REVIEW

Structure and mechanism for DNA lesion recognition

Wei Yang[1]

[1]*Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, 9000 Rockville Pike, Bethesda, MD 20892, USA*

A fundamental question in DNA repair is how a lesion is detected when embedded in millions to billions of normal base pairs. Extensive structural and functional studies reveal atomic details of DNA repair protein and nucleic acid interactions. This review summarizes seemingly diverse structural motifs used in lesion recognition and suggests a general mechanism to recognize DNA lesion by the poor base stacking. After initial recognition of this shared structural feature of lesions, different DNA repair pathways use unique verification mechanisms to ensure correct lesion identification and removal.

Keywords: base unstacking, MMR, BER, NER, ATPase, nuclease, glycosylase

Introduction

DNA repair is generally categorized into four different types, base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR) and double-strand break repair (DBR) [1]. BER, NER and MMR share a similar repair procedure that includes lesion recognition and excision. Lesions can be modified bases as recognized by BER and NER or normal bases paired with wrong partners as recognized by MMR. Excision includes an incision in the strand containing the lesion and removal of as few as 1-2 nucleotides as in BER or as many as hundreds as in MMR. After excision, these three repair pathways require new DNA synthesis to replace the removed nucleotides and ligation to seal the single-strand nick. DBR is achieved by homologous recombination or non-homologous end joining. For a complete description of each repair pathway, please consult the relevant reviews in this issue. Since DNA double-strand breaks concern the continuity of phospho-backbones, which is distinctly different from mismatched or modified bases, the mechanism of detecting broken DNA ends is beyond the scope of this review. The focus here is detection of base lesions in BER, NER and MMR.

DNA lesions that affect individual bases without grossly changing the double-helix structure are most often repaired by BER. Substrates for BER include spontaneous deamination of Cyt, Ade and Gua, oxidation by reactive oxygen species, e.g. Gua to 8-oxo-Gua (GO), and alkylation of Ade, Gua and Cyt by exogenous agents [2]. Certain alkylation damages can be repaired by direct removal of alkyl adducts [3, 4]. BER removes a damaged nucleotide in two consecutive steps. The first step is catalyzed by a glycosylase, which cleaves the glycosidic bond between the base and deoxyribose (deglycosylation). Glycosylases are endowed with two essential functions for BER, detecting a damaged base (lesion recognition) and removing the damaged base or its mismatched pairing partner (see discussion later; Figure 1A). DNA glycosylases are specialists, each recognizing one or at most a small subset of lesions. Most organisms express multiple glycosylases in order to protect the genome from a broad spectrum of damage. There are two classes of glycosylases, mono-functional (deglycosylation only) and bi-functional (deglycosylation and backbone cleavage at the 3′ side of the lesion) [2]. The second step in BER is catalyzed by an AP (apurinic/apyrimidinic) endonuclease, which recognizes an abasic product of glycosylase and cleaves 5′ to the lesion to generate a normal 3′-OH for DNA re-synthesis [5]. Structures of a large number of glycosylases, two classes of AP endonucleases, and their complexes with respective DNA substrate or substrate analogs are available at atomic resolution [2], which allow us to examine how a modified base is recognized by a glycosylase and an abasic lesion by AP endonuclease.

Bulky DNA adducts like benzo[*a*]pyrene and DNA lesions caused by intra-strand crosslinking agents, e.g. ultraviolet light and cisplatin, are usually removed by the

Correspondence: Wei Yang
Tel: +1-301-402-4645; Fax: +1-301-496-0201
E-mail: Wei.Yang@nih.gov

NER pathway (Shuck *et al.* in this issue). NER may also be involved in protection against interstrand crosslinking compounds, psoralen and mitomycin C. Unlike BER, which depends on specialized glycosylases to remove different types of damaged bases, NER uses the same set of proteins to recognize and remove various bulky adducts and crosslinked bases that severely distort the DNA duplex. Interestingly, lesion recognition proteins in the NER pathway do not possess nuclease or glycosylase activity and cannot remove lesions directly. Instead, lesion recognition leads to recruitment of specialized nuclease(s) to excise 12-30 nucleotides including the lesion [6, 7] (Figure 1B). The challenge of finding a variety of lesions without a defined shape or chemical nature in NER is hence more confounding than specific lesion recognition in BER. The task of detecting and excision of a lesion by NER requires at least three proteins (UvrA, UvrB and UvrC) in *Escherichia coli* and over a dozen proteins in humans [8]. Unlike

glycosylases and AP endonucleases, which are conserved throughout three domains of life (bacteria, archaea and eukarya), bacterial and eukaryotic NER proteins share no discernible similarity. Recently, structures of NER proteins, bacterial UvrB and yeast Rad4, complexed with damaged DNAs have been reported [9, 10]. Our understanding of lesion recognition is also gleaned from structural studies of proteins outside of the NER pathway yet capable of repairing a typical NER substrate, cis-syn thymine dimers, e.g. T4 endonuclease V [11] and photolyase [12].

The MMR pathway mainly removes nucleotides misincorporated by DNA polymerase and thereby improves the overall fidelity of replication (see the review by Li in this issue). Similar to NER, MMR utilizes the same set of proteins with broad substrate specificity to remove all possible mispaired or unpaired bases. As in NER, a mismatch recognition protein does not have enzymatic activity towards DNA, and its role is to recruit endo- and exo-nucleases to excise the newly synthesized daughter strand but not the template, thus removing errors of DNA replication [13, 14] (Figure 1B). The challenges for MMR are two-fold: to identify "lesions" that could be any one of the four normal nucleotides instead of damaged bases and to target repair to the daughter strand. MMR proteins are well conserved from bacteria to humans. MutS and its eukaryotic homologs MutSα and MutSβ are essential components of MMR and specialized in recognizing and binding to mispaired or unpaired bases in DNA duplexes. Crystal structures of *Taq*, *E. coli* and human MutS complexed with a variety of unpaired and mispaired bases [15-18] reveal how MutS recognizes a mismatch surrounded by random sequence. Biochemical studies of interactions between MutS and mismatch DNA and MutS with downstream effector proteins in *E. coli* and humans further illuminate the mechanism for MMR specificity [19].

By surveying literature and structural database of lesion DNA and repair protein complexes, the aim of this review is to discern how each DNA glycosylase recognizes a narrow range of slightly altered bases and how MMR and NER proteins find their wide ranges of repair targets with high specificity. Interestingly, these three repair pathways are not completely independent. Sometimes they compete for the same repair substrate. For example, a deaminated 5-methyl Cyt base paired with Gua is recognized by thymine glycosylases in BER and MutS in MMR. Occasionally, the different repair pathways work together. For example, uracil glycosylase and MutS are reported to cooperatively generate somatic hypermutation and class switching in lymphocytes [20]. We will discuss how BER, NER and MMR find DNA lesions and how they may compete and cooperate with each other.
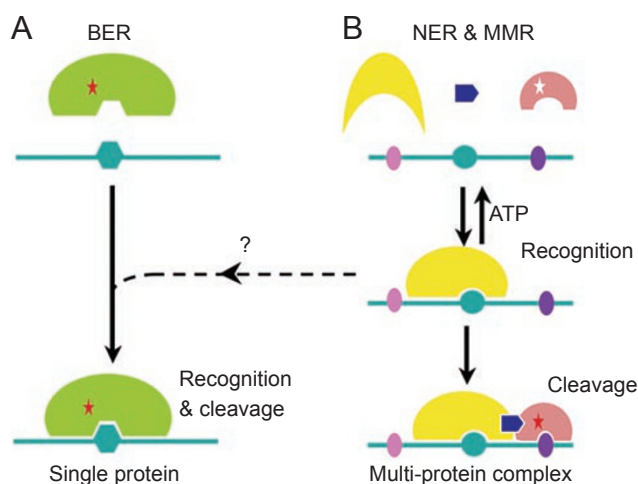

Figure 1 Diagrams of the lesion recognition step in BER, NER and MMR pathways and their potential connections. **(A)** Base modification due to oxidation, deamination or alkylation (shown as a hexagon) is recognized and excised by a glycosylase, which has an intrinsic catalytic activity (represented by the red star) and an active site specifically matches the shape and hydrogen bonding potential of the excised base. **(B)** Lesions recognized by MMR or NER are represented by an oval. Lesion-recognition proteins, e.g. MutS and UvrA (shown in yellow), do not contain any catalytic activity towards DNA, but they have an intrinsic ATPase activity for kinetic proofreading and often undergo an induced-fit conformational change upon association with lesion. Excision of lesion requires endonuclease, e.g. MutH, UvrC, XPF or XPG, shown in pink with a white star (latent activity). These endonucleases often require activation (red star) by the lesion recognition proteins and a mediator or molecular matchmaker (shown as a blue bullet), and cleave on either side of the lesion at a specific sequence (e.g. hemimethylated GATC) or at a defined distance (magenta and purple ovals).

