Three-dimensional structural views of damaged-DNA recognition: T4 endonuclease V, *E. coli* Vsr protein, and human nucleotide excision repair factor XPA

Kosuke Morikawa a, *, Masahiro Shirakawa b

a Department of Structural Biology, Biomolecular Engineering Research Institute (BERI), 6-2-3 Faruuedai, Suita, Osaka 565-0874, Japan
b Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0101, Japan

Abstract

Genetic information is frequently disturbed by introduction of modified or mismatch bases into duplex DNA, and hence all organisms contain DNA repair systems to restore normal genetic information by removing such damaged bases or nucleotides and replacing them by correct ones. The understanding of this repair mechanism is a central subject in cell biology. This review focuses on the three-dimensional structural views of damaged DNA recognition by three proteins. The first protein is T4 endonuclease V (T4 endo V), which catalyzes the first reaction step of the excision repair pathway to remove pyrimidine-dimers (PD) produced within duplex DNA by UV irradiation. The crystal structure of this enzyme complexed with DNA containing a thymidine-dimer provided the first direct view of DNA lesion recognition by a repair enzyme, indicating that the DNA kink coupled with base flipping-out is important for damaged DNA recognition. The second is very short patch repair (Vsr) endonuclease, which recognizes a TG mismatch within the five base pair consensus sequence. The crystal structure of this enzyme in complex with duplex DNA containing a TG mismatch revealed a novel mismatch base pair recognition scheme, where three aromatic residues intercalate from the major groove into the DNA to strikingly deform the base pair stacking but the base flipping-out does not occur. The third is human nucleotide excision repair (NER) factor XPA, which is a major component of a large protein complex. This protein has been shown to bind preferentially to UV- or chemical carcinogen-damaged DNA. The solution structure of the XPA central domain, essential for the interaction of damaged DNA, was determined by NMR. This domain was found to be divided mainly into a (Cys)4-type zinc-finger motif subdomain for replication protein A (RPA) recognition and the carboxyl terminal subdomain responsible for DNA binding. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Damaged-DNA recognition; Vsr protein; XPA

1. Introduction

DNA, the carrier of genetic information, is constantly injured by various mutagens, including UV irradiation, chemical carcinogens, and other environmental agents. DNA repair, including base excision repair, constitutes a central cellular defense mechanism to maintain genetic stability. Thus, the understanding of this repair mechanism is ranked as the most important subject in cell biology. One of the most crucial problems in DNA repair is how repair proteins specifically recognize lesions within DNA
duplexes. This review focuses on damaged DNA recognition, for which a three-dimensional structural view should be elucidated at an atomic level by X-ray crystallography and NMR. The targets of the first two crystallographic studies are a pyrimidine-dimer (PD) specific glycosidase and a mismatched base recognition endonuclease, both of which act on particular lesions at the initial stages of the excision repair pathways. The third target is a major component of a large protein complex, which plays crucial roles in eukaryotic DNA excision repair.

2. T4 endonuclease V

2.1. DNA-free form

T4 endonuclease V (T4 endo V) is a representative base excision enzyme, which long has been a major target for biochemical and biophysical studies to elucidate the mechanism of damaged DNA recognition by proteins [1]. This enzyme recognizes both cyclobutane-type PDs and abasic sites within DNA duplexes, and carries two distinct catalytic activities; cleavage of the glycosyl bond at the 5’ side of PD and subsequent scission of the phosphodiester bond at the 3’ position of an abasic site through β-elimination. It is also known that the enzyme scans nontarget sequences to search for lesions in DNA duplexes.

We successfully crystallized the T4 endo V protein purified from an Escherichia coli overproducing strain [2]. The first crystal structure of a DNA repair enzyme was determined at 1.6-Å resolution [3], and was subsequently refined at 1.45-Å resolution [4]. Despite the two distinct activities, the enzyme possesses a single compact domain, consisting of three α-helices and connecting loops (Fig. 1). The remarkable feature of the folding is the penetration of the NH2 terminal segment between the two major H1 and H3 helices to prevent their direct contact (Fig. 1). This folding scheme contradicts the close packing category for α-helices. The electrostatic surface of the protein displays an extensive basic area with dimensions of about 50 × 40 Å². This area, contributed by about 10 basic residues, exceeds the diameter of a B-DNA duplex. This surface appears to be responsible for the scanning of non-target sequences [1].

Extensive mutation analyses identified the catalytic residues for the glycosylase [5]. The E23Q (replacement of Glu by Gln), E23D, and R3Q mutations completely abolished the activity, while two

![Fig. 1. Ribbon model of T4 endonuclease V. The catalytic center of the glycosylase is shown by the side chains of the four catalytically important residues, Arg3, Glu23, Arg22, and Arg26, and the amino terminus (N), which forms an imino-covalent intermediate with C1’ of the 5’ side of the thymidine-dimer. This catalytic center occupies a small area in the basic concave interface with DNA.](image-url)
mutants, R22Q and R26Q, had remarkably decreased activity. Notably, the former two mutants exhibited full DNA substrate binding ability, although R3Q completely lost this ability. Thus, it is concluded that Arg3, Arg22, Glu23, and Arg26 constitute the glycosylase catalytic center [3,5] (Fig. 1). These catalytic residues were found to be mostly conserved in two T4 endo V homologs, which were identified in Chlorella cells infected with the PBCV-1 virus [6] and in Micrococcus luteus cells [7].

