

# Essential dynamics of DNA containing a *cis.syn* cyclobutane thymine dimer lesion

Hiroshi Yamaguchi\*, Daan M. F. van Aalten<sup>1,+</sup>, Miroslav Pinak<sup>2</sup>, Akira Furukawa<sup>3</sup> and Roman Osman<sup>4</sup>

Space and Particle Radiation Science Research Group, National Institute of Radiological Sciences, Anagawa 4-9-1, Inage-ku, Chiba 263, Japan, <sup>1</sup>Department of Biochemistry and Molecular Biology, University of Leeds, Leeds, UK, <sup>2</sup>Comenius University, Mlynska dolina F1, Bratislava, Slovakia, <sup>3</sup>Division of Radiobiology and Biodosimetry, National Institute of Radiological Sciences, Anagawa 4-9-1, Inage-ku, Chiba 263, Japan and <sup>4</sup>Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York, NY 10029, USA

Received December 31, 1997; Revised and Accepted February 20, 1998

## ABSTRACT

Conformational properties of a UV-damaged DNA decamer containing a *cis.syn* cyclobutane thymine dimer (PD) have been investigated by molecular dynamics (MD) simulations. Results from MD simulations of the damaged decamer DNA show a kink of  $\sim 21.7^\circ$  at the PD damaged site and a disruption of H bonding between the 5'-thymine of the PD and its complementary adenine. However, no extra-helical flipping of the 3'-adenine complementary to the PD was observed. Comparison to two undamaged DNA decamers, one with the same sequence and the other with an AT replacing the TT sequence, indicates that these properties are specific to the damaged DNA. Essential dynamics (ED) derived from the MD trajectories of the three DNAs show that the backbone phosphate between the two adenines complementary to the PD of the damaged DNA has considerably larger mobility than the rest of the molecule and occurs only in the damaged DNA. As observed in the crystal structure of T4 endonuclease V in a complex with the damaged DNA, the interaction of the enzyme with the damaged DNA can lead to bending along the flexible joint and to induction of adenine flipping into an extra-helical position. Such motions may play an important role in damage recognition by repair enzymes.

## INTRODUCTION

Ultraviolet light produces cyclobutane-type pyrimidine dimers (PD) in DNA. Such lesions can cause mutations and they strongly correlate with UV-induced cancer. T4 endonuclease V is a repair enzyme specific for PD which is encoded by the T4 *denV* gene and consists of 138 amino acids (1). The enzyme associates in a non-specific way with DNA that does not contain PD. It scans

non-target DNA by electrostatic unidimensional diffusion, as indicated by the salt dependence of the processive component of enzymatic catalysis (2). The interaction becomes specific upon encounter of the enzyme with the target PD. The enzyme catalyzes excision of the damage by two distinct processes. First, it hydrolyzes the glycosyl bond on the 5'-side of the PD (3), then the enzyme proceeds to incise the phosphodiester bond at the 3' position of the abasic site through  $\beta$ -elimination, resulting in an  $\alpha,\beta$ -unsaturated aldehyde and a 5'-terminal monophosphate (4,5). Mutational studies of the enzyme established that the N-terminus acts as the nucleophile and Glu23 is important in enzymatic activity (6). Arg3, Arg22, Arg26, Arg117 and Lys121 have been demonstrated to participate in non-target recognition (6–8). It appears that the enzymatic mechanism is reasonably well understood and represents a rather general mechanism in base excision repair.

In contrast to the enzymatic mechanism, the selectivity of damage recognition is still sketchy. Certain aspects of damage recognition could be inferred from the crystal structure of the enzyme (9–11) and the enzyme–DNA complex (12). The central depression in the protein is surrounded by four lysines and five arginines, generating a strong positive potential on the surface of the protein. This observation is fully consistent with the electrostatic mechanism of non-specific recognition of undamaged DNA. Determination of the crystal structure of the complex between endonuclease V and damaged DNA (12) provided an additional range of observations about the DNA–protein interaction. As predicted by a previous study (13), the protein interacts with the DNA in the minor groove. The protein in the complex has a very similar structure to the free enzyme. The only substantial differences occur in the positions of two arginines (Arg22 and Arg26), a loop on the opposite side of the recognition surface (residues 83–91) and a loop near the C-terminus (residues 125–130). The two arginines are observed to come close to the PD and the C-terminus loop partially overlaps with the WYKYY

\*To whom correspondence should be addressed. Tel: +81 43 251 2111; Fax: +81 43 251 4531; Email: yamag@nirs.go.jp

<sup>+</sup>Present address: Keck Structural Biology Group, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA

sequence, which was demonstrated to play a role in DNA binding (14–16).

An important discovery in the complex is the flipped out adenine complementary to the 5' thymine of the PD. The adenine is positioned in a cavity made up of mostly polar residues (Tyr21, Asp87, Thr89, Gln71 and Gln91), but forms no hydrogen bonds with any polar atoms. Interestingly, molecular dynamics simulations of a PD-containing DNA show that hydrogen bonding between the adenine and the complementary 5' thymine of the PD is grossly perturbed (17). Also, the DNA has a substantial kink in the helical axis near the PD, but of a smaller magnitude than DNA in the complex. This suggests a possible mechanism of base flipping as an element of recognition of DNA damage. Interaction of the protein with DNA may produce an enhanced kink which could facilitate base flipping.

To explore the relationship between the kinked structure, perturbed hydrogen bonding and base flipping we applied the method of essential dynamics (ED) (18) to reveal dynamic motions that could be responsible for base flipping. Such an approach has been applied previously to study structure–function relations in proteins. The method identified several principal concerted motions among atoms of the active site of the protein and attributed these motions to function of the enzyme (19–22).

## MATERIALS AND METHODS

### Molecular dynamics (MD) simulation of DNA

Three DNA decamers were simulated in this work; two native DNA sequences and one with a thymine dimer (T<sup>^</sup>T). The sequence of DNA I was d(GCGGATGGCG)<sub>2</sub> and of DNA II d(GCGGTTGGCG)<sub>2</sub>. DNA III had the same sequence as DNA II but the two thymines were replaced with a T<sup>^</sup>T representing the PD. These are the same sequences as those investigated by NMR by Lee *et al.* (23).