A survey of protein-DNA base lesion complexes

Damaged base recognition by DNA glycosylases

From *E. coli* to humans, each organism has a collection of diverse glycosylases, and each glycosylase specializes in recognizing and removing one lesion, or at most a small subset of lesions [2]. They can be classified based on substrate preference (deamination, oxidation and alkylation), cleavage mechanism (mono- or bi-functional) or three-dimensional structures. Structurally, glycosylases are divided into five subfamilies: (1) monofunctional uracil/thymine glycosylase (UDG or UNG, TDG, MUG and SMUG); (2) bifunctional MutM (also known as Fpg) and Nei family specialized for oxidative damage; (3) functionally diverse HhH-GPD family, and two single member families; (4) alkyladenine glycosylase (AAG) (specialized for alkylation damage); and (5) T4 endonuclease V (for pyrimidine dimers by UV damage). These glycosylases are monomeric

and contain less than 400 residues and no more than two structural subdomains.

Interestingly, the catalytic mechanism of glycosylases does not segregate with either structural class or substrate specificity. A single lesion, e.g. GO, may be recognized by several different glycosylases, which have different structures and use different catalytic mechanisms [2]. A collection of glycosylases that repairs GO base paired either with Cyt or Ade is shown in Figure 2A. When paired with Cyt, a GO can be removed by MutM in *E. coli* and by hOGG1 (HhH-GPD family) in humans. GO prefers to be in the *syn* conformation and can form hydrogen bonds with Ade to form a Hoogsteen base pair rather than with Cyt forming the Watson-Crick pair. Ade is often misincorporated opposite GO during DNA replication. MutY or its homolog MYH, an adenine glycosylase in the HhH-GPD family, recognizes a GO/A mispair and removes the Ade specifically. When complexed with these repair enzymes,
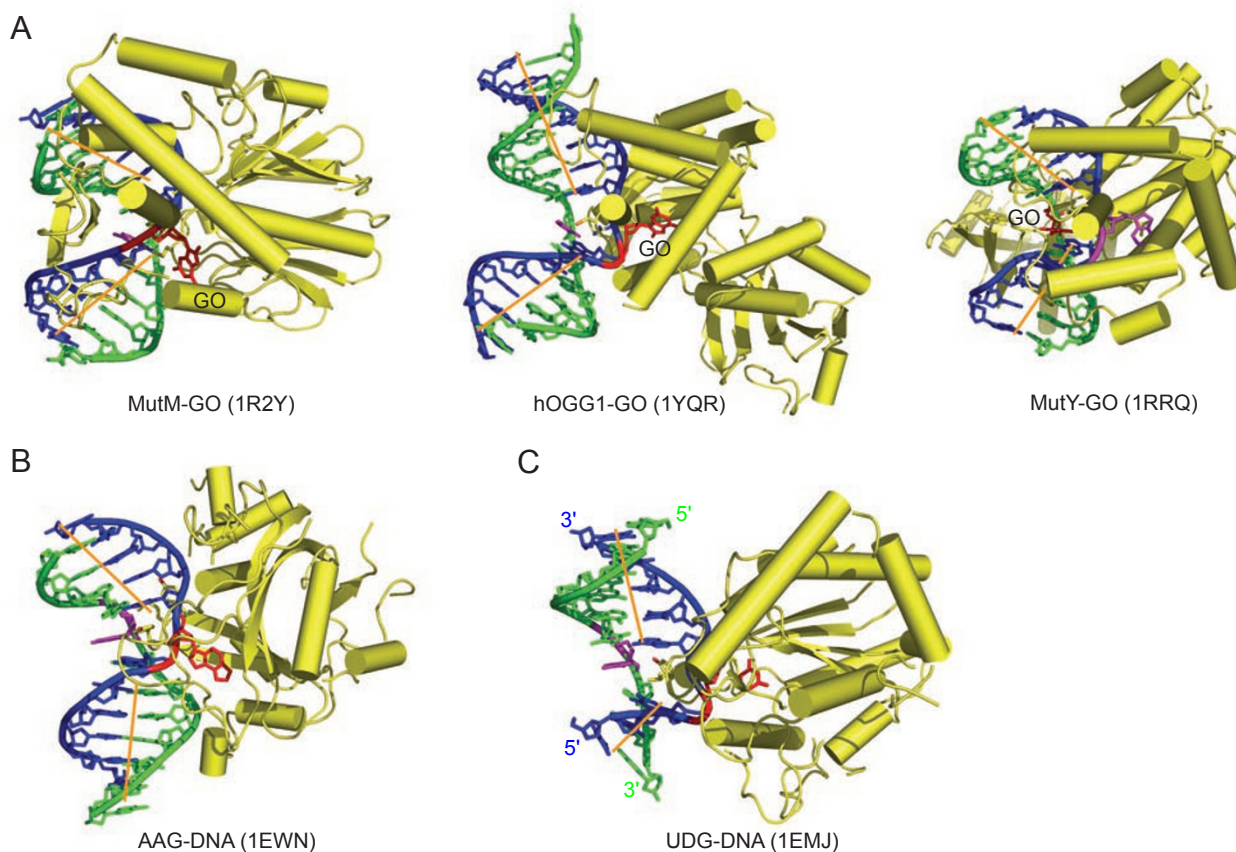

Figure 2 Nucleotide flipping observed with DNA glycosylases. **(A)** Glycosylases complexed with 8-oxo-Gua (GO): MutM (PDB: 1R2Y), hOGG1 (PDB: 1YQR) and MutY (PDB: 1RRQ). Each uses a different binding pocket for the extruded base. **(B)** Human AAG-DNA complex (PDB: 1EWN). **(C)**.UDG-DNA complex (PDB: 1EMJ). Proteins are shown as ribbons (β strands), cylinders (α helices) and coils (loops); side chains forming "reading head" are also shown. Proteins are shown in yellow, DNA strand with a lesion in blue, the complementary strand in green, lesion in red, and its base pairing partner in magenta. DNA helical axes are depicted in orange. The green strands run 5′ to 3′ from top to bottom, and the blue strands run in opposite direction. **Figures 2-7**, except for **Figure 6B**, are generated using PyMol (Pymol.sourceforge.net).

the nucleotide designated for deglycosylation, that is GO in MutM and hOGG1 and its base-pairing partner Ade in MutY, is flipped out from DNA duplexes (Figure 2A). Despite the different protein structures and nature of the extruded nucleotide in three cases, every hydrogen bonding potential of the damaged base, GO, is sampled by the repair proteins. Another common feature is that base stacking is disrupted and the DNA duplex is segmented due to nucleotide flipping accompanied by a sharp kink (MutM) or both a kink and unwinding (MutY and hOGG1) (Figure 2A).

Nucleotide flipping, DNA kinking and duplex segmentation (discontinuous base stacking) are also common features of glycosylases that repair alkylation damage [21] and UDG and MUD that repair uracil and thymine resulting from deamination of Cyt or 5-methyl-Cyt [22, 23] (Figure 2B and 2C). In the crystal structure of the UDG-uracil DNA complex, uracil is extruded from the DNA duplex and inserted into a pocket in the enzyme ready for deglycosylation. The Ade opposite this uracil is separated from its 3′ neighbor (Figure 2C). In the co-crystal structure of human AAG and 1,N(6)-ethenoadenine, the modified Ade is flipped out and its base-pairing partner Thy is unstacked with neighboring bases.

In all glycosylase-DNA complexes, damaged bases are recognized based on their shape, hydrogen-bonding potential and electric charge distribution that are different from normal bases, and an extruded nucleotide, whether normal or damaged, has to fit in the active site of a glycosylase. If a normal Gua is forced into the complex with hOGG1 and flipped out of the duplex, it does not occupy the 8-oxo-G-binding pocket and is not cleaved by the glycosylase [24]. Similarly, when a Thy is trapped into a complex with UDG, it becomes extrahelical but remains outside of the active site and is not a substrate for UDG [25]. In all cases, the void left in the DNA duplex due to nucleotide flipping out is filled by protein residues, e.g. Phe, Tyr, Leu, Ile, Met or Arg, which form a "reading head" protruding into the DNA base stack. It has been debated whether such a "reading head" actively pushes and flips a lesion nucleotide out of the double helix or passively stabilizes the kinked and segmented DNA after the base flipped out.

