Molecular Dynamics Simulations Provide a Structural Basis for the Experimentally Observed Nucleotide Preferences for DNA Interstrand Cross-Links Induced by Aziridinylbenzoquinones†

Ian S. Haworth,* Chong-Soon Lee, Masayuki Yuki, and Neil W. Gibson‡

School of Pharmacy and Comprehensive Cancer Center, University of Southern California, Los Angeles, California 90033

Received May 10, 1993; Revised Manuscript Received September 9, 1993®

ABSTRACT: Two-electron reduction of the structurally related aziridinylbenzoquinones DZQ and MeDZQ to their hydroquinone forms, DZHQ and MeDZHQ, respectively, generates species which interact and cross-link DNA at distinct nucleotide sequences. Within single target site duplex oligonucleotides, DZHQ was found to cross-link DNA at 5′-GC-3′ and 5′-GNNC-3′ sequences, whereas MeDZHQ was found to cross-link predominantly at 5′-GNC-3′ within a 5′-GTCA-3′ sequence. In a multitarget site duplex oligonucleotide, which contains either the target sequence 5′-TGCAC-3′ or 5′-TGCTC-3′, DZHQ was found to cross-link at both a 5′-GC-3′ (a 1,2 cross-link) and a 5′-GNNC-3′ (a 1,4 cross-link) site with approximately equal efficiency. Molecular dynamics simulations were able to accurately reproduce the experimental results and provide a structural basis for the alkylation preferences. Calculations were performed to determine the mobility of the hydroquinone species following guanine N7 (G) alkylation at 5′-TGCAC-3′ and 5′-TGTCA-3′ sequences. Conformations consistent with the formation of both 1,2 and 1,4 cross-links were observed when DZHQ was placed within a 5-TGCAC-3′ sequence. The 1,2 cross-link orientation was more stable and thermodynamically favored. For MeDZHQ at the same site the ligand was unable to form stable 1,2 or 1,4 cross-linking conformations, primarily due to clashes with thymine methyl groups. In contrast, the MeDZHQ monoadduct with a 5′-TGTCA-3′ sequence adopted a very stable conformation consistent with formation of a 1,3 cross-link.

Aziridinylbenzoquinones, such as DZQ[1] and MeDZQ, are thought to require enzymatic reduction in order to display antitumor activity (Siegel et al., 1990; Gibson et al., 1992). Enzymatic reduction of such quinones can take place via both one- and two-electron mechanisms (Siegel et al., 1990; Gibson et al., 1992). One-electron reduction of quinones generates semiquinone radicals which, dependant upon their redox potential, may interact with oxygen to produce superoxide and other reactive oxygen species (Gutierrez, 1989; Powis, 1987). Two-electron reduction converts the parent quinone to its hydroquinone derivative and in the case of antitumor quinones generates a reactive intermediate which is capable of alkylating and cross-linking DNA (Lee et al., 1992; Siegel et al., 1992). Previous work with DZQ and MeDZQ has shown that it is the formation of DNA interstrand cross-links and not DNA strand breaks which is the most toxic lesion to cells (Gibson et al., 1992). Ross and colleagues, however, have shown that although the initial rates of two-electron reduction of the aziridinylbenzoquinone, AZQ, under aerobic and hypoxic conditions are similar, the fate of the reduced metabolites can be considered different (Siegel et al., 1990; Gibson et al., 1992). We have identified an important difference in the fate of the reduced metabolites which can help explain the observed differences in cytotoxicity obtained between DZQ and MeDZQ (Gibson et al., 1992). That is the manner in which they alkylate and cross-link DNA. DZQ and MeDZQ reacted with guanines, as measured by Maxam and Gilbert sequencing (Hartley et al., 1991; Lee et al., 1992), with a sequence selectivity similar to that of the nitrogen mustard class of antitumor agent. Enzymatic reduction of DZQ and MeDZQ by DTD, however, was found to alter their sequence selective alkylation (Lee et al., 1992). Reduced DZQ (DZHQ) showed enhanced guanine alkylation in 5′-TGC-3′ sequences and new sites of adenine alkylation in 5′-(A/T)-AA-3′ sequences. Reduced MeDZQ (MeDZHQ) only showed new sites of adenine alkylation at 5′-(A/T)AA-3′ sequences but no enhancement of guanine alkylation (Lee et al., 1992). More recently, we have determined the sequence selectivity of DNA interstrand cross-linking in defined duplex oligonucleotides (Berardini et al., 1993). This work utilized "multitarget" duplex oligonucleotides, that is, oligonucleotides which contained two or more potential cross-linkable sites. DZHQ was found to cross-link DNA at 5′-GC-3′ and 5′-GNNC-3′ sequences, whereas MeDZHQ was found to cross-link DNA preferentially at 5′-GNC-3′ sequences (Berardini et al., 1993).

In this study we have compared a series of experimental observations obtained in both single-target and multitarget duplex oligonucleotides with those predicted by a computational approach within the same target sequence. Our data suggest that the nucleotide preferences obtained for cross-linking induced by DZHQ and MeDZHQ not only reflect their initial differences in DNA alkylation but also depend upon the ability of the alkylated monoadduct to achieve the correct orientation favorable for DNA interstrand cross-link formation. The theoretical basis for the observed experimental results is discussed.

MATERIALS AND METHODS

Chemicals and Reagents. DZQ and MeDZQ were synthesized according to published methods (Dzielendziak &

† This work was supported in part by Grant CA 51210 (N.W.G.) and by a NATO grant to N.W.G.

* To whom correspondence should be addressed.

‡ Current address: Cancer Research Group, Pfizer, Inc., Central Research Division, Eastern Point Road, Groton, CT 06340.

® Abstract published in *Advance ACS Abstracts,* November 1, 1993.

[1] Abbreviations: DZQ, 2,5-diaziridinyl-1,4-benzoquinone; MeDZQ 3,6-dimethyl-2,5-diaziridinyl-1,4-benzoquinone; NADH, nicotinamide adenine dinucleotide (reduced form); DTD, DT-diaphorase.