The native DNA sequences were constructed as idealized B-DNA structures using the NUCGEN module of AMBER 4.1 (24). The coordinates and partial charges of PD in DNA III were the same as those of Miaskiewicz (17). To neutralize the DNA, 18 K<sup>+</sup> ions were placed 5 Å from the phosphorus atom on the bisector of the O-P-O angle. The van der Waals parameters of the K<sup>+</sup> ions were the same as those of Na<sup>+</sup>. However, because the mass of the K<sup>+</sup> ion is larger than that of Na<sup>+</sup> the kinetic energy distribution of the K<sup>+</sup> ions in MD simulations has a somewhat different profile and may contribute better stability to the trajectory. The DNA was solvated by a rectangular box of TIP3P water with a cut-off of 18 Å and an exclusion distance of 2.9 Å from the oxygen and 2.1 Å from the hydrogen of the water. This procedure yielded 7973 water molecules for DNA I, 7995 for DNA II and 7960 for DNA III.

All the calculations were performed with AMBER 4.1. Periodic boundary conditions were applied with a cut-off of 12 Å for the van der Waals interactions and the Particle Ewald Mesh method was used to calculate the electrostatic interactions.

Initially the DNA was kept frozen and only the water molecules were minimized. This was followed by a minimization of the entire system. Initial velocities were assigned from a Maxwellian distribution at 30 K and the system was heated at constant volume to 300 K over a period of 10 ps with a coupling constant of 0.2 ps. To equilibrate the density of the system a constant pressure MD run was performed at 300 K with a coupling constant of 0.4 ps.

The equilibrated densities were 0.989 g/cm<sup>3</sup> for DNA I, 0.983 g/cm<sup>3</sup> for DNA II and 0.981 g/cm<sup>3</sup> for DNA III. The density stabilized within 7–10 ps. The final box sizes were: DNA I, 70.729 × 59.451 × 60.158 Å<sup>3</sup>; DNA II, 70.660 × 59.916 × 59.913 Å<sup>3</sup>; DNA III, 71.076 × 59.683 × 59.878 Å<sup>3</sup>. Production runs were performed at 300 K and constant volume for 500 ps. The trajectory was recorded every 20 fs. The simulations were executed on a DEC A600 Workstation at the National Institute of Radiological Sciences (NIRS).

### Essential dynamics of DNA

The method of essential dynamics (ED) (18) extracts essential degrees of freedom in molecular motions from a MD trajectory. These essential degrees of freedom correspond to correlated vibrational modes or collective motions of groups of atoms in normal mode analysis (25). The translational motions of MD trajectories are removed by a translation to the average geometrical center of the molecule and the rotational motions are removed by a least squares fit superimposition on a reference structure. To examine the correlated motions a covariance matrix, constructed from the trajectories, is defined as

$$M_{ij} = \langle (x_i - x_{i0})(x_j - x_{j0}) \rangle \quad \mathbf{1}$$

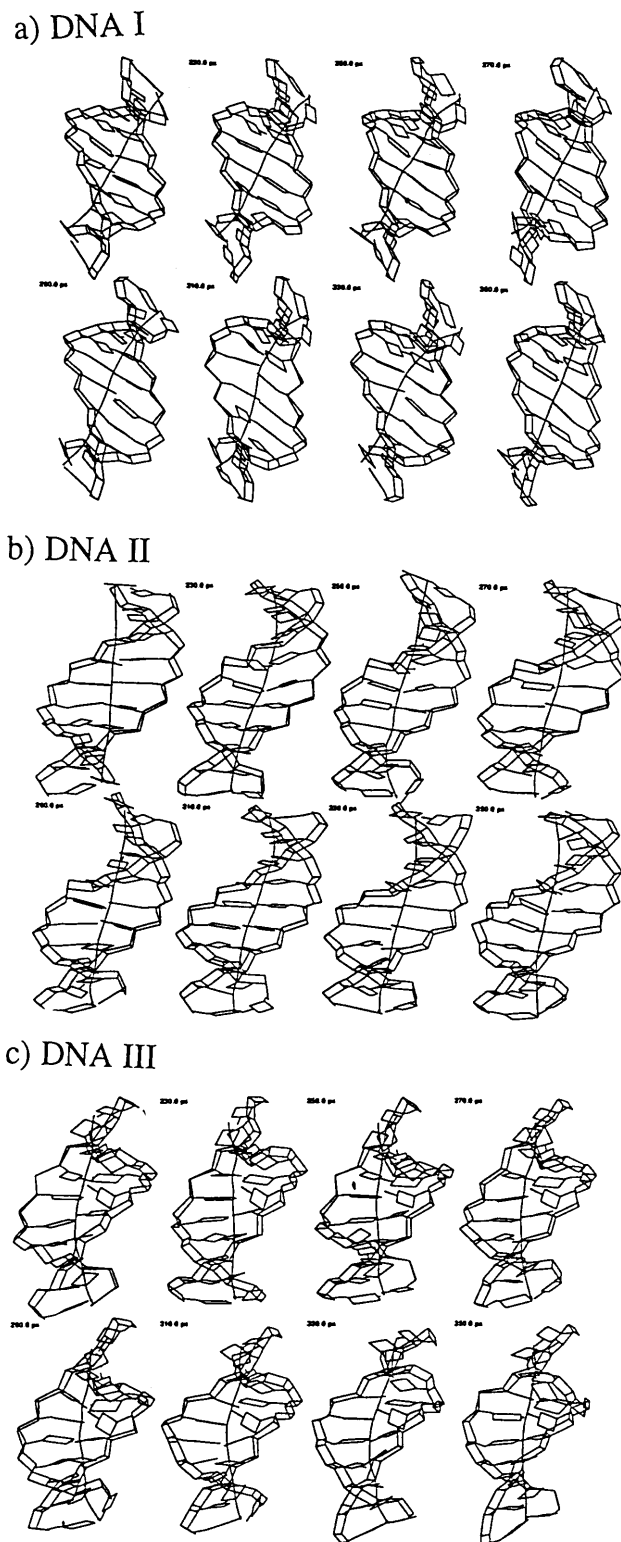
where  $x_i$  are the separate  $x, y, z$  coordinates of atoms in DNA and  $x_{i0}$  are the average coordinates derived from the trajectory.  $\langle \rangle$  designates time averaging over a specified period, which is usually taken from the stabilized portion of the simulation. Diagonalization of the covariance matrix yields a set of eigenvectors and eigenvalues. The eigenvectors indicate the directions of motions of the atoms and the eigenvalues represent the displacement variance of the particular eigenvector. Projection of the trajectory on a particular eigenvector shows the time-dependent motions that the atoms perform in the particular vibrational mode. A time average of the projection shows the contribution or components of the atomic vibrations to this mode of concerted motion.

