2506–2517 Nucleic Acids Research, 2012, Vol. 40, No. 6 Published online 1 December 2011
doi:10.1093/nar/gkr1087

Structural, energetic and dynamic properties of guanine(C8)–thymine(N3) cross-links in DNA provide insights on susceptibility to nucleotide excision repair

Shuang Ding[1], Konstantin Kropachev[2], Yuqin Cai[1], Marina Kolbanovskiy[2], Svetlana A. Durandina[2], Zhi Liu[2], Vladimir Shafirovich[2], Suse Broyde[1,*] and Nicholas E. Geacintov[2,*]

[1]Department of Biology and [2]Department of Chemistry, New York University, 100 Washington Square East, New York, NY10003, USA

Received September 9, 2011; Revised October 31, 2011; Accepted November 2, 2011

ABSTRACT

The one-electron oxidation of guanine in DNA by carbonate radical anions, a decomposition product of peroxynitrosocarbonate which is associated with the inflammatory response, can lead to the formation of intrastrand cross-links between guanine and thymine bases [Crean *et al*. (Oxidation of single-stranded oligonucleotides by carbonate radical anions: generating intrastrand cross-links between guanine and thymine bases separated by cytosines. *Nucleic Acids Res.* 2008; 36: 742–755.)]. These involve covalent bonds between the C8 positions of guanine (G*) and N3 of thymine (T*) in 5′-d(...G*pT*...) and 5′-d(...G*pCpT*...) sequence contexts. We have performed nucleotide excision repair (NER) experiments in human HeLa cell extracts which show that the G*CT* intrastrand cross-link is excised with approximately four times greater efficiency than the G*T* cross-link embedded in 135-mer DNA duplexes. In addition, thermal melting studies reveal that both lesions significantly destabilize duplex DNA, and that the destabilization induced by the G*CT* cross-link is considerably greater. Consistent with this difference in NER, our computations show that both lesions dynamically distort and destabilize duplex DNA. They disturb Watson–Crick base-pairing and base-stacking interactions, and cause untwisting and minor groove opening. These structural perturbations are much more pronounced in the G*CT* than in the G*T* cross-link. Our combined experimental and computational studies provide structural and thermodynamic understanding of the features of the damaged duplexes that produce the most robust NER response.

INTRODUCTION

The inflammatory response plays an important role in the progression of human cancers (1,2). It produces oxidizing species that include carbonate radical anions derived from nitrosoperoxycarbonate, an unstable and reactive intermediate that stems from the combination of carbon dioxide and peroxynitrite (3). The primary target of oxidatively generated damage in DNA is guanine, the most easily oxidizable DNA base (4). It has been shown experimentally that photochemically generated carbonate radical anions (5–7) selectively oxidize guanine in DNA by a one-electron abstraction mechanism, producing novel intrastrand cross-links between guanine and thymine bases (5′-dG*pT* and 5′-dG*pCpT*, Figure 1) together with other known guanine oxidation products (8). The cyclic nature of the 5′-d(G*pCpT*) cross-link, in which the guanine C8-atom is bound to the thymine N3-atom (Figure 1), has been determined by LC-MS/MS and NMR studies (8). These cross-linked lesions are also found when native double-stranded DNA is exposed to peroxynitrite in aqueous carbon dioxide/bicarbonate solutions (9). In all cases, the yields are higher for the 5′-G*CT* than for the 5′-G*T* lesion. These G*(C8)-(N3)T* cross-links are formed via O_2-dependent nucleophilic reactions of guanine radicals with nearby thymine bases in DNA. They differ from previously discovered cross-links that are formed in deoxygenated environments, involving a covalent bond between C8 of

*To whom correspondence should be addressed. Tel: +1 212 998 8231; Fax: +1 212 995 4015; Email: broyde@nyu.edu
Correspondence may also be addressed to Nicholas E. Geacintov. Tel: +1 212 998 8407; Fax: +1 212 998 8421; Email: ng1@nyu.edu

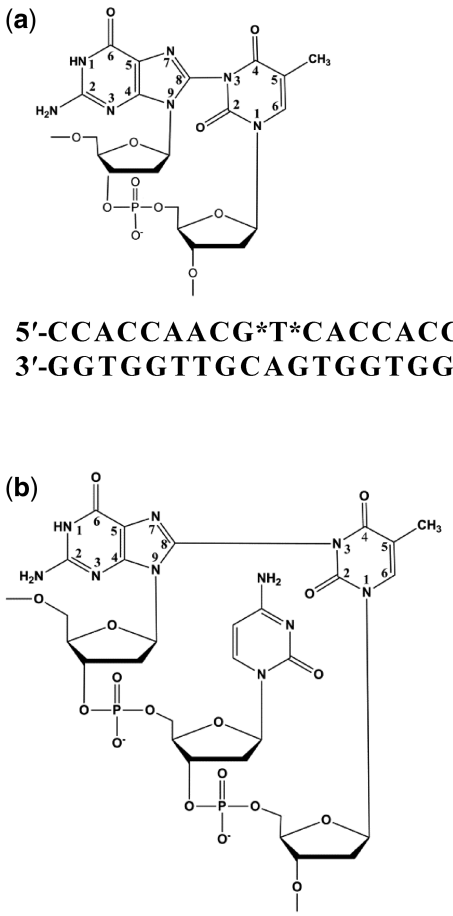

Figure 1. Chemical structures of the cross-links and the sequences of the B-DNA duplexes used in the experimental studies. The starred bases denote the modified sites. (**a**) G*T* cross-link; (**b**) G*CT* cross-link.

guanine and C5 of thymine, which have been investigated extensively (10–15).

However, there is no information available on the structural and biological impact of these highly constrained G(C8)-(N3)T lesions. Here, we demonstrate that both the G*CT* and G*T* lesions are substrates of the nucleotide excision repair (NER) system in human cell extracts, and that the excision of the G*CT* lesion is approximately four times more efficient than that of the G*T* cross-link. We have carried out an extensive molecular modeling study with MD simulations and free energy calculations, in order to obtain insight on the structural, energetic and dynamic properties of these two cross-linked DNA duplexes, and their relationship with the repair susceptibility and thermal melting properties of these lesions. Our results reveal that both cross-links distort the B-DNA duplex, but that the G*CT* cross-link is much more dynamic and distorting than the G*T* cross-link. These results are consistent with the observed greater thermal destabilization and NER susceptibility of the G*CT* cross-link.

MATERIALS AND METHODS

Modeling, MD simulations and free energy calculations

*G*T* cross-link*. Two initial models were built utilizing the following criteria: (i) the C8(G*) and N3(T*) had to be close enough to each other (less than ~5 Å) so that they would be capable of forming a covalent bond upon minimization of the energy with AMBER 9 (16) and (ii) there would be minimal collision between G*or T* and nearby DNA atoms. The search for such structures began with an unmodified B-DNA duplex 12-mer. We then rotated the glycosidic bond of G* continuously from 0° to 360°, seeking positions with minimal close contacts. From these positions, the T* glycosidic bond was then rotated over its entire 360° span without altering the B-DNA backbone. Once an optimal position for T* was found, the G* glycosidic bond was further adjusted to find its optimal position. This survey produced two initial models that met our criteria. Model 1 had both G* and T* in the *syn* glycosidic bond domain with C8(G*) and N3(T*) within the bonding range, ~1.4 Å. Model 2 had both G* and T* *anti* and C8(G*) and N3(T*) were 5.1 Å apart, without collision. These two initial models were then energy minimized to produce the starting structures for the 30 ns MD simulations. Both of these minimized starting structures have a proper covalent bond which constitutes the cross-link (about 1.4 Å). Supplementary Figure S1 shows these energy minimized models and Supplementary Table S1 gives their glycosidic torsion angles. The AMBER force field terms (17,18) that govern achievement of proper bond lengths, bond angles and torsion angles for chemically specific atom types is responsible for generating these chemically correct models for both the G*T* and the G*CT* cross-link, described below.