Lesion recognition by AP endonucleases

Crystal structures of *E. coli* Endo IV [26] and human APE1 [27], which represent the two classes of AP endonucleases, have been determined in complex with abasic analog-containing DNA (Figure 3). Although the two AP endonucleases share no structural similarity, in both complexes the abasic site is flipped out into the active site of the enzymes. As in the cases of glycosylases, the base opposite the abasic lesion is unstacked with neighboring bases, and in the case of Endo IV it is even extruded from
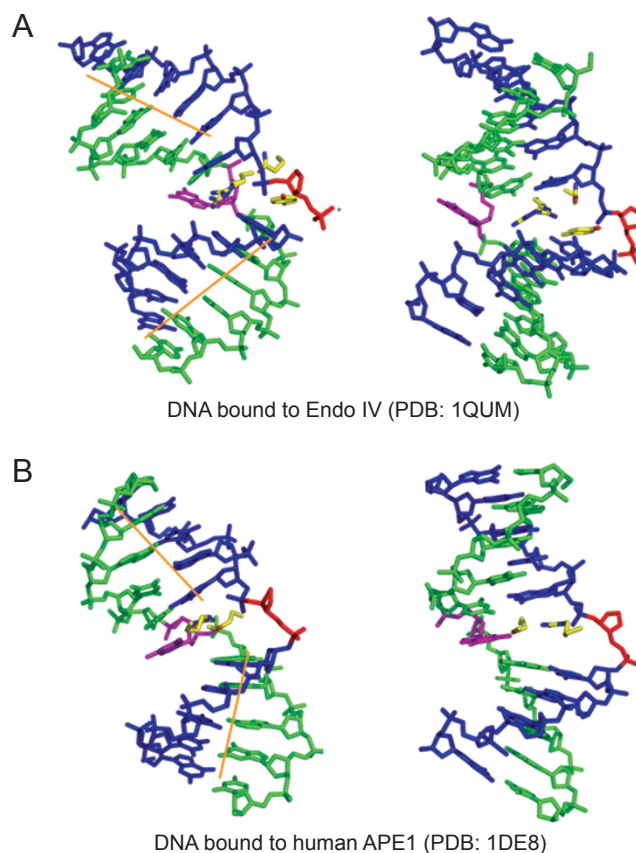
DNA bound to Endo IV (PDB: 1QUM)

DNA bound to human APE1 (PDB: 1DE8)

Figure 3 DNA kinking and nucleotide flipping observed with AP endonucleases. **(A)** Endo IV-DNA complex (PDB: 1QUM). **(B)** APE1-DNA complex (PDB: 1DE8). Proteins are omitted for clarity except for the side chains directly interacting with lesion base pairs. Each DNA is shown in two orientations.

the duplex (Figure 3A). Accompanying nucleotide flipping and base unstacking, the DNA is segmented to two parts due to severe kinking at the lesion site. The helical axes of the two DNA segments do not intersect (Figure 3), indicating that base stacking is truly discontinuous. Similar to glycosylases, AP endonucleases possess a protruding "reading head", which occupies the void in the DNA duplex.

Structures of repair proteins complexed with a cis-syn thymine dimer

Cis-syn thymine dimers, also known as cyclobutane pyrimidine dimer, CPD, are adjacent thymine bases covalently linked by C5-C5 and C6-C6 bonds upon UV irradiation. Because of the covalent bonds between adjacent pyrimidines, the bases in a CPD retain no double bond and are thus no longer planar. CPD disrupts normal base pairing and stacking, and if not repaired it can block replication and transcription. CPD can be repaired by three different mechanisms. First, in bacteria, low eukaryotes and plants,

CPDs are efficiently repaired by photolyases, which use blue light to directly break the covalent bonds between the adjacent pyrimidines and restore the native structure of thymine bases [28]. Second, the entire CPD can be excised by phage T4 enodnuclease V (Endo V), a bi-functional glycosylase, in BER [29]. Third, CPDs are generally removed by NER in bacteria, archaea and eukaryotes [7, 30].

The most recently reported crystal structure of CPD-containing DNA bound to yeast Rad4, a homolog of human XPC (xeroderma pigmentosum complementation group C), reveals that the CPD is flipped out of the DNA duplex and is not in contact with the Rad4 protein [10]. The normal nucleotides on the complementary strand are also flipped out, and they are recognized by the protein in the unpaired and unstacked form. Similarly, bacterial UvrB appears to recognize the normal strand complementary to the lesion DNA in the single-stranded form [9]. Recognition of such bulky distorting lesions appears to be achieved by strand separation and exclusion of lesions.

Similar unwinding, kinking, and nucleotide flipping are also observed in other protein-CPD complexes. In the crystal structure of Endo V-DNA complexes, the DNA is severely unwound and kinked around the CPD [11] (Figure 4A). Interestingly, the Ade opposite the 5′ thymine of the CPD is extruded, while the CPD remains intrahelical. The CPD to be cleaved by Endo V is unstacked on both 5′ and 3′ sides. The vacant space left by the extruded Ade is occupied by protein side chains (two Arg, a Gln and Pro) as in every glycosylase-DNA substrate complex (Figure 4A). In the structure of CPD-photolyase complex [12], however, the CPD is extruded from the DNA duplex and inserted into the active site of photolyase (Figure 4B). The two Ade's opposite the CPD are intrahelical, but the DNA is unstacked and severely kinked by ~60° between them (Figure 4B). Unlike Endo V and all glycosylases, the open space left by the flipped out CPD is not occupied by side chains of the photolyase. Nearby Trp, Arg and Pro residues are approximately stacked between the flipped out CPD bases and the phosphosugar backbone of the opposite Ade. These residues are located on a concave DNA-binding surface and appear to stabilize rather than induce the DNA distortion. The absence of a protruding "reading head" makes it unlikely that photolyases actively push DNA bases out. Interestingly, there is no indication that the photolyase is less capable or less efficient in finding the DNA lesions than glycosylases possessing a "reading head".

Structures of MutS and mismatched DNA complexes and the kinetic proofreading mechanism

Nearly a dozen high-resolution crystal structures of *Taq* and *E. coli* MutS-DNA complexes have been determined with unpaired or mispaired bases [15-17]. Recently crystal structures of human MutSα complexed with mismatched or damaged DNA have also become available [18]. MutS proteins are made of two polypeptide chains and form homodimers in bacteria and heterodimers in eukaryotes (see the review by Li in this issue; Figure 5A). Relative to the BER proteins, MutS is large and complex. Each MutS subunit contains at least 800 aa and five structural domains: two for DNA binding (I and IV), one with ATPase activity (V) and the remaining two forming connecting domains (II and III) (Figure 5B). Mismatched DNA is fully encircled by the four DNA-binding domains from both MutS subunits (Figure 5A), and the mispaired or unpaired base, whether Thy, Ade, Cyt or Gua, is separate from its 3′ neighboring base and stacked instead with a Phe side chain of MutS (Figure 5C). The DNA-binding domain containing the Phe is wedged in the minor groove where the mismatch resides. Concurrently the DNA duplex is kinked by 60° towards the major groove. As a result, the DNA duplex is segmented at the mismatch site.

Based on the crystal structures and biochemical studies of mismatched DNA, we propose that despite varied shapes and hydrogen bonding potentials mispaired or unpaired G, A, T and C bases share the similarity of weakened base stacking and susceptibility to kink, and that MutS recog-
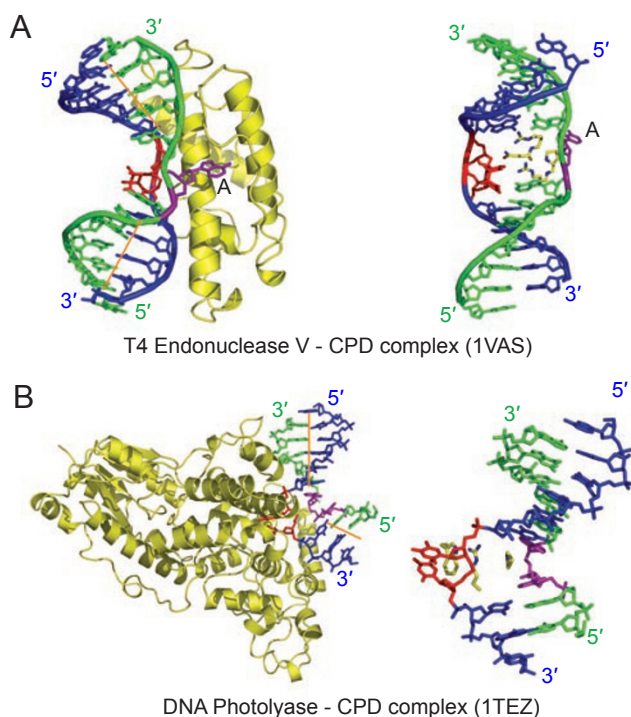