We applied ED (26) to native DNA I and DNA II and damaged DNA III. Each DNA consists of 630 atoms. Trajectories of those 630 atoms from 200 to 420 ps were used (11 050 frames). We used the final structure after heating to 300 K as the reference structure. We applied ED to two cases: (i) all atoms except hydrogens (a total of 404 atoms); (ii) only phosphorus atoms (a total of 18 atoms). An SGI Workstation at the University of Leeds and a DEC A600 Workstation at the NIRS were used for ED calculations.

## RESULTS

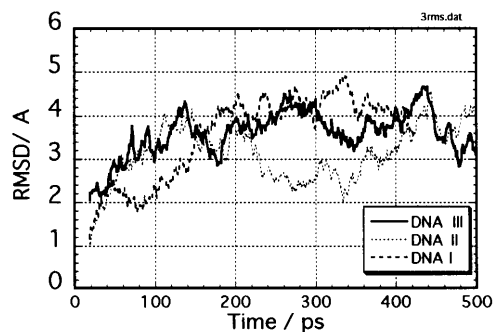
### Molecular dynamics

The MD trajectories of the three DNA sequences, DNAs I–III, produce three different structures and their trajectories exhibit a somewhat different behavior. A qualitative difference between them can be observed from plate images produced in the interval 220–350 ps with the program Dials\_and\_Windows (27). To focus on the difference in the area of the thymine dimer (in DNA III) Figure 1 shows the plate images for the three DNAs in the same view, such that the T5-A16 and T6-A15 base pairs are in the plane of the figure. It can be seen from Figure 1 that during this time period the conformation of the three DNAs does not change much. DNA I and DNA II show normal pairing of all bases, whereas DNA III shows distorted base pairing around the thymine dimer. It is not apparent from this presentation that DNA



**Figure 1.** Plate image output from the program 'Dials\_and\_Windows' (27). The three panels (a) DNA I, (b) DNA II and (c) DNA III show structures every 10 ps from 200 to 350 ps.

III has a significant kink around the PD (see below), but the helical axis appears to be consistently bent towards the PD.

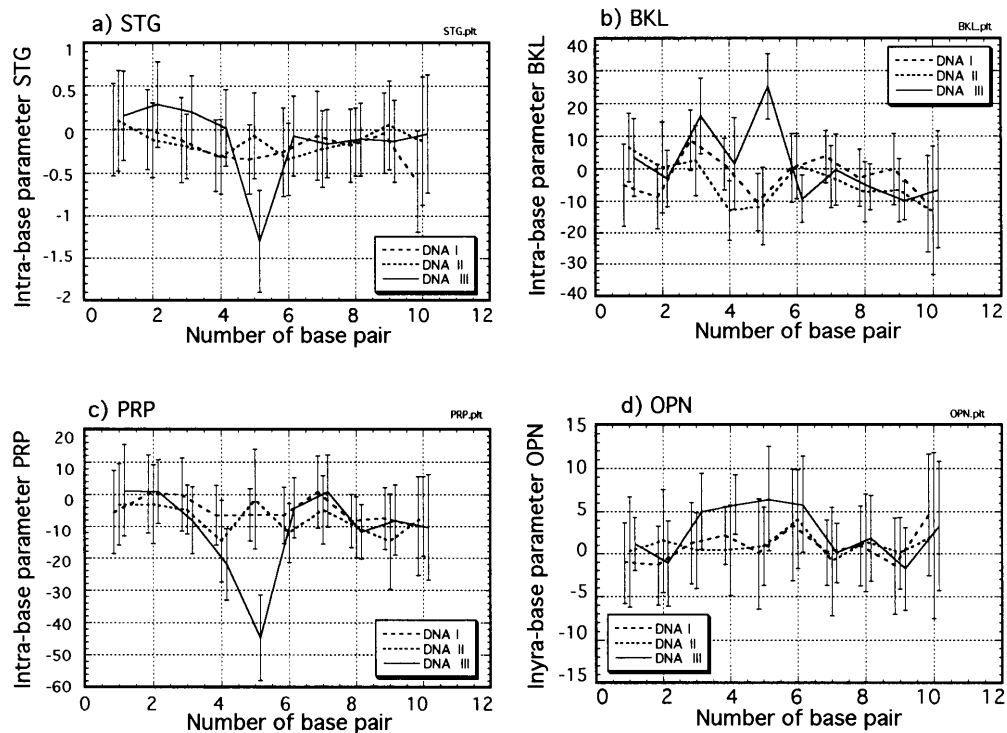


**Figure 2.** Root mean square deviation of DNA structures measured from a reference at  $t = 0$  ps of the MD run.

The energies of the different systems stabilized rather quickly, as they reach a constant value within 20 ps. A similar analysis of the changes in structural parameters as a function of time shows that the structures of the DNAs stabilize at a slower rate. The root mean square (r.m.s.) deviation of the DNA structures along the trajectory with respect to the initial structure after equilibration are shown in Figure 2. DNA I stabilized only after 200 ps at a r.m.s. deviation of  $3.93 \pm 0.61$  Å (200–420 ps), whereas DNA III stabilized after only 100 ps at a r.m.s. deviation of  $3.76 \pm 0.34$  Å. Within the 500 ps simulation time presented in this work DNA II oscillates, but the average r.m.s. deviation is smallest at a value of  $3.08 \pm 0.51$  Å.

