

Supporting Information

Lesion sensing during initial binding by yeast XPC/Rad4: toward predicting resistance to nucleotide excision repair

Hong Mu¹, Yingkai Zhang^{2,3}, Nicholas E. Geacintov², and Suse Broyde^{1}*

¹ Department of Biology and ² Department of Chemistry, New York University, New York, NY
10003, USA

³ NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China

* To whom correspondence should be addressed. Suse Broyde, Biology Department, New York
University, 100 Washington Square East, 1009 Silver Center, New York, NY 10003. Tel: +1 212
998 8231; Fax: +1 212 995 4015; Email: broyde@nyu.edu

Table of Contents

SUPPLEMENTARY FIGURES	4
Figure S1 MD simulation initial models, DNA base sequences and definition of twist and untwist angles.....	4
Figure S2 The initial binding of Rad4 to the <i>cis</i> -B[<i>a</i>]P-dG lesion site with pre-flipped partner C based on the previously obtained trajectory ³	6
Figure S3 The initial binding of Rad4 to the PhIP-C8-dG lesion site with pre-flipped partner C.	8
Figure S4 The initial binding of Rad4 to the 14 <i>R</i> -DB[<i>a,l</i>]P-dG lesion site with ruptured partner C. ...	10
Figure S5 The initial binding of Rad4 to the <i>cis</i> -B[<i>a</i>]P-dG in the deletion duplex with its partner nucleotide missing.	12
Figure S6 The initial binding of Rad4 to the <i>cis</i> -B[<i>a</i>]P-dG lesion site with major groove contacting mismatched partner A.	14
Figure S7 The initial binding of Rad4 to the 14 <i>R</i> -DB[<i>a,l</i>]P-dA lesion site with Watson-Crick paired partner T.	16
Figure S8 The initial binding of Rad4 to the PhIP-C8-dG in the deletion duplex with its partner nucleotide missing.	18
Figure S9 The initial binding of Rad4 to the <i>cis</i> -B[<i>a</i>]P-dG lesion site with major groove-contacting mismatched partner T.	20
Figure S10 The initial binding of Rad4 to the (+) <i>trans</i> -B[<i>a</i>]P-dG-containing duplex.	22
Figure S11 The initial binding of Rad4 to the (-) <i>trans</i> -B[<i>a</i>]P-dG-containing duplex.....	24
SUPPLEMENTARY MOVIE TITLES	26
Movie S1. Initial binding state of the (+) <i>cis</i> -B[<i>a</i>]P-dG:dC duplex shown in Figure 3.	26
Movie S2. Initial binding state of the PhIP-C8-dG:dC duplex shown in Figure 3.	26
Movie S3. Initial binding state of the 14 <i>R</i> -DB[<i>a,l</i>]P-dG:dC duplex shown in Figure 3.....	26
Movie S4. Initial binding state of the (+) <i>cis</i> -B[<i>a</i>]P-dG:deletion duplex shown in Figure 3.....	26
Movie S5. Initial binding state of the (+) <i>cis</i> -B[<i>a</i>]P-dG:dA duplex shown in Figure 3.....	26
Movie S6. Initial binding state of the 14 <i>R</i> -DB[<i>a,l</i>]P-dA:dT duplex shown in Figure 3.	26
Movie S7. Initial binding state of the PhIP-C8-dG:deletion duplex shown in Figure 3.	26
Movie S8. Initial binding state of the (+) <i>cis</i> -B[<i>a</i>]P-dG:dT duplex shown in Figure 3.	26
Movie S9. Initial binding state of the (+) <i>trans</i> -B[<i>a</i>]P-dG:dC duplex shown in Figure 3.	26
Movie S10. Initial binding state of the (-) <i>trans</i> -B[<i>a</i>]P-dG:dC duplex shown in Figure 3.	26
SUPPLEMENTARY METHODS	27
Molecular modeling and structure preparation	27
Force field and MD simulation protocols	27

Structural analyses	28
SUPPLEMENTARY REFERENCES.....	30

SUPPLEMENTARY FIGURES

A. Initial models (docking complex)

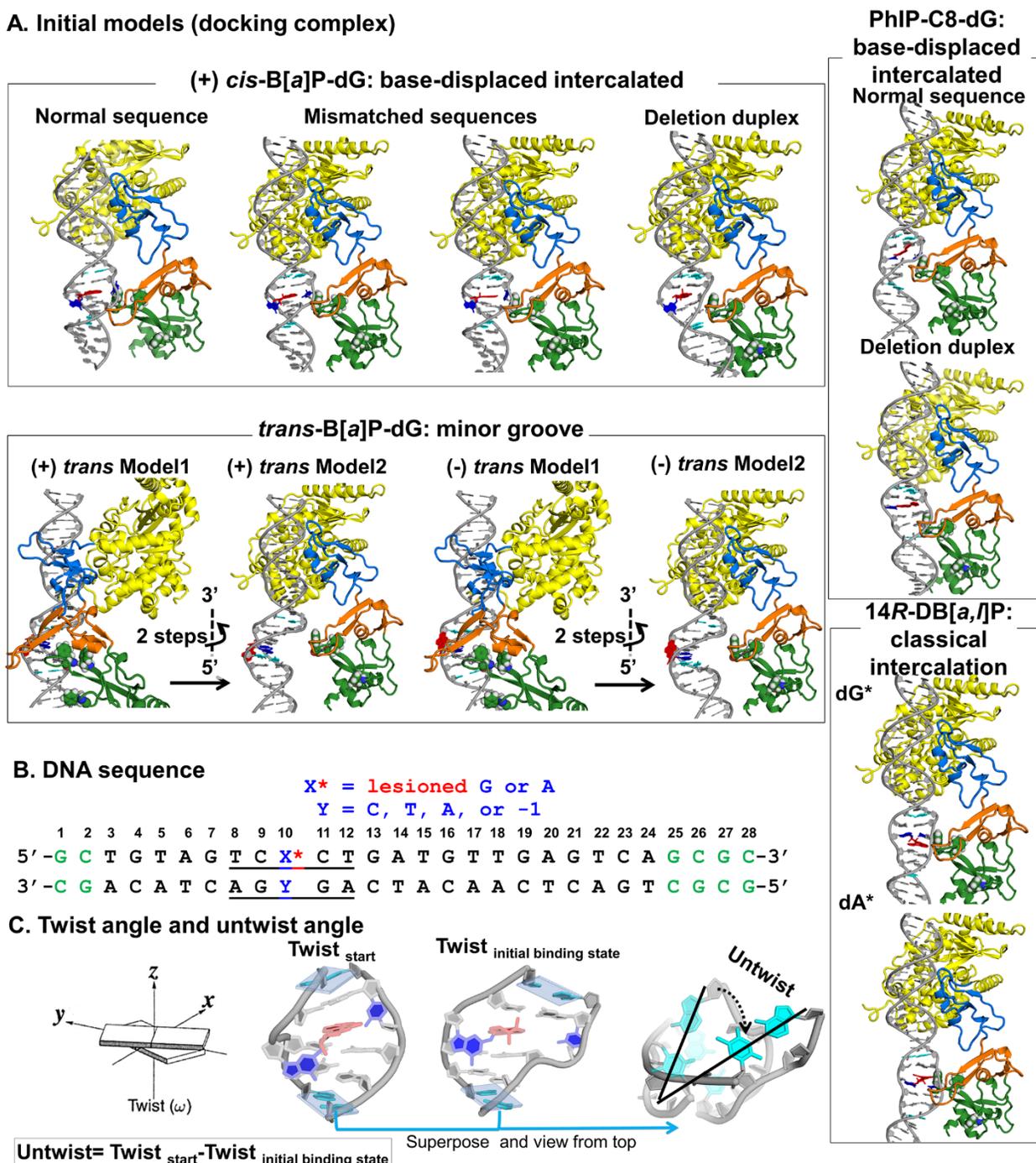


Figure S1 MD simulation initial models, DNA base sequences and definition of twist and untwist angles.

(A) The initial models (docking complex) that provide the starting structures for the MD simulations. The (+) *cis*-B[a]P-dG adduct adopts a base-displaced intercalated conformation in all sequence contexts: the B[a]P aromatic ring system is intercalated from the minor groove and the adducted dG* is displaced into

the minor groove. With normal partner C the base is dynamic and extruded into the major groove. With mismatched partner T the base is displaced from the helix and is in contact with the major groove via its methyl group; it is somewhat dynamic and extrudes on occasion. With mismatched partner A the base is bound to the major groove surface, intercalates into the helix in a minor population, but does not extrude. With the partner nucleotide missing, the neighboring base pairs form a tightly stacked wedge around the B[a]P rings. For the *trans*-B[a]P-dG adducts, the B[a]P rings are in the minor groove with Watson-Crick pairs maintained, oriented 5' on the lesion-containing strand for (+) *trans* and 3' for (-) *trans*. The PhIP-C8-dG adopts a base-displaced intercalated conformation similar to the (+) *cis*-B[a]P-dG, but contains a smaller ring system and a mobile phenyl ring that rotates dynamically in the minor groove. With normal partner C the base is extruded into the major groove; with a missing partner nucleotide the tight wedge-like stacking seen in the (+) *cis*-B[a]P-dG:deletion duplex is inhibited by the phenyl ring rotation. The 14*R*-DB[a,l]P adducts are classically intercalated without displacement of the adducted base. For the guanine adduct, the intercalation is from the narrow minor groove to the 3' side of the modified guanine and Watson-Crick pairing with partner C is ruptured although the C remains stacked into the helix but overlaps poorly with the DB[a,l]P rings. For the adenine adduct, the intercalation is from the major groove to the 5' side of the modified adenine and the A*:T Watson-Crick pair is maintained. This lesion is marked by very strong stabilizing van der Waals stacking interactions between the DB[a,l]P ring system and its neighboring base pairs¹, manifested by its thermally stabilizing impact on the DNA duplex of $\sim +10$ °C². (B) The base sequence context of the lesion containing duplexes utilized in the simulations. The lesion-containing base pair is blue with the damaged nucleotide denoted by a red *, and extended sequences are green. (C) Definition and illustrations of twist and untwist angle.

