Functions of base flipping in E. coli nucleotide excision repair

ARTICLE *in* DNA REPAIR · OCTOBER 2008

Impact Factor: 3.11 · DOI: 10.1016/j.dnarep.2008.06.011 · Source: PubMed

6 AUTHORS, INCLUDING:

Gijs A van der Marel
Leiden University
882 PUBLICATIONS **22,928** CITATIONS
SEE PROFILE

Nora Goosen
Leiden University
89 PUBLICATIONS **2,983** CITATIONS
SEE PROFILE

Chapter 1

Introduction

1.1 General introduction

DNA was first isolated in the year 1869 by the Swiss physician Friedrich Miescher, who obtained it in large quantities from the pus of infections (Dahm *et al.* 2005). Since he isolated the substance from the nucleus he called it nuclein. It took until the next century however before its structure was finally solved by James D. Watson and Francis Crick (Watson and Crick*,* 1953) and it was recognized that this macro-molecule is the carrier of all our genetic information. At that time it was believed that DNA is a stable structure in itself but now we know that DNA is susceptible to various kinds of chemical alterations. Every day around 18000 purines are lost from each cell by hydrolysis of the bond connecting the purine to the ribose unit of the DNA helix and 100-500 times per day a cytosine residue gets deaminated and turns into a uracil residue (Friedberg, 2006). Both of these modifications are potentially disastrous to the cell and have to be restored.

In addition to these spontaneous alterations, nature produces various agents that have the property to alter our DNA. These agents can be divided into two distinct classes; either they are endogenous or exogenous. The group of endogenous agents comprises Reactive Oxygen Species (ROS) and other reactive agents caused by metabolic reactions such as lipid peroxidation. Exogenous agents are numerous and can be subdivided in electromagnetic radiation and chemicals.

If left unrepaired these modifications might lead to cell death by blocking vital processes like replication and/or transcription. Another possibility is that these DNA lesions can cause mutations by changing the geometries and/or base pairing properties of the base. Consequently the DNA polymerase might mistakenly recognize it for another base and incorporate the wrong nucleotide in the new DNA strand. After another round of replication this then leads to a dsDNA molecule carrying a mutation at the site where the lesion was located.

The accumulation of damage within cells, especially double strand breaks and base lesions blocking replication and transcription, are known to evoke an intricate DNA damage response, which triggers several cellular pathways to preserve the cell's integrity. These pathways include cell cycle checkpoint control (in eukaryotes), which slows down cell cycle progression in order to give the cell time to repair the damage, a transcription response (i.e. the prokaryotic SOS response) leading to expression of numerous genes, DNA repair pathways and chromatin remodeling pathways. In case a cell experiences too much DNA damage to efficiently restore its genetic information the cell might sacrifice itself by inducing a programmed cell death called apoptosis to preserve the integrity of the organism. Another possibility is that the cell goes into senescence; an irreversible state where the cell no longer divides and can therefore no longer harm the organism.

As pointed out, DNA repair is one of the major mechanisms that counteract the deleterious effects of DNA damage. Since DNA exists in a double helical form and more than one copy of this information carrier is present in the cell DNA repair mechanisms

are able to restore the DNA to its original form and sequence after DNA damage has been inflicted. This allows DNA repair systems to obtain all the necessary sequence information from the undamaged strand or its additional copy. Important DNA repair mechanisms are: direct reversal (DR), base excision repair (BER), mismatch repair (MMR), double-strand break repair (DSBR) and nucleotide excision repair (NER). However in order for these repair mechanisms to access the formed DNA damages at sites of dense chromatin structure specific chromatin remodeling mechanisms have evolved to open up the DNA structure.

In addition to these different DNA repair mechanisms cells have also devised mechanisms enabling it to just tolerate DNA damage without repairing it. Examples of these responses are postreplicative gap filling, and translesion synthesis (TLS). Both processes start with the stalling of DNA replication at a site of damage but differ in subsequent steps. In the former mechanism replication re-initiates at a site further downstream and the intervening gap is filled up by recombinational events making this process error-free. In translesion synthesis however, the blocked DNA polymerase is removed from the DNA and replaced by a low-fidelity translesion polymerase. Following synthesis of a short patch over the lesion, possibly containing an error, the polymerase exchange is reversed and the replicative polymerase resumes synthesis. This makes this pathway an error-prone pathway.

1.2 DNA repair mechanisms

1.2.1 Direct reversal

1.2.1.1 Alkyl-transferase

Of all DNA repair mechanisms characterized direct reversal (DR) is the most efficient one since it directly catalyzes the reversal of a modified nucleotide to its native form. This process therefore doesn't cleave the DNA backbone of either of the two DNA strands. Two examples of direct reversal repair are alkyl-transferases and photoreactivation. Alkyl transferases catalyze the removal of an alkyl group attached to a DNA nucleotide. The most commonly encountered alkyl modification is methylation. This modification occurs through the reaction of the DNA with certain methylating agents in the environment, like methylmethane sulfonate (MMS), N-methyl-N-nitrosourea (MNU) as well as the endogenous compound S-adenosylmethionine (SAM). These compounds induce several deleterious lesions in the DNA of which the major ones are: O^6-methylguanine, N^3-methyladenine, N^1-methyladenine and N^3-methylcytosine, all of which affect base pairing properties of the altered nucleotides.

Methylation also occurs enzymatically and is for example involved in epigenetic silencing of genes (Bird, 2002), strand discrimination in mismatch repair (Wagner and Meselson, 1976) or protection against restriction enzymes (Arber and Dussoix, 1962). However, these methylation events occur at distinct non-disturbing positions on the nucleobases, preventing them from being recognized and removed by DNA repair processes.

Cells have evolved three distinct strategies to deal with methylation modifications: two direct reversal mechanisms and a mechanism following the BER pathway. The first strategy to remove methylation modifications by direct reversal involves the removal of methyl groups by enzymes containing nucleophilic cysteine residues. These enzymes recognize the modified base (O^6-methylguanine) and use their catalytic cysteine residue to irreversibly accept the methyl moiety from the DNA. This strategy is followed by for example the human O6-alkylguanine-DNA alkyltransferase (AGT), also known as O6-methylguanine-DNA methyltransferase (MGMT). In *E. coli* two examples are the O6-methylguanine DNA methyltransferase N-Ada (Sedgwick, 2004), which was named after its expression in the adaptive response to alkylating agents, and O6 guanine transferase, Ogt (Potter *et al.* 1986, 1987). Co-crystal structures of the human AGT-protein bound to O6-methylguanine showed that the protein obtains access to the modified guanine by flipping it out into a specific binding pocket of the protein (Daniels *et al.* 2004; Duguid *et al.* 2005). This brings the modified nucleotide in close vicinitv to the catalytic cysteine residue (Fig. 1).

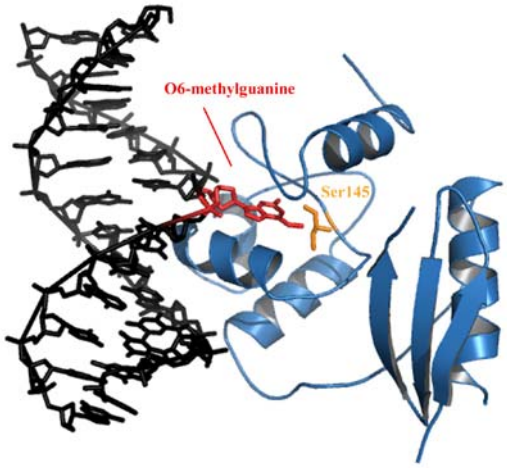

Fig. 1 Crystal structure of the human alkylguanine-DNA alkyltransferase (AGT) bound to dsDNA containing an O6-methylguanine adduct (pdb accession number 1T38). The protein is indicated in blue and the DNA in black. The extrahelical nucleotide is highlighted in red and is in close vicinity to the catalytic cysteine, which in the protein used for the co-crystal has been replaced by serine (orange).

The extrahelical conformation is induced and stabilized by a tyrosine residue that rotates the 3' phosphate due to steric repulsion. This mechanism is used by multiple base flipping proteins to flip out the damaged nucleotide. AlkA (Hollis *et al.* 2000), APE1 (Mol *et al.* 2000), Endonuclease IV (Hosfield *et al.* 1999) and UDG (Parikh *et al.* 1998) all use an aromatic residue to rotate the 3' phosphate thereby facilitating base flipping. In addition the AGT protein uses a conserved arginine to fill up the place in the DNA helix left by the flipped-out nucleotide (Daniels *et al.* 2004; Duguid *et al.* 2005). This arginine might therefore be involved in active base flipping by pushing the nucleotide out of the helix and prevent the extrahelical nucleotide to fall back to its original position. Also this strategy is often used by base flipping enzymes.

The extrahelical nucleotide is packed in a hydrophobic cleft that confers specificity by specific hydrogen bond formation with guanine and steric exclusion of the other residues.

The difference in affinity for O6-methylguanine over guanine is only 3-fold (Rasimas *et al.* 2003) and seems to be only due to a larger hydrophobic surface emanating from the methyl group. This lack of specificity towards its substrate is however not cytotoxic in this direct reversal pathway since no reaction can occur on non-damaged bases.

A second pathway to remove methyl groups from the DNA has been discovered much later and effectively repairs several methylated nucleobases like 1-methyladenine and 3-methylcytosine. Its mechanism involves an oxidative dealkylation dependent on the presence of iron(II), α-ketoglutarate and O_2 (Trewick *et al.* 2002; Falnes *et al.* 2002). The endpoint of the repair reaction is identical to the mechanism of AGT but the overall repair pathway is completely different and involves the release of the methyl group as formaldehyde (Fig. 2).

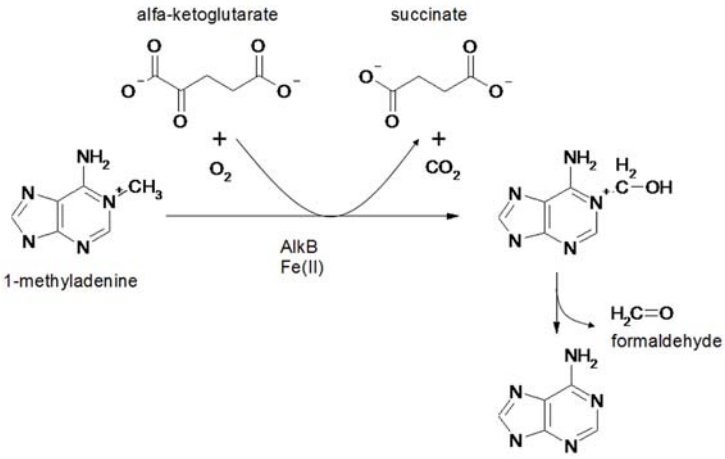

Fig. 2 Reaction mechanism of the oxidative dealkylation pathway. The schematic representation indicates the activity of the AlkB protein on 1-methyladenine, but the same sequence of events is used by other proteins from the same group and on other alkylated nucleobases.

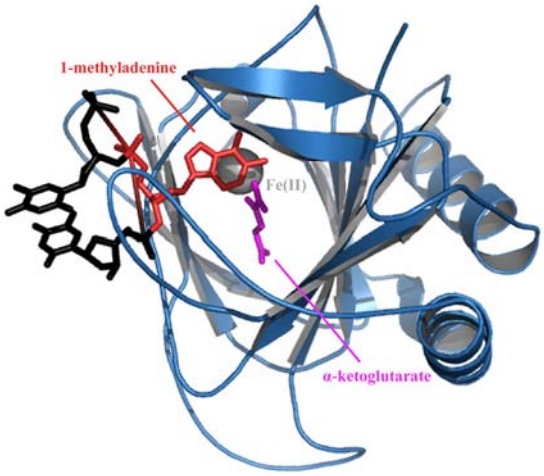

Fig. 3 Crystal structure of the E. coli AlkB protein bound to a trinucleotide containing a centrally located 1-methyladenine adduct (pdb accession number 2FD8). The protein is indicated in blue and the DNA in black. The extrahelical nucleotide is highlighted in red and is located in close vicinity to the Fe^{2+} ion (gray sphere) and the 2-oxoglutarate cofactor (magenta).

This group of proteins has been named DNA-dioxygenases and its most well-known constituent is the *E. coli* AlkB protein. A co-crystal structure of this protein bound to a

1-methyl-adenine containing trinucleotide showed that yet again the modified nucleotide becomes extrahelical relative to its neighboring nucleotides (Yu *et al.* 2006). The alkylated base is bound in a deep, predominantly hydrophobic cavity. Since AlkB recognizes alkylation modifications at endocyclic ring nitrogens in all four nucleotide bases it is clear that flipping the modified nucleotide into the pocket does not confer damage recognition but serves to bring the methylated base in close vicinity to the 2-oxoglutarate and Fe^{2+} metal ion required for the oxidation reaction (Fig. 3). A mechanism for damage recognition is however not known but it can be envisioned that the presence of the methyl group facilitates flipping due to destabilization of the helical structure of the DNA, thereby conferring (slight) selectivity for modified DNA nucleotides. However, high selectivity is not required since the enzyme's reaction mechanism does not affect undamaged DNA.

The third pathway to remove alkylation damages does not occur through direct reversal but instead involves the activity of DNA glycosylases and will therefore be discussed in the base excision repair section (paragraph 1.2.2.2). This pathway is used by the alkyladenine glycosylase (AAG) which efficiently repairs cytotoxic 3-methyladenine lesions.

1.2.1.2 Photolyase

The only proteins belonging to the group of direct reversal proteins not involved in repairing alkyl modifications are photolyases. These proteins specifically recognize either cyclobutane pyrimidine dimers (CPDs) or 6-4 pyrimidine-pyrimidones (6-4PPs), the two major UV-light induced DNA lesions (Fig. 4). Also the Dewar isomer of the 6-4PP is repaired by a photolyase, but with a severely reduced efficiency (Zhao *et al.* 1997).

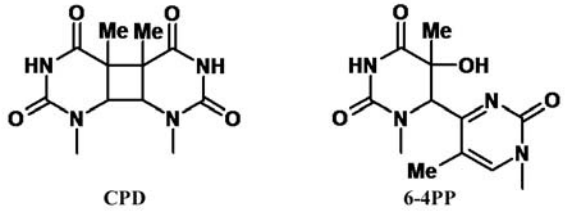

Fig. 4 The two major UV-induced DNA damages: cyclobutane pyrimidine dimer (CPD) and (6-4) pyrimidine pyrimidone (6-4PP).

In both lesions two adjacent pyrimidines are covalently linked to each other by either a cyclobutane ring (CPD) or a single covalent bond (6-4PP). Photolyase can revert these dimers to their native monomeric form at an energetic cost. This energetic cost is paid by the absorption of light from the violet-blue range of the spectrum (Rupert *et al.* 1958) by a cofactor, either methenyltetrahydrofolate (MTF) or 8-hydroxy-7,8-didemethyl-5-deazariboflavin (8-HDF) (Park *et al.* 1995; Tamada *et al.* 1997). These cofactors serve as an antenna and are only required under conditions of limiting light (Hamm-Alvarez *et al.* 1989). The excited chromophore transfers its energy to a second chromophore present in each photolyase, the $FADH^-$ molecule. This cofactor proved to be essential for the function

of photolyase and the photochemically excited form is capable of reverting the photoproduct to its native form. Crystal structures of two distinct CPD photolyases have shown that the $FADH^-$ cofactor is deeply buried within the protein and can only be accessed through a small cavity (Park *et al.* 1995; Tamada *et al.* 1997). The dimensions of this cavity are perfectly suited to accommodate a CPD lesion to bring it in close vicinity of the catalytic $FADH^-$. The presence of the cavity suggested that when photolyase is bound to the DNA the CPD lesion is flipped out of the DNA helix into the cavity, or pocket, of the photolyase protein.

More recently a crystal structure of photolyase was obtained in the presence of a 14 nucleotide DNA duplex containing a central CPD analog modification (Mees *et al.* 2004). The crystal structure revealed however that the DNA substrate does not contain the originally used CPD modification but instead harbours a repaired thymine dinucleotide at this position (Fig. 5). Probably, the presence of the protein in combination with synchrotron radiation resulted in efficient photoreactivation of the CPD. In addition, the structure revealed that the protein undergoes extensive contacts with the DNA phosphodeoxyribose backbone as can be expected for a sequence-independent DNA-binding protein (Fig. 5).

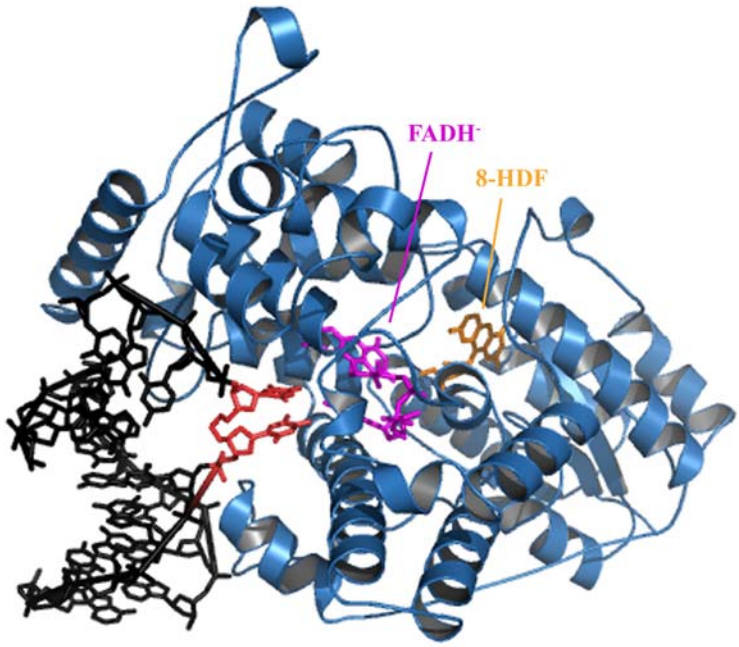

Fig. 5 Co-crystal structure of the *Anacystis nidulans* DNA photolyase bound to a dsDNA substrate containing a CPD analog after in-situ repair (pdb accession number 1TEZ)(see text). The protein is indicated in blue and the DNA in black. The flipped-out nucleotides representing the CPD analog are highlighted in red. In close vicinity of the extrahelical nucleotides the $FADH^-$ cofactor is visible (magenta). The second chromophore, 8-HDF, is shown in orange.

The lesion itself is indeed flipped-out of the DNA helix into a pocket of the protein where it has direct access to the $FADH^-$ cofactor. The open space left by the flipped-out CPD is not occupied by side chains of the photolyase protein, but results in a sharp kink in the DNA structure of approximately 60^o. However, the opposite might also be true; the protein-induced kink in the DNA helix might facilitate base flipping.

Summarizing, the crystal structure provides evidence for the damage recognition

mechanism of photolyase. Since it is known that the presence of a lesion reduces base stacking interactions in the DNA it can be expected that the protein initially senses this property: bending of the DNA is facilitated and base flipping occurs more readily at the site of damage. Verification of the presence of a lesion and determination of the nature of this lesion occurs through incorporation of the modification into a highly specific binding pocket. This pocket only facilitates entrance of the specific photoproduct thereby excluding other modifications and non-damaged nucleotides.

