10912 *Biochemistry* **2010,** *49,* 10912–10924
DOI: 10.1021/bi1012035

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Modeling the Interactions of the Nucleotide Excision Repair $UvrA_2$ Dimer with DNA[†]

Tsvetan G. Gantchev* and Darel J. Hunting

Center for Research in Radio-Oncology (CR2), Department of Nuclear Medicine and Radiobiology, Faculty of Medicine and Life Sciences, Université de Sherbrooke, Sherbrooke, Québec, J1H 5N4 Canada

Received July 29, 2010; Revised Manuscript Received November 19, 2010

ABSTRACT: The UvrA protein initiates the DNA damage recognition process by the bacterial nucleotide excision repair (NER) system. Recently, crystallographic structures of holo-$UvrA_2$ dimers from two different microorganisms have been released (Protein Data Bank entries 2r6f, 2vf7, and 2vf8). However, the details of the DNA binding by $UvrA_2$ and other peculiarities involved in the damage recognition process remain unknown. We have undertaken a molecular modeling approach to appraise the possible modes of DNA−$UvrA_2$ interaction using molecular docking and short-scale guided molecular dynamics [continuum field, constrained, and/or unrestricted simulated annealing (SA)], taking into account the three-dimensional location of a series of mutation-identified UvrA residues implicated in DNA binding. The molecular docking was based on the assumptions that the $UvrA_2$ dimer is preformed prior to DNA binding and that no major protein conformational rearrangements, except moderate domain reorientations, are required for binding of undamaged DNA. As a first approximation, DNA was treated as a rigid ligand. From the electrostatic relief of the ventral surface of $UvrA_2$, we initially identified three, noncollinear DNA binding paths. Each of the three resulting nucleoprotein complexes (C1, C2, and C3) was analyzed separately, including calculation of binding energies, the number and type of interaction residues (including mutated ones), and the predominant mode of translational and rotational motion of specific protein domains after SA to ensure improved DNA binding. The $UvrA_2$ dimer can accommodate DNA in all three orientations, albeit with different binding strengths. One of the $UvrA_2$−DNA complexes (C1) fulfilled most of the requirements (high interaction energy, proximity of DNA to mutated residues, etc.) expected for a natural, high-affinity DNA binding site. This nucleoprotein presents a structural organization that is designed to clamp and bend double-stranded DNA. We examined the binding site in more detail by docking DNAs of significantly different (AT- vs CG-enriched) sequences and by submitting the complexes to DNA-unrestricted SA. It was found that in a manner independent of the DNA sequence and applied MD protocols, $UvrA_2$ favors binding of a bent and unwound undamaged DNA, with a kink positioned in the proximity of the Zn3 hairpins, anticollinearly aligned at the bottom of the ventral protein surface. It is further hypothesized that the Zn3 modules play an essential role in the damage recognition process and that the apparent existence of a family of DNA binding sites might be biologically relevant. Our data should prove to be useful in rational (structure-based) mutation studies.

Nucleotide excision repair (NER) is a universal cellular mechanism committed to the recognition and removal of bulky and structurally unrelated DNA lesions (thymidine photodimers, psoralen, mitomycin C and fluorescein DNA conjugates, DNA intra- and interstrand cross-links, etc.) (*1*−*6*). In prokaryotes, this multistep process is initiated by an ATP-dependent enzyme complex, the ABC nuclease, comprised of three proteins, UvrA, UvrB, and UvrC with relative molecular masses, M_r (*Escherichia coli*), of 103, 76, and 66 kDa, respectively (*7*). Bacterial NER includes the following steps: (1) initial DNA damage recognition (formation of the $UvrA_2$−DNA nucleoprotein), (2) damage verification (recruitment of UvrB by the $UvrA_2$−DNA complex), (3) dissociation of UvrA from the tertiary complex (formation of a stable UvrB−DNA preincision complex), and (4) binding of UvrC to the UvrB−DNA complex and DNA strand incision (the nuclease reaction, per se). The remaining steps of DNA repair are completed by UvrD (DNA helicase II) and DNA polymerase II, while the newly synthesized nucleotide is sealed to the parental strand by DNA ligase (*8*, *9*). Insights into the primary processes of DNA damage recognition are obtained from X-ray diffraction studies of the damage-sensing UvrABC components, the UvrA and UvrB proteins. The crystal structure of UvrB from different microorganisms has been known for more than a decade (*10*−*12*). More recently, crystal structures of UvrB−DNA complexes have been determined (*13*, *14*). However, only holo-$UvrA_2$ dimer crystal structures (without bound DNA) have been reported. These structures [Protein Data Bank (PDB) entries 2r6f (*15*) and 2vf7 and 2vf8 (*16*)] are derived from two homologous but not identical proteins, isolated from two different sources, *Bacillus stearothermophilus* (bstUvrA) and *Deinococcus radiodurans* (drUvrA), which together with the recently released structure of the isolated partial UvrA−UvrB complex [interaction domains only; PDB entry 3fpn (*17*)] provide new knowledge about the intriguing yet largely obscure mechanism of DNA damage recognition by the ABC excinuclease. Because the explicit three-dimensional (3D) molecular structure of the $UvrA_2$−DNA complex is still unavailable, we have undertaken this study to build

[†]This research was supported by the Cancer Research Society (Montreal, QC) and the National Sciences and Engineering Research Council (NSERC, Ottawa, Canada).

*To whom correspondence should be addressed. Telephone: (819) 564-5401. Fax: (819) 564-5442. E-mail: tsvetan.gantchev@usherbrooke.ca.

 Published on Web 11/23/2010

a 3D structural model of the $UvrA_2$−DNA complex by applying molecular modeling and restricted molecular dynamics [simulated annealing (SA)]. Our primary goal was to determine the preferential undamaged DNA binding site(s) within the $UvrA_2$ ventral surface and to evaluate the structural determinants engaged in DNA binding. Guidelines for our docking experiments were the general assumptions outlined by the authors of the first bst$UvrA_2$ crystallographic study (*15*), namely, the $UvrA_2$ electrostatic surface and groove curvature coupled with the point mutations indicating the role of selected basic protein residues in the signature 2 (SGN2) domains. It was assumed that DNA passes between the insertion domains (ID) of the two protomers of the dimer. Because the IDs in this crystallographic structure are in an "open state", there are several possible DNA sites that could be derived simply by DNA rotation in approximately the same plane, which may represent a family of binding sites. In contrast, the more recent dr$UvrA_2$ study by Timmins et al. (*16*) depicts a "closed ID" configuration and indicates the DNA binding role of several conserved residues (point mutations) within the same domain. These findings presuppose a more restricted DNA positioning within the $UvrA_2$ ventral side coupled with interdomain rearrangements. We examined three significantly distinct DNA orientations along the ventral surface of the bst$UvrA_2$ dimer as possible paths of DNA alignment. To elucidate the potential role of coordinated domain motion upon DNA binding, we applied restricted short-scale molecular dynamics (SA). Our results indicate that there is one preferential high-affinity site, while a number of weaker binding sites may represent transient modes of DNA binding.

METHODS

All molecular modeling tasks were performed using the Tripos, Inc., molecular modeling package Sybyl version X.1.1 (and earlier) using standard computational and analytical tools: MaxiMin, FlexiDock, Simulated Annealing, Prodat, Protein Loops, Protein Table, Sequence Viewer, Structure Comparison (homology alignment), Connolly electron-density, etc., as incorporated within the Sybyl Base, Advanced Computing, Biopolymer, Dynamics, and Molcad modules. The starting $UvrA_2$ structure was that of *B. stearothermophilus* [PDB entry 2r6f (bstUvrA, class I UvrA protein, 3.2 Å resolution) (*15*)], which at different simulation stages was compared with multiple structures of *D. radiodurans* dr$UvrA_2$ [PDB entries 2vf7 and 2vf8 (resolved at 2.3 Å) (*16*)]. The latter is representative of the class II UvrA proteins, lacking the UvrB binding domain. A standard double-strand B-DNA molecule 41 bp in length was built with Sybyl Biopolymer, which employs the Arnott algorithm. The Amber99 force field with dictionary-defined atomic charges was used throughout this study. All force field computations were performed using separate evaluation of steric and electrostatic interactions with a distance-dependent dielectric function ($D = 2$) to compensate for the lack of counterions and solvent (*18*). The chosen electrostatic interaction cutoff was 10 Å.

Domain Organization of the $UvrA_2$ Dimer and Preparation for DNA Docking. An exhaustive description of the bst$UvrA_2$ dimer structure (PDB entry 2r6f) is presented in ref *15*. In brief, bstUvrA assembles as a head-to-head dimer, where each of the 949-amino acid monomer chains (residue coordinates in certain flexible loops are missing) is comprised of 40 α-helices and 30 relatively short β-strands and contains three Zn modules (Figure 1A). Two of the Zn sites (Zn-1 and Zn-2) do not adopt a classical Zn finger structure. The dimer was cocrystal-

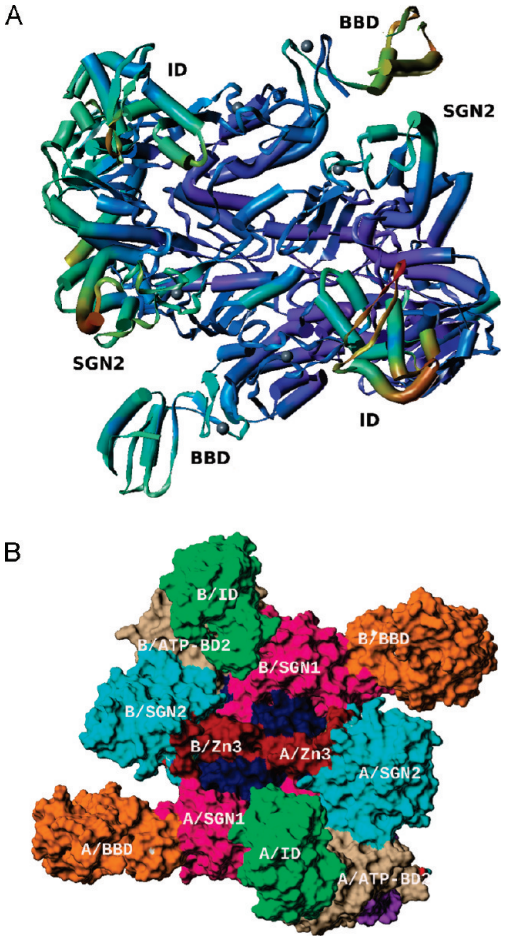


FIGURE 1: (A) Structural elements of the original (none repaired) $UvrA_2$ dimer as defined in PDB entry 2r6f (ribbon and tube) and colored by *R* factor (red and yellow for high and blue and violet for low), with the six Zn atoms (gray spheres). (B) Domain structure of the repaired holo-$UvrA_2$ dimer (Connolly surface). Colors and residue definition: ATP-BD1 (indigo, 1−87, 503−590), SGN1 (pink, 88−117, 257−286, 399−502), ATP-BD2 (beige, 609−686, 843−949), SGN2 (cyan, 687−842), BBD (orange, 118−256), ID (green, 287−398), ATP−BD1−BD2 linker (magenta, 591−608), and Zn3 hairpin (brown, 736−762).

lized with two Mg-ADP ligands at each of the two ATP-binding sites. Each of the two UvrA protomers has Walker A and B motifs and Q, D, and H loops that are conserved within all ABC-type ATPases. The Q loops join signature domains to the corresponding ATP-binding domains and possibly transmit interdomain allosteric interactions during ATP binding and hydrolysis. The $UvrA_2$ structural elements have been designated as nucleotide-binding domains, NBD-I and NBD-II, connected by a flexible linker and were further divided into ATP-binding domains (ATP-BD1 and ATP-BD2) and signature domains (SGN1 and SGN2). Inserted in the amino terminus of NBD-I are two sequences that are specific for the UvrA-type ATPase: the UvrB-binding domains (BBD) and the insertion domains (ID).

