and −G adducts (Figure 1) at the single base (*18*) and duplex DNA level (*19*−*21*). A key finding was that in the DNA duplex the pairs of stereoisomers with near mirror image symmetry are oriented oppositely with respect to the modified base with the D-ring of the estrogen moiety directed toward the 3′- or the 5′-end of the modified strand. The adducts can adopt *syn* or *anti* conformations, with the equilenin ring system positioned in the major or minor grooves; the stereochemical properties and the nature of the modified base govern the specific conformations adopted. Moreover, these computational approaches predict that the distortions to the DNA structures imposed by the lesions should be different for each stereoisomeric adduct and depend on the base modified. This suggests that the processing of these lesions by replicative and repair enzymes might also be different, as has been observed in certain cases (*15*, *22*, *23*). In vitro primer extension studies conducted with the Y-family bypass polymerase pol η showed that the bypass frequency differed by 2 orders of magnitude in the members of a pair of 4-OHEN−dC stereoisomers whose CD spectra were opposite in sign (*15*). In addition, NER experiments with human HeLa cell extracts have shown that the 4-OHEN−C3 adduct is excised more efficiently, by a factor of ~2.5, than 4-OHEN−C4 (*17*, *22*).

Motivated by the possibilities of ultimately relating the conformational properties of these adducts with their biological effects, we have utilized a combined NMR and computational approach to deduce the structures of the 4-OHEN−C3 and 4-OHEN−C4 adducts in the GC*G sequence context of the (5′-GGTAGC*GTAGG)·(3′-CCATCGCTACC) duplex (C* designates the damaged cytosine). The equilenin ring systems are found to be oriented on opposite sides of the modified cytosine residue in the 4-OHEN−C3 and 4-OHEN−C4 adducts, being on the 5′-side in the former case and the 3′-side in the latter. Moreover, distortions of the DNA duplex structure are also more pronounced on the 5′- and 3′-sides, respectively. A comparison of the structural features for the present GC*G sequence context with the GC*T sequence previously investigated by computational approaches suggested that base sequence context may influence the conformational properties of the 4-OHEN−dC adducts as shown for other bulky adducts (*24*, *25*). Understanding the impact of adduct stereochemistry and sequence context effects that may modulate nucleotide excision repair of these DNA adducts may lead to new insights into the genotoxic properties of the DNA damage associated with equine estrogens present in HRT formulations.

METHODS

Preparation of 4-OHEN−C Adducts in DNA Duplexes. Oligonucleotide sequences were synthesized using automated DNA synthesis techniques and purified by reversed-phase HPLC methods, and their masses were verified by matrix-assisted laser desorption ionization with time-of-flight detection (MALDI-TOF) MS. The modified oligonucleotides 5′-GGTAGC*-GATGG containing a single 4-OHEN−C3 or 4-OHEN−C4 adduct at position C17 (the numbering system is also defined in Figure 1) were prepared as previously described in detail (*11*). The 4-OHEN was reacted directly with the oligonucleotides in aqueous solutions following procedures described by Shen et al. (*9*). Briefly, 0.10 mL of a 3.6 mM solution of 4-OHEN in DMSO (3.6 mM) was added to 1 mL of a 25 mM sodium phosphate buffer solution, pH 7, containing the 11mer oligonucleotides (0.050 mM). This mixture was incubated for 12 h at 37 °C. The covalently modified oligonuclotides were separated from the unmodified oligonucleotides by a Waters model 510 solvent delivery HPLC system using a 250 mm × 10.0 mm (5 μm) Phenosphere C18 column (Phenomenex) and acetonitrile/triethylammonium acetate (50 mM) solution gradients as described in detail elsewhere (*11*). The different oligonucleotide fractions were isolated and further purified and desalted by subsequent HPLC cycles. The modified oligonucleotides 5′-GGTAGC*GATGG were annealed with their natural complementary strands 5′-CCATCGCTACC to form fully complementary duplexes. The absolute configurations of the 4-OHEN−C lesions were established as described in detail by Ding et al. (*17*). The 4-OHEN−C adducts studied have either 1′*S*,2′*R*,3′*R* (4-OHEN−C3) or 1′*R*,2′*S*,3′*S* (4-OHEN−C4) absolute configurations (Figure 1).

NMR Methods. The NMR studies were carried out for the duplexes formed by both 4-OHEN−C3 and 4-OHEN−C4 adduct modified strands with their complementary strands. All one- and two-dimensional NMR experiments were carried out using a Varian INOVA 600 MHz spectrometer. The one-dimensional proton NMR spectra of the exchangeable protons were collected in aqueous buffer (100 mM NaCl, 10 mM phosphate, pH 6.8) at different temperatures (0, 4, 7.5, 15, 20, and 25 °C) for the 4-OHEN−C3 and 4-OHEN−C4 duplexes. Under the same buffer condition, two-dimensional NOESY (for both exchangeable and nonexchangeable protons) and COSY and TOCSY (for nonexchangeable protons only) NMR spectra were collected at 7.5 °C for the 4-OHEN−C3 and 15 °C for the 4-OHEN−C4 adduct duplexes. The NMR data sets were processed using Varian software and analyzed using the FELIX program (Accelrys, Inc.).

NMR Distance Restrained MD Simulations. *(A) Starting Models*. We started with an energy-minimized B-DNA structure, whose sequence is given in Figure 1, and replaced the cytosine with the quantum mechanically optimized 4-OHEN−C adduct structures (*18*). The modified cytosine was rotated through the glycosidic torsion to locate an initial conformation with minimal collisions within the *anti* domain, since the data unequivocally showed that the adducted cytosine was *anti* (see Results).

In order to elucidate the solution structures of the 4-OHEN−C adducts in the 11mer DNA duplexes, NMR distance-restrained molecular dynamics simulations were carried out using the AMBER 8.0 package (*26*), the parm99.dat parameter set, and the same force field parameters employed previously for the 4-OHEN−C adducts (*19*). First, 2 ns of unrestrained simulations was performed. Then, the DNA duplex structure was subjected to distance restraints based on the obtained NMR intermolecular NOEs between the adduct and the DNA, which were classified as of medium and weak intensity (Tables 1 and 2); proton pairs were then restrained to distances of 3.5−5 Å (medium NOEs) and 5−7.5 Å (weak NOEs). The restrained MD simulations were initially carried out at 293 K with a force constant of 2 kcal mol^{-1} Å^{-2}. The force constants employed in the computations of the distances, constrained by the experimental bounds, were slowly scaled up to 20 kcal mol^{-1} Å^{-2} for all proton distances and for all hydrogen bonds over a period of 50.0 ps. Hydrogen-bonding distance restraints were imposed to align the experimentally identified intact Watson−Crick pairs within ±0.10 Å of ideal target values, including all base pairs except the lesion in 4-OHEN−C4, and except at the lesion site and the G18 C5 pair in 4-OHEN−C3 in accord with the upfield shift and broadening of the imino proton resonance of G18 for this case (see Results). The restrained MD simulations were carried out for 1 ns at 293 K with retention of the full set of distance restraints. The last 0.5 ns of the trajectory was employed for structural analyses.

(B) Unrestrained MD Simulations. We have also carried out unrestrained MD simulations for 4-OHEN−C3 and 4-OHEN−C4 adducts in the same sequence as in the NMR-restrained simulations (Figure 1). Our purpose was to obtain relative free energies for the *syn* and *anti* structures in order to gain insight on the NMR observed *anti* preference. We utilized the same initial models as for the NMR solution structures for the *anti* domain. For the *syn* domain models we rotated the modified cytosine through the glycosidic torsion to locate an initial conformation with minimal collisions within the *syn* domain.

We utilized the same AMBER 8.0 package (*26*) and force field parameters as for the NMR restrained computations. Details of the MD simulation protocols are provided in Supporting Information. MD simulations of the 4-OHEN−C adducts and the unmodified control were carried out for 10 ns. The structures generally fluctuate stably after 6 ns, and the last 4 ns time frame was employed for our further analyses. Plots of root-mean-square deviations (rmsd) of the current structure relative to the average structure, as a function of time, are shown in Supporting Information Figure S1.

Table 1: Experimental Intermolecular NOEs and Trajectory-Averaged Distances from the NMR Restrained Simulation of the 4-OHEN−C3 Adduct

4-OHEN−C3 proton	chemical shift[a] (ppm)	intermolecular NOEs	trajectory-averaged distance (Å)
H1	5.54		
H2	4.01		
H6	7.53	G16(NH1) (m)	4.2 (0.3)
		C7(NH_2-a) (m)	4.1 (0.3)
		C7(NH_2-b) (m)	4.6 (0.3)
		C7(H6) (w)	5.9 (0.3)
		C7(H5) (w)	5.3 (0.3)
H7	6.23	G16(NH1) (m)	4.7 (0.3)
		C7(NH_2-a) (w)	5.3 (0.3)
		C7(NH_2-b) (w)	5.8 (0.3)
		C7(H6) (w)	5.0 (0.2)
H11-a	2.93		
H11-b	2.08		
H12-a	0.85		
H12-b	1.46		
H14	3.99		
H15-a	1.17	G16(NH1) (w)	5.2 (0.2)
H15-b	1.27	G18(NH1) (w)	7.0 (0.5)
H16-a	1.80		
H16-b	2.30	G16(NH1) (w)	7.2 (0.3)
CH_3	−0.45	G16(NH1) (m)	4.5 (0.4)
		C7(H6) (w)	6.5 (0.3)
		C7(H5) (w)	7.4 (0.2)

[a] The chemical shifts of the adduct were recorded at 7.5 °C in H_2O and D_2O buffered solution.