Photolyases have been identified both in prokaryotic as well as in eukaryotic cells, but human cells do not express the photolyase protein. However, the presence of two genes sharing high homology to the photolyases has been shown but their functional role has remained cryptic for a long time hence resulting in their name, cryptochromes (van der Spek *et al.* 1996; Hsu *et al.* 1996). Nowadays several cryptochromes have been identified and the term has acquired a precise meaning: a photolyase-like protein with no DNA repair activity but with known/presumed blue-light receptor functions. Deletion of the two cryptochrome genes in mice (mCry1 and mCry2) has resulted in the mice becoming behaviourally arrhythmic as measured by their locomotor activity in a running wheel (van der Horst *et al.* 1999). In addition, cyclic expression of Per genes (two of the core circadian rhythm proteins) is abolished in a $mCry1^{-/-}mCry2^{-/-}$ background (Okamura *et al.* 1999). These results indicated their involvement in the general regulatory mechanism of the circadian clock. This clock is generated by an autoregulatory transcriptional- and post-translational feedback loop that exhibits a periodicity of approximately 24 hours, without the need for exogenous cues. This means that its periodicity is retained even in constant darkness and temperature.

Another role in the circadian clock for animal-type cryptochromes was first discovered in insects, where a mutation of the *Drosophila melanogaster* cryptochrome (Dmcry), resulted in a defect in light-mediated clock entrainment (Emery *et al.* 1998; Stanewsky *et al.* 1998). Additionally, CRY protein levels themselves have been shown to be controlled by light (Emery *et al.* 1998). This shows that this cryptochrome (CRY) protein from *D. melanogaster* uses its blue-light photoreceptor function to reset the circadian clock and synchronize it to the exogenous light-dark cycle. The clock-resetting function of cryptochromes has however never been shown for human cryptochromes and it was postulated that it has lost its photoreceptive function in humans (Berson, 2003).

In plants, cryptochromes have been found to be involved in several distinct blue-light dependent cell-growth and development processes like early seedling development, leaf and stem expansion, initiation of flowering, and gene regulation.

In addition to classical photolyases and cryptochromes a third class of this family has been identified in *Synechocystis* sp. PCC6803 (Hitomi *et al.* 2000) and in *Vibrio cholerae* (Worthington *et al.* 2003). Members of this family exhibited no or very little DNA repair activity and therefore were considered bacterial cryptochromes. However, these cryptochromes showed higher sequence homology to defined eukaryotic cryptochromes than to bacterial

photolyases and therefore they were called Cry-DASH proteins (*Drosophila*, *Arabidopsis*, *Synechocystis*, Human). Proteins from this class were shown to exhibit photolyase activity with a high specificity for CPDs in ssDNA (Selby and Sancar, 2006).

1.2.2 Base excision repair

Base excision repair (BER) is a ubiquitous DNA repair mechanism targeting a wide variety of different kinds of damages. Since this pathway does not only remove the damaged group but excises the entire damaged nucleotide this results in an intermediate containing a disruption in the DNA backbone. The BER pathway is initiated by a specialized site of damage-recognizing proteins, glycosylases (Krokan *et al.* 1997). These proteins recognize only one or a small group of related DNA modifications and hydrolyze the N-glycosidic bond connecting the modified base to the DNA backbone (Fig. 6).

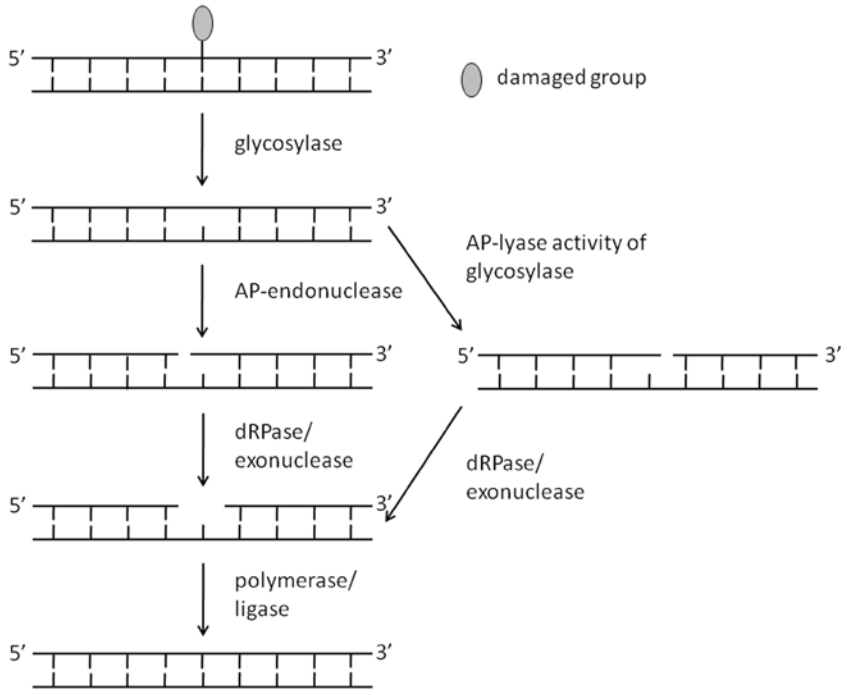

Fig. 6 Schematic representation of the (short-patch) base excision repair mechanism. The damaged group is indicated in gray. The enzyme types that are involved in the process are indicated. For a description of the process see text.

This step results in an abasic site (AP-site) intermediate. Since this intermediate also often occurs spontaneously and is non-coding in DNA replication, cells express a specific class of enzymes to remove the AP-site. These enzymes, apurinic/apyrimidinic (AP) endonucleases, catalyze the hydrolytic cleavage of the phosphodiester bond directly 5' to the lesion (Fig. 6). The resulting intermediate containing a 5' terminal deoxyribose-phosphate

moiety is further processed by a DNA-deoxyribophosphodiesterase (dRPase) or exonuclease. This results in a one nucleotide gap in the DNA strand that contained the damaged base. The pathway is further completed by the repair synthesis steps involving a DNA polymerase to fill in the gap and a ligase to finally seal the DNA strand.

However the BER pathway can also take an alternative route to repair DNA (Fig. 6). A subset of DNA glycosylases exhibit not only their glycosylase function but also have an associated AP lyase activity (Krokan *et al.* 1997). This means that after the AP-site has been created the enzyme also cleaves the DNA strand 3' to this site by a β-elimination step. This incision yields an unstable intermediate and the 3' abasic site will undergo a ring-opening reaction resulting in a 3'-α,β-unsaturated aldehyde (Nash *et al.* 1996). Sometimes these enzymes can also perform a δ-elimination step. This yields a substrate lacking a nucleotide but still containing both the 3' and 5' phosphates (Zharkov *et al.* 2003). The resulting 3' modified nucleotide is subsequently removed by a phosphodiesterase before repair synthesis can take place.

Both of these BER pathways eventually replace only a single nucleotide and for that reason are referred to as the short-patch repair pathway of BER. In mammalian cells, repair involving mono-functional glycosylases (lacking AP-lyase activity) can also take place via a long–patch repair pathway (Izumi *et al.* 2003). In this pathway the AP-site is further processed by an AP-endonuclease resulting in a free 3'-OH site which is available for a DNA polymerase for extension. The extension of the DNA strand displaces the native DNA strand containing the 5' AP-site. The resulting 5' flap is then removed by a flap-endonuclease (FEN1). Finally, ligation restores the DNA again to its undamaged form.

In BER glycosylases are responsible for the initial damage-recognition step. These enzymes specifically recognize only one specific type of damage or a small group of related DNA modifications and therefore vary in their substrate specificity. The principles of damage recognition however are highly similar among members of this group and this aspect will be discussed via two examples: 8-oxoguanine DNA glycosylase 1 (OGG1) and the human alkyladenine glycosylase (AAG). These two enzymes are both well-studied and represent a glycosylase recognizing only one lesion (OGG1) and a glycosylase recognizing multiple DNA modifications (AAG) respectively.

1.2.2.1 8-oxoguanine DNA glycosylase I (OGG1)

Reactive oxygen species (ROS) are produced as intermediates and as by-products of aerobic respiration, but also through exposure to ionizing radiation and other agents that generate free radicals. These oxidants are highly reactive towards organic molecules including DNA and upon reaction with the DNA form numerous genotoxic lesions and DNA strand breaks (Lindahl, 1993). One of the most abundant lesions formed by ROS is 7,8-dihydro-8-oxoguanine (oxoG) (Michaels and Miller, 1992; Grollman and Moriya, 1993). Due to its ability to pair with adenine in Hoogsteen mode, these lesions give rise to GC → TA

transversions (Shibutani *et al.* 1991). To prevent the formation of these mutations cells have evolved three levels of defense (Michaels and Miller, 1992; Grollman and Moriya, 1993). The first one is MutT (hMTH1 in humans), which hydrolyzes oxo-dGTP thereby depleting the cell of this compound and preventing its incorporation into the DNA. The second level is constituted by the glycosylase/β-lyase protein MutM also known as Fpg (hOGG1 in humans) which recognizes and cleaves the glycosidic bond between the oxoG and the ribose unit in the 8oxoG-C basepair. The last protein in this regard is MutY, which recognizes and removes the adenine in a 8oxoG-A mismatch.

OxoG differs from normal guanine in only two positions (C8 and N7, Fig. 7).

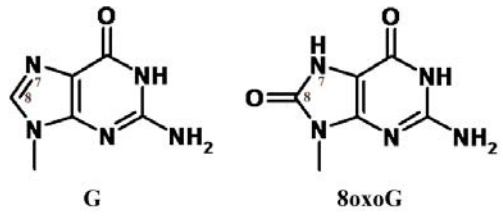

Fig. 7 Chemical structure of 8-oxoguanine (8oxoG) compared to guanine (G). The modified purine differs only at two positions from guanine.

As a consequence this moiety forms a normal base pair with cytosine and its presence only leads to a minor helical distortion (Lipscomb *et al.* 1995). To localize such a lesion amidst about $3 \cdot 10^9$ bases of normal DNA (in a haploid human cell) therefore poses a formidable challenge to these proteins. To reveal the damage recognition mechanism of the hOGG1 protein an inactive mutant of the protein was crystallized in three different stages of the base extrusion pathway (Bruner *et al.* 2000; Banerjee *et al.* 2005*,* 2006). Using disulfide cross-linking technology to target the protein not only a structure of hOGG1 bound to 8oxoG (Bruner *et al.* 2000) could be obtained but also a structure of hOGG1 bound to undamaged DNA (Banerjee *et al.* 2005) and a structure of the protein sampling the G-C basepair adjacent to the oxoG lesion (Banerjee and Verdine, 2006).

The structure of hOGG1 bound to oxoG showed that the oxoG moiety becomes extrahelical and is located in a pocket of the protein (Bruner *et al.* 2000; Fig. 8A). The vacated space inside the DNA helix is filled by an asparagine residue suggesting an active base flipping mechanism. Extensive contacts between the pocket and oxoG prevent incorporation of nucleotides other than G, while only one contact is established that would differ for G as compared to oxoG (H-bond between the carbonyl of Gly42 and the N7-H). The estranged C on the other hand remains intrahelical but has lost base stacking interactions due to a kink in the DNA. This enables the protein to undergo extensive contacts with this nucleotide to ensure that repair only takes place when oxoG is paired with cytosine.

In the structure of hOGG1 bound to undamaged DNA the DNA is drastically kinked again thereby extruding an undamaged G from the DNA helix (Banerjee *et al.* 2005; Fig. 8B). This G is however rejected by the pocket and instead lies against the protein surface at an exo-site in close vicinity to the pocket. Free energy calculations have shown that the pocket

favors oxoG over G by 6.8 kcal/mol corresponding to a roughly 10^5-fold preference to insert oxoG versus G into the pocket. This energy difference arises mainly from a more favorable interaction between Gly42 and the N7-H and a more favorably aligned dipole moment of the pocket for the oxoG lesion (Banerjee *et al.* 2005). Comparison of the two structures shows that contacts between the protein and the DNA at the 3' side of the lesion are formed early in the base-extrusion pathway while interactions on the 5' side and with the extrahelical nucleotide occur late.

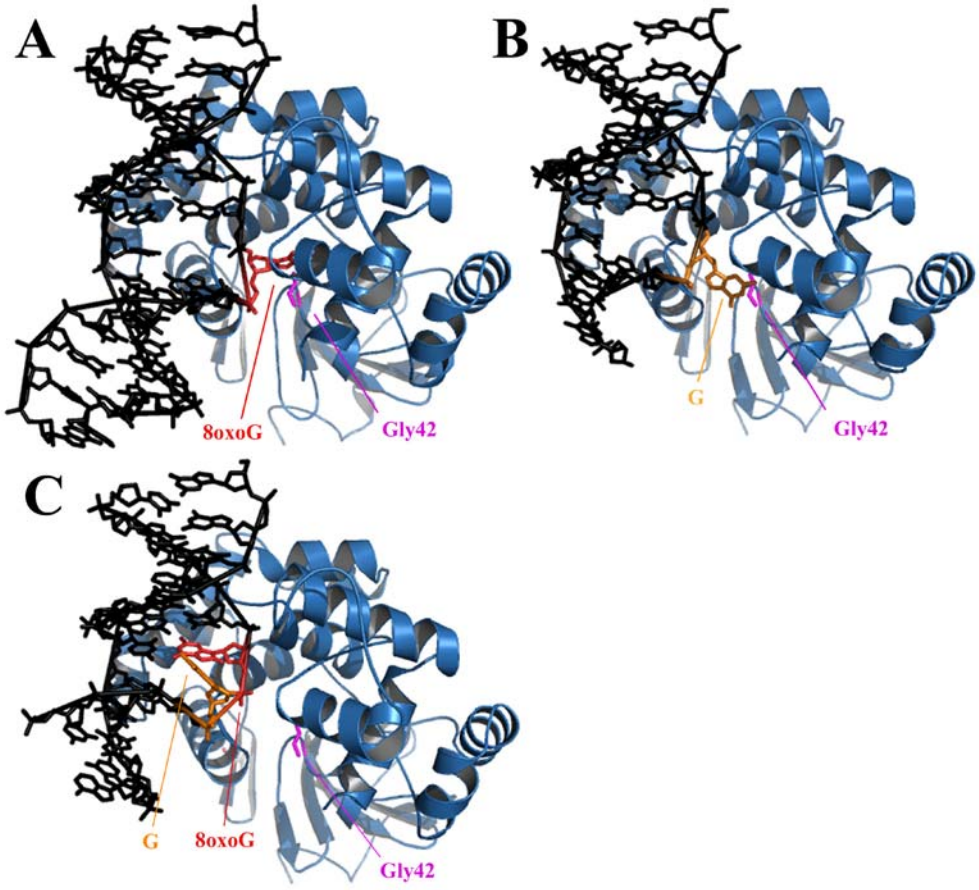

Fig. 8 Co-crystal structures of hOGG1 bound to various DNA substrates. (A) hOGG1 recognizing the 8oxoG damage (pdb accession number 1EBM). The extrahelical 8oxoG moiety is shown in red, Gly42 in magenta. (B) hOGG1 bound to undamaged DNA (pdb accession number 1YQK). The extrahelical G is highlighted in orange, Gly42 in magenta. (C) hOGG1 sampling the undamaged guanine (orange) adjacent to the 8oxoG damage (red) (pdb accession number 2I5W). Colors are the same as for (A). All structures show the protein in blue and the DNA in black.

In a complex where hOGG1 is captured to flip out the guanine adjacent to an oxoG modification the nucleotide does not become fully extrahelical (Banerjee and Verdine, 2006;

Fig. 8C). Instead the nucleotide is folded into the major groove and forms a basepair with the neighboring, intrahelical oxoG. The nucleotide cannot be fully extrahelical and therefore cannot be flipped into the exo-site since in this situation two phosphodiester O-atoms point inwards and clash with the carbonyl group of oxoG. Again in this structure the contacts at the 3' side of the lesion are similar to the previous structures while contacts on the 5' side differ considerably.

Taking into consideration the drastic nature of the DNA remodeling performed by hOGG1 it is highly likely that this occurs through the formation of multiple intermediates. In this regard it is plausible to consider the two structures of hOGG1 exhibiting an extruded G as analogous but not necessarily identical to such intermediates. The oxoG flanked G complex can be considered as an early intermediate in which the target nucleotide is extruded from the DNA helix into the major groove but cannot escape the confines of this major groove yet (Banerjee and Verdine, 2006).The G complex can then be regarded as a late intermediate in which the nucleotide is extrahelical but not yet fully incorporated into the active site pocket (Banerjee *et al.* 2005). Consistent with this, in the oxoG flanked G complex the conformation of the DNA much more resembles B-DNA than in the other two structures. These three structures therefore nicely show three snapshots of the base extrusion pathway of hOGG1.

1.2.2.2 alkyladenine glycosylase (AAG)

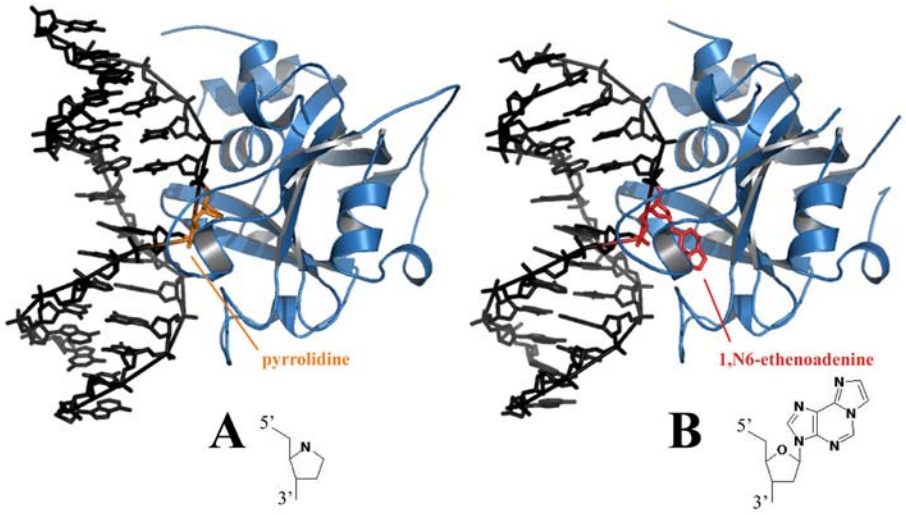

Fig. 9 Co-crystal structures of hAAG bound to various DNA substrates. (A) hAAG bound to a pyrrolidine (pyr) abasic nucleotide; transition-state mimic of glycosylase reaction (pdb accession number 1F6O). The pyr moiety is indicated in orange. (B) hAAG bound to a DNA substrate containing a 1,N6-ethenoadenine (εA) modification indicated in red (pdb accession number 1F4R). Both structures show the protein in blue and the DNA in black. The chemical structures of the DNA modifications are indicated.

Human alkyladenine glycosylase (AAG) recognizes several different types of

lesion including 3-methyladenine, 7-methyladenine and 7-methylguanine as well as different types of modified bases, hypoxanthine (Saparbaev and Laval, 1994) and 1,N^6-ethenoadenine (Saparbaev *et al.* 1995).