First, a standard check of amino acid residue structure and atoms was performed using the Sybyl Biopolymer preparation tool. The selenomethionine residues were replaced with regular methionine. Next, gaps within individual $UvrA_2$ chains were analyzed, and missing residues in thermally flexible loops were filled using the Protein Loops of Sybyl Biopolymer. To increase the number of possible matches, the standard Prodat database of Sybyl was appended with single-chain structures extracted from PDB entry 2r6f itself, PDB entry 3fpn (contains a long loop,

missing in the UvrB-binding domain of PDB entry 2r6f), and the homologous, multistructure protein complexes of PDB entries 2vf7 and 2vf8. The gap-sealing loop selection criteria included homology score and root-mean-square (rms) fit for the six existing residues flanking both sides of the gap window. Not surprisingly, loops from the added sequences ranked at the top of the search lists. Hydrogens have been added; amides have been examined and fixed if needed, and rotamers have been analyzed and corrected if necessary. All basic residues were kept protonated and acidic ones deprotonated. The repaired $UvrA_2$ molecule was assigned Amber99 atom types and charges and was energy-minimized using the Powell gradient algorithm, reaching a potential energy of approximately −22000 kcal/mol. The original bstUvrA2 structure is not fully symmetrical. For example, backbone alignment of A and B chains gives a total rms deviation (rmsd) of 1.7 Å; ATP-BD1 (A) and ATP-BD1 (B) alignment gives a rmsd of 0.3 Å, while ID(A) and ID(B) are significantly displaced from each another (rmsd = 4.6 Å), under the conditions in which dimer core domains of the two chains (ATP-BD1 and SGN1) were aligned. Following initial examination of the interchain domain positions, flexibility, and motion after trial simulated annealing runs (see Results, holo-1), we concluded that for the DNA docking it would be advantageous to use a completely symmetrical starting $UvrA_2$ structure. Because of the more complete assignment of the structural elements of the B chain, an extracted B chain was aligned with the A chain of the original dimer. Thereafter, the A chain was deleted, and the fitted B chain was merged with the original molecule with all cofactors in place. Thus, the new (working) $UvrA_2$ dimer is composed by two identical B chains (protomers) and was subjected to multistep minimization (conjugated gradient), below termed the "holodimer" and used in the subsequent simulations.

DNA Docking. We kept two assumptions in mind while performing the DNA docking experiments. (i) No major (domain) reorientations are necessary to fit long chain DNA in the preselected sites. (ii) As a first approximation, DNA was treated as a rigid body ligand, while possible DNA deformations within the binding site were addressed subsequently. For the initial docking experiments, we built a random 41 bp DNA double helix with an AT-enriched central segment, 5′d(GGGGGGAACGTACGTACGTTTTTACGTACGTTTTTTCCCCC)·5′d(GGGGGAAAAAACGTACGTAAAAACGTACGTACGTTCCCCCC), termed DNA(1). The energy-refined DNA structure was slightly bent (6 ± 2°). In this sequence, the central T tract was included as a presumable damage modification site to be studied in the future, e.g., construction of T-T photodimers, fluorescein-adducted dT, etc. The DNA is 8 bp longer (four at each end) than the known $UvrA_2$-protected DNA based on DNase-I footprinting (~33 bp) (*19*). Docking a 41 bp double-strand DNA helix (catenated ligand) presents a challenge by itself, and automatic docking, e.g., via application of the FlexX program, failed. Therefore, initially, we performed manual positioning of DNA along the hypothetical DNA binding paths (see Results). The adjustment of DNA position was executed using the interactive Sybyl7.2 docking utility that employs creation of a precomputed box region surrounding the ligand, where steric and electrostatic contributions of the interaction energy are computed and large repulsive forces indicated as vectors. After the initial DNA alignment over the proposed binding sites, full-scale docking experiments were performed using FlexiDock. The program allows selection of H-bonding donor−acceptor partners between the protein pocket residues and the ligand, as well as assignment of flexible bonds for rotations in both the site and the ligand. In the experiments described here, no rotational bonds in the DNA backbone were selected. The multiple docking results were examined via creation of Moltable spreadsheets and docked ligands ranked using the Consensus Score (CS) algorithm of Sybyl. The DNA binding properties of the high-affinity binding site (site 1) were further examined by docking another, more rigid DNA ligand [DNA(2)], entirely composed of CG segments [5′d(GGGGGCCCGGCCGGCCGGGGGGCCGCGCGGGGGGGGCCCCC)·5′d(GGGGGCCCCCCCCGCGCGGCCCCCCGGCCGGCCGGGCCCCC)]. The energy-refined structure of DNA(2) exhibits a negligible overall bending angle of approximately 2 ± 1°.

Simulated Annealing. Because the $bstUvrA_2$ holoprotein crystal structure exhibits "open ID" features, in contrast to the "closed ID" $drUvrA_2$ structure (*16*), and the IDs contain some of the major DNA binding determinants, we hypothesized that by allowing limited protein domain motion of the DNA-bound bstUvrA, it may further tighten its grasp on DNA, thus providing deeper insight into formation of the $UvrA_2$−DNA complex and possible DNA deformations. To achieve this, we selected the technique of simulated annealing (SA) in which initially some structural components were kept as still aggregates. In the trial runs, we examined the motion and the structural deformations of the individual components in the holo dimer and the whole DNA−protein complex. We have noted that protein domains in the holo dimer (mainly SGN2, IDs, and BBDs) tend to move in a reversible manner, when one to three subsequent SA runs were performed, while other domains, especially ATP-BD1, remained practically rigid. In the DNA−protein complexes, the domain motion was unidirectional (toward DNA). In the initial MD runs, we attempted a "guided" molecular dynamics simulation, either by keeping certain components of the system as unmovable (unheated) aggregates (DNA and/or some protein domains, ATP-BD-1 and -2 and SGN-1) or by imposing distance, e.g., C1′−C1′, and torsion (δ) angle constraints on DNA, while at the same time optimizing the molecular dynamics parameters (heating and cooling phases). We found that very good reproduction of the results (overall rmsd of < 1.5 Å) takes place if DNA motion is subdued, at least during the initial SA run. In addition, because ATP-BD1 behaved as a rigid structure (and more distant from DNA), we have chosen to keep them as aggregates (for reference purposes), unless otherwise stated. The molecular dynamics (MD) parameters were as follows. The rapid heating to the 320 K phase has a duration of 2.7 ps (500 fs/50 K), followed by gradual cooling in 2 ps steps to 50 K (total of ~57 ps). The last step at 50 K was 4 ps for better equilibrium (for a typical thermodynamic trajectory, see Figure 2S of the Supporting Information). Other parameters were as follows: integration step, 1 fs; NB update, 25 fs; NB cutoff, 10 Å; thermal coupling, 100 fs (cooling phase). In subsequent runs, when DNA was unrestrained the Shake algorithm was applied to side chain heavy atom−hydrogen bonds. Heating to 320 K resulted in an effective temperature of ~300 K according to the system kinetic energy achieved. This protocol is similar to that previously described (*20*), except that the molecules used here were not hydrated. After the SA runs, the $UvrA_2$−DNA complexes were again energy-minimized. Interestingly, one step of SA was sufficient to reach a stable, more tightly bound DNA−$UvrA_2$ complex as judged by the rmsd values of the aligned structures and energy evaluations. The results from DNA docking along the proposed DNA binding paths followed by SA (with restricted DNA motion) indicated that none of these paths could accommodate the DNA(1) helix in

addition to satisfying the combined results of all available mutation studies (*15*, *16*, *19*), i.e., that the DNA passes within 6.0−6.5 Å of these residues (the range for efficient electrostatic interactions). Interestingly, unbent or slightly bent DNA leaves a relatively hollow space located in the middle of the dimer, encompassing the tips of the two Zn3 hairpin (Zn3-HP) motifs. Although it is unknown if undamaged DNA bends significantly when bound to the $UvrA_2$ dimer, it has long been hypothesized that the interaction with $UvrA_2$ forces damaged DNA to bend (to promote helix opening) and thus to facilitate the handoff to the UvrB protein for further damage verification (*21*). To address this question, for the path 1 DNA-binding site after the initial SA run with restrained DNA(1), we performed series of additional experiments, including fully unrestrained SA, and repeated all simulations after loading DNA(2) in the same binding site. Independently, the evaluation of DNA(1) bending in complex 1 was studied via application of a "split docking" approach: after binding site 1 has been determined and the binding efficiency increased following SA, the straight (rigid) DNA was removed from the complex and split into three parts [two of them with 18 bp each (denoted "left-hand" and "right-hand") and a middle portion of 9 bp]. The left and right segments were docked (rigid body) separately within the two preidentified and symmetry-related DNA binding sites of complex 1. The middle DNA portion that consisted of 9 bp from the original DNA with 2 bp overlapping the left and right segments at each end was docked separately within the wide cleft facing the Zn3-HPs. The successfully docked split DNA helices were examined, and the overlapping end nucleotides of each of the three segments (left, right, and middle) were evaluated in terms of rmsd. Ligands with the lowest rmsd values of their overlapping edge nucleotides were selected, and the extra 2 bp at each end of the middle segment were removed before the DNA segments were sealed to construct the 41 bp DNA. The reconstructed full-length DNA was energy-minimized alone or when $UvrA_2$ bound; in both cases, it remained bent and unwound.