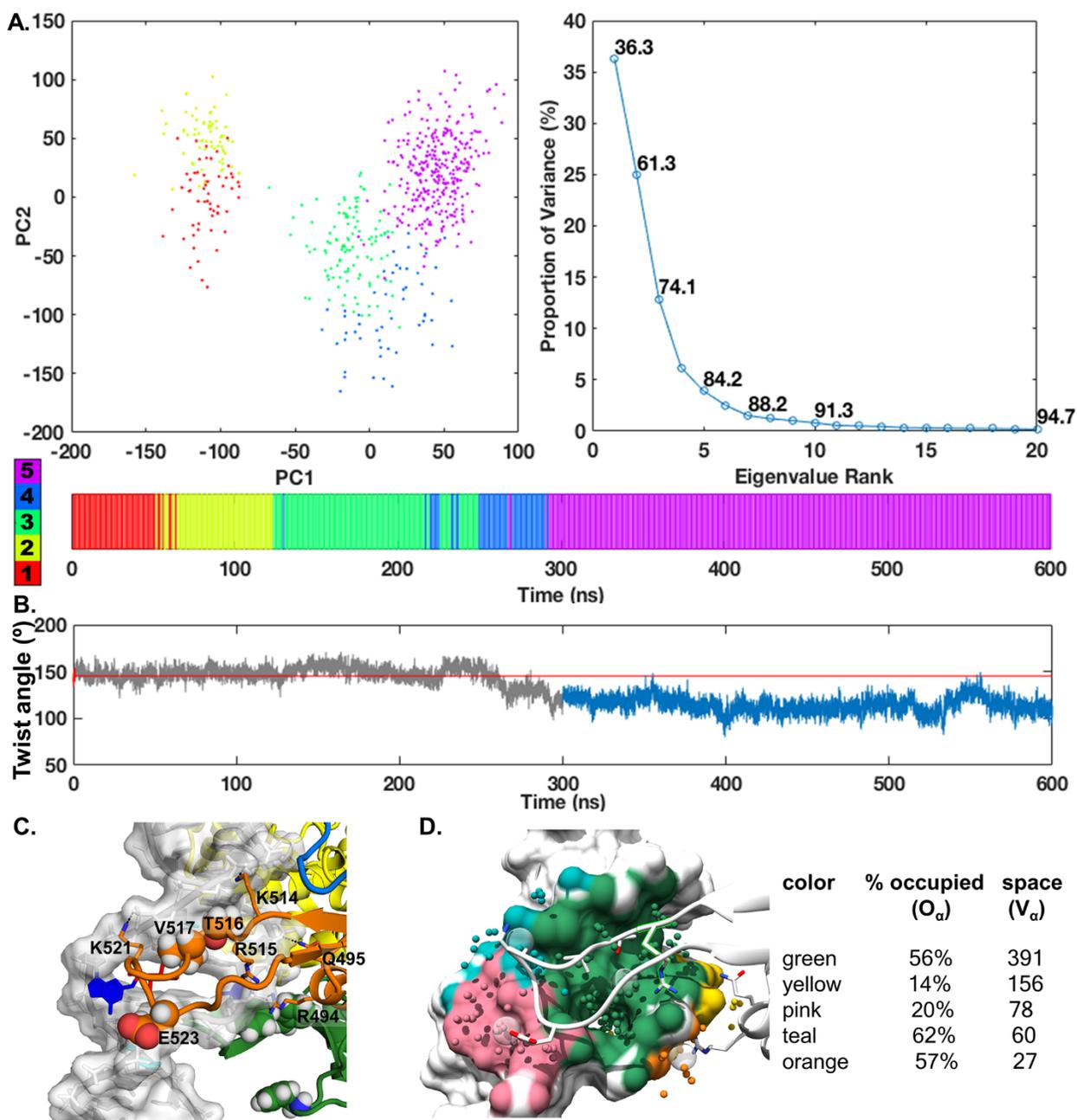


Figure S2 The initial binding of Rad4 to the *cis*-B[a]P-dG lesion site with pre-flipped partner C based on the previously obtained trajectory³.

(A) Clustering for the structural ensemble of the initial binding state. The first two principle components (PC) are plotted against each other. The proportions of variance are plotted for the first twenty PCs, where the cumulative proportions are labeled. The clusters are also given in color code along the time axis. (B) Twist angle between the two end base pairs of the lesion-containing 6-mer (cyan base pairs). The time dependent values are plotted against the time axis. The initial binding state values are in blue (1 to 1.5 μs), the starting state values (0 to 1 ns) are in red and their mean value is shown in red line. (C) Best-representative structure of the initial binding state ensemble. The structure is rendered as in Figure 3 of

the main text and is viewed into the minor groove. The DNA residues are also shown in transparent surface, except the lesion-containing base and its partner base. The DNA backbone heavy atoms and the amino acids that form hydrogen bonds (black dash lines) with DNA are shown in sticks, and the side chains of Thr 516, Val 517 and Glu523 are shown in spheres. These amino acids are color coded by atom type with carbon atom in the same color as the domain. At the initial binding state, the captured partner C stacks with Phe 599 of BHD3, Thr 516 and Val 517 insert into the minor groove, Glu523 interacts with DNA backbone via van der Waals interaction, Lys 514 and Lys 521 form hydrogen bonds with lesion containing strand backbone while Arg 494, Gln 495 and Arg 515 hydrogen bond to the unmodified partner strand DNA backbone. **(D)** Alpha space analyses of BHD2 binding to the lesion-containing DNA. The DNA is shown in surface, with BHD2-contacting pockets color-coded as in the table. The alpha sphere centers are shown as small spheres in the same color as respective pocket, and the alpha cluster centroids are shown in transparent spheres.

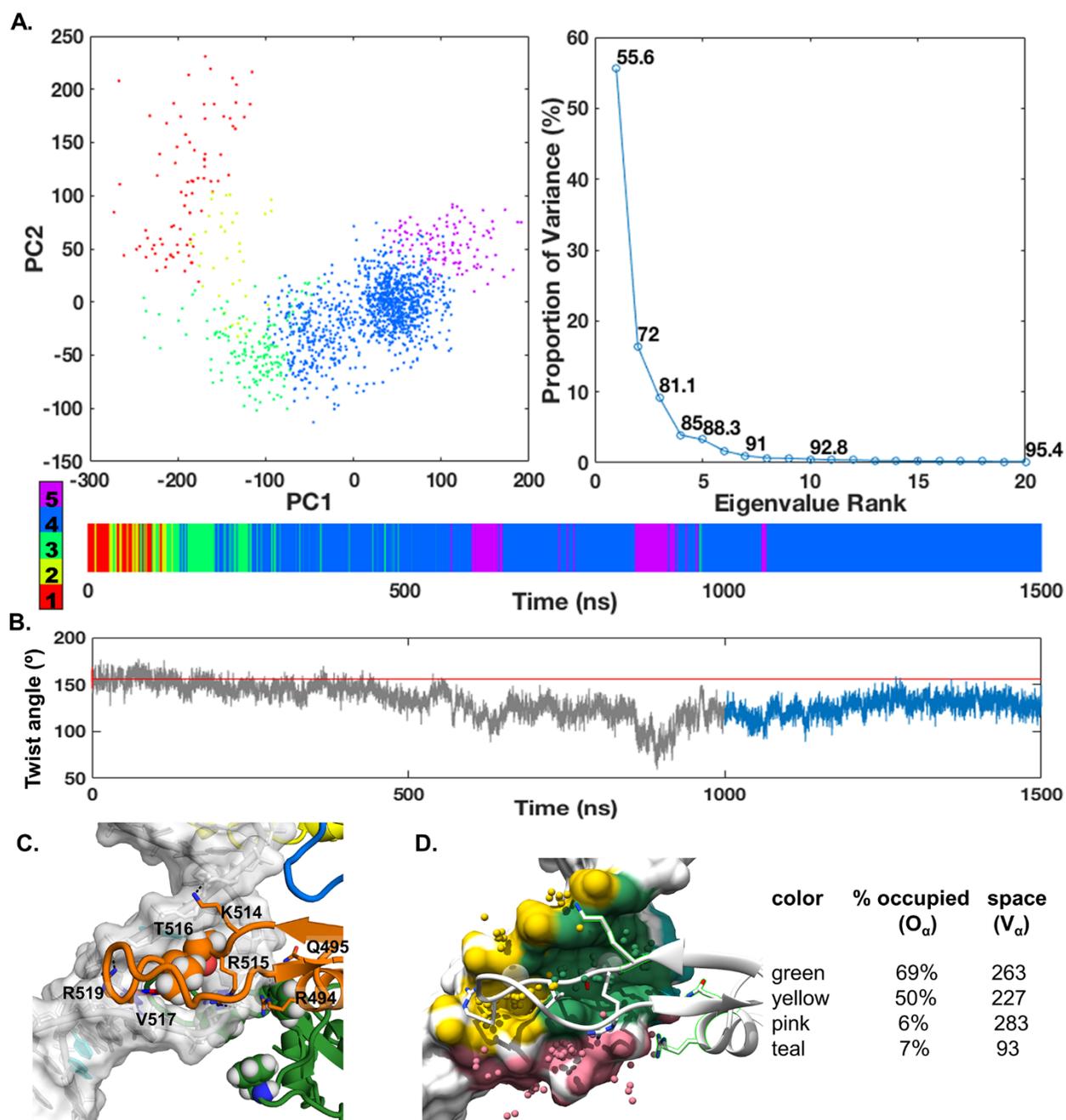


Figure S3 The initial binding of Rad4 to the PhIP-C8-dG lesion site with pre-flipped partner C.