0006-2960/93/0432-12857$04.00/0

NAME	SEQUENCES (5' to 3')
A:	1 CTTCCAAGAT**GC**ATCAGATG GAAGGTTCTA**CG**TAGTCTAC 2
B:	1 CTTCCAAGAT**GT**CAACGATG GAAGGTTCTA**CA**GTTGCTAC 3
C:	1 CTTCCAAGAT**GTT**CACGATG GAAGGTTCTA**CAA**GTGCTAC 4
D:	1 CTTCCAAGAT**GCA**CACGATG GAAGGTTCTA**CGT**GTGCTAC 2 4
E:	1 CTTCCAAGAT**GCT**CACGATG GAAGGTTCTA**CGA**GTGCTAC 2 4

FIGURE 1: Sequences of the duplexes used in this study with the location of the guanines involved in DNA interstrand cross-linking being highlighted.

Butler, 1989; Dzielendziak et al., 1990). NADH (grade IV) was obtained from Sigma. Klenow fragment of DNA polymerase I and restriction endonucleases *Eco*RI, *Taq*I, and *Bam*HI were purchased from Bethesda Research Laboratories. [γ-^{32}P]ATP was purchased from NEN. All other reagents were at least of analytical grade.

Purification of Rat Hepatic DT-Diaphorase. Rat hepatic DTD was purified according to a previously published method (Hojeberg et al., 1981) from uninduced rats, yielding material with a specific activity of 660 nmol/(min·μgof protein).

Preparation of Oligonucleotides. A series of oligonucleotides (Figure 1) was synthesized on an automated DNA synthesizer (Applied Biosystem 391) and deprotected with saturated ammonium hydroxide at 55 °C overnight. After ethanol precipitation, DNA oligonucleotides were purified on a 20% denaturing polyacrylamide gel. These duplex oligonucleotides were designed to contain a centrally located target sequence. For clarity in comparing the experimental and computational studies we have numbered only the individual bases within the target sequence. Therefore, the major guanines which can participate in DNA interstrand cross-linking, within our defined target sites, are G^1 on the top strand and G^2 (duplexes A, D and E), G^3 (duplex B), and G^4 (duplexes C, D, and E) on the bottom strand. A cross-link which involves G^1–G^2 is defined as a 1,2 cross-link (5′-GC-3′), G^1–G^3 a 1,3 cross-link (5′-GNC-3′) and G^1–G^4 a 1,4 cross-link (5′-GNNC-3′).

Purification of DNA ISC. Approximately 10 μg of each oligonucleotide was 5′-end-labeled with T4 polynucleotide kinase and [γ-^{32}P]ATP. After removal of unincorporated [γ-^{32}P]ATP by ethanol precipitation, an equal amount of complementary strand was added, heated to 65 °C, and then cooled to room temperature overnight to form annealed duplex in 30 μL of distilled water. Drug (200 μM) was incubated at 37 °C for 10 min (DZQ) or 30 min (MeDZQ) in 50 μL of 10 mM potassium phosphate (pH 5.8) containing 100 μM NADH, 0.173 μg of DTD, and 1 mM EDTA. DNA interstrand cross-links induced by reduced quinones were purified by running a 20% denaturing polyacrylamide gel (mono to bis acrylamide ratio = 29:1, 8 M urea) until the xylene cyanol marker had migrated 10 cm. After the gel was exposed to X-ray film, each adduct was excised, soaked at 4 °C overnight, filtered through a Spin X centrifuge filter unit, and precipitated with ethanol and sodium acetate. The percentage of DNA ISC in each duplex oligonucleotide was determined by microdensitometry using an LKB ultra scan XL laser densitometer.

Determination of Cross-Linking Site. The purified DNA ISC was resuspended in freshly diluted 1 M piperidine and heated at 90 °C for 20 min to convert quantitatively sites of guanine N^7 alkylation into strand breaks (Mattes et al., 1986). Samples were lyophilized, resuspended in tracking dye containing 80% formamide and 1 mM EDTA, heated at 90 °C for 2 min, chilled on an ice bath, and then subjected to sequencing electrophoresis in parallel with Maxam and Gilbert base specific reactions (Lee & Gibson, 1993). Electrophoresis was achieved on 0.4 mm × 31 cm × 38.5 cm 8% polyacrylamide gels containing 8 M urea. Running time was approximately 3 h at 1500 V, 50 °C. Gels were transferred to filter paper, dried and then autoradiographed.

Computational Method. Molecular dynamics simulations were performed using the AMBER 4.0 force field (Pearlman et al., 1991; Weiner et al., 1986). Two DNA 11-mer helices, d($N_3TG^1CACM_3$)·d($M'_3G^4TG^2CAN'_3$) and d(N_3-TG^1TCAM_3)·d($M'_3TG^3ACAN'_3$) were used. These are henceforth referred to as TGCAC and TGTCA. In each case the monoadduct was formed by alkylation at G^1 in the above sequences. The guanines involved in such potential DNA interstrand cross-links on strand 2 are denoted by G^n. The first sequence has two such sites at G^4 (a 1,4 cross-link) and G^2 (a 1,2 cross-link) while the second has one site at G^3 (a 1,3 cross-link). The numbering of the relevant guanines is consistent with that described above. Other bases referred to in the description of the computational results are described in terms of their sequential relationship to the four guanine bases.

A detailed description of the derivation of the force field parameters for the free ligands and a validation of the model for noncovalent DNA/ligand complexes has been given elsewhere (Yuki & Haworth, 1993). A similar approach was used to obtain force field parameters and charges for the monoadduct complexes. Hence AM1 calculations (Dewar et al., 1985) using the MOPAC6.0 package were performed on the deoxyguanosine monoadducts with DZHQ and MeDZHQ. The nucleoside geometry was taken from AMBER and was not minimized further in the AM1 calculation. The nucleoside was terminated by 5′ and 3′ hydroxyl groups. Full energy refinement was performed on the hydroquinone, using default criteria for optimization [a gradient lower than 1.0 kcal/(mol·Å) and a change in ΔH_f of < 0.1 kcal/mol between successive cycles]. The unreacted aziridinyl group was in the ring-opened form in these calculations and an overall charge of +2 was assigned to the system. The molecular electrostatic potential (mep) of the monoadducts was derived from the AM1 wave function (Ferenczy et al., 1990) using the program RATTLER. The atomic point charge distributions were evaluated from the mep (Singh et al., 1984) using the same program.