RESULTS

bst$UvrA_2$ Ventral Surface and Selection of Binding Paths for DNA Docking. After the ATP-driven dimerization of the UvrA monomers, a saddle-shaped (ventral) surface is formed with the participation of the all four NBDs (two per subunit). The subunit interaction interface involves mostly residues from the ATP-BD1 and SGN1 domains. Unlike other ABC-ATPases, the nucleotide binding pockets are not located at the dimer interface. ATP-BD1 structures are solvent-exposed at the opposite, dorsal side with only a few residues protruding into the ventral part (some interact with the Zn3 hairpin of SGN2). SGN1 and ATP-BD2 form most of the semiplanar part of the ventral surface, while the SGN2 domains rise above that plane, rendering the surface more uneven. The two UvrB-binding domains (BBD) stand apart from the dimer core and stretch outward in opposite directions (Figures 1 and 2). The two IDs are located also at the opposite sides of the ventral dimer surface, significantly rising above the central (semiplanar) part. Remarkably, the nontypical Zn1- and Zn2-binding modules are components of loops that connect the most flexible BBD and ID with the relatively hindered SGN1 domain, which itself joins and partially overlays the most rigid ATP-BD1. The two Zn3 modules (hairpins) are counter-collinearly aligned at the center bottom of the ventral surface and are in steric contact with the underlying ATP-BD1. Their tips (Phe751) are in a van der Waals interaction range. Only a portion of the SGN2 domain, close to the Zn3 HP base, displays an

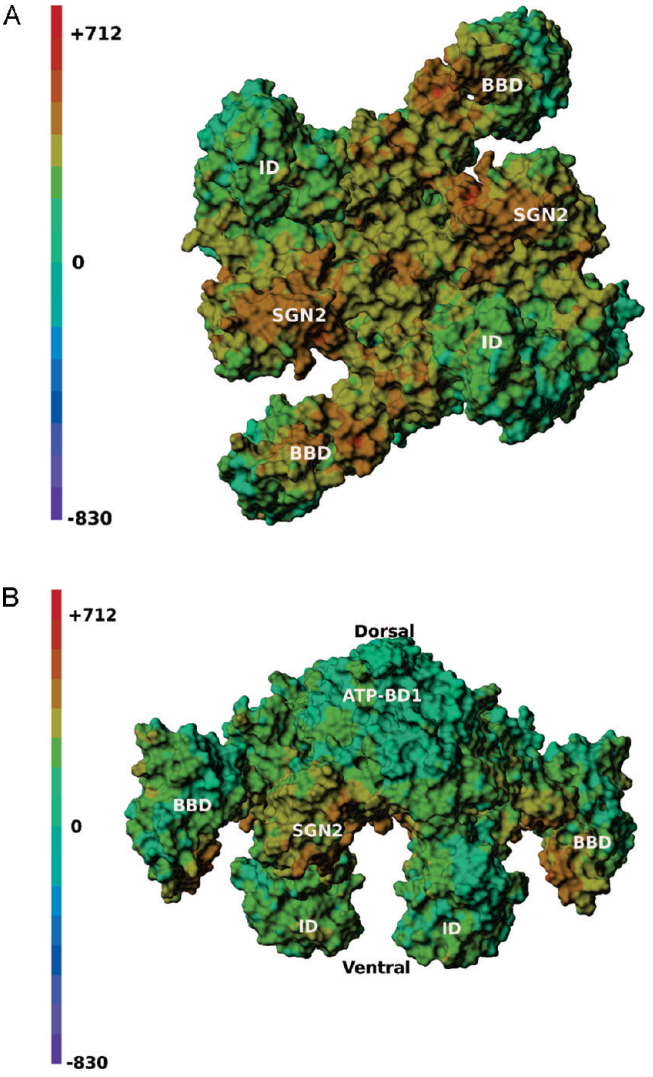


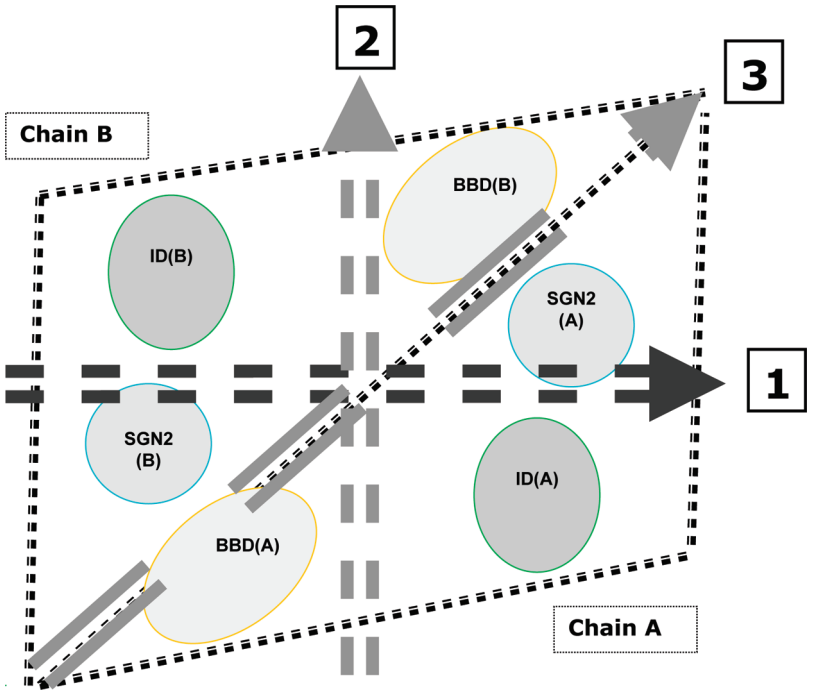
FIGURE 2: (A) Front view of the ventral surface of holo-$UvrA_2$ and (B) view flipped by 90° along the equatorial axis. The surface was built with Molcad and colored by electrostatic potential (EP) values. Some structural elements are labeled. The color ramp is in kilocalories per mole, per electron charge (left).

increased thermal flexibility (Figure 1A). Thus, the six Zn sites provide hingelike structural elements and probably are involved in coordinated interdomain motion upon cofactor binding (MgATP), DNA binding, and/or thermal activation.

Each UvrA protomer has 147 basic versus 138 acidic residues. The dimer, including the cofactors, four ADP molecules, and 10 metal ions (six Zn and four Mg atoms), bears a total of +27 AMBER7 extra charges. The electrostatic surface potential of the holo dimer is presented in Figure 2. The ventral surface is largely positive, whereas the dorsal surface is more negatively charged and its central part, formed mainly by the ATP-BD1, all negative. ATP-BD1 contains in total 38 basic versus 52 acidic residues. Distinct from the basal layer of the ventral surface, which is less positive, patches of high positive potential (brown and red in Figure 2) are located at the tips of the SGN2 domains and the stem regions of BBDs. The inward surfaces of the IDs are also entirely positively charged.

As mentioned, the ventral surface is not smooth. Apart from the clearly separated BBDs and IDs, the edges of ATP-BD2 and SGN2 are significantly out of the basal plane formed by the SGN1 domain and ATP-BD1 and are clearly distinct from the neighboring

Scheme 1: Three Possible DNA Tracks (arrows 1−3) along the Ventral Side of the UvrA2 Dimer (dashed-line triangles depict A and B protomers)[a]



[a]SGN1, ATP-BD1, and ATP-BD2 are not shown. In path 1, DNA is in a plane that makes an angle of ~45° with the plane through the dimer interface and symmetrically interacts with the SGN2 domain and ID of each protomer, but not with BBDs. In path 2, DNA is positioned in a direction nearly perpendicular to that of path 1 and interacts symmetrically with the SGN1 domain and BBD of each dimer subunit. In path 3, DNA lies in the A−B protomer interface plane and at each site interacts with the SGN2 domain of one protomer and the BBD from the other. In the two later paths, the "open-state" IDs are more distant from the DNA.

components (Figure 1B). Thus, the ventral surface relief displays one deep and two more shallow groovelike structures. These grooves were proposed as possible paths for DNA binding across the ventral side of the dimer, because they appear as long transdimer (positively charged) tracks with sufficiently large widths, capable not only of accommodating a 30−40 bp long DNA double helix but also of offering open DNA access. These three DNA directions, designated as paths 1−3, are presented in Scheme 1 as arrows and intersect in the center of the dimer. In all three situations, the $UvrA_2$ dimer provides two symmetric interaction sites, mainly formed by combinations between the three distinct and movable domains (SGN2, ID, and/or BBD), while the SGN1 domain and ATP-BD2 provide the basal support. In addition to the binding surface properties (electrostatic potential and relief), further arguments for the selection of these possible DNA binding paths arise from the estimated initial proximity of the known positions of point mutations implicated in undamaged or damaged DNA binding, 15 residues per monomer, divided into three sets. (i) Set 1 includes the three groups of SGN2 basic residues, according to Pakotiprapha et al. (*15*): (A) Lys732 and Arg735; (B) Lys765, Arg766, and Arg769; and (C) Arg708, Lys718, Arg720, and Arg726. (ii) Set 2 consists of the two residues, identified by Croteau et al. (*19*); one (Arg691) is located at the Zn3 finger base (SGN2) and the other (Lys680) in ATP-BD2. (iii) Set 3 is comprised of mutated residues shown to play a role in DNA binding by $drUvrA_2$ and located in the core region of the insertion domains (*16*). In $bstUvrA_2$, they were identified by homology alignment and are Arg381, Arg382, Arg384, and Arg392. During the initial positioning, not all of the mutation sites mentioned above fulfill the requirement of being situated within 6 Å of the DNA (approximate range for efficient electrostatic interactions). However, an additional guideline kept in mind during docking was the concept of "open and closed IDs" (*16*), i.e., their ability to undergo thermal motion synchronously. This property of $UvrA_2$ dimers is likely not limited to IDs only but possible for other hinge-bound (via Zn sites) structural elements, such as BBDs and SGN2 domains, as supported by our subsequent MD simulations both with the holoprotein and with the DNA complexes (see below).

The DNA(1) helix (41 bp) was positioned manually over the ventral surface in any of the three outlined directions as deep as possible without distortion of the backbone structure, although a few protein side chains clashed with DNA. Double-strand DNA was treated as a rigid body and its position refined interactively, followed by a full-scale FlexiDock execution. From the numerous output results, molecules at the top of the Sybyl MolTable, as ranked by the Consensus Score algorithm, were selected and analyzed. The best-fit protein−DNA(1) complexes were named complexes 01, 02, and 03 (abbreviated C01, C02, and C03, respectively), according to the three DNA main directions (the zeros distinguish these complexes from those obtained after SA). The DNA(1) complex resulting from docking in site 1 is shown in Figures 3A,B as domain-colored backbone ribbons in two views: face on the ventral side and in a perpendicular direction, along the DNA axis (for similar C02 and C03 graphics, see Figure 1S of the Supporting Information). A superimposed presentation of the all three studied DNA binding orientations within $UvrA_2$, together with some examples of low-ranking docked DNA, is also shown in Figure 1S.

General Description of the Three DNA Complexes (before SA). Table 1 and Table 1S of the Supporting Information list the number of residues proximal to DNA (at a distance of < 6 Å) per dimer and the estimated binding energies, respectively.

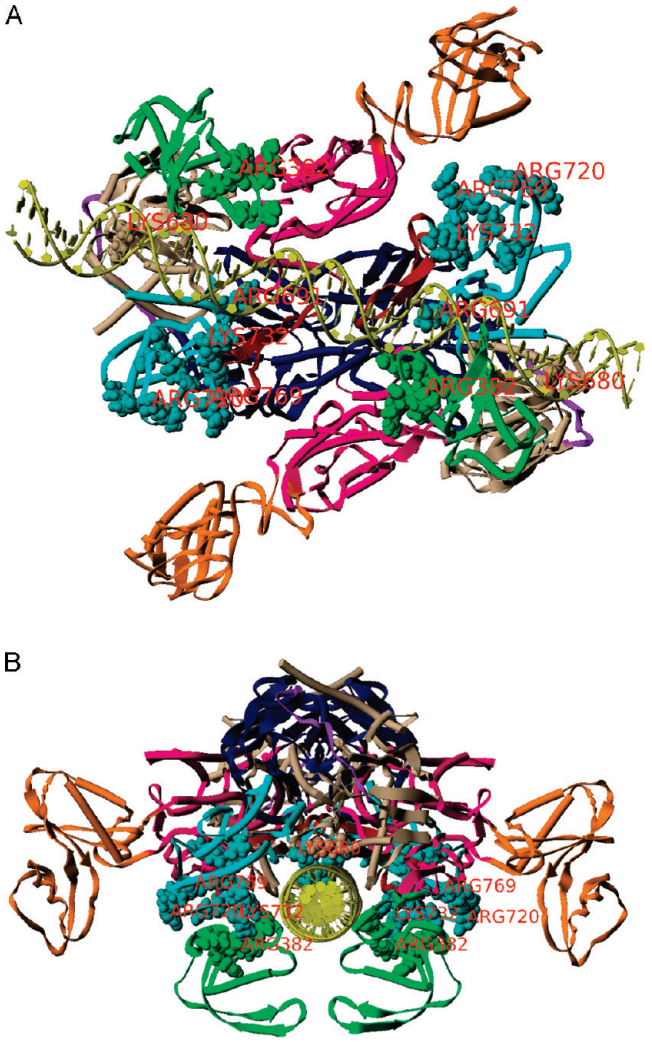


FIGURE 3: Best-scored DNA−$UvrA_2$ complex (C01) from docking along path 1 and subsequent energy minimization. The complex is presented in two views: on the ventral surface (A) and along the DNA axis (B). Domain colors as in Figure 1B. The protein is shown in ribbon−tube presentation and DNA in snake-symbol (yellow); residues from the three mutation sets are shown as space-filling models. To avid text overlap, only some are labeled. In C01, the $UvrA_2$ dimer covers ~32 bp of DNA (see also Figure 4S of the Supporting Information).