Table 2: Experimental Intermolecular NOEs and Trajectory-Averaged Distances from the NMR Restrained Simulation of the 4-OHEN−C4 Adduct

4-OHEN−C4 proton	chemical shift[a] (ppm)	intermolecular NOEs	trajectory-averaged distance (Å)
H1	5.99	C7(H5) (m)	5.2 (0.1)
H2	4.06	C7(H5) (w)	6.6 (0.2)
H6	7.22	G18(NH1) (m)	3.7 (0.2)
		C5(NH2-a) (m)	3.7 (0.3)
		C5(NH2-b) (m)	4.0 (0.3)
		C5(H6) (w)	5.0 (0.2)
		C5(H5) (m)	4.4 (0.3)
H7	6.29	G18(NH1) (m)	3.4 (0.1)
		C5(NH2-a) (w)	5.0 (0.2)
		C5(H6) (w)	5.6 (0.3)
H11-a	2.45	G16(NH1) (m)	3.8 (0.2)
H11-b	2.16	G16(NH1) (w)	5.3 (0.3)
H12-a	1.26	G16(NH1) (w)	6.6 (0.2)
H12-b	1.38	G16(NH1) (w)	5.3 (0.3)
H14	2.30	G18(NH1) (m)	3.9 (0.3)
H15-a	1.58	G18(NH1) (m)	4.0 (0.2)
H15-b	1.87	G18(NH1) (m)	5.2 (0.1)
H16-a	2.14	G18(NH1) (w)	5.9 (0.4)
H16-b	2.20	G18(NH1) (w)	6.8 (0.3)
CH_3	0.19	G16(NH1) (m)	4.7 (0.3)
		C7(NH2-a) (m)	4.1 (0.3)
		C7(NH2-b) (m)	3.8 (0.2)
		C7(H5) (w)	4.0 (0.2)
		C7(H6) (m)	3.8 (0.2)
		G6(H8) (w)	5.8 (0.3)

[a] The chemical shifts of the adduct were recorded at 15 °C in H_2O and D_2O buffered solution.

(C) Free Energy Analyses. The molecular mechanics Poisson−Boltzmann surface area (MM-PBSA) method (*27*, *28*) in AMBER was employed for thermodynamic analyses. Details of the free energy calculations are given in Supporting Information.

(D) Structural Analyses. The best representative structure in the MD trajectory for each adduct was obtained by using the cluster analysis option with 2.0 Å rmsd criterion in MOILView (*29*). The five representative structures were obtained with a 1.2 Å rmsd criterion.

The PTRAJ module of AMBER 8 (*26*) was employed for structural analyses. Stacking interactions were evaluated by computing the van der Waals interaction energies between adjacent base pairs, including the damaged cytosine and partner G pair, with the program ANAL of AMBER 8. The equilenin moiety, nearly perpendicular to the attached cytosine residue, is not included in the base stacking interaction evaluation. The interaction energy between the 4-OHEN−C adduct and the DNA was computed with the program ANAL from the molecular mechanical energy, including internal energy, van der Waals energy, and electrostatic energy. In addition, we employed our hydrogen bond quality index (*30*), I_H, to quantitatively assess the deviation from ideal Watson−Crick hydrogen-bonding distances and angles for base pairs:

$$I_H = \sum_{D-H\cdots A} [(d_{DA} - {d^0}_{DA})^2 + (1+\cos\gamma)^2]$$

d_{DA} is the instantaneous donor−acceptor distance, ${d^0}_{DA}$ is an ideal donor−acceptor distance (N4 (C) to O6 (G) is 2.91 Å, N1 (G) to N3 (C) is 2.95 Å, N2 (G) to O2 (C) is 2.86 Å, N6 (A) to O4 (T) is 2.95 Å, and N1 (A) to N3 (T) is 2.82 Å), and γ is the instantaneous D−H···A bond angle, which has an ideal value of 180°. The summation is over all the Watson−Crick hydrogen bonds in a given pair and over the selected trajectory frame (0.5−1 ns for the NMR restrained simulations and 6−10 ns for the unrestrained simulations). I_H adopts a value of 0 when ideal Watson−Crick bonding is maintained. DNA duplex groove dimensions and helicoidal parameters for the ensemble of structures of each stereoisomer were computed with MD Toolchest (*31*, *32*).

RESULTS AND DISCUSSION

NMR Characteristics of 4-OHEN−C3 and 4-OHEN−C4 Duplexes. The properties of the 4-OHEN−C3 and 4-OHEN−C4 adducts embedded in the oligonucleotide 5′-GGTAGC*GATGG at position C17 (Figure 1), annealed with their unmodified complementary 11mer strands to form duplexes, were analyzed by a combination of NMR and computational methods.

(A) Exchangeable Proton Spectra. The exchangeable proton 1D NMR spectra (10.0−14.0 ppm) of the 11mer duplexes containing 4-OHEN−C3 and 4-OHEN−C4 in H_2O buffer solution, pH 6.8 at 7.5 and 15 °C, respectively, are shown in panels A and B of Figure 2, respectively. The base numbering scheme is also defined in this figure. In the case of the unmodified duplex (5′-CCATCGCTACC)·(3′-GGTAGCGATGG), all observable imino proton resonances appear in the 12.5−14.0 ppm range (Supporting Information Figure S2) as expected for Watson−Crick hydrogen-bonded duplexes (*33*). The 4-OHEN−C3 duplex exhibits two well-resolved but broadened upfield-shifted imino proton resonances at 9.8 and 11.8 ppm. In the case of the 4-OHEN−C4 duplex, only one upfield shifted imino proton resonance (G6) is observed at 10.3 ppm.

Expanded regions of the NOESY contour plot (200 ms mixing time) for the 4-OHEN−C3-containing duplex in H_2O buffer,

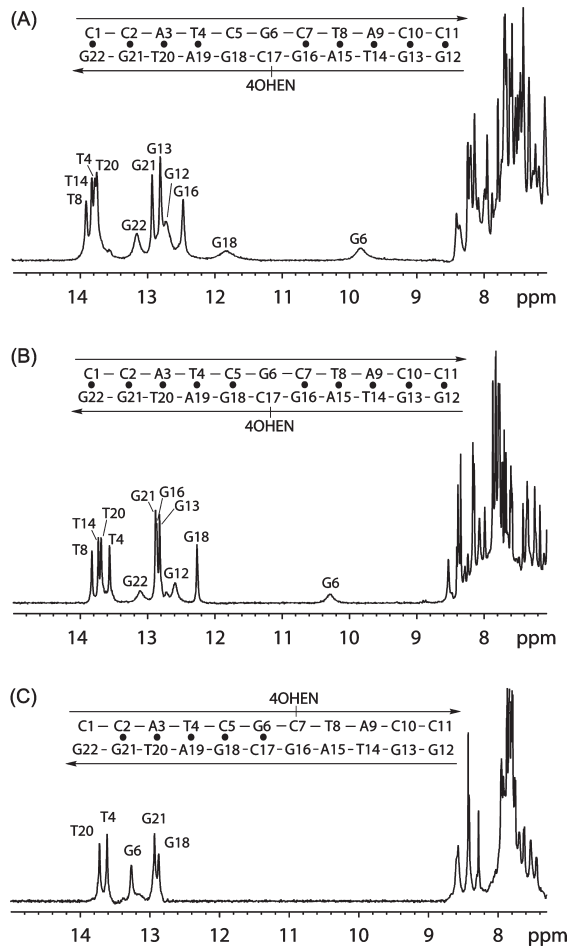

FIGURE 2: Sequence of the 11mer duplex containing the 4OHEN−C adduct (established Watson−Crick base pairs are indicated by solid black dots) and its imino proton NMR spectra (7.0−15.0 ppm). The imino proton assignments are labeled. (A) The 4-OHEN−C3 lesion positioned at site C17*. The spectrum is in 100 mM NaCl, 10 mM phosphate, and H_2O, pH 6.8 at 7.5 °C. (B) The 4-OHEN−C4 lesion positioned at site C17*. The spectrum is in 100 mM NaCl, 10 mM phosphate, and H_2O, pH 6.8 at 15 °C. (C) The 4-OHEN−C4 lesion positioned at site C7*. Spectra are in 100 mM NaCl, 10 mM phosphate, and H_2O, pH 6.8 at 12.5 °C.

pH 6.8 at 7.5 °C, are plotted in Figure 3. The imino and amino protons in the duplex have been assigned by standard procedures (*34*, *35*). Panel A identifies NOEs between imino protons of adjacent base pairs from G16 to G13 on the 5′-side of C17* to G13 and from T20 to G21 on the 3′-side. In Figure 3B, the NOEs between imino protons (12.4−14.0 ppm) and amino protons (5.0−9.0 ppm) are shown. The observed imino to amino proton connectivities from T4·A19 to C2·G21, and from C7·G16 to C10·G13, indicate Watson−Crick base pairing at all nonterminal sites except the central C5·G18 and G6·C17* base pairs. The observable NOEs between imino protons and nonexchangeable protons in the −0.6 to 2.8 ppm region are shown in Figure 3C. The NOEs p to s represent the connectivities between thymine imino protons and their own methyl group protons, while NOEs 1 to 5 (Figure 3B,C) represent intermolecular NOEs between 4-OHEN−C3 and DNA base protons. The NOEs include the 4-OHEN−C3 protons H6, H7, H15, and H16 and $-CH_3$ with G16(NH1). Additional intermolecular NOEs are observable in an expanded NOESY (200 ms mixing time) contour plot (Supporting Information Figure S3) that include connectivities between the C7-NH_2 and nonexchangeable C7-H5, H6 protons