Canonical B-DNA coordinates from the QUANTA 3.2.3 package were used to construct the DNA helices. The adduct was formed by overlaying the helix G^1 and the guanine base from the AM1-minimized guanosine–hydroquinone monoadduct. Subsequent rotation around the N7–CH_2 bond gave a

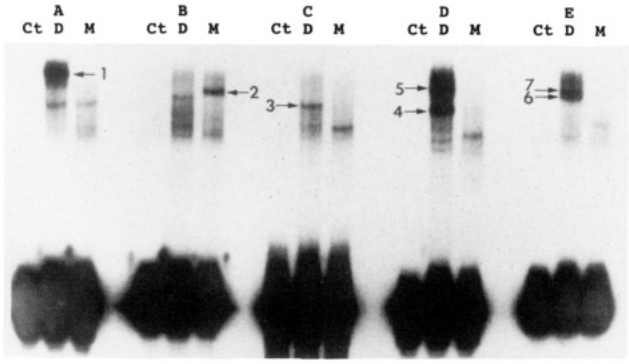

FIGURE 2: Autoradiography of a 20% denaturing PAGE showing DNA cross-links induced by DTD reduction of DZQ and MeDZQ in duplex oligonucleotides A, B, C, D, and E. All lanes show the data obtained when each duplex was 5′-end labeled on the lower strand. Lane Ct represents control non-drug-treated duplex oligonucleotide. Lanes D represent duplex incubated with 200 μM DZQ with 0.173 μg DTD and NADH (see Materials and Methods for details), and lanes M represent each duplex incubated with 200 μM MeDZQ with 0.173 μg of DTD and NADH. Cross-link bands which are numbered indicate those bands which provided unambiguous sequencing data (see Figure 3).

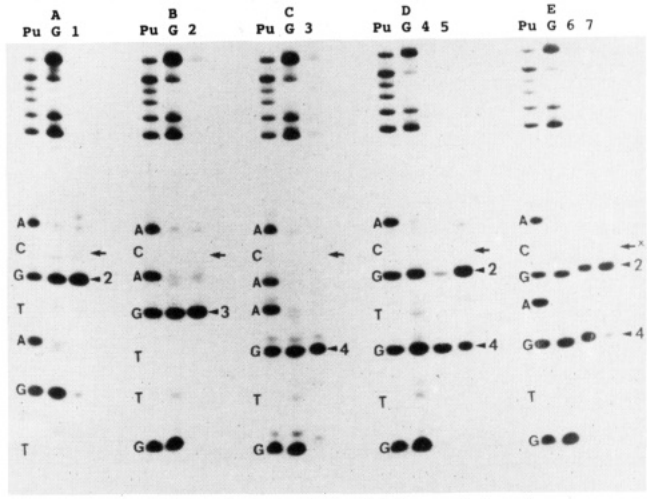

FIGURE 3: Autoradiogram of an 8% denaturing PAGE showing piperidine-induced DNA strand cleavage at the cross-linked sites. Pu represents the purine sequencing lane, and G represents the guanine sequencing lane. Each numbered lane represents the number of the cross-linked band which was isolated (see Figure 2). The sequence of the target site within each duplex is indicated, and the guanines involved in the cross-linking site; G^2, G^3, and G^4 as well as G^1 (arrow ^{x}) are indicated.

starting conformation in which the end carbon, C(az), of the unreacted, ring-opened aziridinyl group was located at a maximum distance from the helix with the "long" axis of the ligand approximately orthogonal to the helical axis. This initial conformation was unbiased toward either interstrand cross-linking or interaction of the aziridinyl group with the phosphodiester backbone. Simulations were performed for monoadducts of DZHQ and MeDZHQ at both TGCAC and TGTCA sites. Specific water molecules were not included, and a distance-dependent dielectric of $4r_{ij}$ was used to reproduce the effect of bulk solvent on the electrostatic interactions (Orozco et al., 1990). To maintain the helical conformation of the DNA, light positional restraints [2 kcal/(mol·Å)] were applied to all the helix atoms (but not to the ligand). An all atom representation of DNA was used with standard AMBER parameters and charges for all nucleotides except the deoxyguanosine of the monoadduct (see above).

The initial structures were relaxed using a 4000-step conjugate gradient energy minimization performed under conditions similar to those described above for the dynamics simulations. Each system was then subjected to 400 ps of molecular dynamics using a step length of 0.002 ps and a temperature of 298 K. This included a heating period of 4 ps to raise the temperature from 0 to 298 K. Following an initial decrease, the potential energy of the system remained essentially constant throughout each simulation. The coordinates of structures generated in the dynamics trajectory were saved every 0.4 ps, giving a total of 1000 configurations. The motion of the hydroquinone in the monoadduct was assessed in terms of the distance of the end carbon of the ring-opened aziridinyl group, C(az), to the guanine N^7 cross-linking target sites in each configuration. For the TGCAC site, the data were analyzed for both the 1,2 and 1,4 cross-linking sites. The internal energy of the DNA and the hydroquinone and a hydroquinone/DNA interaction energy were evaluated over each of the simulations.

