The cis–syn cyclobutane pyrimidine dimer (CPD) (Fig. 1A) and the pyrimidine(6–4)pyrimidone photoproduct [(6–4)PP] constitute the two major classes of cytotoxic, mutagenic and carcinogenic DNA photoproducts induced by ultraviolet (UV) irradiation (1–3). The CPD lesion is repaired by various kinds of DNA repair enzymes (4–7). The T4 endonuclease V (endo V) is a repair enzyme that catalyzes the first step of the CPD-specific base excision repair pathway (6). The Escherichia coli CPD-specific photolyase binds specifically to the CPD and, upon excitation by blue light, splits the cyclobutane ring and restores the intact base (7). In in vitro assays, the CPD among DNA photoproducts is recognized with the lowest affinity by the E. coli uvrA protein, which initiates repair by the uvr(A)BC excinuclease complex (4). Therefore, this type of lesion is repaired about nine times more slowly by the uvr(A)BC excinuclease than is the (6–4)PP (8).

Various structural studies of CPD-containing DNA duplexes have suggested mechanisms of binding of the repair proteins to CPDs (6,9–12). The crystal structure of the CPD–endo V complex reveals that the DNA duplex exhibits a sharp kink at the central CPD site (6). This kink causes the adenine base opposite the 5′ T of the CPD to flip out of the DNA duplex and be trapped in a pocket of the endo V enzyme (6). Results of a structural study of a DNA dodecamer duplex that contained a CPD suggest that the unique structural features of the DNA double helix (that is, helical bending, flexible backbone conformation, and significant changes of the major and/or minor grooves) might be important factors in determining the binding affinity of the XPC–hHR23B complex to DNA.

INTRODUCTION

The cis–syn cyclobutane pyrimidine dimer (CPD) (Fig. 1A) and the pyrimidine(6–4)pyrimidone photoproduct [(6–4)PP] constitute the two major classes of cytotoxic, mutagenic and carcinogenic DNA photoproducts induced by ultraviolet (UV) irradiation (1–3). The CPD lesion is repaired by various kinds of DNA repair enzymes (4–7). The T4 endonuclease V (endo V) is a repair enzyme that catalyzes the first step of the CPD-specific base excision repair pathway (6). The Escherichia coli CPD-specific photolyase binds specifically to the CPD and, upon excitation by blue light, splits the cyclobutane ring and restores the intact base (7). In in vitro assays, the CPD among DNA photoproducts is recognized with the lowest affinity by the E. coli uvrA protein, which initiates repair by the uvr(A)BC excinuclease complex (4). Therefore, this type of lesion is repaired about nine times more slowly by the uvr(A)BC excinuclease than is the (6–4)PP (8).

Various structural studies of CPD-containing DNA duplexes have suggested mechanisms of binding of the repair proteins to CPDs (6,9–12). The crystal structure of the CPD–endo V complex reveals that the DNA duplex exhibits a sharp kink at the central CPD site (6). This kink causes the adenine base opposite the 5′ T of the CPD to flip out of the DNA duplex and be trapped in a pocket of the endo V enzyme (6). Results of a structural study of a DNA dodecamer duplex that contained a CPD suggest that the unique BII type backbone
Affinity results from the increased structural distortion caused by the double T-G mismatches of the CPD and that additional factors and steps might be required for CPD recognition (19).

In order to elucidate how the T-G mismatches of the CPD affect the binding affinity of XPC–hHR23B to a CPD-containing DNA duplex, we determined the solution structures of two DNA decamer duplexes. One contained a mismatched base pair between the 3′ T of the CPD and the opposite G residue (referred to as the CPD/GA duplex; PDB accession no. 1pib), and the other contained two mismatched base pairs between two consecutive T residues that are part of a CPD and their opposing G residues (referred to as the CPD/GG duplex; PDB accession no. 1snh). These structures were then compared with that of a properly matched CPD (the CPD/AA duplex), which was described in our previous study (9). The structural differences among the various duplexes observed in this study can account for the observed enhancement of the XPC–hHR23B complex binding affinity for CPDs that contain double T-G mismatches. This structural comparison thus provides insight into the mechanism of the damage recognition step initiated by XPC–hHR23B during NER.