*G*CT* cross-link*. For the G*CT* cross-link, creating initial models was even more challenging because the G* and T* were far away from each other in the B-DNA structure. To ensure that all glycosidic bond domains for G* and T* would be considered, we created eight initial models which contained glycosidic torsion bond combinations for G* and T* of *anti*/*syn*, *anti*/*anti*, *syn*/*syn* and *syn*/*anti*. These were generated by beginning with *anti* structures with χ at around 250° (*anti*) or 60° (*syn*) and then rotating the two glycosidic bonds alternately to search for structures with G* and T* as close to each other as possible. Subsequently, the glycosidic torsion of the intervening C* was rotated to locate a position which avoided collision with the potential covalent bond between G* and T*. Two possible C* conformational regions were located for each G*/T* combination, yielding eight initial models. At this stage, the covalent bond between G* and T* was not yet formed in any of the initial models. The next stage was energy minimization which yielded covalently linked structures without any collisions. Supplementary Figure S2 shows these energy-minimized models and Supplementary Table S2 gives their glycosidic torsion angles. Subsequently, 30 ns of MD simulations were performed using the AMBER 9 simulation package (16). The Cornell *et al*. force field (18)

and the *parm99.dat* (17) parameter set modified by *parmbsc0* (19) were employed for all simulations.

For each model, the root-mean-square deviation (RMSD) of each snapshot in the trajectory relative to the first structure was plotted as a function of time and is shown in Supplementary Figure S3A for the G*T* and Supplementary Figure S3B for the G*CT* cross-link. For all cases, the MD achieved good stability after 15 ns, and we employed the structural ensembles from 15 to 30 ns time frame for further analyses. The sequences used for the modeling studies are the duplex 12-mer 5′-GTAGCG*T*TGGTG-3' for the G*T* cross-link and the duplex 13-mer 5′-CCAACG*CT*ACCAC-3′ for the G*CT* cross-link. Unmodified control 12- and 13-mer duplexes were also simulated for 30 ns for comparison with the damaged duplexes.

Molecular modeling was carried out with InsightII (Accelrys Software, Inc.), and MD simulations and free energy calculations using the MM-PBSA method (20) were carried out with the AMBER 9 package (16). The Ptraj, Anal and Canal modules of AMBER 7 (21), Curves+, and Toolchest (22) were used for structural analyses. The best representative structures in the MD trajectory were obtained by using the cluster analysis option in MOIL-View (23). The figures and movies were prepared with Pymol (24). Full details of the force field parameters (Supplementary Tables S3–S5) for the cross-links, as well as all other protocols, are given in Supplementary Data.

NER experiments

The 17-mer oligonucleotides containing the G*T* and G*CT* cross-link lesions in the sequence contexts 5′-CC ACCAACG*T*CACCACC-3′ and 5′-CCACCAACG*C TACCACC-3′, respectively, were synthesized as described in Crean *et al.* (8). The oligonucleotide containing the stereochemically defined 10*R* (+)-*cis*-*anti*-B[*a*]P-N^2-dG adduct (*cis*-B[*a*]P-N^2-dG), which was generated by reacting the racemic (±)-*anti*-7,8-dihydroxy-9,10-epoxy-7,8,9,10-tetrahydrobenzo[*a*]pyrene diol epoxide with the 11-mer sequence 5′-CCTACG*CTACC as described earlier (25), was utilized as a positive control of NER activity in the HeLa cell extract experiments. All oligonucleotides with either G*T* or G*CT* lesions were purified by reversed phase HPLC and by gel electrophoresis as described (8). The modified 11-mer oligonucleotides were ^{32}P-end labeled at the 5′ ends, and ligated into double-stranded 135-mer oligonucleotides duplexes (the details are described in Supplementary Data). In these duplexes, the internal ^{32}P label was placed at the 67th position from the 5′-ends of the modified 135-mer strands. The exact sequences of the full 135-mer duplexes are provided in Supplementary Data. The full details of the ligation methods and NER experiments in HeLa cell extracts are described in detail elsewhere (26). Briefly, the 135-mer duplexes were incubated in the cell extracts for varying amounts of time. Following incubation, the oligonucleotide excision products and intact DNA were desalted by precipitation with an aqueous 80% methanol solution and subjected to denaturing 12% acrylamide gel electrophoresis and analyzed by gel autoradiography using a Storm 840 phosphorimager.

RESULTS AND DISCUSSION

The G*T* and G*CT* cross-links are substrates of human NER: the G*CT* cross-link is a better substrate than G*T*

Typical results of NER dual incision assays are depicted in Figure 2a. The 135-mer duplexes were separately incubated in different aliquots of the same NER-active cell extracts for varying amounts of time. The lanes marked M define the mobilities of unmodified oligonucleotide markers of different lengths that are indicated by the numbers adjacent to the bands. The hallmark of successful NER activity in cell extracts is the appearance of ^{32}P-labeled oligonucleotides 24–32 nt in length containing the lesions. The *cis*-B[*a*]P-N^2-dG lesion embedded at the 67th nucleotide in the 135-mer duplexes is an excellent NER substrate (27,28) and was used as a positive control of NER activity. A series of bands are evident that co-migrate with the markers ~25–30-nt long, and their intensities increase as a function of incubation times in the range of 2–40 min (Figure 2a).

In the case of the G*CT* cross-linked adducts, a similar pattern of dual incision bands is evident but the dual excision products are characterized by mobilities expected for 23–27-mer oligonucleotides, as shown in the densitometry tracings of the 40-min lanes (Figure 2b). The amounts of dual excision products are significantly smaller in the case of the duplexes with the G*T* cross-linked lesions, although the apparent lengths of the excised oligonucleotides are closer to those observed for the *cis*-B[*a*]P-N^2-dG adducts. The apparent lengths of the dual excision products could be due to a faster mobility of the single-stranded oligonucleotides with G*CT* than with G*T* or *cis*-B[*a*]P-N^2-dG lesions, rather than to true differences in the sizes of the dual excision products. However, this hypothesis remains to be verified.

A quantitative analysis of the 40-min lanes (Figure 2b) indicates that efficiencies of dual incisions of duplexes with *cis*-B[*a*]P-N^2-dG and G*CT* are similar, while the efficiencies of dual incisions of the G*T*-containing duplexes are about four times smaller. The kinetics of dual incisions for the results shown in Figure 2a are depicted in Supplementary Figure S4; after a 30–40-min incubation time, the efficiency of dual incisions approaches ~16% in the case of G*CT* and ~4% in the case of the G*T* intrastrand lesion. Some cleavage at the site of the cross-linked lesion is evident as shown by the bands indicating 67-nt-long cleavage products; these are generated as a result of the incubation of the duplexes with the G*T* and G*CT* cross-linked lesions, but not with the *cis*-B[*a*]P-N^2-dG control samples. To further verify the difference in NER dual incision efficiencies of the G*CT* and G*T* intrastrand cross-linked lesions, the NER experiments with 135-mer duplexes containing either of these two lesions were determined in eight independent experiments using incubation times of 10, 20 and 30 min. The results are