In C01, DNA spans the dimer from ATP-BD2 of chain A to that of chain B, passing between the IDs and partially overlying SGN2 basal portions. The side core residues of IDs and SGN2 domains, together with the tip surface residues of ATP-BD2 contribute to the two distinct binding regions (sites) in each protomer. In addition, DNA interacts with some residues from the Zn3 hairpins (SGN2), albeit more weakly. The total number of the residues found less than 6 Å from DNA is 162 (Table 1). A full list of residues in steric interaction with DNA (82 in total), i.e., with at least one backbone or side chain atom within < 3.5 Å of DNA, is given in Table 2S of the Supporting Information. In C01, DNA is in the proximity (< 6 Å) with nine of the mutated positively charged residues, with six of them in van der Waals (< 3.5 Å) contacts. At least one residue from each of the three mutation sets (see above) falls in the < 6 Å range. The only exception is set 1(B), with its Arg and Lys residues located at the tip of the SGN2 domain that remain relatively distant. Remarkably, DNA lies exactly along the line connecting Lys680 and Arg691 (set 2) of protomer A with the same residues of protomer B. All four of these residues are within van der Waals (steric) contacts (< 3.5 Å) with DNA.

Table 1: Number of $UvrA_2$ Residues within 6 Å of DNA before (C01, C02, and C03) and after Simulated Annealing (C1, C2, and C3) per Dimer Domain and Number of Residues That Belong to Any of the Three Mutation Sets (Mut. Res.; see the text)

domain	C01	C1	C02	C2	C03	C3
ATP-BD1	2	2	0	0	0	0
ATP-BD2	42	43	0	0	0	0
SGN1	6	23	21	26	0	6
SGN2	64	78	0	2	55	66
ID	48	57	10	23	0	0
BBD	0	0	1	38	21	46
total	162	203	32	89	76	118
Mut. Res.	9	13	4	8	12	12

In C02, DNA traverses the dimer in a wide shallow groove defined by the ID and SGN2 domain from the one protomer and the BBD from the other protomer. However, the underlying surface, partly comprised of SGN1, ATP-BD1, and SGN2 (Zn3 hairpin) residues, is uneven, which prevents deeper positioning of the rigid DNA helix, without perturbations of the backbone of these domains (Figures 1S and 4S of the Supporting Information). The protein residues close to the DNA are mainly from SGN1 and IDs [total of 32 (Table 1)], with only four mutation residues falling within the 6 Å interaction range, and none from the SGN2 set 1 (A and B) and set 2 mutation groups. There are only 13 residues in close van der Waals contacts with DNA (Table 2S of the Supporting Information). The estimated interaction energy is also modest, as compared to that of C01 (Table 1S of the Supporting Information).

In C03, DNA is positioned along the interface plane dividing the two protomers of the dimer. The total number of residues that surround DNA in the 6 Å radius cylinder is 76 (mainly from SGN2 and BBD). The number of residues in a van der Waals contact with DNA is 38, of which eight are mutated. As seen in Figures 1S and 4S of the Supporting Information, the binding site involves a vast majority of the SGN2 mutation set residues (total of 12 within the 6 Å range), but none from sets 2 and 3. Figure 1S of the Supporting Information (panels G and H) shows also that ID mutation sites (set 3) and Lys680 located in ATP-BD2 (set 2) are quite remote from DNA. The binding energy is the lowest among the three complexes (because of strong repulsion) (Table 1S of the Supporting Information). Several attempts have been made to improve the binding strength in this case, but the current interdomain configuration would allow deeper positioning only of bent DNA, or after interdomain rearrangement (displacement), especially of the BBDs.

DNA−Protein Complexes after the Simulated Annealing (SA). In line with the thermal mobility map presented in Figure 1A, one to three subsequent SA runs with the holo dimer (for a typical thermodynamic trajectory, see Figure 2S of the Supporting Information) confirmed that BBDs and IDs are the most prone to dynamic reorientations, and less so the SGN2 domains, whereas other domains remained quite rigid. The examination of the domain motion within the holo dimer shows that BBDs and IDs tend to move reversibly (not necessarily symmetrically in the two subunits) in directions toward (closing) and away from (opening) the neighboring SGN2 domains. This motion may essentially narrow or augment the gap between

Table 2: Comparison of $UvrA_2$−DNA Complexes (site 1) with Docked DNA(1) or DNA(2) before and after DNA-Restrained and DNA-Relaxed SA

	energy (kcal/mol)					
molecule[a]	*E*(total)	*E*(DNA)	*E*(protein)	*E*(interact)[b]	rmsd[c] (Å)	DNA bending (deg)
Holo-0	−23369.8	−	−23369.8	−	−	
C1/DNA(1)-0	−25618.8	−857.0	−23740.6	−1021.2	0.0	6 ± 2
C1/DNA(1)-A	−28525.9 ± 61.3	−856.3 ± 8.2	−26209.2 ± 58.6	−1460.4 ± 16.7	5.7 ± 0.4	6 ± 2
C1/DNA(1)-ARlx	−29893.4 ± 58.8	−626.0 ± 11.6	−26787.9 ± 44.5	−2479.5 ± 24.3	8.9 ± 0.7	20 ± 3
C1/DNA(1)-TRlx	−28914.8 ± 82.7	−338.2 ± 14.5	−26369.3 ± 59.4	−2207.3 ± 31.2	7.5 ± 0.9	24 ± 4
C1/DNA(2)-0	−28408.6	−1169.9	−25815.5	−1423.2	0.0	2 ± 1
C1/DNA(2)-A	−28774.7 ± 49.3	−1162.2 ± 6.0	−25764.1 ± 39.4	−1848.4 ± 12.8	5.2 ± 0.6	2 ± 1
C1/DNA(2)-ARlx	−29826.6 ± 61.6	−696.1 ± 12.4	−26385.9 ± 56.3	−2744.6 ± 23.3	7.9 ± 0.8	22 ± 6
C1/DNA(2)-TRlx	−28877.0 ± 89.6	−640.1 ± 14.6	−25802.8 ± 67.3	−2434.1 ± 26.8	7.2 ± 0.7	26 ± 7

[a]The docking default structure, Holo-0, is the initial dimer (Methods). C1/DNA(1)-0 and C1/DNA(2)-0 are the complexes before SA. C1/DNA(1)-A and C1/DNA(2)-A are molecules after one step of SA with DNA and ATP-BD1 restrained (aggregates). C1/DNA(1)-ARlx and C1/DNA(2)-ARlx are molecules subjected to subsequent (DNA-unrestrained) SA runs. C1/DNA(1)-TRlx and C1/DNA(2)-TRlx are molecules after one step of completely unrestrained (total relaxed) SA. C1/DNA(1)-0 and C1/DNA(1)-A are the same as C01 and C1, respectively (Table 1). [b]*E*(interact) = *E*(total) − [*E*(DNA) + *E*(protein)]. [c]The rmsd reference is Holo-0.

SGN2 and IDs from the same protein chain, or between SGN2 and BBDs from the opposite protomers, eventually either opening or closing two of the DNA presupposed sites (paths 1 and 3), while path 2 appears to be less affected by the interdomain reconfiguration in holo-$UvrA_2$ (Figure 3S of the Supporting Information). In contrast, SA with the DNA-loaded protein complexes showed mostly unidirectional domain motion toward the DNA, including the BBDs in C1 that tend to orient toward the central DNA part. The subsequent SA runs resulted in a tighter grasp on DNA (see below), instead of further pronounced interdomain reorientation. Therefore, for the sake of simplicity, only single-SA run treated DNA−protein complexes are presented, except in the case of complex 1. The structures of the DNA−$UvrA_2$ complexes after SA and energy minimization, named C1−C3, are shown in the Supporting Information (superimposed on the staring complexes, Figure 4S). In all three cases, the flexible domains (IDs and BBDs) adopt more closed interchain configurations. The DNA binding energies improve significantly after SA, as did the protein potential energies (Table 1S of the Supporting Information). However, the DNA potential energy increased in some of the complexes because of intra-DNA strains (helix distortion) and close unfavorable interactions with the environment, including a few penetrations of protein side chains into the DNA duplex (in both the major and minor grooves). To characterize the interdomain displacement after SA runs, we measured (i) translational domain motions, i.e., rmsd (backbone atoms only) as compared with their starting configurations (before SA), and (ii) rotational domain motion defined as interdomain backbone plane angle changes after SA. An example of domain motion is provided in Figure 4, which shows the absolute displacement of the center of mass of the BBDs from chains A and B of complex 3. In particular, the graph shows that the near equilibrium is reached at the early SA stages of the annealing phase, and that the BBDs from the opposite protomers move similarly, although not strictly equally. A graphical example of global domain translational and rotational motion is presented in Figure 5S of the Supporting Information. It shows the superimposed structures of the starting C02 complex with that obtained at the end of the SA run and after minimization, C2. Tables 3S and 4S of the Supporting Information list data of domain displacement (rmsd) and intra- and interprotomer domain rotations (plane angles before and after SA). Graphical excerpts from Table 4S of the Supporting Information are presented in Figure 5

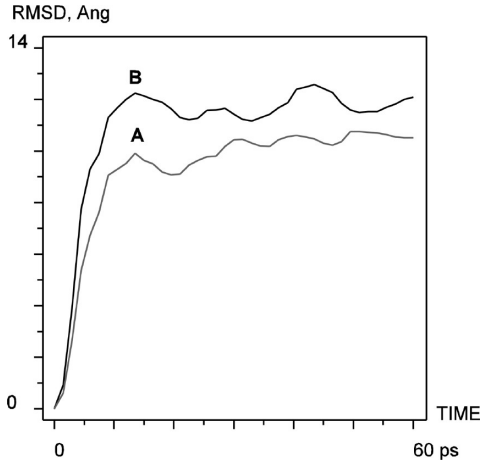


FIGURE 4: Trajectory of the center of mass (rmsd, in angstroms) of chains A and B of the C3 BBDs during a single-step SA. The beginning coincides with the BBD positions in C03 (and Holo-0); the trajectory ends are the BBD coordinates as displaced during the formation of C3 (before energy refinement).