In general the backbone torsional angles of the three DNA sequences show that they belong to the B-DNA class. Similar results were obtained in previous simulations of a dodecamer with a different sequence (17), suggesting that the distortion introduced by the thymine dimer is local and rather small. With the exception of the backbone angles around the PD, the values of the torsional angles do not differ significantly among the sequences. Table 1 summarizes the average values and their fluctuations as measured by the standard deviation. Because some backbone torsional angles near the PD were significantly changed from the average they were not included in the averaging. The main change appears in the glycosidic dihedral angle  $\chi$  at PD5, which is larger by  $64^\circ$  than the average values of this angle in other sites. This places the conformation of the 5' thymine of the PD as high-*anti*, in good agreement with NMR results (28). The other difference in DNA III is observed in the pseudorotation angle. The changes are localized near the PD, showing values around C1'-*endo* ( $136.3 \pm 31.6$ ) for PD5 and near O4'-*endo* ( $80.1 \pm 21.2$ ) for PD6. Also, the sugar of the complementary adenine A14 is found mostly in the C1'-*endo* conformation.

Local conformational changes induced by the thymine dimer can be seen in the intrabase parameters. These describe the relative position of the two bases in a base pair and are expressed as stagger (STG), buckle (BKL), propeller twist (PRP) and opening (OPN) (27). These parameters are shown in Figure 3a–d. The negative STG of  $-1.3$  Å, the large positive BKL of  $25^\circ$  and the large negative PRP of  $-45^\circ$  are the consequence of formation of the cyclobutane ring that brings the thymines together on the side of the major groove. An important change is the increase in the OPN parameter near the PD. This parameter is increased to  $7^\circ$  and possibly reports breaking of hydrogen bonds between the adenines and the thymines of the PD (see below). The local



**Figure 3.** Intra-base parameters (a) stagger (STG; Å) (b) buckle angle (BKL; °), (c) propeller twist angle (PRP; °) and (d) opening angle (OPN; °) of each base step: 1 corresponds to base pair G1-C20 and the rest follow in sequence.

**Table 1.** Average backbone torsional angle ( $\pm$ SD)

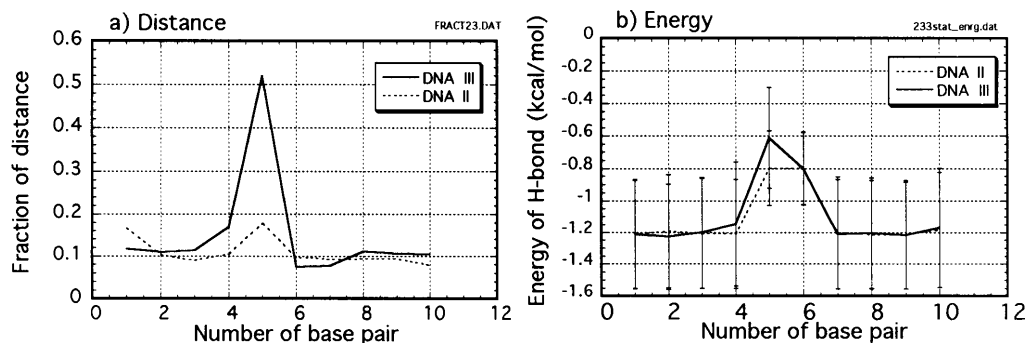
Angle	DNA I		DNA II		DNA III	
	ls	rs	ls	rs	ls	rs
$\delta$	104 $\pm$ 32	107 $\pm$ 33	108 $\pm$ 32	114 $\pm$ 32	111 $\pm$ 34	111 $\pm$ 34
$\epsilon$	180 $\pm$ 42	186 $\pm$ 49	177 $\pm$ 41	187 $\pm$ 50	195 $\pm$ 57	191 $\pm$ 55
$\zeta$	260 $\pm$ 59	257 $\pm$ 67	258 $\pm$ 59	243 $\pm$ 67	237 $\pm$ 77	243 $\pm$ 74
$\alpha$	278 $\pm$ 63	257 $\pm$ 67	279 $\pm$ 63	276 $\pm$ 63	276 $\pm$ 63	275 $\pm$ 63
$\beta$	164 $\pm$ 38	162 $\pm$ 39	164 $\pm$ 38	162 $\pm$ 39	160 $\pm$ 40	160 $\pm$ 40
$\gamma$	53 $\pm$ 16	52 $\pm$ 15	55 $\pm$ 16	52 $\pm$ 15	50 $\pm$ 17	51 $\pm$ 16
$\chi$	215 $\pm$ 51	219 $\pm$ 53	217 $\pm$ 51	223 $\pm$ 53	225 $\pm$ 57	224 $\pm$ 54
$\phi$	106 $\pm$ 43	111 $\pm$ 44	116 $\pm$ 45	122 $\pm$ 43	117 $\pm$ 43	113 $\pm$ 47

changes produced by the PD also affect interbase parameters which describe the relative shift (SHF), tilt (TLT) and roll (ROL) of one base pair with respect to the other. A large positive SHF of 0.85 Å at the PD step (the value at the TT step of DNA II is -0.14 Å), a large positive TLT of 18° (5° in DNA II) and a large positive ROL of 38° (3.8° in DNA II) were observed in DNA III.

Finally, these local changes lead to a kinked structure, which can be described by the axis base pair parameter. Significant changes were observed in the tip of the axis (TIP) and, in particular, in the polar angles of the axis with respect to the  $y$ - $z$  plane (AIN) (the long axis of DNA is positioned along the  $x$ -axis) and to the  $x$ - $z$  plane (ATP) of the PD5-A16 and PD6-A15 base pairs. When compared with DNA II the AIN was 8° larger and ATP 20° larger in DNA III. Thus the kink angle at the thymine dimer was estimated to be 21.7° in terms of vector sum of the two angles AIN and ATP. The present result is in agreement with our previous result that the average curvature magnitude increased by

21° at the PD site compared with a native dodecamer using a global DNA curvature analysis (17). Interestingly, approximately the same kink/bend angle was suggested in two oligonucleotides with different sequences, i.e. the decamer d(GCGGT<sup>^</sup>TGGCG)<sub>2</sub> in the present work and the dodecamer d(CGGAAT<sup>^</sup>TCGCG)<sub>2</sub> in our previous work (17). It seems that the distortion produced by the thymine dimer is a local property that does not depend on DNA sequence.