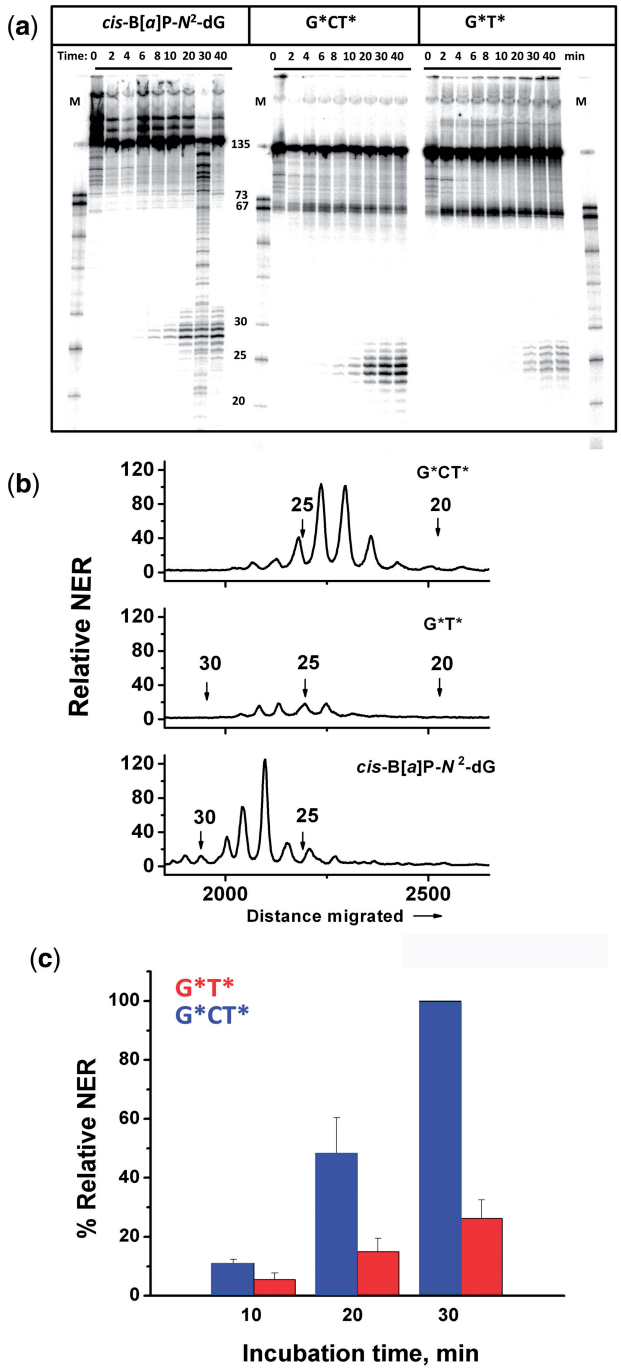

Figure 2. (**a**) Nucleotide excision repair in HeLa cell extracts. Denaturing gel showing the appearance of dual incision products elicited by the *cis*-B[*a*]P-N^2-dG (positive control) and the G*T* and G*CT* cross-link-containing duplexes as a function of incubation time in HeLa cell extracts. The lanes M represent oligonucleotide size markers. (**b**) Representative densitometry tracings of the 40-min lanes, adjusted for the total radioactivity in each lane to compensate for loading factors and differences in the radioactivity of the samples. (**c**) Incision kinetics of the internally labeled 135-mer modified duplexes in HeLa cell extracts. The incision efficiencies of the G*T* cross-link were normalized in each of the eight independent experiments to the value obtained with the G*CT* cross-link (relative value of 100) at the 30-min incubation time point. The averages and standard deviations shown were obtained from these normalized values.

summarized in Figure 2c. In each of these individual experiments, the NER efficiency of the G*CT* sample after an incubation time of 30 min was assigned a value of 100. The efficiencies of all other results in the same experiment were then calculated relative to this value. In this manner, variations in the NER activities of different cell extracts prepared at different times were taken into account as shown by the acceptable error bars in Figure 2c. Based on these results, it is concluded that the NER efficiency of the G*T* lesion is 3.8 ± 0.4 greater for the G*CT* than the G*T* lesion. A similar phenomenon is observed for *cis*-platinum cross-links: the G*G* cross-link is excised much less efficiently by the human NER system in cell extracts than the G*TG* lesion in double-stranded DNA (29).

Modeling, molecular dynamics and free energy calculations show that the G*CT* cross-link is more dynamically disturbed than the G*T* cross-link

We investigated the G*T* and G*CT*-cross-links in duplex 12- and 13-mers, respectively (see 'Materials and Methods'). First, molecular modeling was utilized to explore possible conformations that could form a chemically feasible covalent bond for each of the two cross-links. Those structures that appeared capable of forming the covalent bond due to the close proximity of the atoms that were to be linked, were then subject to energy minimization. This created the covalent cross-link in the selected starting structures (Supplementary Figures S1 and S2). Subsequently, 30 ns of MD simulations were performed. The results of these studies provided two possible structures for the G*T* cross-link (Models 1 and 2) and eight for the G*CT* cross-link (Models 1 through 8). We then computed free energies for each of these models, based on the ensembles derived from the MD simulations; this yielded relative free energy rankings for the different conformations obtained for each cross-link.

The G*T* cross-link: hydrogen bond disruption, disturbed stacking interactions, helix untwisting, minor groove opening and increased duplex dynamics

Two G*T* cross-linked structures containing chemically correct bonding between C8(G*) and N3(T*) were obtained (Supplementary Figure S5). Supplementary Figure S6 shows an example of the bond lengths and angles. The single bond that forms the cross-link has torsional flexibility. However, the flexibility is severely restrained by the interconnecting DNA backbone and by steric crowding from nearby bases. We computed the torsion angle C4(T*)–N3(T*)–C8(G*)–N7(G*) (Figure 1a) for Models 1 and 2 of the G*T* cross-link, and obtained ensemble average values of $58.1 \pm 12.0^\circ$ and $74.6 \pm 11.2^\circ$, respectively. The narrow range that these values can adopt (Supplementary Figure S5) reflects the constraint that the cross-link imposes on the relative orientations of the G* and T*. These two cross-linked structures have unusual glycosidic bond conformations: except for G* in Model 1 which is in the standard *syn* domain; the other glycosidic bonds (T* of Model 1 and

G* and T* of Model 2) are in the *syn*/*anti* interface region (30), as shown in Supplementary Table S1.

Our free energy calculations showed that Model 2 is the energetically preferred conformation for the G*T* cross-link, with a free energy stabilization of ~14 kcal/mol for this model relative to Model 1 (Supplementary Table S6). Although severely distorted in comparison to the unmodified control, the distortion is less than for Model 1, particularly in stacking and hydrogen bonding. Since it is much more energetically stable and relatively less distorted, we consider Model 2 (Figures 3a and b) to be a plausible conformation for the G*T* cross-link. The initial model for MD that provided this best representative structure (see Supplementary Data, Methods) is very similar (Figures 3c and d). Supplementary Table S7 summarizes structural properties of the higher energy Model 1, illustrated in Supplementary Figure S5. Model 2 is described below.

Watson–Crick pairing between G* and its partner C7 is maintained (occupancy of the hydrogen bonds ranges from 96% to 99%) (Figures 3a and b), but the pair is severely buckled (Figure 4). Furthermore, the duplex

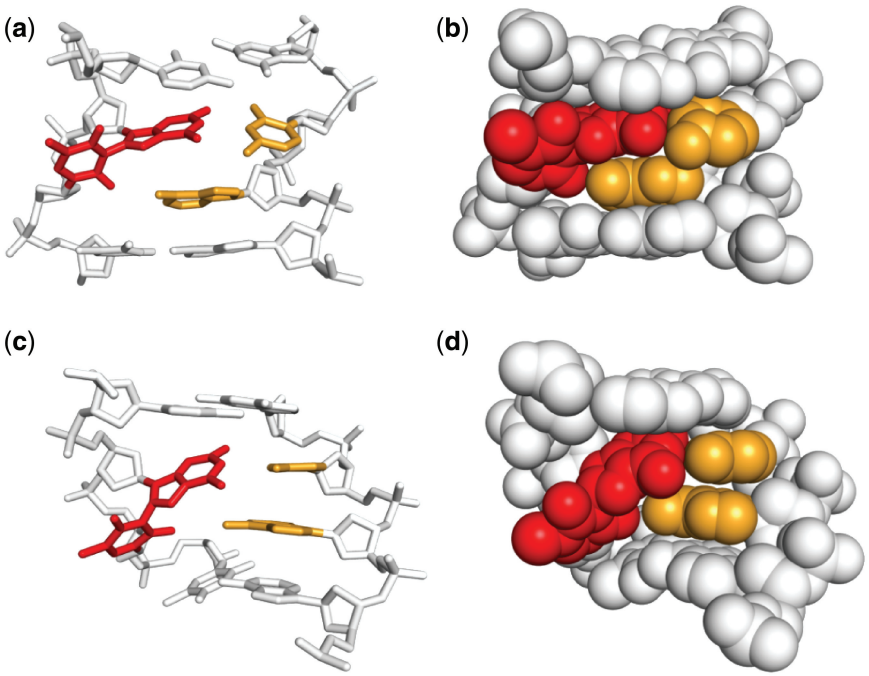

Figure 3. (**a**) Central tetramers of the best representative structure (23) for the G*T* cross-link modified duplexes. See Supplementary Movie S1. (**b**) Central tetramers in CPK view. The covalently linked model used to initiate the MD simulation that produced this structure (see 'Materials and Methods' section) is given in (**c**) and (**d**) with similar rendering as (a) and (b), respectively. Full duplexes are shown in Supplementary Figures S1 and S5. The color code is as follows: DNA, gray; G*T* cross-link bases, red; partner bases A6 and C7, bright orange.