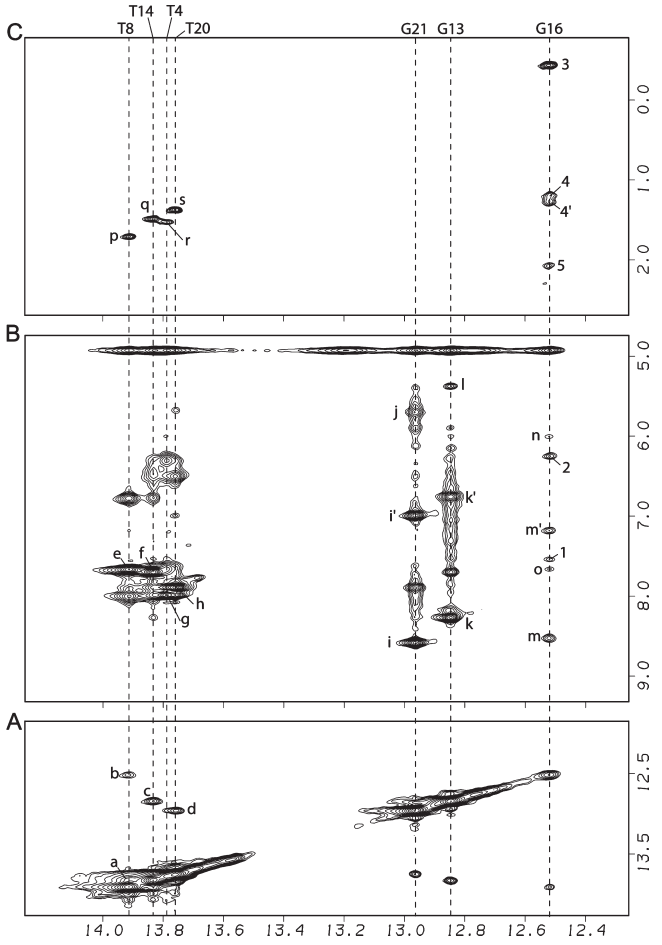

FIGURE 3: Expanded NOESY (200 ms mixing time) contour plots for the 4OHEN−C3 adduct in the G-[4OHEN]C-G sequence context at the 11mer duplex level. Spectra are in 100 mM NaCl, 10 mM phosphate, and H_2O, pH 6.8 at 7.5 °C. Panel A identifies NOEs between imino protons, with the cross-peaks labeled a to d, which are assigned as follows: a, T8(NH3)-T14(NH3); b, T8(NH3)-G16(NH1); c, T14(NH3)-G13(NH1); d, T20(NH3)-G21(NH1). Panels B and C identify NOEs between imino protons (12.0−14.0 ppm) and amino and nonexchangeable protons (4.8−9.0 ppm in (B) and −0.6 to 2.8 ppm in (C)), with NOE cross-peaks between DNA protons characteristic of Watson−Crick base pairs labeled e to o in (B) and p to s in (C), while NOE cross-peaks between carcinogen and DNA protons labeled 1 to 2 in (B) and 3 to 5 in (C). Cross-peaks e to s are assigned as follows: e, T8(NH3)-A15(H2); f, T14(NH3)-A9(H2); g, T4(NH3)-A19(H2); h, T20(NH3)-A3(H2); i, i′, G21(NH1)-C2(NH2); j, G21(NH1)-C2(H5); k, k′, G13(NH1)-C10(NH2); l, G13(NH1)-C10(H5); m, m′, G16(NH1)-C7(NH2); n, G16(NH1)-C7(H5); o, G16(NH1)-A15(H2); p, T8(NH3)-T8(CH_3); q, T14(NH3)-T14(CH_3); r, T4(NH3)-T4(CH_3); s, T20(NH3)-T20-(CH_3). Cross-peaks 1 to 5 are assigned as follows: 1, G16(NH1)-4OHEN-(H6); 2, G16(NH1)-4OHEN(H7); 3, G16(NH1)-4OHEN-(CH_3); 4, 4′, G16(NH1)-4OHEN(H15 × 2); 5, G16(NH1)-4OHEN(H16).

(6.6−8.7 ppm) with 4-OHEN−C3 protons H6 and H7 (7.53 and 6.23 ppm, respectively). A list of the identified equilenin residue−DNA intermolecular NOE cross-peaks for the 4-OHEN−C3 duplex is summarized in Table 1.

The upfield-shifted imino protons were identified as G6 in the case of the 4-OHEN−C4 duplex, and to G18 (11.8 ppm) and G6 (9.8 ppm) in the 4-OHEN−C3 duplex. Watson−Crick hydrogen bonding is not observed at the modified base pairs G6·C17*, since the cyclic 4-OHEN adducts obstruct the Watson−Crick hydrogen-bonding edge of the modified cytosine. These particular guanine imino proton resonances are broadened by rapid exchange with solvent, which is also the case for the terminal G12

and G22 imino protons. Furthermore, the significant upfield shifts of the G6 imino protons, but not the terminal G12 and G22 imino protons, suggest that upfield shifts of the G6 imino protons may be due to ring current effects that arise from stacking interaction with nearby 4-OHEN residues. The broadening of the imino proton resonances at base pair G18·C5 in 4-OHEN−C3 indicate enhanced exchange of the G18 imino proton with solvent water for this base pair. In contrast, the G18 imino proton resonance, on the 3′-side of the 4-OHEN−C4 adduct at C17*, is rather sharp and is less upfield-shifted. All other base pairs in this 11mer duplex have intact Watson−Crick pairing, characterized by the imino protons resonating between 12.5 and 14 ppm, and are indicated by the black dots in the numbered base pairs shown in Figure 2.

There are also differences in the expanded regions of the NOESY (200 ms mixing time) contour plots for the 4-OHEN−C4 duplex (Figure 4). NOE cross-peaks are observed between all thymine imino protons and those of their neighboring base pairs (cross-peaks a−f, Figure 4C). Further, the imino protons of all four thymine imino protons exhibit NOEs to adenine H2 protons (cross-peaks g−j in Figure 4B), and the nonterminal guanine imino protons (except G6) exhibit NOEs to the hydrogen-bonded cytosine amino or H5 protons (cross-peaks k−r, Figure 4B). These NOE patterns establish that all nonterminal A·T and all G·C base pairs, except for G6, adopt Watson−Crick alignment in the 4-OHEN−C3 duplex. The patterns of intermolecular NOE cross-peaks between the 4-OHEN residue and DNA base protons is also different in the 4-OHEN−C4 (cross-peaks 1−8, Figure 4A,B) from the 4-OHEN−C3 adduct (Figure 3A,B). These NOEs include the 4-OHEN−C4 protons H6, H7, H14, H15, and H16 with imino protons of G18 (cross-peaks 1−5) and protons H2, H3, and $-CH_3$ with imino protons of G16.

Additional intermolecular NOEs are observable in an expanded NOESY (200 ms mixing time) contour plot (Supporting Information Figure S4) that include connectivities between the C5-NH_2, C5(H5) protons with H6 of the equilenin residue, C7 (H5) with H1, and C7(−NH2, H5, H6) and G6(H8) with the methyl group protons of the 4-OHEN−C4 adduct. A complete list of the identified equilenin residue−DNA intermolecular NOE cross-peaks for the 4-OHEN−C4 duplex are summarized in Table 2.

(B) Nonexchangeable Protons. The expanded regions of the NOESY contour plots (200 ms mixing time) for the 4-OHEN−C3 and 4-OHEN−C4 duplexes in D_2O buffer, pH 6.8 at 7.5 and 15 °C, respectively, are plotted in Figure 5. The NOE connectivities between the base protons (purine H8 or pyrimidine H6, 6.6−8.6 ppm) and their own and 5′-flanking sugar H1′ protons (5.1−6.4 ppm) are traced as red lines for the unmodified strand from C1 to C11 and as blue lines for the modified strand from G12 to G22. In the case of the 4-OHEN−C3 duplex, the intramolecular NOEs between the adjacent H5 and H6 proton of the modified cytosine residue (4-OHEN−C17*) and the other cytosine are indicated by asterisks. In the case of the unmodified strand of the 4-OHEN−C3 duplex, clearly defined NOEs are observed from C2 to T4 and from T8 to C10. The C7-T8 cross-peak occurs in a crowded region of the spectrum, but is still discernible, while the C5-G6-C7 cross-peaks also occur in crowded regions of the 2D plot and are weak. These connectivities are consistent with a right-handed alignment for the C2−C5 and C7−C10 segments of the 4-OHEN−C3 duplex on the 5′- and 3′-sides of the 4-OHEN−C3 residue.