(A) Clustering for the structural ensemble of the initial binding state. The first two principle components (PC) are plotted against each other. The proportions of variance are plotted for the first twenty PCs, where the cumulative proportions are labeled. The clusters are also given in color code along the time axis. (B) Twist angle between the two end base pairs of the lesion-containing 6-mer (cyan base pairs). The time dependent values are plotted against the time axis. The initial binding state values (1 to 1.5 μ s), the starting state values (0 to 1 ns) are in red and their mean value is shown in red line. (C) Best-representative structure of the initial binding state ensemble. The structure is rendered as in Figure 3 of the main text and is viewed into the minor groove. The DNA residues are also shown in transparent

surface, except the lesion-containing base and its partner base. The DNA backbone heavy atoms and the amino acids that form hydrogen bonds (black dash lines) with DNA are shown in sticks, and the side chains of Thr 516 and Val 517 are shown in spheres. These amino acids are color coded by atom type with carbon atom in the same color as the domain. The captured partner C stacks with the bottom of the pocket formed by Phe 599 of BHD3. Arg 494, Lys 514, Arg 515, and Arg 519 form hydrogen bonds with the DNA backbone, while Thr 516 is inserted into the minor groove and Val517 displays van der Waals interactions with the lesion. **(D)** Alpha space analyses of BHD2 binding to the lesion-containing DNA. The structure is rendered as in Figure S2D.

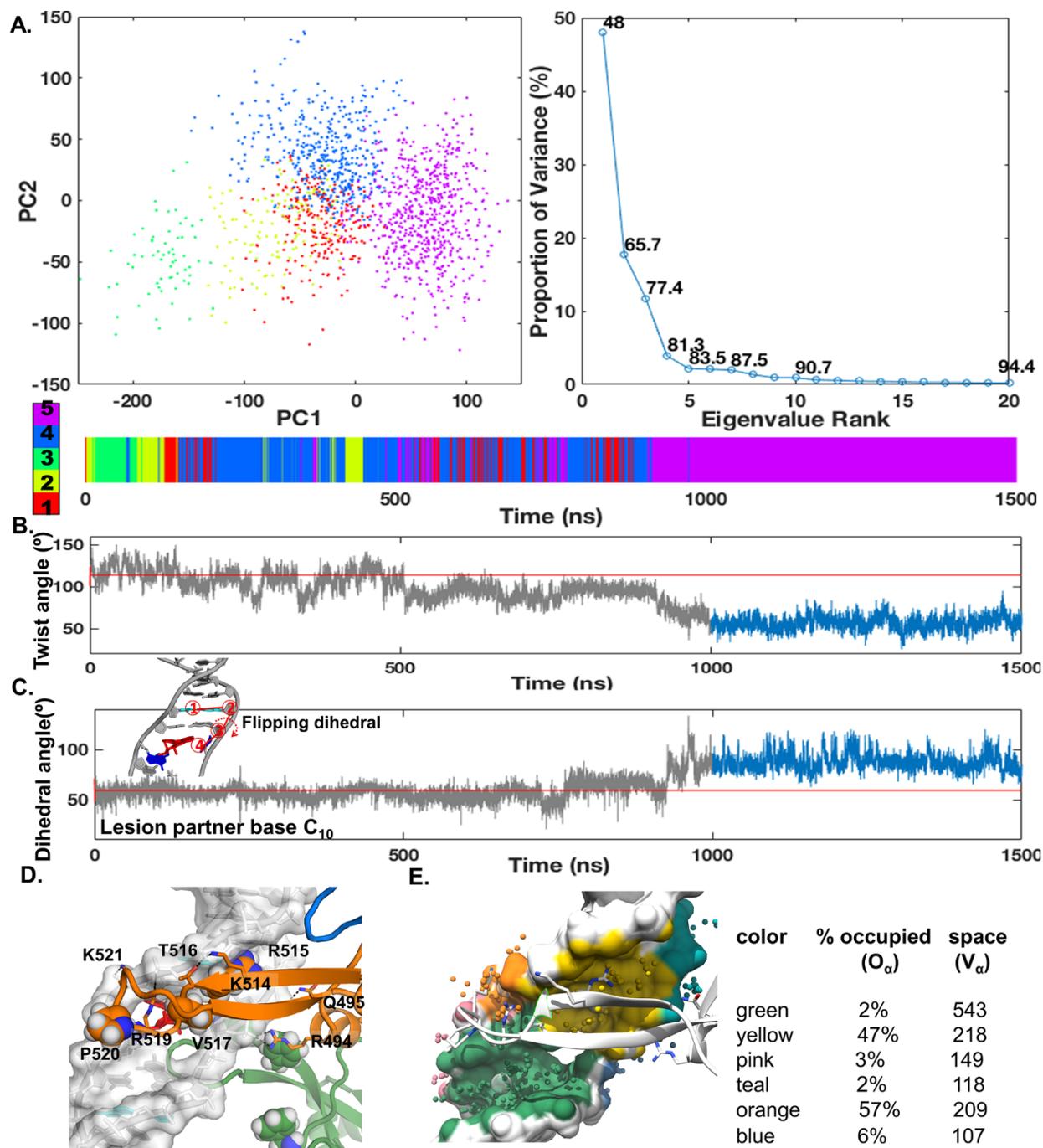


Figure S4 The initial binding of Rad4 to the 14R-DB[a,l]P-dG lesion site with ruptured partner C.

(A) Clustering for the structural ensemble of the initial binding state. The first two principle components (PC) are plotted against each other. The proportions of variance are plotted for the first twenty PCs, where the cumulative proportions are labeled. The clusters are also given in color code along the time axis. (B) Twist angle between the two end base pairs of the lesion-containing 6-mer (cyan base pairs). The time dependent values are plotted against the time axis. The initial binding state values are in blue (1 to 1.5 μ s), the starting state values (0 to 1 ns) are in red and their mean value is shown in red line. (C) Flipping

dihedral angle of lesion partner base. The definition of pseudo-dihedral angle that reflect lesion partner flipping is given in the inset: 1) the center of mass (COM) of cyan base pair, 2) and 3) COM of sugar rings, and 4) COM of partner base. The time dependent values are plotted against the time axis. Increased values show extrusion to major groove. **(D)** Best-representative structure of the initial binding state ensemble. The structure is rendered as in Figure 3 of the main text and is viewed into the minor groove. The DNA residues are also shown in transparent surface, except the lesion-containing base and its partner base. The DNA backbone heavy atoms and the amino acids that form hydrogen bonds (black dash lines) with DNA are shown in sticks, and the side chains of Arg 515, Val 517 and Pro520 are shown in spheres. These amino acids are color coded by atom type with carbon atom in the same color as the domain. Upon Rad4 initial binding, the ruptured partner C flips out and stacks with Phe 599 of BHD3 in its binding pocket, Arg 494, Gln 495, Thr 516, Arg 519, and Lys 521 form hydrogen bonds with the DNA backbone, while Val 517 and Arg 515 insert into the minor groove and Pro 520 interacts with the DNA backbone via van der Waals interactions. **(E)** Alpha space analyses of BHD2 binding to the lesion-containing DNA. The structure is rendered as in Figure S2D.

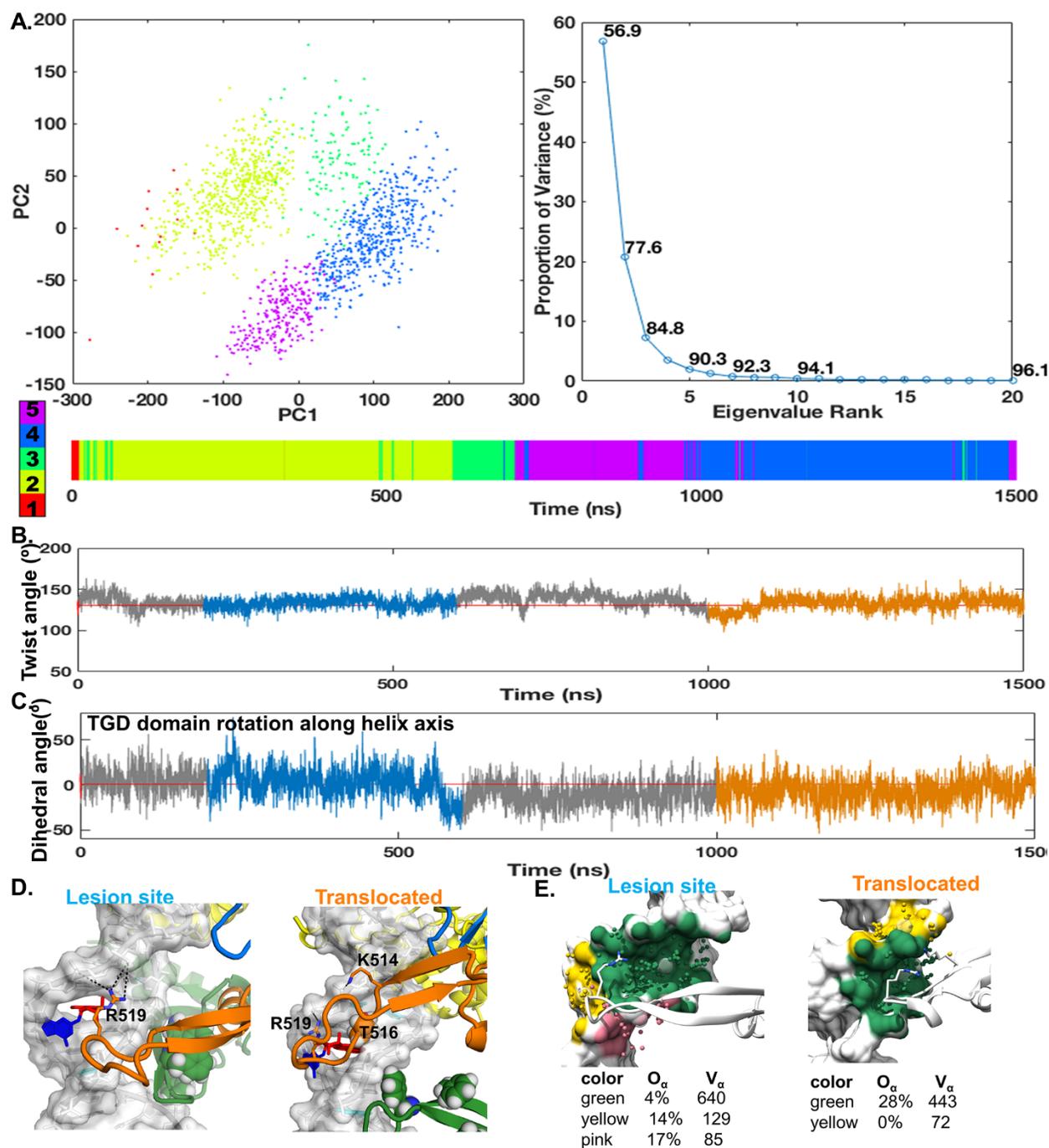


Figure S5 The initial binding of Rad4 to the *cis*-B[a]P-dG in the deletion duplex with its partner nucleotide missing.