(backbone rmsd) and Figures 6S of the Supporting Information (A and B, inter- and intrachain backbone rotations, respectively). The comparison of domain mobility in the same complex and between different DNA−protein complexes underlines the fact that (i) IDs and BBDs tend to reorient to a higher degree, as compared to other domains (ATP-BD and SGN), (ii) depending on the DNA position within the complex, domains move to ensure tighter DNA binding, and (iii) an example of the holoprotein subjected to one-step (Holo-1) or two-step (Holo-2) SA shows that the overall translational reorientation (Figure 5) is similar after subsequent SA runs.

Detailed Examination of DNA Binding in Complex 1: Proposed Damage Recognition Site. The characteristics of complex 1, such as the overall topology (well-defined, narrow, and positively charged DNA-binding groove), the high DNA−protein interaction energy, and the largest number of amino acids and mutated residues interacting with DNA (Tables 1 and 2 and Table 1S of the Supporting Information), substantiate that in this case DNA likely occupies a naturally preformed, high-affinity binding site. This is also evident from the data in Tables 3S and 4S of the Supporting Information, Figure 5, and Figure 6S of the Supporting Information showing that the IDs in C1 do not need

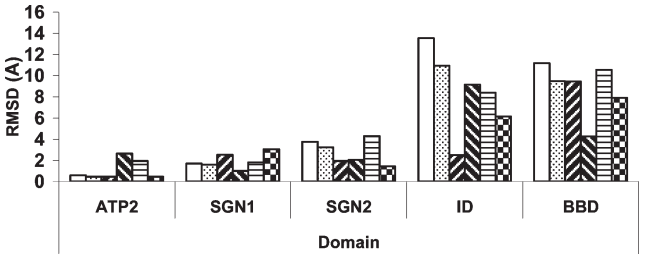


FIGURE 5: Comparative translational domain motion (backbone rmsd) after SA. The starting coordinates are of Holo-0. Holo-1 (blank) and Holo-2 (dotted) denote the holo-$UvrA_2$ submitted to one and two step SA runs, respectively (Figure 2S of the Supporting Information). C1 (left hatch), C2 (horizontal hatch), and C3 (checkered) are the complexes derived from C01, C02, and C03, respectively (after SA and energy refinement). C1T, also labeled as C1/DNA(1)-ARlx (right hatch), is C1 submitted to a second (totally unrestricted) SA run. Averaged data over A and B subunits (from Table 3S of the Supporting Information).

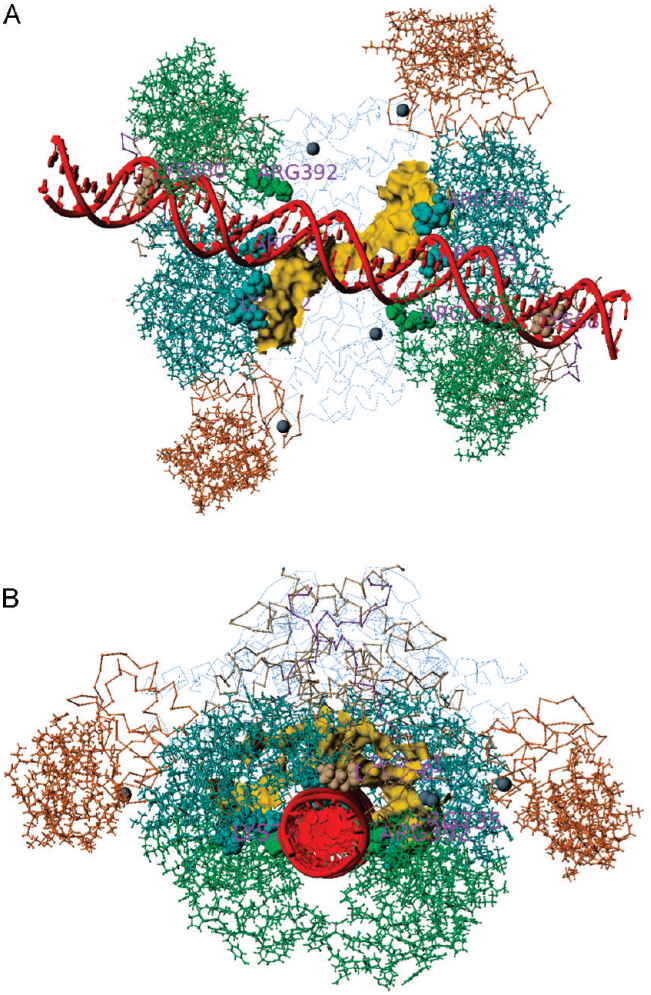


FIGURE 6: (A) C1 DNA binding site (SGN2, ID, and ATP-BD2) and proximity of the Zn3 HP (yellow surface) to the central DNA part. (B) Orthogonal view of panel A. Domain colors, as in Figure 1B. Gray spheres denote Zn atoms in sites Zn1 and Zn2; ATP-BD1 and SGN1 are in the background (Cα trace, blue). Several mutation-identified DNA-binding residues (representative from sets 1−3; see the text) are shown in space-filling mode and labeled.

to undergo excessive translational motion in the presence of DNA. In C1, DNA is tightly and symmetrically side-clamped between the ID and SGN2 domain from the two subunits (Figure 6). The third (basal) component that firmly supports the DNA clamping by the ID and the SGN2 domain is provided by the tip residues of ATP-BD2. At the same time, the central DNA part remains partially exposed to the solvent and is much less sterically hindered (Figures 6A and 7). The molecular surface models shown in Figure 7 and Figure 7S of the Supporting Information demonstrate the nearly perfect 3D shape fit of DNA in the $UvrA_2$ dimer. Because of the chosen electrostatic force field parameters (and the absence of solvent and counterions), the negative electrostatic potential created by DNA expands over the protein surface (compare with Figure 2). The number of basic residues in the 6 Å radius cylinder around DNA is large, including the basic residues, known to influence DNA binding [Figure 8, Figures 8S and 9S of the Supporting Information (panels A and B), Table 1, and Table 1S of the Supporting Information]. Notably, some of the positively charged protein side chains are located in the duplex groves (Figure 8 and Figures 7S and 8S of the Supporting Information). There are also several close to DNA acidic residues, but in the presence of counterions and/or water, these are probably neutralized or H-bond-bridged to the DNA backbone. The next interesting phenomenon in C1 is that the Zn3 HPs residues (Figure 6) from the two protein chains more closely interact with DNA (13 in total, including the basic ones, His740, Lys746, and His750). The close interaction with DNA (< 3.5 Å) also involves the mutation-determined important residues (from set 1A) Lys732 and Arg735 located near the base of the Zn3 hairpin (Figure 8 and Table 2S of the Supporting Information).

The popular hypothesis concerning the DNA structure change upon binding by $UvrA_2$ is that the bound DNA is bent and/or unwound to facilitate the damage proofreading by the coupled UvrB protein. This prompted us to examine the properties of complex 1 in more detail. In the models of $UvrA_2$−DNA complexes described above, the DNA(1) was treated as a rigid body ligand. To investigate possible deformations of DNA(1) structure upon binding to UvrA2, we performed additional simulation experiments based on the C1 complex: (i) submitting the already SA-treated C1 nucleoprotein to a subsequent SA run (after minimization), where the whole complex, DNA included, was unrestrained; (ii) once the DNA position had been determined in C1, performingd independent split DNA docking experiments at the same site (Methods), and (iii) C01 complex being submitted to a completely unrestrained SA. To elucidate possible DNA sequence-dependent effects, we docked the CG-enriched DNA(2) 41 bp duplex in the same site (site 1) and repeated the SA experiments as described above.

If C1 was submitted to a second SA run with unrestrained DNA(1) [the new complex named C/DNA(1)-ARlx, or C1T], IDs additionally moved to wrap around DNA even tighter (Figures 5, 9, and 10), while DNA partially melted and moved closer "down" to the basal UvrA2 surface (SGN1, ATP-BD1, and the Zn3 hairpins). DNA(1), which was initially bent to ~6°, is now bent (in the same direction) to ~20° and exhibits a kink facing the Zn3 hairpins, unwound, and many bases are flipped out of the helix (Figure 9C and Figure 10S of the Supporting Information). The number of close contact residues from the ATP-BD2 domains also increased [Table 2S and Figures 9S (panels C and D) of the Supporting Information], while the two BBDs rotated (Figure 5 and Figure 6S of the Supporting Information) to point toward the center of the DNA−protein complex but do not interact with DNA (Figure 9). The results clearly indicate improved DNA−protein interaction energy, apparently because of the increased number interacting with DNA protein residues (Table 2). These comprise several additional

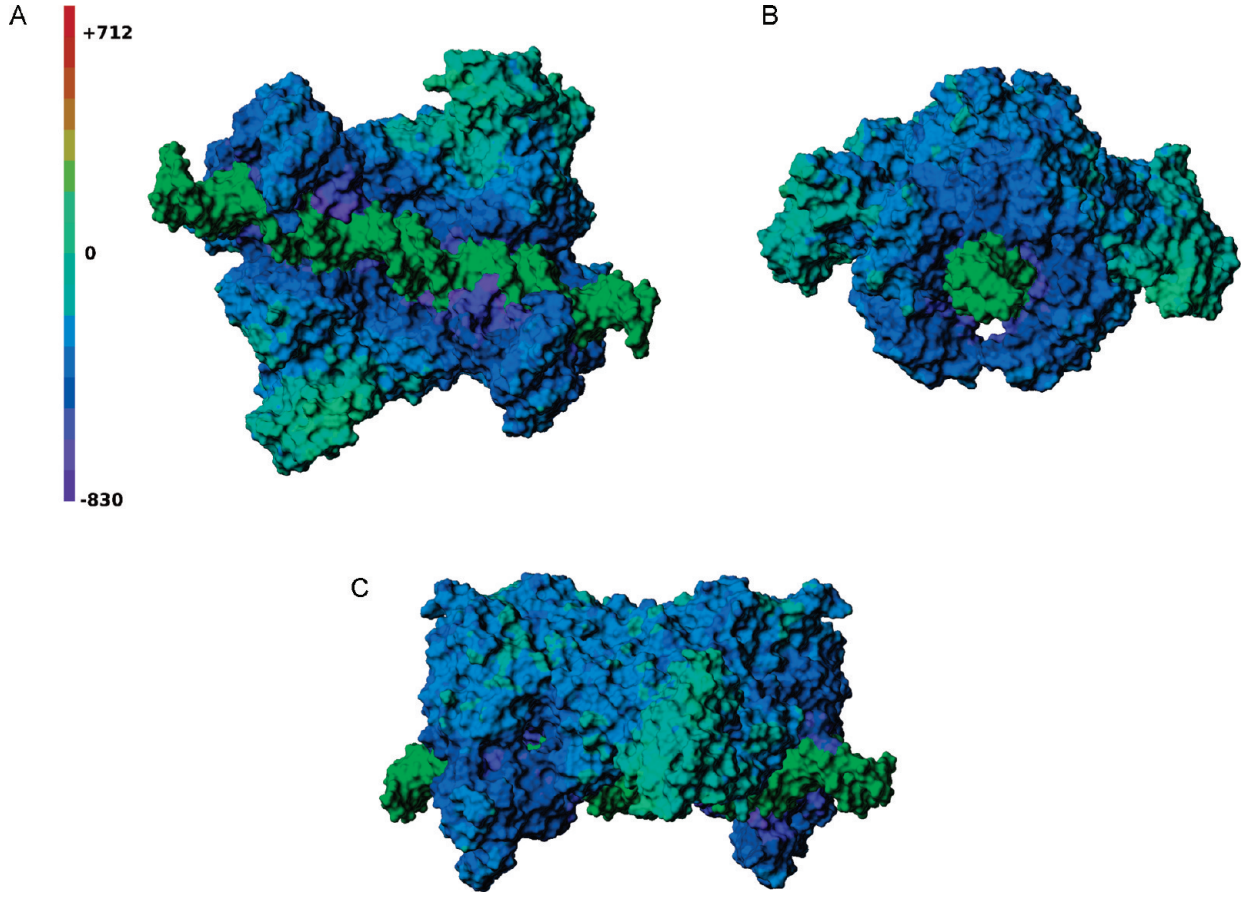


FIGURE 7: Molecular surface of the C1 complex. The protein is colored by EP, and the DNA surface is uniform (green). Views are of the dorsal site (A), along the DNA axes (B), and after a 90° rotation of the latter (C). DNA is strongly clamped mainly because of IDs wrapping around it; the middle DNA portion is accessible to solvent from the dorsal side. The rotation of BBDs renders these domains aligned in the perpendicular to the ventral surface diagonal plane that bisects the DNA axis at the center. The expansion of the negative potential around DNA is obvious (compare with Figure 2; the color ramps are identical). Additional C1 molecular surface models are shown in Figure 7S of the Supporting Information.