**MATERIALS AND METHODS**

**Sample preparation**

The CPD-containing DNA decamer was prepared by direct 254-nm UV irradiation of a DNA oligomer in an aqueous solution and purified as described (11). The CPD/GA and CPD/GG duplexes (Fig. 1B) were prepared by dissolving the lesion-containing DNA strands and the complementary strands at a 1:1 stoichiometric ratio in an aqueous solution containing 20 mM sodium phosphate (pH 7.0) and 100 mM NaCl.

**NMR experiments**

All NMR data sets generated with the CPD/GA and CPD/GG duplexes were collected with a Varian Inova 600 MHz spectrometer (KAIST, Daejon). Details of the NMR experiments and data processing can be found in our earlier published studies of a photoproduct-containing DNA duplex (11,20,21). Nuclear Overhauser effect (NOE) distance restraints from nonexchangeable protons were obtained from two-dimensional NOE spectroscopy (NOESY) experiments with mixing times of 80, 150 and 300 ms in a D2O buffer solution. Exchangeable proton NOEs were determined using NOESY spectra in H2O buffer with 100 and 300 ms mixing times. Watson–Crick-type hydrogen bonding restraints were imposed on each base pair, except the T5–Y16 and T6–X15 base pairs. The torsion angle restraints were derived by analyzing the data from DQF-COSY (22). Other backbone torsion angle restraints in the flanking C–G base pair region were equivalent to the values of a normal B-form helix. Torsion angle restraints in the CPD lesion and opposite nucleotides were only from the experimental data.

Scalar 1JCH and dipolar 1DCH couplings of the CPD/GG duplex were derived from natural abundance sensitivity enhancement 1H–13C HSQC experiments with and without 19F (~15 mg/ml) at 25°C (see table 1 in Supplementary Material). The data were collected on a Bruker 800 MHz DRX spectrometer equipped with a 1H [13C, 15N] triple resonance

![Chemical structures analyzed in this study. (A) Chemical structure of the CPD lesion. (B) DNA sequence contexts of the CPD/GA, CPD/GG and CPD/AA duplexes.](https://academic.oup.com/nar/article-abstract/32/8/2474/2904544/download)
cryoprobe (Osaka University, Osaka). Pf1 filamentous bacteriophage was purchased from ASLA, Ltd.

Structure calculation
The structures of the CPD/GA and CPD/GG duplexes were initially calculated using the program X-PLOR 3.1 (23) with restrained molecular dynamics. We initially generated the normal A- and B-form starting structures with modification of the CPD at the T5±T6 position. These structures were subjected to restrained molecular dynamics and simulated annealing protocols. Details of the structure calculation protocols can be found in our earlier published studies of a photoproduct-containing DNA duplex (15,20,21). Fifteen structures of the CPD/GA duplex and 13 structures of the CPD/GG duplex were chosen on the basis of the lowest total energies.

The ensemble of the CPD/GG duplex was then refined with RDCs in addition to NOE and torsion angle restraints. The calculation was performed by XPLOR–NIH version (24). Alignment tensor analysis of the observed residual dipolar coupling was performed as before (25,26). The values $D_a = -13.0$ Hz and $R = 0.4$ were used in the calculation. The refinement step consists of an initial equilibration stage where the dipolar coupling force constants were increased from 0.01 to 5.0 kcal mol$^{-1}$ Hz$^{-2}$ over 50 cycles, corresponding to a 15 ps molecular dynamics run. This was followed by a 20 ps restrained molecular dynamics calculation at 300 K. The final structures were generated after 2000 cycles of energy minimization (25).

The helical parameters of the refined structures were calculated using the program CURVES 5.3 (27) and Figures 2 and 7 were prepared with the program MOLMOL (28).

