DNA-repair enzymes
Dmitry G Vassylyev* and Kosuke Morikawa†

Recent crystallographic studies of DNA-repair enzymes have provided the structural basis for the recognition of damaged DNA. The results imply that flipping out of the base is a common and crucial event in DNA repair. Two classes of repair enzymes that recognize distinct types of damage may exist. DNA-repair enzymes that share similar folds and DNA binding motifs have been proposed to belong to a superfamly.

Introduction
Genetic information preserved in DNA is used by all living organisms to ensure their normal function and developement. In contrast with globular proteins, DNA molecules are extremely thin and long, so that even the bases are exposed to the solvent. DNA, therefore, is easily modified and damaged by various chemical agents and different types of radiation to yield mutagenic lesions [1]. Mutagenesis of DNA also results in the spontaneous formation of mismatched base pairs in the genome during replication [1]. If not repaired, these DNA lesions disturb the genomic integrity and could consequently become lethal to living cells. All organisms, therefore, contain various repair mechanisms that remove the improper nucleotides from DNA and replace them with the correct ones.

DNA-repair enzymes recognize and excise the damaged portions of the DNA strand. This major step is carried out by a large variety of enzymes or multienzyme complexes, which can be classified according to the different lesions and the mechanisms of repair. In the base excision repair (BER) pathway [1], DNA glycosylases directly excise the modified-base moieties from the DNA to produce apurinic-apyrimidinic (AP) sites, and the phosphate backbone at the abasic site is subsequently cleaved by the actions of AP-endonucleases. Enzymes involved in the direct reversal (DR) pathway [1] catalyze the reverse reaction to the one that originally caused the base modification; thus, excision is not required. Nucleotide excision repair (NER) is the most important and universal DNA-repair mechanism in living organisms [1,2], and it allows the excision of almost all kinds of DNA lesions relevant to chemical modification. NER removes a segment of the DNA strand including the damaged base and several nucleotides on either side of the lesion. Enzymes involved in the mismatch repair (MMR) systems have a similar mechanism of action to that of the BER and NER systems [3,4]. The major distinction of MMR from the other repair systems is that it can discriminate an improper base between two normal nucleotides forming a mismatched base pair.

DNA-repair systems are highly conserved among different species, ranging from prokaryotes to eukaryotes. Malfunctional DNA-repair mechanisms often cause cancer and, in most cases, are lethal to organisms [2,3]. Thus, the structural study of DNA repair is one of the most important areas, aiding cell biology and the medical sciences. Successful crystallographic studies of several DNA-repair enzymes in the past two years have greatly contributed to the understanding of how damaged DNA is recognized and excised at the molecular level. It would be impossible to describe all of the aspects of DNA repair available from recent structural studies; therefore, this review focuses on the mechanism by which repair enzymes bind specifically to damaged DNA and comments on the general view of damage recognition.

Endonuclease V
Endonuclease (endo) V from bacteriophage T4 has high affinity for both cyclobutane-type pyrimidine dimers (PDs) and AP sites within DNA and possesses two distinct catalytic activities, PD N-glycosylase and AP-lyase [5]. The crystal structure of this enzyme (from the inactive mutant, Glu23→Gln) in complex with a DNA duplex containing a PD reveals a strikingly deformed DNA conformation [6**, although the protein conformation remains unchanged (Fig. 1) [7]. The DNA in the complex kinks sharply at the position of the PD. The adenine base complementary to the 5′ moiety of the PD is completely
"flipped out" from the DNA; it is accommodated into a cavity on the surface of the enzyme by weak and nonspecific van der Waals interactions. The flipping out of the base opens up free space inside the DNA for the catalytic residues of the enzyme to access their target. Thus, this extrahelical base seems to be required for facilitation of the enzymatic reaction, rather than for specific damage recognition. The other surprising feature of the structure is that specific binding to the PD occurs through an 'indirect readout' mechanism. Endo V makes no direct interactions with the modified moiety of the PD bases. Instead, the deformed DNA backbone in the close vicinity of the lesion is recognized using a number of basic protein residues.

Figure 1

Ribbon model of T4 endo V in complex with a 13-mer oligonucleotide duplex containing a PD (α helices and loop regions are shown in red and green, respectively). The trace of the phosphate backbone is shown in cyan, the DNA bases are shown in yellow, and the thymine dimer and the flipped-out base are highlighted in white. Figure generated using the program O [48].

We believe that the unusual DNA conformation observed in the endo V-DNA complex is induced by the enzyme upon binding, as the enzyme-free DNA duplexes containing a PD do not have such striking deformations of the DNA [8-11].