Different cocrystal structures show the human AAG protein bound to a DNA substrate containing either a pyrrolidine abasic (pyr) nucleotide (Lau *et al.* 1998, 2000; Fig. 9A) or a 1,N^6-ethenoadenine (εA) modification (Lau *et al.* 2000; Fig. 9B). In both cases the protein contacts the DNA on both sides of the lesion and induces a small bend in the DNA of approximately 22°. The protein also contacts the non-damaged DNA strand explaining the protein's selectivity for dsDNA substrates. Furthermore, as was the case for the hOGG1 protein, the structures reveal the flipping out of the nucleotide targeted for base excision with a conserved aromatic residue filling up the vacated space. The presence of such a residue, as was argued before, might indicate an active flipping mechanism. Again here, the modified nucleotide is flipped out and secured in an extrahelical position by steric repulsion of protein residues towards the flanking phosphate groups. Not only the target nucleotide undergoes a conformational change upon protein binding but also its estranged base pairing partner is slightly pushed into the major groove by the intercalation of the aromatic residue. This reduces base stacking interactions of this base with its nearest neighbors.

Proteins capable of recognizing various different DNA modifications are not expected to achieve damage recognition by inserting these groups into a pocket of the protein since these sites have to be tailor-made to the specific modified base and should exclude non-damaged nucleotides. However, in the case of AAG the target nucleotide (εA) can be seen to be incorporated in a pocket of the protein where it is sandwiched between two tyrosine and one histidine side chain (Lau *et al.* 2000). The localization of this nucleotide into the pocket of the AAG protein brings the glycosidic bond in close vicinity to the present water molecule. After deprotonation this molecule forms the hydroxyl nucleophile required for bond cleavage. The orientation of the residues inside the pocket is the same for the abasic pyr modification, suggesting that the pocket is not induced by substrate binding but already exists in this configuration. Shape complementarity between the pocket and its substrates alone does however not completely account for its catalytic selectivity. Unfavorable interactions with the exocyclic amino groups of purine residues do slow down the reaction with undamaged purines. The exocyclic group is protonated in some methylated (e.g. 7-methylguanine) purine residues and it was proposed that its developed positive charge pulls it strongly enough into the active site to overcome the unfavorable interactions. Inosine, which lacks this amino group altogether, consequently selectively binds to the protein as well. Selectivity for other substrates can be achieved by the occurrence of a delocalized positive charge on some alkylated DNA bases like 3-methyladenine and 7-methylguanine. This charge might enable tight binding interactions with an aromatic side chain of the AAG active site and at the same time makes the modified base a good leaving group with a weakened glycosidic bond. This weakened bond might enable the glycosylase to dispose of the modified base more efficiently

than a normal base. From the crystal structure it was postulated that specificity for the neutral εA modification is achieved by its enlarged Van der Waals surface area compared to normal adenine and the formation of a hydrogen bond between a conserved histidine and the N^6 of εA which offers a lone pair that only exists in the alkylated form (Lau *et al.* 2000).

Base flipping is also believed to contribute to AAG's specificity for modified nucleotides as it has been shown that destabilized mismatched base pairs are better substrates than stable Watson-Crick pairs (O'Brien and Ellenberger, 2004). However, due to unfavorable interactions of the exocyclic amino group with residues in the pocket, AAG's catalytic activity on these mismatches is still much lower than on substrates containing methylated bases. Some of the lesions recognized by the AAG protein are compromised in their base pairing ability thereby promoting damage recognition by energetically favoring base flipping. Some other modifications that are being recognized are known to cause little or no distortion of the DNA helix making it impossible for the protein to find the lesion by merely scanning the DNA surface. Localization must therefore occur through either partial or complete unstacking of sequential nucleotides of the DNA helix while the protein slides along the DNA. The same model was proposed for the bacterial ortholog AlkA (Verdine and Bruner 1997).

1.2.3 Ultraviolet Damage Endonuclease (UVDE) repair

In contrast to proteins involved in direct reversal repair and base excision repair, the UVDE protein has the property to exhibit a broad substrate specificity. Not only does this protein recognize UV-induced DNA lesions (Bowman *et al.* 1994), as its name indicates, but it also recognizes other non-UV-induced lesions like abasic sites and cis-platin intrastrand crosslinks (Avery *et al.* 1999). Also nucleotide mismatches are recognized by UVDE, but in an inefficient and sequence specific manner (Kaur *et al.* 1999). This suggests that UVDE recognizes a common distortion in the DNA helix rather than the structure of the DNA modification itself. Damage recognition by the UVDE protein results in the occurrence of an incision directly 5' to the lesion (Bowman *et al.* 1994). This incision results in the formation of a 3'-hydroxyl and 5'-phosphate group and requires further processing of the lesion to restore the DNA to its undamaged form. *In vitro* reconstitution of the UVDE pathway implicated several other proteins in this process in addition to UVDE: Rad2 (the *S. pombe* FEN-1 homolog), pol δ (capable of strand-displacement synthesis), PCNA (Proliferating Cell Nuclear Antigen), RFC (Replication Factor C) and ligase (Alleva *et al.* 2000). Two different pathways can be envisioned involving these proteins. In the first pathway Rad2 performs an incision several nucleotides 3' to the lesion after UVDE-mediated incision has taken place. The resulting gap is subsequently filled in by the polymerase. In the other pathway it is the polymerase that first synthesizes a piece of DNA from the nick in the 5' - 3' direction thereby displacing the damaged DNA strand and resulting in a flap which is then removed by Rad2. The resulting nick in both pathways is then closed by the ligase. How the repair process exactly takes place *in vivo* remains however still elusive.

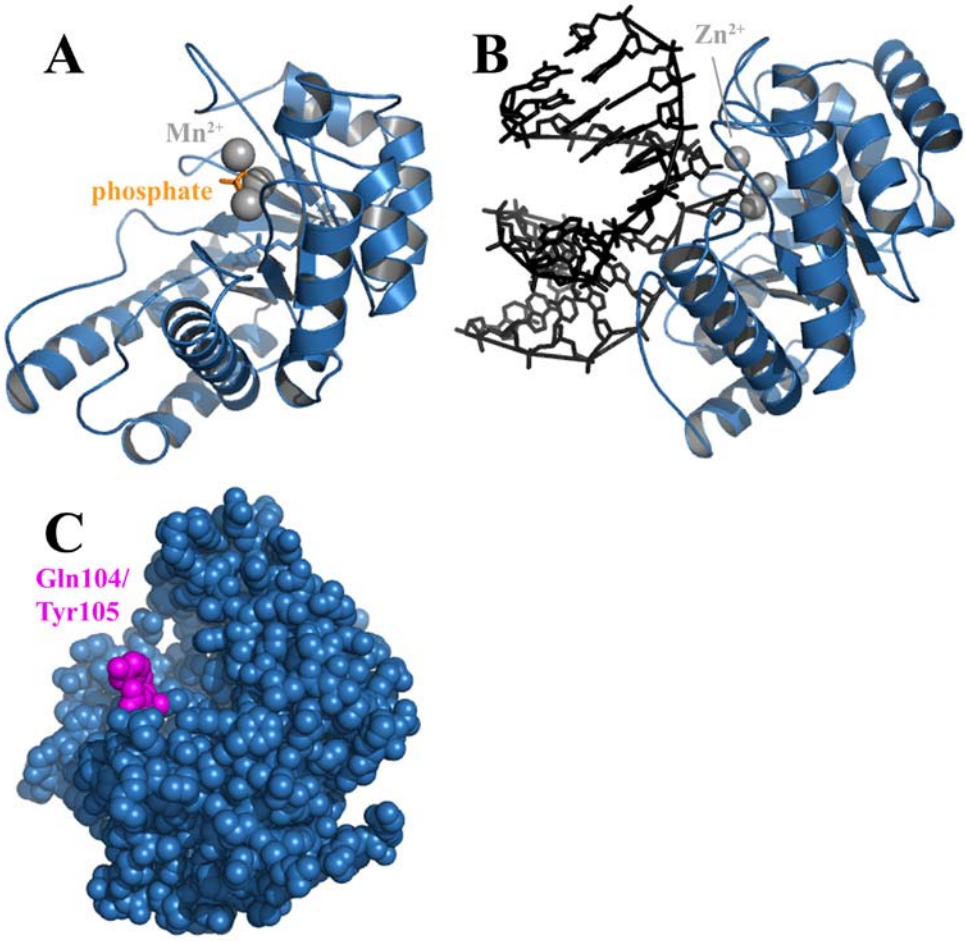

Fig. 10 Crystal structures of UVDE and Endonuclease IV. (A) UVDE containing three Mn^{2+} metal ions (gray spheres) and a phosphate molecule (orange) (pdb accession number 2J6V). (B) Endonuclease IV bound to dsDNA containing tetrahydrofuran as a synthetic abasic site (pdb accession number 1QUM). The substrate exhibits a nick directly 5' to the lesion due to cleavage by the protein. The structure shows three Zn^{2+} metal ions (gray). The DNA is shown in black. (C) Different view of the UVDE protein showing the DNA-binding groove and the two residues Gln104 and Tyr105 in magenta. Proteins are indicated in blue.

Recently the crystal structure of the UVDE protein from *T. thermophilus* was solved and it revealed the protein to be a single-domain TIM barrel (eight β/α motifs folded into a barrel structure) containing three metal ions (Paspaleva *et al.* 2007; Fig.10A). The nature of the three metal ions could not be determined but it was shown that for optimal incision the presence of Mn^{2+} was compulsory indicating manganese to be a required cofactor. UVDE's closest structural neighbor proved to be the TIM barrel containing protein Endonuclease IV (Endo IV). This protein is an apurinic/apyrimidinic (AP) endonuclease that in analogy to

UVDE primes DNA repair synthesis by cleaving the DNA backbone 5' of AP sites (Barzilay and Hickson, 1995). In addition it was shown to contain three divalent Zn ions directly involved in phosphodiester cleavage (Hosfield *et al.* 1999; Fig. 10B). Superimposing the UVDE protein on Endo IV showed that the three metal atoms occupy the same positions in both proteins indicating that also for UVDE these metal ions are involved in catalysis. The metal ions are located at the bottom of a wide groove present in either protein, whose extensive positive charge is suited to accommodate dsDNA. In the structure of Endo IV in complex with abasic site containing dsDNA nicked 5' to the lesion it can be seen that the DNA helix adopts a kinked conformation. The abasic residue is flipped out of the DNA helix towards the protein whereas the opposing nucleotide is flipped into solution (Hosfield *et al.* 1999). A similar conformation for the DNA can be envisioned for the UVDE protein where the DNA is located in the groove of the protein. This localization also implies an important function for two conserved amino acids (Gln104 and Tyr105) sticking out from the groove into the solvent in close vicinity to the metal-binding site (Fig. 10C). Since similarly located residues of Endo IV were proposed to stabilize the kinked structure of the DNA and probe the DNA for damage a similar function for these residues in UVDE can be envisioned.

The major difference between the two homologous proteins is their divergent substrate specificity, where Endo IV mainly, but not solely, recognizes AP sites. UVDE recognizes several structurally unrelated DNA modifications, including UV-induced pyrimidine dimers, implying that UVDE flips out (at least) two nucleotides. These nucleotides will be rotated inside the protein but the interactions with a potential pocket are not expected to be highly selective. Also the observation that UVDE can recognize and repair nucleotide mismatches (Kaur *et al.* 1999) underlines the fact that it does not use a pocket to discriminate damaged from undamaged nucleotides.

1.2.4 Nucleotide excision repair

Just like UVDE the NER system has the property to be capable of repairing various structurally unrelated types of damage. These damages include intrastrand crosslinks like cisplatin adducts, bulky mono-adducts like benzo(a)pyrene diolepoxide (BPDE) and N-2-acetylaminofluorene (AAF), and UV-induced lesions (Truglio *et al.* 2006a). Three different NER-specific proteins are involved in the bacterial repair process (UvrA, UvrB and UvrC) and at least 30 different proteins in man (Araujo *et al.* 2000). Even though the level of complexity of the human NER system is much higher than that of bacteria the overall principles are evolutionary conserved and can be dissected in the following steps. (A) DNA damage recognition and assembly of the pre-incision complex. The broad substrate specificity of the system excludes recognition of the chemical structure of the lesion itself and therefore a common alteration in the DNA induced by all these different damages must be recognized. Since also undamaged DNA can take up different conformations, discrimination between damaged and undamaged DNA can be quite a challenge. To prevent gratuitous repair the

NER system therefore uses multiple proteins to scout for the presence of a lesion, each sampling a different parameter of the DNA helical structure. (B) Dual incision on both sides of the damage. The incision positions are located several nucleotides away on either side of the damage. In bacteria both incisions are exerted by a single enzyme (UvrC; Verhoeven *et al.* 2000) whereas in eukaryotic NER the incisions are made by separate nucleases (XPG and XPF•ERCC1 for the 3' and 5' incision respectively.; Evans *et al.* 1997). (C) DNA repair synthesis and ligation. After the dual incision event, the removal of the excised damaged oligonucleotides by a helicase results in the presence of a single strand DNA gap. This gap has to be filled in by a DNA polymerase and the resulting nick is subsequently closed by DNA ligase. These steps thereby result in the restoration of the original DNA sequence.

Since the Nucleotide Excision Repair system is the main topic of this study both the bacterial and the human NER mechanisms will be described in further detail in section 1.4 (prokaryotic NER) and 1.5 (eukaryotic NER).

1.3 Initial damage sensing mechanisms

One of the most intriguing questions in DNA repair is how a modified nucleotide is identified within billions of undamaged nucleotides. To find the damage in different phases of the cell cycle and in different chromatin environments among so many non-damaged nucleotides represents the classical needle-in-a-haystack problem. Previously the discussion has focused on how the substrate discrimination step takes place but does not include the overwhelming initial search of the genome that these enzymes must undertake to pinpoint the damaged site. Several studies dealt with this topic of which three different theories will be discussed below.

1.3.1 Poor base stacking governs damage recognition

A common feature of all repair protein-DNA complexes is a conformational change in the DNA at the site of lesion due to base flipping, DNA bending and/or strand separation (paragraph 1.2). These structural alterations are energetically more likely to occur at a damaged site due to a reduction in base pairing and stacking interactions at this site. This makes these protein-induced DNA alterations an efficient tool to scout for the presence of a lesion. It seems however unlikely that repair proteins sample each nucleotide individually. For this reason it was proposed that DNA lesion recognition occurs in two distinct steps, not only for NER but also for other DNA repair mechanisms (Yang 2006). Initially the repair protein(s) recognize the weakened base stacking interactions at the site of damage which acts as some kind of hinge rendering the DNA more flexible. Not only damages however render the DNA more flexible, also specific DNA sequences increase the flexibility and bendability of the DNA. TA-rich sequences for example are prone to bend the DNA and are therefore recognized by the TATA box binding protein (TBP), a well-known eukaryotic transcriptional

activator (Parvin *et al.* 1995). Also in bacteria, histone-like proteins HU and IHF are believed to recognize and bind to DNA at sites of increased flexibility (Grove *et al.* 1998). Regardless of the intrinsic flexibility of DNA, checking for the presence of such flexibility instead of sampling all nucleotides separately can increase the process of damage recognition several fold.

Localization of a hinge region subsequently invokes a damage verification step where the protein(s) check for the presence of a lesion by sampling yet another parameter of the damaged DNA (i.e. lesion structure, base pairing interactions). If no apparent aberration can be found this is expected to lead to dissociation of the protein(s) from the DNA.

In NER these two steps can be envisioned to occur by two (or more) different proteins since both prokaryotic and eukaryotic NER employ multiple proteins for damage recognition. In bacteria both the UvrA and UvrB proteins can distinguish damaged from undamaged DNA whereas in humans this is accomplished by the UV-DDB complex, the XPC-hHR23A complex, XPA, and RPA. In this scenario the initial damage recognition protein, UvrA for bacteria and UV-DDB (or XPC•HR23B, XPA•RPA) for man, then might probe the DNA for an acquired flexibility of the DNA helix.

1.3.2 Charge transport as a method for damage recognition

The double-helical structure of the DNA with its π-stacked heterocyclic aromatic DNA bases allows the DNA to serve as a medium for DNA charge transport (CT) (Kelley and Barton, 1999). This CT has been shown to be effective over long distances in the DNA of at least 200 Å (~60 bp; Nunez *et al.* 1999) and is highly sensitive to the presence of DNA damages (Boal and Barton, 2005a) and mismatches (Bhattacharya and Barton, 2001). Even though it has been shown that CT can occur under biological conditions (in cell nuclei (Nunez *et al.* 2001) and histone-wrapped DNA (Nunez *et al.* 2002)) it is still unclear whether this property serves any biological role.

The presence of a [4Fe-4S] cluster of unknown function in several different DNA repair proteins (i.e. Endonuclease III (Cunningham *et al.* 1989), MutY (Guan *et al.* 1998) and XPD (Rudolf *et al.* 2006; further discussed in paragraph 1.6.1)) suggests a possible role of the cluster in CT. Initially the cluster was found to be redox inert and was therefore relegated a structural function (Cunningham *et al.* 1989). However, when bound to DNA both EndoIII and MutY were shown to be redox active displaying midpoint potentials of +50-100 mV versus NHE (Normal Hydrogen Electrode, defined as 0 mV) typical of high-potential iron proteins that can adopt both a 2+ and 3+ oxidation state (Boal *et al.* 2005b). This oxidation from the $[4Fe\text{-}4S]^{2+}$ to the $[4Fe\text{-}4S]^{3+}$ state has been shown to be dependent on CT through the π-stack (Boon *et al.* 2003; Boal *et al.* 2005b). Due to a shift in redox potential upon DNA binding it was furthermore estimated that the proteins have a higher affinity towards the DNA in their oxidized state (Boal *et al.* 2007), suggesting indeed a role for this process in the biological function of the protein.

From these data a model was proposed in which electron release by one protein containing a [4Fe-4S] cluster leads to reduction of a distally bound protein in the $[4Fe\text{-}4S]^{3+}$ state (Boal *et al.* 2007). This process releases the reduced protein from the DNA, due to its reduced affinity. The electron transfer can however only occur when the intervening DNA is well-stacked and does not contain any DNA modifications. When a lesion is present CT is impaired and both proteins remain bound near the lesion. This mechanism thereby enables the proteins to quickly scout the DNA for the presence of DNA damage and allows them to relocalize themselves to damaged DNA sites. Charge transport might then be another mechanism by which a subset of DNA repair proteins, the redox-active proteins, initially scouts for the presence of damage.

The drawback of this model is that it involves the action of more than one protein located at a distance close enough to allow charge transport and implies that the proteins can translocate along the DNA helix. Distances of charge transfer up to 200 Å (~ 60 bp) have been shown to be within reach but with decreasing efficiency. This means that in order to constantly safeguard the genome against DNA damage a huge amount of protein bound to the DNA is required. For the [4Fe-4S] cluster containing protein XPD the model seems even more unlikely since it is not the initial damage-recognizing protein of the NER system. This renders a damage localizing mechanism involving a translocation of the XPD protein unlikely.