The crystal structures of the E23Q, E23D and R3Q mutants were determined at an atomic resolution [4]. The E23Q and R3Q mutants maintained the same conformation, except for the substituted side chains and some water molecules around the catalytic center, whereas the E23Q mutation induced a significant change in the backbone, such as an increase in the central kink of the H1 helix at Pro25.

2.2. Complex of the E23Q mutant with a substrate DNA

The E23Q mutant, which retains the full substrate binding ability, is assumed to be the most appropriate for crystallization with a DNA substrate duplex. Indeed, it yielded atomic resolution crystals of the complex with the DNA duplex.

ATCGC{T}TGCGCT, AGCGCAACGC,GAT, which has a thymine dimer (bold letters) at the center [8].

The crystal structure of the complex (Fig. 2) was determined at 2.75-Å resolution by the molecular replacement method [8,9]. The conformation of the enzyme in the complex agrees well with that of the

Fig. 2. Structure of the DNA substrate bound to T4 endonuclease V. The protein and the DNA are indicated by a GRASP surface representation and a licorice model [54], respectively. The flipped out adenine is shown at the center of the protein surface.
DNA-free enzyme, except for a loop (residues 125–130) near the carboxyl terminus and a flexible segment (83–91).

The DNA duplex bound to the concave basic surface is sharply kinked (60°) at the central thymine dimer moiety which divides the entire duplex into two B-DNA regions (Fig. 2). Half of the basic surface covers the L-strand through plenty of direct and water-mediated hydrogen bonds with the sugar-phosphate backbone, and likewise, the remaining half does so with the R-strand, although no interaction was observed between the DNA bases and the protein atoms.

2.3. DNA deformation includes a flipped-out base

Most remarkably, the adenine base complementary to the 5’ side thymine of the photo-dimer is completely flipped out of the B-DNA duplex interior, while the dimer itself remains inside the duplex (Fig. 2). The apparent hole created by this DNA deformation is expanded by the concavo-convex distortion of the two adjacent base pairs, although their hydrogen bonds are still retained. The flipped-out adenine base is accommodated into a cavity on the enzymatic surface. Unexpectedly, this adenine makes no polar interaction with the protein atoms within the cavity, and it appears to be sandwiched between two layers, consisting of protein atoms and water molecules, through van der Waals interactions [8,9]. This is consistent with the efficient cleavages of the modified DNA substrate duplexes [10]. The kink of the DNA duplex is accompanied by a shortened spacing between the two adjacent two phosphates, and these phosphates are recognized through seven direct polar interactions with five basic residues, including Arg3. Likewise, the neighboring phosphates pair with a shortened spacing to form water-mediated hydrogen bonds with the enzyme.

2.4. Functional significance of base flipping-out

NMR analyses [11–13] demonstrated that synthetic DNA duplexes containing a PD maintain entire base pairs, although the hydrogen bonds formed by the PD are weakened. These results indicate that the remarkable kink in the DNA is induced by the binding to the enzyme [9,10]. Thus, we proposed the following model for damaged DNA recognition (Fig. 3). While scanning DNA duplexes through the positively charged, concave surface, the enzyme may pause at PD sites, possibly because of an electrostatic disturbance around them [14,15]. At this instant, presumably, the electrostatic force between the positively charged surface of the enzyme and the negatively charged DNA backbone may suppress the PD moiety, in which stacking interactions are weakened [9,10]. This model assumes the direct coupling between

![Fig. 3. Process for damaged DNA recognition by T4 endonuclease V. (a) Sliding (b) pause at a pyrimidine dimer (c) specific recognition.](image-url)
between the base flipped-out and the kink of the DNA duplex. Then, the functional role of the base flipping-out would partly be to alleviate the local tension generated by the kink. The efficient binding of the enzyme with an abasic site within DNA duplexes supports this interpretation [16]. Recently, an analogous example has been reported, in which a kinetic study of the EcoRI DNA methyltransferase-DNA interaction demonstrated the coupling of base flipping-out with DNA bending [17].

Another important aspect of the base flipping-out is that the hole generated within the DNA duplex is filled with protein atoms that all participate in the catalytic reaction. It appears that the hole provides a room for the catalytically essential atoms of the NH2 terminus, Arg22, Glu23 and Arg26, to access the target bond for the glycosylase reaction [8,18,19]. The enclosure of the hole may also contribute to protecting a transition state from unfavorable contacts with water molecules during hydrolysis. In addition, the flipped-out base may prevent the slippage of the DNA substrate along the protein interface. These interpretations imply that base flipping-out would be a universal mechanism for DNA recognition by base excision repair enzymes or DNA modifying enzymes. Indeed, similar flipped-out bases were found in various repair proteins, including mismatch repair enzymes [20–24].

Transcription and replication involve multitude of reaction steps, where huge molecular machineries, consisting of dozens of protein subunits, act on DNA molecules. Consequently, they need to interact with a large number of atoms buried within DNA duplexes through extensive interfaces, and therefore, segmental unwinding of the DNA is essential. On the other hand, the targets for base excision repair enzymes are so small that base flipping-out provides a sufficient space for the protein atoms to access the target bonds inside or outside of the DNA duplexes.

3. Very short patch repair (Vsr) endonuclease

3.1. Vsr endonuclease

The aforementioned implications of the base flipping-out raised the new question of how DNA endonucleases other than base excision repair enzymes recognize mismatched base pairs. The E. coli Vsr protein is a good target to address this question in terms of three-dimensional structures.