(A) Clustering for the structural ensemble of the initial binding state. The first two principle components (PC) are plotted against each other. The proportions of variance are plotted for the first twenty PCs, where the cumulative proportions are labeled. The clusters are also given in color code along the time axis. (B) Twist angle between the two end base pairs of the lesion-containing 6-mer (cyan base pairs). The time dependent values are plotted against the time axis. The initial binding state values are in blue (1 to 1.5 μ s),

the starting state values (0 to 1 ns) are in red and their mean value is shown in red line. (C) TGD domain rotation angle around the helix axis. Decreased values show translocation along the helix axis toward 3' side of lesion containing strand. (D) Best-representative structures of the initial binding states at the lesion site and the translocated site. The structures are rendered as in Figure 3 of the main text and are viewed into the minor groove. The DNA backbone heavy atoms and the amino acids that form electrostatic interactions/hydrogen bonds (black dash lines) with DNA are shown in sticks. The amino acids are color coded by atom type with carbon atom in the same color as the domain. At the lesion site, Arg 519 interact with backbone phosphate groups of the damaged strand via electrostatic interactions (black dash lines). At the translocated site, Thr 516 inserts into the minor groove, while Lys 514 and Arg 519 form episodic hydrogen bonds with the backbone phosphate groups of the damaged strand. (E) Alpha space analyses of BHD2 binding to the lesion-containing DNA. The structures are rendered as in Figure S2D.

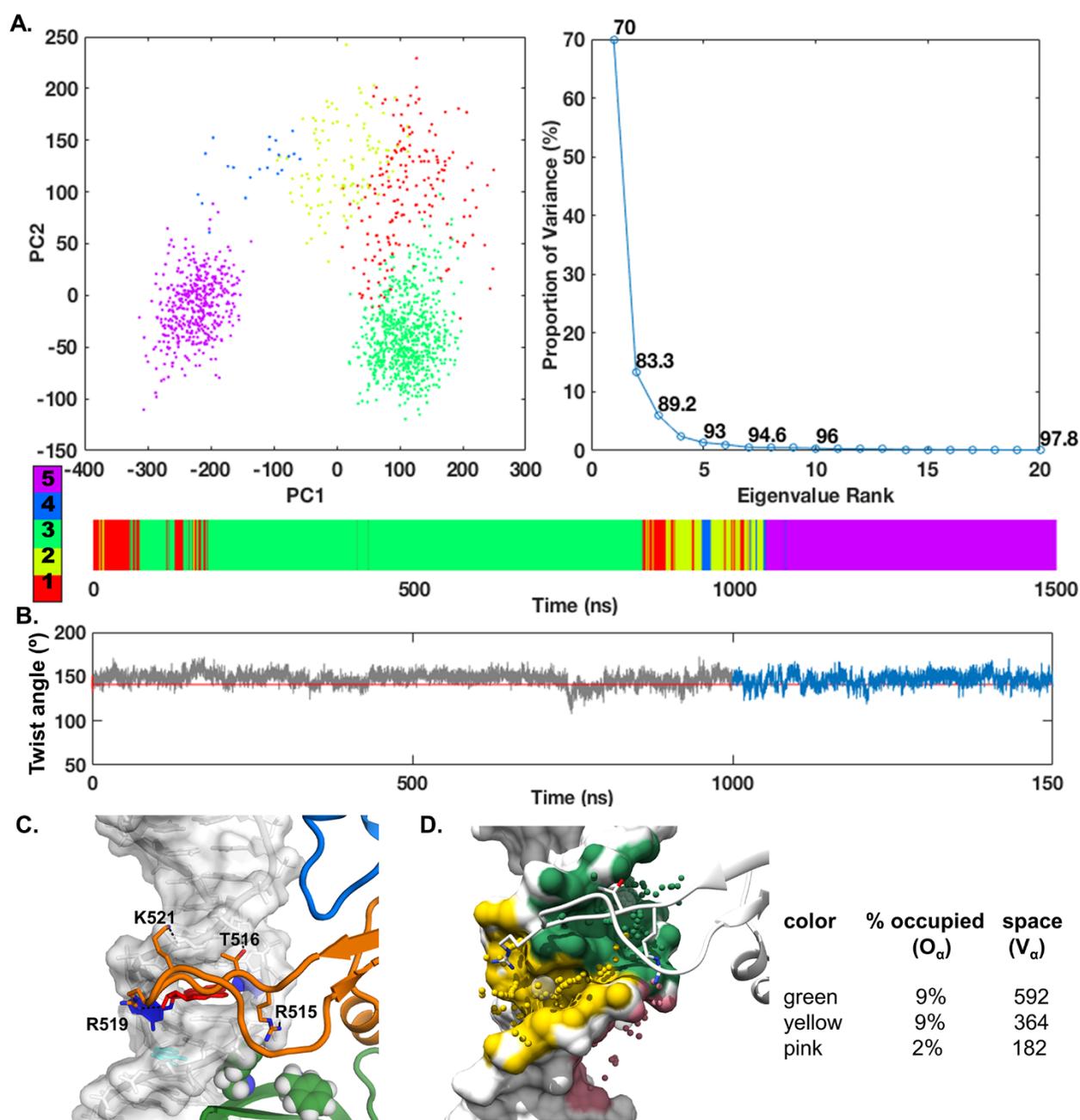


Figure S6 The initial binding of Rad4 to the *cis*-B[a]P-dG lesion site with major groove contacting mismatched partner A.

(A) Clustering for the structural ensemble of the initial binding state. The first two principle components (PC) are plotted against each other. The proportions of variance are plotted for the first twenty PCs, where the cumulative proportions are labeled. The clusters are also given in color code along the time axis. (B) Twist angle between the two end base pairs of the lesion-containing 6-mer (cyan base pairs). The time dependent values are plotted against the time axis. The initial binding state values are in blue (1 to 1.5 μ s),

the starting state values (0 to 1 ns) are in red and their mean value is shown in red line. **(C)** Best-representative structure of the initial binding state ensemble. The structure is rendered as in Figure 3 of the main text and is viewed into the minor groove. The DNA residues are also shown in transparent surface, except the lesion-containing base and its partner base. The DNA backbone heavy atoms and the amino acids that form hydrogen bonds (black dash lines) with DNA are shown in sticks. The amino acids are color coded by atom type with carbon atom in the same color as the domain. At the initial binding state, Thr 516, Lys 521 and Arg 515 form dynamic hydrogen bonds with the DNA backbone, and Arg 519 forms dynamic hydrogen bonds with the hydroxyl group of the benzylic ring. **(D)** Alpha space analyses of BHD2 binding to the lesion-containing DNA. The structure is rendered as in Figure S2D.