RESULTS AND DISCUSSION
Overall features of the solution structures of the CPD/GA and CPD/GG duplexes
In the CPD/GA duplex, a converged subset of 15 structures was identified on the basis of low NOE violations and total energies. These structures exhibited pairwise root-mean-squared deviation (r.m.s.d.) values of $1.05 \pm 0.24$ Å for all heavy atoms. The 13 superimposed refined structures of the CPD/GG duplex exhibited pairwise r.m.s.d. values of $1.10 \pm 0.43$ Å for all heavy atoms (Table 1).

Figure 2A and B shows stereo views of the mean structures of the CPD/GA and CPD/GG duplexes, respectively. Views of individual structures superimposed on the mean structures of the CPD/GA and CPD/GG duplexes show they are reasonably well converged (Fig. 2C). A ribbon representation was applied to the phosphate backbone of each duplex in order to emphasize the distortion of the phosphate backbone caused by formation of the cyclobutane ring and the T–G wobble pairs.

The T residues of the CPD formed stable wobble pairs with the opposite G residues
The H6 and 5-methyl resonances of the T5 and T6 residues were assigned from a unique NOE feature produced by four substituents of the cyclobutane ring of the CPD (Fig. 3A). The strong NOE cross-peak between the imino proton resonance of T5 and the H2 resonance of the opposite A16 in the CPD/GA duplex indicates that the T5-A16 base pair maintains the normal Watson–Crick T–A base pair (Fig. 4A). As was the case for the CPD/AA duplex (9), in the CPD/GA duplex, formation of the CPD caused some structural distortion of the 5′ T–A base pair, which was confirmed by the unusual values of the propeller twist angle. The T6-G15 wobble pair of the CPD formed hydrogen bonds between the T6-imino and G15-O6 and between the G15-imino and T6-O2 (Fig. 4A). This hydrogen bonding feature is consistent with the observation of a strong NOE cross-peak between the two imino protons of the T6-G15 mismatch (see figure 1 in Supplementary Material).
The calculated structure of the CPD/GA duplex showed that the T6–G15 base pair was slightly distorted, as confirmed by the propeller twist (~24°) value. In the CPD/GG duplex, the two imino proton resonances of the T5 and T6 residues showed strong NOEs with the imino protons of the opposite G16 and G15 residues, respectively. The two T (T5, T6) residues of the CPD in the CPD/GG duplex formed wobble base pairs with the opposite G residues, similar to the T6–G15 base pair in the CPD/GA duplex (Fig. 4B). However, the calculated structure of the CPD/GG duplex showed that the T6–G15 base pair was distorted more severely than that of CPD/GA as confirmed by the unusual propeller twist (~24°) and buckle (~44°) angles. The T5–G16 wobble pair was also distorted severely; it showed a large negative propeller twist value.

In temperature-dependent imino proton spectra conducted in H2O buffer (Fig. 5), all imino resonances except those of the terminal base pairs in both duplexes were intact over 25°C, indicating that the CPD and CPD/GA duplexes form stable double helices at room temperature. However, the melting temperature of the CPD/GG duplex was 10°C lower than that of the CPD/GA duplex (Fig. 5). This indicates that the double T–G mismatches of the CPD slightly destabilize the DNA double helix. In both duplexes, the imino proton resonances of the mismatched T and G residues located at the central CPD site show that these residues have the same melting temperature as that of the overall helix. This means that the T residues of the CPD form stable base pairs with the opposite G residues rather than bubble-like structures which may be caused by weak base pairing.

**Base stacking interactions at the CPD sites differ between the two duplexes**

We also studied the base stacking interactions at the CPD sites and the 5′ and 3′ flanking regions of the CPD/GA and CPD/GG duplexes. The purine bases (A16 in CPD/GA and G16 in CPD/GG) opposite the 5′ T (T5) of the CPD were nicely stacked with the 5′ flanking T17 bases (Fig. 3B and C). The chemical shift (0.56 p.p.m.) of the T5–methyl resonance in the CPD/GA duplex was similar to that of the CPD/AA duplex (0.57 p.p.m.) (Fig. 3A). As was the case with the CPD/AA duplex, structure calculation revealed that the A4/T5 base step has a short rise (~2.4 Å) value and, thus, the distance between the methyl carbon of T5 and the C8 molecule of base A4 is very short (Fig. 3B). This structural feature and the location of the T5-methyl group can explain the unusual upfield-shift of the T5-methyl resonance, which is caused by a shielding effect of the A4 base ring current.