RESULTS

In order to determine the sequence selectivity of DNA cross-linking at single-nucleotide resolution, we have used end-labeled duplex oligonucleotides which contain either a single target site (duplexes A, B, and C) or multitarget sites (duplexes D and E). Figure 2 shows a denaturing 20% acrylamide gel which allows the separation of cross-linked DNA from single strand DNA. We found numerous bands with a mobility consistent with cross-linked DNA (Figure 2). DZHQ was found to cross-link duplexes A, C, D, and E with one cross-link being observed in duplexes A and C and two potential cross-links being involved in duplexes D and E (Figure 2). In contrast, MeDZHQ was found to cross-link duplex B (Figure 2). The bands indicated 1–7 refer to those instances where unambiguous identification of the cross-link sites was achieved (Figure 3). Other bands and/or radioactive smears, not numerically designated, which migrated between our single-stranded oligonucleotide and cross-linked DNA were also analyzed. The sequencing patterns obtained from these bands were complex and ambiguous (data not shown). Due to such ambiguity the remainder of the experimental description will focus upon those bands where the identification of the nucleotides involved in DNA interstrand cross-linking was unambiguous.

Analysis of the duplex oligonucleotides, which had been end-labeled on the top strand by Maxam and Gilbert sequencing electrophoresis, implicated the guanine N^7 position of G^1 in all duplexes as being involved in the drug-induced DNA interstrand cross-link (data not shown). All sequencing data shown on Figure 3 refer to duplexes which had been labeled on the 5′-end of the bottom strand. In this case the major bands associated with piperidine cleavage were the guanines at either the G^2, G^3, or G^4 positions (Figure 3). In terms of the single target site duplexes, sequencing analysis confirmed that the DZHQ-induced cross-link in duplex A is a direct result of linkage between the two guanine N^7 positions in the 5′-GC-3′ sequence (lane 1, Figure 3). DZHQ was also found to induce a DNA interstrand cross-link in duplex C within a 5′-GNNC-3′ sequence (lane 3, Figure 3). As will be described below, such a result is important in understanding the data obtained with the multitarget duplex oligonucleotides, duplexes D and E. MeDZHQ, in contrast to DZHQ, induced a DNA interstrand cross-link in duplex B in a 5′-GNC-3′ sequence (lane 2, Figure 3). MeDZHQ was also found to result in the formation of two bands whose electrophoretic mobility was consistent with DNA interstrand cross-links (lane

M, duplex C and D, Figure 2). Sequencing analysis of such bands failed to identify any predominant guanines involved in such a putative cross-link (data not shown).

In terms of the multitarget duplex oligonucleotides, DZHQ was found to induce two major cross-link bands in duplexes D and E (lanes 4–7, Figure 3). These duplexes differ only in that the target contains a 5′-TGCAC-3′ (duplex D) in contrast to a 5′-TGCTC-3′ sequence (duplex E). Sequencing analysis of band 4 (5′-TGCAC-3′ target; Figure 2) when the bottom strand was labeled showed two bands corresponding to the guanines at the G^2 and G^4 positions (lane 4, Figure 3). This would suggest a linkage site at 5′-GC-3′ (G^1–G^2) sequence and possibly either a monofunctional adduct at G^4 or that the cross-link isolated from the gel contained two distinct cross-links, a 5′-GNNC-3′ (G^1–G^4) in addition to the 5′-GC-3′ cross-link. Evidence for the latter possibility was obtained when piperidine cleavage of the DZHQ cross-link identified by band 5 in Figure 2 showed a pattern of cleavage similar to that obtained with the band 4 cross-link [i.e., G^1 band on the upper strand and G^2 and G^4 bands on the lower strands (lane 5, Figure 3)]. The only difference between the two cross-link bands was that the ratio of the intensities of G^2 to G^4 was reversed. This may suggest that in addition to a 5′-GC-3′ cross-link (G^1–G^2), DZHQ is capable of inducing a 5′-GNNC-3′ cross-link (G^1–G^4) within this target sequence.

Results similar to those described above were obtained in the duplex containing the 5′-TGCAC-3′ target sequence (lanes 6 and 7, Figure 3). It is of interest to note the differences in mobilities between similar drug-induced DNA interstrand cross-links (compare mobility of bands 4 and 5 with bands 6 and 7, Figure 2). This may reflect sequence-dependent conformational changes. It is also likely in these experiments that we were unable to isolate two distinct DZHQ cross-link molecules with a 100% purity. The results obtained may therefore reflect the percentage of each cross-link molecule obtained in the excised band. Indeed, as will be presented below, molecular dynamics simulations support the concept that DZQ, upon enzymatic reduction to DZHQ, can induce both a 5′-GC-3′ and a 5′-GNNC-3′ cross-link. Further, direct experimental confirmation of the ability of DZHQ to induce a DNA interstrand cross-link within a 5′-GNNC-3′ sequence (duplex C) is described above (lane 3, Figure 3).

Computational Results. (a) DZHQ/TGCAC. Geometrical and energetic data taken from the molecular dynamics simulation of the DZHQ/TGCAC monoadduct are shown in Figure 4. Following an initial 15-ps period, the adduct essentially retains a conformation consistent with 1,2 cross-link formation for the next 200 ps of the trajectory until about 215 ps, as judged by the N7(G^2)–C(az) interatomic distance (Figure 4A). During this period the only significant deviation from this conformation occurs between 60 and 80 ps. Analysis of structures during this 20-ps period showed that the DZHQ adopts a conformations which would allow 1,4 cross-linking to occur (Figure 4B). At 215 ps a further conformational change takes place and several more 1,4 cross-link orientations occur (over a second 20-ps period). Following this, the hydroquinone does not revert to the 1,2 cross-link location but instead moves to a non-cross-linking position.