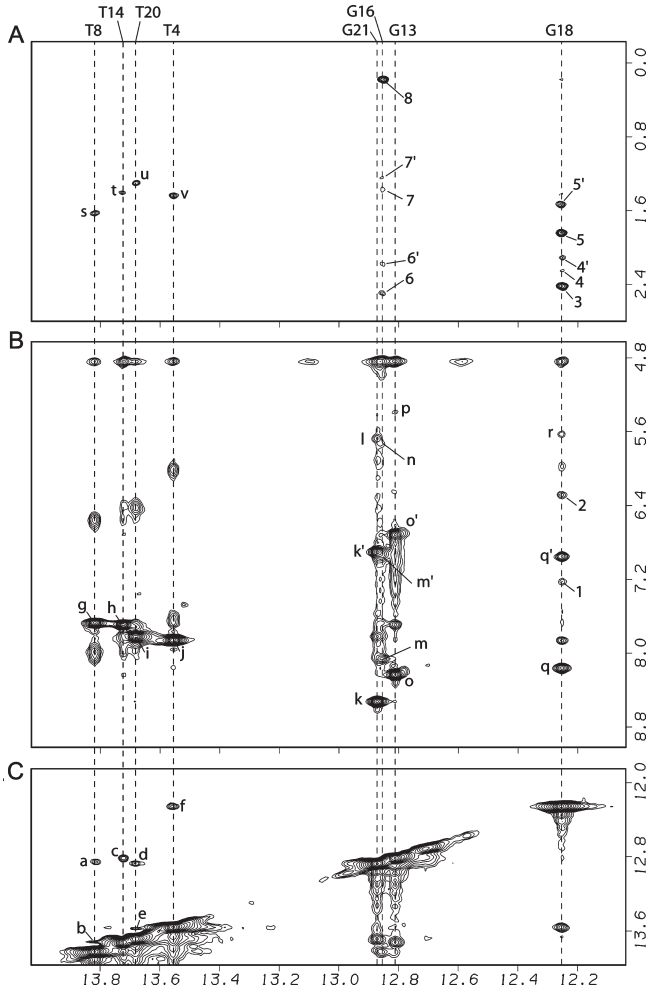

FIGURE 4: Expanded NOESY (200 ms mixing time) contour plots for the 4OHEN−C4 adduct in the G-[4OHEN]C-G sequence context at the 11mer duplex level. Spectra are in 100 mM NaCl, 10 mM phosphate, and H_2O, pH 6.8 at 15 °C. Panel C identifies NOEs between imino protons, with the cross-peaks labeled a to f, which are assigned as follows: a, T8(NH3)-G16(NH1); b, T8(NH3)-T14(NH3); c, T14(NH3)-G13(NH1); d, T20(NH3)-G21(NH1); e, T20(NH3)-T4(NH3); f, T4(NH3)-G18(NH1). Panels A and B identify NOEs between imino protons (12.0−14.0 ppm) and amino and nonexchangeable protons (4.8−8.8 ppm in (B) and 0.0−2.8 ppm in (A)), with NOE cross-peaks between DNA protons characteristic of Watson−Crick base pairs labeled g to r in (B) and s to v in (A), while NOE cross-peaks between carcinogen and DNA protons labeled 1 to 2 in (B) and 3 to 8 in (A). Cross-peaks g to v are assigned as follows: g, T8(NH3)-A15(H2); h, T14(NH3)-A9(H2); i, T20(NH3)-A3(H2); j, T4(NH3)-A19(H2); k, k′, G21(NH1)-C2(NH_2); l, G21(NH1)-C2(H5); m, m′, G16(NH1)-C7(NH_2); n, G16(NH1)-C7(H5); o, o′, G13(NH1)-C10(NH_2); p, G13(NH1)-C10(H5); q, q′, G18(NH1)-C5(NH_2); r, G18(NH1)-C5(H5); s, T8(NH3)-T8(CH_3); t, T14-(NH3)-T14(CH_3); u, T20(NH3)-T20(CH_3); v, T4(NH3)-T4(CH_3). Cross-peaks 1 to 8 are assigned as follows: 1, G18(NH1)-4OHEN-(H6); 2, G18(NH1)-4OHEN(H7); 3, G18(NH1)-4OHEN(H14); 4, 4′, G18-(NH1)-4OHEN(H16 × 2); 5, 5′, G18(NH1)-4OHEN(H15 × 2); 6, 6′, G16(NH1)-4OHEN(H11 × 2); 7, 7′, G16(NH1)-4OHEN(H12 × 2); 8, G16(NH1)-4OHEN(CH_3).

In the case of the 4-OHEN−C4 duplex, the spectral crowding in the central part of the duplex containing the 4-OHEN−C4 lesion (Figure 5B) is somewhat less pronounced than in the case of the 4-OHEN−C3 duplex. For the unmodified strand, the cross-peaks between base protons and the 5′-flanking H1′ protons can be traced within the C2−C5 and the C7−C10 segments. The cross-peaks in the central C5-G6-C7 region occur in overlapping spectral regions with weak, or nondiscernible,

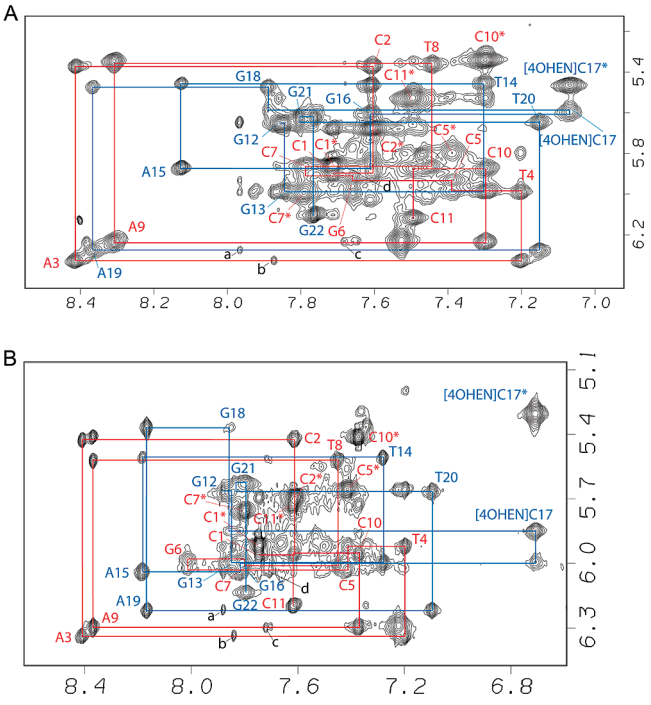

FIGURE 5: Expanded NOESY (200 ms mixing time) contour plots for the [4OHEN]−C adduct in the G-[4OHEN]C-G sequence context at the 11mer duplex level. The NOE connectivities between base protons and their own (indicated as solid triangles) and 5′-linked sugar H2′ as well as H2″ protons are traced as red lines for the lesion of the unmodified strand (C5 to C7) and as blue lines for the lesion of the modified strand (G16 to G18). The NOE connectivities between thymine base H6 protons and their 5′-linked base protons and thymine methyl protons are traced. (A) 4OHEN−C3. Spectra are in 100 mM NaCl, 10 mM phosphate, and D_2O, pH 6.8 at 7.5 °C. (B) 4OHEN−C4. Spectra are in 100 mM NaCl, 10 mM phosphate, and D_2O, pH 6.8 at 15 °C.

cross-peaks. In the case of the modified strand, nonterminal base pair cross-peaks are clearly identifiable in the G22−G18 segment and are observable in the 4-OHEN−C17−G13 segment, except those resonances involving G12, G13, and G16 that are positioned in spectrally overlapping regions and are not well resolved.

The *anti* or *syn* orientations about the 4-OHEN−C17 glycosidic bond in the 4-OHEN−C3 and 4-OHEN−C4 11mer duplexes have been probed by recording the NOESY spectrum in D_2O at a short mixing time of 50 ms. In the damaged duplexes we do not detect the NOE cross-peak between the H6 and H1′ protons of the C17* (Supporting Information Figure S5) compared to the clearly detectable NOEs between the H6 and H5 of the C17* residue (designated by asterisks in Figure 5; the fixed interproton separation is 2.45 Å). This result establishes that the 4-OHEN−C3 and 4-OHEN−C4 lesions in the 4-OHEN−C3 and 4-OHEN−C4 duplexes adopt *anti* glycosidic torsion bond conformations. The base to sugar H1′ NOE would be comparable in intensity to the cytidine H6−H5 NOE if the glycosidic torsion bond were in the *syn* conformation (see, for example, refs *33*, *36*, and *37*). These observations indicate that the modified residue adopts the *anti* conformation in both the 4-OHEN−C3 and 4-OHEN−C4 adduct containing duplexes.

(C) NOEs Observed at the Lesion Site. We have identified and assigned aliphatic and aromatic protons of the equilenin moiety in the 11mer 4-OHEN−C3 and 4-OHEN−C4 duplexes. NOE cross-peaks between the equilenin rings and the flanking base and amino protons are observed for both the 4-OHEN−C3 and 4-OHEN−C4 adducted duplexes (Tables 1 and 2). These intermolecular NOEs involve protons on both the modified and the unmodified strands. The set of intermolecular NOEs between the equilenin moiety and the DNA that we obtained clearly identified the positioning of the equilenin rings in each stereoisomer, notably the opposite orientations of the two stereoisomers with distal D-ring protruding into the minor groove.

In the 4-OHEN−C3 case, medium intensity NOEs are observed between the H6 and H7 of the equilenin aromatic B-ring and the imino proton of G16 and the amino proton of C7. These intermolecular NOEs between the adduct and 5′-flanking base pair G16·C7 reveal that the equilenin rings of 4-OHEN−C3 are oriented in the 5′-direction of the modified strand. Weak NOEs between the H6 and H7 of the equilenin aromatic B ring and H6/H5 of C7 also support this directionality. By contrast, in 4-OHEN−C4, medium intensity NOEs are observed between H6 and H7 of the equilenin moiety and the G18 imino proton as well as medium and weak NOEs between H6 and H7 and the C5 amino protons, respectively; these NOEs reveal that the equilenin rings of 4-OHEN−C4 are oriented in the 3′-direction along the modified strand. Weak NOEs between H6 and H7 of the equilenin moiety and H6/H5 of C5 also support the 3′-directionality. In addition, there are medium intensity NOEs between the methyl protons of the equilenin D-ring and the protons of the base G16 and its partner C7 in 4-OHEN−C4; however, in 4-OHEN−C3 the equilenin methyl protons show a medium intensity NOE only with the G16 imino proton. Thus, as a consequence of the opposite orientations, the methyl group is closer to the partner strand in 4-OHEN−C4, while in 4-OHEN−C3 the methyl group is closer to the damaged strand (Figure 7).