Figure 4 Structures of CPD-containing DNA and repair protein complexes. **(A)** Endo V-CPD complex (PDB: 1VAS). **(B)** Photolyase-CPD complex (PDB: 1TEZ). Each protein-DNA complex is shown in two orientations, one with protein and the other without. Proteins and DNA are shown in similar scheme as in **Figure 2**.
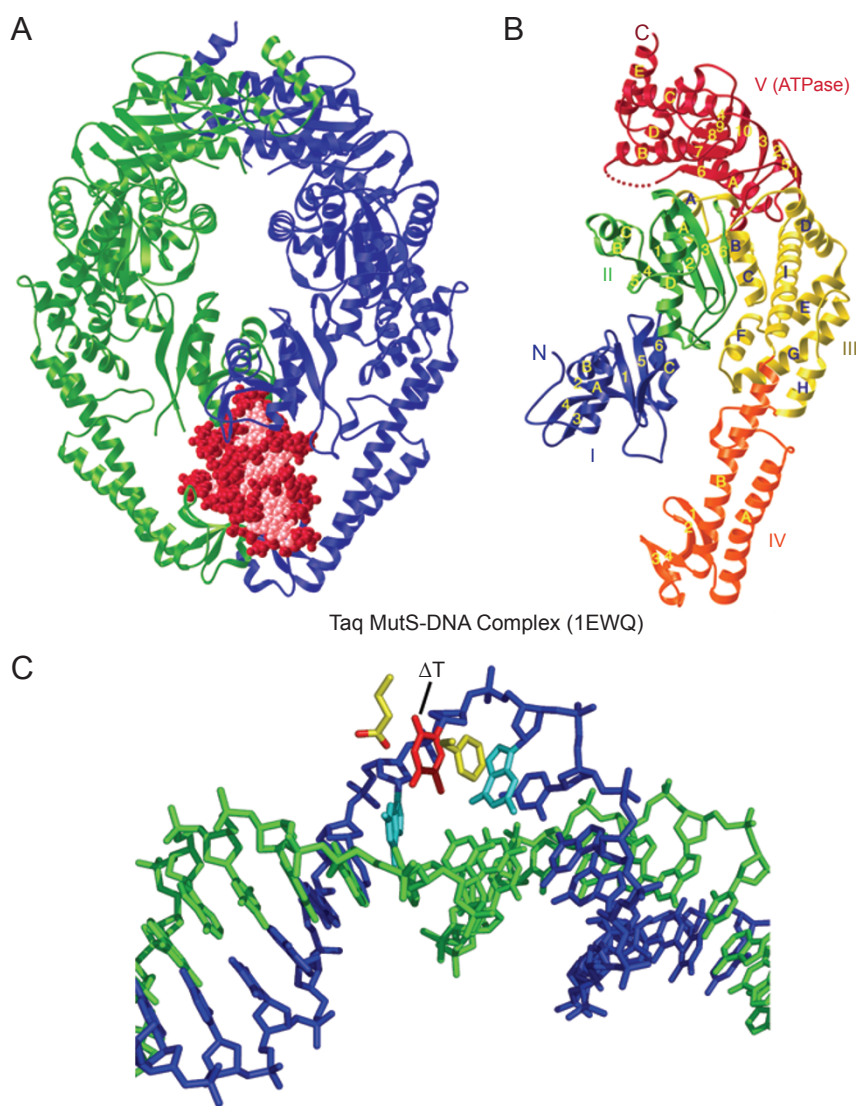
A

B

Taq MutS-DNA Complex (1EWQ)

C

Figure 5 Structure of MutS-DNA complexes. **(A)** Taq MutS-DNA complex (PDB: 1EWQ). The two protein subunits are shown in green and blue ribbon diagrams, and DNA in red and pink space-filling model. **(B)** A MutS subunit consists of five subdomains, each of which is shown in a different color. **(C)** The kinked DNA when bound to MutS. The unpaired base is shown in red (ΔT), its neighboring bases in cyan. Protein side chains (Phe and Glu) that facilitate base unstacking and DNA kinking are shown in sticks (red for oxygen atoms).

nizes a broad range of mismatches by its preferential binding to the flexible and potentially kinked DNA. However, the energetic difference between a normal and mismatched base pair in base stacking cannot be more than 2-3 kcal/mol, the total free energy contributed by a base pair to a DNA duplex [31]. This difference is translated to a difference of 100- to 1 000-fold in MutS-DNA binding constant, which agrees precisely with the binding constants of normal and heteroduplex DNAs determined experimentally [32]. Therefore, at the level of lesion recognition, MutS binds a mismatch (heteroduplex) at best 100-1 000 times better than a perfectly paired DNA (homoduplex). But, at the level of repair with excision by nucleases, the specificity of MMR must match the rate of mismatch occurrence, which is 10^{-6} to 10^{-8}. How does MMR achieve such high specificity with a broad substrate range?

As mentioned in the Introduction, the difference between

BER and MMR is that glycosylases, which have narrow substrate range but high specificity, carry out both lesion recognition and excision, while MutS has no cleavage activity towards DNA and must recruit a nuclease for mismatch removal. MutS, however, is endowed with an enzymatic activity, that is the ATPase (Figure 5B). The MutS ATPase does not directly enhance MutS-mismatch association, but it enables MutS to verify mismatch recognition by kinetic proofreading [33]. In the presence of homoduplex DNA, MutS hydrolyzes ATP quickly, but in the presence of a mismatch the burst of ATP hydrolysis is inhibited, which allows the MutS-DNA-ATP complex to form [34]. Only when associated with both ATP and a mismatch is MutS able to recruit a downstream repair protein MutL (a molecular matchmaker) and activate nucleases [33, 35].

Usage of a high-energy cofactor to increase substrate specificity of a macromolecular machine that has a broad

substrate range is exemplified in protein synthesis by ribosomes [36]. During protein synthesis a ribosome needs to accept 20 to 100 similar amino-acyl-tRNAs (aa-tRNAs), yet in each reaction cycle it has to reject all but one correct aa-tRNA determined by anti-codon and codon match. To achieve this high specificity, ribosomes use GTP as a high-energy cofactor and a special GTPase (EF-Tu) that delivers aa-tRNA to the ribosome to proofread. A matched codon and anticodon triggers quick hydrolysis of GTP and release of EF-Tu to complete the aa-tRNA delivery. A mismatched codon and anticodon inhibits GTP hydrolysis and allows wrong aa-tRNA and EF-Tu to dissociate from the ribosome [36]. The high-energy cofactor may be different (GTP or ATP), and the kinetics upon substrate recognition may even be opposite (inhibited ATP hydrolysis in MMR versus enhanced GTP hydrolysis in protein synthesis), but the effect of improving substrate specificity is the same. Kinetic proofreading may also play a key role in lesion recognition in NER.

DNA distortion during repair, transcription and replication

Base unstacking is a common feature in repair protein-DNA complexes

The survey of the available BER, MMR and CPD repair proteins complexed with substrate DNA reveals a common feature: each DNA duplex is discontinuous at the lesion due to base unstacking, severe kinking, and sometimes unpairing, unwinding and nucleotide extrusion. An obvious question is whether segmented DNA structures exist prior to association with a repair protein or they are induced by the presence of repair proteins. DNA lesions evidently destabilize and distort short linear DNAs as revealed by X-ray and NMR structural studies, decreased DNA melting temperature and molecular dynamic simulation [37, 38]. Structures of lesion-containing DNA alone obtained by X-ray crystallography or NMR, however, are invariably less distorted than when complexed with repair proteins [12, 15, 16, 18, 22, 39, 40]. This may suggest that gross distortions of DNA are induced by repair proteins, but one needs to consider that such characterization is limited by the detection methods. Crystal structures represent an ensemble of time-averaged energy-minimum states, and they do not reflect the thermodynamic nature. High-energy intermediates and unstable conformational states can easily escape detection by X-ray crystallography and NMR.

A DNA double helix is stabilized by three factors: hydration, hydrogen bonds between base pairs, and intra- and inter-strand base stacking. Energetically, base stacking is the predominant force for double helix formation [31]. Base stacking is also the predominant force leading to the rigidity of DNA and thereby the persistence length [41]. The distribution of π-electrons in a purine or pyrimidine leads to a slight negative charge at the center of the base and positive charge at the rim. Adjacent base pairs are stacked by rotating ~36° in the B-form DNA to maximize the hydrophobic and charge interactions. Wobble pairing between mismatched bases requires base displacement (a shift in the plane of base pair), which interferes with optimal base stacking. Indeed in the crystal structures of oligonucleotides containing a wobble base pair, base stacking in the neighborhood of the mismatch is perturbed and helical parameters are altered [39]. NMR studies find that a G/T mismatch causes local flexibility [42], and electrophoresis using uncrosslinked polyacrylamide gel revealed that DNA with a single mismatch migrates differently from normal DNA and resembles kinked species [43].

Poor base stacking enhances DNA binding by TBP and histone-like proteins

Normal DNAs are not uniformly stiff and have a persistence length of 450-500 Å [44]. Depending on local sequence variation DNA can bend on its own as well as be distorted by proteins [45]. For example, TA-rich sequences are prone to bend and are recognized by the transcription activator TBP (TATA binding protein) [46]. The crystal structures of TBP-TATA box complexes reveal a dramatically curved DNA (Figure 6A). TBP has the same structure before and after binding to the TATA box, and its concave surface perfectly complements the bent DNA. Compared with lesion DNA and repair protein complexes, bending of the TATA box is relatively smooth without complete disruption of base stacking (Figure 6B). Large hydrophobic residues facilitate sharp bending at two locations, but none of these protein side chains are inserted between adjacent DNA base pairs or directly stacked with a base. Interestingly, insertion of a 4-nt (unpaired) loop at the sites of TBP-mediated kink increases the affinity of TBP by more than 100-fold [47]. This clearly implicates that unpaired DNA facilitates TBP binding by inducing base unstacking and thus increasing DNA local flexibility or deformability.