A good correlation is obtained between the conformational changes and the DZHQ/DNA interaction energy (Figure 4C). We use this component energy as a measure of the adduct stability, rather than a total energy that includes the DNA and DZHQ internal energy terms. The DNA internal energy is essentially constant over the whole simulation, due to the applied positional restraints (data not shown). We do

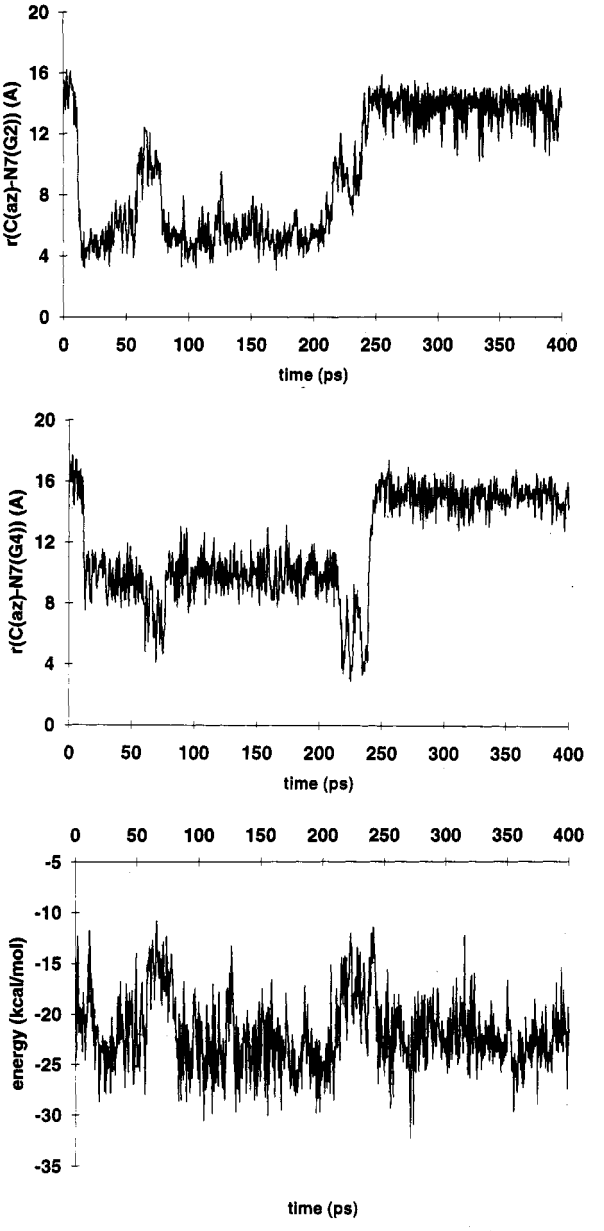

FIGURE 4: Conformational and energetic data taken from a 400-ps molecular dynamics simulation of the DZHQ monoadduct at a 5′-TG^{1}CAC-5′-G^{4}TG^{2}CA duplex. (A, top) Interatomic distance between the end carbon of the ring-opened aziridinyl group, C(az), and N^7(G^2), (B, middle) the C(az)–N^7(G^4) distance, and (C, bottom) the DZHQ/DNA interaction energy.

not observe a significant fluctuation of the hydroquinone internal energy, except between 240 and 250 ps (data not shown). This reflects a potential energy barrier due to distortion of the molecular structure as the drug moves out of the major groove and adopts the backbone-bound, non-cross-linking conformations observed after 250 ps. Conformations consistent with 1,2 cross-linking are more stable than those allowing formation of a 1,4 cross-link. This difference in interaction energy is estimated to be 8 kcal/mol. The 1,2 orientations are also about 2 kcal/mol more stable than the backbone-bound conformations that occur in the latter part of the trajectory. On the basis of these computed stabilities, we might expect primarily 1,2 cross-linking to occur in the DZHQ/TGCAC monoadduct. We recognize that the exclusion of specific water molecules from the simulation could introduce significant errors into these numbers. Furthermore, in contrast to this prediction, experimental analysis suggested that the 1,2 and 1,4 cross-links are formed with about equal

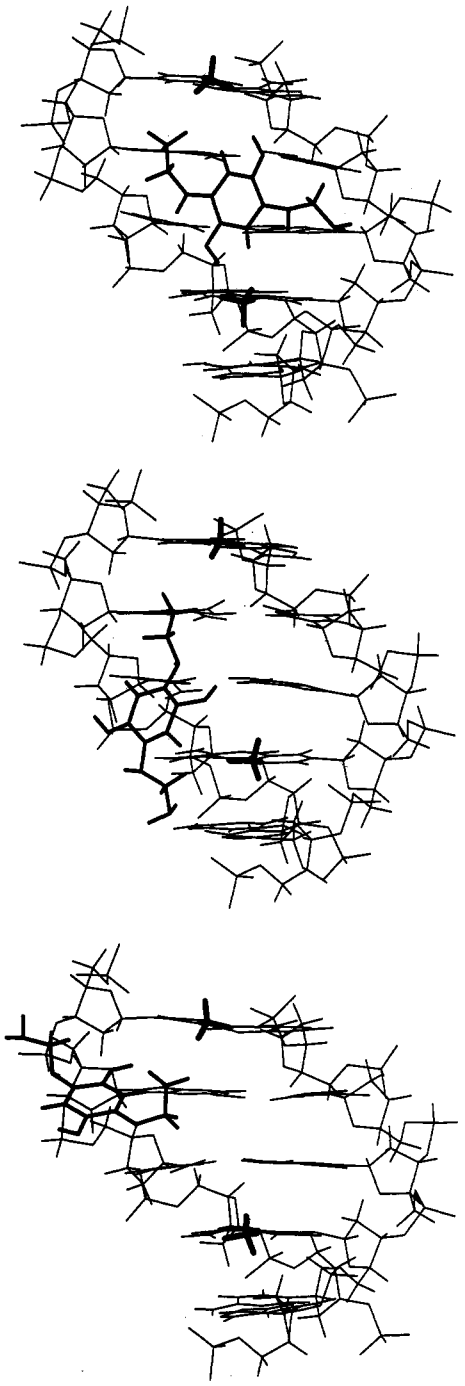

FIGURE 5: Representative structures taken from a 400-ps molecular dynamics simulation of the DZHQ monoadduct at a 5′-TG[1]CAC·5′-G[4]TG[2]CA duplex. In each case the monoadduct is viewed into the major groove with the thymine base to the 5′ side of the alkylated guanine base (G[1]) in the top left of the figure. The thymine methyl groups and DZHQ are shown in bold. The structures are representative conformations consistent with (A, top) 1,2-cross-link formation, taken after 120 ps of the trajectory, (B, middle) 1,4-cross-link formation (225 ps), and (C, bottom) interaction/reaction with the phosphodiester backbone (340 ps).

efficiency (Figure 2). We return to this apparent discrepancy between the theoretical and experimental results below.

