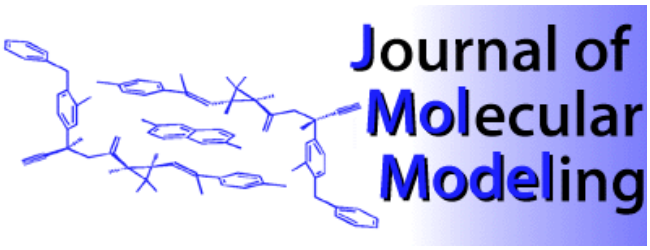
J. Mol. Model. **1999**, *5*, 302 – 316



FULL PAPER

Homology Modeling and Molecular Dynamics Simulations of PBCV-1 Glycosylase Complexed with UV-damaged DNA

Hongyao Zhu, Catherine H. Schein, and Werner Braun

Sealy Center for Structural Biology, Department of Human Biological Chemistry & Genetics, University of Texas Medical Branch Galveston, TX 77551-1157. Phone: (409) 747-6810; Fax: (409) 747-6850; E-mail: werner@newton.utmb.edu

Received: 01 July 1999/ Accepted: 11 October 1999/ Published: 10 December 1999

Abstract The UV-light damage specific DNA glycosylase from *Chlorella* virus strain PBCV-1 (pyrimidine dimer glycosylase; PDG) incises DNA at sites containing UV-induced thymidine dimers by catalyzing the breakage of the N-C-1' glycosyl bond. As the amino acid sequence of PDG is 41 % identical to that of T4 endonuclease V (Endo V), and potential key active site residues are conserved, we used coordinates from a crystal structure of an Endo V complexed with DNA containing a *cis-syn* cyclobutane thymidine-dimer as a template to model a similar complex of PDG. Quantum mechanical calculations of the damaged base pair and the distance geometry based program DIAMOD were used to generate a PDG/DNA model whose backbone root mean square deviation (RMSD) to the Endo V/DNA structure was 0.5 Å, 0.5 Å, and 0.8 Å for DNA, protein, and the whole complex, respectively. To better understand structural details that could account for differences in activity of the two enzymes, molecular dynamics simulations were used to follow protein-DNA interactions in an aqueous environment. The simulations of the Endo V/DNA complex indicate new roles for Arg22 and Arg26 in the active site in recognizing irregular pairing and maintaining the strand separation needed for incision of the damaged bases. The model for the PDG/DNA complex and simulations thereof indicate a similar mechanism for DNA binding by this enzyme despite significant differences in residues maintaining the flipped-out adenine and strand separation in the area of damage. According to our model, PDG's increased affinity for substrate is probably due to a higher surface charge. Further, reduced packing density in the active site could account for PDG's activity on trans-syn II cyclobutane dimers.

Keywords Endonuclease V, DNA glycosylases, Repair of damaged DNA, Flipped out base, DIAMOD, Protein-DNA recognition

Running Title 3D modeling of the PBCV-1 UV Glycosylase.

Correspondence to: W. Braun

Introduction

Better understanding of the enzymes involved in repair of damaged DNA is useful for developing treatments for diseases in which cellular repair mechanisms are absent or impaired.[1-3] Precise definition of their substrate binding mechanism can aid in designing inhibitors of these enzymes to use when their activity limits the effectiveness of chemotherapy.[4] Several enzymes have been identified that repair damaged bases by nucleotide excision pathways.[5] One of the most extensively studied of those catalyzing the first step in this pathway is the thymidine dimer specific endonuclease* from bacteriophage T4 (Endo V).[6-12] The enzymatic mechanism of base removal and strand cleavage has been worked out at the atomic level, based on extensive mutagenesis studies to document residues in its active site [7] and a crystal structure of the enzyme with its DNA substrate at below 3Å resolution.[11] Most of the residues indicated by mutagenesis to affect the activity of Endo V are located at the DNA/protein interface. As has been determined for several other DNA modifying enzymes,[13] Endo V "flips out" an (adenine) base from the ordered helical structure on the opposite strand from the damaged bases. This space in the DNA is occupied by protein residues which cleave the N-C-1' glycosidic bond joining the base to the sugar phosphate backbone.[8]

Recently, an enzyme with 41 % sequence identity to Endo V was found in extracts of *Chlorella* cells infected with the *Chlorella* virus strain, PBCV-1, and the cloned product shown to protect bacterial cells against UV radiation.[14,15] Residues important for the activity of Endo V are conserved in the *Chlorella* UV-glycosylase. The two recombinant enzymes, after similar purification from *Escherichia coli*, both recognize *cis-syn* cyclobutane pyrimidine dimers (PD) and effect glycolytic cleavage via formation of an imino (Schiff base) intermediate with damaged DNA that can be trapped by adding $NaBH_4$. They differ in that the PDG enzyme is also active on DNA substrates containing a trans-syn II PD, which is not recognized by Endo V, and that PDG binds its substrate better at a higher salt concentration.[16]

As PDG cleaves the same substrates and has high residue identity to Endo V particularly around the DNA binding site,[16] we can assume these enzymes share a basic catalytic mechanism and common protein fold.[17,18] We have developed a series of programs that use distance geometry to prepare 3D models for protein sequences, based on the known structure of homologous proteins. These have been used to develop a model of the measles virus receptor CD46 and identify residues that alter haemagglutinin binding.[19,20] In this paper, we use improved variants of these initial programs to calculate a model structure for PDG based on the crystal structure of Endo V. Molecular dynamics simulations of both the crystal structure of the Endo V complex and our model structure were then used to determine residues potentially responsible for the glycosylase's distinct substrate specificity. These "molecular movies" can offer tantalizing clues to how these enzymes establish their catalytic platform within damaged DNA.

Methods

Protein sequence alignment

The amino acid sequence of PDG from *Chlorella* virus [14] and Endo V (Glu23Gln) were aligned with the program CLUSTALW [21] (Figure 1). The amino acid sequence of PDG is approximately 41 % identical and 65 % similar to that of Endo V. Residues known to be in the active site of Endo V (particularly the N-terminal residue Thr2 and the area around Glu23) and areas critical for the recognition of the cyclobutane PD are highly conserved in PDG. In this paper, the sequence numbering system for PDG/DNA is the same as that for the Endo V complex. Two additional residues in PDG have been numbered Lys45' and Val86'. DNA is numbered from 5' to 3', with the strand containing the PD numbered 201-213 and the opposite strand 214-226. Nucleic acids are referred to by the one-letter code followed by the residue number (e.g. A221 is the "flipped out" adenine).

Homology modeling of the PDG/DNA complex

We used the crystal structure of Endo V Glu23Gln mutant complexed with a duplex DNA substrate containing a cis-syn cylcobutane PD determined at 2.75 Å resolution (entry 1vas in the Protein Data Bank [22]) as a template structure. Our distance geometry based program DIAMOD [23,24] was used to calculate the homology model of PDG.

Distance and dihedral angle constraints were extracted from the x-ray crystal coordinates (after regularization of the structure, meaning all bond lengths and angles were converted to the standard values used in the library of DIAMOD). For the DNA, distance constraints were extracted between all the heavy atoms of a base and the heavy atoms in those bases

* This enzyme and PDG are more properly called glycosylases with AP-lyase activity; both excise the damaged base pairs in a two step mechanism that begins with cleavage of the N-C-1' bond of the base to the sugar backbone, in a reaction with a Schiff base intermediate and subsequent nucleophilic cleavage of the phosphate backbone.

** Abbreviations used: pyrimidine dimer (PD); *Chlorella* UV-glycosylase (PDG); thymidine dimer specific endonuclease from bacteriophage T4 (Endo V); Molecular dynamics (MD), radius of gyration (RG); root mean square deviation (RMSD)

which are equal or more than 4 bases away within each strand and between P, C_4, C_1', and C_4' of each base pair between strands. For the pyrimidine dimer PD**, the distances between the C_5 and C_6 of one thymine and O_2, O_4, C_5, and C_6 of the other thymine were extracted. The x-ray coordinates of the PD in the crystal structure of Endo V were added to the library of DIAMOD so as to reproduce that local conformation in our model.

Additional distance constraints were extracted from the Endo V coordinates for the three regions of high identity in the two proteins: Thr2-Arg36 (51 %), Ile46-Ile86 (48%), and Asp103-Ala126 (50 %). The resulting distance constraints include the C_α-C_α distances for residues 2 or more apart and the C_β-C_β distances between residues 3 or more apart in each of the three regions, and the C_α-C_α distances between residues in different regions which are identical in the two proteins. Distances less than or equal to 16 Å between the P atoms of the DNA and C_α atoms of the protein within the three regions of the PDG sequence were used to constrain the relative position of protein and DNA substrate in the complex. The resulting approximately 6000 distance constraints were used to set the limits of permissible distances between the residues involved. In addition, angle constraints were extracted for each strand of the DNA substrate and those three regions of the protein. The constraint ranges for the dihedral angles were 4° and 10° for the DNA and the protein, respectively.