1.3.3 Passive vs. active flipping mechanisms

As mentioned before, base flipping is a conformational change of a DNA base from its intrahelical hydrogen-bonded and base-stacked conformation to an extrahelical conformation, a process resulting in a higher accessibility of the base to for example DNA repair enzymes. Not so clear however is the mechanism by which the nucleotide becomes extrahelical. The two major pathways that have been proposed are (1) a passive mechanism where spontaneous protein-independent base flipping is followed by the protein binding to the DNA and (2) an active mechanism where base flipping and binding occur simultaneously.

The passive mechanism implies that even though its intrahelical localization is energetically more favorable sometimes a nucleotide temporarily assumes an extrahelical conformation. The frequency with which this occurs is dependent on the energy difference between the low (intrahelical) and high energy (extrahelical) state. Studies using imino-proton exchange measurements by NMR spectroscopy have reported an equilibrium constant for spontaneous base flipping of a guanine residue in a GCGC tetramer of 3.3×10^{-7} (Dornberger *et al.* 1999). This corresponds to a free energy difference of ~9 kcal/mol. Potential of mean force (PMF) calculations resulted in even bigger free energy barriers (~15 kcal/mol, depending on the identity of the base) for spontaneous base flipping (Guidice *et al.* 2001, 2003), making spontaneous base flipping of undamaged nucleotides an unlikely process. However, damaged DNA bases are not properly base stacked and hydrogen-bonded in the DNA helix and thereby

the energy barrier to spontaneous base flipping might be decreased. A recent computational study using molecular dynamics simulations indicated that the energy required to flip out a CPD from the DNA is ~6.25 kcal/mol (O'Neill *et al.* 2007), much lower than the previously calculated value of ~15 kcal/mol for undamaged bases. This destabilization could allow for the thymine dimer to undergo spontaneous base flipping, leading to a passive mechanism of base flipping and damage recognition for enzymes involved in CPD repair (i.e. photolyase). A passive flipping mechanism has also been proposed for the uracil DNA glycosylase (UNG) protein where it was shown that the protein binds 8000x more tightly to a uracil residue opposite a modified adenine residue (4-methylindole) not capable of hydrogen-bond formation than when paired with a normal adenine residue (Krosky *et al.* 2005).

hOGG1 on the other hand has been shown to very rapidly slide along the DNA, an activity excluding the possibility that it captures a spontaneously formed extrahelical oxoG since this species is expected to occur at low frequency (Blainey *et al.* 2006). Also the active extrusion of each consecutive nucleotide is unlikely to occur given this result. Consequently, the hOGG1 enzyme must somehow be able to recognize intrahelical oxoG residues and facilitate their extrusion out of the confinements of the DNA helix.

1.4 Nucleotide excision repair in Escherichia coli

1.4.1 Introduction

Nucleotide excision repair is characterized by the feature that it is capable of recognizing various types of damage. Not only does it recognize and repair the UV lesions that were originally used to identify the individual proteins of the bacterial NER system (Howard-Flanders *et al.* 1966), but it also recognizes for example other intrastrand crosslinks (Beck *et al.* 1985; Sancar *et al*. 1985), abasic sites (Snowden *et al.* 1990), bulky lesions (Moolenaar *et al.* 2002; Gomez-Pinto *et al.* 2004; DellaVecchia *et al.* 2004) and protein-DNA crosslinks (Minko *et al.* 2002).

In *E. coli* nucleotide excision repair is mediated by three NER-specific proteins: the damage recognition proteins UvrA and UvrB, and a NER-specific nuclease, UvrC. These proteins were first discovered by isolation of *E. coli* mutants on which plaque formation is reduced after a phage (T1 or λ) is UV irradiated (Howard-Flanders *et al.* 1966). All these mutants mapped at the loci now designated *uvrA, uvrB* and *uvrC*. Nucleotide excision repair, mediated by the corresponding proteins, exists in bacteria and a subset of archaea. Much later another protein has been characterized that is involved in NER in *E. coli* and several other bacteria, Cho (Moolenaar *et al.* 2002). This protein corresponds to a second NER-specific nuclease.

The NER mechanism in *E. coli* (Fig. 11) is a multi-step ATP dependent reaction starting with the formation of the $UvrA_2B_2$-complex in solution (Verhoeven *et al.* 2002a). The $UvrA_2B_2$-complex then binds to the DNA in search of a damaged site and wraps the

DNA around one of the UvrB subunits (Verhoeven *et al.* 2001). Once a damaged site is located both UvrA subunits are lost from the DNA, resulting in the formation of the $UvrB_2$-DNA preincision complex where one of the subunits is stably associated with the damaged site, while the other is more loosely bound (Moolenaar *et al.* 2005). In this complex the DNA remains wrapped around the damage-bound UvrB protein (Verhoeven *et al.* 2001). The more loosely bound UvrB subunit is displaced by UvrC resulting in the incision proficient UvrBC-DNA complex (Verhoeven *et al.* 2002a; Moolenaar *et al.* 2005). In this complex UvrC first performs an incision at the 4th or 5th phosphodiester bond 3' to the lesion (Sancar and Rupp, 1983; Verhoeven *et al.* 2000), which is subsequently followed by an incision at the 8th phosphodiester bond at the 5' side of the lesion (Lin and Sancar, 1992). Often a second UvrA-independent 5' incision occurs 7 nucleotides further upstream, which has been shown to be the result of recognition of the nick made in the first 5' incision event (Moolenaar *et al.* 1998a). The resulting oligonucleotide containing the damage is then released together with UvrC by the action of UvrD (helicase II) (Orren *et al*. 1992). Finally, DNA synthesis and ligase close the resulting gap restoring the DNA to its undamaged state.

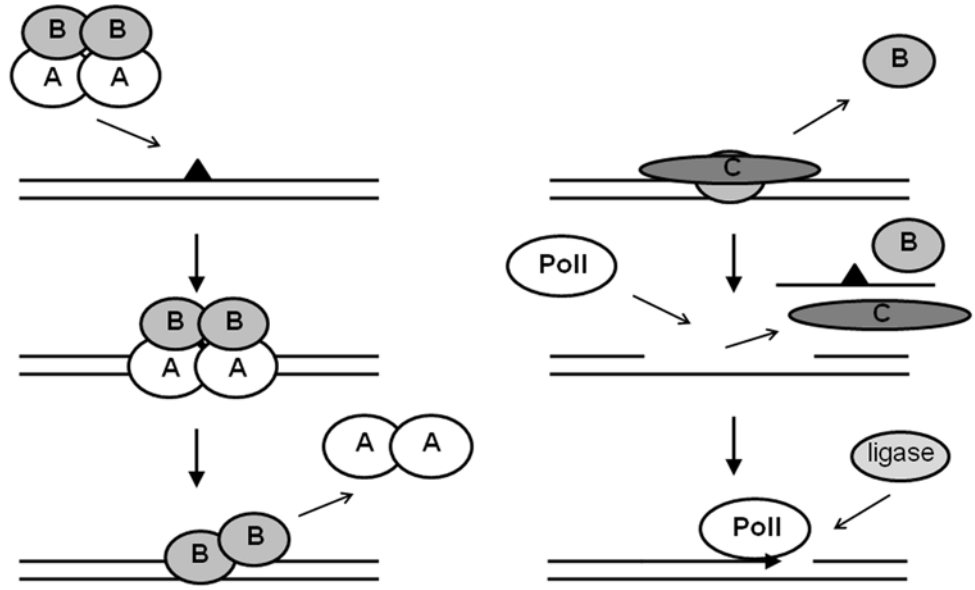

Fig. 11 Schematic representation of the bacterial nucleotide excision repair mechanism.

The three proteins involved in the steps leading to incision (UvrA, UvrB and UvrC) have been studied in great detail and crystal structures of the complete protein (UvrA and UvrB) or protein domains (UvrC) have become available.

1.4.3 The UvrA protein

UvrA is the largest protein of the *E. coli* NER system. It contains 940 amino acids and has a molecular mass of 110 kDa. The protein has been shown to bind preferentially to damaged DNA and thereby serves as the initial damage sensor of the prokaryotic NER system (Truglio *et al.* 2006a). In solution the UvrA protein forms a dimer which is believed to be its functional form (Myles *et al.* 1991).

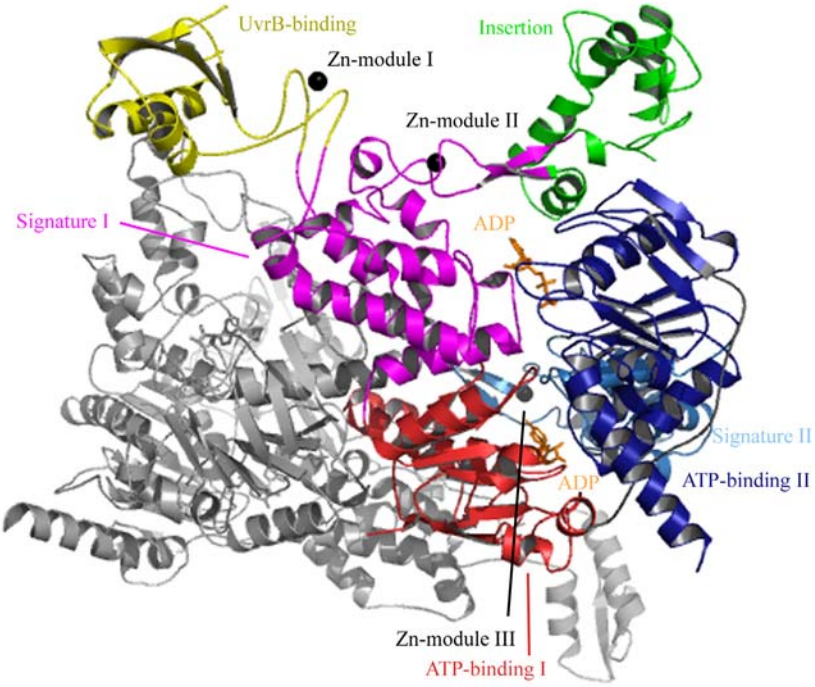

Fig. 12 Crystal structure of the UvrA dimer from *B. stearothermophilus* (pdb accession number 2R6F). In the structure the different domains of UvrA are indicated. NBD-I consists of ATP-binding domain I and Signature domain I, whereas NBD-II consists of ATP-binding domain II and Signature domain II. Zn-ions are shown as black spheres and the ADP molecules in orange. The second UvrA monomer is indicated in grey.

Each of the UvrA monomers contains two ATPase sites that belong to the ATP-binding cassette (ABC) superfamily of ATPases. Other members of this group include the DNA repair proteins Rad50 and MutS (Gorbalenya and Koonin, 1990). Proteins belonging to this group contain a 200-250 residue nucleotide binding domain (NBD), which can be sub-divided in a ATP-binding domain and a signature domain (Linton, 2007). These proteins exhibit composite nucleotide-binding sites, where the first NBD provides the ATP-binding domain and the second one provides the signature domain together forming an active nucleotide-binding site. This requires the presence of two NBDs in one functional unit to form two active nucleotide binding sites. In some proteins the two different NBDs are located

on two different polypeptide chains requiring the formation of a protein dimer to bring them together. These proteins bind the nucleotide at their dimeric interface and nucleotide-binding is thereby required to facilitate dimerization. In other, monomeric proteins the two NBDs are located within the same chain. UvrA contains two NBDs (NBD-I and II) per monomer resulting in a total of four composite nucleotide-binding sites in the dimer, making the protein a unique case.

Recently the crystal structure of UvrA from *Bacillus stearothermophilus* bound to ADP has been solved (Pakotiprapha *et al.* 2008; Fig. 12). This structure revealed UvrA in its dimeric form containing four ADP molecules, six Zn atoms and four water molecules. In figure 12 the dimeric complex of the UvrA protein is shown, clearly indicating the positions of the Zn^{2+} metal ions and ADP molecules. The composite nucleotide-binding sites are, in contrast to other dimeric proteins containing ABC-type ATPases, formed by NBDs that are present within the same polypeptide chain. This means that one of the nucleotide-binding sites is composed by the ATP-binding domain of NBD-I and the signature domain of NBD-II and the second site by ATP-binding domain II and signature domain I, all present in one UvrA monomer. In the structure all four nucleotide binding sites contain an ADP molecule but lack electron density directly adjacent to the β-phosphate normally occupied by Mg^{2+}. Electron density was however shown next to the α-phosphate. Since this is an unusual place for a Mg^{2+}-ion water was modeled at these positions, but the exact nature of the present compound is still unknown. A closer look at the dimer shows that, unlike for other structures of dimeric ABC ATPases (Obmolova *et al.* 2000; Higgins and Linton, 2004), the dimeric interface does not contain any bound nucleotides, a consequence of the intramolecular composition of the NBDs. This explains the previous observation that dimerization of UvrA is not strictly dependent on the presence of a nucleotide (Oh *et al.* 1989). However, the structure did show that ATP binding and hydrolysis might indirectly affect UvrA dimerization (Pakotiprapha *et al.* 2008).

Inserted in signature domain I each monomer contains two domains specific for UvrA. Mutational analysis revealed the first domain to be a UvrB binding domain responsible for association of UvrA to UvrB (Pakotiprapha *et al.* 2008). The domain contains several conserved, solvent-exposed residues. Some of these residues are charged whereas others are hydrophobic suggesting that the UvrA-UvrB interaction is both hydrophobic and charged in nature.

The second domain, referred to as the insertion domain, is of unknown function. Deletion of this domain did not induce an obvious phenotype, suggesting that it is not critical for UvrA function *in vitro*. Several bacterial species other than *E. co*li encode a UvrA homolog in which this domain is lacking (Goosen and Moolenaar, 2008). However, in addition all of these organisms do also express a UvrA protein that does contain this domain suggesting that the domain is important for the repair function of UvrA.

The presence of three Zn atoms per UvrA monomer was surprising since sequence

analysis and biochemical studies suggested the presence of two Zn atoms per monomer, each located in a Zn finger domain (Doolittle *et al.* 1986; Navaratnam *et al.* 1989; Visse *et al.* 1993; Wang *et al.* 1994). However, none of the three Zn atoms shown in the crystal structure is present in a classical Zn finger. The first zinc atom (Zn module 1) is coordinated by Cys120, His123, Cys250 and Cys253 and is located between signature domain I and the UvrB-binding domain. The second zinc atom (Zn module 2) is coordinated by Cys274, Cys277, Cys404 and Cys407 and bridges signature domain I with the insertion domain. The third zinc atom (Zn module 3) is coordinated by Cys736, Cys739, Cys759 and Cys762 and connects the helical region of signature domain II to the dimer interface. The coordination of this Zn module is the only one that is in agreement with the previously published model (Doolittle *et al.* 1986). All three Zn modules are proposed to play a structural role.

In the study by Pakotiprapha *et al* (2008) biochemical experiments and amino acid conservation have indicated the DNA binding surface of the UvrA protein (Fig. 13, backside of Fig. 12). This surface contains several highly conserved positively charged residues and could interact with approximately 30 basepairs of B-DNA which is in good agreement with DnaseI footprint studies. In these studies it was shown that binding of UvrA to the DNA protects a stretch of 33 basepairs from DnaseI-mediated digestion (Van Houten *et al.* 1987).

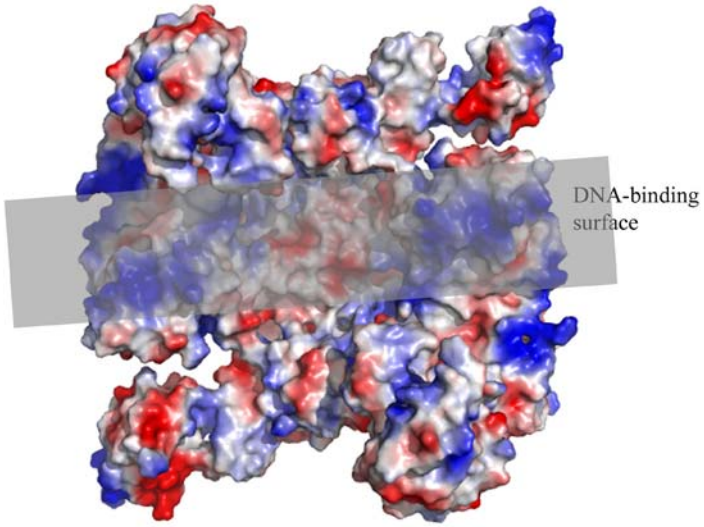

Fig. 13 Structure of the UvrA protein indicating the predicted DNA-binding surface. Blue regions correspond to positively charged regions whereas red regions are negatively charged.

1.4.3.1 Role of ATPase activity still largely unknown

In solution the UvrA protein forms a dimeric $UvrA_2$-complex. UvrA dilution studies have shown that lowering the UvrA concentration promotes monomer formation and causes a drastic decrease in UvrA ATPase activity (Myles *et al.* 1991). These experiments implicate that the UvrA dimer is the functional form of the protein. Dimerization occurs in the absence of cofactor (Wagner, manuscript in preparation) and can be stimulated by the presence of ATP

(Oh *et al.* 1989). Initially it was found that ATPγS further enhances dimerization indicating that ATP hydrolysis leads to a disruption of the dimerization contacts. However, recently it was found using single molecule analysis that dimerization in the presence of ATP is higher than with ATPγS (Wagner, manuscript in preparation).

The previously described ATPase sites of the UvrA protein not only bind and hydrolyze ATP, but also GTP in a DNA-independent fashion (Caron and Grossman, 1988). Addition of UvrB on the other hand was shown to result in a decrease of ATPase activity by UvrA.

In the presence of ATP the affinity of UvrA for damaged DNA is 10^3-10^4-fold higher than that for undamaged DNA, showing that UvrA can detect the presence of a lesion on its own (i.e. in the absence of UvrB). With ATPγS the affinity of UvrA for undamaged DNA is greatly enhanced (Seeberg and Steinum, 1982). It was therefore proposed that ATP hydrolysis induces dissociation of UvrA from undamaged DNA.

Mutations in either the N-terminal (residue Lys37) or C-terminal (residue Lys646) ATPase domain, making the site incapable of ATP hydrolysis but still enabling ATP binding, have shown to render the UvrA protein impaired in discriminating damaged from undamaged DNA (Thiagalingam and Grossman, 1991). This implies that ATP hydrolysis by both ATPase domains per UvrA monomer is required for damage recognition.

These data combined with the observation that UvrA preferentially binds ssDNA over dsDNA (Seeberg and Steinum, 1982) might suggest that UvrA is capable of recognizing a damage due to the induction of a small stretch of ssDNA by the damage. However, not all lesions give rise to ssDNA which would mean that UvrA also contains other damage recognition determinants. After damage recognition the UvrA protein hands over the DNA to UvrB which will initiate the second step in damage recognition.

1.4.4 The UvrB protein

The *E. coli* UvrB protein contains 673 amino acids and has a molecular mass of 76 kDa. It is considered as the central DNA damage recognition protein in bacterial NER as it interacts with all other components of the repair system: UvrA, UvrC, Cho, UvrD (helicase II), PolI and DNA (reviewed in Van Houten *et al.* 2005 and Truglio *et al.* 2006a). UvrB contains six helicase motifs (I-VI) that have been found in a superfamily of putative DNA and RNA helicases. Of these domains domain I and II correspond to the Walker A and B motifs making up a functional ATPase site.