The Vsr endonuclease is involved in the initial reaction for the repair of mismatched TG base pairs generated through the spontaneous deamination of a methylated cytosine. This enzyme recognizes a TG mismatch within the duplex 5′CT(A/T)GG, where the second T forms the mismatch and all of the other bases are in standard Watson-Crick base pairing. It catalyzes the cleavage at the 5′ side of the thymine, leaving a 5′ phosphate and a 3′ hydroxyl at the termini [25–28]. Subsequently, DNA polymerase I [27,29] and presumably DNA ligase are involved in the final stages of repair, where less than 20 bases on either side of the cleavage site need be excised [28]. It was shown that MutL and MutS are required for efficient repair of these TG mismatches [30–32], although the detailed mechanism remains unclear. It was recently reported that MutL enhances Vsr binding to DNA, indicating that the Vsr endonuclease may recognize the T–G mismatch present in the DNA loop created by the MutL–MutS action [33,34].

In order to understand how the Vsr endonuclease recognizes the TG mismatch and cleaves the specific site, we have undertaken its atomic structure determination by X-ray crystallography. We describe here the unexpected recognition mechanism for mismatch base pair and the catalytic mechanism, in comparison with those for base excision repair enzymes.

3.2. Crystal structure of the Vsr truncated form

The crystallization of the full length Vsr molecule was not successful in the absence of DNA, despite extensive screening. Limited proteolysis defined the core domain missing 20 residues at the NH2 terminus. This fragment, which retains substantial DNA cleavage activity, yielded crystals suitable for structure determination. The crystal structure was determined at 1.8-Å resolution by the multiple isomorphous replacement method [35]. In the course of the structural analysis, we found, from the anomalous Patterson map of the substituted Cd atom, that the enzyme contains a single Zn atom.

The Vsr structure consists of a single domain, in which a central five-stranded β-sheet is flanked on
each side by one and two α-helices [35]. The protein surface on one side represents a deep cleft. A zinc atom is coordinated with Cys66, His71, Cys73, and Cys117 (Fig. 4). A similar coordination scheme was found in the Ppo1 endonuclease [36], although the loop structure, including the former three residues, is conformationally different from the corresponding loop of Ppo1.

A striking finding is that the Vsr main chain folding belongs to the Type II restriction enzyme family [37], although they have no significant sequence similarity. The central portion of the Vsr β-sheet superimposes well onto the catalytic β-strand region of type II restriction endonucleases, such as EcoRV [38] and Fok1 [39], with a range of 1.2 to 1.6 Å for the average root mean square displacement values. This restriction enzyme family bears a common catalytic motif, DX(6–30)DEXK [37], which is located at the interface of the short and long strands and is involved in catalytic metal binding. The comparison between Vsr and the restriction enzymes revealed a good agreement in the backbone structures in the motif region. However, Vsr has no acidic or basic residues in the (D/E)XK sequence, and the second acidic residue in Vsr is equivalent to the invariant Phe62 (Fig. 4), the side chain of which cannot participate in metal binding.

### 3.3. Alanine scanning mutational analyses

In parallel with the crystallographic studies, single alanine mutations were introduced at conserved aspartate or glutamate residues. The mutants were purified and their relative endonuclease activities were compared to that of the wild type. There were three notable mutants: the D51A mutant completely lost the activity, while E25A and D97A showed reduced activity (Fig. 4).

Soaking experiments indicated that the Mn$^{2+}$ atom is most likely to coordinate with His69. It also lies in the proximity of His64, and is 12.1 Å away from the Zn site. It was confirmed by cleavage analysis that

---

**Fig. 4.** Sequence alignments and secondary structure of Vsr. Multiple alignment and secondary structure of *E. coli* Vsr and Vsr-related proteins *Xanthomonas oryzae*, *Nocardia aerocolonigenes*, and *Haemophilus parainfluenzae*. Invariant residues and residues partially conserved with *E. coli* Vsr are framed in black and grey. The aromatic residues inserted into the proximity of the GT mismatch are indicated by solid circles. The residues involved in the catalytic reaction are also shown by solid squares.
Mn$^{2+}$ is as effective in catalytic efficiency as Mg$^{2+}$. Consistent with these results, the His69A mutant had hardly any activity, while the His64A mutation decreased the activity.

3.4. Crystal structure of Vsr complexed with DNA

A docking examination of the truncated Vsr with hypothetical DNA substrates allowed us to design DNA duplexes, which would be suitable for crystallization of the full-length Vsr–DNA complexes [40]. Finally, we successfully obtained atomic resolution crystals in the presence of 20 mM MgCl$_2$ from the mixture of Vsr and the blunt end 12-base pair DNA duplex:

AGCTAGGTACGT
TCGGTCCATGCA

where the underlined and the bold bases indicate the recognition sequence and the TG mismatch, respectively. The crystal structure of the complex (Fig. 5a, b, and c) was determined by the molecular replacement method, where the truncated Vsr structure was used as a search model. The structure refined at 2.3-Å resolution provided $R_c = 0.20$ and $R_{free} = 0.23$ [40]. The final model includes all 155 amino acids of the full-length protein, except for the N-terminal methionine removed in vivo, one Zn$^{2+}$ atom, two Mg$^{2+}$ atoms, and three DNA fragments (Fig. 5a).