Restrained MD Simulation Models. (A) Duplex-Spanning anti Conformation with Opposite Orientations of 4-OHEN−C3 and 4-OHEN−C4. We utilized the NOE distance restraints of Tables 1 and 2 to obtain structures for the 4-OHEN−C3 and 4-OHEN−C4 adducts that are consistent with the experimental NOEs; this approach is particularly useful because the NOEs in the central, overlapping region of the duplex are weak or missing entirely. In modeling these structures, we utilized the *anti* conformations of the glycosidic linkage, since the NOESY spectra showed that the distances between the H6 proton of the modified cytosine and the H1′ of its own sugar are characteristic of the *anti* conformation (*33*). Initial models were constructed as described in the Methods section, and restrained molecular dynamics simulations were carried out with the AMBER 8 package as detailed in the Methods section.

Tables 1 and 2 list the interproton distances obtained within the last 500 ps of the 1 ns restrained simulations. We note that all restrained distances are within the bounds of the experimental NMR data. Figure 6 shows an ensemble of five structures selected from the last 500 ps of the trajectory, and Figure 7 and Supporting Information Figure S6 (stereoview) depict one of the structures selected from the ensemble.

While the 4-OHEN−C3 and 4-OHEN−C4 adducts adopt the *anti* glycosidic bond conformations, there are no hydrogen bonds between G6 and the modified cytosine since its Watson−Crick edge is obstructed. The modified cytosine is tilted with respect to the adjacent bases to accommodate the bulky equilenin moiety. The nonplanar D-ring with its bulky methyl group protrudes to the minor groove. The aromatic B-ring is shielded from the solvent because of its insertion into the duplex, and the two hydroxyl groups of the bridge ring are solvent-exposed on the major groove side. Thus, the equilenin moiety actually spans the duplex from the major to the minor groove side of the

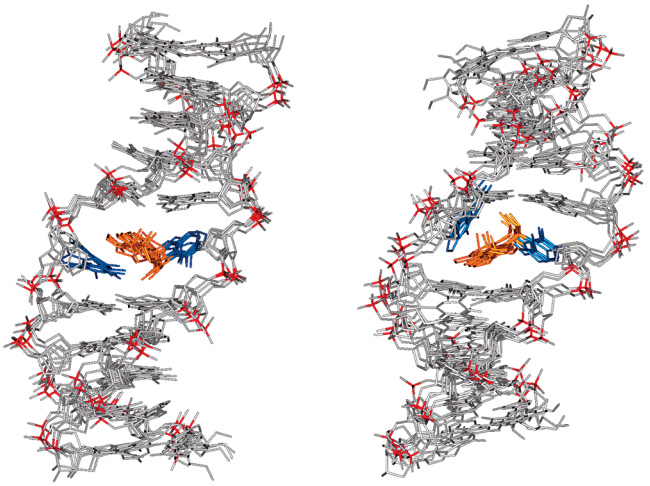

FIGURE 6: View into the minor groove of superpositioned NMR restrained structures of the 4-OHEN−C3 and 4-OHEN−C4 adducts at the 11mer duplex level. The structures shown are five representative structures, computed from MOIL-View (*29*), for each stereoisomer from the final 0.5 ns of restrained MD simulations. The equilenin rings are in orange, the modified base cytosine and its partner guanine are in blue, and the rest of the DNA is in gray except for the phosphorus atoms, which are colored red. Hydrogen atoms are not displayed.

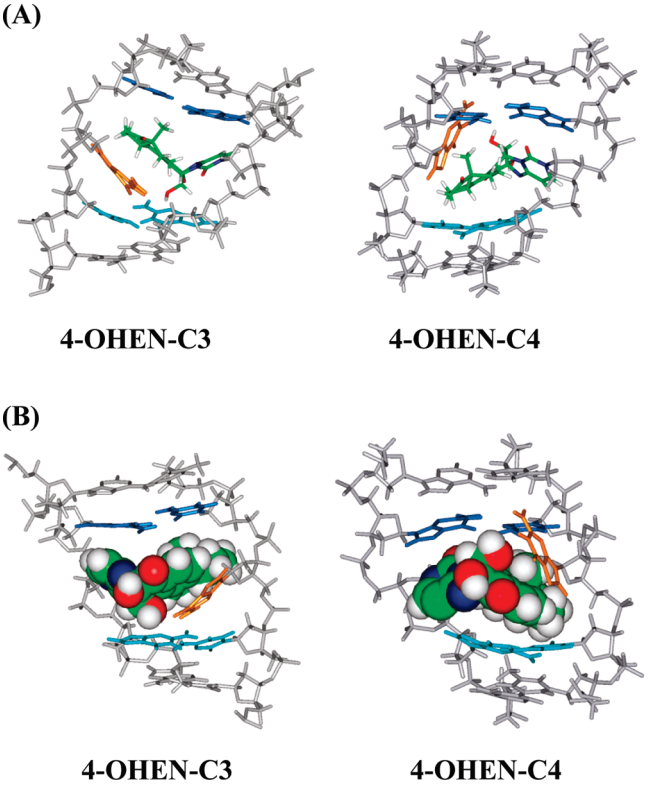

FIGURE 7: Central 5mers in the best representative structure (*29*) for the 4-OHEN−C modified duplexes in the NMR restrained simulations. The color code is as follows: 4-OHEN−C, colored by atom; C, green; O, red; N, blue; H white; partner base G, orange; 5′-side G16 C7 pair, light blue; 3′-side G18·C5 pair, cyan. (A) View into the minor groove; (B) view into the major groove. Stereoviews are given in Supporting Information. Closest distances between the equilenin methyl group proton and C7 O2: 4-OHEN−C3, 2.5 Å, and 4-OHEN−C4, 3.0 Å (see text).

double helix. The G6 partner base is quite mobile in the restrained simulations for both stereoisomers (Figure 6), which is consistent with the experimental NMR data. In the case of 4-OHEN−C4, only one weak NOE is observed between the methyl group of the equilenin and the H8 of G6, while in the case of 4-OHEN−C3 NOEs with G6 were not observed. The position of G6 in both cases is governed by the position of the equilenin methyl group; in 4-OHEN−C3 this methyl group is directed toward the modified strand and away from G6, while in 4-OHEN−C4 it is directed toward the partner strand and contacts G6 as manifested by the single NOE (Table 2 and Figure 7). Accordingly, in 4-OHEN−C4, G6 is severely tilted and unstacked with respect to the adjacent bases; this avoids collision with the methyl group on the minor groove side, and G6 therefore protrudes into the major groove. However, in 4-OHEN−C3, the G6 is modestly stacked with the equilenin aromatic B-ring, consistent with the upfield shift observed for the G18 imino proton (Figure 2); here the equilenin methyl group is on the equilenin face that does not interact with G6 (Figure 7). Stacking interactions between G6 and C5 are also substantial. In addition, the restrained MD simulations show that the equilenin C2′ OH forms hydrogen bonds that alternate between the amino and the imino group of the G6, which aids in anchoring this G6.

The stereoisomeric 4-OHEN−C3 and 4-OHEN−C4 adducts adopt opposite orientations along the modified strand. In 4-OHEN−C3 with *R* stereochemistry at the C3′ position, the equilenin rings are oriented toward the 5′-direction of the modified strand (Figure 7). In 4-OHEN−C4 with *S* stereochemistry at C3′, the equilenin rings are 3′-directed. At the same time, the modified cytosines are essentially perpendicular to the equilenin rings and are tilted oppositely, with 3′-directionality for 4-OHEN−C3 and 5′-directionality for 4-OHEN−C4. One face of the modified cytosine is largely exposed to the major groove. In both 4-OHEN−C3 and −C4, the equilenin C2′ OH group is directed toward the partner strand. These opposite orientations are manifested by the medium intensity NOEs between the equilenin H6 and H7 protons and the 5′-flanking base pair G16·C7 in 4-OHEN−C3, but in 4-OHEN−C4 the equilenin H6 and H7 protons showed medium NOEs with the 3′-flanking base pair G18·C5. Furthermore, the equilenin methyl group is in van der Waals contact with C7, as indicated by the large upfield shift for the methyl protons (Table 1) in 4-OHEN−C3; this is also consistent with the 5′-directionality of the equilenin ring system. By contrast, the methyl group in the case of the 4-OHEN−C4 adduct is significantly further removed from the G16·C7 base pair, and the upfield chemical shift for these methyl protons is therefore modest (Figure 7).

(B) Equilenin Direction-Governed Distortions in the 4-OHEN−C3 and 4-OHEN−C4 Duplexes. The insertion of the bulky 4-OHEN−C adducts into the DNA duplexes distorts the DNA structures in various ways, and these distortions reflect the directions adopted by the equilenin rings in each of the modified duplexes. Experimental thermal melting data for these adducts in the same DNA duplex 11mer sequence reveal that they each destabilize the duplex by about ~26 °C relative to the melting point of unmodified duplexes of 53 °C (*11*), consistent with their distorting effects.