IHF (integration host factor) and its homolog HU are bacterial histone-like proteins. They are highly positively charged and function in various cellular processes [48]. IHF is heterodimeric and has strong sequence preferences (often AT-rich). The crystal structures of IHF-DNA complexes show that IHF induces two near 90° kinks in a 35 bp DNA (Figure 6C). One kink occurs between a direct repeat of A/T base pairs that are conserved among all IHF-binding sites, and the other occurs at a strand break (introduced for crystallization) where base stacking is interrupted [49]. A conserved Pro from each subunit wedges between the base pairs at each kink. The consensus sequence for IHF bind-
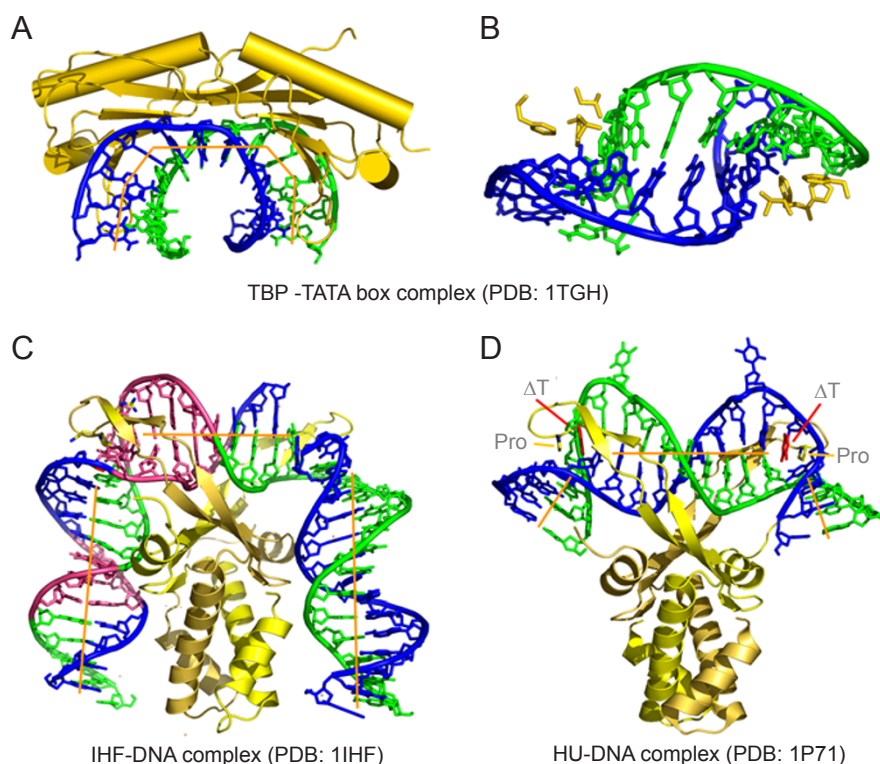
Figure 6 Structures of normal DNA distorted in protein complexes. (A) Structure of TBP-TATA box complex. The curved DNA helical axis is depicted in orange. (B) Side chains of TBP that interact with and sharply bend the DNA are shown in yellow sticks. None is inserted between base pairs. (C) The IHF-DNA complex (PDB: 1IHF). (D) The HU-DNA complex (PDB: 1P71).

TBP -TATA box complex (PDB: 1TGH)

IHF-DNA complex (PDB: 1IHF)

HU-DNA complex (PDB: 1P71)

ing consists of two segments (highlighted in pink in Figure 6C), and for both segments to contact IHF, the intervening sequence has to kink by ~90°. The conserved Pro and surrounding protein residues may induce and stabilize DNA kinking within the consensus sequence and at a second site 9 bp away in a non-specific sequence (Figure 6C) [49].

It is clear that IHF is capable of sharply kinking normal DNA, but is the IHF recognition facilitated by the intrinsic flexibility of the AT-rich sequence? The answer to the question comes from studies of Hu protein. The homodimeric HU has no sequence preference and a much weaker affinity (in the μM range) for the IHF-binding sites than IHF (in the nM range) [48]. But HU has a high affinity (nM) for locally flexible DNA that contains mismatched base pairs or unpaired loops separated by 8-9 bp [50]. Crystal structures of HU and mismatched DNA complexes reveal that two sets of unpaired nucleotides are precisely positioned at the HU-mediated kinks (Figure 6D) [48]. Similarly placed mismatches at the sites of kink also facilitate the association of IHF and DNA [50]. Based on these observations, Grove et al. [47] proposed a connection between inherent DNA flexibility due to mismatched base pairs and preferred binding by IHF, HU and related proteins.

DNA polymerases reject mismatches because of poor base stacking

Poor base stacking due to mismatch was directly ob-

served in a DNA polymerase-substrate complex when the incoming nucleotide was incorrect [51]. *Sulfolobus Solfataricus* Dpo4 is an error-prone and lesion-bypass Y-family DNA polymerase. In the crystal structure of Dpo4 complexed with an incorrect dGTP opposite a templating dT, two alternate conformations for the incoming and primer terminal nucleotides were observed [51]. When the mismatched bases formed a wobble pair, the 3′ terminal nucleotide of the primer strand was disordered and did not properly stack with the incoming dGTP (Figure 7A, left). Alternatively, when the primer terminus was ordered and base paired with the template strand, the incoming dGTP did not form hydrogen bonds with the templating dT and is more than 5 Å away from its base stacking partner (Figure 7A, right).

Unlike Y-family polymerases, which have a preformed and solvent exposed active site, replicative DNA polymerases undergo a large conformational change upon association with DNA substrate and a correct incoming nucleotide, and the active site becomes closed around the replicating base pair prior to catalysis (Figure 7B). A mismatched replicating base pair has not been crystallized with a replicative DNA polymerase for direct visualization, but it prevents the closing of the active site in *Thermus aquaticus* (*Taq*) DNA polymerase as measured by FRET experiments [52]. The previous understanding has been that a DNA polymerase discriminates against a wrong incoming nucleotide by

Figure 7 Poor base stacking at a mismatched replicating base pair observed with DNA polymerases. **(A)** Two conformations of a mismatched replicating base pair when complexed with Dpo4 (PDB: 2AGP). Dpo4 is shown as yellow ribbons, template strand is shown as green sticks, primer blue, the templating base purple and the incoming dGTP red. Blue spheres represent divalent cations. **(B)** Illustration of induced-fit finger domain movement of replicative DNA polymerases. Finger domain movement in the presence of a mismatch is not detectable by FRET.

Dpo4 complexed with mismatched replicating basepair (PDB: 2AGP)
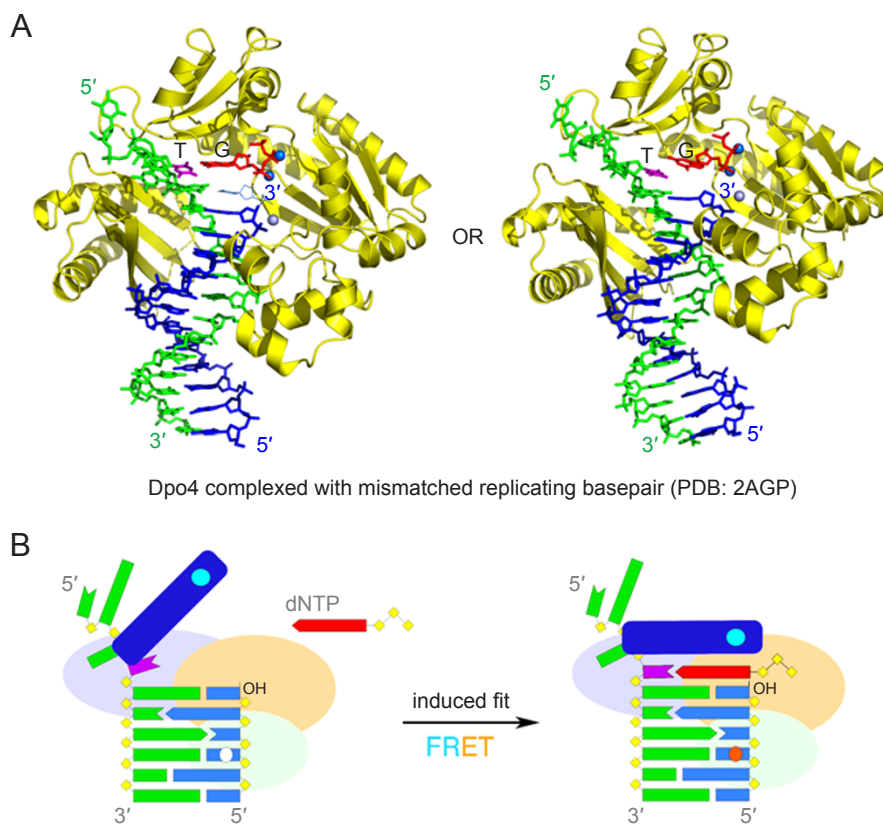
steric clashes due to its altered hydrogen bonding potential from Watson-Crick base pairs and thus altered shape [53]. To sense such shape and hydrogen bond differences, a polymerase has to be in the closed conformation and make intimate contacts with the DNA substrate. How does polymerase detect a mismatch without the conformational change? Poor base stacking between the incorrect incoming nucleotide and primer terminus as observed in the Dpo4 structures probably hinders the conformational change and thus the closure of the active site. A replicative polymerase therefore may reject a wrong incoming nucleotide in two steps: first, by the poor base stacking propensity, which prevents the necessary conformational change and, second, by the improper shape and hydrogen bonding potential, which prevents the proper alignment of metal ions and reactants in the active site and thus the chemical step of nucleotide incorporation (Figure 7B). The rate-limiting step of the catalysis appears to be the latter and not the large conformational change [52, 54].