In agreement with the results from limited proteolysis, the N-terminal 20 residue segment, which is eliminated in the truncated Vsr protein, forms a short $\alpha$-helix AA (Lys7–Arg15) and a loop preceding the N-terminal $\alpha$-helix A (Fig. 5a and c). Apart from the insertion of this AA helix into one of the DNA grooves and a slight shift of the A helix, the architecture of the intact Vsr complexed with DNA is essentially identical to that of the DNA-free truncated form. The highly basic interface of the protein intimately contacts the DNA duplex. The strong acidic patch located between the two basic ridges corresponds to the catalytic center, which binds essential divalent metal cations. This feature implies that the strong repulsive force between the phosphate backbone and the negatively charged active center is alleviated by metal binding, and partly explains the DNA mobility shift assay result that the presence of divalent metals was essential to form the protein–DNA complex (unpublished data).

3.5. Novel DNA recognition scheme

The most striking feature of the complex is the intercalation of three invariant aromatic residues, Phe67, Trp68, and Trp86 (Fig. 4), between the TG mismatch and the adjacent AT base pair [40] (Fig. 5a and b). The former two side chains are stacked with the central AT base pair in the recognition sequence and the mispaired thymidine, respectively, while the Trp86 side chain makes a stacking interaction with the two sequential sugar rings of the strand opposite the cleaved strand. These aromatic residues approach the bases on the major groove side. Met14 and Ile17 around the C-terminus of the AA helix penetrate from the minor groove to interact with the TG mismatched bases and the neighboring guanine base through peptide atoms. Consequently, small spaces, which allow only the phosphate backbone to pass through, are left around the TG mismatch. These extensive contacts generate dramatic changes in the DNA geometrical parameters, most typically, the 6 Å VRISE spacing between TG and AT base pairs and the 57° VROLL, and thus an overall bend of 44° is produced with decreased twist angles near the intercalation site. In spite of this serious conformational strain, all of the base pairs remain intact within the DNA duplex.

To date, a number of DNA–protein complex structures have shown a DNA bend induced by the insertion of aromatic side chains [41]. However, the recognition schemes in these complexes are clearly different from that of the Vsr–DNA complex. In DNA–protein complexes, including the TATA box binding protein [42,43], the high mobility group (HMG) protein [44,45], and the integration host factor [46], the penetration of aromatic residues generates a small wedge through the partial disruption of base stacking interactions. Thereby, the deformed state is induced, resembling A-type DNA with a larger minor groove and a compressed major groove. On the other hand, the penetration of the aromatic residues in the Vsr–DNA complex takes place on the major groove side (Fig. 4a), in contrast with the insertion on the minor groove side in the other complexes. Furthermore, Met14 and Ile17 in the
Fig. 5. Recognition between Vsr endonuclease and DNA. (a) Stereoview of the Vsr/DNA complex structure, with a ribbon representation of the protein and licorice model of the DNA, where carbon, oxygen, nitrogen, and phosphate are depicted in white, red, blue, and yellow, respectively [54]. The structural Zn$^{2+}$ atom and the two Mg$^{2+}$ atoms at the active site are represented by spheres. The mismatched thymine and guanine bases are denoted and form a wobble base pair. (b) Electrostatic and structural surface complementarity of Vsr endonuclease to the DNA. The protein surface is represented and colored according to positive blue, neutral white, and negative red electrostatic potential [54]. The N-terminus (residues 2–20) has been removed for clarity. The DNA is represented as in (a). The three residues intercalating into the DNA duplex are denoted. Metals are not represented in this figure. (c) Vsr endonuclease tightly clamps the DNA duplex. GRASP surface representation of the Vsr endonuclease with the DNA represented as in Fig. 1 [54]. The perspective is also similar to (a).

Vsr–DNA complex are inserted so as to form direct and indirect hydrogen bonds with the bases, and thereby substantially widen the minor groove to a size large enough to accommodate the N-terminal AA helix.

3.6. Structural basis for TG mismatch recognition

The Vsr–DNA complex does not show a flipped-out base, which is universally employed for mismatched base pair recognition by base excision re-
pair enzymes, such as the MutY glycosylase [23] and the uracil DNA glycosylase [20,21,24]. In the complex, the TG mismatch forms a classical wobble base pair, which does not cause serious strain against DNA duplexes. Both of the TG bases form direct hydrogen bonds with the invariant Asn93 side chain and the main chain carbonyl of Met14, in addition to water mediated hydrogen bonds with residues near the C-terminus of the AA helix. Steric effects, including van der Waals contacts, also play major roles in the recognition; the significant shift of the TG mismatch to the normal CG pair within the DNA duplex is recognized by various surrounding protein atoms. In addition, the slightly deformed DNA backbone appear to partially contribute to the recognition by Vsr.

The Vsr endonuclease requires a five base-pair recognition sequence. These nucleotides make a multitude of direct and indirect polar interactions with the protein atoms, although the other DNA region is hardly recognized by the protein. This feature is in good agreement with the experimental results for the substrate specificity requirements [47,48].

Why is base flipping-out not required for Vsr–DNA recognition? The base flipping has two significant functions the specific recognition of lesions through the indirect 8–10 or direct read-out mechanism [24] and the generation of catalytic space. The structural basis for site specific DNA recognition by Vsr is the minor structural difference ascribed to the GT mismatch, in addition to the many interactions between the recognition sequence and the protein atoms. Furthermore, the catalytic residues of the Vsr endonuclease attack the particular phosphodiester bond lying outside of the DNA duplex. Thus, neither context requires the base flipping-out for Vsr–DNA recognition.