DIAMOD distance geometry calculations in torsion angle space were carried out on a SGI/R10000 workstation and a Cray J90. A modified target function for a violated distance constraint was constructed to prevent large violations of distance constraints from dominating the non-linear fit procedure.[23] Finally, 50 PDG/DNA structures were calculated from random initial conformations. The minimizations were performed with 37 levels for the variable target function,[25,26] where the final minimization level covers all distance constraints.

Molecular dynamics simulations

The model resulting from the DIAMOD calculations was then energy minimized with the CHARMM polar hydrogen force field parameters [27] in vacuum for 500 steps, keeping back-

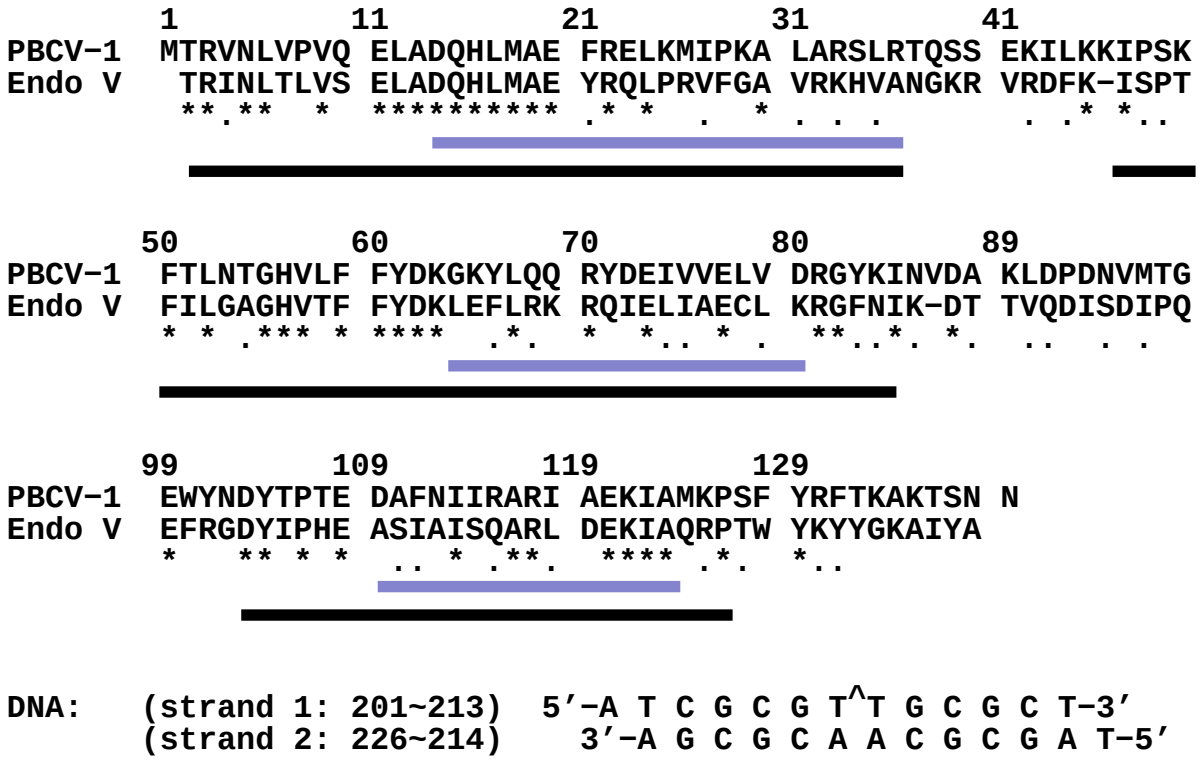


Figure 1 *Sequence alignment of Endo V and PDG indicating identical (stars) and similar (dots) residues (as used in Clustal W). The homology modeling was based on three areas of high identity between the two proteins (solid bars). Three α-helical areas in Endo V are indicated as gray bars. The sequence of PDG was numbered so as to agree with Endo V, with two additional residues in PDG labeled Lys45' and Val86'. The DNA strands (bottom) are numbered from 5' to 3', as shown, and the position of the cyclobutane cis-syn thymidine dimer is indicated by ^.*

bone atoms fixed in position, followed by another 500 steps after restraints were removed. The stereochemical quality of the structure was checked with PROCHECK [28] . The backbone dihedral angles of most residues (98.4 %) in the protein were within the allowed regions of the Ramachandran plot, the peptide bond planarity, the number of bad non-bonded interactions, the α-carbon tetrahedral distortions and the hydrogen bond energies were all within standard deviations of a structure with assumed resolution of 2 Å. This structure was used for subsequent MD simulation in a water environment. The PDG/DNA complex was solvated first in a cubic box of water. Water molecules with close contacts to the solute atoms (within 2.8 Å) were excluded from the solution. All water molecules beyond 10 Å from the heavy solute atoms were deleted. The initial configuration of the system consisted of 2015 atoms from the protein/DNA complex and 2485 water molecules.

The X-PLOR program [29] with the CHARMM polar hydrogen force field parameters [27] was used for molecular dynamics simulations of the solvated PDG/DNA complex. For the *cis-syn* UV-induced PD, modified atom types, partial charges, and force field parameters were prepared for the simulations. The geometry of a N-methanolated model molecule with PD conformation was optimized with AM1 [30] using MOPAC93 [31] followed by Hartree-Fock calculations with the 6-31G* basis set. [32] All revised parameters for the cis-syn cyclobutane PD are given in Table 1. Water molecules were treated as TIP3P residues.[33]

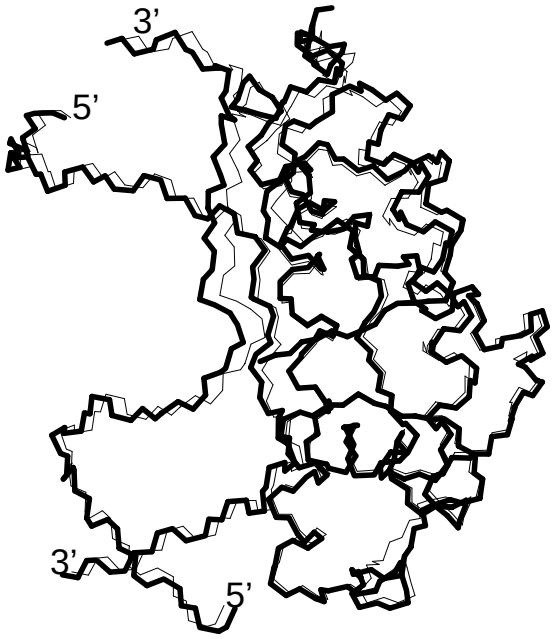


Figure 2 *Protein and DNA backbone superposition of the PDG/DNA model obtained from homology modeling using DIAMOD calculations (thick line) on the Endo V/DNA crystal structure (thin line)*

Table 1 *Geometry and revised force field parameters for the* cis-syn *cyclobutane pyrimidine dimer (PD) used in the energy minimization and MD simulations. Atom partial charges (e) were: N1:-0.11; C4: 0.30; C5: -0.07; C5A: 0.07; C6: 0.11. The PD geometry was obtained from* ab initio *optimization with the 6-31G* basis set for the N-methanolated model system. The atom types, partial charges and force constants were adopted from the CHARMM polar hydrogen force field.*

Angles	k_a **(kcal·mol^{-1}·deg^{-2})**	θ **(deg)**
C5-C4-C5A	116	108.8
C5-C5-C5A	116	113.8
C5-C6-C5A	116	113.5
C4-C5-C5	116	114.1
C4-C5-C6	116	115.5
C5-C4-N3	116	118.7
C5-C4-O4	86	121.4
C5-C5-C6	116	90.5
C5-C6-C6	116	89.5
C5-C6-N1	116	115.7
C6-N1-C1'	70	115.2
C4-N1-C6	70	125.3
C6-C6-N1	116	114.8
Bonds	k_b **(kcal·mol^{-1}·Å^{-2})**	r_0 **(Å)**
C5-C5A	201	1.53
C4-C5	201	1.51
C5-C5	201	1.58
C5-C6	201	1.55
C6-C6	201	1.58
C6-N1	300	1.45

The assemblies were minimized prior to dynamics runs to relax the local strain in the initial configurations. Five hundred steps of conjugated gradient minimization were performed with all heavy atoms fixed, and then 500 minimization steps followed without constraints. Nonbonding interactions were cut off at 10 Å. To prevent water evaporating during the MD simulation runs, a harmonic potential with a 2.0 kcal mol^{-1} force constant was used to restrain the water molecules in the 4 Å outer shell. All simulations used a 0.001 ps time step and assumed a constant dielectric of 1.0. All hydrogen-involved bonds were constrained by the SHAKE algorithm [34] with a tolerance of 1.0×10^{-6} nm. The MD simulations on the entire system started with assigned random velocities and gradually warmed up to 300 K using 20 K temperature steps of 30 ps. The temperature was held by means of the Berendsen coupling algorithm [35] with a coupling constant of 0.2 ps^{-1}. An additional 50 ps equilibration followed to stabilize the system. During heating and equilibration runs for PDG/DNA system, H-bonding distance con-

Table 2 *RMSD and potential energy values for the initial DIAMOD generated model and the dynamic average conformation of the PDG/DNA complex*

	Model [a]	Dynamic-mean [b]
RMSD (Å) [c]		
Backbone atoms	0.4	3.4
Heavy atoms	0.7	3.9
Energy (kcal·mol^{-1})		
Total	-1334.6	-1418.7
Bond	28.2	29.0
Angle	322.9	274.7
Dihedral	383.5	335.7
Improper	36.5	42.1
VDW	-2081.4	-2085.5
Electrostatic[d]	-24.3	-32.8

[a] Energy-minimized conformation of the model structure from DIAMOD calculations.
[b] Energy-minimized conformation averaged over all structures between 150 and 300 ps.
[c] The reference conformation was the model structure of the PDG/DNA complex.
[d] A dielectric constant of 80 was used for energy minimization.

straints for base pairs in DNA were applied with square-well potential at 2.0 ± 0.3 Å, and then were removed during the MD simulations. No such constraints were used during simulations of the Endo V/DNA complex. A 300 ps simulation was performed at 300 K as production runs. Configurations were recorded every 0.5 ps.