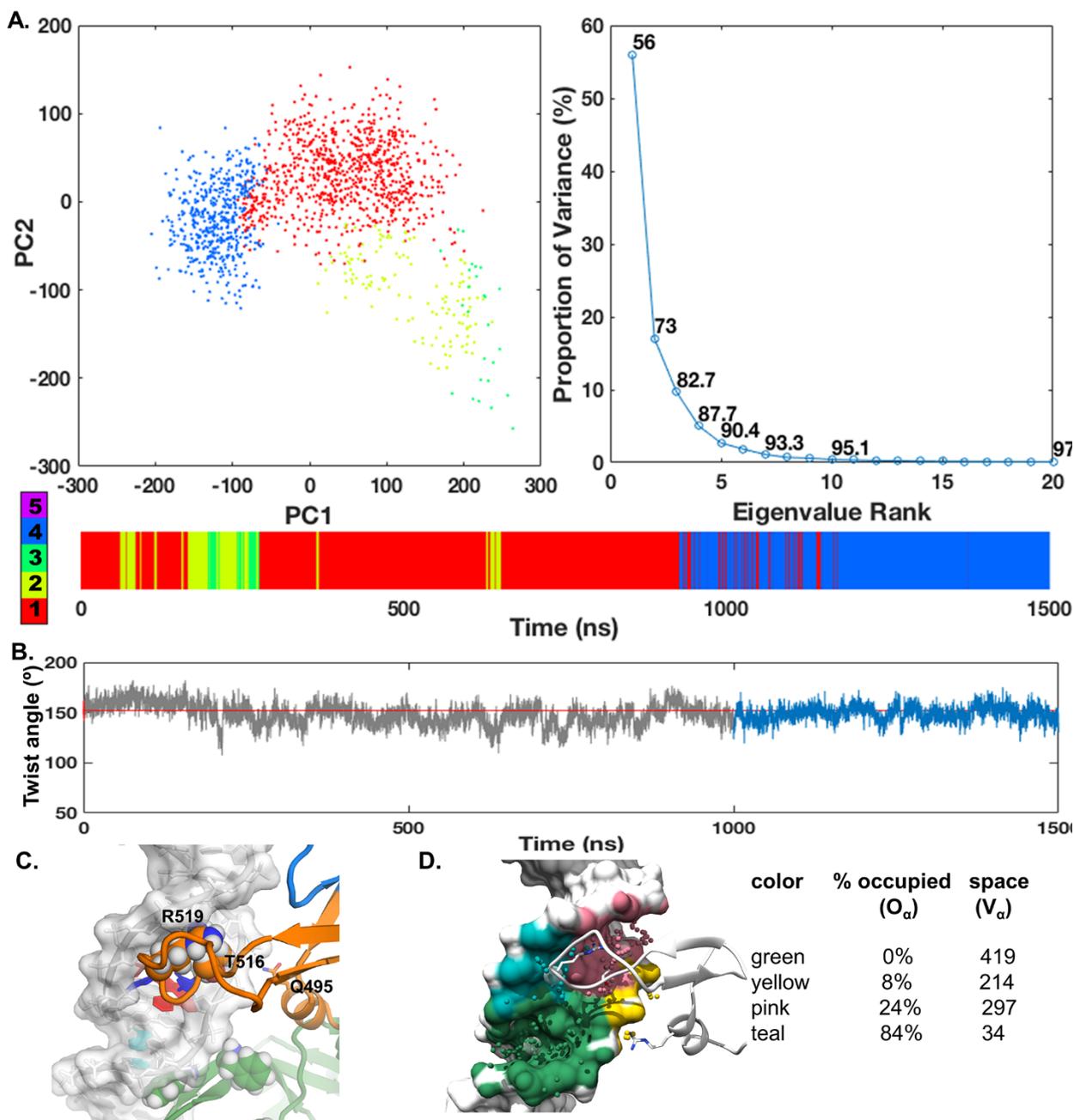


Figure S7 The initial binding of Rad4 to the 14R-DB[a,l]P-dA lesion site with Watson-Crick paired partner T.

(A) Clustering for the structural ensemble of the initial binding state. The first two principle components (PC) are plotted against each other. The proportions of variance are plotted for the first twenty PCs, where the cumulative proportions are labeled. The clusters are also given in color code along the time axis. (B) Twist angle between the two end base pairs of the lesion-containing 6-mer (cyan base pairs). The time dependent values are plotted against the time axis. The initial binding state values are in blue (1 to 1.5 μs), the starting state values (0 to 1 ns) are in red and their mean value is shown in red line. (C) Best-representative structure of the initial binding state ensemble. The structure is rendered as in Figure 3 of

the main text and is viewed into the minor groove. The DNA residues are also shown in transparent surface, except the lesion-containing base and its partner base. The DNA backbone heavy atoms and the amino acids that form hydrogen bonds (black dash lines) with DNA are shown in sticks, and the side chains of Arg 515, Val 517 and Pro520 are shown in spheres. These amino acids are color coded by atom type with carbon atom in the same color as the domain. At the initial binding state, Thr 516 and Arg 519 insert into the minor groove, while Gln 495 hydrogen bonds to the DNA backbone. **(D)** Alpha space analyses of BHD2 binding to the lesion-containing DNA. The structure is rendered as in Figure S2D.

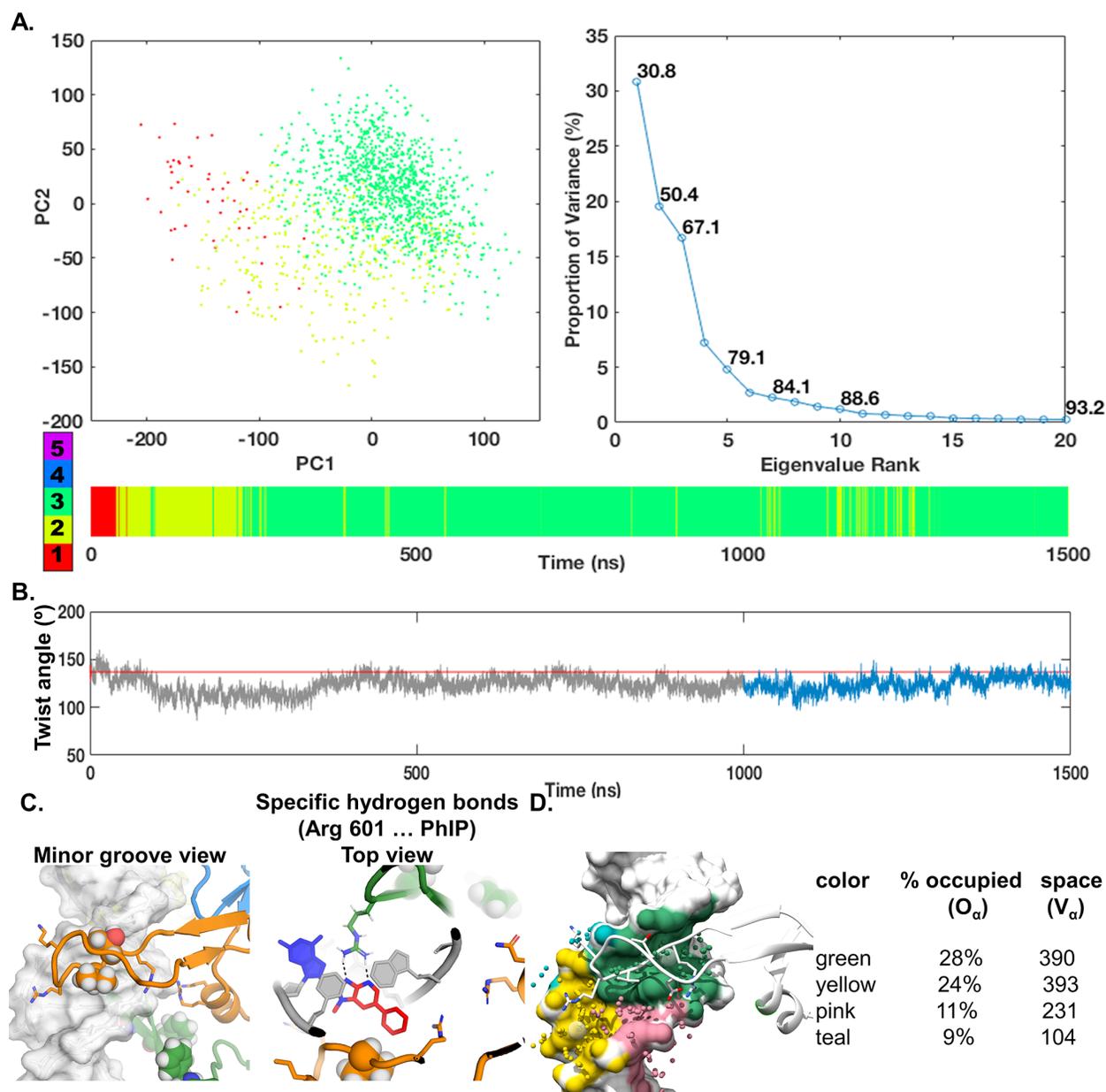


Figure S8 The initial binding of Rad4 to the PhIP-C8-dG in the deletion duplex with its partner nucleotide missing.

(A) Clustering for the structural ensemble of the initial binding state. The first two principle components (PC) are plotted against each other. The proportions of variance are plotted for the first twenty PCs, where the cumulative proportions are labeled. The clusters are also given in color code along the time axis. **(B)** Twist angle between the two end base pairs of the lesion-containing 6-mer (cyan base pairs). The time dependent values are plotted against the time axis. The initial binding state values are in blue (1 to 1.5 μ s), the starting state values (0 to 1 ns) are in red and their mean value is shown in red line. **(C)** Best-representative structure of the initial binding state ensemble. The structure is rendered as in Figure 3 of the main text and is viewed into the minor groove. The DNA backbone heavy atoms and the amino acids that form hydrogen bonds (black dash lines) with DNA are shown in sticks, and the side chains of Thr

516 and Val 517 are shown in spheres. These amino acids are color coded by atom type with carbon atom in the same color as the domain. At the initial binding state, Thr 516, Arg 519 and Lys 521 hydrogen bond with DNA backbone of the lesion-containing strand, while Arg 494, Gln 495 and Arg 515 hydrogen bond with DNA backbone of the undamaged strand. Furthermore, a lesion specific hydrogen bonding interaction with the BHD3 hairpin is observed: Arg 601 of BHD3 forms hydrogen bonds with the imidazole ring of PhIP. The PhIP rings protrude further into the minor groove, stabilized by van der Waals interactions with Thr 516 and Val 517. **(D)** Alpha space analyses of BHD2 binding to the lesion-containing DNA. The structure is rendered as in Figure S2D.