Structures representative of those observed at different points in the trajectory are shown in Figure 5. In Figure 5A the 1,2 cross-link conformation shows how the DZHQ molecule is able to fit between the methyl groups of the flanking thymine bases to the 5′ side of G[1] and G[2]. We have previously shown that this is a favored orientation in the noncovalent DZHQ complex at this site (Yuki & Haworth, 1993). The methyl

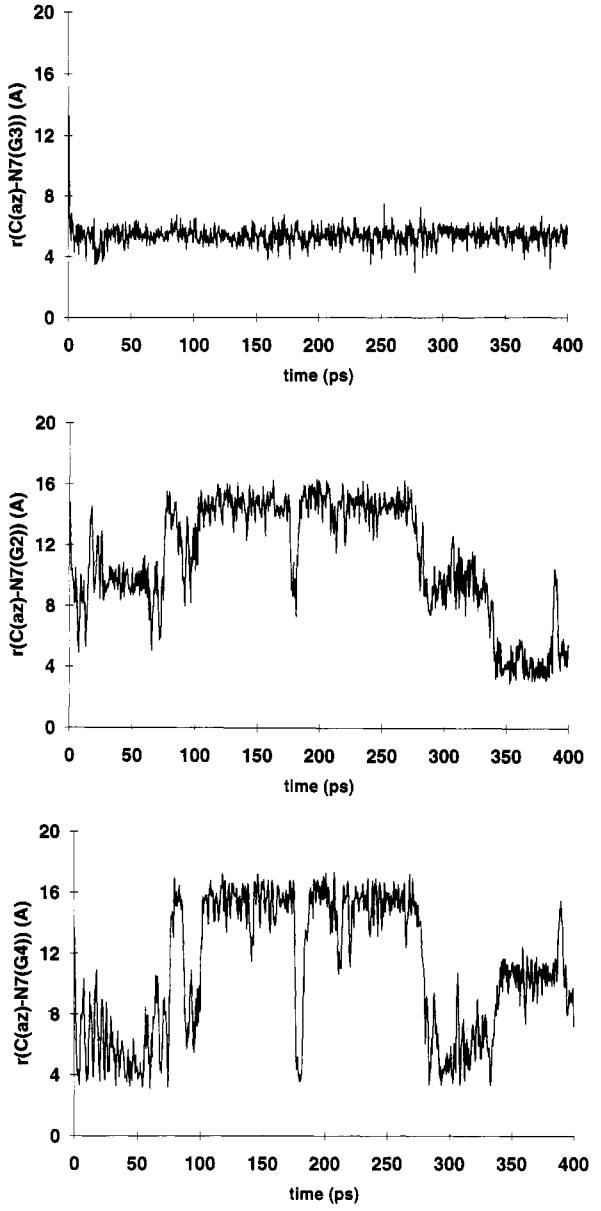

FIGURE 6: Conformational data taken from molecular dynamics simulations of MeDZHQ monoadducts with DNA. (A, top) Interatomic distance between the end carbon of the ring-opened aziridinyl group, C(az), and N7(G[3]) in the 5′-TG[1]TCA·5′-TG[3]ACA monoadduct, (B, middle) the C(az)-N[7](G[2]) distance in the 5′-TG[1]CAC·5′-G[4]TG[2]CA monoadduct, and (C, bottom) the C(az)-N[7](G[4]) distance from the same simulation.

group on strand 2 becomes a hindrance in the formation of the 1,4 cross-link conformation (Figure 5B), and it is remarkable that this orientation is observed in the trajectory and found to occur experimentally. Finally, in Figure 5C, the ligand has folded back to the phosphate backbone close to the initial alkylation site.

(b) MeDZHQ/TGTCA. Simulations performed on the MeDZHQ monoadducts provided a remarkable contrast between the conformational rigidity of the dimethylated hydroquinone at a TGTCA site and the relative conformationally flexibility at a TGCAC site (Figure 6). This reflects the greater stability of the MeDZHQ/TGTCA monoadduct, which retains a conformation consistent with 1,3 cross-link formation throughout the trajectory. The MeDZHQ/DNA interaction energy (data not shown) over the trajectory is about −28 kcal/mol, compared to an average of about −20 kcal/mol for multiple conformational forms of the MeDZHQ/

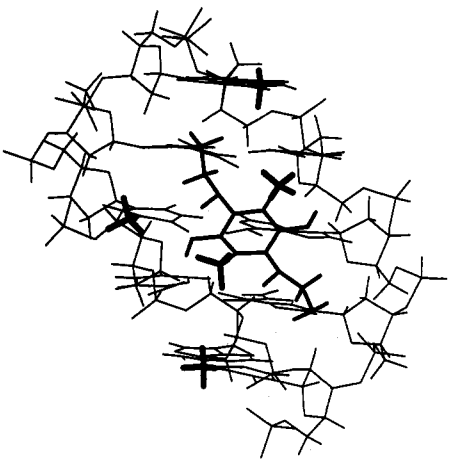

FIGURE 7: Representative structure taken from the molecular dynamics simulation of the MeDZHQ monoadduct with 5′-TG[1]-TCA·5′-TG[3]ACA, showing a 1,3-cross-linking conformation. Thymine methyl groups of MeDZHQ and the helix are shown in bold.

TGCAC monoadduct (see below). It is recalled that experimentally the formation of MeDZHQ cross-links is essentially restricted to 5′-GNC sites (Figure 2).

The MeDZHQ adduct with the TGTCA sequence immediately adopts a 1,3 cross-link orientation and remains in essentially the same orientation for the full 400 ps of the trajectory. As for DZHQ, the location of the thymine methyl groups plays a role in determining the alkylation characteristics of the molecule (Figure 7). In this case the three methyl groups (of thymines both 5′ and 3′ of G^1 and to the 5′ side of G^3) effectively guide MeDZHQ into the correct orientation for 1,3 cross-link formation and then prevent any movement away from this conformation. The C–C distance between ligand and helix methyl groups averages about 4–5 Å over the trajectory. We observe no tendency for 1,2 cross-linking at the N^7 atom of the adenine base on the 3′ side of G^3 (data not shown). Conformations consistent with this cross-link would require much closer approach of the ligand and helix methyl groups.