For MD simulations of the PD-DNA alone in an aqueous environment, the starting configuration was that in the Endo V crystal complex, with the A221 flipped out. A 15 Å water shell with 5 Å outer layer of water molecules constrained was used for the 500 ps simulations at 300 K. Configurations were recorded every 1 ps.

The dynamic averaged structures for both complexes were obtained by averaging the coordinates of the structure in all frames between 150 and 300 ps, followed by 200 energy minimization steps with all heavy atoms fixed, 200 steps with backbone atoms fixed, and 1000 steps with no constraints. In the dynamic averaged structures all water molecules were removed and a constant dielectric of 80 was used for the energy minimizations.

All RMSD values were calculated with MOLMOL [36], which was also used to create the molecular graphics presented in this paper.

Table 3 *Positively charged protein side chains in the dynamic averaged structures that are closer than 6.5 Å to the negatively charged DNA phosphate backbone*

Endo	DNA	PDG	DNA
Arg3	T208	Arg3	C209
Arg22	T207	Arg22	A221
Arg26	A220	Lys29	G211
Lys33	C212		C212
Lys86	C222	Arg33	G211
Lys121	T207	Lys84	G224
	T208	Arg117	T207
Arg125	A226	Lys121	G206
Lys130	G225	Lys125	A226
		Arg131	G225
		Lys133	V224
			G225

Results

Model structure of PDG/PD containing DNA

Well defined structures were obtained from DIAMOD calculations with the homology-derived distance and angle constraints. The mean global RMSD for the narrow bundle of 40 of the 50 structures was 1.3 Å for all heavy atoms (0.5 Å if only backbone atoms are considered). The backbone RMSD of the PDG/DNA structure to the Endo V/DNA are 0.5 Å, 0.5 Å, and 0.8 Å for DNA, protein, and the whole complex, respectively. The structure with the lowest target function value in the bundle of models was chosen for molecular dynamics simulations. Figure 2 shows the backbone superposition of this PDG/DNA model with the Endo V/DNA crystal structure. The three helical segments are nearly identical in the PDG/DNA model and the Endo V/DNA crystal structure, and the PDG model is particularly similar to the Endo V crystal structure in the helix-loop-helix between Thr37-Lys60. Some variations between the model and the template are observed in the less conserved regions, especially around the two insertions Lys45' and Val86' (see Figure 1).

Global shape of the simulated structures

MD simulations (300 ps) using X-PLOR started from the DIAMOD generated model of PDG/DNA, or from the crystal coordinates for the Endo V/DNA complex. The initial configuration of the PDG/DNA with a 10 Å water shell is shown in Figure 3. Backbone RMSD plots for the complete MD simulations of both complexes are shown in Figure 4a. In both simulations the RMSDs maintained a steady average value (i.e., reached a stable structure) only after 100 ps, comparable to results with unrestrained MD of photodamaged

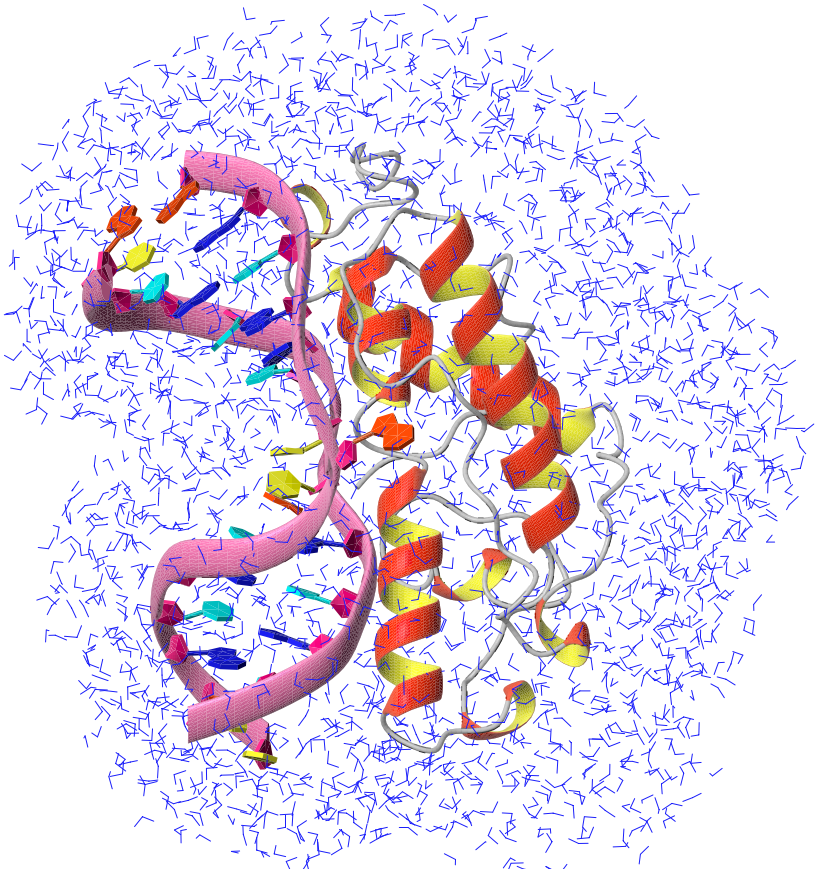


Figure 3 *The initial configuration of the PDG/DNA complex (model structure surrounded by a ~10 Å water shell) used for the molecular dynamics simulation*

DNA in aqueous solution [37] using the Cornell et al. [38] force field. The steady state average RMSD backbone values to the starting conformations for Endo V/DNA and PDG/DNA were 2.6 Å and 3.4 Å respectively. The higher relative deviation of the PDG complex from its initial structure during the simulation suggests that some protein/DNA contacts may be sub-optimal in the initial model. Table 2 summarizes the RMSD and energy values for the initial and dynamic averaged structures of the PDG complex.

Figure 4b shows the dynamic fluctuations of the radius of gyration (RG) averaged over all heavy atoms of each complex. The RG values for both complexes during the simulations are similar, indicating that the PDG/DNA model has a global shape similar to that of the Endo V/DNA complex. The dynamic RG values for Endo V/DNA are 0.2 Å higher than for the corresponding X-ray structure, which is in keeping with results for MD simulations of other proteins.[39]

Charge distributions on the surface of PDG

Figure 5 shows the charge distributions on the surface of the dynamic averaged structures for Endo V and PDG. The lack of the negatively charged group near the PD in the Endo V complex is due to the Glu23Gln mutation. The higher surface positive charge (indicated by blue color) in PDG is evident. As Table 3 shows, there are more electrostatic interactions less than 6.5 Å between the protein and the negatively charged DNA phosphate backbone in the PDG complex than in Endo V. Most of these interactions are near the C terminus of the protein. Our model indicates that the PDG complex should have higher salt stability than the Endo V one.

Major areas of contact between protein and DNA

The MD simulations allow us to categorize the major interactions between the protein and DNA in stable and transient interactions. Protein residues lying less than 8 Å from the DNA in both complexes for the initial and dynamics averaged conformations were determined using MOLMOL [36] (Figure 6). For the Endo V/DNA, most of the short contacts seen in the crystal structure are conserved during the MD simulations (Figure 6a, left and right); these include amino acids inferred by mutational analysis to participate in DNA binding and catalysis as discussed below. The four major areas of the protein close to the DNA (for Endo V, residues Thr2, Arg3 and Gln15-Arg26, Lys33, His34, Gly55-Tyr61, Gln71, Gly82-Gln91 and Gln124-Lys130) are similarly located in the PDG complex's initial and dynamic averaged structure (Figure 6b, left and right).