Watson−Crick hydrogen bonding is maintained in all base pairs except at the lesion site for both 4-OHEN−C3 and −C4. However, the 3′-side G18· C5 base pair in 4-OHEN−C3 is more flexible as manifested by the line broadening at the G18 imino proton resonance (Figure 2). The time dependence of the

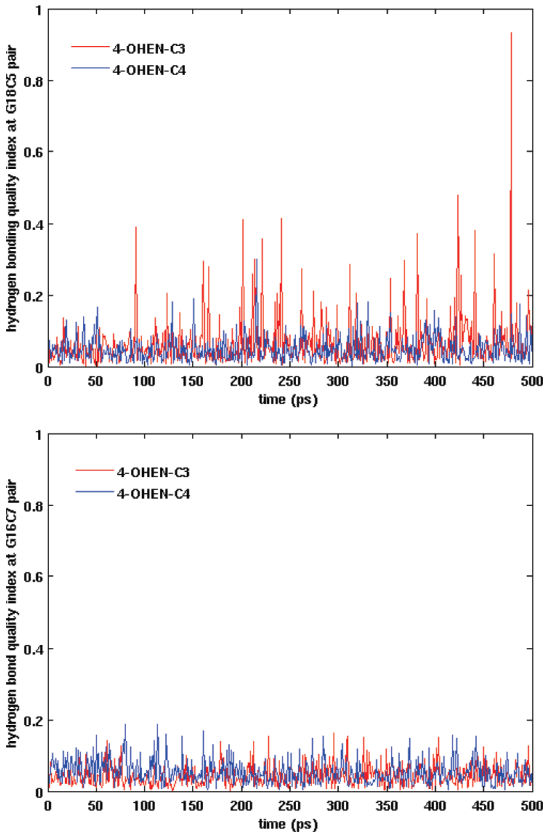

FIGURE 8: Hydrogen bond quality index versus time in NMR restrained simulations for 5′- and 3′-side base pairs adjacent to the lesion as defined in Figure 1.

calculated hydrogen bond quality index clearly reveals the flexibility (see Methods) of the conformation of this G18·C5 pair: it is apparent from Figure 8 that the G18·C5 pair in 4-OHEN−C3 is much more dynamic than the G16·C7 pair, as well as being more dynamic than either of these base pairs in 4-OHEN−C4. This figure clearly shows the large oscillations in the hydrogen bond quality index only in the case of the G18·C5 base pair for the 4-OHEN−C3 adduct (the more disturbed the hydrogen bond, the larger the hydrogen bond quality index).

Base−base stacking interactions are particularly impaired. We computed the van der Waals interaction energies between adjacent base pairs to estimate the degree of disturbance of the stacking interactions caused by the lesions. We found that the stacking interactions are greatly weakened in the vicinity of the lesion as compared to the unmodified control (by ~10 kcal/mol) for both stereoisomers. In addition, the perturbations in the stacking interactions are related to the orientations of the adducts (Figure 9). The greatest weakening in stacking involves the 5′-side base pair (G16·C7) of 4-OHEN−C3 and the 3′-side base pair (G18·C5) of 4-OHEN−C4.

Both 4-OHEN−C3 and 4-OHEN−C4 have enlarged major grooves as compared to the unmodified duplex (Figure 10). This enlargement results from the positioning of the two equilenin hydroxyl groups in the major groove. The orientations of the adducts also produce directional effects on the distortions of the grooves. For the 5′-directed 4-OHEN−C3, the major groove is more open on the 5′-side of the modified strand. However, for the 3′-directed 4-OHEN−C4, the major groove is more open on the 3′-side of the modified strand. In addition, in 4-OHEN−C4 the minor groove is also more open as compared to 4-OHEN−C3 and the unmodified duplex (Figure 10). The strong tilt of G6,

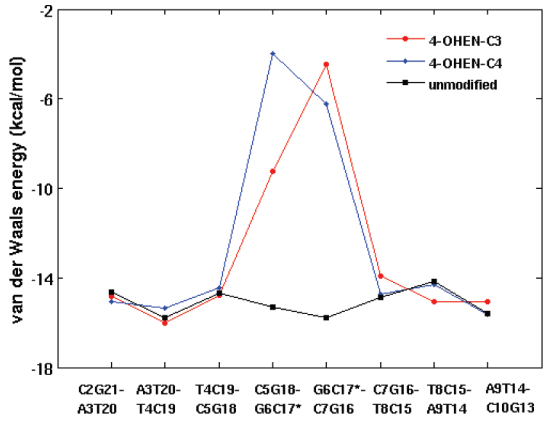

FIGURE 9: Trajectory-averaged van der Waals interaction energy (0.5−1 ns) vs base pair step plot in NMR restrained simulations. Lesion modification sites are designated with asterisks.

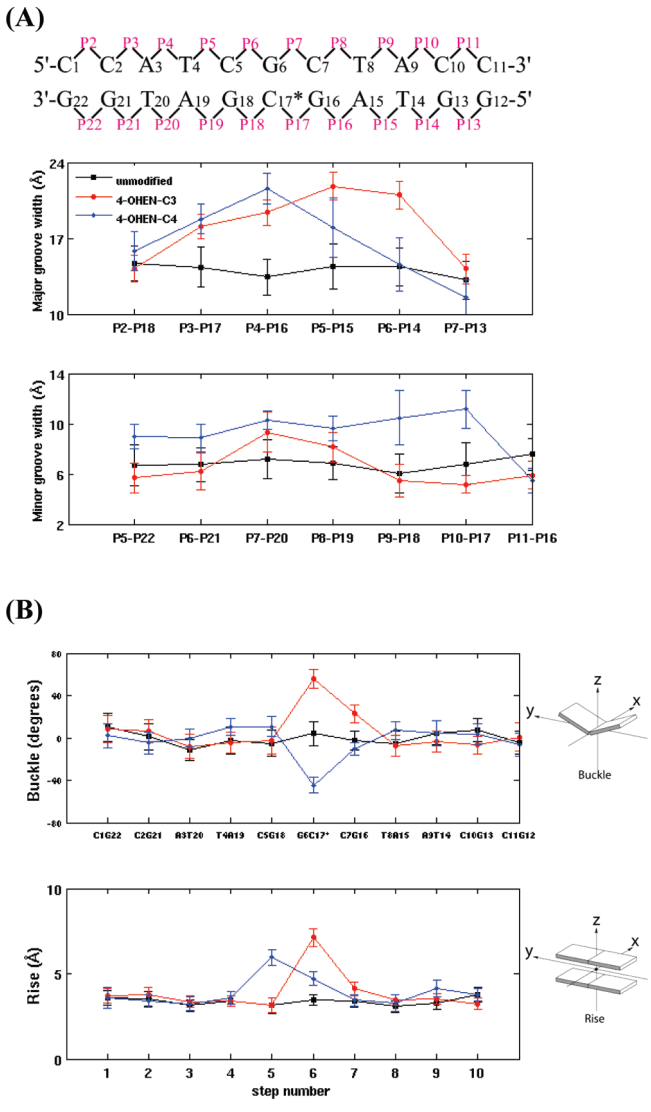

FIGURE 10: (A) Trajectory-averaged groove dimensions of the duplexes. Error bars indicate standard deviations. (B) Trajectory-averaged helicoidal parameters buckle and rise in NMR restrained simulations. Illustrations of the parameter definitions are adapted from Lu and Olson (*40*). The numbering scheme for the nucleotide base pair steps is that C1·G22 to C2·G21 is step 1, C2·G21 to A3·T20 is step 2, and so on.

with its positioning on the major groove side, forces the 5′-side base pair G16·C7 toward the minor groove to avoid collision, thus causing the groove to open.

The modified base pair is greatly buckled as compared to other base pairs of the duplexes in both stereoisomeric adducts. The opposite orientation of the adducts is also manifested in the distortions of this helical parameter (Figure 10). Thus, the buckle parameter of the modified base pair is positive for 4-OHEN−C3 and negative for 4-OHEN−C4, as seen in the oppositely buckled structures (Figure 7).

The rise parameter between the modified base pair and adjacent base pairs is also greatly increased and governed by the directionalities of the equilenin rings (Figure 10). The enlargement in rise between the 5′-side G16·C7 and the modified base pair is greatest in the 5′-directed 4-OHEN−C3, while in the 3′-directed 4-OHEN−C4, rise between the 3′-side G18·C5 and the modified base pair is greatly increased. The directionally tilted insertion of the equilenin rings produces this directional rise increase. Other helicoidal parameters, twist, tilt, roll, and propeller, also show severe deformations from the unmodified control near the lesion in both stereoisomers (Supporting Information Figure S7).

*Free Energy Calculations for the 4-OHEN−C3 and 4-OHEN−C4 Adducts Show That anti Conformations Are Energetically Preferred in the GC*G Sequence Context*. In order to gain further insight into the preference for the *anti* conformation in the NMR-restrained structures in the GC*G sequence context, we performed de novo MD simulations with free energy calculations for the 4-OHEN−C3 and 4-OHEN−C4 adducts beginning with *syn* and *anti* initial models (see Methods).

Overall, the *anti* unrestrained models are very similar to the NMR-restrained structures. Supporting Information Figure S8 shows the most representative structures from the last 4 ns of the 10 ns simulations for both stereoisomers. Supporting Information Tables S1 and S2 show a comparison of the trajectory-averaged values (6−10 ns) for the NMR-derived NOEs in these unrestrained simulations. Interproton distances that provided NOEs are within or close to NOE range with very few exceptions.

In the *syn* structures, the equilenin rings are situated externally in the DNA major groove, with the equilenin moiety exposed to solvent. Supporting Information Figure S9 shows the most representative *syn* structures for both stereoisomers. Details are given in Supporting Information and Figure S10−S12. These major groove structures are high energy compared to the *anti*, duplex-spanning structures. Their relative free energies were calculated as described in Methods, and results show that the *anti* conformation is favored energetically for both adducts, by ∼10 kcal/mol in 4-OHEN−C3 and by ∼15 kcal/mol in 4-OHEN−C4 (Supporting Information Table S3). Thus, no substantial population of a *syn* conformation is indicated from the relative free energies. This is consistent with the NMR data in the same sequence context, which showed only a single *anti* conformation.

To gain additional understanding of the energetic preference for the *anti* conformation, we computed interaction energies (see Methods) between the adduct and the DNA for both *syn* and *anti* conformations. Our results show that the *anti* NMR-restrained structures, which are inserted through the helix, have much more favorable interactions between the adduct and the DNA, by ∼17−25 kcal/mol (Supporting Information Table S4), than the simulated *syn*/major groove conformers. Thus, these favorable adduct/DNA interaction energies play an important role in the preference for the *anti* conformations observed in the NMR-restrained structures.