In contrast, the T5-methyl resonance in the CPD/GG duplex was significantly downfield-shifted to 1.33 p.p.m. (Fig. 3A). In the CPD/GG duplex, the A4/T5 base step feature is significantly different from that of the CPD/GA duplex (Fig. 3B and C). In the CPD/GG duplex, the T5–G16 wobble pair changed the rise value of the A4/T5 base step (~2.9 Å) and the location of the T5-methyl (Fig. 3C). This structural change disrupted the shielding effect of the A4 base on the T5-methyl resonance, as was evident by the downfield shift of the T5-methyl resonance (Fig. 3A). In the CPD/GG duplex, the base stacking interaction at the CPD site and 3′ flanking region was disrupted, as evidenced by its negative roll (~24°) angle; this observation is similar to that made for a DNA duplex decamer that contained a 3′ T–T base pair in its CPD (15) and in contrast to the CPD/GA duplex, which was well stacked. This indicates that the 5′ T–G wobble pair of the CPD causes structural distortions not only in the 5′ flanking region, but also in the 3′ flanking region of the CPD.

**The CPD/GG duplex exhibits significant helical bending**

In the previous study, we found that formation of the CPD caused a slight bending of the DNA helix in the CPD/AA duplex (9). Our structural calculation herein showed that the 3′ T–G mismatch of the CPD maintained this helical bending (~10°) in the CPD/GA duplex. However, more helical bending was observed in the calculated structures of CPD/GG duplex. Recently, it was found by scanning force spectroscopy that the XPC–hHR23B complex induces a bend in the DNA helix upon binding and that this bend stabilized XPC–hHR23B binding at the damaged site (29). It was also reported that the helical bending angle caused by a (6–4)PP lesion, which is a good substrate for binding to the XPC–hHR23B complex, is quite large (~44°) (9,30). Unusual DNA structures, such as those that contain bubbles and junctions between the single- and double-strands, yield flexible helices that are easily bent by the binding of a protein. DNA damage and unusual DNA structures, which are efficiently recognized by the XPC–hHR23B complex, appear to have the helical bending properties or allow helices to be bent easily at the binding site of a protein complex. Thus, we hypothesize that the helical bending properties observed in the CPD/GG duplex may enhance the binding affinity of the XPC–hHR23B complex for the mismatched CPD lesion.

**The backbone conformation of the CPD/GG duplex is distorted and flexible**

In both structures, the backbone conformation involving the phosphorous atom between the 3′ T (T6) of the CPD and its 3′ neighboring A7 residue is distorted, as confirmed by the gauche− orientation of the ϵ angle and the trans orientation of the ζ angle in the calculated structures. Because H2′−H3′, H2′−H4′, H2′−H3′ and H3′−H4′ cross-peaks of T6 of both duplexes are absent in the DQF-COSY spectra (data not shown), we can deduce that the ϵ angles of T6 bases are in either the trans or gauche− region (ΣϵT6′ < 10 Hz) (31). Even though we did not
put any artificial dihedral restraints on these damaged bases and opposite ones, the calculated structures showed the ε angles of T6 bases had gauche− conformations. This B II conformation is in accord with the analysis of scalar coupling patterns and is expected to play a crucial role in the recognition of a CPD by repair enzyme, such as endo V (10).