The dynamic averaged structures (right) indicate far more close contacts between DNA and protein than the crystal or model structures (left) do. There are 602 contacts within 3.5Å from the DNA in the Endo V crystal structure, 435 for the PDG complex model, compared to 1023 and 1102 respec-

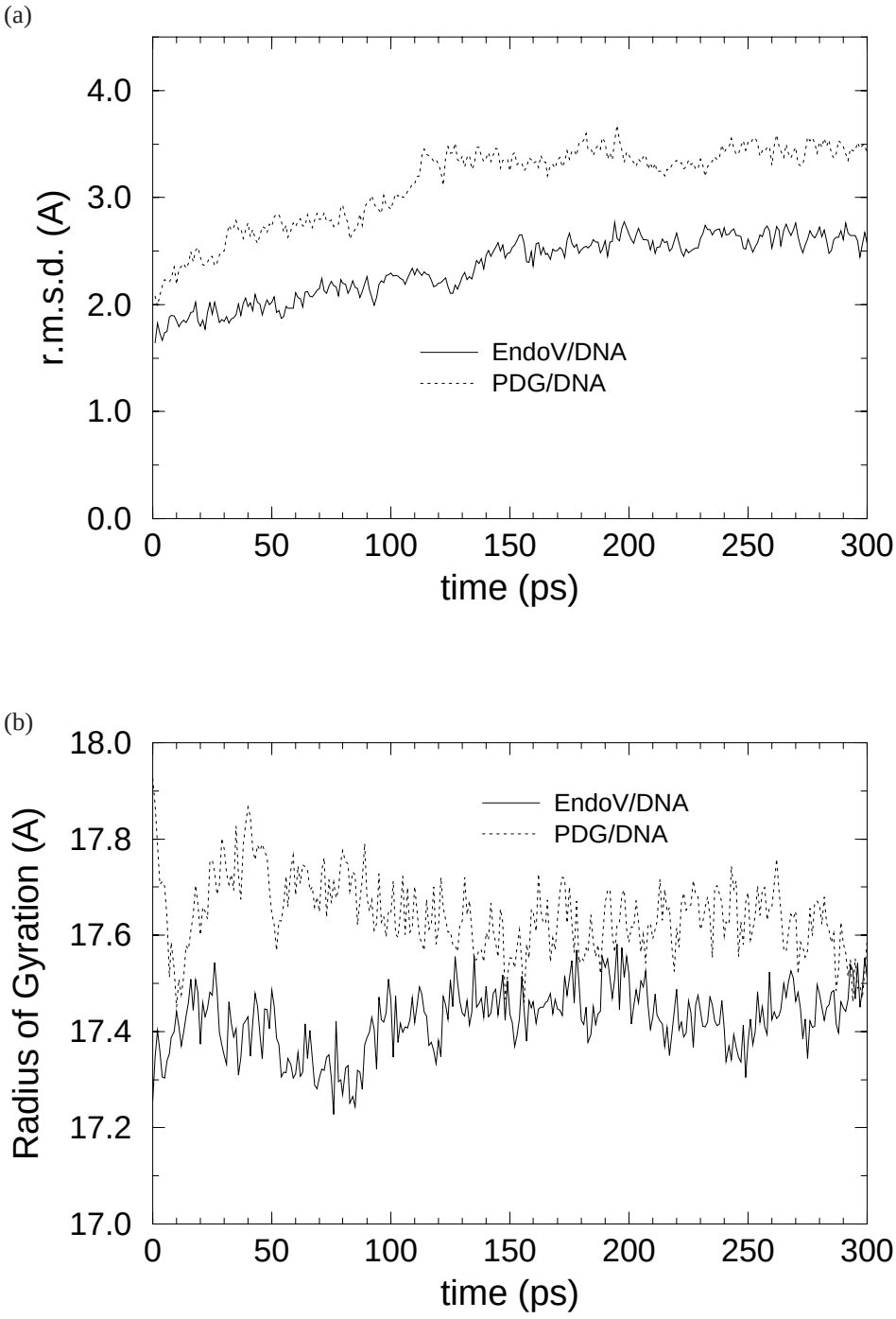


Figure 4 *Fluctuations in (a) the backbone RMSD and (b) the radius of gyration (RG) during the 300 ps MD trajectories for the Endo V and PDG complexes*

tively in the dynamic averaged structures. The time course of formation of some of these contacts throughout the simulations, a measure of their stability, and their possible role in establishing a transition state structure for each enzyme are discussed below.

Base pairing adjacent to the pyrimidine dimer

Figure 7 shows the distances between atoms in the CG base pairs adjacent to the PD which should be involved in Watson-Crick hydrogen bonds in normal B-DNA as a function of simulation time. The CG base pair to the 5' end of the dimer (Figure7, first two columns) is much more unstable in the

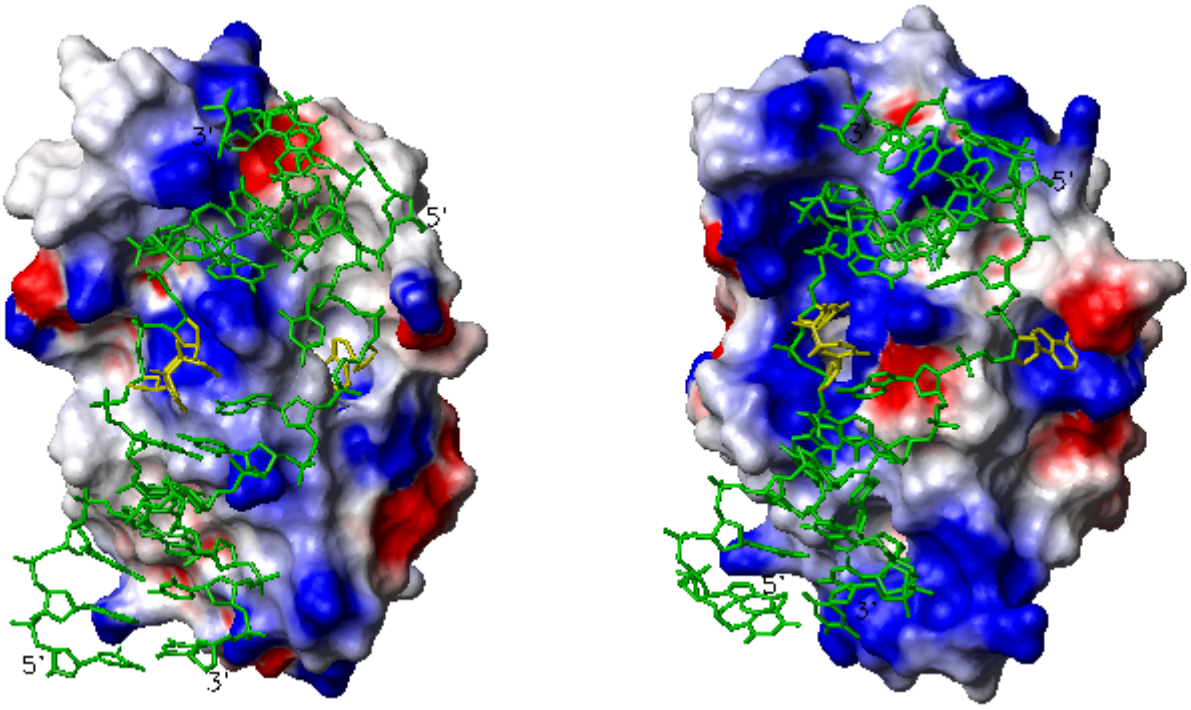


Figure 5 *Charge distributions on the surfaces of the dynamic averaged structures for Endo V/DNA (left), and PDG/DNA (right). Residues with positive (Arg and Lys) and negative (Asp and Glu) charges are shown in blue and red respectively. The PD and flipped out bases are yellow. Note that due to mutation of Glu23 to Gln in the crystal structure of Endo V/DNA, there is no peak of negative charge near the PD.*

Endo V complex (Figure 7a) than in the PDG model (Figure 7b). In both complexes, the base pairs 3' to the catalytic complex are also disturbed, and the disturbance seems to propagate through to the next base pair (Figure 7 a and b, last two columns).

Detailed inspection of the MD simulation run identified the side chains of Gln 15, Met18 and Arg22 which could disturb the G206-C222 pairing at the 5' end of the dimer. The guanidinium moiety of the Arg26 side chain extends in the opposite direction and, with Thr2, interferes with the base pairing of G209-C219 base pair and possibly the subsequent C-G pair (last columns of Figure 7a). In PDG, Arg22 forms less contacts with the G206-C222 pair, which is more stable than in the Endo V structure (Figure 7b, first columns). Met26 lies outside the cleft, closer to the flipped out adenine, and the base pairing 3' to the PD is less disturbed (Figure 7b, last columns).