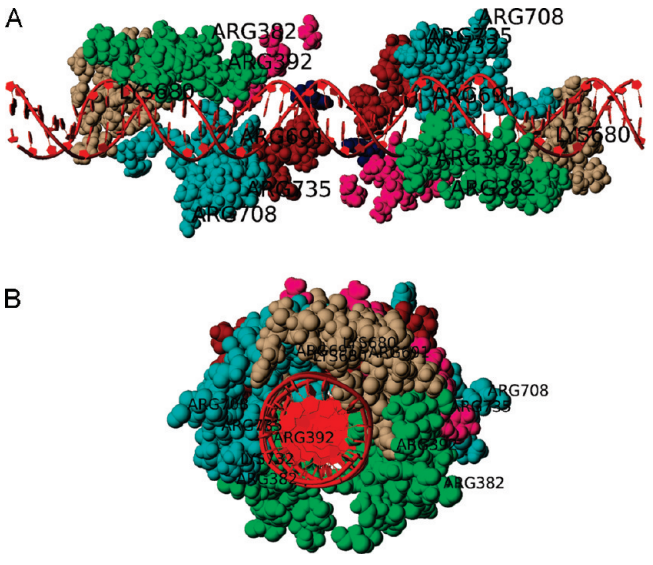


FIGURE 8: C1 protein residues engaged in the 6 Å DNA binding cleft (space filling, domain color as in Figure 1B). Views (A) of the dimer ventral side and (B) along the DNA axis. Several mutation-identified residues are labeled. Shown is the domain composition of the two binding sites (clamps) with the participation of ATP-BD2, ID, and SGN2. A few side chains penetrate the DNA grooves, and the Zn3 hairpins (brown) are within interaction range of the less restricted central DNA segment. The acid−base residue composition of the 6 Å DNA-binding site is presented in Figure 8S of the Supporting Information.

basic residues falling within the 6 Å cylinder around the DNA. Interestingly, 20 residues from the Zn3 hairpins (from chains A and B) now are in a van der Waals contact (< 3.5 Å) with DNA (from 13 in C1). However, the overall Zn3 hairpin conformation remains practically unchanged. The grip on DNA is supported by an increased number of residues closely interacting with DNA from the ATP-BD2 (29 in total, from 24 in C1). Apart from Lys680 (from the mutation set 2), these include basic residues Lys658, His660, Arg661, Lys663, His667, and Arg847, all falling within the 3.5 Å van der Waals interaction range with DNA (for the full list of van der Waals interacting residues, see Table 2S of the Supporting Information). The C/DNA(1)-ARlx complex also presents a large number of residues in the 6 Å range from DNA belonging to the most distant domain, ATP-BD1 (15, but no basic ones among them); in the C/DNA(1)-A complex, ATP-BD1 presents only one residue (Tyr60) per chain that falls in this range. The interaction of DNA and the ventral side domains with the underlying structures is graphically demonstrated by the holes appearing in the back Connolly surfaces of C1/DNA(1)-ARlx components shown in Figure 10B. In general, the binding site(s) and the entire complex are more compact, but as expected, the distortion of the DNA helix is coupled with a significant increase in the DNA potential energy at the expense of increased binding energy (Table 2). Apart from the "melting" of the DNA ends, which should be considered highly artificial and the twisted central DNA part (facing Zn3 hairpins), the helix is also distorted at the sites positioned directly within the clamps formed by SGN2

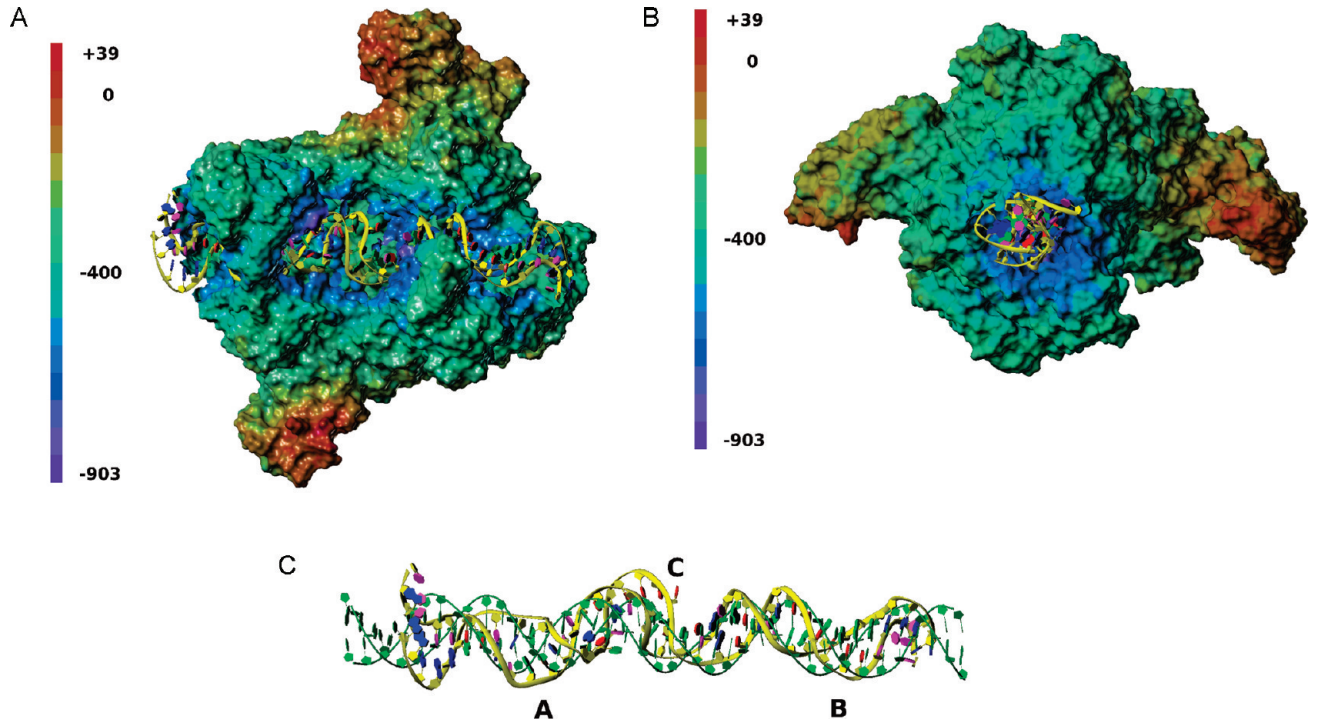


FIGURE 9: Surface model of the C1/(DNA(1)-ARlx (also labeled, C1T) complex obtained from C1/DNA(1) (C1) after a second round of SA with unrestricted DNA. Viewed (A) on the ventral side and (B) along the DNA axis. The DNA surface is removed and DNA shown as a ribbon symbol. The ventral surface with the exception of BBDs is all negatively charged. The whole structure is more compact than that of C1 (Figure 7), and the DNA is bent in the middle. (C) Superposition of the original (straight, green) and bent (yellow) DNA extracted from complexes C1 and C1/DNA(1)-ARlx (C1T), respectively. Letters A and B in panel indicate the positions of groove opening due to protein side chain penetration at the clamping sites; the letter C in panel C indicates the kink in the middle of the DNA facing the Zn3 hairpin tips. The two ends of C1/DNA(1)-ARlx DNA are also melted, which might be artificial (finite DNA length).

and IDs (Figure 10 and Figure 10S of the Supporting Information). The DNA bending angle was estimated to be 20 ± 3°. The results obtained by split docking experiments with three cutoff (left, right, and center) and subsequently resealed DNA segments were similar to those obtained via SA and confirmed that the site occupation by twisted and slightly unwound DNA is beneficial, although from the split docking method the DNA emerged twice less bent {~9° [Figure 10S of the Supporting Information (panel B)]}. Submitting C/DNA(1)-0 to a completely unrestricted SA run resulted in a similar structure (C/DNA(1)-TRlx, the totally relaxed complex), albeit DNA(1) in the latter case exhibited a larger bending angle [24° (Table 2)]. Although the overall backbone rmsd between C1/DNA(1)-ARlx and C1/DNA(1)-TRlx is ~3.5 Å, the number of residues falling in the 3.5 or 6.0 Å distance range from DNA(1) (including the mutated ones) is essentially the same. Similar conformational rearrangements were observed if the SA procedures described above were repeated with DNA(2) bound to UvrA2 at the same site 1. The data summarized in Table 2 indicate similar trends in energy and conformational changes in C1/DNA(1) and C1/DNA(2) complexes when they were treated under identical conditions. The suggestion that the protein (and DNA) moved to ensure the formation of similar complexes with DNA(1) and DNA(2) is supported by the identical number of residues falling in the examined interaction ranges [e.g., 263 and 264 in the 6 Å range for C1/DNA(1)-ARlx and C1/DNA(2)-ARlx, respectively]. Importantly, unrestrained SA applied to the C1/DNA(2) complex also led to DNA bending and formation of a kink located between the two clumping sites and above the Zn3 hairpin (Figure 10C). From the examination of both complexes, it appears that a distinctive structural motive in the insertion domains (ID) is specially designed to tackle major and minor DNA grooves simultaneously (at the clumping sites). This three-finger (tripod) structural motive (well seen in Figure 10C, right bottom, A/ID) encloses the loop immediately following the α10 helix, the loop connecting sheets β12 and β13, and a third component, built of helix α14 and the loop preceding it. However, only helix α14 contains DNA interacting basic residues (Arg392 and Lys397).