We also found that some of the calculated backbone torsion angles for the CPD/GG duplex are not within a single conformational range. This applies mainly to the damaged bases and their neighbors. In the calculated structures, the γ angles of G15 and G16 were measured in either the gauche+ or the trans range. There were no H3′−H4′, H4′−H5′ and H4′−H5″ cross-peaks of G15 and G16 in the DQF-COSY spectra, which implies that the γ angles of both bases may be in either the gauche+ or the anti− range. However, the possibility of the anti− range can be excluded by analyzing NOESY data obtained at 1°C (31). This means that the γ angles of G15 and G16 may have the gauche+ conformation at least at the low temperature. Because the sequential NOE cross-peaks and even the intra-NOE cross-peaks of G15 between base and sugar at these steps were not observed or were very weak in NOESY spectra recorded at 17°C, whereas these peaks exist in NOESY spectra recorded at 1°C, it is expected that the residues may experience a kind of slow dynamical process (Fig. 6). Although the proposal of a flexible DNA backbone from experimental data and calculation results is a reasonable one, it is rather dangerous to propose this flexibility solely on the basis of observed torsion angles inter-conversion in the course of restrained molecular dynamics. Additionally, the previous simulation study also showed that the phosphate that bridges the bases complementary to the CPD had flexibilities in the CPD/AA duplex (32).

Even though the sugar-backbone conformations at these step are somewhat flexible in nature, the structural conformations of these base pairs are well converged in the ensemble of our calculated structures, as evidenced by strong imino resonances in the temperature-dependent 1-D spectra (Fig. 5).

It was reported that the XPC–hHR23B complex shows higher affinity for single-stranded DNA than for double-stranded DNA (33). Also, certain secondary DNA structures were strongly preferred by the complex, such as single- and double-stranded junctions (33) and three- or five-nucleotide bubble structures (19). However, Batty et al. clearly showed single-stranded character by itself was an insufficient explanation for binding of XPC–hHR23B complex to DNA by the competition experiments (34). These results imply that the DNA-binding affinity of the XPC–hHR23B complex may correlate with the presence of the flexibility provided by the unpaired bases in a DNA double helix. In the three CPD-containing DNA duplexes we studied here, all T residues of the CPDs formed stable hydrogen bonds with the opposite A or G residues. However, even though the two T residues of the CPD in the CPD/GG duplex were able to form stable wobble
pairs with the opposite G residues, the binding affinity of the XPC–hHR23B complex for this CPD and the repair rate for this CPD via NER were increased significantly. These findings imply that the XPC–hHR23B complex does not recognize unpaired bases directly, but, rather, recognizes the structural distortion caused by bases that are unpaired or mismatched.

We mentioned that the CPD/GG duplex has a highly flexible backbone conformation, which is consistent with the fact that we observed no sequential NOE cross-peaks at the CPD-damaged site at 17°C (Fig. 6). The backbone conformations of DNA structures that have bubbles or junctions between single- and double-strands are extremely flexible in the single-stranded regions. The (6–4)PP lesion also has a flexible backbone conformation at its 3’ T site and melts at room temperature (13,30). Thus, we suggest that the flexibility of the distorted backbone conformation of double-stranded DNA provides a crucial factor in determining the binding affinity of the XPC–hHR23B complex for DNA.

**Widening of the major groove and narrowing of the minor groove at the CPD site in the CPD/GG duplex**

Figure 7 shows the major and minor grooves of the three DNA duplex decamers that contained a CPD and a (6–4)PP-containing DNA duplex decamer with the same flanking sequences. The CPD caused widening of the minor groove and narrowing of the major groove at the damage site in both the CPD/AA and CPD/GA duplexes (Fig. 7). This is consistent with the result from the crystal structure of CPD–endo V complex which shows that the protein interacts with the DNA in the minor groove (6). In contrast, for the CPD/GG duplex, the major groove at the CPD site was significantly wider than that of an average B-DNA structure (Fig. 7C). In addition, the minor groove at the CPD site of the CPD/GG duplex narrowed greatly to 2–4 Å (Fig. 7G). The groove width of a DNA double helix can play a crucial role in the formation of DNA–protein complexes, and it has been suggested that the change in groove width influences the specific binding affinity of proteins for CPDs in the context of the genome (6). As with the CPD/GG duplex, a (6–4)PP caused the enlargement of the major groove and compression of the minor groove at the damaged site (Fig. 7D and H). It is possible that this significant change in groove width aids the XPC–hHR23B complex in detection of the CPD site. However, the major and minor groove widths of CPD in the CPD/GG duplex are less deviated from the normal B-DNA than that of (6–4)PP.
Implications for DNA damage recognition by the XPC–hHR23B complex