Endonuclease III

Endo III from Escherichia coli has a very broad substrate specificity for various oxidized pyrimidine derivatives within the DNA [12-17]. It possesses N-glycosylase and AP-lyase activities similar to those of endo V [18]. In combination with mutational analyses, the 1.8 Å resolution structure of this enzyme [19**] revealed not only the catalytic site and the DNA-binding interface, but also two novel DNA-binding motifs (Fig. 2). The protruding [4Fe-4S] cluster loop (FCL) and the helix-hairpin-helix (HhH) motif, previously identified as the binding site of thymine glycol [20], were shown to participate in DNA binding. The HhH motif contains a putative catalytic residue, Lys120, which lies near the deep cleft formed between two domains. Another catalytic residue, Asp138, is completely buried within this cleft. Thus, the global view of the binding cavity suggests that the damaged base should be extrahelical upon the binding of endo III to the damaged DNA duplexes.

Figure 2

Overview of the endo III structure. The two major domains of the enzyme are shown in green and red, the HhH motif is shown in yellow, and the FCL loop and [4Fe-4S] cluster are shown in magenta and cyan, respectively. Figure generated using the program O [48].

3-Methyladenine DNA glycosylase II (AIkA)

3-Methyladenine DNA glycosylase II (AIkA) is an E. coli monofunctional enzyme, that exhibits only N-glycosylase activity and broad substrate specificity for alkylated DNA [21–26]. Crystallographic studies of this enzyme [27**•,28**•] show that it folds into a compact globular structure, consisting of three domains that have approximately equal sizes (Fig. 3). Domain I is topologically identical to
one half of the DNA-binding domain of the TATA box binding protein (TBP) [29]. The folds and the arrangements of domains II and III have high structural similarity to endo III, although no significant homology in the sequence is present. The endo-III-like domains of AlkA are directly involved in DNA binding and catalysis, whereas the biological function of the TBP-like domain is not clear. AlkA lacks the [4Fe-4S] cluster and therefore the FCL DNA-binding motif that was found in endo III. The HhH motif, however, is highly conserved both in conformation and in sequence across the two proteins. There is no lysine residue which is crucial for the AP-lyase activity in endo III inside the HhH motif of AlkA, which implies that the HhH motif is more important for the recognition and binding of damaged DNA than for catalysis. In AlkA, a catalytic residue, Asp238, projects into the cleft between the two endo-III-type domains, as Asp138 in endo III does. This also suggests that the damaged base would be flipped out of the DNA duplex upon binding to the protein.

**Endonuclease III superfamily**

The structure of the 26kDa catalytic domain of the MutY protein, the DNA-glycosylase which possesses a BER mechanism against adenine bases in A–G mispairs (MMR) [30,31], has recently been solved (John Tainer, personal communication) and was found to be very similar to that of endo III, including the FCL and HhH motifs. A high level of structural similarity is observed between endo III and AlkA, whereas the sequence homology is limited to only the HhH motif and the adjacent active-site loop. Thus, the presence of this structural homology in a number of other BER enzymes (e.g. yeast 8-oxoguanine DNA glycosylase [Ogg1], PD specific endonuclease from *Micrococcus luteus*) provides evidence for the existence of an endo III superfamily, which adopts a common fold but which has extremely broad substrate specificity [32,33]. Notably, the HhH motif was also observed in the NER enzymes [34], which highlights its general significance for recognition of damaged DNA.

**Uracil-DNA glycosylase**

Uracil-DNA glycosylase (UDG) has a N-glycosylase activity that is highly specific to the uracil base [35] and its analogous derivative, hydroxy-2'-deoxyuridine [17] within the DNA. Two crystal structures of UDG, herpes simplex virus type-1 (HSV-1) and human (h), have recently been reported with respect to both the free-enzyme form and the enzyme in complex with either a trinucleotide and uracil (for HSV-1UDG) or 6-aminouracil (for hUDG) [36,37]. In agreement with the 39% sequence similarity between the two enzymes, both structures have a similar fold and have a remarkably conserved 3D arrangement of the uracil-binding pocket. As the uracil base is completely buried inside the pocket, it must to flip out from DNA duplexes during complex formation. The specific hydrogen bond network between the protein and the uracil allows UDG to discriminate uracil from structurally similar cytosine, while the steric collision with protein residues prevents binding of the thymine to UDG.

The crystallographic studies of UDG were expanded to include structural determinations of HSV-1UDG and hUDG in complex with a protein inhibitor, UGI [38,39]. These structures revealed that the negatively charged residues of UGI bind to a positively charged DNA-binding groove in UDG and mimic the interactions between UDG and the phosphate backbone of the DNA. Leu272, in a protruding loop of hUDG, is packed into the hydrophobic pocket at the center of the negatively charged surface of UGI. Leu272 was proposed to penetrate into the dsDNA at the position of the uracil base, thereby facilitating the flipping-out of the base.
Very recently, John Tainer and coworkers [40**] have succeeded in solving the structure of a hUDG–dsDNA complex using a double mutant, Leu272→Arg/Asp145→Asn, of hUDG. The DNA in this complex essentially adopts the normal B-form (15° bend) with the local strand separation on the 5'-side to the uracil and the uracil base flipped-out (Fig. 4). The DNA conformation is very similar to that observed in the Hhal methyltransferase–DNA complex [41]. The sidechain of Arg272 is inserted into the abasic site in place of uracil, which is in good agreement with the role of Leu272 predicted for the wild-type enzyme. The nonspecific interactions between the protein and the DNA phosphate backbone are found within a very limited region in one strand only around the lesion. This observation suggests how UDG may bind to both dsDNA and ssDNA in a similar manner.