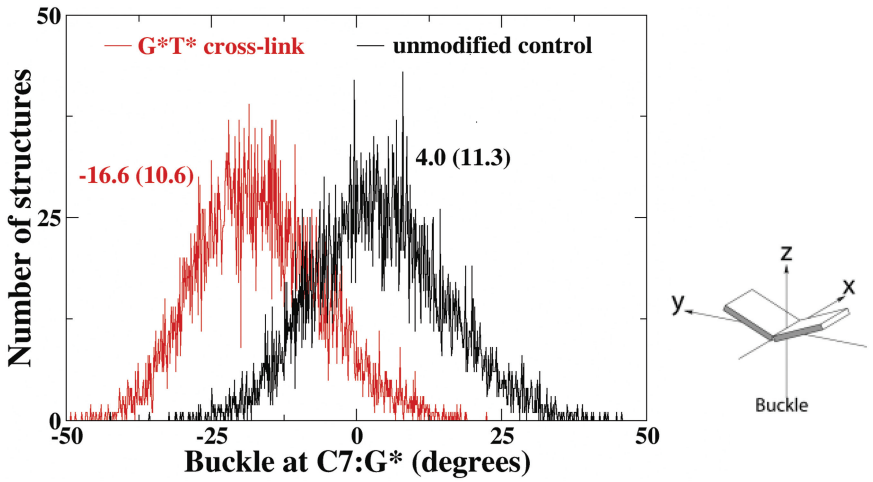

Figure 4. Population distribution of the Buckle in the G*T* cross-link (red) and the unmodified control (black). Ensemble average values and standard deviations (in parentheses) are given. The cartoon is reprinted with permission from Lu and Olson (49). Copyright 2003 Oxford University Press.

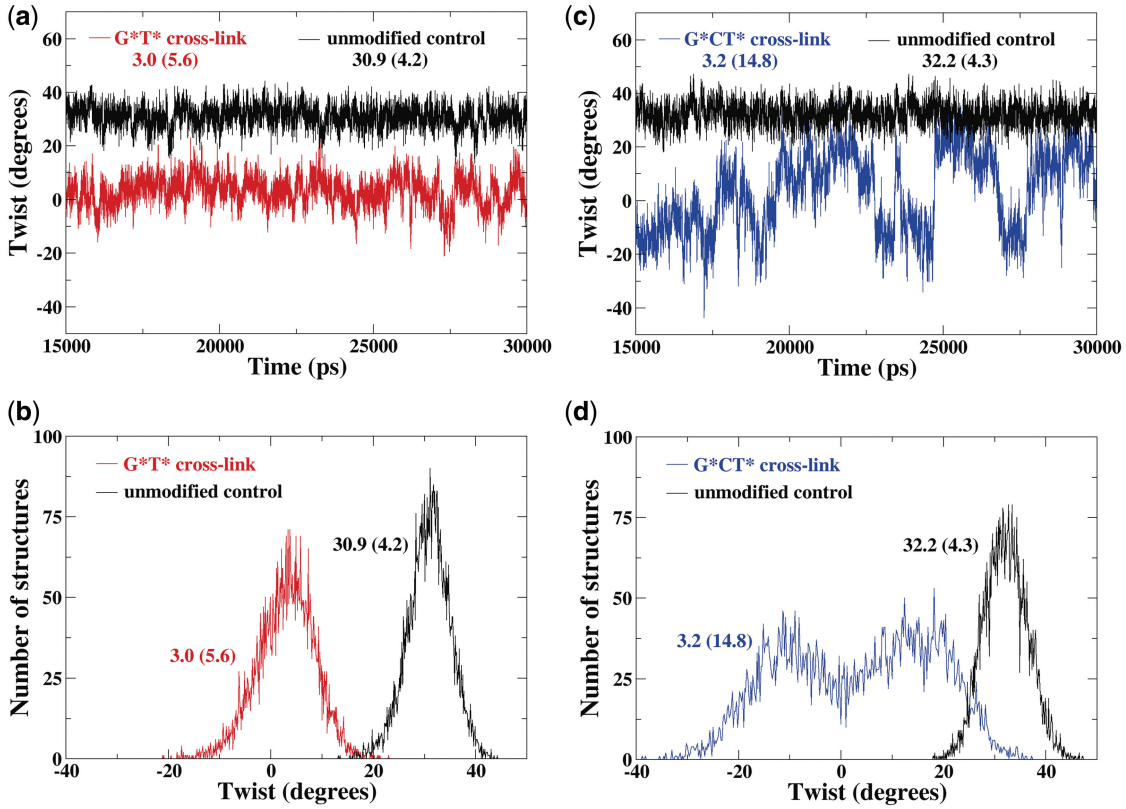

Figure 5. (**a**) Time-dependence and (**b**) Population distribution of the Twist angle in the G*T* cross-link (A6:T* to C7:G* step) (red) and the unmodified control (black). (**c**) Time-dependence and (**d**) population distribution of the Twist angle in the G*CT* (G*:C21 to C*:G20 step) cross-links (blue) and unmodified control (black). Ensemble average values and standard deviations (in parentheses) are given. These results illustrate the greater and more dynamic untwisting of the G*CT* cross-link, which displays a bimodal population distribution whose twist angle spans a much greater range than the G*T* cross-link.

Table 1. Dynamic distortions, melting temperatures and NER for the energetically preferred G*T* and G*CT* cross-link models in comparison with the unmodified control duplexes.[a]

	Number of Watson–Crick hydrogen bonds maintained[b]	Base-pair-stacking interactions (kcal/mol)[c]	Most dynamic minor groove width (Å)[d]	Most enlarged minor groove width (Å)[d]	Twist angle (°)[e]	T_m (°C)[f]	Relative NER[g]
G*T*	3	−119.4	11.0 (2.0)	17.6 (1.4)	3.0 (5.6)	53.3 (1.1)	Moderate ~26%
Unmodified control	5	−136.0	7.2 (1.3)	7.5 (1.3)	30.9 (4.2)	70.4 (1.2)	
G*CT*	3	−125.9	14.7 (2.3)	19.0 (2.0)	3.2 (14.8)	45.4 (1.9)	Excellent ~100%
Unmodified control	8	−150.5	7.2 (1.5)	7.3 (1.4)	32.2 (4.3)	68.3 (1.8)	

[a]Standard deviations are given in parentheses.
[b]At the G*T* lesion site, the unmodified control has three Watson–Crick hydrogen bonds for G and two for T. In the G*T* cross-link, all three hydrogen bonds of G are maintained, but the bonds involving T are broken. Similarly for the G*CT* unmodified control there are eight Watson–Crick hydrogen bonds, but with the lesion present only the three involving C remain intact (see Figures 3 and 9). Figure 6 shows the dynamic nature of the Watson–Crick base pairing in the damaged duplexes.
[c]Van der Waals-stacking interactions between adjacent base pairs are computed for all pairwise interactions, except that one base pair at each end was eliminated. See Figure 6, showing the dynamic nature of base pair stacking in the damaged duplexes.
[d]See Figure 7, showing the dynamic nature of the minor groove dimensions in the damaged duplexes.
[e]The most untwisted value in the damaged duplex is given. See Figure 5, showing the dynamic nature of the untwisting in the damaged duplexes.
[f]Mean of two trials for duplex 17-mers (5′-CCACCAACG*T*CACCACC-3′ and 5′-CCACCAACG*CT*ACCACC-3′).
[g]Data is taken from Figure 2C.

is severely untwisted locally; at the G*T* step, the ensemble average mean twist angle is 3.0° compared to 30.9° in the unmodified control (Figure 5 and Table 1). In addition, while the G* is stacked with A6, the T* is tilted and extruded into the major groove to form the intrastrand cross-link (Figures 3a and b). This disrupts Watson–Crick hydrogen bonding of T* with its partner A6, although there is a single, non-Watson–Crick