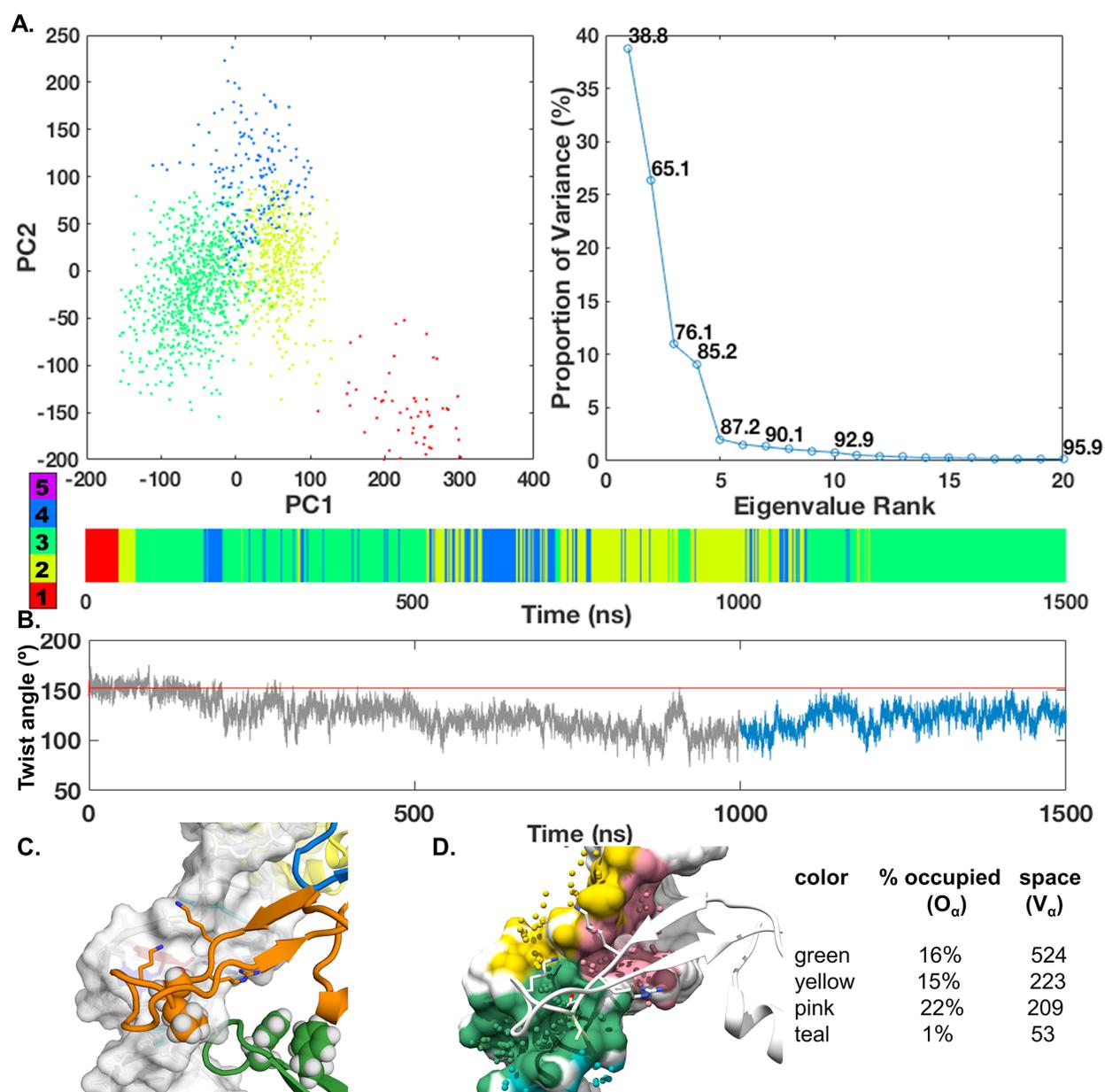


Figure S9 The initial binding of Rad4 to the *cis*-B[a]P-dG lesion site with major groove-contacting mismatched partner T.

(A) Clustering for the structural ensemble of the initial binding state. The first two principle components (PC) are plotted against each other. The proportions of variance are plotted for the first twenty PCs, where the cumulative proportions are labeled. The clusters are also given in color code along the time axis. (B) Twist angle between the two end base pairs of the lesion-containing 6-mer (cyan base pairs). The time dependent values are plotted against the time axis. The initial binding state values are in blue (1 to 1.5 μs), the starting state values (0 to 1 ns) are in red and their mean value is shown in red line. (C) Best-representative structure of the initial binding state ensemble. The structure is rendered as in Figure 3 of the main text and is viewed into the minor groove. The DNA residues are also shown in transparent surface, except the lesion-containing base and its partner base. The DNA backbone heavy atoms and the

amino acids that form hydrogen bonds (black dash lines) with DNA are shown in sticks, and the side chains of Thr 516 and Val 517 are shown in spheres. These amino acids are color coded by atom type with carbon atom in the same color as the domain. At the initial binding state, the major groove-contacting partner T stacks in with its methyl group facing the B[a]P rings. Thr 516 inserts into the minor groove and Val 517 adopts van der Waals interactions with DNA backbone, while Lys 514, Arg515, Thr 516, and Lys 521 form hydrogen bonds with the DNA backbone around the lesion site. **(D)** Alpha space analyses of BHD2 binding to the lesion-containing DNA. The structure is rendered as in Figure S2D.

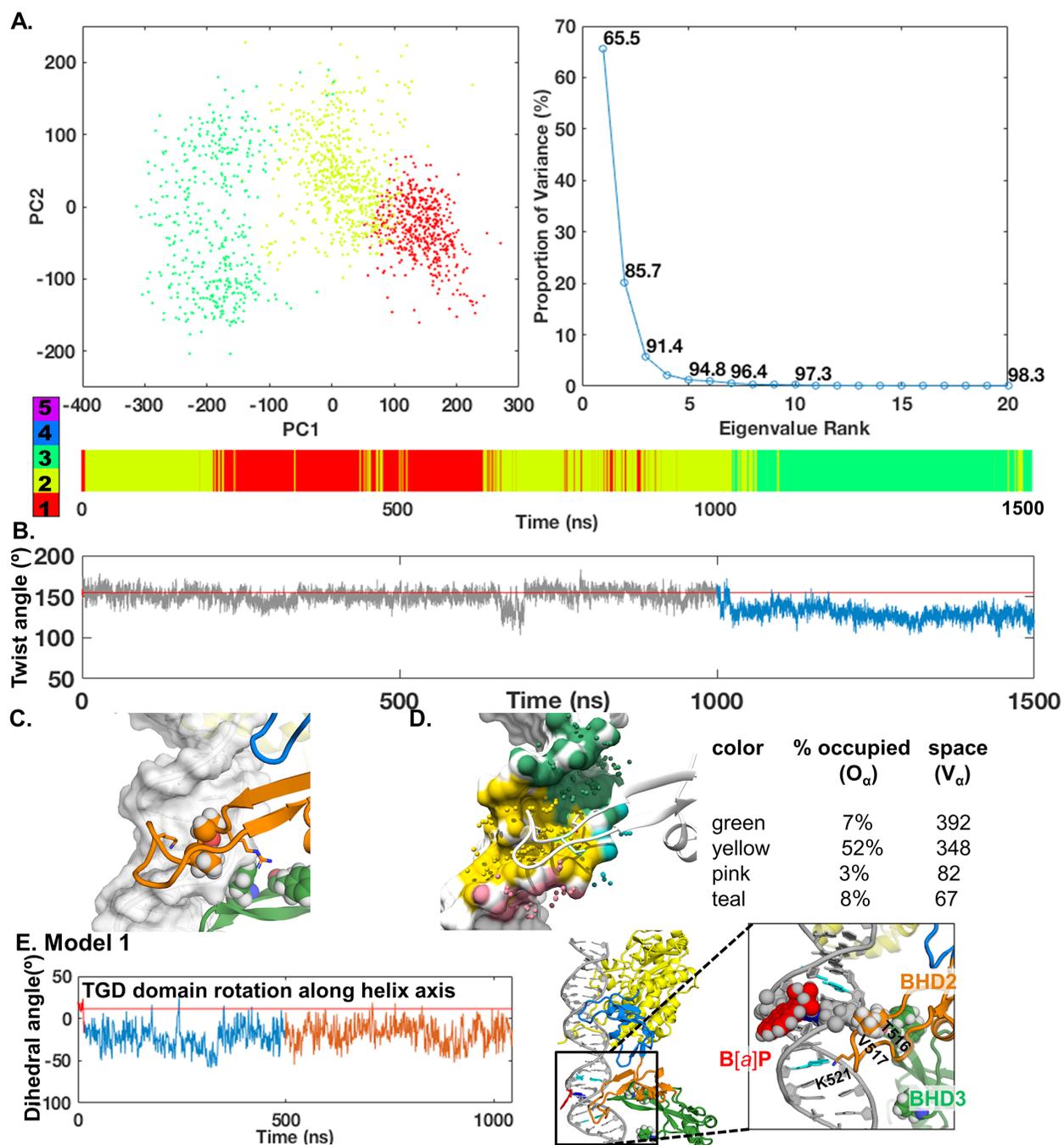


Figure S10 The initial binding of Rad4 to the (+) *trans*-B[a]P-dG-containing duplex.

(A) Clustering for the structural ensemble of the initial binding state. The first two principle components (PC) are plotted against each other. The proportions of variance are plotted for the first twenty PCs, where the cumulative proportions are labeled. The clusters are also given in color code along the time axis. (B) Twist angle between the two end base pairs of the lesion-containing 6-mer (cyan base pairs). The time dependent values are plotted against the time axis. The initial binding state values are in blue (1 to 1.5 μs), the starting state values (0 to 1 ns) are in red and their mean value is shown in red line. (C) Best-representative structure of the initial binding state ensemble derived from Model 2. The structure is

rendered as in Figure 3 of the main text and is viewed into the minor groove. The DNA residues are also shown in transparent surface, except the lesion-containing base and its partner base. The DNA backbone heavy atoms and the amino acids that form hydrogen bonds (black dash lines) with DNA are shown in sticks, and the side chains of Thr 516 and Val 517 are shown in spheres. These amino acids are color coded by atom type with carbon atom in the same color as the domain. At the initial binding state. Thr 516 inserts into the minor groove and Val 517 interacts with DNA backbone via van der Waals interactions, while Lys 521 and arg 494 hydrogen bond to the phosphate group of the DNA backbone. **(D)** Alpha space analyses of BHD2 binding to the lesion-containing DNA. The structure is rendered as in Figure S2D. **(E)** MD simulation started from Model 1: translocation of Rad4 and best representative structure. TGD domain rotation angle around the helix axis is plotted against time axis. Red line shows the average value before translocation. Decreased values show translocation along the helix axis toward 3' side of lesion containing strand. The best representative structure is obtained for the last 500 ns (orange ensemble). The zoomed in view of the best representative structure shows that the insertion of the BHD2 hairpin at the lesion site is blocked by the bulky B[a]P rings in the minor groove.