**DNA photolyase**

DNA photolyase (PHL) is a unique DNA-repair enzyme with a narrow spectrum of specificity for the cyclobutane-type PD. In contrast to endo V, which repairs the PD using the BER mechanism, PHL uses the DR pathway, reverting the PD to normal pyrimidines by splitting the cyclobutane ring using photoreactivation [42].

The molecular structure of PHL from *E. coli* in complex with two organic cofactors has been solved [43**]. PHL consists of two major (N- and C-terminal) domains. The PD-binding site is located in a hole, which lies at the center of the C-terminal domain in close proximity to the catalytic cofactor. The proposed mode of PD binding, in which the lesion is accommodated deeply within this hole,
provides evidence that the PD must be flipped out of the dsDNA.

**Exonuclease III**

Exo III is a major AP endonuclease in *E. coli*, and it possesses a number of additional catalytic activities against undamaged DNA and RNA substrates [44]. In terms of DNA repair, however, exo III is only highly specific to AP sites in the DNA. The enzymatic activities of exo III require a single divalent-metal ion. It was not surprising, therefore, that the 3D structure of this enzyme [45] revealed a fold similar to that found in a number of other metal-dependent DNA-processing enzymes (e.g. *E. coli* RuvC resolvase, RNase H, HIV-1, integrase) [46]. The crystal structure of exo III in complex with Mn$^{2+}$ and dCMP determined the active site of the enzyme (Fig. 5). The arrangement of the residues in the vicinity of the active site raises the possibility that the DNA base opposite the AP site may be flipped out of the double helix and bound to the enzyme, in a manner similar to that observed in the endo V–DNA complex.

**Conclusions**

The crystal structures of endo V in complex with a dsDNA substrate and the UDG–dsDNA complex revealed three possible steps of damage recognition: first, protein-induced DNA kinking around a lesion; second, indirect readout of the target damage via the deformed phosphate backbone; and third, flipping out the base from the DNA. The first two steps are inherent only in endo V and seem to be most important for the enzyme to distinguish damaged DNA from normal DNA. The third step is important in the mechanisms of both enzymes. The flipping out of the base and its subsequent accommodation into a highly specific cavity on the protein surface seem to be the major factors used by UDG to discriminate damage. Endo V probably uses the flipping of the base to achieve effective catalysis, rather than to enhance specific recognition.

Why do endo V and UDG, both DNA N-glycosylases, recognize their substrates buried within the DNA in such different manners? One reason may be the difference in the specificities of these enzymes. Endo V recognizes both PD and AP sites with almost the same affinity, although the two types of damage are very different in terms of their 3D architectures. Moreover, the different violations of the intact DNA double helix associated with these lesions are likely to provide two distinct local conformations of DNA in the absence of the enzymes. Thus, in principle, endo V should have two different modes for DNA recognition and binding for the modified bases. For endo V, the induced-fit conformational change of the DNA, once bound to the enzyme, and the indirect readout of the lesion form a very elegant solution to this problem. In contrast, UDG has a very specific affinity for uracil bases only, and, like many other sequence-specific DNA-binding proteins, this enzyme essentially recognizes B-form DNA. The outward resemblance of uracil to other bases, however, requires it to be flipped out to enable accurate identification by the enzyme.

These two distinct recognition mechanisms may be used to classify other DNA-repair enzymes with high and broad substrate specificities into two groups (Table 1). It is worthwhile mentioning that all of the enzymes containing the HhH motif belong to the endo V class (Table 1), and hence damage recognition should be associated with enzyme-induced DNA deformations. Thus, the HhH motif may have a role as a universal 'DNA deformation' motif for DNA-repair enzymes that have broad specificity.