The structural changes described so far can be attributed to the steric distortion introduced by the thymine dimer. In a previous simulation of another sequence that contained a thymine dimer (17) we observed that one of the hydrogen bonds between adenine and the complementary PD on the 5'-side was significantly disrupted by the structural distortions in the DNA. This adenine has been observed to be flipped out by the protein in a complex between the repair enzyme endonuclease V and damaged DNA. We have therefore investigated the nature of the hydrogen bond disruption to gain a better understanding of the relation between weakening of the hydrogen bond and flipping of the adenine into an extra-helical position. We have calculated the distribution of distances between donor and acceptor of a hydrogen bond of each base pair in DNA II and DNA III along the trajectory. The fraction of hydrogen bond disruption was calculated as all the occurrences in which the hydrogen bond distance was larger than the average + 1 SD observed in DNA II. The results presented in Figure 4a show that on average 10% of the time hydrogen bonds are dissociated beyond the specified distance in normal hydrogen bonding pairs. However, near the thymine dimer 50% of the distances between PD5 and A16 in DNA III are greater than the distance of a normal hydrogen bond,



**Figure 4.** (a) Fraction of distances between hydrogen bond donor and acceptor atoms in DNA III and DNA II at each base pair step which are longer than the sum of the average and 1 SD of the distances at each base step of DNA II. (b) Potential energy component of the hydrogen bond estimated from the distances according to equation 2 (29).

that of T5-A16 in DNA II. The N(1)...H(3)-N(3) hydrogen bond of PD5-A16 is stretched by  $\sim 0.4$  Å on average and it was calculated that  $\sim 82\%$  of the time this hydrogen bond exceeded the specified distance. To estimate the energetic loss due to disruption of this hydrogen bond we calculated the average of the potential energy with a hydrogen bonding function of the form

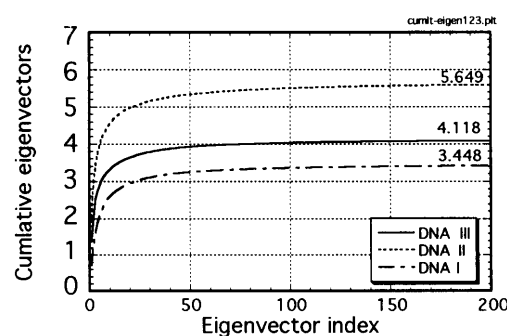
$$E = C/r^{12} - D/r^{10} \quad 2$$

where  $C = 7557$  and  $D = 2385$  for the H...O hydrogen bond and  $C = 10238$  and  $D = 3071$  for the H...NC hydrogen bond (29). The results are presented in Figure 4b. Clearly, the hydrogen bond between PD5 and A16 is weakened by 25% in DNA III as compared with DNA II. The potential energy of the N(1)...H(3)-N(3) hydrogen bond is smaller ( $-0.23$  kcal/mol) by  $\sim 56.8\%$  than that of other hydrogen bonds, which are  $\sim -0.40$  kcal/mol. The other hydrogen bonds of PD6-A15 were the same as those of the T5 and T6 steps of DNA II. These observations are in agreement with previous results (17) and suggest that weakening of the hydrogen bond may give rise to easier base flipping. These results do not provide evidence on whether this process can occur spontaneously, but it is clear that if the protein induces base flipping it should be easier to extrude a base whose hydrogen bonding has been weakened by the structural changes produced by the thymine dimer.

### Essential dynamics

The structural distortions in DNA III and partial disruption of the hydrogen bond between PD5 and its complementary adenine suggest that PD-containing DNA may have increased flexibility. We have therefore analyzed the three DNA sequences simulated in this work using the method of ED (18). In particular, we were interested in the question of whether the vibrational properties of the damaged DNA suggest a mechanism by which the adenine complementary to PD5 could flip into an extra-helical position.

ED can be used to analyze the concerted motions of macromolecules as well as their overall flexibility based on a MD trajectory. Concerted motions, in direction, among atoms are expressed by the eigenvectors of the covariance matrix and the extent of concerted fluctational motion described by each eigenvector is expressed by the eigenvalue. It has been shown that a large portion of the overall fluctuations of the macromolecule



**Figure 5.** Cumulative eigenvalues ( $\text{nm}^2$ ) of DNA I, DNA II and DNA III as a function of the eigenvector index.

can often be accounted for by a few low frequency eigenvectors (25) with large eigenvalues. The cumulative sum of the eigenvalues is a convenient way of expressing this property. The cumulative sum as a function of the number of eigenvalues derived from the ED analysis of the trajectories of all atoms except the hydrogens (404 atoms) of the three DNAs are shown in Figure 5. It can be seen that concerted motions specified by the first six eigenvectors can account for  $\sim 75\%$  of the overall fluctuations, which are expressed by the asymptotic cumulative values. Clearly, DNA II is the most flexible of the three, with an asymptotic value of  $5.649 \text{ nm}^2$ , followed by DNA III with a value of  $4.118 \text{ nm}^2$  and DNA I, which is the most rigid, with a value of  $3.448 \text{ nm}^2$ . The difference between DNA II and DNA I can be partly explained by the estimated base pair stacking energies (30). The estimated base pair stacking energies of the central 4 bp, 5'-GTTG (DNA II) and 5'-GATG (DNA I), were calculated as  $-18.52$  and  $-22.95$  kcal/mol respectively.