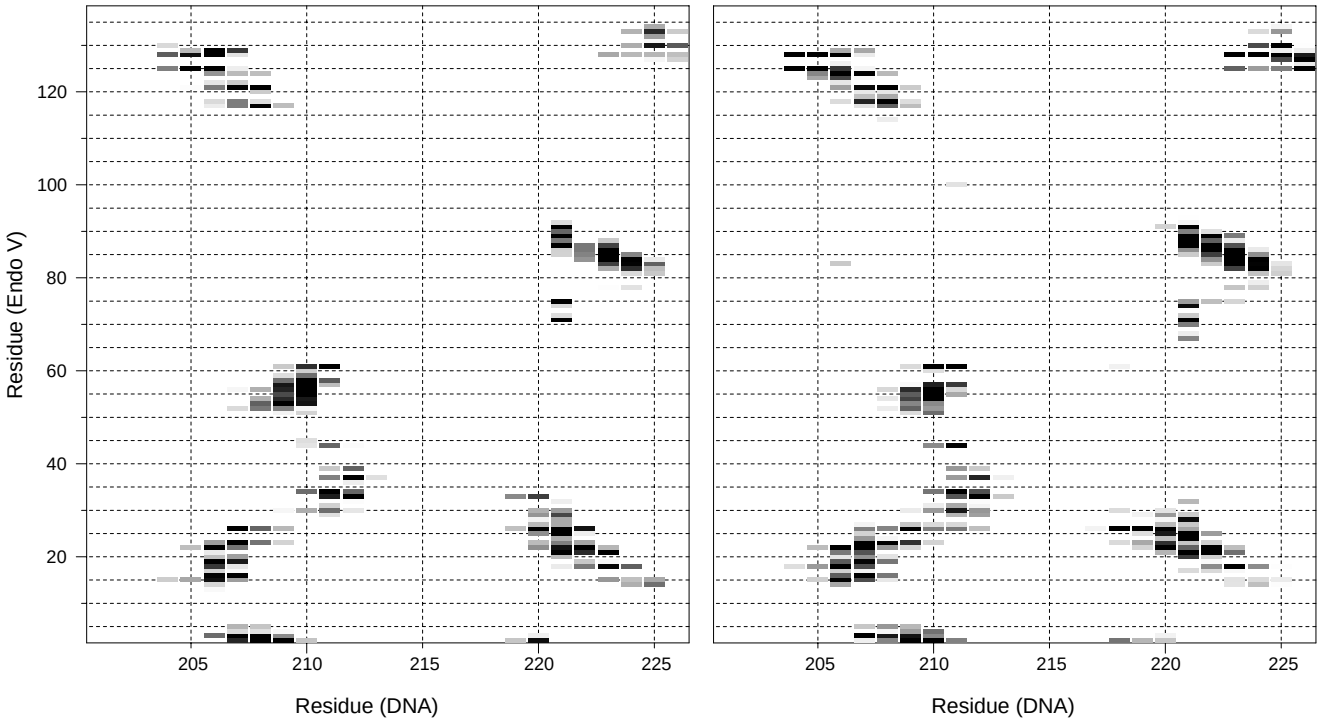
Residues in the pocket for the flipped out base

In Figure 8 we monitor some important interactions between the extra-helical base A221 and protein residues during the dynamics run. Most of the stable interactions are hydrophobic in both complexes, but we also could identify some stable hydrogen bonds in the PDG/DNA complex. Both Tyr21/Phe21 in Endo V/PDG are near the base A221, but whereas Phe21 (in PDG) is in close van der Waals contact with the base during a substantial time, the hydroxyl group of Tyr21 does not form an H-bond in Endo V. The non-polar interaction of Pro25 in Endo V has been assumed by the aliphatic arm of Lys25 in PDG. A221 lies within a few angstrom of Gln71, Thr89 and Gln91 in Endo V. The terminal amides of the two glutamines come alternatively close to the ring (Figure 8a) but do not establish stable hydrogen bonding.

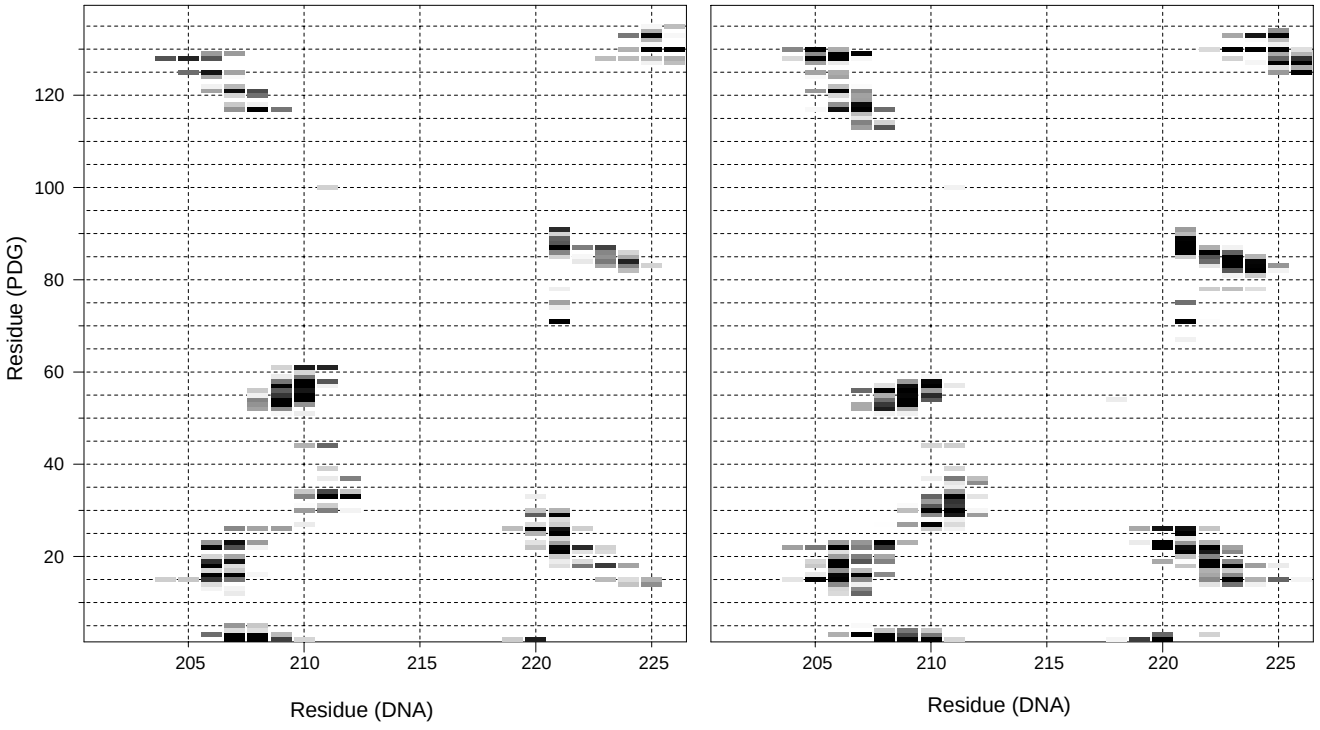
In the PDG model, A221 forms a stable H-bond with the hydroxyl group of Tyr71, and the side chains of Asn86, Asp87 and Lys25 are all within 3.5Å of the base (Figure 8b). Our model predicts that the extra-helical base in the PDG model is less mobile than in the Endo V complex. Figure 9a illustrates the polar environment of the extra-helical base A221 in our PDG model.

Active site dynamics

Amino acids involved in recognition and catalytic function in both enzymes Thr2, Arg3, His16, Glu20, Arg22, Glu23, lie within the pocket of the DNA near the PD as shown in Figures 9b and 9c. In both complexes, His16, which is a possible proton donor in the lyase reaction (His16Arg mutants of Endo V have enhanced non-target DNA binding but diminished catalytic activity [40]), remains behind the phosphate backbone at T207. The non-flipped base A220 is able



a



b

Figure 6 *Protein side chains lying within 8.0 Å of the DNA for: (a) Endo V, and (b) PDG; plots (generated with MOLMOL) to the left are for the X-ray or model structure, respectively, and to the right for the dynamic averaged structures (obtained as described in Methods). Distances <3.5 Å are indicated with black boxes; the boxes become progressively lighter with increasing distance.*