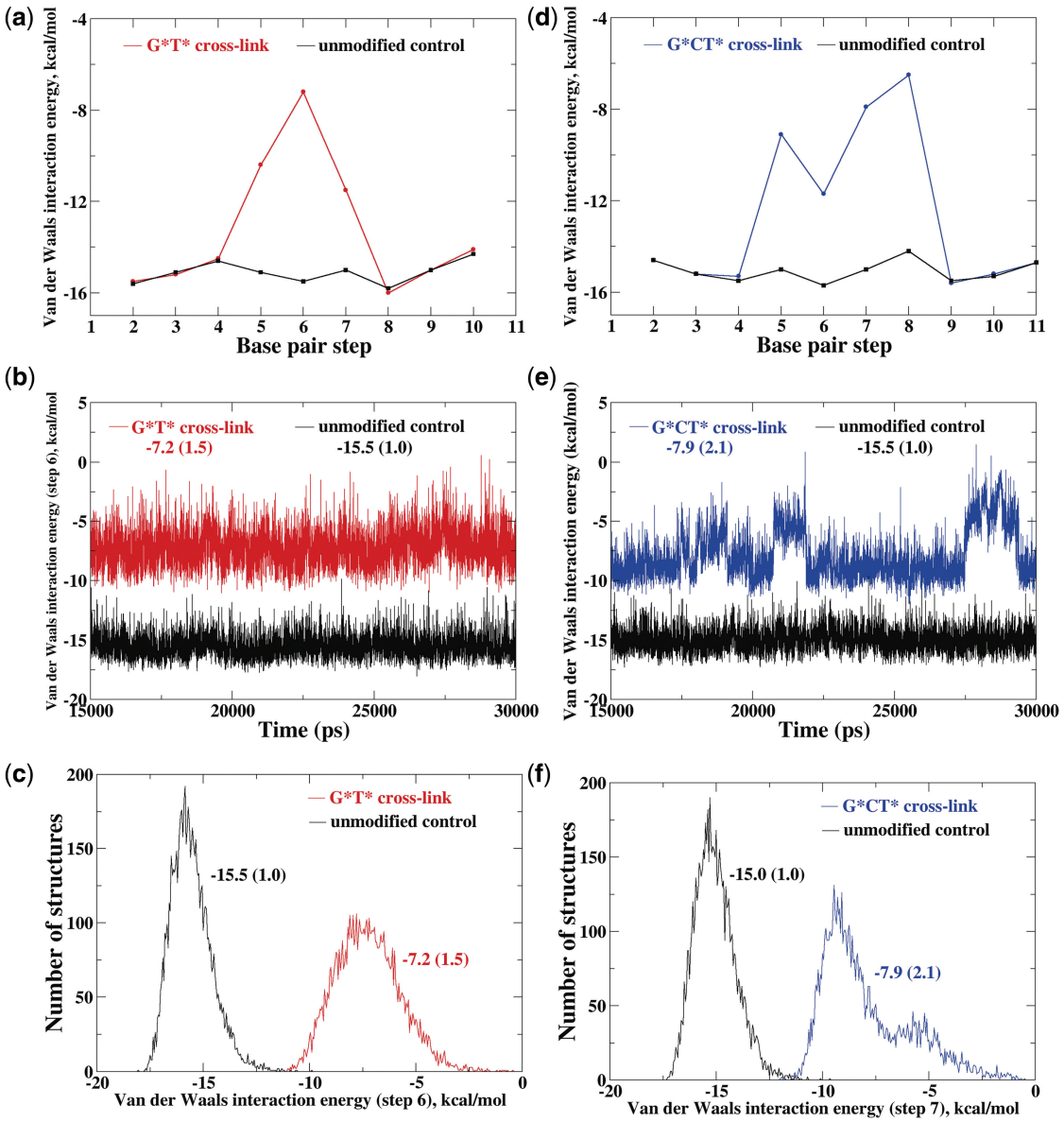

Figure 6. (**a**) Ensemble average van der Waals-stacking interaction energy (kcal/mol) vs base pair step plot for the G*T* cross-link; A2:T23 to C3:G22 is step 2, C3:G22 to C4:G21 is step 3, and so on; (**b**) time dependence and (**c**) population distribution of stacking interactions in the G*T* cross-link (Step 6). (**d**–**f**) Stacking interactions for the G*CT* cross-link, C2:G25 to A3:T24 is step 2, A3:T24 to A4:T23 is step 3, and so on. Step 7 of the G*CT* cross-link is given in (e) and (f). Ensemble average values in (a) and (d) are shown in Supplementary Table S8. The most dynamic steps, with largest standard deviations, are shown in (b) and (c) for the G*T* cross-link, and in (e) and (f) for the G*CT* case. Ensemble average values and standard deviations (in parentheses) are given in (c) and (f).

hydrogen bond between O2(T*) and N6(A6). The extrusion of T* causes a significant loss of base-stacking interactions in the vicinity of the cross-link site (Figure 6a). The van der Waals interaction energies between adjacent base pairs show that the stacking interactions between G*:C7 and T*:A6 are weakened, by 8 kcal/mol compared to the unmodified control (Supplementary Table S8). The total stacking interactions in the central 10-mer are weakened by 17 kcal/mol relative to the unmodified control (Table 1). Furthermore, the DNA bends toward the major groove to shield one face of the T* from the solvent, which greatly opens the minor groove (Figure 7a).

The G*T* cross-link also causes the modified duplex to be more dynamic than the unmodified control. This is revealed in the dynamics of a number of structural parameters: the local helical twist is more dynamic than the unmodified control, occupying a range from ~−21° to +23° compared with from ~13° to 44° for the unmodified duplex (Figures 5a and b). Stacking interactions (Figure 6b and c) and minor groove dimensions (Figures 7b and c) show similar dynamic properties, with greater fluctuations in time and larger ranges than the unmodified control. Watson–Crick hydrogen bonding is also somewhat more dynamic, as shown in Figure 8a and b.

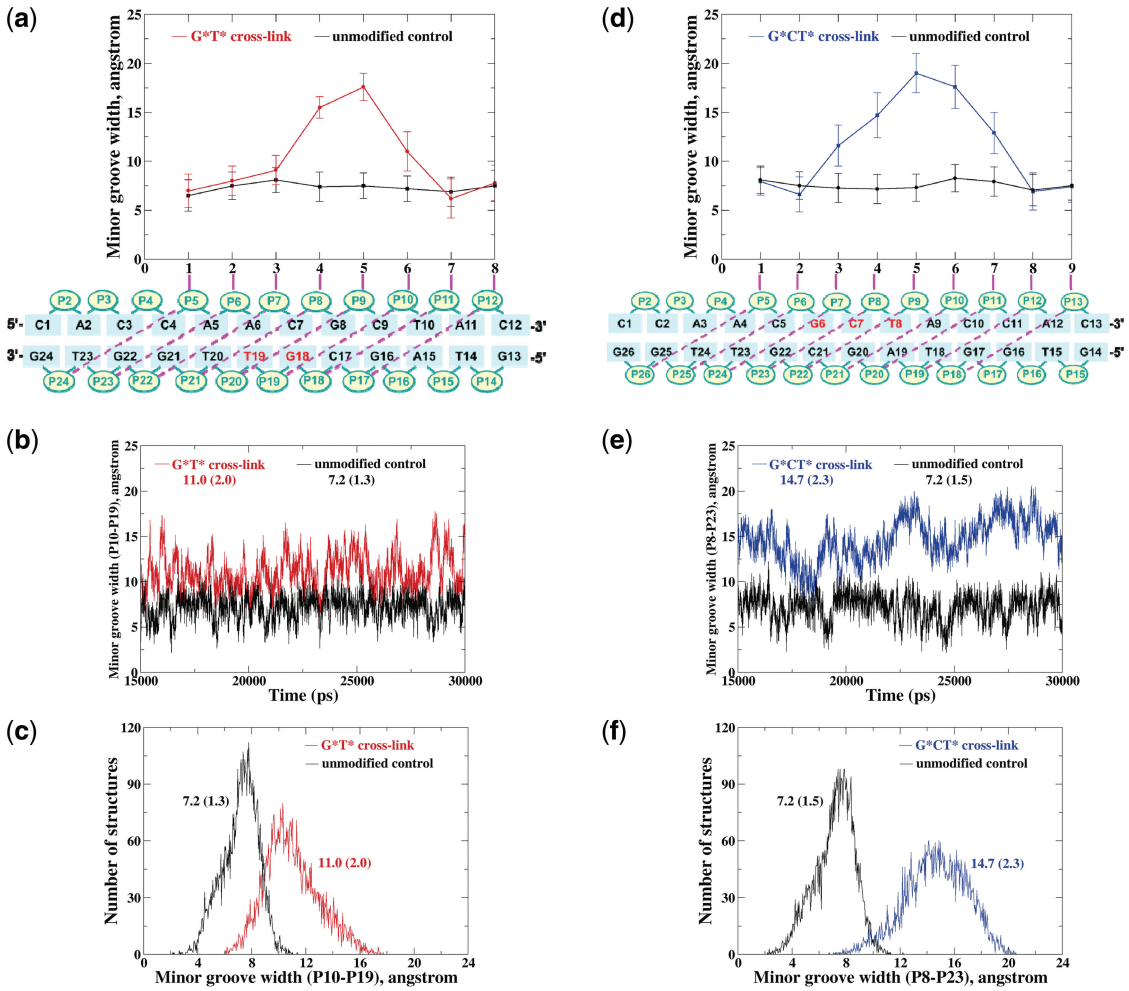

Figure 7. (**a**) Ensemble average minor groove widths (Å) with standard deviations for the G*T* cross-link. (**b**) Time dependence and (**c**) population distribution of minor groove width at P10–P19 in the G*T* cross-link. This is the most dynamic minor groove dimension with the largest standard deviation. (**d**) Minor groove dimensions in the G*CT* cross-link. The minor groove width at P8–P23 in the G*CT* cross-link is shown in (**e**) and (**f**). Ensemble average values and standard deviations (in parentheses) are given in (c) and (f).