(c) MeDZHQ/TGCAC. We have shown experimentally (Lee et al., 1991) and theoretically (Yuki & Haworth, 1993) that formation of the MeDZHQ monoadduct at a 5′-TGCA site is unfavorable relative to other sites, although the discrimination is weak. It is apparent from Figure 6 that, in addition, the TGCAC monoadduct has difficulty in adopting cross-linking conformations, in good agreement with the experimental cross-linking preferences. From about 100 to 300 ps a non-cross-linking conformation is retained. In the periods before and after this (and briefly at 160 ps), potential 1,4 cross-linking conformations are formed, but these are transient structures. 1,2 cross-link conformations are completely absent until the final 50 ps of the trajectory, in which some stability is attained. However, the rapid conformational change at about 390 ps (Figure 6B) suggests that continuance of the trajectory would have resulted in further erratic motion.

The clash between the ligand and helix methyl groups is the most important factor preventing conformations consistent with 1,2 cross-linking, although a stable arrangement of methyl groups does ultimately occur and permit such conformations. Energetically, these 1,2-cross-link conformations are favored over those in the earlier part of the trajectory (data not shown). Methyl–methyl clashes do not seem to obviously detract from formation of the 1,4 cross-link. However, contrasting the 1,4 MeDZHQ/TGCAC monoadduct with the 1,3 MeDZHQ/TGTCA monoadduct, it became apparent that the ordered arrangement of ligand and helix methyl groups in the latter (Figure 7) is not present in the former conformation. Hence van der Waals (or hydrophobic) interactions between the methyl groups which drive the formation of the 1,3 cross-link are absent in the MeDZHQ/TGCAC monoadduct.

(d) DZHQ/TGTCA. For completeness, a simulation was also performed on the DZHQ monoadduct at a TGTCA site (data not shown). We consider this to be essentially hypothetical, since both experimental (Lee et al., 1992) and theoretical (Yuki & Haworth, 1993) work has shown formation of this monoadduct to be a relatively uncommon event. Over the trajectory, we observed no tendency for cross-linking to occur. The monoadduct essentially adopted a backbone oriented conformation similar to that shown in Figure 5C, although significant motion occurred about this average location. We performed a second simulation in which a starting monoadduct conformation suitable for 1,3 cross-linking was used. In this case, no major deviation from this conformation occurred and an $N^7(G^3)$–C(az) interatomic distance plot resembling that shown for MeDZHQ in Figure 6A resulted. Hence cross-link formation is highly dependent upon the initial orientation of the monoadduct, and the formation of 1,3 cross-links by DZHQ at the TGTCA site is unlikely to occur.

DISCUSSION

The molecular dynamics simulations and experimental results are in excellent agreement regarding the sequence selective DNA cross-linking exhibited by DZHQ and MeDZHQ. The structural data in this paper and the results obtained for the noncovalent complexes of DZHQ and MeDZHQ (Yuki & Haworth, 1993) at the TGCAC and TGTCA binding sites provide us with a molecular level description of the formation of the cross-linked diadducts. Broadly, this suggests that the DZHQ preference for 1,2 cross-linking is due to both a noncovalent and a cross-linking preference at this site. For MeDZHQ, the preference for the formation of 1,3 cross-links is more dependent on the behavior following monoalkylation, although there may also be a weak preference for the noncovalent interaction at the GNC site.

For DZHQ, our noncovalent binding data showed a significant preference for interaction at a GC and specifically a TGCA site, primarily due to the presence of the thymine methyl groups "guiding" the ligand into the GC step (Yuki & Haworth, 1993). These same methyl groups are also responsible for stabilizing the 1,2 cross-linking conformation of the monoadduct and hence the TGCA site is favorable for both the noncovalent and monoadduct species. Previous experimental and theoretical data provide a strong case for suggesting that 1,2 cross-link formation should be significantly favored over 1,4 cross-linking for DZHQ at the TGCAC site. Monoalkylation at both guanines of the 1,2 cross-link is equally favorable due to the palindromic nature of the sequence (5′-TGCA) and specifically is favored over alkylation of other guanine bases due to the 5′-TG dinucleotide steps (Lee et al., 1992; Yuki & Haworth, 1993). Furthermore, monoalkylation of G^4 within the 5′-TGCAC duplex is especially hindered by the 5′-GT sequence on strand 2 (Yuki & Haworth, 1993). Thus the formation of such a monoadduct capable of 1,4 cross-linking is significantly less likely than the similar formation of the G^2 monoadduct leading to 1,2 cross-link formation.

In the absence of contradictory experimental data, the molecular dynamics simulation of the DZHQ/TGCAC monoadduct presented in this work are supportive of a preference for 1,2 cross-link formation. Thermodynamically,

conformations consistent with 1,4 cross-linking are considerably less favorable than those leading to potential 1,2 cross-linking. However, experimentally there seems to be about equal proportions of the two types of cross-links formed. Since, as described above, we can reasonably discard the possibility of formation of the 1,4 cross-link from a strand 2 monoadduct as significantly affecting the 1,2:1,4 ratio, we are left with two explanations for the different experimental and theoretical results. These relate to either defects in the computational approach or inherent features of the cross-linking reaction.

First, the specific assumptions made in the calculations may be incorrect. Incorporation of explicit solvent molecules could change the relative conformational energies, making the 1,4 cross-link conformation more favorable. This does not necessarily seem likely, however, since the solvent-accessible surface of the monoadduct is not significantly changed between the 1,2 and 1,4 cross-link conformations. Another problem may be that the length of the simulation is insufficient to sample all conformations and that more energetically favorable 1,4 cross-linking orientations might be possible outside of this time frame. However, while 400 ps is a relatively short period, we believe that all the significant conformational space was covered in the simulation. A more serious problem could be the relatively restricted motion of the DNA helix. More flexibility of the helix might reasonably be expected to favor the 1,4 cross-linking conformation, since the N^7–N^7 distance between the G^1 and G^4 in an essentially canonical B-DNA is at the limit of the span of the bis-(aziridinyl)hydroquinone moiety.