3.7. Catalytic mechanism

The Vsr–DNA complex crystal structure [40] confirmed that the catalytic site, which was suggested from the DNA-free Vsr structure [35] and the alanine scan mutational analyses [35], was indeed correct. The imidazole ring of His69 coordinates with the cleaved phosphate on the 5’ side of the mismatched thymine and with one of the two Mg$^{2+}$ ions. These two Mg$^{2+}$ ions, with locations similar to those found in the classical two metal binding mechanism [49], were identified by the clear octahedral-shaped electron densities conforming well with the magnesium–water clusters. Surprisingly, the cleaved deoxyribose-3'-oxygen was found at a distance of 2.3 Å from one of the Mg$^{2+}$ ions, suggesting that the DNA is trapped at an intermediate stage just before proton donation to the oxygen. As suggested previously [35], the carboxyl side chain of Asp51 is coordinated with both Mg$^{2+}$ ions.

The importance of magnesium–water clusters in endonuclease catalytic mechanisms was pointed out in a study of the apoenzyme crystal structure of Serratia endonuclease [50]. Consistent with this report, Glu25 and His64, whose mutations reduced activity, form hydrogen bonds with the clusters in the complex structure. The equivalent occupancy of the two metal sites is supported by the B-values of 24.9 and 22.5 Å$^2$ and by the peak heights of 13.7 and 13.8 σ, respectively, on the simulated annealed $F_o-F_c$ omit map. In addition to the coordination scheme, these values exclude the possibility that metals other than Mg$^{2+}$ occupied the same positions.

Finally, it should be stressed that these structural features do not necessarily support the two metal catalytic mechanism, as proposed for various nucleases and polymerases [51]. In fact, it took 3 weeks at room temperature to obtain the crystal of the complex containing a reaction intermediate, implying that the reaction rate is extremely slow. It remains unclear whether this slow rate indeed reflects the catalytic efficiency in vivo. It is possible that other proteins, such as DNA polymerase I, may function to facilitate the release of the cleaved product from the Vsr endonuclease. On the other hand, we wonder if the crystal structure with the two metals may correspond to the scene locked at a certain stage during the hydrolytic reaction. In reference to this question, it is interesting that the crystal structure of the non-productive BglI endonuclease–Ca–DNA ternary complex showed a coordination scheme with two Ca$^{2+}$ ions similar to that in the Vsr–DNA complex [52]. Likewise, Ca$^{2+}$ ions may inhibit the BglI endonuclease activity by simultaneous binding to the two sites. As proposed in the previous study on RNase H [53], the “metal ion shift mechanism”.

where a bound metal shift from one site to the other during the catalytic reaction, may also apply to the Vsr catalytic mechanism.

4. Human nucleotide excision repair factor XPA

4.1. Nucleotide excision repair

Nucleotide excision repair (NER) is a most versatile and ubiquitous DNA repair pathway by which a broad spectrum of structurally unrelated forms of DNA, damaged by UV and chemical carcinogens, is removed from the genome. The NER pathway recognizes DNA lesions, repairs them by excising a 24–32-nucleotide long oligomer carrying the damaged bases, and synthesizes a repair patch [55,56]. The biological significance of NER has been highlighted by studies on the human inherited disease, xeroderma pigmentosum (XP), which is characterized by a 1000-fold higher frequency of skin cancer on sun-exposed areas, and by neurological complications [57]. Cells from XP patients have defects in NER and are hypersensitive to UV irradiation. Cell fusion studies have revealed the presence of seven complementation groups (A–G) and a variant form in the XP cells, reflecting different NER genes [58]. With the identification of the protein factors corresponding to these NER genes, in vitro NER reactions have been reconstituted, and showed that the asymmetric double incisions are accomplished by the sequential and cooperative actions of 14 polypeptides in six repair factors, XPA, RPA, XPC, TFIIH, XPG and XPF-ERCC1 [59–62].

4.2. XPA

XPA is a 273 amino acid zinc-finger protein, which has been shown to bind preferentially to UV- or chemical carcinogen-damaged DNA in vitro, and thus has been suggested to be involved in the damage recognition step of NER [61,63–68]. XPA binds directly to the following repair factors: replication protein A (RPA), XPF-ERCC1 heterodimer, and TFIIH, which contains XPB and XPD as its subunits [69–76]. Therefore, it has been suggested that XPA might play a role in loading the incision protein complex onto a damaged site, as a multifunctional protein that coordinates the early steps of NER processes [71,73]. Although XPA is implicated in recognizing and/or verifying lesions in damaged DNA, its actual role in the damage recognition is not well understood. Of the six essential NER factors, RPA [77,78], XPA [66], and XPC [79] have preferential binding affinity for damaged DNA, although none has sufficient independent selectivity for damaged DNA [80–82]. For the highly specific damage recognition, XPA, RPA, TFIIH-XPC-hHR23B, XPG, and ATP were shown to be indispensable [61]. On the other hand, Sugawara et al. [80] suggested a two-step mechanism in which damage is detected by XPC-hHR23B and verification by XPA follows. In contrast, Wakasugi and Sancar [81] proposed that XPA and RPA are the initial damage sensing factors of the human excision complex, based on their analyses of the repair kinetics under a variety of “order of addition” conditions.