DISCUSSION

The structure-based determinants by which the simplest (prokaryotic) NER molecular machinery recognizes damaged DNA remain largely enigmatic. The recent release of the crystallographic structures of bstUvrA$_2$ (*15*) and drUvrA$_2$ (*16*) dimers opens new avenues to studying DNA damage recognition, including 3D molecular modeling. Because the reported dimers have been crystallized without DNA, the first step in the modeling approach is to attempt to identify the (undamaged) DNA-binding sites in the protein and once the structure of the nucleoprotein (the DNA−UvrA$_2$ complex) is predicted to continue with detailed investigation of the damage recognition process. Here we performed a first approximation study of the possible modes of interactions between undamaged double-stranded DNA and the bstUvrA$_2$ dimer. Applying SA, we also tackled the problem of protein domain rearrangements in the presence or absence of DNA to ensure increased binding strength and specificity. The examination of the ventral side of the bstUvrA$_2$ dimer [relief/curvature, domain organization and mobility, and the abundance of basic residues, including the three sets of mutation-determined residues influencing DNA binding and located mainly in the SGN2 domain and IDs (*15*, *16*, *19*)] led us to the conclusion that there might be at least three possible and significantly different

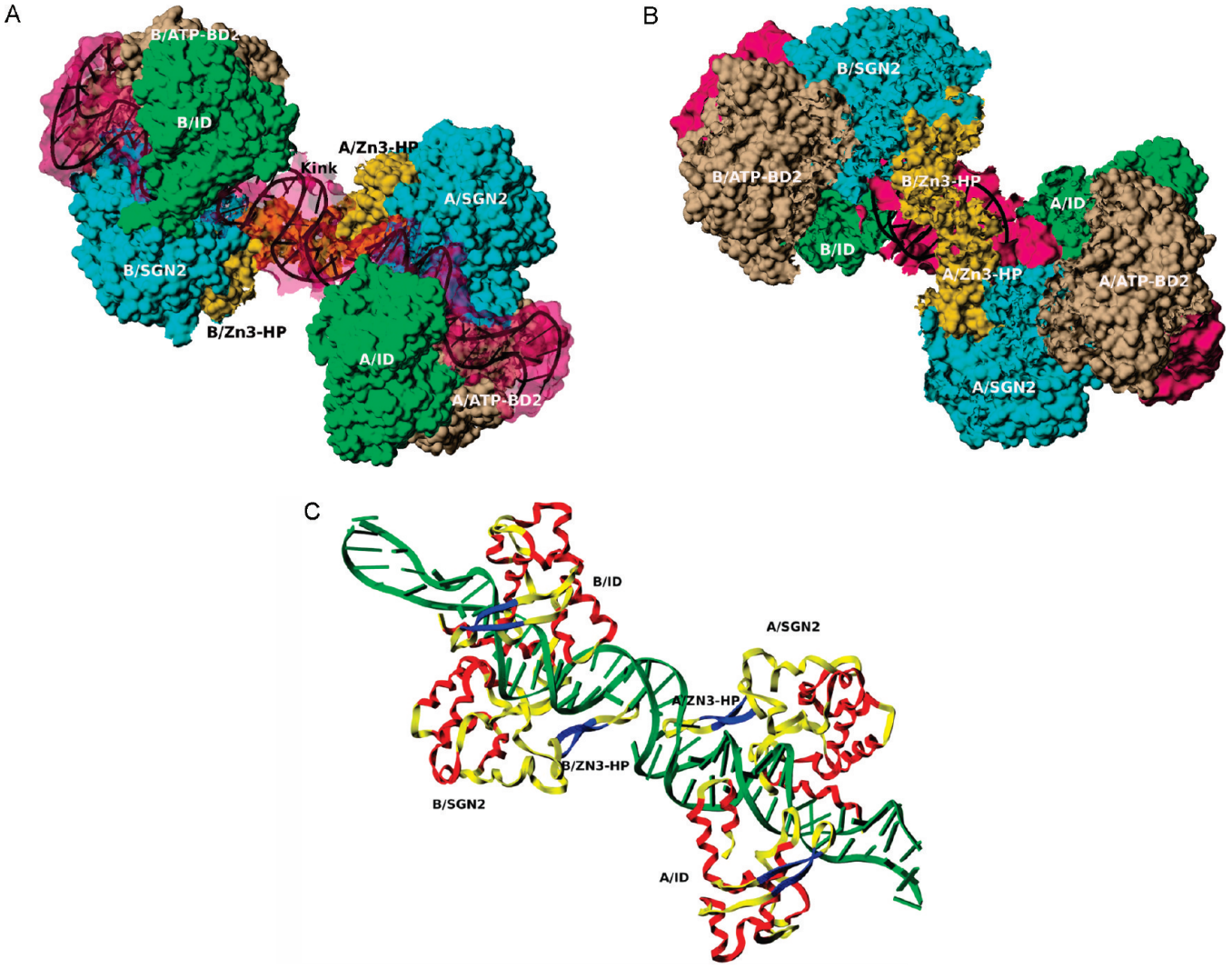


FIGURE 10: Zoomed surface model of the C1/DNA(1)-ARlx DNA binding site. (A) Front view on the ventral site. (B) Back view from the dorsal side. The left clamp (B chain) is completely closed by ID, while the right clamp (A chain) is semiclosed. Surfaces: DNA (pink transparent, in panel A), ID (green), Zn3 hairpins (yellow), SGN2 (cyan), and ATP-BD2 (beige). The DNA central part (with the kink in the middle) overlays and interacts with the Zn3 hairpins. DNA is shown as a black ladder. The DNA surface in panel B is opaque to help visualize the large surface hole above the Zn3 hairpins. All surfaces (on the dorsal side) have perforations indicating close van der Waals interactions (side chain penetrations) with the hidden (background) $UvrA_2$ domains (ATP-BD1 and SGN1). (C) Ribbon presentation of the DNA clumps (ID and SGN2) as observed in the C1/DNA(2)-TRlx complex colored by secondary structure (red for helices, blue for sheets, and yellow for loops). DNA(2) ribbon (green) with a central kink and melted ends. The DNA clumping three-finger (tripod) structural motive is well seen in A/ID (bottom right).

(noncollinear) initial DNA binding paths (Scheme 1). As expected (*22*), the dimer in all three studied DNA orientations presents two separate DNA-binding sites. In the three passages, DNA lies between the IDs, as previously suggested (*15*, *16*), but the DNA binding engages different combinations among other ventral surface-exposed protomer components, the SGN1, SGN2, ATP-BD2, and/or BBDs in a grossly symmetrical manner. Following docking, the best-scored DNA−protein complexes were energy-minimized and subjected to SA to further investigate the protein dynamics upon DNA binding. The applied continuum MD FF protocol was similar to that of Gantchev et al. (*18*). Because of system heterogeneity as a precaution to avoid artifacts, e.g., irreversible DNA melting during the thermal jump, we have chosen to keep DNA restrained (rigid) in the initial simulated annealing experiments. This simplified approach (apart from being CPU cost-effective) allowed us to examine systematically only the protein behavior when DNA was docked in any of the three originally proposed (1−3) binding paths and helped in the selection of the high-affinity site, in a manner independent of the DNA local structure. At later stages (with complex 1), we addressed the problem of DNA structure deviations upon $UvrA_2$ binding by performing DNA-unrestrained SA runs and with DNA of different sequence.

The obtained results indicate that paths 2 and 3 (the corresponding DNA−$UvrA_2$ complexes 2 and 3, respectively) are less consistent with the experimental data (proximity of point mutations involved in DNA binding and/or damage recognition) than path 1 ($UvrA_2$−DNA complex 1). However, because the mutated residues are positively charged (unspecific), it is unknown if all of them are involved in direct interactions with DNA or have a supportive (domain stabilization) structural role. The later follows from the fact that mutation sets 1 and 3 exhibit their strong DNA binding inhibition only in combination (*15*, *16*), while the two mutated residues from set 2 (*19*), which belong to different domains, are effective as single-residue mutations, as well. At the same time, it is obvious that in C2 and C3 large conformational rearrangements would be needed to engage ATP-BD2 in DNA binding and/or damage recognition. Complexes 2

and 3 also show modest to low DNA binding energies, although better DNA binding characteristics were achieved after SA. Apart from the clearly higher DNA binding energy of complex 1 (Table 1S of the Supporting Information) and much larger electrostatic DNA binding contributions from the residues that belong to the three mutational sets (Table 1, Figure 8, and Figures 8S and 9S of the Supporting Information), C1 also presents two prearranged high-affinity DNA-binding sites with symmetric contributions from protomers A and B (Figures 5−7 and Tables 1S−3S of the Supporting Information). As in the other complexes, the DNA binding parameters following simulated annealing were additionally improved, mainly because of DNA wrapping by the IDs. In addition, the C1 nucleoprotein binds ~32 bp DNA, corresponding to known experimental data (*19*). These observations led us to the conclusion that the DNA position in the C1 nucleoprotein represents the natural DNA damage recognition site. Further investigations of the same complex by allowing free motion of DNA showed that DNA binding at this site favors bent and unwound DNA [C1/DNA(1,2)-ARlx and C1/DNA(1,2)-TRlx complexes], as previously hypothesized (*21*). DNA bending also promotes base flipping, a phenomenon already known in the damage recognition by the UvrB protein (*14*, *23*). This disordered DNA topology massively exposes the middle part of the duplex to a large number of Zn3 hairpin residues within the van der Waals interaction range [< 3.5 Å (Table 2S of the Supporting Information and Figure 10B)]. Thus, it is very likely that together with the specific DNA helix deformations, the Zn3 structural motives in $UvrA_2$ play role in DNA lesion recognition. However, the elucidation of the structural dynamics of this process requires special investigation. Our C1 model is consistent with the previous findings that Zn3 hairpins are dispensable in the undamaged DNA binding, but their deletion leads to a complete loss of the ability of the $UvrA_2$ dimer to distinguish between undamaged and damaged DNA (*19*, *24*, *25*). However, the previously published UvrA−DNA binding site model, built before the release of the $UnrA_2$ crystal structure, wrongly depicts the Zn3 hairpins pointing outward from the DNA (*19*).

All constructed nucleoprotein models (particularly, C1) provide an insight into why IDs and Zn3 hairpin modules are dispensable in undamaged DNA binding (*16*, *19*). In C1, apart from IDs, the DNA is firmly bound by SGN2 and ATP-BD2 and several basic residue side chains are located in both the major and minor DNA grooves. The middle DNA portion approximately one helical turn in length, facing the Zn3 hairpins, is less restricted. This configuration provides an example of molecular machinery designed to bend DNA clamped at its two ends. However, it appears that the damage recognition might not depend on a single DNA property only, i.e., flexibility differences, but also may depend on other factors. For example, in types of damage such as intrahelical DNA adducts and cross-links (*4*, *26*, *27*), the altered intrinsic DNA flexibility (e.g., resistance to bending and/or unwinding) may signal nucleoprotein protein conformational rearrangements and the recruitment of UvrB, whereas in the presence of bulky covalent adducts (e.g., FldT), the putative interactions between the DNA extrahelical group and the protein (probably involving the Zn3 hairpin) can restrict the DNA motion to trigger the next damage recognition step. The simulations were performed with undamaged AT- and CG-enriched DNAs. In both cases, MD-unrestricted DNA bends to a similar extent of ~20−26°. On the other hand, $UvrA_2$ is known to have a relatively low affinity for undamaged DNA (*4*). Because our models relate to the forward (association) reaction of $UvrA_2$ and DNA, it follows that the dissociation of undamaged DNA should be fast (compared to the damaged DNA). It can be further hypothesized that apart from possible ATPase-dependent protein rearrangements, the fast undamaged DNA dissociation is thermodynamically coupled with the bound (bent) DNA free energy balance.