The inherent nature of the cross-linking reaction may be such that only a transient formation of conformations leading to cross-linking is necessary. Hence the mere observance of conformations consistent with 1,4 cross-linking in the dynamics simulation of the DZHQ/TGCAC monoadduct may be of more significance than the thermodynamic comparison between the 1,2 and 1,4 cross-linking conformations. It is essential that the cross-link formation occurs relatively quickly following monoalkylation and aziridinyl ring opening, since the ring-opened species is highly reactive and particularly susceptible to hydration (Salvati et al., 1992). Hence transient occurrences of potential cross-linking conformations may be sufficient for cross-link formation.

For MeDZHQ, the preference for 1,3 over 1,2 or 1,4 cross-linking seems to be more dependent upon events occurring after the formation of the monoadduct. This provides a good rationale for the lower cross-linking efficiency of MeDZHQ compared to DZHQ (Lee et al., 1992). For DZHQ, monoalkylation occurs preferentially at sites which can go on to form cross-linked species efficiently. For MeDZHQ, monoadducts are formed with less sequence selectivity (Yuki & Haworth, 1993), but diadduct formation is significantly favored at only 5′-GNC sites. We should emphasize here that these arguments do not account for the significant monoalkylation at poly(purine) tracts, which cannot produce a cross-linked adduct and may be of more significance for MeDZHQ (Lee et al., 1992).

In contrasting the potential for MeDZHQ to form 1,2 and 1,3 cross-links, we have focused on the interaction (unfavorable and favorable respectively) with flanking thymine methyl groups. While this interaction seems to be of primary importance, there may also be a further basis for the selectivity for 1,3 over 1,2 cross-linking. In the 1,3 cross-linking conformation the ligand methyl groups occupy locations between base pairs, whereas in the 1,2 conformation the same groups are much closer to the base pairs. Hence the 1,3 selectivity may be a more general phenomena for this reason, although it seems likely that the presence of correctly positioned methyl groups enhances the effect. Several sequences could be used to examine this effect in more detail. A particularly interesting case would be the 5′-AGCCT sequence, with which both 1,2 and 1,3 cross-links could be formed. The flanking methyl groups do not interfere with the 1,2 conformation and do not direct the formation of the 1,3 conformation. Hence the significance of correctly located thymine bases could be evaluated by the observed ratio of 1,3:1,2 cross-links. This ratio should be small if the thymine methyl groups are essential for the 1,3 cross link formation and large if the more general location of the ligand methyl groups with respect to the base pairs is primarily responsible.

REFERENCES

Berardini, M. D., Lee, C.-S., Gibson, N. W., & Hartley, J. A. (1993) *Biochemistry 32*, 3306.

Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., & Stewart, J. J. P. (1985) *J. Am. Chem. Soc. 107*, 3902.

Dzielendziak, A., & Butler, J. (1989) *Synthesis*, 643.

Dzielendziak, A., Butler, J., Hoey, B. M., Lea, J. S., & Ward, T. H. (1990) *Cancer Res. 50*, 2003.

Ferenczy, G. G., Reynolds, C. A., & Richards, W. G. (1990) *J. Comput. Chem. 11*, 159.

Gibson, N. W., Hartley, J. A., Butler, J., Siegel, D., & Ross, D. (1992) *Mol. Pharmacol. 42*, 531.

Gutierrez, P. L. (1989) *Free Radical Biol. Med. 6*, 405.

Hartley, J. A., Berardini, M. D., Ponti, M., Gibson, N. W., Thompson, A. S., Thurston, D. E., Hoey, B. M., & Butler, J. (1991) *Biochemistry 30*, 11719.

Hojeberg, B., Blomberg, K., Stenberg, S., & Lind, C. (1981) *Arch. Biochem. Biophys. 207*, 205.

Lee, C.-S., & Gibson, N. W. (1993) *Biochemistry 32*, 2592.

Lee, C.-S., Hartley, J. A., Berardini, M., Butler, J., Siegel, D., Ross, D., & Gibson, N. W. (1992) *Biochemistry 31*, 3019.

Mattes, W. B., Hartley, J. A., & Kohn, K. W. (1986) *Biochim. Biophys. Acta 868*, 71.

Orozco, M., Laughton, C. A., Herzyk, P., & Neidle, S. (1990) *J. Biomol. Struct. Dyn. 8*, 359.

Pearlman, D. A., Case, D. A., Caldwell, J. C., Seibel, G. L., Singh, U. C., Weiner, P., & Kollman, P. A. (1991) AMBER4.0, University of California, San Francisco.

Powis, G. (1987) *Pharmacol. Ther. 35*, 57.

QUANTA version 3.2.3, Molecular Simulations Inc., 16 New England Executive Park, Burlington, MA 01803-5297.

RATTLER, a program for the computation of atomic point charges from molecular electrostatic potentials, Oxford Molecular. Ltd., Magdalene Science Park, Oxford, U.K.

Salvati, M. E., Moran, E. J., & Armstrong, R. W. (1992) *Tetrahedron Lett. 33*, 3711.

Siegel, D., Pacheco, D. Y., Gibson, N. W., & Ross, D. (1990) *Cancer Res. 50*, 7293.

Siegel, D., Senekowitsch, C., Beall, H., Kasai, M., Arai, H., Gibson, N. W., & Ross, D. (1992) *Biochemistry 31*, 7879.

Singh, U. C., & Kollman, P. A. (1984) *J. Comput. Chem. 5*, 139.

Weiner, S. J., Kollman, P. A., Nguyen, D. T., & Case, D. A. (1986) *J. Comput. Chem. 7*, 230.

Yuki, M., & Haworth, I. S. (1993) *Anti-Cancer Drug Des.* (in press).