XPA consists of several distinct functional domains. Its amino-terminal part contains regions for binding to a subunit of RPA, RPA34, and for ERCC1 [71,72,74–76]. The carboxy-terminal part of XPA has been shown to bind to TFIIH and to recruit it to the damaged site [72,73]. The central domain (residues 98 to 219) has been identified as the minimal region essential for the preferential binding to damaged DNA, using limited proteolysis and deletion analyses [67]. This domain also includes the region for binding to RPA70 [70,71].

The central domain encompasses a C4-type zinc-binding sequence of Cys-X-X-Cys-(X)4–Cys-X-X-Cys in its amino-terminal region. EXAFS spectroscopy [83] and ¹¹¹Cd-NMR analyses combined with site-directed mutagenesis and atomic absorption [84] have shown that one zinc ion is tetrahedrally coordinated by these four cysteine residues. A mutation at each of the four cysteines causes a drastic reduction in the UV resistance of cells [85]. In contrast, a mutant at Cys-153, located outside of the zinc-binding sequence, retains almost the same UV resistance as that of the wild-type XPA. These results indicate that XPA has a Cys₄-zinc-finger motif, and that this zinc-finger is essential for the function of XPA.
4.3. Structure of the central domain of XPA

The solution structure of the central domain of XPA has been determined using multidimensional heteronuclear NMR spectroscopy [86–88]. The central domain is made up of a zinc-containing subdomain (residues 102 to 129), a carboxyl-terminal subdomain (residues 138 to 209), and an eight-residue linker sequence that connects these two subdomains (Fig. 6). The backbone conformations are well defined in the structure determination, except for the linker sequence and two long loops (L1: residues 148–163, L2: residues 166–179) in the carboxyl-terminal subdomain.

The zinc-containing subdomain is composed of a β hairpin (residues 103–112) and a helical turn (residues 116–121). The loop that connects the two β-strands in the hairpin encompasses the zinc-coordinating Cys-105 and Cys-108 residues, while the loop that follows the helical turn provides the other two zinc-coordinating cysteine residues, Cys-126 and Cys-129. The overall folding of this subdomain is largely defined by the zinc ion, which is coordinated to the four cysteines in an S configuration [89]. In addition to this tight packing around the zinc ion, the subdomain is further stabilized by a hydrophobic core, thus fixing the relative orientation of the β-sheet and the helical turn.

The zinc-containing subdomain has a hydrophobic patch on the outer surface of the helical turn (Fig. 7a). This patch has an extended interaction with the well-conserved hydrophobic residues in the carboxyl-terminal subdomain, tightly packing the helical turn against a three-stranded antiparallel β-sheet of the carboxyl-terminal subdomain. These remote contacts between the subdomains are probably required to maintain at least the local structure of both the zinc-containing and carboxyl-terminal subdomains. Thus, neither the zinc-containing nor the carboxy-terminal subdomains are structurally independent. In contrast, structural independence is a common feature of the zinc-fingers whose tertiary structures have been determined: these Zn-fingers exist as highly stable, independently folded domains [90].

The overall folding of the zinc-containing subdomain shows a similarity with that of C_{2}H_{2}-type zinc-fingers: each zinc-finger contains a β hairpin and a helical conformation at the N- and C-terminus, respectively [87,90]. In the zinc-containing subdomain of XPA, a series of extensive hydrogen bond networks has been found around the four cysteine residues [87]. These structure elements are common in the (Cys)_{4} type zinc-fingers of the erythroid transcription factor GATA-1 and the glucocorticoid receptor [90], although a comparison of the zinc-binding sequences of XPA and GATA-1 shows that only the positions of the four cysteines and a proline at XPA position of 124 are identical.

The acidity is a feature of the zinc-containing subdomain of XPA that makes it distinct from the zinc-fingers of DNA-binding transcription factors (Fig. 8a). The zinc-containing subdomain and the subsequent linker sequence both include many conserved glutamate and aspartate residues (Fig. 7a). Many of them are confined in the region containing the β-hairpin, residues 99 to 114, forming a large superficial acidic patch.

The carboxy-terminal subdomain is made up of two modules: a sheet-helix-loop region and a helix-turn-helix region (Fig. 7b). The sheet-helix-loop region consists of a three-stranded antiparallel β-sheet, helix α1, and two long loops, L1 and L2. Loop L2 is highly mobile in solution relative to the overall tumbling rate of the molecule (see Section 4.6). The helix-turn-helix region contains two long helices, α2 and α3, connected by the well-conserved Gly–Ser sequence forming a characteristic turn. The angle between helices α2 and α3 is kept to about 60° by hydrophobic contacts between Leu-200, Ala-203, and highly conserved Trp-194. These hydrophobic residues also serve as part of a large and well-defined hydrophobic core formed by packing the helix-turn-helix region against the sheet-helix-loop region.

The association of the sheet-helix-loop region and the helix-turn-helix region creates a shallow cleft. A color display of the electrostatic potentials shows that many positively charged side chains are present in the cleft (Fig. 7b). These residues are primarily located in helices α1 and α3. These helices also have many conserved glutamate residues, but none of the negatively charged side chains of these residues are located in the cleft, thus giving a bipartite feature to helices α1 and α3 (Fig. 7b). These acidic residues form negatively charged patches on the opposite side of the basic cleft.
Fig. 6. Solution structure of the central domain of human XPA [86,87]. Schematic ribbon drawing of the NMR structure of the central domain of human XPA in stereo. The domain consists of a zinc-containing subdomain (residues 102 to 129) and a C-terminal subdomain (residues 138 to 209), connected by a linker sequence (PDB code: 1xpa) [86]. The following secondary structure elements are indicated: helices \( \alpha_1 \) (residues 141 to 147), \( \alpha_2 \) (residues 183 to 194), \( \alpha_3 \) (residues 197 to 209), strands \( \beta_1 \) (residues 103 to 104), \( \beta_2 \) (residues 111 to 112), \( \beta_3 \) (residues 139 to 140), \( \beta_4 \) (residues 164 to 165), and \( \beta_5 \) (residues 180 to 181), and loops L1 (residues 148 to 163) and L2 (residues 166 to 179). The figure was produced with MOLSCRIPT [102] and RASTER3D [103].