The G*CT* cross-link: More severely distorting and more dynamic than the G*T* cross-link

For the G*CT*cross-link, we computed free energies of all eight models that we had obtained which contained proper chemical bonds between the C8(G*) and N3(T*) (Supplementary Figure S7). The relative free energies are given in Supplementary Table S9. Our results showed that the lowest energy structure, Model 1, was severely distorted relative to the unmodified control. The other models (~4–15 kcal/mol higher in energy) were even more distorted, and the stacking and hydrogen bonding are particularly affected. These models are depicted in Supplementary Figure S7 and their structural properties are summarized in Supplementary Table S7. We, therefore, consider Model 1 (Figure 9a and b) as the most plausible structure for the G*CT* cross-link. The initial model for MD that provided this best representative structure (See Supplementary Data, Methods) required somewhat more rearrangement during the MD simulation to achieve this representative structure (Figure 9c and d) than for the G*T* case (Figure 3). The realignment in G*CT* is facilitated through the flexibility permitted in the DNA backbone by the intervening deoxycytidine nucleotide and is illustrated in Supplementary Movie S3. Possibly the higher reaction yield for the G*CT* cross-link than for the G*T* case (8) is related to the greater flexibility available to form this cross-link. We computed the torsion angle governing the flexibility of the single bond that creates the cross-link, C4(T*)-N3(T*)-C8(G*)-N7(G*) (Figure 1b) for this model, and obtained ensemble average values of $62.7 \pm 10.5°$, in the same range as for the G*T* cross-link.

By comparison with the G*T* cross-link, the G*CT* cross-link impairs Watson–Crick hydrogen bonding and base-stacking interactions much more severely and more dynamically. In the G*CT* structure, only one Watson–Crick base pair remains of the three in the unmodified control (Table 1). The G* is stacked with C*, and the T* is extruded into the major groove and is unstacked (Figure 9). Watson–Crick hydrogen bonding between C*

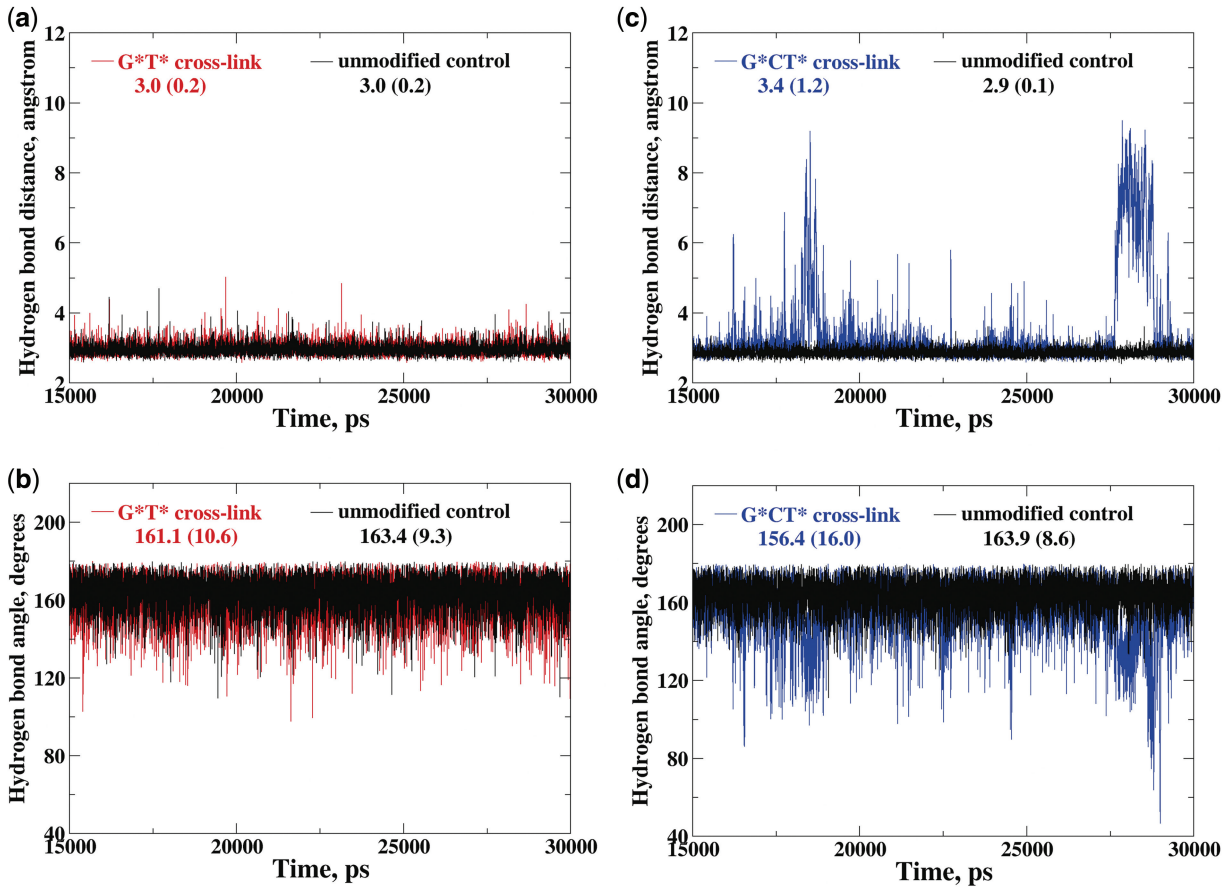

Figure 8. Hydrogen bond distances and angles versus time for base pairs C7:G*of G*T* and C*:G20 of G*CT*-modified duplexes, showing that dynamic disruption of hydrogen bonding is much greater for the G*CT* cross-link. The hydrogen bonds shown are the most dynamic ones in these Watson–Crick pairs; they are O6(G*)–N4(C7) and O6(G*)–H42(C7)–N4(C7) for the G*T* cross-link and N2(G*)–O2(C7) and N2(G*)–H22(G*)–O2(C7) for the G*CT* case.

and its partner G20 is episodically ruptured, seen in the plot of the hydrogen bond distances and angles (Figures 8c and d); occupancies of the three hydrogen bonds are in the range of 77–88% (Supplementary Table S10), while occupancies for the unmodified control are around 99%. Furthermore, G* is displaced too far toward the major groove to permit any Watson-Crick pairing with its partner C21. The bridging C* is partly cross-strand stacked with A19, to form a tilted stacking triad comprising G*, C* and A19 (Figure 9). However, the cross-link greatly disturbs base-stacking interactions, and this is shown in the total stacking interaction energy. Specifically, the total van der Waals stacking interaction energy of the central 11-mer is −126 kcal/mol compared to −151 kcal/mol for the unmodified control, revealing a destabilization of about 25 kcal/mol by the cross-link (Supplementary Table S8 and Figure 6). The cross-link causes the duplex to bend into the major groove (Supplementary Figure S7), which greatly opens the minor groove, as in the G*T* cross-link. However, the G*CT* cross-link produces larger groove opening (Figure 7). Additionally, the G*CT* cross-link severely unwinds the duplex at the lesion site, with untwisting that is much more dynamic than for the G*T* case (Figure 5).