Stereochemistry-Governed Orientations. Our earlier studies, beginning with quantum mechanical investigations with base adducts and subsequent MD simulations with DNA duplexes, showed opposite orientations in 4-OHEN stereoisomers when the C3′ atom had opposite *R* and *S* stereochemistries (*18*, *21*). For DNA duplexes in 4-OHEN−C3 and 4-OHEN−C4 with *anti*/minor groove conformations we had predicted that the equilenin rings would be directed 5′ in 4-OHEN−C3 and 3′ in 4-OHEN−C4. In the present study the identified NOEs between the equilenin moiety and the DNA together with the *anti* conformation of the glycosidic bond, determined from the long distance between the H6 proton of the modified cytosine and the H1′ of its own sugar (*33*), clearly revealed these structural features. In addition, they placed the equilenin rings so as to span through the helix, with the nonplanar D-ring containing the bulky methyl group protruding into the minor groove and the two hydroxyl groups of the bridge ring exposed to solvent on the major groove side in both stereoisomers. Evidence of aliphatic ring conformational flexibility is indicated from line broadening in the one-dimensional NMR spectra (the line widths of protons on the adduct C- and D-rings are broader than those of the B ring) and for the G6 partner to the lesion by the complete absence of NOEs between it and the lesion in 4-OHEN−C3 and only one weak NOE for 4-OHEN−C4. In addition, the imino proton line widths of the lesion-flanking Watson−Crick base pairs are more sensitive to temperature variation than other base pairs (Supporting Information Figure S13), reflecting dynamics of the DNA lesion site.

Base Sequence Context Affects the Conformational Preference of the 4-OHEN−C Adducts. In previous work we have computationally investigated the 4-OHEN−C3 and 4-OHEN−C4 stereoisomeric adducts in the same 11mer duplex but with the modification at the C7 position (Figure 1). Thus, the lesion was in the GC*T sequence context, while in the present work the sequence context is GC*G. Hence, this affords an opportunity for exploring the impact of the nearest neighbors on the adduct conformations. Our prior work showed *syn* conformers with the equilenin rings in the major groove and *anti* conformers where the equilenin rings are in the minor groove for these two stereoisomers. Overall, we found that the advantages for the *anti* conformers were their favorable interactions with the DNA duplexes with low solvent exposure, although they were more distorting. The *syn* conformers were less distorting but had the disadvantage of solvent exposure of the equilenin moiety and less stabilizing interactions with the DNA. The energy differences between the *syn* and *anti* conformations were modest, leading to the suggestion that sequence context and environmental conditions could influence the conformational preference. In the present studies of the effects of the GC*G sequence context the strong preference for the *anti* conformation exemplifies this effect. Specifically, a hallmark of the *anti* conformers is the stabilizing interactions between the adduct and the DNA residues that surround the lesion. These interactions are optimized in the GC*G sequence context where the equilenin actually protrudes through the duplex. With the methyl-containing D-ring on the minor groove side, the aromatic B-ring partly stacks within the duplex, and the bridge ring hydroxyl groups protrude from the duplex on the major groove side, in the case of both stereoisomeric adducts. However, in the GC*T sequence context,

the bulky methyl group of the thymine obstructs this insertion of the equilenin into the duplex for both stereoisomeric adducts, and the equilenin is positioned more predominantly in the minor groove (*19*).

While we were unable to fully characterize the structure of the 4-OHEN−C4 adduct in the GC*T sequence context by NMR because of significant local destabilization of the duplex, we were able to determine the 1D exchangeable proton NMR spectrum (Figure 2C). This spectrum indicates that the impact of the 4-OHEN−C4 adduct at C7* (Figure 2C) rather than at C17* (Figure 2B) in otherwise identical duplexes is entirely different. With the adduct at C7*, none of the imino proton resonances from G12 to G16, from the site of the lesion and its entire 3′-side, are discernible. This suggests that these imino proton resonances are significantly broadened by rapid exchange with solvent, thus suggesting a significant destabilization of the duplex on the 3′-side of C7* in the GC*T sequence context. This observation is consistent with the 3′-orientation established for the identical 4-OHEN−C4 adduct positioned at C17* in the GC*G sequence context. Thus, subtle sequence context effects clearly do play a role in determining the lesion conformational characteristics and the local structural destabilization that it engenders.

CONCLUSION

We have utilized combined NMR and computational methods to elucidate the structures of the 4-OHEN−C3 and 4-OHEN−C4 adducts in 11mer DNA duplexes. The experimental NMR studies reveal the opposite orientations adopted by this stereoisomeric pair of cyclic 4-OHEN−C3 and 4-OHEN−C4 adducts derived from the reactions of 4-OHEN with cytosine in DNA. In both of these stereoisomeric adducts the damaged cytosine adopts the *anti* glycosidic bond conformation, and the equilenin rings span the width of the helix, with the nonplanar distal ring containing the bulky methyl group protruding into the minor groove and the two hydroxyl groups of the bridge ring exposed to solvent on the major groove side. In the case of 4-OHEN−C3, the equilenin rings are oriented toward the 5′-end of the damaged strand, while they are 3′-directed in the 4-OHEN−C4 adduct. While the double helices are significantly distorted by this invasive adduct conformation, stabilization is provided by favorable interactions between the estrogen ring system and DNA moieties. Notably, the distortions are propagated in the direction of the equilenin ring orientation in each case. As a consequence of these differences, it is plausible to assume that the processing of these two stereoisomeric adducts by cellular enzymes should be different as well. Indeed, differences in mutagenic activities (*15*, *16*, *22*, *38*, *39*) have been observed. Furthermore, the results of our studies indicate that base sequence context influences the extent of local destabilization of the DNA duplex and this effect may modulate the DNA repair of the 4-OHEN−C lesions.

SUPPORTING INFORMATION AVAILABLE

Details of MD simulation protocol, free energy analyses, and structural description of *syn*/major groove conformations from unrestrained MD simulations; Tables S1 and S2 give experimental intermolecular NOEs and trajectory-averaged distances from the unrestrained simulation of the 4-OHEN−C3 and 4-OHEN−C4 adducts; Table S3 gives MM-PBSA calculated free energies for the 4-OHEN−C3 and −C4 modified duplexes derived from the unrestrained simulations; Table S4 gives interaction energies between the adduct and the DNA for the 4-OHEN−C3 and −C4 modified duplexes; Figure S1 shows rmsd vs time plot for each MD simulation; Figure S2 shows the exchangeable proton 1D NMR spectra of the unmodified duplex; Figures S3 and S4 show expanded NOESY (200 ms mixing time) contour plots for 4-OHEN−C3 and −C4 in the G-[4OHEN]C-G sequence context at the 11mer duplex level; Figure S5 shows expanded NOESY (50 ms mixing time) stacked plots correlating NOEs between base and sugar H-1′ protons for 4-OHEN−C3 and −C4 in the G-[4OHEN]C-G sequence context at the 11mer duplex level; Figure S6 shows central 5mers in the representative structure for the 4-OHEN−C modified duplexes in the NMR restrained simulations, in stereoview; Figure S7 shows helicoidal parameters vs time for NMR restrained simulations; Figures S8 and S9 show central 5mers in the representative structures for the *anti*/minor groove and *syn*/major groove conformations of the unrestrained MD simulations; Figure S10 shows five superpositioned representative structures of the 4-OHEN−C4 adduct at the 11mer duplex level in the major groove conformation of the unrestrained simulation; Figure S11 shows the glycosidic torsion angle and hydrogen bond distance and angle in the unrestrained simulation for the 4-OHEN−C4 *syn*/major groove conformation; Figure S12 shows trajectory-averaged van der Waals interaction energies, groove dimensions, and helicoidal parameters for the *syn*/major groove conformation of the unrestrained simulation; Figure S13 shows the expanded imino proton region (7.0−15.0 ppm) of the 1D ^{1}H spectra of the [4OHEN]C adduct in the G-[4OHEN]C-G sequence context at the 11mer duplex level at different temperatures. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Hersh, A. L., Stefanick, M. L., and Stafford, R. S. (2004) National use of postmenopausal hormone therapy: annual trends and response to recent evidence. *JAMA, J. Am. Med. Assoc. 291*, 47–53.
2. Rossouw, J. E., Anderson, G. L., Prentice, R. L., LaCroix, A. Z., Kooperberg, C., Stefanick, M. L., Jackson, R. D., Beresford, S. A., Howard, B. V., Johnson, K. C., Kotchen, J. M., and Ockene, J. (2002) Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA, J. Am. Med. Assoc. 288*, 321–333.
3. Ravdin, P. M., Cronin, K. A., Howlader, N., Berg, C. D., Chlebowski, R. T., Feuer, E. J., Edwards, B. K., and Berry, D. A. (2007) The decrease in breast-cancer incidence in 2003 in the United States. *N. Engl. J. Med. 356*, 1670–1674.
4. Yager, J. D., and Davidson, N. E. (2006) Mechanisms of disease: Estrogen carcinogenesis in breast cancer. *N. Engl. J. Med. 354*, 270–282.
5. Okamoto, Y., Chou, P. H., Kim, S. Y., Suzuki, N., Laxmi, Y. R. S., Okamoto, K., Liu, X. P., Matsuda, T., and Shibutani, S. (2008) Oxidative DNA damage in Xpc-knockout and its wild mice treated with equine estrogen. *Chem. Res. Toxicol. 21*, 1120–1124.
6. Bolton, J. L. (2002) Quinoids, quinoid radicals, and phenoxyl radicals formed from estrogens and antiestrogens. *Toxicology 177*, 55–65.
7. Zhang, F., Chen, Y., Pisha, E., Shen, L., Xiong, Y., van Breemen, R. B., and Bolton, J. L. (1999) The major metabolite of equilin, 4-hydroxyequilin, autoxidizes to an o-quinone which isomerizes to the potent cytotoxin 4-hydroxyequilenin-o-quinone. *Chem. Res. Toxicol. 12*, 204–213.
8. Shen, L., Qiu, S. X., vanBreemen, R. B., Zhang, F. G., Chen, Y. M., and Bolton, J. L. (1997) Reaction of the Premarin metabolite 4-hydroxyequilenin semiquinone radical with 2'-deoxyguanosine: Formation of unusual cyclic adducts. *J. Am. Chem. Soc. 119*, 11126–11127.
9. Shen, L., Qiu, S., Chen, Y., Zhang, F., van Breemen, R. B., Nikolic, D., and Bolton, J. L. (1998) Alkylation of 2′-deoxynucleosides and DNA by the Premarin metabolite 4-hydroxyequilenin semiquinone radical. *Chem. Res. Toxicol. 11*, 94–101.
10. Embrechts, J., Lemiere, F., Van Dongen, W., and Esmans, E. L. (2001) Equilenin-2′-deoxynucleoside adducts: analysis with nano-liquid