CPDs are produced directly by sunlight at a 5 to 10 times higher rate than the other major photoproduct, (6–4)PP (2,3). In the NER pathway, most DNA damage binding proteins show lower binding affinities for CPD than for other DNA photoproducts (4,12). The XPC–hHR23B complex, which is involved in the DNA damage recognition step in human cells, also rarely binds to CPD lesions (17,19). Thus, because of its high abundance and low repair rate, the CPD might be the most mutagenic photoproduct in eukaryotic cells. One unique feature of the NER pathway is that the XPC–hHR23B complex binds to a variety of DNA lesions and unusual structures with different affinities (1–3).

The XPC–hHR23B complex strongly prefers to bind to single-stranded DNA, specifically, secondary DNA structures that contain a single- and double-stranded junction and three- or five-nucleotide bubble structures (19,33). The complex also shows high binding affinity for DNA duplexes that contain (6–4)PP lesions (17). DNA photoproducts, base mismatches, and single-stranded regions all destabilize the DNA double helix and thus decrease the melting temperature of the DNA duplex (35,36).

Sugasawa et al. found that double T–G wobble pairs in CPDs enhance the binding affinity of the XPC–hHR23B complex for the CPD (19). It was believed that an understanding of the relationship between this enhanced binding affinity and the structural features induced by the double T–G wobble pairs would lead to a deciphering of the molecular mechanisms behind the damage recognition step initiated by the XPC–hHR23B complex. In the CPD/GA duplex, the single G residue, which forms stable hydrogen bonds with the opposite 3′ T residues of the CPD, causes a slightly distorted backbone conformation similar to that of the CPD duplex with two A residues. However, having double T–G wobble pairs in the CPD caused not only an unusual base pair geometry, but also backbone distortion, even though the two T residues of the CPD lesion formed stable hydrogen bonds with the opposite G residues (CPD/GG duplex). Among these structural features, we found three distortions related to the DNA backbone in the CPD/GG duplex that may correlate with binding of the XPC–hHR23B complex to DNA. First, the widths of the major and minor grooves in the CPD/GG duplex were greatly widened and narrowed, respectively, in contrast to the CPD/AA and CPD/GA duplexes. Second, the double T–G wobble pairs of the CPD formed stable hydrogen bonding, but induced flexibility into the sugar-backbone conformation of the DNA helix. The third feature is the helical bending caused by the double T–G wobble pairs in the CPD/GG duplex, which is more severe than those observed in CPD/AA and CPD/GA duplexes. This structural distortion may correlate with the greater affinity of the XPC–hHR23B complex for the CPD in the CPD/GG duplex, versus that in duplexes with matched or single T–G mismatched base pairs.

As discussed above, these structural features can be found in DNA duplexes that contain a (6–4)PP, which is the best DNA substrate for the XPC–hHR23B complex among the various DNA photoproducts. Some flexible DNA structures such as three- or five-nucleotide bubbles and partially single-stranded DNA duplexes, which are good substrates for the XPC–hHR23B complex, were able to obtain these structural features easily after binding to the complex. Thus, we suggest that the XPC–hHR23B complex recognizes DNA damage by searching for unusual groove width, flexible backbone distortion, and/or a bent helix during damage recognition in NER.
Supplementary Material

Supplementary Material is available at NAR Online.

Acknowledgements

We thank Dr Sung-Hun Bae for useful discussions. This work was supported by the Creative Research Initiative Program from the Ministry of Science and Technology, Korea. C.-J.P. and J.-S.S. were supported partially by the BK21 project.

References