We can envisage that a replicative high-fidelity polymerase uses the large conformational change and active site closure to sense the stability of base stacking and reject non-Watson-Crick base pairing before the chemical reaction can take place. For a translesion DNA polymerase, which is specialized for bypassing lesions, good base

stacking is often absent. In this case, a preformed active site is advantageous to accommodate a damaged template or mismatched base pair, otherwise the chemical step may never take place. For translesion DNA polymerases, the determinant for substrate selection and catalytic efficiency is the fit between the active site and substrates, and the rate-limiting step is also the alignment of the two metal ions and reactants relative to the catalytic residues and the chemistry of phosphoryl bond formation [55].

Local flexibility may recruit repair proteins to DNA lesions

Based on molecular dynamics, damaged bases likely destabilize a DNA duplex as does a mismatched base pair. For example, oxidized C8 of GO clashes with the sugar-phosphate backbone, and GO/C and GO/A are more flexible than a G/C pair and prone to extrusion [37]. As determined by X-ray crystallography, a CPD-containing DNA without protein is only gently bent at the lesion [40], but molecular dynamic simulation reveals that the absence of base stacking and reduced DNA rigidity due to crosslinking between adjacent pyrimidines should lead to local flexibility and a reduced energy barrier for distortion [38]. Local flexibility is detected even at A/U base pairs manifested by a lower melting temperature than A/T base pairs [56]. In A-form

dsRNA, which has better base stacking and a 2-3 times longer persistence length than B-form DNA [57], A/U pairs are stable and fine. The advantage of replacement of Uri with Thy in DNA, in addition to differentiation from deaminated Cyt by the methyl group, may be the increased helical stability of the B-form helix.

All BER enzymes (glycosylase and AP endonuclease) studied to date possess a "reading head" that inserts into the DNA duplex at a lesion site and a binding pocket to accommodate an extruded base (Figures 2-4A). The "reading head" and binding pocket have been interpreted to actively "interrogate" DNA by "pushing" and "pulling" a nucleotide out of the duplex, and the enzyme engages in catalysis only when an extruded base fits well in the binding pocket [24, 25]. However, if identifying a lesion were accomplished after nucleotide extrusion by examination of shape and hydrogen bonding potential of each and every nucleotide, repair would be very inefficient and its success rate would approximate the frequency of lesion occurrence, which is one repair out of millions of binding events. As discussed earlier, the presence of a lesion increases the local flexibility of DNA. The increased DNA flexibility leads to an increased frequency of the kinked and unstacked conformations favored by repair proteins and thus increased probability of protein-DNA association. Since the persistence length of normal B-form DNA is ~450 to 500 Å, increased flexibility at a lesion site allows glycosylases and AP endonucleases to skip hundreds of normal base pairs and selectively bind and examine potential lesion sites.

As demonstrated by TBP, IHF and HU proteins, increased localized flexibility due to mismatches facilitates protein-DNA association that requires DNA to kink. Local flexibility may play an essential role for lesion recognition by repair proteins that do not possess a "reading head". In the absence of a "reading head", photolyases (Figure 4) are able to stabilize the distorted structure by specific protein-CPD interactions, but the DNA is likely to be kinked around a CPD lesion prior to its capture. A lesion "reading head" is also absent in the mismatch recognition protein MutS. Its DNA-binding domains are flexible and disordered in the absence of DNA [15] and thus are unlikely to actively unstack bases or force DNA to kink. As shown by crystallographic and NMR studies, an unpaired nucleotide in a DNA duplex can be stacked in or flipped out of a DNA duplex on its own [58, 59]. The alternative conformations are in accordance with the notion of weakened base stacking and increased flexibility. MMR protein MutS may be attracted to poor base stacking and local flexibility in DNA and stabilize one out of hundreds of possible flexing DNA structures, the one that complements the shape and recognition site of MutS. Similar approaches to identify DNA lesions may be used by NER proteins [60].

Recognition of weakened base stacking first and specificity shape and feature second

Recognition of the localized flexibility is likely the first step of lesion identification. The mechanism effectively reduces non-specific binding to normal DNA by perhaps a hundred fold. But not only all lesions appear to destabilize base stacking, normal DNA also show sequence-dependent local flexibility. After initial localization to a potential lesion site, repair proteins must scrutinize the flexible joint in search of a cognate substrate. Repair proteins that have a "reading head" and a pocket for extruded base, i.e. glycosylases and AP endonucleases, can readily differentiate a cognate substrate from non-substrate. For example, Endo V has low affinity for 8-oxo-G, and vice versa MutY cannot form a stable complex with CPD or even 8-oxo-G paired with C. After initial sampling of local flexibility, repair proteins may dissociate from an improper substrate and undergo a few rounds of trial-and-error before achieving specific association with a cognate lesion.

For MMR and NER proteins, which have a broad substrate range, the danger of binding and repairing a non-cognate lesion is greater than for BER proteins. However, lesion recognition proteins in MMR and NER do not have enzymatic activity towards DNA, and therefore their non-specific association with DNA is most likely inconsequential. In particular, the MutS ATPase activity makes the association of MutS and DNA highly reversible (Figure 1B). Binding of ATP actively dissociates MutS from DNA (see the review by Li in this issue). Only in the presence of ATP and another repair protein MutL, MutS is stabilized on a mismatch [61, 62]. The ATPase activity of UvrA is also suggested to perform a similar kinetic proofreading function [6]. Such a proofreading function ensures that non-specific interactions of MMR and NER proteins due to local flexibility of DNA result in no immediate repair.

This raises the question of what else MMR and NER proteins recognize besides local flexibility. It is well known that different mismatches are repaired with different efficiency, and repair efficiency is also influenced by local DNA sequence [60, 63]. Structural, biochemical and biophysical analyses of MutS-DNA complexes provide some clues. First of all, a mismatched base is unstacked from neighboring bases but not extruded from the DNA duplex. Therefore, some flexibility and kinking are necessary, but nucleotide extrusion [64] appears to be unfavorable [65]. Since both the chemical nature of a mismatch and surrounding sequences can influence stability of base stacking, they can also influence MutS binding and repair efficiency. Secondly, mismatch recognition requires a mismatched base to stack with a conserved Phe in MutS and consequently results in a 60° kink of the DNA (Figure 5C), which leads to base shearing and unusual inter-strand interactions [15-17].

Local sequence may influence whether base shearing and inter-strand interactions are feasible. Moreover, a damaged base might not be able to stack well with the Phe just as it cannot stack well with normal DNA bases. Furthermore, it may present steric hindrance and prevent proper DNA kinking. These may be reasons why MutS normally does not bind BER or NER substrates. Determinants of NER specificity are largely unknown except for the preference for single-stranded DNA. The lack of distinct structural features of NER substrates suggests that it may take care of lesions that are not repaired by BER and MMR.

Implications for crosstalk among repair pathways

The hypothesized local flexibility-dependent lesion recognition provides a shared recognition target and thus a means for crosstalk among MMR, BER and NER. In the previous section, repair proteins are treated as non-interactive and able to independently assess the nature of a lesion until a best fit is found. Since a lesion could be initially recognized by many repair proteins, competition between different repair pathways is implicit. For example, a G/T mismatch in *E. coli* may be recognized by MMR (MutS) as well as BER (MUG, uracil/thymine glycosylase). When a G is misincorporated opposite a T by DNA polymerases, it would be disastrous for MUG to remove the template T. In a similar scenario, a GO/A mismatch can be recognized by both MutY and MutS, and removal of GO by MMR would be mutagenic. Solutions to the competition and repair of a wrong strand may be cell-cycle-dependent regulation of protein expression, post-translational modification and degradation, which may favor one repair pathway over others.

Interactions between different pathways are not necessarily competitive, and they can be complementary and cooperative. For example, uracil glycosylase and MutS appear to work at different stages of the cell-cycle downstream of cytosine deamination to facilitate somatic hypermutation and class switching in immune cells [20]. In a broad sense, there are two possible ways for repair proteins to interact cooperatively. One is by direct protein-protein interaction, and the other is by interaction mediated by DNA. For instance, binding of one protein at a flexible joint may accentuate the DNA flexibility and recruit other repair proteins. HMG proteins, which are not specialized in DNA repair but have general affinity for kinked DNA, have been found to influence and participate in MMR and NER pathways both positively and negatively [66]. One possibility is that non-specific binding of HMG to a DNA flexible joint may facilitate loading of a cognate repair protein. Occasionally binding of HMG proteins may inhibit DNA repair due to the competitive nature.