The crystal structures of UvrB from *Bacillus caldotenax* (Theis *et al.* 1999) and *Thermus thermophilus* (Machius *et al.* 1999; Nakagawa *et al.* 1999) have been solved. The UvrB protein consists of five domains (domains 1a, 1b, 2, 3 and 4) of which domain 1 to 3 are visible in the various crystal structures. The C-terminal domain 4 is highly disordered and is therefore missing from these structures. However, the structure of this domain from *Escherichia coli* has been studied separately by both X-ray crystallography (Sohi *et al.* 2000)

and NMR spectroscopy (Alexandrovich *et al.* 1999). These studies show that this domain can form dimers in solution, strongly suggesting that also the full-length protein dimerizes in solution. The structure of UvrB from *Bacillus caldotenax* in the presence of ATP is depicted in Fig. 14A showing the domain build-up of the UvrB protein (PDB entry 1D9Z). The missing part of the crystal structure, the C-terminal domain of UvrB, is depicted in Fig. 14B in the dimeric form (PDB entry 1OOJ).

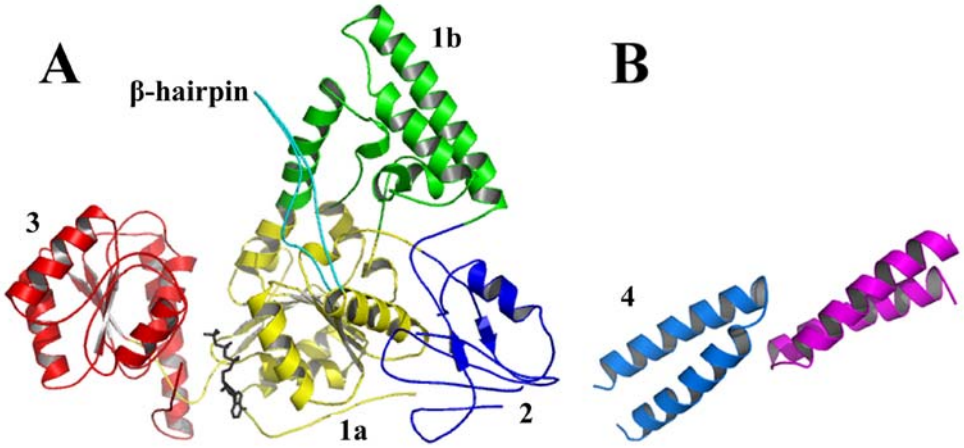

Fig. 14 Crystal structure of UvrB. (A) Crystal structure of monomeric UvrB from *Bacillus caldotenax*. Domains 1a (yellow), 1b (green), 2 (blue) and 3 (red) are indicated (pdb accession number 1D9Z). The β-hairpin moiety is shown in cyan and the bound ATP in black. The C-terminal domain 4 is missing. (B) The missing C-terminal part of the structure of (A), crystallized from *Escherichia coli* (pdb accession number 1QOJ). The domain forms a dimer under crystallization conditions.

The helicase motifs are located in domains 1a and 3 with the ATP wedged between these two domains. The fold of these domains is similar to the domains of other helicases belonging to the same family, PcrA (Velankar *et al.* 1999), Rep (Korolev *et al.* 1997), and NS3 (Kim *et al.* 1998). Structures of helicases in the absence or presence of cofactor have shown divergent orientations of the domains that bind ATP indicating that ATP hydrolysis is coupled to domain motion which is subsequently used for DNA unwinding (Velankar *et al.* 1999). The similarity to certain helicases suggests that also for UvrB ATPase-mediated domain motions can lead to DNA unwinding. However, the protein does not possess a 'true' helicase activity although the UvrAB complex does displace small oligonucleotides not larger than 22 nucleotides and depending on the DNA's melting temperature (Gordienko and Rupp, 1997; Moolenaar *et al.* 1994). This activity is also referred to as UvrB's strand-destabilizing activity.

Domain 2 of UvrB, which is located between domains 1a and 1b, shows clear sequence homology to a domain present in the transcription repair coupling factor (Selby and Sancar, 1993), the Mfd protein (also described in paragraph 1.4.7). The association of

both proteins with UvrA suggested the domain's involvement in an interaction with UvrA. This hypothesis was later verified by a study showing that a fusion protein of the maltose-binding protein and residues 116-251 from *E. coli* UvrB (domain 2) still binds to UvrA (Hsu *et al.* 1995). In addition, mutation of conserved residues in domain 2 impaired UvrAB complex formation (Truglio *et al.* 2004). Upon determination of the crystal structure of Mfd it was found that structural homology not only encompasses sequence-homologous domain 2 of UvrA but also parts of domains 1a and 1b (Assenmacher *et al.* 2006; Deaconescu *et al.* 2006).

Initially co-sedimentation studies indicated that UvrA and UvrB interact only in the presence of ATP, but not in the presence of ADP, ATPγS or without cofactor (Orren and Sancar 1989). However FRET analysis revealed that not only in the presence of ATP but also in the presence of ATPγS an interaction is established (Chapter 2), suggesting that ATP-binding by UvrA and/or UvrB is sufficient for the UvrAB interaction. Moreover, filter-binding assays have shown that salt-resistant protein-DNA complexes, characteristic of UvrB-containing complexes, can be formed in the presence of GTP (Caron and Grossman 1988). Since competition experiments have shown that UvrB cannot accommodate GTP into its nucleotide-binding site (Caron and Grossman, 1988), whereas UvrA can, this means that the presence of a cofactor in UvrA alone is sufficient for UvrA and UvrB to interact.

As discussed before, the crystal structure of the C-terminal domain 4 has been solved separately (Alexandrovich *et al.* 1999; Sohi *et al.* 2000; Fig. 14B) and reveals that this domain adopts a helix-loop-helix conformation stabilized by hydrophobic interactions and salt bridges between the two helices. Two domain 4 molecules form a dimer in solution by interacting in a head-to-head fashion and establishing both hydrophobic and ionic interactions (Fig. 14B). Oligomerization analysis by Hildebrand and Grossman (1999) has shown that the C-terminal domain 4 is required for optimal dimerization of UvrB in solution since a mutant UvrB protein lacking this domain lowers dimerization. Also on the DNA this domain is involved in stabilizing the dimer of UvrB (Verhoeven *et al.* 2002a). Dimer formation however does still occur albeit in reduced quantity, indicating that other as yet unidentified domains of the UvrB protein are involved in the contact between two UvrB's as well.

Domain 4 has also been shown to be an important UvrC-binding domain (Hsu *et al.* 1995) and deletion of this domain (resulting in UvrB*) abolishes 3' incision by UvrC (Moolenaar *et al.* 1995). 5' UvrC incision or Cho incision (Moolenaar *et al.* 2002) on the other hand can still take place indicating that an interaction with domain 4 is not required for these incision events.

Finally domain 4 might also be involved in the interaction with UvrA. A fusion of the maltose-binding protein with domain 4 and part of domain 3 from *E. coli* UvrB was shown to bind to UvrA (Hsu *et al.* 1995). Crystal structure analysis has shown the domain 4 dimer to pack against domain 3 and to extend across the entrance to the ATP-binding site towards domain 2 (Eryilmaz *et al.* 2006). UvrB mutants lacking this C-terminal domain

(UvrB*) exhibit a UvrA- and DNA-independent cryptic ATPase activity whereas wildtype UvrB does not hydrolyze ATP under these conditions (Caron and Grossman 1988; Wang *et al.* 2006). Addition of UvrA and DNA to wildtype UvrB however induces its ATPase activity suggesting that interaction of UvrA with domain 4 regulates the ATPase activity of UvrB by repositioning this domain and thereby altering the ATP binding pocket (Wang *et al.* 2006).

One of the most striking features of the UvrB structure is its β-hairpin structure that extrudes from domain 1a (Fig. 14A). Deletion of the tip of the hairpin results in a protein that is no longer capable of damage recognition (Skorvaga *et al*. 2002). The hairpin is rich in conserved hydrophobic residues both at the tip and bottom of the structure and mutation of these residues has underlined its importance in damage recognition (Moolenaar *et al.* 2001, 2005; Skorvaga *et al.* 2004). Both hydrophobic and charged residues in the tip of the hairpin were shown to interact with domain 1b of the protein. This interaction leaves a space between the hairpin and the body of the UvrB protein big enough to accommodate a single strand of DNA. From the structure and homology to other hairpin-containing DNA binding proteins a padlock model was proposed (Theis *et al.* 1999). In this model the hairpin inserts in between the DNA strands, where it interacts with residues at the bottom of the hairpin. The interaction between the tip and domain 1b strongly locks one of the DNA strands behind the hairpin leading to stable binding as long as the DNA is flanked by double stranded DNA on both sides. Mutation of residue Tyr96, located at the bottom of the hairpin, results in a protein that is impaired in preincision complex formation and therefore is incapable of repair (Skorvaga *et al.* 2004). Mutation of hairpin-residues Tyr92 and Tyr93 into alanine residues on the other hand resulted in a protein that binds more efficiently to undamaged DNA. However the mutant protein can still form UvrB-DNA complexes at a damaged site indicating that the residues are not required for the damage recognition process but prevent association of the protein with undamaged DNA. Further analysis revealed that these residues sterically hinder binding of undamaged DNA by clashing with the undamaged nucleotides (Moolenaar *et al.* 2001, 2005).

Insertion of the β-hairpin between the two DNA strands was confirmed by the determination of a crystal structure of UvrB bound to a small DNA loop where it recognizes a single-strand double-strand DNA junction (Truglio *et al.* 2006b, Fig. 15).

Indeed one of the DNA strands can be seen to pass behind the β-hairpin where there is just enough space to accommodate this strand. The nature of the UvrB-DNA interaction and its homology to the previously mentioned DNA helicases suggest that UvrB is able to translocate along the inner DNA strand in a 3’ → 5’ direction. Extrapolation of the complementary strand shows that it will pass in front of the hairpin. In the structure it was also shown that one of the nucleotides rotates behind the β-hairpin and is incorporated into a tight planar pocket of the UvrB protein where it stacks on Phe249 (Fig. 15 B,C). Whether it is the damaged DNA strand that is located behind the β-hairpin structure could however not be determined in this study.

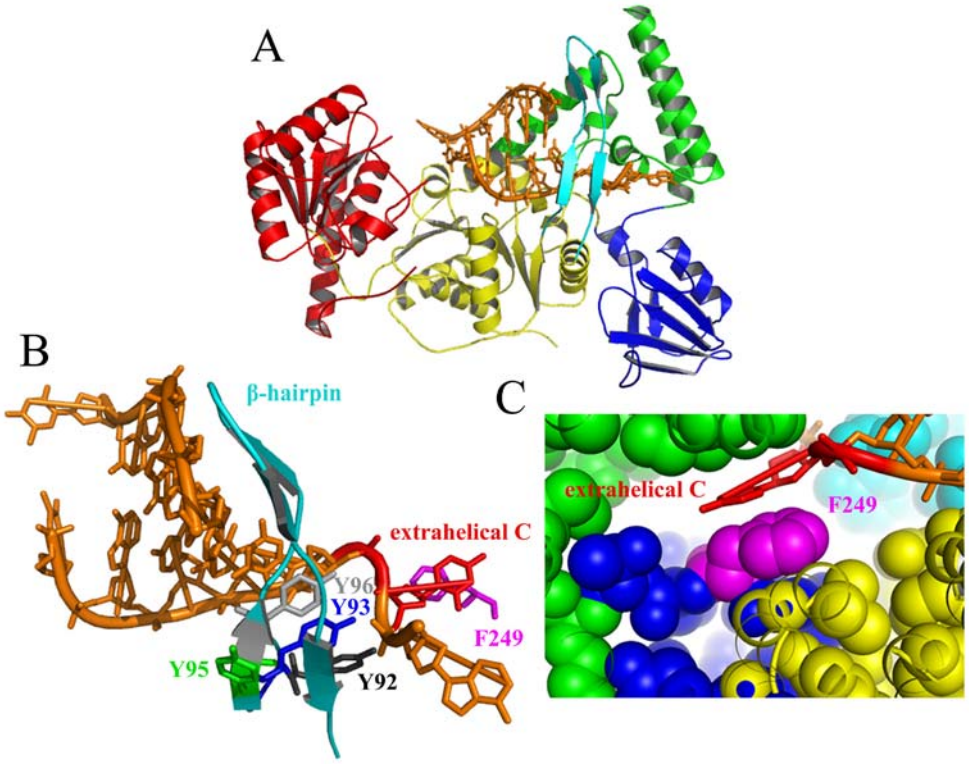

Fig. 15 Crystal structure of the UvrB protein bound to a small DNA loop (pdb accession number 2FDC). (A) Crystal structure of the UvrB protein bound to DNA. Colors are identical to figure 14 and denote the domain organization of UvrB. DNA is indicated in orange. (B) Zoom-in representation of the structure showing the β-hairpin (cyan) and residues Tyr92 (black), Tyr93 (blue), Tyr95 (green), Tyr96 (gray) and Tyr249 (magenta). The extrahelical cytosine is indicated in red. (C) Space-filling representation showing the planar pocket of the UvrB protein. The extrahelical cytosine (red) undergoes a stacking interaction with Tyr249 (magenta).

A co-crystal structure of UvrB bound to a pentanucleotide containing a single fluorescein-adducted thymine did shed some light on the nature of the DNA strand located behind the β-hairpin (Waters *et al.* 2006). In this structure the damage-containing oligonucleotide is located behind the β-hairpin suggesting that also on damaged dsDNA it is the damaged strand that threads behind the hairpin of UvrB. This would then suggest a possible function for the small planar pocket of the UvrB protein. Possibly during translocation along the DNA each consecutive nucleotide is flipped out of the DNA into the protein pocket. This process would then stall at the site of damage since the dimensions of the pocket only allow incorporation of a planar nucleotide and not of a base containing a crosslink or chemical adduct. This would imply an involvement of the pocket of the UvrB protein in the damage recognition process.

However a closer look at the structure of Waters *et al* (2006) did not reveal an

extrahelical nucleotide located in the pocket of the UvrB protein stacking on Phe249. The phenylalanine in this structure even adopts a different conformation, suggesting that the shape of the pocket is induced by the nucleotide that is inserted (Fig. 16).

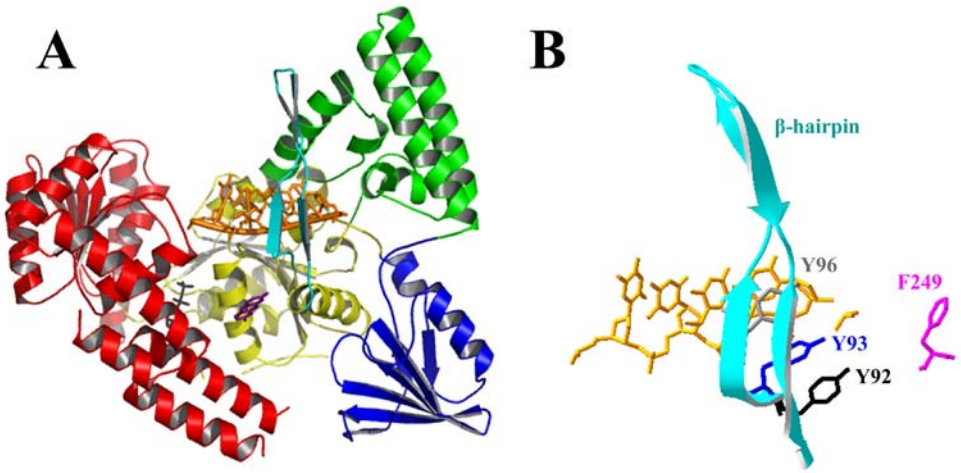

Fig. 16 Crystal structure of the UvrB protein bound to a single-strand pentanucleotide containing a centrally located fluorescein modification (pdb accession number 2NMV). (A) Crystal structure of the entire UvrB protein bound to DNA. Colors are identical to figure 14 and denote the domain organization of UvrB. DNA is indicated in orange and ADP in black. The triple-ring system of the fluorescein moiety is indicated in purple. No density for the rest of the fluorescein moiety was observed. (B) Zoom-in representation of the structure showing the β-hairpin (cyan) and residues Tyr92 (black), Tyr93 (blue), Tyr96 (gray) and Phe249 (magenta).

We will show in chapter 4 that the pocket of the UvrB protein plays an important role in preventing binding to undamaged DNA. From the data presented in this chapter we derived an alternative model where stacking of an undamaged nucleotide in the pocket of UvrB positions the DNA behind the hairpin in such a way that residues Tyr92 and Tyr93 clash with the 3' flanking undamaged nucleotides. On damaged DNA however the presence of the lesion blocks translocation because this lesion cannot pass behind the hairpin. Positioning of the DNA behind the hairpin in this arrested conformation subsequently prevents incorporation of a nucleotide in the pocket. As a consequence the undamaged nucleotides 3' to the lesion will have more conformational freedom and under the influence of Tyr92 and Tyr93 they will be able to occupy a position where they will no longer clash with these aromatic residues. This results in stable binding and enables UvrC-mediated incision.

As mentioned previously the UvrB protein alone does not hydrolyze ATP, but in the presence of UvrA the protein exhibits a DNA-dependent ATPase activity. This ATPase activity is associated with a limited DNA unwinding activity (Oh and Grossman 1987) which was shown to be important for loading of UvrB onto the site of damage (Moolenaar *et al.* 1994). Most likely the observed unwinding activity is the result of insertion of the β-hairpin between the two DNA strands. Subsequent rounds of ATP hydrolysis might then lead to

limited translocation along the DNA until this translocation is blocked by the damage. After binding of the damage renewed ATP uptake is required to result in a preincision complex that can be incised by UvrC (Moolenaar *et al.* 2000). In this complex a DnaseI hypersensitive site appears 5' to the lesion suggesting that the DNA is highly distorted at this position (Van Houten *et al.* 1987).

Combining structural and biochemical data the following model for damage recognition by the bacterial NER-proteins can be presented. First of all the $UvrA_2B_2$-complex binds to the DNA and wraps it around one of the UvrB subunits in an ATP-dependent manner (Verhoeven *et al.* 2001, 2002a). In this complex the presence of a possible lesion is initially detected by the UvrA protein, possibly by recognizing a stretch of ssDNA induced by the damage. Wrapping of the DNA around UvrB might promote DNA unwinding thereby assisting damage recognition by UvrA. This initial damage recognition by UvrA brings the UvrB protein in close proximity to the lesion in order for UvrB to verify the presence of a damage. Further opening of the DNA by UvrA could facilitate the insertion of the β-hairpin of UvrB in between the two DNA strands resulting in a hand-off of the DNA from UvrA to UvrB. Now the complex of UvrB bound to the DNA is in a conformation similar to that of the NS3 RNA helicase (Kim *et al.* 1998) and subsequent cycles of ATP hydrolysis will lead to (limited) translocation along the DNA to properly position the protein at the site of damage. The presence of two copies of the UvrB protein might serve to alternately scan both strands for the presence of a lesion. At the site of damage further translocation is arrested by preventing the damage from passing behind the hairpin. Finally this arrested complex of UvrB on the damaged site takes up a new ATP molecule to induce a specific DNA conformation that can be incised by UvrC.