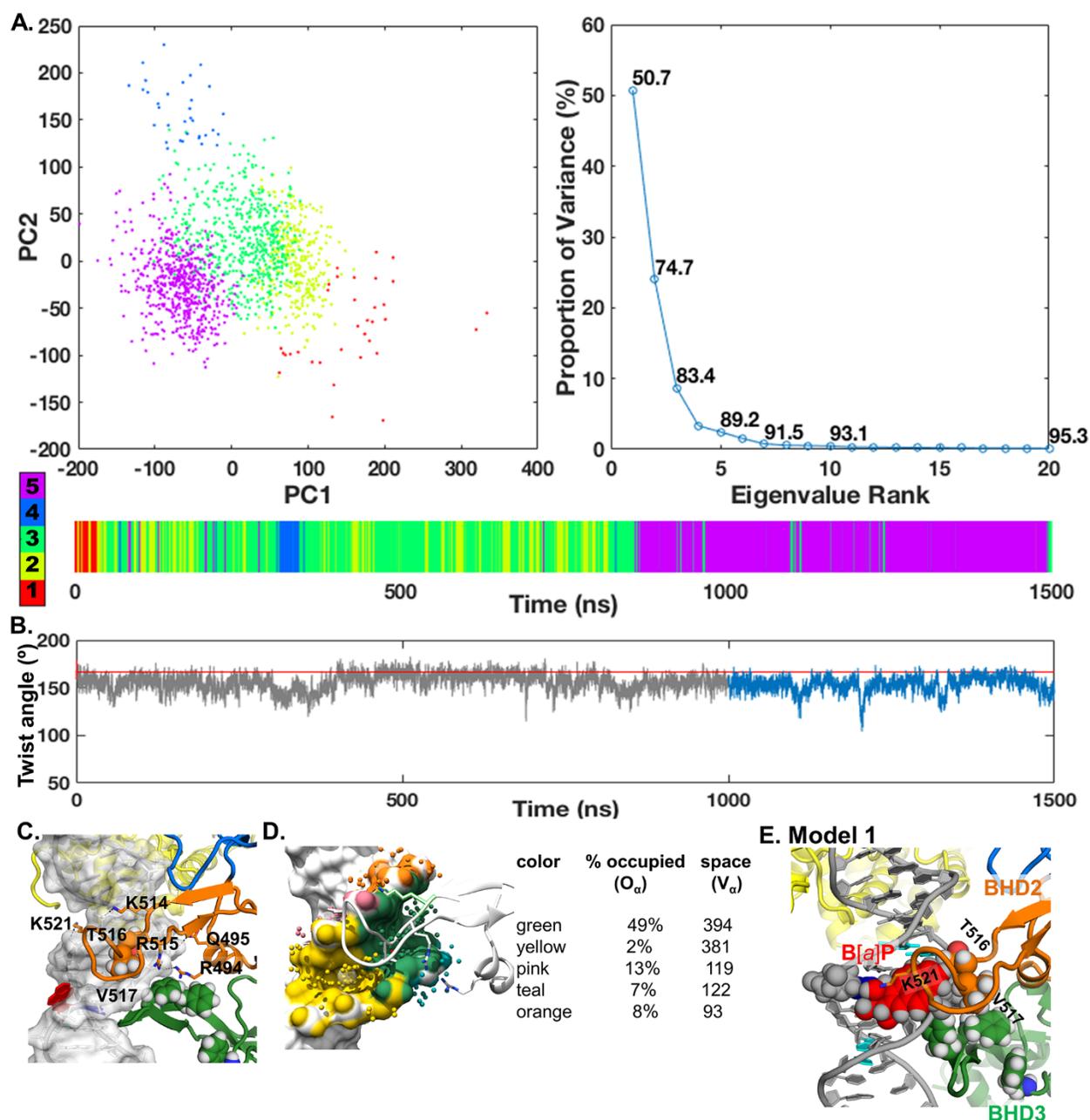


Figure S11 The initial binding of Rad4 to the (-) *trans*-B[a]P-dG-containing duplex.

(A) Clustering for the structural ensemble of the initial binding state. The first two principle components (PC) are plotted against each other. The proportions of variance are plotted for the first twenty PCs, where the cumulative proportions are labeled. The clusters are also given in color code along the time axis. (B) Twist angle between the two end base pairs of the lesion-containing 6-mer (cyan base pairs). The time dependent values are plotted against the time axis. The initial binding state values are in blue (1 to 1.5 μ s), the starting state values (0 to 1 ns) are in red and their mean value is shown in red line. (C) Best-representative structure of the initial binding state ensemble derived from Model 2. The structure is rendered as in Figure 3 of the main text and is viewed into the minor groove. The DNA residues are also

shown in transparent surface, except the lesion-containing base and its partner base. The DNA backbone heavy atoms and the amino acids that form hydrogen bonds (black dash lines) with DNA are shown in sticks, and the side chains of Thr 516 and Val 517 are shown in spheres. These amino acids are color coded by atom type with carbon atom in the same color as the domain. At the initial binding state, Thr 516 and Val 517 insert into the minor groove and moderately unwind the DNA duplex. Arg 494, Gln 495, Lys 514, Arg 515 and Lys 521 hydrogen bond with the DNA backbone. **(D)** Alpha space analyses of BHD2 binding to the lesion-containing DNA. The structure is rendered as in Figure S2D. **(E)** Blocking of BHD2 by the B[a]P rings in Model 1. The best representative structure is obtained for the last 500 ns ensemble from the MD simulation started with Model 1.

SUPPLEMENTARY MOVIE TITLES

Movie S1. Initial binding state of the (+) *cis*-B[*a*]P-dG:dC duplex shown in Figure 3.

Movie S2. Initial binding state of the PhIP-C8-dG:dC duplex shown in Figure 3.

Movie S3. Initial binding state of the 14*R*-DB[*a,l*]P-dG:dC duplex shown in Figure 3.

Movie S4. Initial binding state of the (+) *cis*-B[*a*]P-dG:deletion duplex shown in Figure 3.

Movie S5. Initial binding state of the (+) *cis*-B[*a*]P-dG:dA duplex shown in Figure 3.

Movie S6. Initial binding state of the 14*R*-DB[*a,l*]P-dA:dT duplex shown in Figure 3.

Movie S7. Initial binding state of the PhIP-C8-dG:deletion duplex shown in Figure 3.

Movie S8. Initial binding state of the (+) *cis*-B[*a*]P-dG:dT duplex shown in Figure 3.

Movie S9. Initial binding state of the (+) *trans*-B[*a*]P-dG:dC duplex shown in Figure 3.

Movie S10. Initial binding state of the (-) *trans*-B[*a*]P-dG:dC duplex shown in Figure 3.

SUPPLEMENTARY METHODS

Molecular modeling and structure preparation

The initial model. The ‘docking complex’ models were based on a previously obtained MD equilibrated docking complex for Rad4-Rad23 with a CPD containing DNA duplex with mismatched partner thymines⁴. In brief, an apo Rad4 model, based on the crystal structure of apo Rad4 (PDB ID: 2QSF⁵), was docked on to the damaged DNA with its TGD and BHD1 domains positioned as in the crystal structure of the productive open complex (PDB ID: 2QSG⁵). The docked complex was subjected to MD simulation with a restraint on the distance between BHD2-BHD3 domains and the lesion site, which provided an equilibrated ‘docking complex’ of the Rad4 with CPD-containing DNA⁴. For the present study, the DNA duplex of the equilibrated ‘docking complex’ was modified to the lesion-containing duplexes under investigation. The lesion sites were replaced with their NMR/MD-derived structures (Figure 2 in the main text). In addition, the DNA ends were extended by four base pairs to the 5’ end and by two base pairs to the 3’ end of the lesion-containing strand (Figure S1) using standard B-DNA in Discovery Studio 2.5 (Accelrys Software Inc.).

Protonation. The protonation states of charged residues were determined by the H++⁶ and pdb2pqr⁷ methods. When the two methods disagreed, protonation states were assigned based on the following criteria: the potential H-bonding network, solvent exposure of the ionizable residues, potential steric clashes if the proton was added, and preservation of the crystal structure. The results are given in⁴.

Water box and counterions. The protein-DNA complex was neutralized with Na⁺ counterions and solvated with explicit TIP3P water⁸ in a cubic periodic box with side length of 125.0 Å using the tLEAP module of the AMBER16⁹ suite of programs.

Force field and MD simulation protocols

We used the ff14SB force field¹⁰ for all MD simulations. Partial charges and parameters for the lesions were obtained from^{2, 11-13}.

All MD simulations were carried out using the AMBER16⁹ suite of programs. The Particle-Mesh Ewald method¹⁴ with 9.0 Å cutoff for the non-bounded interactions was used in the energy minimizations and MD simulations. Minimizations were carried out in three stages. First, 500 steps of steepest descent minimization followed by 500 cycles of conjugate gradient minimization were conducted on the water molecules and counterions with a restraint force constant of 50 kcal/(mol·Å²) on the solute molecules (DNA and protein complex). Then, 500 steps of steepest descent minimization followed by 500 cycles of conjugate gradient minimization were carried out on the water molecules and counterions with a restraint force

constant of 10 kcal/(mol·Å²) on the solute molecules. In the last round, 500 steps of steepest descent minimization followed by 500 cycles of conjugate gradient minimization were carried out on the whole system without restraints. The minimized structure was then subjected to three rounds of equilibration. First, each system was equilibrated at constant temperature of 10K for 30ps with the solute molecules fixed with a restraint force constant of 25 kcal/(mol·Å²). Then the system was heated from 10 K to 300 K over 30 ps with the solute molecules fixed with a restraint force constant of 10 kcal/(mol·Å²) at constant volume. In the last round of equilibration, the restraint force constant on the solute was reduced through three steps: at 10 kcal/(mol·Å²) for 30 ps, at 1 kcal/(mol·Å²) for 40 ps, then at 0.1 kcal/(mol·Å²) for 50 ps with constant pressure at 300K. Following equilibration, production MD simulations for each system were carried out in an NPT ensemble at 300K and constant pressure of 1 Atm. The temperature was controlled with a Langevin thermostat¹⁵ with a 5 ps⁻¹ collision frequency. The pressure was maintained with the Berendsen coupling method¹⁶. The production MDs were carried out with 1 kcal/(mol·Å²) restraints on the distances between each pair of hydrogen bonded heavy atoms at the end base pairs, but not on any other residues of the complex. A 2.0 fs time step and the SHAKE algorithm¹⁷ were applied in all MD simulations.