Finally, reversible binding of MMR and NER proteins to DNA lesions may give BER proteins, which have high affinity for specific lesions, an opportunity to interrogate and remove damaged bases (Figure 1, the dashed arrow). Evidence for NER proteins facilitating glycosylases has emerged in last few years. XPG, which functions mainly in the NER pathway and has high affinity for bubbled DNA and possibly low affinity for kinked DNA, greatly improves the DNA binding and catalytic activity of hNth1 (a BER glycosylase). This activation appears to be mediated by the oxidized DNA [67]. Alternatively, XPC, a lesion recognition protein in the NER pathway, directly interacts with TDG (a T/G mismatch-specific glycosylase) and facilitates product (abasic site) release from TDG [68]. NER may be the last resort for lesion repair when BER and MMR have failed.

Implications for cell-cycle signaling

Based on the above sketch of trafficking around DNA lesions, and competitive or cooperative interactions of various repair proteins, the outcome of lesion recognition, whether repair or signaling for cell-cycle arrest, may be a result of collective actions by many repair proteins. It is well established that MMR proteins, MutS and MutL, mediate alkylation and damage-induced senescence and apoptosis [69]. The hypothesis of recognition of localized flexibility by MMR proteins immediately provides a possible mechanism for DNA damage signaling. For example, when alkylation damage is beyond the repair capacity of a normal cell, lesion DNA initially sampled by MMR proteins cannot be passed onto a cognate repair protein. The abnormal persistent association by MMR proteins with the lesions may activate cell-cycle arrest directly [70] or through DNA recombination and replication [69]. In accordance with this "balancing" act, increased expression of a cognate BER protein does reduce the occurrence of cell-cycle arrest [71]. On the other hand, overexpression of MMR protein MLH1, which stabilizes MutS-mismatch DNA association, causes a mutator phenotype in yeast [72]. The high mutation rate may be due to increased association of MMR protein with DNA lesions normally repaired by the BER or NER pathways and thus inhibiting appropriate repair. A similar explanation can be applied to the dominant mutator phenotype due to a MutS mutant defective in ATP binding and thus incapable of ATP-dependent DNA dissociation [73, 74]. The proofreading-defective MutS mutants not only are defective in MMR but also may inhibit BER. It will be interesting to test whether these cell lines fail to repair certain damaged bases.

Concluding remarks

Our survey and analysis of repair protein and DNA le-

sion interactions have uncovered a common feature: the DNA double helix is discontinuous at a lesion site due to base unstacking, kinking and/or nucleotide extrusion. Studies of proteins that bind normal DNA but cause sharp bending suggest that these proteins benefit from local flexibility and their association with DNA is enhanced by the deliberate introduction of mismatched base pairs. Lesion-induced destabilization and distortion of short linear DNAs have been detected. Negatively supercoiled DNA is under-wound and presumably could augment the reduced stability caused by a lesion. A hypothesis is thus put forward that a DNA lesion weakens base stacking and shortens the persistence length of DNA, and the resulting flexible hinge is a common feature initially recognized by all repair proteins. The initial sampling of the general flexibility of DNA leads to a scrutiny of the lesion itself. If a lesion and the recognition site of a repair protein do not match perfectly, they dissociate. Thus, a single lesion can be sampled by more than one repair protein until repaired. Not surprisingly, MutS, which has a broad range of substrate specificity, actively dissociates from DNA via an ATP-dependent proofreading mechanism, thus allowing a cognate enzyme to access the lesion. This proposition immediately suggests a mechanism for crosstalk between different repair and signaling pathways. It also raises the possibility that sampling of a lesion by one protein could facilitate loading of another by direct protein-protein or DNA-mediated interactions.

Acknowledgment

References

1 Friedberg EC, Walker GC, Siede W, Wood RD, Shchultz RA, Ellenberger T. In: Friedberg EC, ed. DNA repair and mutagenesis. Wahsington, DC: ASM Press, 2005.

2 Fromme JC, Verdine GL. Base excision repair. *Adv Protein Chem* 2004; **69**:1-41.

3 Sedgwick B. Repairing DNA-methylation damage. *Nat Rev Mol Cell Biol* 2004; **5**:148-157.

4 Drablos F, Feyzi E, Aas PA, *et al*. Alkylation damage in DNA and RNA – repair mechanisms and medical significance. *DNA Repair (Amst)* 2004; **3**:1389-1407.

5 Mol CD, Hosfield DJ, Tainer JA. A basic site recognition by two apurinic/apyrimidinic endonuclease families in DNA base excision repair: the 3′ ends justify the means. *Mutat Res* 2000; **460**:211-229.

6 Sancar A, Reardon JT. Nucleotide excision repair in *E. coli* and man. *Adv Protein Chem* 2004; **69**:43-71.

7 Truglio JJ, Croteau DL, Van Houten B, Kisker C. Prokaryotic nucleotide excision repair: the UvrABC system. *Chem Rev* 2006; **106**:233-252.

8 Hanawalt PC, Ford JM, Lloyd DR. Functional characterization of global genomic DNA repair and its implications for cancer. *Mutat Res* 2003; **544**:107-114.

9 Truglio JJ, Karakas E, Rhau B, *et al*. Structural basis for DNA recognition and processing by UvrB. *Nat Struct Mol Biol* 2006; **13**:360-364.

10 Min JH, Pavletich NP. Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* 2007; **449**:570-575.

11 Vassylyev DG, Kashiwagi T, Mikami Y, *et al*. Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: structural basis for damaged DNA recognition. *Cell* 1995; **83**:773-782.

12 Mees A, Klar T, Gnau P, Hennecke U, Eker AP, Carell T, *et al*. Crystal structure of a photolyase bound to a CPD-like DNA lesion after *in situ* repair. *Science* 2004; **306**:1789-1793.

13 Iyer RR, Pluciennik A, Burdett V, Modrich PL. DNA mismatch repair: functions and mechanisms. *Chem Rev* 2006; **106**:302-323.

14 Jun SH, Kim TG, Ban C. DNA mismatch repair system. Classical and fresh roles. *FEBS J* 2006; **273**:1609-1619.

15 Obmolova G, Ban C, Hsieh P, Yang W. Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature* 2000; **407**:703-710.

16 Lamers MH, Perrakis A, Enzlin JH, Winterwerp HH, de Wind N, Sixma TK. The crystal structure of DNA mismatch repair protein MutS binding to a G × T mismatch. *Nature* 2000; **407**:711-717.

17 Natrajan G, Lamers MH, Enzlin JH, Winterwerp HH, Perrakis A, Sixma TK. Structures of *Escherichia coli* DNA mismatch repair enzyme MutS in complex with different mismatches: a common recognition mode for diverse substrates. *Nucleic Acids Res* 2003; **31**:4814-4821.

18 Warren JJ, Pohlhaus TJ, Changela A, Iyer RR, Modrich PL, Beese LS. Structure of the human MutSalpha DNA lesion recognition complex. *Mol Cell* 2007; **26**:579-592.

19 Yang W. Human MutLalpha: the jack of all trades in MMR is also an endonuclease. *DNA Repair (Amst)* 2007; **6**:135-139.

20 Rada C, Di Noia JM, Neuberger MS. Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol Cell* 2004; **16**:163-171.

21 Hollis T, Lau A, Ellenberger T. Crystallizing thoughts about DNA base excision repair. *Prog Nucleic Acid Res Mol Biol* 2001; **68**:305-314.

22 Parikh SS, Putnam CD, Tainer JA. Lessons learned from structural results on uracil-DNA glycosylase. *Mutat Res* 2000; **460**:183-199.

23 Stivers JT. Site-specific DNA damage recognition by enzyme-induced base flipping. *Prog Nucleic Acid Res Mol Biol* 2004; **77**:37-65.

24 Banerjee A, Yang W, Karplus M, Verdine GL. Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *Nature* 2005; **434**:612-618.

25 Parker JB, Bianchet MA, Krosky DJ, Friedman JI, Amzel LM, Stivers JT. Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature* 2007; **449**:433-437.

26 Hosfield DJ, Guan Y, Haas BJ, Cunningham RP, Tainer JA. Structure of the DNA repair enzyme endonuclease IV and its DNA complex: double-nucleotide flipping at abasic sites and three-metal-ion catalysis. *Cell* 1999; **98**:397-408.

27 Mol CD, Izumi T, Mitra S, Tainer JA. DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination [corrected]. *Nature* 2000; **403**:451-456.

28 Sancar A. Photolyase and cryptochrome blue-light photoreceptors. *Adv Protein Chem* 2004; **69**:73-100.

29 Lloyd RS. The initiation of DNA base excision repair of dipyrimidine photoproducts. *Prog Nucl Acid Res Mol Biol* 1999; **62**:155-175.

30 Tang J, Chu G. Xeroderma pigmentosum complementation group E and UV-damaged DNA-binding protein. *DNA Repair (Amst)* 2002; **1**:601-616.

31 Crothers DM, Zimm BH. Theory of the melting transition of synthetic polynucleotides: evaluation of the stacking free energy. *J Mol Biol* 1964; **116**:1-9.

32 Schofield MJ, Brownewell FE, Nayak S, Du C, Kool ET, Hsieh P. The Phe-X-Glu DNA binding motif of MutS. The role of hydrogen bonding in mismatch recognition. *J Biol Chem* 2001; **276**:45505-45508.