1.4.5 The UvrC protein

The *E. coli* UvrC protein is composed of 610 amino acids and has a molecular weight of 67 kDa. The protein is the nuclease of the *E. coli* NER system and can interact with both UvrB and the DNA to form the UvrBC-DNA incision complex (Orren and Sancar, 1989, Tang *et al.* 2001). It contains two endonuclease active sites, a cysteine-rich region, a UvrB-interaction domain, and a tandem Helix-hairpin-Helix (HhH_2) domain (Fig. 17).

One of the two endonuclease active sites is located in the N-terminal half of the protein and is responsible for the 3' incision whereas the other one is located in the C-terminal half and is responsible for the 5' incision and extra 5' incision. The catalytic domain in the N-terminal half of the UvrC protein from both *Thermotoga maritima* and *Bacillus caldotenax* has been crystallized and shows homology with the catalytic domain of the GIY-YIG family of intron-encoded endonucleases (Truglio *et al.* 2005). This catalytic domain makes use of a novel one-metal mechanism to cleave the DNA backbone. The metal-ion is octahedrally coordinated and is directly bound to only a single residue, Glu76. In order to cleave the phosphodiester backbone three chemical entities are required: 1. a general base to activate the

nucleophile for attack of the 5' phosphate, in this case a conserved tyrosine residue (Tyr29). This tyrosine is located in a region of high positive charge thereby lowering its pKa and enabling it to serve as a base. 2. a general acid to protonate the 3' leaving group, in this case a metal-coordinated water molecule and 3. a Lewis acid to stabilize the pentacovalent phosphoanion transition state, the metal. Furthermore in this study it has been shown that the catalytic domain by itself is not able to bind or incise the DNA. This shows that interactions of other domains of UvrC with UvrB and/or the DNA are required to line up the catalytic site to its required incision position.

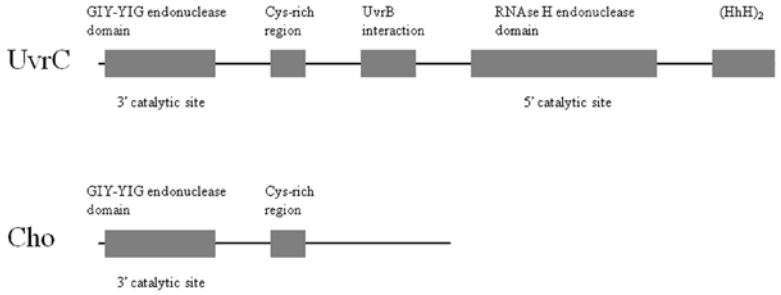

Fig. 17 Domain architecture and alignment of the UvrC and Cho proteins from *Escherichia coli*.

The UvrB-interaction domain of UvrC exhibits homology with the C-terminal domain 4 of UvrB (Fig. 14B) and is involved in interaction with this domain of UvrB (Moolenaar *et al*. 1995, 1997). Since the same domain of UvrB is also involved in dimerization this means that in order for UvrC to interact with UvrB, the second UvrB subunit has to be released. This specific UvrBC interaction has been shown to be critical for 3' incision but not for 5' incision (Moolenaar *et al.* 1998b).

Furthermore the N-terminal half of the protein contains a region with multiple conserved cysteines with unknown function, but which are likely to be also involved in the 3' incision event.

The structure of the C-terminal catalytic domain of *Thermotoga maritima* UvrC has also been solved recently (Karakas *et al.* 2007). This domain shares structural homology with RNaseH despite a lack of sequence homology and it contains an uncommon catalytic DDH triad. The same catalytic triad has only been found in the Argonaute protein (Rivas *et al.* 2005), a protein involved in small RNA-induced gene silencing. In *E. coli* UvrC the catalytic site is composed of two aspartates and one histidine residue. The triad was shown to coordinate a single manganese ion, but it was predicted by comparison to the structure of theTn5-transposase (Lovell *et al.* 2002) and RnaseH (Nowotny *et al.* 2005) that the functional form of the protein contains two ions. A mechanism for DNA cleavage was proposed in which the first metal activates the attacking water for nucleophilic attack and the other stabilizes the negative charge formed on the pentacovalent intermediate. Both ions act as Lewis acids and

stabilize the expected pentacovalent transition state (Steitz and Steitz 1993).

In addition the structure of the C-terminal domain exhibits a tandem helix-hairpin-helix domain that is implicated in DNA binding. The two HhH motifs together form one functional HhH_2 domain whose structure has also been solved separately by NMR (Singh *et al.* 2002). The domain binds to a single-strand-double-strand DNA junction, but has a strong preference for dsDNA containing an unpaired bubble region of six or more nucleotides. Such a structure might mimic the DNA in the UvrB-DNA preincision complex (Truglio *et al.* 2006a).

The HhH_2 domain has been shown to be involved in either 3' or 5' incision events depending on the sequence environment at the site of damage (Verhoeven *et al.* 2002b). Interaction of the HhH_2 domain with the DNA has been proposed to stabilize specific DNA structures required for the two incisions thereby contributing to the flexibility of the UvrABC repair system. The same HhH_2 motif is also present in other DNA repair proteins like ERCC1, XPF (Tripsianes *et al.* 2005), AlkA (Hollis *et al.* 2000) and hOGG1 (Bruner *et al.* 2000) (see also paragraph 1.6.2).

1.4.6 The Cho protein

In addition to UvrC, *E. coli* and several other bacterial species express a second NER-specific endonuclease (Moolenaar *et al.* 2002). This protein is highly homologous to the N-terminal half of UvrC including the catalytic site and the cysteine-rich region (Fig. 17) and was for that reason named UvrC-homolog (Cho). It was found that Cho is only capable of performing an incision at the 3' side of the lesion which is consistent with its homology to the catalytic N-terminal half of UvrC. Interestingly, the Cho protein has been shown to induce 3' incision in a complex with a truncated UvrB lacking the C-terminal domain 4 (Moolenaar *et al.* 2002), showing that in contrast to UvrC, Cho does not need domain 4 of UvrB. Alignment of Cho with UvrC indeed indicated that the UvrC domain interacting with domain 4 of UvrB is not present in the Cho protein (Fig. 17). The exact domains involved in the interaction between UvrB and Cho are still unknown, but the different mode of binding might explain why Cho incises the DNA four nucleotides further away from the damage compared to UvrC (Moolenaar *et al.* 2002). The incision efficiencies of both endonucleases vary on different DNA substrates. Certain DNA substrates are more efficiently incised by UvrC whereas others are more efficiently incised by Cho. This suggests that the function of Cho in nucleotide excision repair is to repair lesions that are not (efficiently) incised by UvrC. After 3' incision by Cho however incision on the 5' side of the damage is still required to excise the damage from the DNA. Indeed it has been shown that substrates incised by Cho can be further processed by UvrC (Moolenaar *et al.* 2002). In some bacteria the Cho protein additionally contains a domain that shows homology to the Epsilon proofreading subunit of DNA polymerase III. These proteins therefore combine 3' incision with a 3'→5' exonuclease activity possibly enabling them to remove the damaged nucleotide without the

need for UvrC.

1.4.7 Transcription-coupled DNA repair in E. coli

It has been shown that NER-mediated repair of UV-induced CPD is faster in the transcribed strand than in the non-transcribed strand (Bohr *et al.* 1985; Mellon and Hanawalt, 1989). The process through which this occurs is referred to as transcription-coupled DNA repair. This 'pathway' starts when RNA polymerase (RNAP) is blocked by the presence of a non-coding DNA lesion. The resulting complex is subsequently recognized by the Mfd protein (also known as TRCF, transcription repair coupling factor) (Selby and Sancar, 1993). This protein makes use of ATP-dependent translocation upstream of the transcription bubble to remove the stalled RNAP (Park *et al.* 2002; Smith *et al.* 2007). Mfd serves a dual function in this pathway: release of RNAP together with its unfinished transcript and recruitment of the DNA excision repair proteins to the site of the lesion (Selby and Sancar, 1990, 1993).

Mfd was originally discovered genetically due to its effect on mutagenesis hence its name: mutation frequency decline (Witkin, 1966). Mfd's sequence homology with domain 2 of UvrB suggests binding of UvrA via this homologous region (Deaconescu *et al.* 2006). After recruitment of UvrA by Mfd, the remainder of the NER machinery will be assembled and repair will take place as described previously. Transcription-coupled repair (TCR) is more efficient than the transcription-independent global genome repair (GGR) discussed before since scanning the DNA for damages by a translocating RNA polymerase is expected to be much faster than scanning by the UvrAB complex which has only very limited translocation activity.

1.5 Eukaryotic nucleotide excision repair

1.5.1 Introduction

In analogy with prokaryotes also eukaryotes contain a NER system. Defects in this NER process in humans have been shown to lead to Xeroderma Pigmentosum (XP), a disorder rendering its patient highly sensitive to UV-light and highly susceptible for skin cancer. The process of NER in eukaryotes is far more complex than in bacteria but the underlying mechanisms and damage repertoire are mostly the same. It encompasses the combined action of over 30 different proteins (Aboussekhra *et al.*1995; Araujo *et al.* 2000), which can be divided in proteins involved in damage recognition and DNA incisions, and proteins involved in repair synthesis and ligation.

The eukaryotic NER system has been studied in a wide variety of organisms including human cells and yeast (*Saccharomyces cerevisiae*). In these two organisms the mechanisms of NER have shown to be highly similar. In humans, mutations in NER proteins may not only cause XP but can also result in two other autosomal recessive disorders: Cockayne Syndrome (CS) and trichothiodystrophy (TTD). These disorders are not always related to repair, but

also to transcription.

Table 1. Nomenclature of the NER factors required for the damage recognition and incisions of NER in humans.

Human NER factors	*Subunits*
XPA•RPA	XPA, RPA1, RPA2 and RPA3
XPC•HR23B	XPC and HR23B
UV-DDB	DDB1 and DDB2 (XPE)
TFIIH	XPB, XPD, p34, p44, p52, p62,TTDA, cdk7, cyclin H and MAT1
XPF•ERCC1	XPF and ERCC1
XPG	

1.5.2 The initial damage recognition step

The minimal set of proteins to confer NER-mediated incision on damaged DNA *in vitro* consists of multiple repair factors (Table 1). These are the XPA•RPA complex, transcription factor IIH (TFIIH, containing multiple subunits including both XPB and XPD), XPC•hHR23B, XPF•ERCC1 and XPG (Mu *et al.* 1996). Of these factors three proteins or protein complexes have been shown to be capable of binding damage-specifically to DNA implicating their involvement in damage recognition: XPA (Jones and Wood, 1993; Asashina *et al.* 1994), XPC (Reardon *et al.* 1996) and the RPA complex (Clugston *et al.* 1992; Burns *et al.* 1996). Also the TFIIH subunits XPB (Fan *et al.* 2006) and XPD (Reardon and Sancar, 2002) have been implicated in discriminating damaged from undamaged DNA. However until now there has been some debate as to which of the factors binds damage first.

In the "XPC first" model it is the XPC•hHR23B complex that first recognizes the damage and is responsible for the subsequent recruitment of the remaining NER factors. It was shown that the centrosomal protein centrin2 also interacts with the XPC•hHR23B complex and thereby stimulates incision of UV-damaged DNA (Nishi *et al.* 2005). However this interaction is not an absolute requirement for incision. *In vitro* the "XPC first" model is supported by a study of Sugasawa *et al.* (1998) showing that pre-incubation of a DNA substrate containing N-acetoxy-2-acetyl-aminofluorene (AAF)-induced adducts in the presence of XPC•hHR23B resulted in an increase in repair kinetics compared to a DNA substrate pre-incubated with XPA. Further support for this model came from the observation that the presence of XPC is not required for the process of transcription-coupled repair (Venema *et al.* 1990). In this pathway stalling of the RNA polymerase at UV-induced DNA lesions serves as the first damage detection signal, rendering the presence of the initial damage sensor (in this case XPC) superfluous. Reconstitution experiments demonstrated that XPC and TFIIH associate with cisplatin-damaged DNA in the absence of ATP, whereas for the association of XPA ATP is essential (Riedl *et al.* 2003). Probably the ATPase activity of

the TFIIH complex is required for XPA recruitment thereby giving another indication that the "XPC first" model is the right one.

However, in contrast to these experiments Wakasugi and Sancar (1999) showed that repair of a DNA fragment containing a single (6-4) photoproduct occurred 5-fold faster for reactions in which the DNA was preincubated with XPA and RPA as compared to reactions where XPC was included first. From these results it was concluded that the XPA•RPA complex is the initial damage recognition factor, supporting an "XPA first" model. One of the differences between this study and the previous study by Sugasawa *et al.* (1998) is the nature of the damage. It might therefore be possible that the order of recruitment in the damage recognition step is not fixed but might vary for different lesions. In this light it is also noteworthy that the XPC•hHR23B complex is not required for incision on a cholesterol-modified DNA substrate (Mu *et al.* 1996), reminiscent of transcription-coupled repair in a XP-C mutant cell line (Venema *et al.* 1990). Also when a lesion is located adjacent to a bubble (mimicking a transcription bubble) repair can take place independent of the presence of XPC (Mu and Sancar, 1997). These two DNA alterations probably induce a similar unwound structure in the DNA that is normally induced by the XPC protein. This is consistent with the high affinity of XPC for ssDNA suggesting its involvement in stabilizing an unwound DNA structure (Masutani *et al.* 1994; Reardon *et al.* 1996).

An *in vivo* study to elucidate the order of subunit association at a damaged site focused on nuclear trafficking in cells (Volker *et al.* 2001). In this study, human cells were locally UV-irradiated and the movement of repair factors was monitored by staining them with fluorescently tagged antibodies. These experiments revealed that XPC•hHR23B accumulates at repair foci even in cells lacking XPA whereas XPC is required for XPA accumulation. These results clearly showed that on UV-lesions the XPC•hHR23B complex is the initial damage recognizing factor.

The *in vivo* studies also enforced the previous *in vitro* studies by ending an ongoing debate as to whether the repair factors are recruited to the DNA consecutively or that they pre-assemble into one big complex, called the "repairosome". Accumulation of repair factors at different time points after irradiation disfavors the repairosome model and shows that factors are recruited sequentially.

While the order of recruitment of NER factors in the damage recognition step remains debated it is widely accepted that XPC does not recognize CPDs (Sugasawa *et al.* 2001; Fitch *et al.* 2003). Since also XPA and RPA are inefficient in recognizing CPD, another initial damage recognition factor is required for this type of lesion. This most likely is the UV-DDB complex, since specific binding of the complex to CPDs was shown to stimulate their repair (Reardon *et al.* 1993; Keeney *et al.* 1994; Hwang *et al*. 1998; Fitch *et al.* 2003). UV-DDB is a heterodimer consisting of p127 (DDB1) and p48 (DDB2) (Dualan *et al.* 1995). Mutations in the 48 kDa subunit are found in all known cases of XP complementation group E, directly implicating a link between these proteins and repair of UV-damage (Rapic-Otrin

et al. 2003).

1.5.3 The mechanism

The currently established model (Fig. 18) for the mechanism of NER starts with the formation of a heterotrimeric complex between XPC, hHR23B (Masutani *et al.* 1994; Volker et al. 2001) and centrin 2 (Araki *et al.* 2001). Both partners have been shown to bind directly to XPC and stabilize the protein by preventing polyubiquitylation (Ng *et al.* 2003), a process known to lead to degradation of proteins by the 26S proteasome (Ortolan *et al.* 2000). It is however unclear whether hHR23B is present upon binding of XPC to the lesion. As discussed, CPDs are poorly recognized by XPC (Sugasawa *et al.* 2001; Fitch *et al.* 2003). To facilitate repair of these highly ubiquitous lesions and to improve repair of other DNA damages (i.e. 6-4PPs) the UV-damaged DNA binding protein (UV-DDB/XPE) improves XPC association at sites of these lesions (Moser *et al.* 2005), either by stimulating or by stabilizing its binding.

In the next step the transcription factor TFIIH joins the complex which contains several protein subunits, including XPB and XPD (Drapkin and Reinberg, 1994). Of this complex, XPD uses its DNA-dependent ATPase and helicase activity respectively to open up the DNA helix around the lesion (Coin *et al.* 2007). This step is proposed to verify the presence of a lesion since its helicase activity stalls at a site of damage when it encounters a lesion (Dip *et al.* 2004). The open complex containing both the XPC•hHR23B and TFIIH complexes serves as a scaffold to facilitate the recruitment of XPA and RPA to the site of damage (Missura *et al.* 2001). RPA, the human single strand binding protein, has been shown to bind stably to the non-damaged DNA strand in this complex (de Laat *et al.* 1998a). This might also explain RPA's proposed damage recognition activity (Clugston *et al.* 1992; Burns *et al.* 1996) since at the site of damage the helical structure of the DNA might be disrupted facilitating strand separation and subsequent binding of RPA. By specifically binding to the undamaged DNA strand RPA is thought to direct the structure-specific endonucleases XPG and XPF-ERCC1 to the single-strand double-strand DNA junctions (de Laat *et al.* 1998a) in such a way that the non-damaged DNA strand is protected from nuclease activity. In this step recruitment of XPF•ERCC1 has been shown to be dependent on the presence of XPA whereas recruitment of XPG is not. XPG binds first and is thought to release the XPC-hHR23B complex from the DNA (Riedl *et al.* 2003). Subsequent recruitment of the XPF-ERCC1 complex enables dual incision, at the 3' side of the lesion by XPG (O'Donovan *et al.* 1994), and at the 5' side by the XPF•ERCC1 complex (Sijbers *et al.* 1996). Probably, after the excised oligonucleotide has been removed PCNA is loaded onto the DNA in a RFC-dependent manner (Kelman, 1997). Finally, repair synthesis followed by DNA ligation closes the gap resulting in a repaired stretch of DNA (Aboussekhra *et al.*1995; Araujo *et al.* 2000).

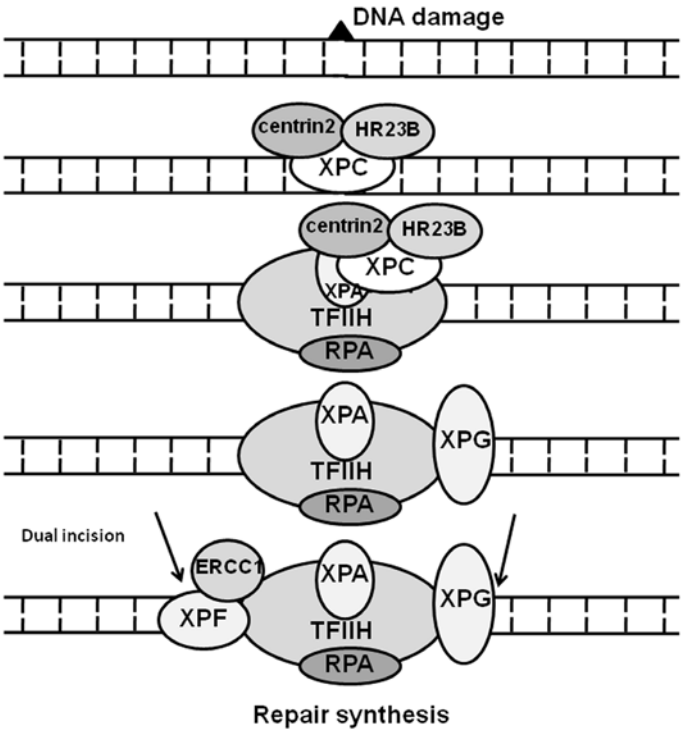

Fig. 18 Schematic representation of the core reaction of the human nucleotide excision repair mechanism.