Our analysis indicates that the entire, positively charged $UvrA_2$ ventral surface provides a platform for nonspecific undamaged DNA binding, and the specificity and affinity are further modulated by the interdomain arrangement (surface relief), e.g., the presence of groovelike DNA-binding elements, coupled with small-scale reorientation that is best exemplified in the case of C1. The availability of a family of DNA binding sites in $UvrA_2$, albeit of different affinity, is interesting and may indicate possible multiple DNA processing steps by $UvrA_2$, e.g., initial weak DNA binding followed by complex readjustment (DNA and/or protein rotation) to probe for DNA damage, e.g., DNA handoff to site 1. The apparent existence of weaker binding sites, such as paths 2 and 3 might be not only relevant to the protein function but also particularly important in special cases. For example, because in path 3, DNA traverses the dimer along the plane dividing the two protomers, hypothetically complex 3 can be assembled after initial DNA binding by a monomeric UvrA with a subsequent recruitment of the second monomer. Another property of the $UvrA_2$ dimer is also quite remarkable; in paths 1 and 2, DNA strands make an angle of ~90° with each other, crossing at the middle of the ventral surface (over the Zn3 hairpins). Such combined DNA configuration resembles that of the four-way junction DNA. Therefore, it can be hypothesized that the $UvrA_2$ dimer is adapted to bind DNA of unusual topology, e.g., crucifix DNA, similar to the human XPA damage recognition protein (*26*). To add to the puzzle of how $UvrA_2$ recognizes such a large variety of bulky DNA damage, Weng et al. (*4*) recently proposed that two $UvrA_2$ dimers bind simultaneously to each strand of an ICL dG-mitomycin-C-dG adducted DNA. We were intrigued by this possibility and attempted to construct a quadruple $[UvrA_2]_2$−DNA complex. It appears that the formation of such a "heretic" complex is not impossible, especially when the NER system encounters excessively balky lesions, as the mitomycin-C ICL (see Figure 11S of the Supporting Information). Finally, our SA results show that BBDs in complex 1 tend to reorient and project toward the center of the nucleoprotein (Zn3 hairpins). This reorientation does not align BBD−UvrB segments [from PDB entry 3fpn (*17*)] with the DNA ends, as shown in ref *28*, which means that a handoff of UvrB to the UvrA−DNA central part is more likely than the tandem recruitment model. One reason for this difference might be the fact that in the later work, high-molecular weight quantum-dot antibody-labeled UvrA and UvrB proteins were used, which by itself may strain the motion of the BBDs in the UvrAB complex. Although further studies are required to understand the damaged DNA recognition by UvrA proteins, we believe that the presented models (especially C1) will be useful in planning rational, structure-based mutation studies.

SUPPORTING INFORMATION AVAILABLE

The additional information is compiled in three separate files: Tables-S.pdf, Charts-S.pdf, and Molecular Graphics-S.pdf. Tables-S contains Table 1S (which lists protein, DNA, and interaction energies for C1−C3), Table 2S (A−C) [which lists residues in van der Waals contact (< 3.5 Å) with DNA in the C1−, C2−, and

C3−DNA(1) complexes before and after SA], Table 3S [which lists data of translational domain displacement (mean rmsd) at the end of the SA run], and Table 4S (A and B) [which presents domain (inter- and intrachain) plane angles and their change after SA]. Charts-S contains Figure 2S (thermodynamic trajectory for holo-$UvrA_2$ subjected to two-step SA) and Figure 6S (comparative domain rotations after SA). Molecular Graphics-S contains figures presenting various conformational features of the studied $UvrA_2$−DNA complexes and the possible quadruple structure, $[UvrA_2]_2$. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Ramaswamy, M., and Yeung, A. T. (1994) Sequence-specific interactions of UvrABC endonuclease with psoralen interstrand crosslinks. *J. Biol. Chem. 269*, 485–492.
2. Barre, F.-X., Asseline, Y., and Harel-Bellan, A. (1999) Asymmetric recognition of psoralen interstrand crosslinks by the nucleotide excision repair and the error-prone repair pathways. *J. Mol. Biol. 286*, 1379–1387.
3. Imoto, S., Bransfield, L. A., Croteau, D. L., Van Houten, B., and Greenberg, M. M. (2008) DNA tandem lesion repair by strand displacement synthesis and nucleotide excision repair. *Biochemistry 47*, 4306–4316.
4. Weng, M.-W., Zheng, Y., Jasti, V. P., Champeil, E., Tomasz, M., Wang, Y., Basu, A. K., and Tang, M.-S. (2010) Repair of mitomycin C mono- and interstrand cross-linked DNA adducts by UvrABC: A new model. *Nucleic Acids Res. 38*, 6976–6984.
5. Lindahl, T., Karran, P., and Wood, R. D. (1997) DNA excision repair pathways. *Curr. Opin. Genet. Dev. 7*, 158–169.
6. Hanada, K., Budzowska, M., Modesti, M., Maas, A., Wyman, C., Essers, J., and Kanaar, R. (2006) The structure-specific endonuclease Mus81-Eme1 promotes conversion of interstrand DNA crosslinks into double-strand breaks. *EMBO J. 25*, 4921–4932.
7. Husain, I., Van Houten, B., Thomas, D. C., and Sancar, A. (1986) Sequences of *Escherichia coli uvrA* gene and protein reveal two potential ATP binding sites. *J. Biol. Chem. 261*, 4895–4901.
8. Grossman, L., and Thiagalingam, S. (1993) Nucleotide excision repair, a tracking mechanism in search of damage. *J. Biol. Chem. 268*, 16871–16874.
9. Truglio, J. J., Croteau, D. L., Van Houten, B., and Kisker, C. (2006) Prokaryotic nucleotide excision repair: The UvrABC system. *Chem. Rev. 106*, 233–252.
10. Nakagawa, N., Sugahara, M., Masui, R., Kato, R., Fukuyama, K., and Kuramitsu, S. (1999) Crystal structure of *Thermus thermophilus* HB8 UvrB protein, a key enzyme of nucleotide excision repair. *J. Biochem. 126*, 986–990.
11. Theis, K., Chen, P. J., Skorvaga, M., Van Houten, B., and Kisker, C. (1999) Crystal structure of UvrB, a DNA helicase adapted for nucleotide excision repair. *EMBO J. 18*, 6899–6907.
12. Machius, M., Henry, L., Palnitkar, M., and Deisenhofer, J. (1999) Crystal structure of the DNA nucleotide excision repair enzyme UvrB from *Thermus thermophilus*. *Proc. Natl. Acad. Sci. U.S.A. 96*, 11717–11722.
13. Waters, T. R., Eryilmaz, J., Geddes, S., and Barrett, T. E. (2006) Damage detection by the UvrABC pathway: Crystal structure of UvrB bound to fluorescein-adducted DNA. *FEBS Lett. 580*, 6423–6427.
14. Truglio, J. R., Karakas, E., Rhau, B., Wang, H., DellaVecchia, M. J., Van Houten, B., and Kisker, C. (2006) Structural basis for DNA recognition and processing by UvrB. *Nat. Struct. Mol. Biol. 13*, 360–364.
15. Pakotiprapha, D., Inuzuka, Y., Bowman, B. R., Moolenaar, G. F., Goosen, N., Jeruzalmi, D., and Verdine, G. L. (2008) Crystal structure of *Bacillus stearothermophilus* UvrA provides insight into ATP-modulated dimerization, UvrB interaction, and DNA binding. *Mol. Cell 29*, 122–133.
16. Timmins, J., Gordon, E., Caria, S., Leonard, G., Acajjaoui, S., Kuo, M.-S., Monchois, V., and McSweney, S. (2009) Structural and mutational analysis of *Deinococcus radiodurans* UvrA2 provide insight into DNA binding and damage recognition by UvrAs. *Structure 17*, 547–558.
17. Pakatoprapha, D., Liu, Y., Verdine, G. L., and Jeruzalmi, D. (2009) A structural model for the damage-sensing complex in bacterial nucleotide excision repair. *J. Biol. Chem. 284*, 12837–12844.
18. Gantchev, T. G., Cecchini, S., and Hunting, D. J. (2005) Dynamic conformational states of DNA containing T·T or BrdU·T mispaired bases: Wobble H-bond pairing versus cross-strand inter-atomic contacts. *J. Mol. Model. 11*, 141–159.
19. Croteau, D. L., DellaVecchia, M. J., Perera, L., and Van Houten, B. (2008) Cooperative damage recognition by UvrA and UvrB: Identification of UvrA residues that mediate DNA binding. *DNA Repair 7*, 392–404.
20. Gantchev, T. G., Girouard, S., Dodd, D. W., Wojciechowski, F., Hudson, R. H. E., and Hunting, D. J. (2009) γ-Radiation induced interstrand cross-links in PNA-DNA heteroduplexes. *Biochemistry 48*, 7032–7044.
21. Wang, H., Lu, M., Tang, M.-S., Van Houten, B., Ross, J. B., Weinfeld, M., and Le, X. C. (2009) DNA wrapping is required for DNA damage recognition in the *Escherichia coli* DNA nucleotide excision repair pathway. *Proc. Natl. Acad. Sci. U.S.A. 106*, 12849–12854.
22. Wagner, K., Moolenaar, G., van Noort, J., and Goosen, N. (2009) Single-molecule analysis reveals two separate DNA-binding domains in the *Escherichia coli* UvrA dimer. *Nucleic Acids Res. 37*, 1962–1972.
23. Malta, E., Verhagen, C. P., Moolenaar, G. F., Filippov, D. V., van der Marel, G. A., and Goosen, N. (2008) Functions of base flipping in *E. coli* nucleotide excision repair. *DNA Repair 7*, 1647–1658.
24. Wang, J., Mueller, K. L., and Grossman, L. (1994) A mutational study of the C-terminal zinc-finger motif of the *Escherichia coli* UvrA protein. *J. Biol. Chem. 269*, 10771–10775.
25. Croteau, D. L., DellaVecchia, M. J., Wang, H., Bienstock, R. J., Melton, M. A., and Van Houten, B. (2006) The C-terminal zinc finger of UvrA does not bind DNA directly but regulates damage-specific DNA binding. *J. Biol. Chem. 281*, 26370–26381.
26. Camenisch, U., Dip, R., Schumacher, S. B., Schuler, B., and Naegeli, H. (2006) Recognition of helical kinks by xeroderma pigmentosum group A protein triggers DNA excision repair. *Nat. Struct. Mol. Biol. 13*, 278–284.
27. Cai, Y., Patel, D. J., Geacintov, N. E., and Broyde, S. (2007) Dynamics of a benzo[a]pyrene-derived guanine DNA lesion in TGT and CGC sequence contexts: Enhanced mobility in TGT explains conformational heterogeneity, flexible bending, and greater susceptibility to nucleotide excision repair. *J. Mol. Biol. 374*, 292–305.
28. Kad, N. M., Wang, H., Kennedy, G. G., Warshaw, D. M., and Van Houten, B. (2010) Collaborative dynamic DNA scanning by nucleotide excision repair proteins investigated by single-molecule imaging of quantum-dot-labeled proteins. *Mol. Cell 37*, 702–713.