A comparison of the amino acid sequences of the central domain of human XPA with those of the XPAs from other eukaryotes shows that the hydrophobic residues are well conserved at the pos-
tions of the residues defining the hydrophobic core and the subdomain interface of the human XPA. The residues of human XPA that contribute to the positive charges in the basic cleft and the carboxyl-terminal flanking sequence (residues 211 to 219) are also well conserved. These findings suggest that the tertiary structure and the positive charges in the basic cleft of XPA are conserved among eukaryotes, and the structure of human XPA will allow for useful modeling of other XPAs. Although the conservation of the acidic residues in the zinc-containing subdomain is clear among the XPAs from human, mouse, chicken, and Xenopus, it is not apparent in either Drosophila XPA or the yeast RAD14.

4.4. The DNA binding surface

The DNA binding site of the central domain of human XPA has been investigated by chemical shift perturbation experiments using $^{15}$N-enriched XPA$_{98-219}$ and damaged DNAs [86,88]. This type of NMR experiment is often used to probe the binding surfaces of proteins and their ligands [91–93]. In the experiment using the cisplatin-treated oligonucleotide as a damaged DNA, most of the signals that showed remarkable chemical shift perturbation or broadening were attributed to the amide residues in the basic cleft of the carboxy-terminal subdomain [86] (Fig. 8b). Similar experiments using DNA with other kinds of lesions, such as dihydrothylmidine or 6-4-thymidine-cytidine, exhibited essentially the same pattern of chemical shift changes [88]. These findings strongly indicate that the basic cleft and the surrounding region, but not the zinc-containing subdomain, are involved in the interaction with the DNA. A visual inspection revealed that the internal

![Fig. 8. Mapping of the electric potential, the interaction sites, and the flexible region of the central domain of human XPA [86,87]. (a) Distribution of the electrostatic potential on the solvent-accessible surface [104]. Blue corresponds to positive potential and red to negative potential. The molecule is viewed in the same orientation in all of the panels of this figure, as in Fig. 6. (b) Mapping of the XPA residues with large chemical shift perturbations or broadening effects in the $^{15}$N-1H HSQC spectra [86]. The residues showing perturbed amide resonances upon complex formation with the cisplatin-damaged DNA are colored magenta, and the residues showing specifically broadened amide resonances upon complex formation with RPA70$_{181-422}$ are colored green. (c) Mapping of the flexible parts of XPA that are related to the DNA binding activity [87]. Regions exhibiting internal motions on a picosecond to nanosecond time scale, characterized by small $^1$H,$^{15}$N NOE values (< 0.55 at 50.7 MHz), small S2 values (< 0.6), large T2 values, slow internal correlation times on the order of 1 ns, and little contribution from Rex values, are indicated in magenta. A region exhibiting conformational exchange on a microsecond to millisecond time scale, characterized by large Rex values (> 2.0 s$^{-1}$ at 50.7 MHz), is indicated in red. These regions are located in the interaction surface for DNA. Residues 133 and 134, with large Rex values (> 3.5 s$^{-1}$ at 50.7 MHz), are indicated in cyan. This region is a linker connecting the zinc-containing subdomain and the C-terminal subdomain, and exhibits conformational exchange on a microsecond to millisecond time scale. The figures in (b) and (c) were produced with the program, MOLMOL [105].

curvature of the basic cleft fits roughly well to the diameter of standard B-form double-stranded DNA. A similar pattern of chemical shift perturbations was observed for complex formation with a non-damaged DNA, indicating the low binding specificity of XPA [86,88].

Deletion of a part of the zinc-containing subdomain or substitution of the zinc-coordinating Cys-108 with Ser resulted in the loss of its DNA binding activities, suggesting that the tertiary structure of the zinc-containing subdomain is necessary for the DNA binding activity of XPA [66,67]. As it is clear from the chemical shift perturbation experiments that the zinc-containing subdomain does not directly interact with the DNA, it is assumed that the zinc-containing subdomain is required for the maintenance of the conformation of the carboxy-terminal subdomain. Thus, hydrophobic contacts between the subdomains are important, and their loss caused by unfolding of the zinc-containing subdomain may result in a conformational distortion of the carboxy-terminal subdomain and a consequent loss of the DNA binding activity.

4.5. A binding surface for RPA70

RPA is a heterotrimeric, single-strand DNA binding protein, which is involved in DNA replication, homologous recombination, and NER, and its interaction with XPA is essential for NER [59,69,70,94]. XPA binds to the two largest subunits of RPA, RPA70 and RPA34 [69–71]. RPA70 is composed of several domains: an N-terminal domain that participates in protein interactions, a middle domain, a putative C4-type Zn finger, and a C-terminal inter-subunit interaction domain [97,98]. The location of XPA binding domain in RPA70 has not been well elucidated from previous deletion analyses. Two groups showed that no particular domain of RPA70 is essential for the interaction with XPA [95,96]. On the other hand, Walther et al. [97] reported that the RPA trimer that has a deletion of residues 1-236 of RPA70 interacted at an intermediate level but a deletion 1–382 of RPA70 caused a very poor interaction with XPA. These observations suggest that the interaction domain includes residues 237–382. The middle domain (residues 181–422) of RPA70, RPA70<sub>181–422</sub>, forms a compact structural domain that has single-stranded DNA binding activity, and its crystal structure has been determined [98,99].