1.5.4 Damage recognition in eukaryotic NER, the Rad4 protein

Recently, several crystal structures have become available of the *Saccharomyces cerevisiae* XPC ortholog Rad4 bound to a domain of Rad23, which is the yeast hHR23B counterpart (Min and Pavletich, 2007). These structures showed the protein complex both in the absence and in the presence of (damaged) DNA. The structure containing damaged DNA revealed Rad4 bound to a CPD lesion located in a stretch of three thymidine-thymidine mismatches known to increase specificity for the CPD lesion (Fig. 19).

The protein structures reveal the presence of an N-terminal domain containing a ~45 residue core transglutaminase fold (transglutaminase-homology domain, TGD) lacking however the catalytic triad that is characteristic of the transglutaminase superfamily. Furthermore, the Rad4 protein contains three structurally similar 50-90 residue domains characterized by the presence of a long β-hairpin structure (beta-hairpin domains 1-3, BHD1-3). The protein interacts with the DNA in two different parts: the TGD and BHD1 domains interact with a stretch of DNA comprising 11 basepairs of non-damaged DNA (bp 1-11) thereby explaining the substantial affinity of the XPC protein for undamaged DNA (Batty *et al.* 2000; Hey *et al.* 2002). BHD2 and BHD3 on the other hand bind to a 4-bp stretch containing the CPD lesion (bp 14-17). These interactions with the protein result in a highly kinked DNA conformation, where the DNA is bent by approximately 42^{o}.

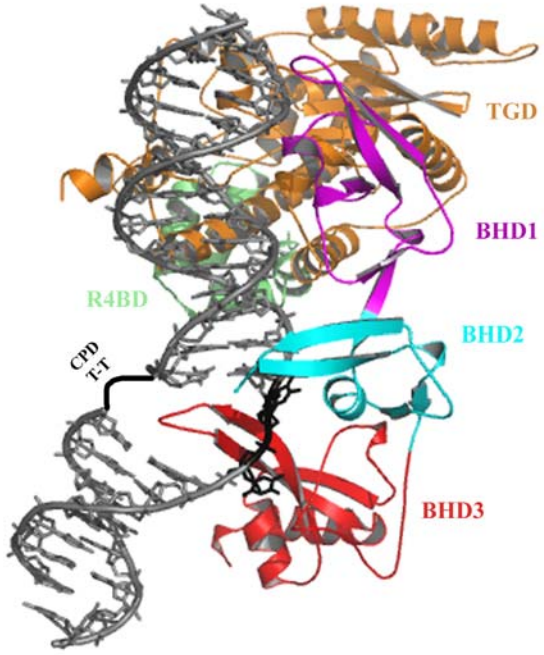

Fig. 19 Co-crystal structure of the *Saccharomyces cerevisiae* Rad4 protein (pdb accession number 2QSG). The transglutaminase domain (TGD; orange) and the three β-hairpin domains (BHD1-3; purple, cyan and red resp.) are indicated. The DNA is indicated in grey. The estranged nucleotides in the non-damaged DNA strand are shown in black. The Rad4 binding domain of the hHR23B protein is also visible (R4BD; green).

At the site of damage the β-hairpin structure of BHD3 inserts in between the two DNA strands resulting in both CPD bases becoming extrahelical but with no apparent interaction between these bases and the protein. This orientation of the hairpin slightly resembles that of the bacterial NER protein UvrB (Truglio *et al.* 2006b). However, in contrast to Rad4 the hairpin of UvrB completely protrudes through the two DNA strands thereby establishing interactions with the DNA by residues located at the bottom of the hairpin. Rad4 on the other hand engages the DNA with the tip of its hairpin.

A groove made up by BHD2 and BHD3 interacts with the backbone of the undamaged strand and in particular with the undamaged flipped-out bases opposite the lesion. This confirms previous biochemical studies by Buterin *et al* (2005) which already suggested that XPC mainly interacts with the non-damaged strand. Also in this regard the interaction of UvrB with the DNA is different, where the main contacts are made with the damaged strand (Waters *et al.* 2006; chapter 4).

The solvent-exposed orientation of the CPD lesion enables the Rad4 protein to exhibit a broad substrate specificity since no damage-specific interactions have to be established. This is in contrast to other UV-damage repair proteins where the protein undergoes extensive interactions with the lesion (Mees *et al.* 2004). The lack of interactions also renders the CPD modification highly disordered and therefore invisible in the crystal structure.

Mapping of sequence conservation between yeast Rad4 and human XPC, which spans along all three BHDs and the TGD domain, indicates that XPC will bind the DNA in a similar way. In addition, most non-truncating mutations of XPC patients are predicted to

destabilize the XPC structure on basis of the available Rad4 structure.

The structure of the Rad4 protein bound to damaged DNA suggests that Rad4/XPC recognizes the presence of a lesion by the propensity of the DNA substrate to adopt a particular structure. Formation of this DNA structure is accompanied by an energetic cost due to loss of base stacking and base pairing interactions. Most NER lesions are expected to reduce these energetic costs and damage-specific binding will take place if the energy cost is lowered beyond the energy gain of the binding reaction. The total energy gained by DNA binding in that case reflects the efficiency of recognition. It is however still unclear whether recognition by Rad4 occurs through a passive mechanism in which the protein recognizes a conformational state of the damaged DNA or through an active mechanism where the protein actively tries to induce the required conformational changes. The extent of kinking induced by the presence of a CPD (~27^{o} into the major groove, Pearlman *et al.* 1985; Husain *et al.* 1988) is slightly lower than that observed in the CPD-containing Rad4-DNA co-crystal structure (~42^{o} into the major groove) indicating that the protein at least partly actively induces a conformational change in the DNA.

Comparison of the apo-Rad4 protein (in the absence of DNA) to the damaged DNA bound structure showed that when the TGDs are aligned the BHDs of the DNA-bound Rad4 are closer to the DNA compared to the apo-structure (Min *et al.* 2007). When undamaged B-DNA is modeled in the damaged DNA-bound protein structure it can be seen that extensive clashing will occur between the DNA and BHD3. In the Apo-structure this would however not be the case suggesting that the protein binds to the undamaged DNA in this particular conformation. Intrinsic flexibility of the DNA and/or a kink induced by the protein in this orientation might bring the DNA close enough to BHD3 to check the DNA for modifications.

In addition to the crystal structures of apo-Rad4 and Rad4 bound to CPD-containing mismatch DNA the study also mentioned a structure of the protein bound to DNA containing only the three nucleotide mismatches (Min *et al.* 2007). Comparing this structure to the one in the presence of the CPD-containing DNA showed that both structures are nearly identical. This strongly suggests that it is not the CPD modification that is recognized by the protein but the distortion in the DNA induced by the mismatches. This is in agreement with earlier studies in which it was shown that CPDs are poorly recognized by the XPC•hHR23B complex (Sugasawa *et al.* 2001; Fitch *et al.* 2003). The asymmetric localization of the mismatch in the DNA probably stabilizes the XPC-DNA complex in this particular orientation, since the DNA flanking the bubble is only long enough on one side to interact with the TGD and BHD1. It remains to be determined whether this structure truly corresponds to the damage recognizing complex and whether or not the damage might be located in the other DNA strand. Possibly *in vivo* the UV-DDB complex determines the orientation of XPC at the site of damage.

1.6 Similarities between prokaryotic and eukaryotic NER proteins

1.6.1 The TFIIH component XPD

The human protein XPD is a subunit of the TFIIH complex, which plays a crucial role in both transcription initiation and nucleotide excision repair. The TFIIH complex consists of 10 polypeptide chains (XPB, XPD, p62, p52, p44, p34, TTDA, cdk7 kinase, cyclin H and MAT1) and its structure has been visualized by electron microscopy (Schultz *et al.* 2000). TFIIH exhibits a ring-like structure with a hole in the middle large enough to accommodate double-stranded DNA. Two of its constituents, XPB and XPD both of which are required for the cell's viability, exhibit an ATP-dependent helicase activity but with opposite polarity. The XPB protein has a 3' → 5' helicase activity and this activity has been shown to be required for transcription but not for the opening and repair of damaged DNA (Coin *et al*. 2007). In contrast XPD is a 5' → 3' helicase whose helicase activity is essential for repair but not for transcription (Winkler *et al*. 2000).

Although the proteins do not share considerable sequence identity, both XPD and UvrB contain seven highly conserved helicase motifs from which it is known that they share three-dimensional structure conservation. This suggests that in the prokaryotic and eukaryotic NER pathways the two proteins have an analogous function. Both proteins are thought to be involved in creating an open conformation in the DNA and serve as a scaffold to recruit the NER specific endonucleases.

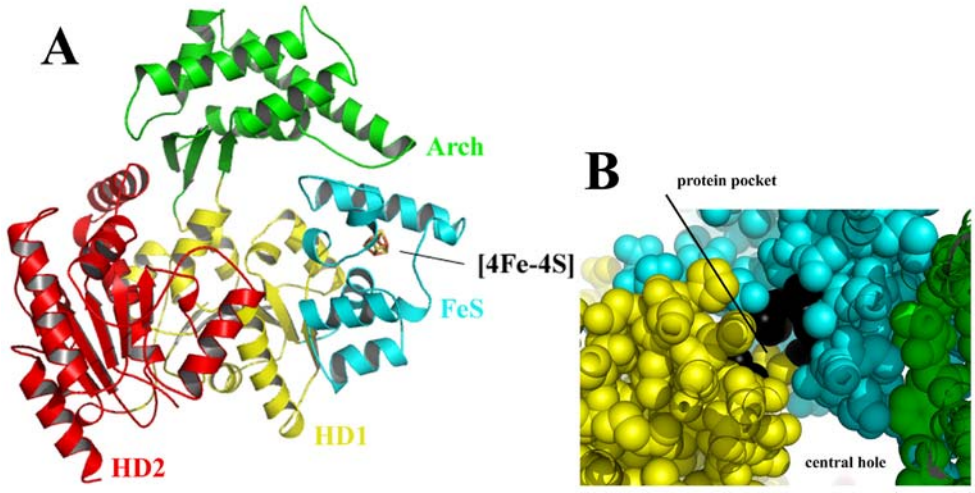

Fig. 20 Crystal structure of the *Thermoplasma acidophilum* XPD protein (pdb accession number 2VSF). (A) Full structure. The two helicase domains (HD1-2, yellow and red), the FeS domain (cyan) and the Arch domain (green) are indicated. The 4Fe-4S cluster is also shown. (B) Close-up view of the pocket present in the crystal structure. Residues forming the pocket (Arg88, Tyr166 and Tyr185) are shown in black.

Recently, various crystal structures have been solved of XPD derived from different archaeal thermophilic organisms: *Sulfolobus acidocaldarius* (Fan *et al.* 2008), *Sulfolobus tokodaii* (Liu *et al.* 2008) and *Thermoplasma acidophilum* (Wolski *et al.* 2008; Fig. 20).

Sequence alignments show that these XPD proteins contain the XPD catalytic core (XPDcc) with a 4Fe-4S cluster and all the helicase motifs conserved with the human XPD. The XPD C-terminal extension, which in the human protein has been shown to interact with p44 and thereby stimulates XPDs helicase activity (Coin *et al.* 1998), is missing from the archaeal XPD proteins. The archaeal genomes however do not encode a p44 homolog, making the C-terminal domain redundant. The catalytic core is comprised of four domains: two Rad51/ RecA-like helicase domains (HD1 and HD2) with two additional domains (the FeS and Arch domains) inserted into HD1 (Fig. 20). The FeS domain coordinates the 4Fe-4S cluster and the Arch-domain has an arch-shaped conformation. In all three studies a ssDNA substrate was modeled into the XPD structure by superimposing known helicase-DNA complex structures. Even though the structures represent the first structures of a helicase with a 5' to 3' polarity it was postulated that the ssDNA is bound in the same orientation as in the 3' to 5' helicases but that the enzyme simply translocates in the opposite direction (Liu *et al.* 2008). This would then place the FeS domain at the site of DNA unwinding.

4Fe-4S clusters are typically found in enzymes performing redox reactions, suggesting a direct role in oxidative stress detection or even direct detection of damaged DNA as has been proposed for other FeS-containing DNA repair proteins (Boal *et al.* 2005a; paragraph 1.3.2). The observation however that FeS domains are also present in various helicases not involved in DNA repair suggests a more generic role in helicase activity for these domains. In XPD (Rad3), the FeS-cluster containing domain specifically recognizes single-strand-double-strand DNA junctions and positions the helicase in an orientation consistent with duplex unwinding (Pugh *et al.* 2008). Since XPD mutants with a disrupted FeS domain lack helicase activity but still bind ssDNA and display a DNA-dependent ATPase activity this suggests that this domain indeed physically separates the DNA strands (Rudolf *et al.* 2006; Pugh *et al.* 2008).

In the structures of Liu *et al* (2008) and Wolski *et al* (2008) the structures of XPD and UvrB have been superimposed and it was concluded that the secondary structure of the two helicase domains (HDs) match well but that there is no significant correspondence between the remaining domains. Remarkably however, the β-hairpin of UvrB coincides with the FeS cluster in XPD suggesting that both domains have a similar function.

In the study of Wolski *et al* (2008; Fig. 20) it was found that helicase domain 1, the FeS domain and the arch domain adopt a donut-shaped structure containing a hole with a diameter of approximately 13 Å (Figure 20A,B). This hole was also visible in the SaXPD structure (3CRV) but with severely reduced dimensions. In the StXPD structure (2VL7) the ring around the hole is breached on one side creating more of a recess. DNA modeling on these structures revealed in all cases that a single strand of DNA can pass through this

hole/recess. However, only in the structure of Wolski *et al.* (2008) the authors identified a narrow pocket made up of two tyrosine residues and one arginine residue located at the side of the hole and directly adjacent to the FeS cluster (Fig. 20B). The dimensions and shape of the pocket are ideal to allow entrance of a non-modified DNA base thereby suggesting a role similar to that of the pocket of the UvrB protein (Truglio *et al.* 2006b). It was postulated that the pocket is used to scan the nature of each nucleotide base by accommodating only unmodified nucleotide bases and excluding damaged ones.

Genome sequencing has revealed that although in addition to XPD archaebacteria also encode homologues of the XPB, XPF and XPG proteins, they do not contain homologues of the essential eukaryotic DNA repair proteins XPA and XPC. Additionally many archaebacteria contain a UvrABC system. It is therefore questionable whether the archaeal XPD plays a role in DNA repair. It is more likely that they are only used for their function in transcription. Consequently, it is doubtful that the pocket is present for a function in damage recognition. It might however play a role in the helicase activity required for strand opening in transcription.

1.6.2 The XPF•ERCC1 complex

The human Excision Repair Cross Complementation group 1 protein (ERCC1) forms a hetero-dimeric complex with XPF in solution, which is recruited to the preincision complex where it binds to XPA, RPA, DNA (Volker *et al.* 2001; Tsodikov *et al.* 2007), and possibly XPB (Evans *et al* .1997; Coin *et al.* 2004). After association of XPF•ERCC1 with the preincision complex the complex performs incision ~20 nucleotides 5' to the lesion (Sijbers *et al.* 1996).

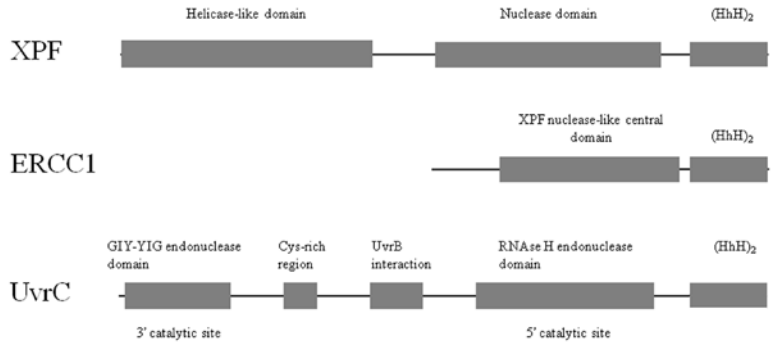

Fig. 21 Domain architecture and alignment of the human XPF and ERCC1 proteins.

Crystal structure analysis (Tsodikov *et al*. 2005) and NMR studies (Tripsianes *et al.* 2005) have shown that part of the domain architectures of the XPF and ERCC1 proteins are very similar suggesting that both proteins evolved from a gene duplication event. Both proteins contain a tandem helix-hairpin-helix C-terminal domain, similar to the bacterial NER protein

UvrC, and a central domain containing a nuclease fold. Only XPF contains an additional N-terminal helicase-like domain (Fig. 21). Both proteins however have divided tasks since XPF comprises the only active nuclease domain of the heterodimer whereas ERCC1 contains the only HhH-domain that binds to DNA. The HhH_2-domains of both proteins have however been shown to be essential for XPF•ERCC1 interaction and stabilization (de Laat *et al.* 1998b; Tripsianes *et al*. 2005). The helicase fold of the XPF protein is not conserved and has been shown to be dispensable for its repair function in NER (Tsodikov *et al.* 2005).

HhH_2-domains are present in several DNA repair proteins (i.e. UvrC, OGG1 and RuvA) where they are known to mediate sequence aspecific DNA contacts and protein-protein interactions. The structure of a complex containing both HhH_2-domains of XPF and ERCC1 has been solved and is shown in figure 22. Characteristic of HhH motifs is a conserved GhG consensus sequence on the tip of the hairpin of each HhH domain in which 'h' denotes a hydrophobic residue (Fig. 22). In the human ERCC1 protein the consensus sequence is only present in the second HhH domain (GLG) whereas the first domain does not contain any glycine residues (Tripsianes *et al.* 2005; Fig. 22). XPF on the other hand does not contain a full GhG sequence in any of its HhH domains, but both hairpins do contain one conserved glycine residue.

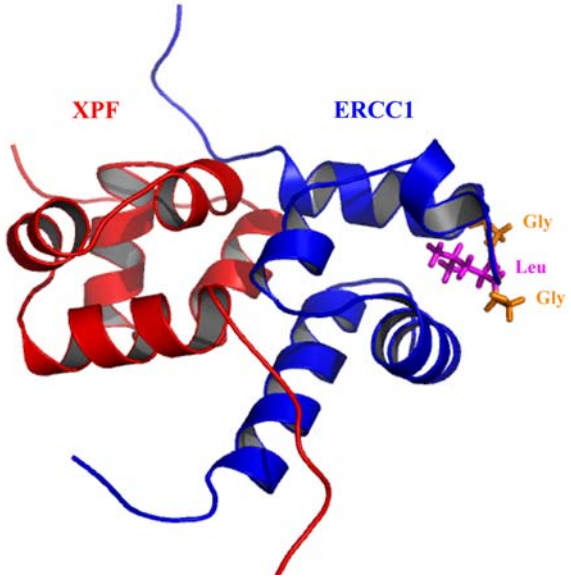

Fig. 22 Crystal structure of the HhH_2-domains of the human XPF/ERCC1 complex (pdb accession number 1Z00). The HhH_2-domain of XPF (red) and ERCC1 (blue) are indicated. Also the GLG consensus sequence is indicated with glycines in orange and leucine in magenta.