33 Junop MS, Obmolova G, Rausch K, Hsieh P, Yang W. Composite active site of an ABC ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair. *Mol Cell* 2001; **7**:1-12.

34 Antony E, Hingorani MM. Asymmetric ATP binding and hydrolysis activity of the *Thermus aquaticus* MutS dimer is key to modulation of its interactions with mismatched DNA. *Biochemistry* 2004; **43**:13115-13128.

35 Wang H, Hays JB. Signaling from DNA mispairs to mismatch-repair excision sites despite intervening blockades. *EMBO J* 2004; **23**:2126-2133.

36 Rodnina MV, Gromadski KB, Kothe U, Wieden HJ. Recognition and selection of tRNA in translation. *FEBS Lett* 2005; **579**:938-942.

37 Cheng X, Kelso C, Hornak V, de los Santos C, Grollman AP, Simmerling C. Dynamic behavior of DNA base pairs containing 8-oxoguanine. *J Am Chem Soc* 2005; **127**:13906-13918.

38 Osman R, Fuxreiter M, Luo N. Specificity of damage recognition and catalysis of DNA repair. *Comput Chem* 2000; **24**:331-339.

39 Kennard O. Structural studies of DNA fragments: the G.T wobble base pair in A, B and Z DNA; the G.A base pair in B-DNA. *J Biomol Struct Dyn* 1985; **3**:205-226.

40 Park H, Zhang K, Ren Y, *et al*. Crystal structure of a DNA decamer containing a cis-syn thymine dimer. *Proc Natl Acad Sci USA* 2002; **99**:15965-15970.

41 Mills JB, Hagerman PJ. Origin of the intrinsic rigidity of DNA. *Nucleic Acids Res* 2004; **32**:4055-4059.

42 Isaacs RJ, Rayens WS, Spielmann HP. Structural differences in the NOE-derived structure of G-T mismatched DNA relative to normal DNA are correlated with differences in (13)C relaxation-based internal dynamics. *J Mol Biol* 2002; **319**:191-207.

43 Pulyaeva H, Zakharov SF, Garner MM, Chrambach A. Detection of a single base mismatch in double-stranded DNA by electrophoresis on uncrosslinked polyacrylamide gel. *Electrophoresis* 1994; **15**:1095-1100.

44 Bednar J, Furrer P, Katritch V, Stasiak AZ, Dubochet J, Stasiak A. Determination of DNA persistence length by cryo-electron microscopy. Separation of the static and dynamic contributions to the apparent persistence length of DNA. *J Mol Biol* 1995; **254**:579-594.

45 Cloutier TE, Widom J. Spontaneous sharp bending of double-stranded DNA. *Mol Cell* 2004; **14**:355-362.

46 Parvin JD, McCormick RJ, Sharp PA, Fisher DE. Pre-bending of a promoter sequence enhances affinity for the TATA-binding factor. *Nature* 1995; **373**:724-727.

47 Grove A, Galeone A, Yu E, Mayol L, Geiduschek EP. Affinity, stability and polarity of binding of the TATA binding protein governed by flexure at the TATA Box. *J Mol Biol* 1998; **282**:731-739.

48 Swinger KK, Rice PA. IHF and HU: flexible architects of bent DNA. *Curr Opin Struct Biol* 2004; **14**:28-35.

49 Rice PA, Yang S, Mizuuchi K, Nash HA. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* 1996; **87**:1295-1306.

50 Grove A, Galeone A, Mayol L, Geiduschek EP. Localized DNA flexibility contributes to target site selection by DNA-bending proteins. *J Mol Biol* 1996; **260**:120-125.

51 Vaisman A, Ling H, Woodgate R, Yang W. Fidelity of Dpo4: effect of metal ions, nucleotide selection and pyrophosphorolysis. *EMBO J* 2005; **24**:2957-2967.

52 Rothwell PJ, Mitaksov V, Waksman G. Motions of the fingers subdomain of klentaq1 are fast and not rate limiting: implications for the molecular basis of fidelity in DNA polymerases. *Mol Cell* 2005; **19**:345-355.

53 Doublie S, Sawaya MR, Ellenberger T. An open and closed case for all polymerases. *Structure Fold Des* 1999; **7**:R31-R35.

54 Zhang H, Cao W, Zakharova E, Konigsberg W, De La Cruz EM. Fluorescence of 2-aminopurine reveals rapid conformational changes in the RB69 DNA polymerase-primer/template complexes upon binding and incorporation of matched deoxynucleoside triphosphates. *Nucleic Acids Res* 2007; **35**:6052-6062..

55 Yang W, Woodgate R. What a difference a decade makes: Insights into translesion DNA synthesis. *Proc Natl Acad Sci USA* 2007; **104**:15591-15598.

56 Howard FB. The stabilizing contribution of thymine in duplexes of (dA)24 with (dU)24, (dT)24, (dU12-dT12), (dU-dT)12, (dU2-dT2)6, or (dU3-dT3)4: nearest neighbor and next-nearest neighbor effects. *Biopolymers* 2005; **78**:221-229.

57 Hagerman PJ. Flexibility of RNA. *Annu Rev Biophys Biomol Struct* 1997; **26**:139-156.

58 Joshua-Tor L, Frolow F, Appella E, Hope H, Rabinovich D, Sussman JL. Three-dimensional structures of bulge-containing DNA fragments. *J Mol Biol* 1992; **225**:397-431.

59 Kalnik MW, Norman DG, Zagorski MG, Swann PF, Patel DJ. Conformational transitions in cytidine bulge-containing deoxytridecanucleotide duplexes: extra cytidine equilibrates between looped out (low temperature) and stacked (elevated temperature) conformations in solution. *Biochemistry* 1989; **28**:294-303.

60 Isaacs RJ, Spielmann HP. A model for initial DNA lesion recognition by NER and MMR based on local conformational flexibility. *DNA Repair (Amst)* 2004; **3**:455-464.

61 Grilley M, Welsh KM, Su SS, Modrich P. Isolation and characterization of the *Escherichia coli* mutL gene product. *J Biol Chem* 1989; **264**:1000-1004.

62 Habraken Y, Sung P, Prakash L, Prakash S. ATP-dependent assembly of a ternary complex consisting of a DNA mismatch and the yeast MSH2-MSH6 and MLH1-PMS1 protein complexes. *J Biol Chem* 1998; **273**:9837-9841.

63 Kramer B, Kramer W, Fritz HJ. Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of *E. coli*. *Cell* 1984; **38**:879-

887.

64 Mariappan SV, Silks LA III, Bradbury EM, Gupta G. Fragile X DNA triplet repeats, (GCC)n, form hairpins with single hydrogen-bonded cytosine.cytosine mispairs at the CpG sites: isotope-edited nuclear magnetic resonance spectroscopy on (GCC)n with selective 15N4-labeled cytosine bases. *J Mol Biol* 1998; **283**:111-120.

65 Su SS, Modrich P. *Escherichia coli* mutS-encoded protein binds to mismatched DNA base pairs. *Proc Natl Acad Sci USA* 1986; **83**:5057-5061.

66 Reeves R, Adair JE. Role of high mobility group (HMG) chromatin proteins in DNA repair. *DNA Repair (Amst)* 2005; **4**:926-938.

67 Klungland A, Hoss M, Gunz D, Constantinou A, Clarkson SG, Doetsch PW, *et al*. Base excision repair of oxidative DNA damage activated by XPG protein. *Mol Cell* 1999; **3**:33-42.

68 Shimizu Y, Iwai S, Hanaoka F, Sugasawa K. Xeroderma pigmentosum group C protein interacts physically and functionally with thymine DNA glycosylase. *EMBO J* 2003; **22**:164-173.

69 Stojic L, Brun R, Jiricny J. Mismatch repair and DNA damage signalling. *DNA Repair (Amst)* 2004; **3**:1091-101.

70 Wang Y, Qin J. MSH2 and ATR form a signaling module and regulate two branches of the damage response to DNA methylation. *Proc Natl Acad Sci USA* 2003; **100**:15387-15392.

71 Calleja F, Jansen JG, Vrieling H, Laval F, van Zeeland AA. Modulation of the toxic and mutagenic effects induced by methyl methanesulfonate in Chinese hamster ovary cells by overexpression of the rat *N*-alkylpurine-DNA glycosylase. *Mutat Res* 1999; **425**:185-194.

72 Shcherbakova PV, Kunkel TA. Mutator phenotypes conferred by MLH1 overexpression and by heterozygosity for mlh1 mutations. *Mol Cell Biol* 1999; **19**:3177-3183.

73 Junop MS, Yang W, Funchain P, Clendenin W, Miller JH. *In vitro* and *in vivo* studies of MutS, MutL and MutH mutants: correlation of mismatch repair and DNA recombination. *DNA Repair (Amst)* 2003; **2**:387-405.

74 Drotschmann K, Yang W, Kunkel TA. Evidence for sequential action of two ATPase active sites in yeast Msh2-Msh6. *DNA Repair (Amst)* 2002; **1**:743-753.