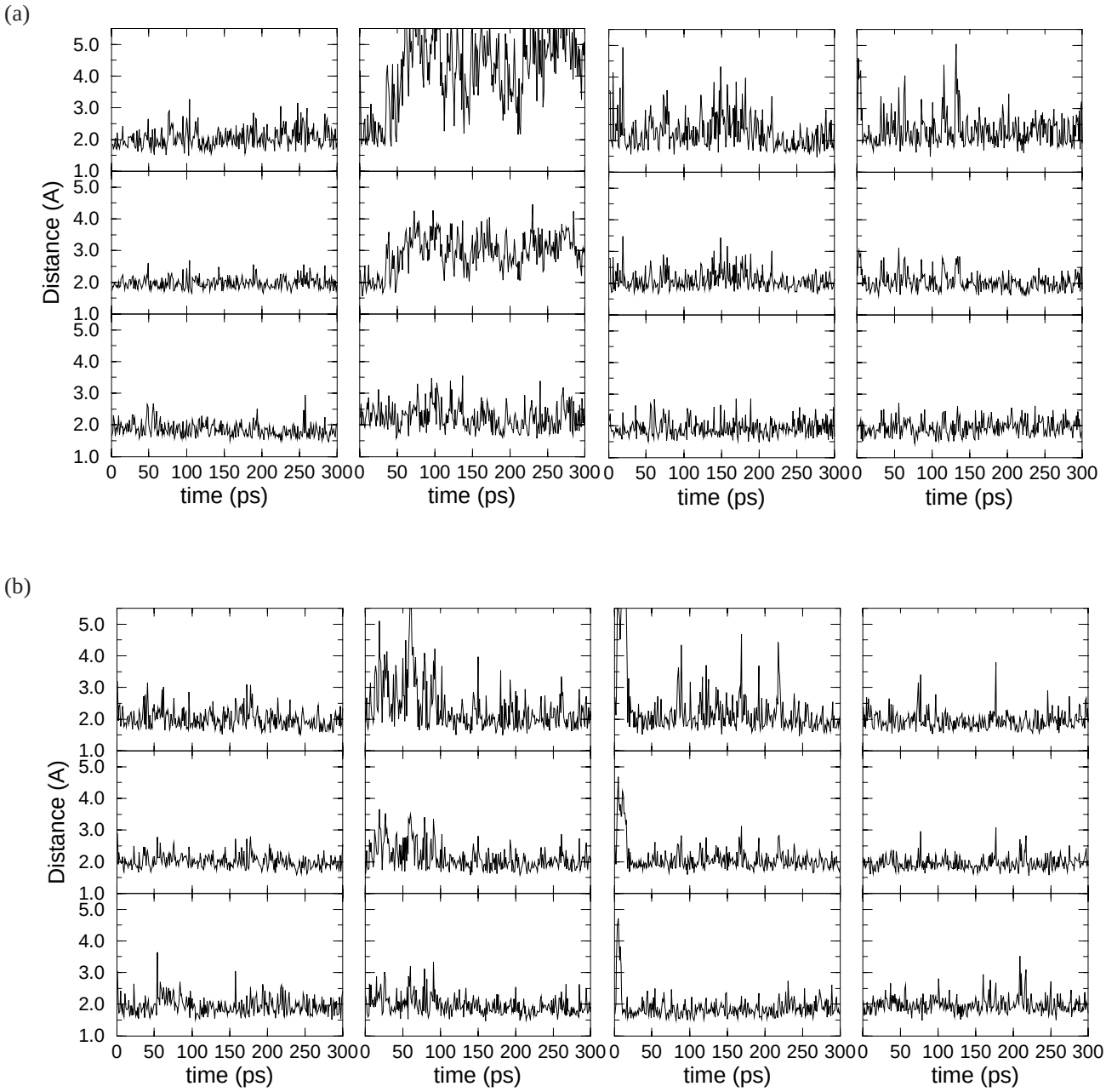


Figure 7 *Distances between H-bond pairs (from top to bottom: O6···H42, H1···N3, and H22···O2) as a function of time during the MD simulation for the CG pairs (from left to right: C205:G223, G206:C222; G209:C219, C210:G218) at both sides of the PD for (a) Endo V/DNA and (b) PDG/DNA.*

to pair with either T207 or T208 (Figures 9b, 9c), but shows greater flexibility in the PDG complex. This suggests that packing density near the PD is less in the PDG complex than in Endo V, and may account for PDG's ability to accept substrates containing a *trans-syn* II cyclobutane dimer.

The main difference in amino acid residues near the PD is the replacement of Pro25-Arg26 in Endo V to Lys25-Met26 in PDG. In the active site of PDG (Figure9c) Met26, at the equivalent sequence position of Arg26 in Endo V, is close to A220. Many of the interactions of Arg26 with the DNA backbone in Endo V are assumed by Gln15 and Arg22. As the simulation progresses, the Met26 side chain moves away from the active site. Another difference between Endo V and PDG near the active site is due to the substitution Glu23Gln in Endo V that was necessary to obtain stable co-crystals. Our simulations indicate two H-bonds at the active site in the PDG model to Glu23 which forms 2 stable H-bonds with Arg3 (OE3–HH22) and Arg22 (OE1–HH11) (Figure 9c).

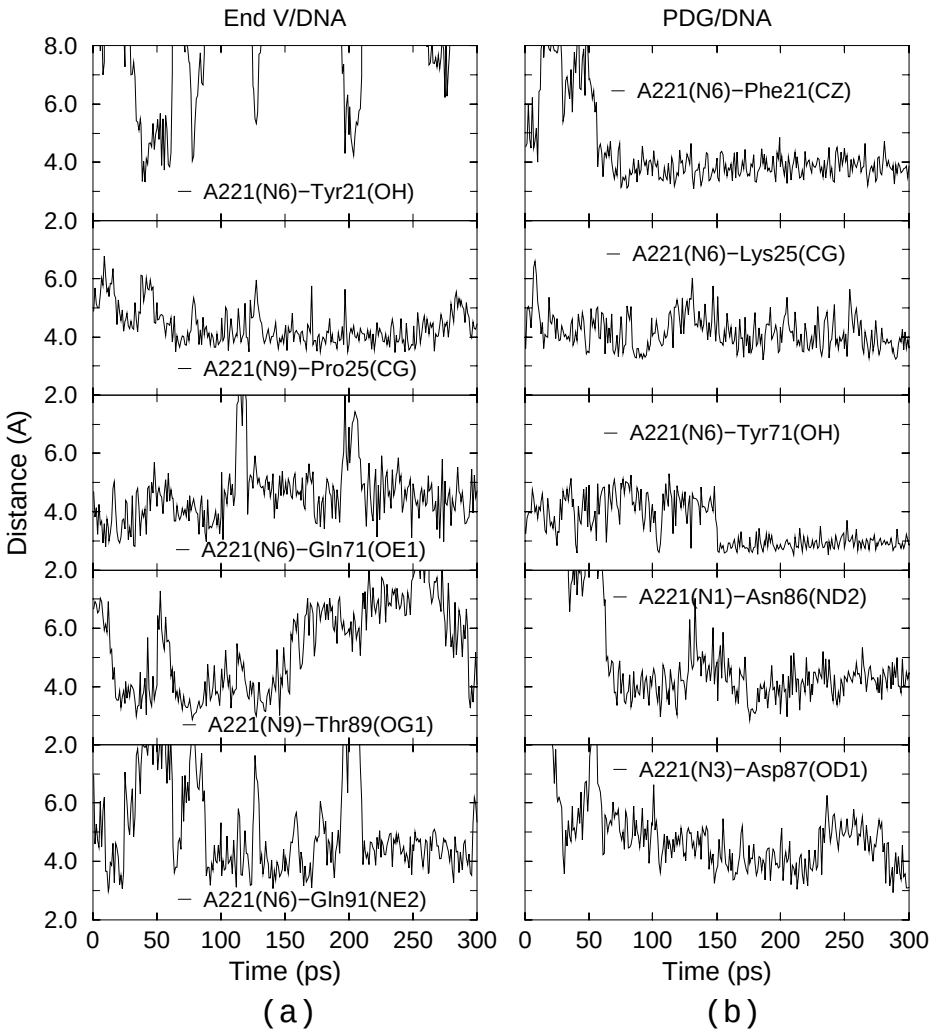


Figure 8 *Distances as a function of simulation time between atoms of protein residues and the flipped out A221 base in the (a) Endo V/DNA and (b) PDG/DNA complexes*

A hydrophobic region near the C-terminus is closer to the DNA in the dynamic averaged structures

Glycosylases and other repair enzymes scan DNA [41] in a fashion similar to that described in detail for restriction enzymes.[42] Logically, residues involved in binding non-target DNA to initiate scanning will be located at the protein surface. We used the GETAREA program [43, 44], developed in this group to locate three major areas with side chains having a high degree of surface exposure in the crystal structures of the free Endo V protein and three mutants. These were: many (but not all) positively charged residues within the first 45 amino acids, Asn84, Lys86, and a 12 amino acid stretch around Trp128 that contains many aromatic residues. As discussed above, many of these surface exposed residues form specific DNA contacts with the active site. On the other hand, the residues surrounding the flipped out base in the complex have little or no surface exposure in the free protein.

Previous NMR and mutational studies of Endo V indicated that the aromatic region around Trp128 was involved in binding both non-target and substrate DNA.[45] While the crystal structure showed this area to be relatively distant from the DNA, the dynamic averaged structure for the Endo V complex indicates 46 possible contacts less than 3.5Å from Arg125; Trp128 lies near 35 different atoms of G204, C205, G206, G223, C224, G225 and A226. In the PDG complex average structure, this area of the protein (the side chains of Lys125, Ser127, Tyr129, and Arg130) is near the same areas of the DNA (Figure 9d). There are 53 possible close DNA contacts for Arg130 alone, 5 % of those for the whole protein. The side chains of Arg117, Lys121 (Figure 9c), and Tyr129 (Figure 9d) lie close to the phosphates behind the PD, further stabilizing the DNA/protein interaction.