The reduced flexibility in DNA III is consistent with the introduction of a constraint in the form of a cyclobutane ring in the PD, but it does not provide a detailed description of the fluctuational properties along the DNA sequence or in the subgroups of the DNA, i.e. bases, sugars and phosphates. To analyze in detail concerted motions of the components along the sequence of the DNA we have calculated atomic fluctuations in the first two eigenvectors. The results are displayed in Figure 6, in which atom numbers 1–202 belong to the leading strand (in

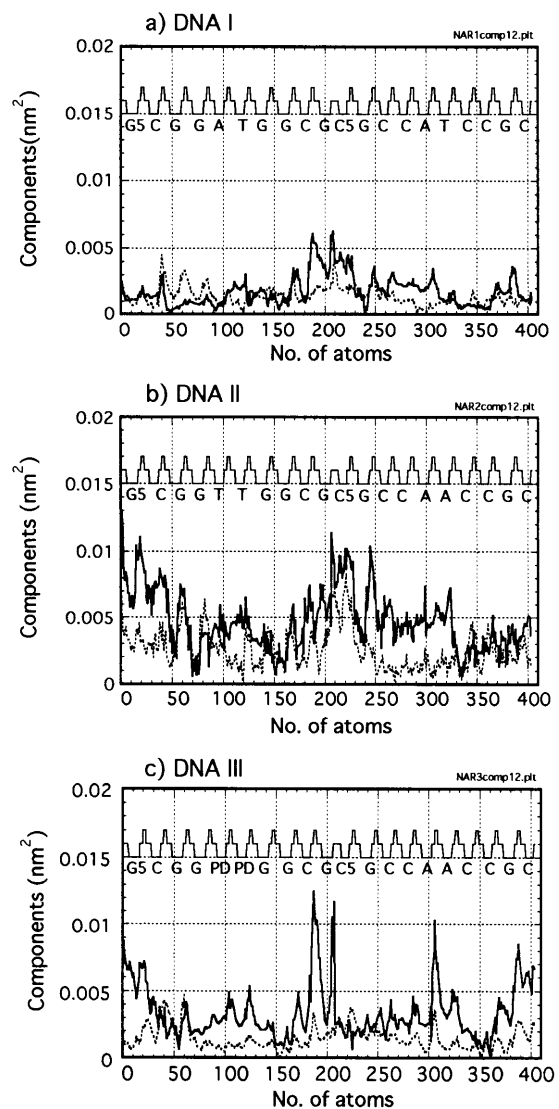
DNA III this strand contains the PD) and 203–404 to the complementary strand. Large fluctuations indicated by sharp peaks usually appear at the positions of phosphates, with much smaller fluctuations at the positions of sugars. Bases show the smallest fluctuations. DNA I (Fig. 6a) shows large concerted fluctuations, mostly at the end of the DNA duplex, suggesting a bending motion as well as a certain degree of fraying at the ends. Similar fluctuational features are seen in DNA II (Fig. 6b), but their size is larger than in DNA I. This is consistent with the fact that DNA II is more flexible than DNA I, as was demonstrated by the asymptotic values of the cumulative sums of the eigenvalues (Fig. 5). DNA III, in addition to a similar general appearance of large fluctuation at the ends, shows a new feature (Fig. 6c). The phosphate that bridges the adenines complementary to the PD (A15 and A16) shows a very large fluctuation in the first eigenvector, as indicated by the sharp peak at this position. This indicates that the fluctuation of the bridging phosphate is strongly correlated with the fluctuations at the ends of the DNA, suggesting coupling between movement of the phosphate and a bending motion.

To emphasize fluctuations at positions of phosphate we have applied ED to the 18 phosphates only. The atomic fluctuations in the first two eigenvectors are presented in Figure 7 for each base pair step. In the panels of Figure 7 the upper curves show the leading strand and the lower curves the complementary strand (to separate the curves the sign of the values in this strand has been changed). Differences in the patterns of concerted fluctuations between DNA I and DNA II suggests that even small differences in sequence of the base pairs (here the central part is AT or TT) may cause quite a difference in concerted motions. DNA II has a concerted motion at the 5'-end of both strands, while in DNA III this concerted motion appears at the 3'-end of both strands. Again, a sharp peak is observed at the phosphate between A15 and A16, suggesting that asymmetrical concerted fluctuations between each strand of the DNA III duplex may be produced by the thymine dimer and partially disrupted hydrogen bonding. It is possible that this flexibility may be important for recognition of the damaged site by the repair enzyme.

## DISCUSSION

This work represents an attempt to generalize the nature of DNA distortions produced by a cyclobutane pyrimidine dimer and to derive the changes in dynamic properties of the damaged DNA. The sequences of DNA studied in this work are different from those studied in our previous work (17) and by others (31), yet the distortions produced by the thymine dimer appear to be very similar. It appears that the thymine dimer produces changes in DNA that are localized to the vicinity of the lesion. They can be summarized by two major changes: one is bending of the DNA, the other is disruption of a hydrogen bond.

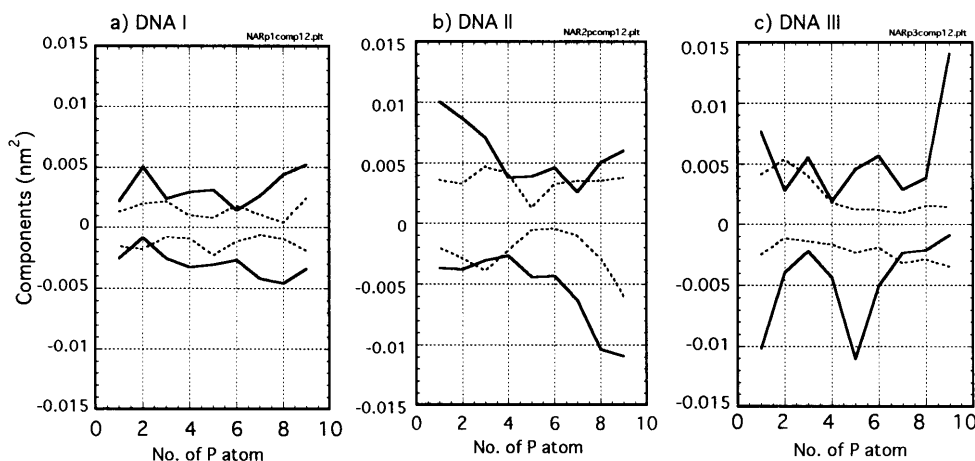
Bending of the DNA is clearly produced by formation of the cyclobutane ring, which draws one thymine to the other from a distance of  $\sim 3.5$  Å in normal DNA to a bonding distance of 1.5 Å. Such a structure also moves the O4 and N3-H3 groups which participate in hydrogen bonding to their complementary adenines closer to each other. The consequence of these changes is distortion of the helical axis. The extent of this bending is directional, as can be seen from the different values of AIN and ATP, which were  $8^\circ$  and  $20^\circ$  respectively. This difference is quite important, as was demonstrated in our previous work (17). We



**Figure 6.** Component ( $\text{nm}^2$ ) of concerted motions of 404 atoms of (a) DNA I, (b) DNA II and (c) DNA III along the directions of eigenvectors 1 (solid line) and 2 (dotted line). The zipper pattern shows the position of phosphorus atoms as a sharp peak, sugars the next lower shoulder and bases the bottom line. Numbers 1–202 belong to the leading strand (in DNA III this strand contains the PD) and 203–404 to the complementary strand.

have previously demonstrated that on average both normal DNA and DNA with a thymine dimer are bent. However, the directionality of bending in a thymine dimer-containing DNA is skewed, whereas that of normal DNA is not. Such behavior is fully consistent with formation of two C-C bonds in the major groove of DNA, giving rise to the observed directionality and to a small compression of the major groove and widening of the minor groove.