The interaction of the central domain of XPA with RPA70<sub>181–422</sub> was analyzed by a chemical shift perturbation experiment using <sup>15</sup>N-enriched XPA<sub>98–219</sub> and RPA70<sub>181–422</sub>. Most of the residues whose signals showed large losses were confined to the zinc-containing subdomain, suggesting that this subdomain serves as one of RPA70-binding surfaces of XPA (Fig. 8b). Larger signal losses were observed for several residues in the β hairpin and its vicinity (residues 101 to 114), where five acidic residues are relatively conserved. The binding between the central domain of XPA and RPA70<sub>181–422</sub> is markedly sensitive to ionic strength. As the KCl concentration in the sample solution was increased from 28 to 77 mM, the signal loss of XPA<sub>98–219</sub>, caused by its binding to RPA70<sub>181–422</sub> was much reduced. These results indicate that the molecular interaction is weak, and may be due primarily to the electrostatic force between the negatively charged patch around the β hairpin of XPA and a positively charged patch in RPA70<sub>181–422</sub>.

By deletion analyses, Saijo et al. [71] have shown that residues 98–187 of XPA, which contain the zinc-containing subdomain and the sheet-helix-loop region of the carboxy-terminal subdomain, are required for binding to RPA70, but deletion of the sheet-helix-loop region from XPA<sub>98–187</sub> would cause the loss of RPA70 binding. These results indicate that the sheet-helix-loop region of the carboxy-terminal subdomain is also required for RPA70 binding. Thus, it is conceivable that the sheet-helix-loop region of XPA would serve as another binding surface with the amino- or carboxy-terminal region, outside of RPA70<sub>181–422</sub>, of RPA70. This was examined by a recent chemical shift perturbation experiment using the central domain of XPA and a another fragment of RPA70, RPA70<sub>1–326</sub> [88]. RPA70<sub>1–326</sub> was shown to contact several residues in the sheet-helix-loop region of the carboxyl-terminal subdomain, as well as those in the zinc-containing subdomain. There is also a possibility that the removal of the sheet-helix-loop region induces minor structural distortion of the zinc-containing subdomain, which causes a reduction in the RPA70 binding of the subdomain, because there are extensive hydrophobic interactions between
the sheet-helix-loop region and the zinc-containing subdomain.

4.6. Dynamic properties

The backbone dynamics of the central domain of XPA have been studied by analyzing $^{15}$N relaxation data obtained at two static magnetic fields [87]. The data were extensively analyzed by means of both the model-free formalism and the spectral density function at the zero frequency [100]. This series of analyses supplied parameters for the internal motions over a wide range of time scales, as well as for the overall rotational diffusion of the domain. The analysis showed that the domain exhibits a statistically significant anisotropy of rotational diffusion.

The zinc-containing subdomain was found to be rigid [87]. This characteristic is probably due to the stabilization of the folding by the zinc coordination with its associated extensive hydrogen bond networks, and by the hydrophobic core. In contrast, several residues on the DNA binding surface in the basic cleft exhibit large extents of either fast (a pico to nanosecond time scale) or slow (a micro to millisecond time scale) internal motions (Fig. 8c). In particular, loop L2 seems to be the most flexible part of the domain, as shown by the steady-state heteronuclear $^{15}$N($^1$H)-NOE values, which are direct measures of protein flexibility [87,100]. NOE values smaller than 0.67 were observed for the residues (157–162) in loop L1 and for all the residues in loop L2 except residues (171–174) whose values could not be obtained due to the lack of their signals in the HSQC spectra [86]. This indicates that loop L2, part of loop L1, and the carboxyl-terminal flanking sequence are highly mobile in solution relative to the overall tumbling rate of the molecule.

The high flexibility in the DNA binding site is consistent with the model of ‘thermodynamic probing of the DNA duplex’, in which the destabilization of the duplex introduced by the lesion is the determinant of the recognition by NER factors [88,101]. The more a lesion causes destabilization, the stronger the NER factors can bind to it [101]. The large internal motion of these residues on the DNA binding domain may play a key role in the interaction of XPA with a variety of types of damaged DNA by fitting the conformation of the interaction surface to the structures of the damaged DNAs. This possibility was partly substantiated by a recent analysis of the backbone dynamics of the domain bound to DNA: the carboxyl-terminal subdomain becomes more rigid upon binding to DNA [88].

Acknowledgements

We thank Dr. S. Tsutakawa and Prof. E. Ohtsuka for discussions. We also thank Dr. T. Ikegami for discussions and figure preparations.

References

[38] P.K. Winkler, D.W. Banner, C. Oefner, D. Tsernoglou, R.S. Brown, S.P. Heathman, R.K. Bryan, P.D. Martin, K. Petrotos, K.S. Wilson, The crystal structure of EcoRV en-


[98] R.A. Pluetzner, A. Bochkarev, L. Frappier, A.M. Edwards,