Discussion

The primary result of this work is a model, based on sequence identity to Endo V, for PDG binding to damaged DNA (Figure 2), that is stable throughout dynamic simulations and can

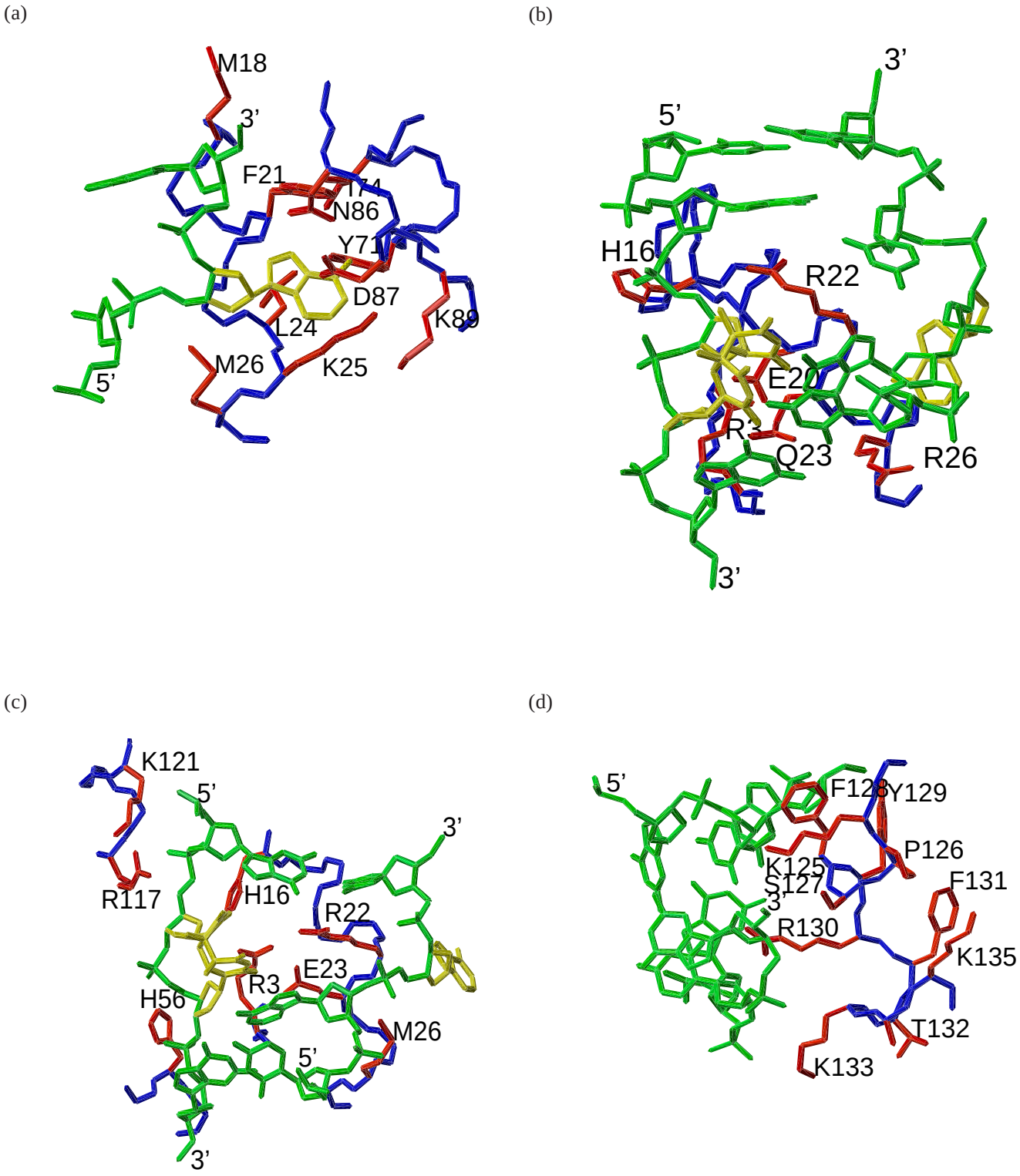


Figure 9 *View of (a) surrounding pocket of the flipped out base A221 in the dynamic averaged conformation for PDG/DNA; (b) the active site of Endo V in dynamic averaged conformation for Endo V/DNA; (c) the recognition and catalytic sites in the dynamic averaged conformation for PDG/DNA; and (d) binding sites near the C-terminus of PDG in the dynamic averaged conformation for PDG/DNA complex. The DNA is green, while the PD and flipped out base (A221) are yellow. The backbone and side chains of protein are blue and red respectively*

explain PDG's catalytic activities. The MD simulations of this model and the Endo V/DNA crystal structure allow us to suggest roles for amino acids that are consistent with mutagenesis studies and NMR of the interaction of the inactive Glu23Gln mutant of Endo V with its substrate in solution.[45] We were able to quantitate many of the similarities and differences in the two structures by following the changes in the distance between key amino acids and areas of the DNA as a function of simulation time.

The higher surface charge of PDG could account for the salt stability of the complex

Endo V has been shown to be processive in its action, i.e. it completes its reaction before releasing its substrate. Processivity is a good indication of how tightly an active enzyme binds to substrates, as DNA binding constants can only be determined accurately for non-active nucleases or for non-cleavable substrates.[41] Mutants of Endo V with reduced processivity in the *in vitro* assay, even if they appear to have a higher catalytic rate, generally cannot restore UV resistance to repair deficient *E.coli* strains.[40] While Endo V becomes distributive at salt concentrations greater than 40mM, PDG retains processivity to concentrations greater than 100mM.[16] The model for PDG indicates a higher surface positive charge than for Endo V (Figure 5) and more charged interactions with the DNA backbone (Table 3), consistent with its ability to bind substrate tightly even at higher ionic strength.

A wider gap in the DNA complexed with PDG could account for binding to a trans-syn *II cyclobutane PD*

The active sites of PDG and Endo V are similar with respect to the positioning of catalytic residues (Figure9 b,c). However, the sites do not match exactly and the angle of the PD is different. One key observation is that the distance between the A220 (the non-flipped adenine) and the PD is longer in the PDG model than in simulations of the Endo V complex. The distance between Arg22 and the PD also increases in PDG, and the large side chain of Arg26 is missing from the active site. Thus the binding area for the PD would be larger, which may allow the trans-syn II dimer to fit in the PDG catalytic site.

While these differences could account for the ability of PDG to bind the *trans-syn* II PD substrate, it is also possible that binding of this alternate substrate induces changes in the site that would only be determined from a structure of this complex.

Arg22 disrupts base pairing near the PD and could be responsible for pushing out A221

In the crystal structure of free Endo V (PDB entry 2end), the side chains of Arg22 and Arg26 (Met26 in PDG) are near one another and highly surface exposed, while in the crystal structure of the complex, they lie within the hole in the DNA and point to opposite sides of the dimer. These two arginines account for 105 (10 %) of the distances less than 3.5Å between protein and DNA atoms in the dynamic averaged structure of the Endo V complex. The Arg22 guanidinium group extends into the G206-C222 pair (Figure 9b,9c). The overall mobility in both structures throughout the simulations, coupled with its exposure in the free Endo V, suggest that this long charged side chain could probe for base stacking irregularities during the scanning. Successful intercalation near the PD could be the driving force for expulsion of A221. The other active site residues would then move into the gap, forcing the DNA further apart and accentuating any preexisting backbone deformations.

While NMR studies of free oligonucleotides containing a *cis-syn* dimer indicate that the structure is B-form DNA, both a slight bend (9° [46], 22° [47]) in the DNA and irregular base stacking are indicated.[48] Our simulations suggest that both enzymes recognize and then, by binding, exaggerate these deviations in the back bone and base pair interactions to push the adenine out of the DNA and establish the catalytic complex. In keeping with their important role in our model, mutagenesis of Arg3, Arg22 or Arg26 in Endo V lowers DNA binding, catalytic activity and the protein's ability to increase the survival of repair deficient *E. coli* subjected to UV irradiation.[11,40]

Mutants that could provide more information about this model

We would suggest from the above analysis that a Met26Arg mutant of PDG would have reduced activity on *trans-syn* II dimer substrates. As Gln15 assumes many more roles in PDG than in Endo V, mutants here with smaller side chains should have a more drastic effect on PDG's activity and substrate preference. A Gln15Arg mutant of Endo V had lower catalytic activity than the wild type but a lower K_m.[40] Other interesting mutations would be those affecting base flipping, such as converting Leu24 to a charged residue or Asp87 to an uncharged one in either protein, mutations of Tyr71 in PDG or Gln71/Gln91 in Endo V, and mutations at Pro25/Lys25 in Endo V and PDG respectively. Finally, mutations near the C-terminus, especially Arg130 in PDG, should have a pronounced effect on binding DNA and possibly on establishing the catalytic site. The appropriate mutants are now being prepared and studied in the group of R. Stephen Lloyd at the UTMB.