Disruption of the hydrogen bond is more difficult to understand. It is clear that the same hydrogen bond is disrupted regardless of sequence, confirming the suggestion that the changes are very local and do not depend on DNA sequence. Since formation of a cyclobutane would align the two thymines and virtually abolish the natural twist in a right-handed DNA,



**Figure 7.** Component (nm<sup>2</sup>) of concerted motions of phosphates in (a) DNA I, (b) DNA II and (c) DNA III along the directions of eigenvectors 1 (solid line) and 2 (dotted line). The numbers of P atoms are those of the leading strand. The component of the complementary P atoms are shown as negative values.

several small adjustments are made to compensate for this undesirable effect. First, the cyclobutane is puckered, which allows a certain degree of relative twist of one thymine with respect to the other. Second, the backbone angles shift by a small amount to absorb the distortion:  $\beta$  by  $\sim 10^\circ$  and  $\delta$  by  $\sim 20^\circ$ . The sugar ring also contributes, by changing its pucker from C3'-endo to C1'-endo and, finally, the base re-orient with respect to the sugar by changing its glycosidic angle from an *anti* to a high-*anti* conformation. These changes are required to re-establish the Watson-Crick hydrogen bonds, but the success is only partial. The O4...H-N6 bond is successfully restored to its original state, but the N3-H...N1 hydrogen bond remains partially disrupted. We estimate that the fraction of intact N3-H...N1 hydrogen bond distances is only  $\sim 20$ – $30\%$ , which is in agreement with the conclusions of Lee *et al.* (23), although it is difficult to estimate from their results a quantitative measure. Our estimate of the energetic loss associated with partial disruption of the hydrogen bond is of the order of 0.2 kcal/mol, which seems very small to be of major significance. However, it is noteworthy that such estimates are strongly dependent on the choice of the C and D constants in equation 2. Furthermore, the importance of partial disruption of the hydrogen bond is not so much in the loss of the energy *per se*, but in the change in vibrational properties of the DNA (see below).

ED has been used to analyze the collective motions of proteins (19–22), but no known studies have applied this method to the analysis of motions in DNA. Our ED analysis shows that different sequences of DNA have a different degree of flexibility. DNA I, with an AT sequence, is significantly more rigid than DNA II, with a TT sequence in the same position. This is consistent with a stronger stacking interaction between the bases, which may contribute to a more rigid structure in the AT sequence. It is also clear that introduction of a thymine dimer in place of the TT sequence reduces the overall flexibility of DNA III. In view of the fact that the cyclobutane introduces two additional bonds and limits the relative mobilities of the adjacent thymines, this result is not surprising. However, a detailed analysis of atomic fluctuations in the low frequency eigenvectors reveals a distinct difference between normal DNA and that with a thymine dimer. DNA III shows a distinct large fluctuation in the phosphate group

that bridges between the adenines on the complementary strand compared with that of the thymine dimer. It is difficult to determine whether this fluctuation is related to partial disruption of the hydrogen bond between PD5 and the complementary adenine. In fact, fluctuations of the base atoms in the first eigenvector are larger than those of other bases, but the connection between the base and the phosphate through the sugar precludes a causal relation. An important observation is that fluctuations in the low frequency eigenvectors are localized to the ends of the DNA sequences, suggesting that this eigenvector describes a bending motion. Furthermore, the fact that fluctuations of the phosphates in DNA III are coordinated with fluctuations of the ends of the DNA suggests that bending is coupled to a large phosphate motion.

Such an analysis leads to an interesting suggestion regarding the importance of such fluctuations in damage recognition by repair enzymes. The crystal structure of the complex between endonuclease V and DNA shows that DNA is bent near the thymine dimer more than in uncomplexed DNA observed in MD simulations. Thus the protein may be responsible for the additional bending observed in the crystal structure. Disruption of the hydrogen bond near the thymine dimer, while minimally reducing stability of the hydrogen bond, may reduce the force constant for bending by a substantial amount. In fact, the observation that the hydrogen bond is disrupted  $\sim 70\%$  of the time, i.e. the hydrogen bond distance is longer, supports the suggestion that the energy required for this stretching is lower. Since DNA bending is coupled to base pair opening and hydrogen bond disruption (32), it follows that in PD-containing DNA it will be easier to bend the DNA and induce extrusion of the adenine into an extra-helical position. Thus the protein is responsible for base flipping through induction of additional bending in the damaged DNA. This may also be the basis for selectivity of the protein in recognition of thymine dimers, as was recently demonstrated by fluorescence studies (33) of DNA with a 2-aminopurine substitution of the adenines in positions complementary to the thymines in thymine dimers. The data support the suggestion that the base opposite the 5' thymine but not that opposite the 3' thymine of the thymine dimer is flipped by the enzyme.

It appears that a general mechanism of selectivity emerges from these studies. Not only is bending of the DNA necessary for surface complementarity (17), but flexibility of the DNA is also important. Since the energy required for bending of undamaged DNA from an unbent conformation may be too large to be induced by the protein or the random distribution of DNA bending may present too many possible conformations to select from, the protein is unable to induce such a process in undamaged DNA. In PD-containing DNA all the required elements are in place: the DNA is bent in the correct direction and the hydrogen bond is partially disrupted, making bending and base flipping an easy process.