chromatography coupled to nano-electrospray tandem mass spectrometry. *J. Mass Spectrom. 36*, 317–328.
11. Kolbanovskiy, A., Kuzmin, V., Shastry, A., Kolbanovskaya, M., Chen, D., Chang, M., Bolton, J. L., and Geacintov, N. E. (2005) Base selectivity and effects of sequence and DNA secondary structure on the formation of covalent adducts derived from the equine estrogen metabolite 4-hydroxyequilenin. *Chem. Res. Toxicol. 18*, 1737–1747.
12. Bolton, J. L., and Thatcher, G. R. (2008) Potential mechanisms of estrogen quinone carcinogenesis. *Chem. Res. Toxicol. 21*, 93–101.
13. Zhang, F., Swanson, S. M., van Breemen, R. B., Liu, X., Yang, Y., Gu, C., and Bolton, J. L. (2001) Equine estrogen metabolite 4-hydroxyequilenin induces DNA damage in the rat mammary tissues: formation of single-strand breaks, apurinic sites, stable adducts, and oxidized bases. *Chem. Res. Toxicol. 14*, 1654–1659.
14. Embrechts, J., Lemiere, F., Van Dongen, W., Esmans, E. L., Buytaert, P., Van Marck, E., Kockx, M., and Makar, A. (2003) Detection of estrogen DNA-adducts in human breast tumor tissue and healthy tissue by combined nano LC-nano ES tandem mass spectrometry. *J. Am. Soc. Mass Spectrom. 14*, 482–491.
15. Suzuki, N., Yasui, M., Santosh Laxmi, Y. R., Ohmori, H., Hanaoka, F., and Shibutani, S. (2004) Translesion synthesis past equine estrogen-derived 2′-deoxycytidine DNA adducts by human DNA polymerases eta and kappa. *Biochemistry 43*, 11312–11320.
16. Yasui, M., Laxmi, Y. R., Ananthoju, S. R., Suzuki, N., Kim, S. Y., and Shibutani, S. (2006) Translesion synthesis past equine estrogen-derived 2′-deoxyadenosine DNA adducts by human DNA polymerases eta and kappa. *Biochemistry 45*, 6187–6194.
17. Ding, S., Wang, Y., Kolbanovskiy, A., Durandin, A., Bolton, J. L., van Breemen, R. B., Broyde, S., and Geacintov, N. E. (2008) Determination of absolute configurations of 4-hydroxyequilenin-cytosine and -adenine adducts by optical rotatory dispersion, electronic circular dichroism, density functional theory calculations, and mass spectrometry. *Chem. Res. Toxicol. 21*, 1739–1748.
18. Ding, S., Shapiro, R., Geacintov, N. E., and Broyde, S. (2003) Conformations of stereoisomeric base adducts to 4-hydroxyequilenin. *Chem. Res. Toxicol. 16*, 695–707.
19. Ding, S., Shapiro, R., Geacintov, N. E., and Broyde, S. (2005) Equilenin-derived DNA adducts to cytosine in DNA duplexes: structures and thermodynamics. *Biochemistry 44*, 14565–14576.
20. Ding, S., Shapiro, R., Geacintov, N. E., and Broyde, S. (2007) 4-hydroxyequilenin-adenine lesions in DNA duplexes: stereochemistry, damage site, and structure. *Biochemistry 46*, 182–191.
21. Ding, S., Shapiro, R., Cai, Y., Geacintov, N. E., and Broyde, S. (2008) Conformational properties of equilenin-DNA adducts: stereoisomer and base effects. *Chem. Res. Toxicol. 21*, 1064–1073.
22. Chen, D. (2006) Nucleotide excision repair and translesion synthesis of DNA adducts derived from the equine estrogen metabolite 4-hydroxyequilenin, in Department of Chemistry, New York University, New York.
23. Chen, D. D., Oum, L., Kolbanovskiy, A., Kuzmin, V., Shastry, A., Chang, M. S., Bolton, J. L., and Geacintov, N. (2004) Translesion synthesis and nucleotide excision repair of site specifically modified oligodeoxyribonucleotides containing single lesions derived from the equine estrogen metabolite 4-OHEN. *Chem. Res. Toxicol. 17*, 1782–1782.
24. Cai, Y., Patel, D. J., Geacintov, N. E., and Broyde, S. (2009) Differential nucleotide excision repair susceptibility of bulky DNA adducts in different sequence contexts: hierarchies of recognition signals. *J. Mol. Biol. 385*, 30–44.
25. Geacintov, N. E., Broyde, S., Buterin, T., Naegeli, H., Wu, M., Yan, S. X., and Patel, D. J. (2002) Thermodynamic and structural factors in the removal of bulky DNA adducts by the nucleotide excision repair machinery. *Biopolymers 65*, 202–210.
26. Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Merz, K. M., Wang, B., Pearlman, D. A., Crowley, M., Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak, V., Cui, G., Beroza, P., Schafmeister, C., and Kollman, P. A. (2004) AMBER 8, University of California, San Francisco, CA.
27. Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., and Case, D. A. (1998) Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate−DNA helices. *J. Am. Chem. Soc. 120*, 9401–9409.
28. Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., and Cheatham, T. E.3rd (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res. 33*, 889–897.
29. Simmerling, C., Elber, R., and Zhang, J. (1995) MOIL-View—a program for visualization of structure and dynamics of biomolecules and STO—a program for computing stochastic paths, in Modeling of Biomolecular Structure and Mechanisms (Pullman, A., Ed.) Kluwer, Amsterdam, The Netherlands.
30. Hingerty, B. E., Figueroa, S., Hayden, T. L., and Broyde, S. (1989) Prediction of DNA structure from sequence: a build-up technique. *Biopolymers 28*, 1195–1222.
31. Ravishanker, G., Swaminathan, S., Beveridge, D. L., Lavery, R., and Sklenar, H. (1989) Conformational and helicoidal analysis of 30 ps of molecular dynamics on the d(CGCGAATTCGCG) double helix: "curves", dials and windows. *J. Biomol. Struct. Dyn. 6*, 669–699.
32. Ravishanker, G., Wang, W., and Beveridge, D. L. MD Toolchest, Wesleyan University, Middletown, CT.
33. Patel, D. J., Kozlowski, S. A., Nordheim, A., and Rich, A. (1982) Right-handed and left-handed DNA: studies of B- and Z-DNA by using proton nuclear Overhauser effect and P NMR. *Proc. Natl. Acad. Sci. U.S.A. 79*, 1413–1417.
34. Patel, D. J., Shapiro, L., and Hare, D. (1987) DNA and RNA: NMR studies of conformations and dynamics in solution. *Q. Rev. Biophys. 20*, 35–112.
35. Vandeven, F. J. M., and Hilbers, C. W. (1988) Nucleic-acids and nuclear magnetic-resonance. *Eur. J. Biochem. 178*, 1–38.
36. Kouchakdjian, M., Marinelli, E., Gao, X. L., Johnson, F., Grollman, A., and Patel, D. (1989) NMR studies of exocyclic 1,N^2-propanodeoxyguanosine adducts (X) opposite purines in DNA duplexes: protonated X(syn)·A(anti) pairing (acidic pH) and X(syn)·G(anti) pairing (neutral pH) at the lesion site. *Biochemistry 28*, 5647–5657.
37. Santos, C. D., Kouchakdjian, M., Yarema, K., Basu, A., Essigmann, J., and Patel, D. J. (1991) NMR studies of the exocyclic 1,N(6)-ethenodeoxyadenosine adduct (epsilon-dA) opposite deoxyguanosine in a DNA duplex: epsilon-dA(syn)·dG(anti) pairing at the lesion site. *Biochemistry 30*, 1828–1835.
38. Yasui, M., Matsui, S., Laxmi, Y. R., Suzuki, N., Kim, S. Y., Shibutani, S., and Matsuda, T. (2003) Mutagenic events induced by 4-hydroxyequilin in supF shuttle vector plasmid propagated in human cells. *Carcinogenesis 24*, 911–917.
39. Yasui, M., Suzuki, N., Liu, X., Okamoto, Y., Kim, S. Y., Laxmi, Y. R., and Shibutani, S. (2007) Mechanism of translesion synthesis past an equine estrogen-DNA adduct by Y-family DNA polymerases. *J. Mol. Biol. 371*, 1151–1162.
40. Lu, X. J., and Olson, W. K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res. 31*, 5108–5121.