Structural analyses

The structures along each trajectory were clustered using the principle component analysis (PCA) method in the Bio3D package¹⁸. The clustering was performed with 1500 frames selected every 1 ns from each 1.5 μs trajectory. The structures from each ensemble were superposed to the first frame at the heavy atoms of the base pairs that are non-specifically bound to TGD and BHD1 (base pair 14 – 24, Figure S1). PCA calculations were performed for the heavy atoms of the lesion-containing 6-mer and the protein backbone atoms of BHD2. Hierarchical clustering of structures in PC space was performed along the first seven PCs using R¹⁹. The clustering results are given in Figures S2-S11.

The initial binding state was defined as a stable conformational cluster with stable twist angle between the end base pairs of the lesion-containing 6-mer. This state was achieved before 1 μs for all lesions (Figures S2-S11). The structural ensembles between 1 and 1.5 μs were collected and analyzed to characterize the initial binding states.

We measured the twist angles for the lesion-containing 6-mer (Figure S1) using the CPPTRAJ module of AMBER16²⁰. The time dependence values of the twist angle for each duplex are shown in Figure S2-S11. To reflect duplex untwisting around the lesion site, we calculated the untwist angles for the initial binding state structures (Untwist = Twist_{start} - Twist_{initial binding state}). Twist_{start} is the ensemble average twist angle of the lesion-containing 6-mer during the first 1 ns of production MD during which significant untwisting was not observed (Figures S2-S11); this ensemble represents the state of the lesion-containing sequence before the engagement of BHD2. Twist_{initial binding state} is the twist angle of the lesion-containing 6-mer for the structures in the initial binding state ensemble (1 - 1.5 μs). Positive values indicate further untwisting and

negative values indicate further twisting. The untwist angle values for each initial binding state ensembles were analyzed using the block averaging method^{21,22}. In brief, the time series data were divided into “blocks” with a block size that exceeds the longest correlation time, 20 ns in our case. The average for each block was computed and termed “block average”. The mean values and the standard deviations of the block averages (Figure 3A in the main text) were used to represent the average and the variance of averages.

Displacement of the partner nucleotide for the 14*R*-DB[*a,l*]P-dG:dC case was measured using a nucleotide flipping pseudo-dihedral angle that reflect its direction and degree of flipping around the DNA backbone. The pseudo-dihedral angle, defined in Figure S4, increases when flipping is toward the major groove.

The best representative structure is defined as the one frame that has the shortest RMSD for the heavy atoms of the lesion-containing 6-mer and the protein backbone atoms of BHD2 to all other frames in each ensemble.

The best representative structure for the initial binding state of each lesion-containing duplex was further analyzed to quantify BHD2’s binding into the DNA minor groove around the lesion site. The alpha space volumes ($V\alpha$) of the binding pockets in the DNA and their occupancies ($O\alpha$) by BHD2 were calculated using AlphaSpace v1.0²³: the DNA duplex was set as the receptor and BHD2 was set as the ligand. The total occupied alpha space volume ($\sum V\alpha \times O\alpha$) was used to quantify the extent of BHD2 binding into the DNA minor groove (Figure S2-S11). The value reflects the curvature and surface area of the DNA minor groove region that is occupied by BHD2.

All molecular structures were rendered using PyMOL 1.3.x (Schrodinger, LLC.). All alpha space figures are render using UCSF Chimera 1.10.2 (developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311)²⁴. All data were plotted using MATLAB R2017b (The MathWorks, Inc.).

SUPPLEMENTARY REFERENCES

- (1) Cai, Y., Geacintov, N. E., and Broyde, S. (2012) Nucleotide excision repair efficiencies of bulky carcinogen-DNA adducts are governed by a balance between stabilizing and destabilizing interactions. *Biochemistry* *51*, 1486-1499.
- (2) Tang, Y., Liu, Z., Ding, S., Lin, C. H., Cai, Y., Rodriguez, F. A., Sayer, J. M., Jerina, D. M., Amin, S., Broyde, S., and Geacintov, N. E. (2012) Nuclear magnetic resonance solution structure of an N(2)-guanine DNA adduct derived from the potent tumorigen dibenzo[a,l]pyrene: intercalation from the minor groove with ruptured Watson-Crick base pairing. *Biochemistry* *51*, 9751-9762.
- (3) Mu, H., Geacintov, N. E., Min, J. H., Zhang, Y., and Broyde, S. (2017) Nucleotide excision repair lesion-recognition protein Rad4 captures a pre-flipped partner base in a benzo[a]pyrene-derived DNA lesion: how structure impacts the binding pathway. *Chem. Res. Toxicol.* *30*, 1344-1354.
- (4) Mu, H., Geacintov, N. E., Zhang, Y., and Broyde, S. (2015) Recognition of damaged DNA for nucleotide excision repair: a correlated motion mechanism with a mismatched *cis-syn* thymine dimer lesion. *Biochemistry* *54*, 5263-5267.
- (5) Min, J. H., and Pavletich, N. P. (2007) Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* *449*, 570-575.
- (6) Anandakrishnan, R., Aguilar, B., and Onufriev, A. V. (2012) H++3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* *40*, W537-W541.
- (7) Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* *32*, W665-W667.
- (8) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* *79*, 926-935.
- (9) Case, D. A., Babin, V., Berryman, J. T., Betz, R. M., Cai, Q., Cerutti, D. S., Cheatham, T. E., 3rd, Darden, T. A., Duke, R. E., Gohlke, H., Goetz, A. W., Gusarov, S., Homeyer, N., Janowski, P., Kaus, J., Kolossváry, I., Kovalenko, A., Lee, T. S., LeGrand, S., Luchko, T., Luo, R., Madej, B., Merz, K. M., Paesani, F., Roe, D. R., Roitberg, A., Sagui, C., Salomon-Ferrer, R., Seabra, G., Simmerling, C. L., Smith, W., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., Wu, X., and Kollman, P. A. (2014) AMBER14, University of California, San Francisco.
- (10) Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* *11*, 3696-3713.
- (11) Mocquet, V., Kropachev, K., Kolbanovskiy, M., Kolbanovskiy, A., Tapias, A., Cai, Y., Broyde, S., Geacintov, N. E., and Egly, J. M. (2007) The human DNA repair factor XPC-HR23B distinguishes stereoisomeric benzo[a]pyrenyl-DNA lesions. *EMBO J.* *26*, 2923-2932.
- (12) Reeves, D. A., Mu, H., Kropachev, K., Cai, Y., Ding, S., Kolbanovskiy, A., Kolbanovskiy, M., Chen, Y., Krzeminski, J., Amin, S., Patel, D. J., Broyde, S., and Geacintov, N. E. (2011) Resistance of bulky DNA lesions to nucleotide excision repair can result from extensive aromatic lesion-base stacking interactions. *Nucleic Acids Res.* *39*, 8752-8764.
- (13) Cai, Y., Ding, S., Geacintov, N. E., and Broyde, S. (2011) Intercalative conformations of the 14R (+)- and 14S (-)-trans-anti-DB[a,l]P-N(6)-dA adducts: molecular modeling and MD simulations. *Chem. Res. Toxicol.* *24*, 522-531.

- (14) Darden, T., York, D., and Pedersen, L. (1993) Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* *98*, 10089-10092.
- (15) Loncharich, R. J., Brooks, B. R., and Pastor, R. W. (1992) Langevin Dynamics of Peptides - the Frictional Dependence of Isomerization Rates of N-Acetylalanyl-N'-Methylamide. *Biopolymers* *32*, 523-535.
- (16) Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., and Haak, J. R. (1984) Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* *81*, 3684-3690.
- (17) Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* *23*, 327-341.
- (18) Skjaerven, L., Yao, X. Q., Scarabelli, G., and Grant, B. J. (2014) Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinformatics* *15*, 399.
- (19) R Core Team. (2017) R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria.
- (20) Case, D. A., Betz, R. M., Cerutti, D. S., Cheatham, I., T.E., Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H., Goetz, A. W., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T. S., LeGrand, S., Li, P., Lin, C., Luchko, T., Luo, R., Madej, B., Mermelstein, D., Merz, K. M., Monard, G., Nguyen, H., Nguyen, H. T., Omelyan, I., Onufriev, A., Roe, D. R., Roitberg, A., Sagui, C., Simmerling, C. L., Botello-Smith, W. M., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., Wu, X., Xiao, L., and Kollman, P. A. (2016) AMBER 16, University of California, San Francisco, San Francisco, CA.
- (21) Flyvbjerg, H., and Petersen, H. G. (1989) Error-Estimates on Averages of Correlated Data. *J. Chem. Phys.* *91*, 461-466.
- (22) Yang, W., Bitetti-Putzer, R., and Karplus, M. (2004) Free energy simulations: Use of reverse cumulative averaging to determine the equilibrated region and the time required for convergence. *J. Chem. Phys.* *120*, 2618-2628.
- (23) Rooklin, D., Wang, C., Katigbak, J., Arora, P. S., and Zhang, Y. K. (2015) Alpha Space: Fragment-Centric Topographical Mapping To Target Protein-Protein Interaction Interfaces. *J. Chem. Inf. Model.* *55*, 1585-1599.
- (24) Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* *25*, 1605-1612.