Conclusions

We developed a stable model of the recently isolated PDG from a *Chlorella* virus bound to its substrate, based on the crystal structure of the complex of the closely related Endo V from T4 bacteriophage bound to a DNA containing a PD,

with the DIAMOD program. In both complexes, the flipped out A221 base is loosely held mainly by hydrophobic interactions but in a more polar environment in the PDG as compared to the Endo V pocket. While many of the amino acid residues binding to DNA and in the catalytic center are identical, we were able to identify key differences in the structures that could account for PDG's unique activity on the *trans-syn* II pyrmidine dimer as well as the *cis-syn* PD substrate recognized by both enzymes.

Acknowledgments We thank Steven Lloyd and his group in the Sealy Center for Molecular Science at the UTMB, particularly Amanda McCullough and John Garvish, for their cooperation in initiating this project and helpful discussions. This work was supported by grants to WB from the National Science Foundation (DBI-9632326) and the U.S. Department of Energy (DE-FG03-96ER62267) and used the infrastructure of the Sealy Center for Structural Biology, which was established with generous funding from the Sealy and Smith Foundation and an award from the Lucille P. Markey Foundation.

Supplementary material available Atomic coordinates of a superposition of the energy refined model structure of PDG/DNA and the unmodified PDB entry (1vas), the energy refined dynamic average structure of PDG/DNA and the energy refined dynamic average structure of Endo V/DNA are available as supplementary material.

References

1. Ellenberger, T. *Chem. Biol.* **1995**, *2*, 351-354.
2. Wei, Q.; Matanoski, G.M.; Farmer, E.R.; Hedayati, M.A.; Grossman, L. *J. Investigative Dermatology* **1995**, *104*, 933-936.
3. Zeng-Rong, N.; Paterson, J.; Alpert, L.; Tsao, M.S.; Viallet, J.; Alaoui-Jamali, M.A. *Cancer Res.* **1995**, *55*, 4760-4764.
4. Koc, O.N.; Phillips, W.P.J.; Lee, K.; Liu, L.; Zaidi, N.H.; Allay, J.A.; Gerson, S.L. *Cancer Treat. Res.* **1996**, *87*, 123-146.
5. Dodson, M.L.; Michaels, M.L.; Lloyd, R.S. *J. Biol. Chem.* **1994**, *269*, 32709-32712.
6. Manuel, R.C.; Latham, K.A.; Dodson, M.L.; Lloyd, R.S. *J. Biol. Chem.* **1995**, *270*, 2652-2661.
7. McCullough, A.K.; Scharer, O.; Verdine, G.L.; Lloyd, R.S. *J. Biol. Chem.* **1996**, *271*, 32147-32152.
8. McCullough, A.K.; Dodson, M.L.; Scharer, O.D.; Lloyd, R.S. *J. Biol. Chem.* **1997**, *272*, 27210-27217.
9. Lloyd, R.S.; Cheng, X. *Biopolymers* **1997**, *44*, 139-151.
10. Latham, K.A.; Lloyd, R.S. *Biochemistry* **1995**, *34*, 8796-8803.
11. Vassylyev, D.G.; Kashiwagi, T.; Mikami, Y.; Ariyoshi, M.; Iwai, S.; Ohtsuka, E.; Morikawa, K. *Cell* **1995**, *83*, 773-782.
12. Vassylyev, D.G.; Morikawa, K. *Curr. Opin. Struct. Biol.* **1997**, *7*, 103-109.
13. Cheng, X.; Blumenthal, R.M. *Structure* **1996**, *4*, 639-645.
14. Lu, Z.; Li, Y.; Zhang, Y.; Kutish, G.F.; Rock, D.L.; L., V.E.J. *Virology* **1995**, *206*, 339-352.
15. Furuta, M.; Schrader, J.O.; Schrader, H.S.; Kokjohn, T.A.; Nyaga, S.; McCullough, A.K.; Lloyd, R.S.; Burbank, D.E.; Landstein, D.; Lane, L.; Van Etten, J.L. *Appl. Enviro. Microbiol.* **1997**, *63*, 1551-1556.
16. McCullough, A.K.; Romberg, M.T.; Nyaga, S.; Wie, Y.; Wood, T.G.; Taylor, J.S.; Van Etten, J.L.; Dodson, M.L.; Lloyd, R.S. *J. Biol. Chem.* **1998**, *273*, 13136-13142.
17. Chothia, C.; Lesk, A.M. *EMBO J.* **1986**, *5*, 823-826.
18. Aszodi, A.; Taylor, W.R. *Folding & Design* **1996**, *1*, 325-334.
19. Mumenthaler, C.; Schneider, U.; Buchholz, C.J.; Koller, D.; Braun, W.; Cattaneo, R. *Protein Sci.* **1997**, *6*, 588-597.
20. Buchholz, C.J.; Koller, D.; Devaux, P.; Mumenthaler, C.; Schneider-Schaulies, J.; Braun, W.; Gerlier, D.; Cattaneo, R. *J. Biol. Chem.* **1997**, *272*, 22072-22079.
21. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. *Nucleic Acids Res.* **1994**, 4673-4680.
22. Bernstein, F.C.; Koetzle, T.F.; Williams, G.J.; Meyer, E.F.J.; Brice, M.D.; Rodgers, J.R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *Eur. J. Biochem.* **1977**, *80*, 319-324.
23. Hänggi, G.; Braun, W. *FEBS Letters* **1994**, *344*, 147-153.
24. Mumenthaler, C.; Braun, W. *Protein Sci.* **1995**, *4*, 863-871.
25. Braun, W.; Go, N. *J. Mol. Biol.* **1985**, *186*, 611-626.
26. Güntert, P.; Braun, W.; Wüthrich, K. *J. Mol. Biol.* **1991**, *217*, 517-530.
27. Brooks, B.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187-217.
28. Laskowski, R.A.; Mac Arthur, M.W.; Moss, D.S.; Thornton, J.M. *J. Appl. Cryst.* **1993**, *26*, 283-291.
29. Brünger, A.T. X-PLOR, Version 3.1; Yale University Press: New Haven, 1992.
30. Dewar, M.J.S.; Zoebisch, E.F.; Healy, E.F.; Stewart, J.J.P. *J. Amer. Chem. Soc.* **1985**, *107*, 3902-3909.
31. Stewart, J.J.P. MOPAC 93, Fujitsu Ltd., Tokyo, Japan.
32. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Gill, P.M.W.; Johnson, B.G.; Robb, M.A.; Cheeseman, J.R.; Keith, T.; Petersson, G.A.; Montgomery, J.A.; Raghvachari, K.; Al-Laham, M.A.; Zakrewski, V.G.; Ortiz, J.V.; Foresman, J.B.; Peng, C.Y.; Ayala, P.Y.; Chen, W.; Wong, M.W.; Andres, J.L.; Replogle, E.S.; Gomperts, R.; Martin, R.L.; Fox, D.J.; Binkley, J.S.; Defrees, D.J.; Baker, J.; Stewart, J.P.; Head-Gordon, M.; Gonzalez, C.; Pople, J.A. Gaussian 94, Revision B.3, 1995.
33. Jorgensen, W.; Chandrasekar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926-935.
34. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H.J.C. *J. Comput. Phys.* **1977**, *23*, 327-341.

35. Berendsen, H.J.C.; Postma, J.P.M.; van Gunsteren, N.F.; DiNola, A.; Haak, J.R. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
36. Koradi, R.; Billeter, M.; Wüthrich, K. *J. Mol. Graphics* **1996**, *14*, 51-55.
37. Spector, T.I.; Cheatham III, T.E.; Kollman, P.A. *J. Am. Chem. Soc.* **1997**, *119*, 7095-7104.
38. Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz Jr., K.M.; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W.; Kollman, P.A. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
39. Arnold, G.E.; Ornstein, R.L. *Proteins* **1994**, *18*, 19-33.
40. Nyaga, S.G.; Dodson, M.L.; Lloyd, R.S. *Biochemistry* **1997**, *36*, 4080-4088.
41. Lloyd, R.S. *Mutation Research* **1998**, *408*, 159-170.
42. Pingoud, A.; Jeltsch, A. *Eur. J. Biochem.* **1997**, *246*, 1-22.
43. Fraczkiewicz, R.; Braun, W. *J. Comput. Chem.* **1998**, *19*, 319-333.
44. The web site for GETAREA is http://www.scsb.utmb.edu/getarea
45. Lee, B.J.; Sakashita, H.; Ohkubo, T.; Ikehara, M.; Doi, T.; Morikawa, K.; Kyogoku, Y.; Osafune, T.; Iwai, S.; Ohtsuka, E. *Biochemistry* **1994**, *33*, 57-64.
46. Kim, J.; Patel, D.; Choi, B. *Photochem. Photobiol.* **1995**, *62*, 44-50.
47. Yamaguchi, H.; van Aalten, D.M.; Pinak, M.; Furukawa, A.; Osman, R. *Nucl. Acid Res.* **1998**, *26*, 1939-1946.
48. McAteer, K.; Jing, Y.; Kao, J.; Taylor, J.-S.; Kennedy, M.A. *J. Mol. Biol.* **1998**, *282*, 1013-1032.