As DNA-repair enzymes may induce substantial deformations of the cognate DNA, structural studies of the protein–DNA complexes will be crucial for a full understanding of DNA repair mechanisms. The design of substrate analogs for DNA repair [47] should facilitate successful crystallization of the enzyme–DNA complexes.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Mechanism</th>
<th>Target DNA</th>
<th>Specificity</th>
<th>HhH motif</th>
<th>Possible recognition event</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endo V</td>
<td>BER</td>
<td>ds</td>
<td>Broad</td>
<td>No</td>
<td>Yes</td>
<td>Endo V</td>
</tr>
<tr>
<td>Endo III</td>
<td>BER</td>
<td>ds</td>
<td>Broad</td>
<td>Yes</td>
<td>Yes</td>
<td>Endo V</td>
</tr>
<tr>
<td>AlkA</td>
<td>BER</td>
<td>ds</td>
<td>Broad</td>
<td>Yes</td>
<td>Yes</td>
<td>Endo V</td>
</tr>
<tr>
<td>MutY</td>
<td>BER/MMR</td>
<td>ds</td>
<td>Broad</td>
<td>Yes</td>
<td>Yes</td>
<td>Endo V</td>
</tr>
<tr>
<td>UvABC</td>
<td>NER</td>
<td>ds</td>
<td>Broad</td>
<td>Yes</td>
<td>Yes</td>
<td>Endo V</td>
</tr>
<tr>
<td>ERCC1</td>
<td>NER</td>
<td>ds</td>
<td>Broad</td>
<td>Yes</td>
<td>Yes</td>
<td>Endo V</td>
</tr>
<tr>
<td>MutHLS</td>
<td>MMR</td>
<td>ds</td>
<td>Broad</td>
<td>Yes</td>
<td>Yes</td>
<td>Endo V</td>
</tr>
<tr>
<td>UDG</td>
<td>BER</td>
<td>ds/ss</td>
<td>High</td>
<td>No</td>
<td>No</td>
<td>UDG</td>
</tr>
<tr>
<td>PHL</td>
<td>DR</td>
<td>ds/ss</td>
<td>High</td>
<td>No</td>
<td>No</td>
<td>UDG</td>
</tr>
<tr>
<td>ADA*</td>
<td>DR</td>
<td>ds</td>
<td>High</td>
<td>No</td>
<td>No</td>
<td>UDG</td>
</tr>
<tr>
<td>Exo III</td>
<td>BER</td>
<td>ds</td>
<td>High</td>
<td>No</td>
<td>No</td>
<td>UDG</td>
</tr>
</tbody>
</table>

*O$^{6}$-methylguanine-DNA methyltransferase [49].
Acknowledgements

We are most grateful to Cliff Mol and John Tainer for providing us with the endo III coordinates and the exo III figure. Yuriko Yamagata for providing the coordinates of AlkA. We also thank Susan Tsutakawa for helpful discussions during the preparation of this manuscript.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest


The crystal structure of an inactive mutant of T4 endo V in complex with a cognate DNA substrate refined to 2.75 Å resolution is reported. Implications for the mechanisms of damage recognition and catalytic reaction at the molecular level are discussed.


The yeast DNA glycosylase (Ogg1) specific to 8-oxoguanine is suggested to be closely related in 3D structure to endo IlL. The sequence motif shared between Ogg1, endo III, and AIkA was found in a number of other enzymes. This finding led to the proposal of a structural superfamily of DNA repair enzymes.


The high-resolution crystal structures of UDG from herpes simplex virus, both in its free form and in complexes with a trinucleotide and uracil, reveal a new protein fold and allow the authors to propose mechanisms of DNA binding and catalysis.


The crystal structures of the free form of glycosylase and its complex with 6-aminouracil, in combination with a mutational analysis, reveal the uracil-binding site, the catalytic residues, and the positively charged groove that is involved in DNA binding. The mechanism of the catalytic reaction is proposed.


The atomic structure of the complex of hUDG with a protein inhibitor, UGI, has been refined to 1.9 Å resolution. UGI was shown to bind to the DNA-binding loop of the hUDG, and to form no contacts with the uracil-binding pocket. A specific role in DNA binding is proposed for Leu272 of UDG.


The crystal structure of HSV-1UDG in complex with a protein inhibitor (UGI) shows that UGI directly mimics the enzyme–DNA interactions observed in the HSV-1UDG–trinucleotide complex. UGI is shown to block access to the active site of the enzyme but to form no contacts within the uracil-binding pocket.


The structure of hUDG(Leu272→Arg/Asp145→Asn)–DNA directly reveals the DNA-binding mechanism of hUDG and its cognate DNA substrate. The mutation Leu272→Arg abolishes the release of the product DNA from the protein but does not affect the glycosylase activity of the enzyme, so that the uracil in the complex is cleaved from the sugar moiety.


The high-resolution (2.3 Å) crystal structure of DNA photolyase complexed with its light-harvesting, MTHF, and catalytic, FAD, cofactors is reported. The binding site for PD is determined and a mechanism of photoreactivation through electron transfer between FAD and PD is proposed.


The atomic structure of exo III has been refined to 1.7 Å resolution. It exhibits a twofold symmetric fold similar to that of RNase H. The ternary complex of the enzyme with Mn2+ and dCMP leads to proposals of enzymatic and DNA-binding mechanisms.