Lesion-induced duplex destabilization and repair susceptibility

Our experimental results have revealed that these DNA intrastrand cross-links are susceptible to NER, and that the G*CT* cross-link is better repaired than the G*T* cross-link, as shown in Figure 2c and Supplementary Figure S4. In addition, thermal melting data reveal that these cross-linked lesions destabilize 17-mer double-stranded DNA by ~17°C in the case of the G*T* cross-link and by ~23°C for the G*CT* intrastrand cross-link. Table 1 summarizes our experimental and computational results. Our structural studies show that the G*CT* cross-link causes more distortion and destabilization to the DNA duplex than the G*T* cross-link, including a more dynamic structure with more disrupted Watson–Crick hydrogen bonding, more disturbed base-pair-stacking interactions, more dynamic untwisting, and a greater opening of the minor groove. The greater structural distortions associated with the G*CT* lesion is consistent with its higher susceptibility to NER.

CONCLUSIONS

The one electron oxidation of guanine in DNA can give rise to unique G(C8)–(N3)T intrastrand cross-linked

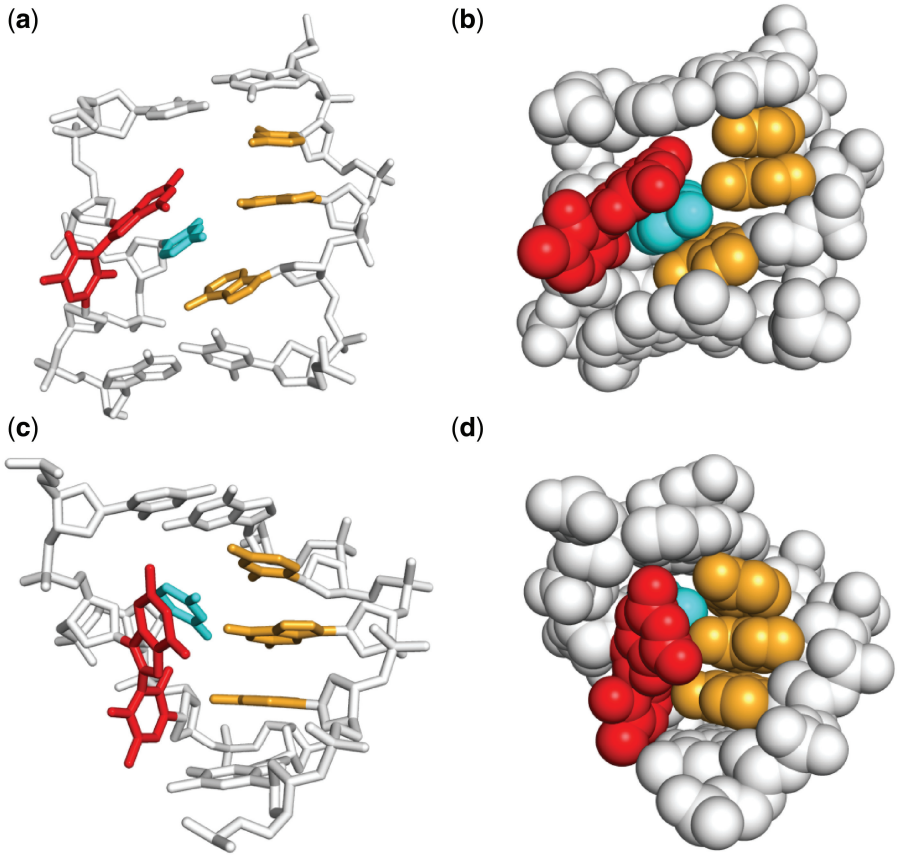

Figure 9. (**a**) The best representative structure (23) for the G*CT* cross-link modified duplex. See Supplementary Movie S2. (**b**) Central 5-mers for the G*CT* cross-link. The covalently linked model used to initiate the MD simulation that produced this structure (see 'Materials and Methods' section) is given in (**c**) and (**d**) with similar rendering as (a) and (b), respectively. Full duplexes are shown in Supplementary Figures S2 and S7. Supplementary Movie S3 shows snapshots from the MD trajectory (from 1 to 15000 psat 500 ps intervals) that illustrates the realignment. The color code is as follows: DNA, gray; G*–T* bases, red; the bridging C base, cyan; partner bases A19, C20 and C21, bright orange.

lesions in duplex DNA. Measurements of the thermal melting points of 17-mer duplexes with these single intrastrand cross-linked lesions indicate that the G*CT* is more destabilizing than the G*T*lesion. Dual incision repair assays in human cell extracts demonstrate that the G*CT* lesion is a better substrate of the human NER apparatus by a factor of ~4. Molecular modeling, molecular dynamics simulations and free energy calculations were utilized to describe the differences in the extent of dynamic structural disturbances to double-stranded DNA caused by these two cross-linked lesions, and it was found that the G*CT* cross-link produces much greater and more dynamic distortions than the G*T* lesion. These greater perturbations are consistent with the experimentally observed more pronounced destabilizing effects of the G*CT* lesions on the thermal melting points of the DNA duplexes and their greater susceptibilities to nucleotide excision repair. The larger structural distortions/destabilizations in the vicinity of the G*CT* cross-link and its greater susceptibility to NER, are consistent with the hypothesis that relative NER dual incision efficiencies of different lesions are a reflection of the extent of the local thermodynamic destabilization that the lesions impose on double-stranded DNA (31).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–10, Supplementary Figures 1–7, Supplementary Methods, Supplementary Movies 1–3 and Supplementary References [16–20,23,32–48].

ACKNOWLEDGEMENTS

The authors gratefully acknowledge TeraGrid resources provided by the Texas Advanced Computing Center supported by the National Science Foundation. S.D. carried out the computer modeling with the help of Y.C., K.K. and M.K. carried out the NER experiments and S. Durandina and Z.L. carried out the thermal melting experiments. V.S. and N.E.G. supervised the experimental work, while S.B. supervised the modeling studies. S.D. and S.B. analyzed the results, and S.D., S.B., Y.C., N.E.G. and V.S. wrote the manuscript.

FUNDING

National Institutes of Health (Grants R01CA75449 to S.B.) and (2 R01 ES 011589 to V.S.). Computational infrastructure and systems management were partially

supported by R01CA28038 (S.B). Funding for open access charge: National Institutes of Health (Grants R01CA75449 to S.B.).

Conflict of interest statement. None declared.