## ACKNOWLEDGEMENTS

H.Y is grateful to the Department of Physiology and Biophysics of Mount Sinai School of Medicine for their hospitality. Professor A.C.T.North of the University of Leeds is acknowledged for his help in ED analysis of DNA and Mr D.Akrigg for his arrangement of the computer environment at the University of Leeds. Mr S.Hongo and Mr H.Takeshita at the National Institute of Radiological Sciences are acknowledged for their help with the computational resources during this work. This work was supported by grants from the Science Technology Agency Japan, the British Council and NIH grant CA 66317 (to R.O.).

## REFERENCES

- 1 Yasuda,S. and Sekiguchi,M. (1970) *Proc. Natl. Acad. Sci. USA*, **67**, 1839–1845.
- 2 Gruskin,E.A. and Lloyd,R.S. (1986) *J. Biol. Chem.*, **261**, 9607–9613.
- 3 Dodson,M.L., Schrock,R.D.d. and Lloyd,R.S.(1993) *Biochemistry*, **32**, 8284–8290.
- 4 Manoharan,M., Mazumder,A., Ranson,S.C., Gerlt,J.A.and Bolton,P.H. (1988) *J. Am. Chem. Soc.*, **111**, 2690–2691.
- 5 Kim,J. and Linn,S. (1988) *Nucleic Acids Res.*, **16**, 1135–1141.
- 6 Doi,T., Recktenwald,A., Karaki,Y., Kikuchi,M., Morikawa,K., Ikehara,M., Inaoka,T., Hori,N. and Ohtsuka,E. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 9420–9424.
- 7 Dowd,D.R. and Lloyd,R.S. (1989) *Biochemistry*, **28**, 8699–8705.
- 8 Nyaga,S.G., Dodson,M.L. and Lloyd,R.S. (1997) *Biochemistry*, **36**, 4080–4088.
- 9 Morikawa,K., Matsumoto,O., Tsujimoto,M., Katayanagi,K., Ariyoshi,M., Doi,T., Ikehara,M., Inaoka,T. and Ohtsuka,E. (1992) *Science*, **256**, 523–526.
- 10 Morikawa,K., Ariyoshi,M., Vassylyev,D., Katayanagi,K., Nakamura,H., Doi,T., Hori,N. and Ohtsuka,E. (1994) *Ann. N.Y. Acad. Sci.*, **726**, 198–207.
- 11 Morikawa,K., Ariyoshi,M., Vassylyev,D.G., Matsumoto,O., Katayanagi,K. and Ohtsuka,E. (1995) *J. Mol. Biol.*, **249**, 360–375.
- 12 Vassylyev,D.G., Kashiwagi,T., Mikami,V., Ariyoshi,M., Iwai,S., Ohtsuka,E. and Morikawa,K. (1995) *Cell*, **83**, 773–782.
- 13 Iwai,S., Maeda,M., Shimada,Y., Hori,N., Murata,T., Morioka,H. and Ohtsuka,E. (1994) *Biochemistry*, **33**, 5581–5588.
- 14 Recinos,A.D. and Lloyd,R.S. (1988) *Biochemistry*, **27**, 1832–1838.
- 15 Stump,D.G. and Lloyd,R.S. (1988) *Biochemistry*, **27**, 1839–1843.
- 16 Latham,K.A., Carmical,J.R. and Lloyd,R.S. (1994) *Biochemistry*, **33**, 9024–9031.
- 17 Miaskiewicz,K., Miller,J., Coony,M. and Osman,R. (1996) *J. Am. Chem. Soc.*, **118**, 9156–9163.
- 18 Amadei,A., Linssen,A.B.M. and Berendsen,H.J.C. (1993) *Proteins*, **17**, 412–425.
- 19 van Aalten,D.M.F., de Groot,B.L., Berendsen,H.J.C. and Findlay,J.B. (1996) *Biochem. J.*, **319**, 543–550.
- 20 Amadei,A., Linssen,A.B., de Groot,B.L., van Aalten,D.M.F. and Berendsen,H.J.C. (1996) *J. Biomol. Struct. Dyn.*, **13**, 615–625.
- 21 de Groot,B.L., van Aalten,D.M.F., Amadei,A. and Berendsen,H.J.C. (1996) *Biophys. J.*, **71**, 1707–1713.
- 22 van Aalten,D.M.F., Findlay,J.B., Amadei,A. and Berendsen,H.J.C. (1995) *Protein Engng.*, **8**, 1129–1135.
- 23 Lee,B.J., Sakashita,H., Ohokubo,T., Ikehara,M., Doi,I., Morikawa,K., Kyogoku,T. Osafune,T. and Otsuka,E. (1994) *Biochemistry*, **33**, 57–64.
- 24 Pearlman,D.A., Case,D.A., Caldwell,J.W., Ross,W.S., Cheatham,T.E., Ferguson,D.M., Siebel,U.C., Weiner,P.K. and Kollman,P.A. (1995) *AMBER 4.1*. University of California, San Francisco, CA.
- 25 Teeler,M.M. and Case,D.A. (1990) *J. Phys. Chem.*, **94**, 8091–8097.
- 26 Vriend,G. (1990) *J. Mol. Graphics*, **8**, 52–56.
- 27 Ravishanker,G. and Beveridge,D.L. (1991) *Molecular Dynamics Analysis Toolchest*. Department of Chemistry, Wesleyan University, Middletown, CT.
- 28 Kemmink,J., Boelens,R., Koning,M.G., Kaptein,R., van der Marel,G., van Boom,J. (1987) *Eur. J. Biochem.*, **162**, 37–43.
- 29 Weinere,S.J., Kollman,P.A., Case,D.A., Chandra Singh,U., Chio,C., Alagona,G., Profeta,S., Jr and Weiner,P. (1984) *J. Am. Chem. Soc.*, **106**, 765–784.
- 30 Ornstein,R.L., Rein,R., Breen,D.L. and MacElroy,R.D. (1978) *Biopolymers*, **17**, 2341–2360.
- 31 Cooney,M.G. and Miller,J.H. (1997) *Nucleic Acids Res.*, **25**, 1432–1436.
- 32 Ramstein,J. and Lavery,R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 7231–7235.
- 33 McCullough,A.K., Dodson,M.L., Scharer,O.D. and Lloyd,R.S. (1997) *J. Biol. Chem.*, **272**, 27210–27217.