REFERENCES

1. Balkwill,F. and Mantovani,A. (2001) Inflammation and cancer: back to Virchow? *Lancet*, **357**, 539–545.
2. Coussens,L.M. and Werb,Z. (2002) Inflammation and cancer. *Nature*, **420**, 860–867.
3. Goldstein,S. and Czapski,G. (1999) Viscosity effects on the reaction of peroxynitrite with CO_2: Evidence for radical formation in a solvent cage. *J. Am. Chem. Soc.*, **121**, 2444–2447.
4. Steenken,S. and Jovanovic,S.V. (1997) How easily oxidizable is DNA? One-electron reduction potentials of adenosine and guanosine radicals in aqueous solution. *J. Am. Chem. Soc.*, **119**, 617–618.
5. Shafirovich,V., Dourandin,A., Huang,W. and Geacintov,N.E. (2001) The carbonate radical is a site-selective oxidizing agent of guanine in double-stranded oligonucleotides. *J. Biol. Chem.*, **276**, 24621–24626.
6. Joffe,A., Geacintov,N.E. and Shafirovich,V. (2003) DNA lesions derived from the site selective oxidation of Guanine by carbonate radical anions. *Chem. Res. Toxicol.*, **16**, 1528–1538.
7. Crean,C., Geacintov,N.E. and Shafirovich,V. (2005) Oxidation of guanine and 8-oxo-7,8-dihydroguanine by carbonate radical anions: Insight from oxygen-18 labeling experiments. *Angew. Chem. Int. Ed Engl.*, **44**, 5057–5060.
8. Crean,C., Uvaydov,Y., Geacintov,N.E. and Shafirovich,V. (2008) Oxidation of single-stranded oligonucleotides by carbonate radical anions: generating intrastrand cross-links between guanine and thymine bases separated by cytosines. *Nucleic Acids Res.*, **36**, 742–755.
9. Yun,B.H., Geacintov,N.E. and Shafirovich,V. (2011) Generation of guanine–thymidine cross-links in DNA by peroxynitrite/carbon dioxide. *Chem. Res. Toxicol.*, **24**, 1144–1152.
10. Box,H.C., Budzinski,E.E., Dawidzik,J.D., Wallace,J.C., Evans,M.S. and Gobey,J.S. (1996) Radiation-induced formation of a crosslink between base moieties of deoxyguanosine and thymidine in deoxygenated solutions of d(CpGpTpA). *Radiat. Res.*, **145**, 641–643.
11. Box,H.C., Dawidzik,J.B. and Budzinski,E.E. (2001) Free radical-induced double lesions in DNA. *Free Radic. Biol. Med.*, **31**, 856–868.
12. Hong,H., Cao,H. and Wang,Y. (2006) Identification and quantification of a guanine–thymine intrastrand cross-link lesion induced by Cu(II)/H_2O_2/ascorbate. *Chem. Res. Toxicol.*, **19**, 614–621.
13. Cao,H. and Wang,Y. (2007) Quantification of oxidative single-base and intrastrand cross-link lesions in unmethylated and CpG-methylated DNA induced by Fenton-type reagents. *Nucleic Acids Res.*, **35**, 4833–4844.
14. Yang,Z., Colis,L.C., Basu,A.K. and Zou,Y. (2005) Recognition and incision of gamma-radiation-induced cross-linked guanine–thymine tandem lesion G[8,5-Me]T by UvrABC nuclease. *Chem. Res. Toxicol.*, **18**, 1339–1346.
15. Raychaudhury,P. and Basu,A.K. (2011) Genetic requirement for mutagenesis of the G[8,5-Me]T cross-link in Escherichia coli: DNA polymerases IV and V compete for error-prone bypass. *Biochemistry*, **50**, 2330–2338.
16. 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. *et al.*(2006) *AMBER 9*. University of California, San Francisco.
17. Cieplak,P., Cornell,W.D., Bayly,C. and Kollman,P.A. (1995) Application of the multimolecule and multiconformational Resp methodology to biopolymers – charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.*, **16**, 1357–1377.
18. Cornell,W.D., Cieplak,P., Bayly,C.I., Gould,I.R., Merz,K.M., Ferguson,D.M., Spellmeyer,D.C., Fox,T., Caldwell,J.W. and Kollman,P.A. (1995) A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
19. Orozco,M., Perez,A., Marchan,I., Svozil,D., Sponer,J., Cheatham,T.E. and Laughton,C.A. (2007) Refinenement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. *Biophys. J.*, **92**, 3817–3829.
20. Kollman,P.A., Massova,I., Reyes,C., Kuhn,B., Huo,S., Chong,L., Lee,M., Lee,T., Duan,Y., Wang,W. *et al.* (2000) Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.*, **33**, 889–897.
21. Case,D.A., Pearlman,D.A., Caldwell,J.W., Cheatham,T.E. III, Wang,J., Ross,W.S., Simmerling,C.L., Darden,T.A., Merz,K.M., Stanton,R.V. *et al.* (2002) *AMBER 7*. University of California, San Francisco, CA.
22. Ravishanker,G., Wang,W. and Beveridge,D.L., (1998) *Molecular Dynamics Tool Chest 2.0*. Wesleyan University, Middletown, CT.
23. 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: Pullman,A. *et al.* (eds), *Modeling of Biomolecular Structure and Mechanisms*. Kluwer Acad. Publishers, Netherlands.
24. The PyMOL Molecular Graphics System, Version 1.3 Schrödinger, LLC.
25. Pirogov,N., Shafirovich,V., Kolbanovskiy,A., Solntsev,K., Courtney,S.A., Amin,S. and Geacintov,N.E. (1998) Role of hydrophobic effects in the reaction of a polynuclear aromatic diol epoxide with oligodeoxynucleotides in aqueous solutions. *Chem. Res. Toxicol.*, **11**, 381–388.
26. Kropachev,K., Kolbanovskii,M., Cai,Y., Rodriguez,F., Kolbanovskii,A., Liu,Y., Zhang,L., Amin,S., Patel,D., Broyde,S. *et al.* (2009) The sequence dependence of human nucleotide excision repair efficiencies of benzo[a]pyrene-derived DNA lesions: insights into the structural factors that favor dual incisions. *J. Mol. Biol.*, **386**, 1193–1203.
27. Hess,M.T., Gunz,D., Luneva,N., Geacintov,N.E. and Naegeli,H. (1997) Base pair conformation-dependent excision of benzo[a]pyrene diol epoxide-guanine adducts by human nucleotide excision repair enzymes. *Mol. Cell. Biol.*, **17**, 7069–7076.
28. Mocquet,V., Kropachev,K., Kolbanovskiy,M., Kolbanovskiy,A., Tapias,A., Cai,Y., Broyde,S., Geacintov,N.E. and Egly,J.M. (2007) The human DNA repair factor XPC-HR23B distinguishes stereoisomeric benzo[a]pyrenyl-DNA lesions. *EMBO J.*, **26**, 2923–2932.
29. Wang,D., Hara,R., Singh,G., Sancar,A. and Lippard,S.J. (2003) Nucleotide excision repair from site-specifically platinum-modified nucleosomes. *Biochemistry*, **42**, 6747–6753.
30. Neidle,S., Schneider,B. and Berman,H.M. (2003) Fundamentals of DNA and RNA structure. In: Bourne,P.E. and Weissig,H. (eds), *Structural Bioinformatics*. Wiley-Liss, Inc, Hoboken, N.J.
31. Cai,Y., Patel,D.J., Broyde,S. and Geacintov,N.E. (2010) Base sequence context effects on nucleotide excision repair. *J. Nucleic Acids*, **2010**, 1–9. doi:10.4061/2010/174252.
32. Cheatham,T.E., Cieplak,P. and Kollman,P.A. (1999) A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.*, **16**, 845–862.
33. Zhang,Q., Broyde,S. and Schlick,T. (2004) Deformations of promoter DNA bound to carcinogens help interpret effects on TATA-element structure and activity. *Philos. T. Roy. Soc. A*, **362**, 1479–1496.
34. Frisch,M.J., Trucks,G.W., Schlegel,H.B., Scuseria,G.E., Robb,M.A., Cheeseman,J.R., Montgomery,J.J.A., Vreven,T., Kudin,K.N., Burant,J.C. *et al.* (2004) *Gaussian 03*. Gaussian, Inc., Wallingford CT.
35. Bayly,C.I., Cieplak,P., Cornell,W.D. and Kollman,P.A. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges – the Resp Model. *J. Phys. Chem.-Us*, **97**, 10269–10280.
36. Jorgensen,W.L., Chandrasekhar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.

37. Essmann,U., Perera,L., Berkowitz,M.L., Darden,T., Lee,H. and Pedersen,L.G. (1995) A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.*, **103**, 8577–8593.
38. Darden,T., York,D. and Pedersen,L. (1993) Particle Mesh Ewald – an N.Log(N) method for Ewald Sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.
39. Ryckaert,J.P., Ciccotti,G. and Berendsen,H.J.C. (1977) Numerical integration of Cartesian equations of motion of a system with constraints – molecular dynamics of *N*-alkanes. *J. Comput. Phys.*, **23**, 327–341.
40. Izaguirre,J.A., Catarello,D.P., Wozniak,J.M. and Skeel,R.D. (2001) Langevin stabilization of molecular dynamics. *J. Chem. Phys.*, **114**, 2090–2098.
41. Berendsen,H.J.C., Postma,J.P.M., Vangunsteren,W.F., Dinola,A. and Haak,J.R. (1984) Molecular dynamics with Coupling to an External Bath. *J. Chem. Phys.*, **81**, 3684–3690.
42. Fratini,A.V., Kopka,M.L., Drew,H.R. and Dickerson,R.E. (1982) Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATTBrCGCG. *J. Biol. Chem.*, **257**, 14686–14707.
43. 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.
44. Honig,B. and Nicholls,A. (1995) Classical electrostatics in biology and chemistry. *Science*, **268**, 1144–1149.
45. Nicholls,A. and Honig,B. (1991) A Rapid Finite Difference Algorithm, Utilizing Successive Over-Relaxation to Solve the Poisson-Boltzmann Equation. *J. Comp. Chem.*, **12**, 435–445.
46. Alberts,B., Bray,D., Lewis,J., Raff,M., Roberts,K. and Watson,J.D. (1994) *Molecular Biology of the Cell*, 3rd edn. Garland Publishing, Inc, New York, NY.
47. Sanner,M.F., Olson,A.J. and Spehner,J.C. (1996) Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers*, **38**, 305–320.
48. Sitkoff,D., Sharp,K.A. and Honig,B. (1994) Accurate calculation of hydration free-energies using macroscopic solvent models. *J. Phys. Chem.-Us*, **98**, 1978–1988.
49. 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.