The importance of the consensus sequence is shown in the co-crystal structure of the RuvA protein bound to DNA (Ariyoshi *et al.* 2000; Fig. 23), an intermediate in the processing of Holliday junctions. In this structure the hairpin glycines contact the DNA backbone of the dsDNA portion from the minor groove side. Additional contacts are made through interaction of positive residues present in the second helix of each HhH domain.

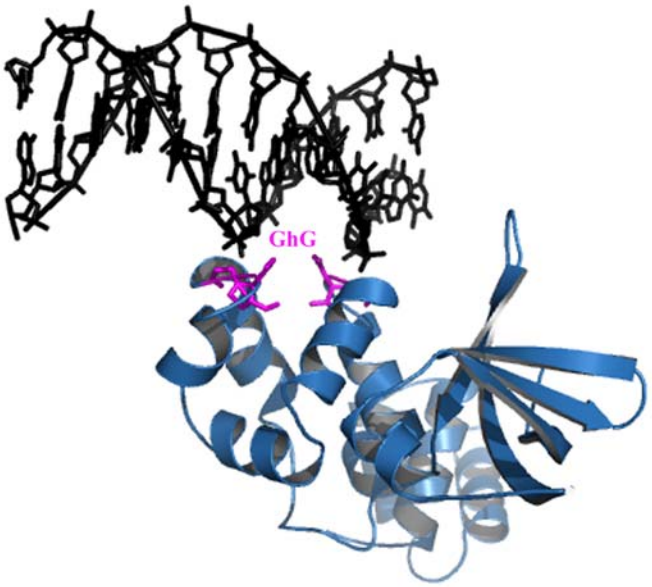

Fig. 23 Crystal structure of the *E. coli* RuvA protein bound to DNA (pdb accession number 1C7Y). The protein is indicated in blue and the DNA in black. Also the GhG consensus sequences are indicated (magenta).

NMR chemical shift analyses have shown that also for ERCC1 both hairpins make contact with the DNA with the strongest contacts involving the glycine residues present in the second HhH domain and some positively charged residues in the second helix of this domain. The absence of any significant chemical shifts for the human XPF hairpin motif indicates that this HhH domain does not participate in DNA binding and probably only mediates interactions with ERCC1 (Tripsianes *et al.* 2005). As in hERCC1 also in *E. coli* UvrC only one of the HhH domains (the first HhH domain) contains the intact consensus sequence (GVG) whereas the other HhH does not (GIS). Both hairpins are however involved in DNA binding (Singh *et al.* 2002) in a way similar to that of ERCC1 involving the glycine residues and positively charged residues in the second helix of the first HhH domain.

Binding of the HhH_2-domains of ERCC1 and UvrC only occurs efficiently on DNA substrates containing a single-strand double-strand DNA junction (Tripsianes *et al.* 2005; Singh *et al.* 2002). This structure mimics the open complex in the preincision complex in both bacterial and human NER. Probably the HhH_2-domain specifically recognizes and binds to this junction in the open complex, thereby stabilizing its structure and positioning the catalytic domain for incision to occur at the appropriate incision position.

1.7 Outline of this thesis

The DNA repair mechanisms described above (DR, BER, UVDE repair, and NER) all play an important role in the maintenance of genomic integrity. The *E. coli* NER system is unique amongst these systems since it requires the action of only three proteins (UvrA, UvrB and UvrC) to recognize various structurally unrelated types of damage and incise the damaged

DNA. After initial damage sensing by UvrA, UvrB verifies the presence of the damage and recruits UvrC, the nuclease. The UvrB protein has been shown to discriminate damaged from undamaged DNA, but the exact mechanism it uses to recognize a DNA damage is until now largely unknown and will be the main topic of research in these studies.

For a long time it was generally believed that the initial damage recognition complex of the bacterial NER system consists of two UvrA subunits and one UvrB subunit. In **chapter 2** however Fluorescence Resonance Energy Transfer (FRET) measurements on GFP- or YFP-tagged UvrB molecules show that this complex does not only contain two UvrA molecules but also two UvrB molecules. In this complex the domains of UvrB involved in UvrC recruitment interact with each other preventing premature UvrC association and gratuitous repair.

In order to study the conformational changes in the DNA upon binding of UvrB to a damaged site we made use of the fluorescent base analog 2-aminopurine. This probe for base stacking interactions is used in **chapter 3** to show that upon binding of UvrB to a lesion two nucleotides become extrahelical: the nucleotide directly 3' to the lesion and its base pairing partner. The extent of base flipping is shown to be dependent on the nature of the cofactor that is bound.

In **chapter 4** it becomes clear that this dual flipping also serves a dual function. Flipping in the damaged DNA strand is shown to be important for the prevention of UvrB binding to undamaged DNA. Flipping in the non-damaged strand on the other hand is required to create a specific conformation to allow binding and subsequent incision by UvrC.

In **chapter 5** different fluorescent damages have been used to monitor conformational changes of the damage itself upon UvrA and/or UvrB binding. We show that upon binding of UvrB to the damage the damaged nucleotide itself remains stacked inside the DNA helix. Furthermore it is shown that occlusion of non-planar groups is an important damage recognition strategy of the UvrB protein.

1.8 References

Aboussekhra, A., Biggerstaff, M., Shivji, M.K., Vilpo, J.A., Moncollin, V., Podust, V.N., Protić, M., Hübscher, U., Egly, J.M. and Wood, R.D. (1995) Mammalian DNA nucleotide excision repair reconstituted with purified protein components. *Cell,* **80,** 859-868.

Alexandrovich, A., Sanderson, M.R., Moolenaar, G.F., Goosen, N. and Lane, A.N. (1999) NMR assignments and secondary structure of the UvrC binding domain of UvrB. *FEBS Lett.,* **451,** 181-185.

Alleva, J.L., Zuo, S., Hurwitz, J. and Doetsch, P.W. (2000) In vitro reconstitution of the Schizosaccharomyces pombe alternative excision repair pathway. *Biochemistry,* **39,** 2659-2666.

Araki, M., Masutani, C., Takemura, M., Uchida, A., Sugasawa, K., Kondoh, J., Ohkuma, Y. And Hanaoka, F. (2001) Centrosome protein centrin 2/caltractin 1 is part of the xeroderma pigmentosum group C complex that initiates global genome nucleotide excision repair. *J. Biol. Chem.,* **276,** 18665-18672.

Araujo, S.J., Tirode, F., Coin, F., Pospiech, H., Syväoja, J.E., Stucki, M., Hübscher, U., Egly, J.M. and Wood, R.D. (2000) Nucleotide excision repair of DNA with recombinant human proteins: definition of the

minimal set of factors, active forms of TFIIH, and modulation by CAK. *Genes Dev.,* **14,** 349-359.
Arber, W. and Dussoix, D. (1962) Host specificity of DNA produced by Escherichia coli. *J. Mol. Biol.,* **5,** 18-36.
Ariyoshi, M., Nishino, T., Iwasaki, H., Shinagawa, H. and Morikawa, K. (2000) Crystal structure of the holliday junction DNA in complex with a single RuvA tetramer. *Proc. Natl. Acad. Sci. USA,* **97,** 8257-8262.
Asahina, H., Kuraoka, I., Shirakawa, M., Morita, E.H., Miura, N., Miyamoto, I., Ohtsuka, E., Okada, Y. and Tanaka, K. (1994) The XPA protein is a zinc metalloprotein with an ability to recognize various kinds of DNA damage. *Mutat. Res.,* **315,** 229-237.
Assenmacher, N., Wenig, K., Lammens, A. and Hopfner, K.P. (2006) Structural basis for transcription-coupled repair: the N terminus of Mfd resembles UvrB with degenerate ATPase motifs. *J. Mol. Biol.,* **355,** 675-683.
Avery, A.M., Kaur, B., Taylor, J.S., Mello, J.A., Essigmann, J.M. and Doetsch, P.W. (1999) Substrate specificity of ultraviolet DNA endonuclease (UVDE/Uve1p) from Schizosaccharomyces pombe. *Nucleic Acids Res.,* **27,** 2256-2264.
Banerjee, A., Yang, W., Karplus, M. and Verdine, G.L. (2005) Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *Nature,* **434,** 612-618.
Banerjee, A. and Verdine, G.L. (2006) A nucleobase lesion remodels the interaction of its normal neighbor in a DNA glycosylase complex. *Proc. Natl. Acad. Sci.,* **103,** 15020-15025.
Barzilay, G. and Hickson, I.D. (1995) Structure and function of apurinic/apyrimidinic endonucleases. *Bioessays,* **17,** 713-719.
Batty, D., Rapic-Otrin, V., Levine, A.S. and Wood, R.D. (2000) Stable binding of human XPC complex to irradiated DNA confers strong discrimination for damaged sites. *J. Mol. Biol.,* **300,** 275-290.
Beck, D.J., Popoff, S., Sancar, A. and Rupp, W.D. (1985) Reactions of the UVRABC excision nuclease with DNA damaged by diamminedichloroplatinum(II). *Nucleic Acids Res.,* **13,** 7395-7412.
Berson, D.M. (2003) Strange vision: ganglion cells as circadian photoreceptors. *Trends Neurosci.,* **26,** 314-320.
Bhattacharya, P.K. and Barton, J.K. (2001) Influence of intervening mismatches on long-range guanine oxidation in DNA duplexes. *J. Am. Chem. Soc.,* **123,** 8649-8656.
Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.,* **16,** 6-21.
Blainey, P.C., van Oijen, A.M., Banerjee, A., Verdine, G.L. and Xie, X.S. (2006) A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc. Natl. Acad. Sci. USA,* **103,** 5752-5757.
Boal, A.K. and Barton, J.K. (2005a) Electrochemical detection of lesions in DNA. *Bioconjug. Chem.,* **16,** 312-321.
Boal, A.K., Yavin, E., Lukianova, O.A., O'Shea, V.L., Davis, S.S. and Barton, J.K. (2005b) DNA-bound redox activity of DNA repair glycosylases containing [4Fe-4S] clusters. *Biochemistry,* **44,** 8397-8407.
Boal, A.K., Yavin, E. and Barton, J.K. (2007) DNA repair glycosylases with a [4Fe-4S] cluster: a redox cofactor for DNA-mediated charge transport? *J. Inorg. Biochem.,* **101,** 1913-1921.
Bohr, V.A., Smith, C.A., Okumoto, D.S. and Hanawalt, P.C. (1985) DNA repair in an active gene: Removal of pyrimidine dimers from the DHFR gene of CHO cells is much more efficient than in the genome overall. *Cell,* **40,** 359-369.
Boon, E.M., Livingston, A.L., Chmiel, N.H., David, S.S, and Barton, J.K. (2003) DNA-mediated charge transport for DNA repair. *Proc. Natl. Acad. Sci. USA,* **100,** 12543-12547.
Bowman, K.K., Sidik, K., Smith, C.A., Taylor, J.S., Doetsch, P.W. and Freyer, G.A. (1994) A new ATP-independent DNA endonuclease from Schizosaccharomyces pombe that recognizes cyclobutane pyrimidine dimers and 6-4 photoproducts. *Nucleic Acids Res.,* **22,** 3026-3032.
Bruner, S.D., Norman, D.P. and Verdine, G.L. (2000) Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature,* **403,** 859-866.
Burns, J.L., Guzder, S.N., Sung, P., Prakash, S. and Prakash, L. (1996) An affinity of human replication

protein A for ultraviolet-damaged DNA. *J. Biol. Chem.,* **271,** 11607-11610.
Buterin, T., Meyer, C., Giese, B. and Naegeli, H. (2005) DNA quality control by conformational readout on the undamaged strand of the double helix. *Chem. Biol.,* **12,** 913-922.
Caron, P.R. and Grossman, L. (1988) Involvement of a cryptic ATPase activity of UvrB and its proteolysis product, UvrB* in DNA repair. *Nucleic Acids Res.,* **16,** 10891-10901.
Clugston, C.K., McLaughlin, K., Kenny, M.K. and Brown, R. (1992) Binding of human single-stranded DNA binding protein to DNA damaged by the anticancer drug cis-diamminedichloroplatinum (II). *Cancer Res.,* **52,** 6375-6379.
Coin, F., Marinoni, J.C., Rodolfo, C., Fribourg, S., Pedrini, A.M. and Egly, J.M. (1998) Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH. *Nat. Genet.,* **20,** 184-188.
Coin, F., Auriol, J., Tapias, A., Clivio, P., Vermeulen, W. and Egly, J.M. (2004) Phosphorylation of XPB helicase regulates TFIIH nucleotide excision repair activity. *EMBO J.,* **23,** 4835-4846.
Coin, F., Oksenych, V., and Egly, J.M. (2007) Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. *Mol. Cell,* **26,** 245-256.
Cunningham, R.P., Asahara, H., Bank, J.F., Scholes, C.P., Salerno, J.C., Surerus, K., Münck, E., McCracken, J., Peisach, J. and Emptage, M.H. (1989) Endonuclease III is an iron-sulfur protein. *Biochemistry,* **28,** 4450-4455.
Dahm, R. (2005) Friedrich Miescher and the discovery of DNA. *Dev. Biol.,* **278,** 274-288.
Daniels, D.S., Woo, T.T., Luu, K.X., Noll, D.M., Clarke, N.D., Pegg, A.E. and Tainer, J.A. (2004) DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nat. Struct. Mol. Biol.,* **11,** 714-720.
Deaconescu, A.M., Chambers, A.L., Smith, A.J., Nickels, B.E., Hochschild, A., Savery, N.J. and Darst, S.A. (2006) Structural basis for bacterial transcription-coupled DNA repair. *Cell,* **124,** 507-520.
de Laat, W.L., Appeldoorn, E., Sugasawa, K., Weterings, E., Jaspers, N.G. and Hoeijmakers, J.H. (1998a) DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair. *Genes Dev.,* **12,** 2598-2609.
de Laat, W.L., Sijbers, A.M., Odijk, H., Jaspers, N.G. and Hoeijmakers, J.H. (1998b) Mapping of interaction domains between human repair proteins ERCC1 and XPF. *Nucl. Acids Res.,* **26,** 4146-4152.
DellaVecchia, M.J., Croteau, D.L., Skorvaga, M., Dezhurov, S.V., Lavrik, O.I. and van Houten, B. (2004) Analyzing the handoff of DNA from UvrA to UvrB utilizing DNA-protein photoaffinity labeling. *J. Biol. Chem.,* **279,** 45245-45256.
Dip, R., Camenisch, U. And Naegeli, H.P. (2004) Mechanism of DNA-damage recognition and strand discrimination in human nucleotide excision repair. *DNA repair (Amst),* **3,** 1409-1423,
Doolittle, R.F., Johnson, M.S., Husain, I., van Houten, B., Thomas, D.C. and Sancar, A. (1986) Domainal evolution of a prokaryotic DNA repair protein and its relationship to active-transport proteins. *Nature,* **323,** 451-453.
Dornberger, U., Leijon, M. and Fritzsche, H. (1999) High base pair opening rates in tracts of GC base pairs. *J. Biol. Chem.,* **274,** 6957-6962.
Drapkin, R. and Reinberg, D. (1994) The multifunctional TFIIH complex and transcriptional control. *Trends Biochem. Sci.,* **19,** 504-508.
Dualan, R., Brody, T., Keeney, S., Nichols, A.F., Admon, A. and Linn, S. (1995) Chromosomal localization and cDNA cloning of the genes (DDB1 and DDB2) for the p127 and p48 subunits of a human damage-specific DNA binding protein. *Genomics,* **29,** 62-69.
Duguid, E.M., Rice, P.A. and He, C. (2005) The structure of the human AGT protein bound to DNA and its implications for damage detection. *J. Mol. Biol.,* **350,** 657-666.
Emery, P., So, W.V., Kaneko, M., Hall, J.C. and Rosbash, M. (1998) CRY, a Drosophila clock and light-

regulated cryptochrome, is a major contributor to circadian rhythm resetting and photosensitivity. *Cell,* **95,** 669-679.

Eryilmaz, J., Ceschini, S., Ryan, J., Geddes, S., Waters, T.R. and Barrett, T.E. (2006) Structural insights into the cryptic DNA-dependent ATPase activity of UvrB. *J. Mol. Biol.,* **357,** 62-72.

Evans E, Moggs JG, Hwang JR, Egly JM, Wood RD. (1997) Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. *EMBO J.,* **16,** 6559-6573.

Falnes, P.Ø., Johansen, R.F. and Seeberg, E. (2002) AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. *Nature,* **419,** 178-182.

Fan, L., Arvai, A.S., Cooper, P.K., Iwai, S., Hanaoka, F. and Tainer, J.A. (2006) Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair. *Mol. Cell,* **22,** 27-37.

Fan, L., Fuss, J.O., Cheng, Q.J., Arvai, A.S., Hammel, M., Roberts, V.A., Cooper, P.K. and Tainer, J.A. (2008) XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. *Cell,* **133,** 789-800.

Fitch, M.E., Nakajima, S., Yasui, A. and Ford, J.M. (2003) In vivo recruitment of XPC to UV-induced cyclobutane pyrimidine dimers by the DDB2 gene product. *J. Biol. Chem.,* **278,** 46906-46910.

Friedberg, E. (2006) DNA repair and mutagenesis. *American society for microbiology, 2nd edition.*

Gomez-Pinto, A., Cubero, E., Kalko, S.G., Monaco, V., van der Marel, G., van Boom, J.H., Orozco, M. and Gonzalez, C. (2004) Effect of bulky lesions on DNA: solution structure of a DNA duplex containing a cholesterol adduct. *J. Biol. Chem.,* **279,** 24552-24560.

Goosen, N and Moolenaar, G.F. (2008) Repair of UV damage in bacteria. *DNA repair,* **7,** 353-379.

Gorbalenya, A.E. and Koonin, E.V. (1990) Superfamily of UvrA-related NTP-binding proteins. Implications for rational classification of recombination repair systems. *J. Mol. Biol.,* **213,** 583-591.

Gordienko, I. and Rupp, W.D. (1997) The limited strand-separating activity of the UvrAB protein complex and its role in the recognition of DNA damage. *EMBO J.,* **16,** 889-895.

Grollman, A.P. and Moriya, M. (1993) Mutagenesis by 8-oxoguanine: an enemy within. *Trends Genet*., **9,** 246-249.

Grove, A., Galeone, A., Yu, E., Mayol, L. and Geiduschek, E.P. (1998) Affinity, stability and polarity of binding of the TATA binding protein governed by flexure at the TATA box. *J. Mol. Biol.,* **282,** 731-739.

Guan, Y., Manuel, R.C., Arvai, A.S., Parikh, S.S., Mol, C.D., Miller, J.H., Lloyd, S. and Tainer, J.A. (1998) MutY catalytic core, mutant and bound adenine structures define specificity for DNA repair enzyme superfamily. *Nat. Struct. Biol.,* **5,** 1058-1064.

Guidice, E.; Va´rnai, P.; Lavery, R. (2001) Energetic and Conformational Aspects of A:T Base-Pair Opening within the DNA Double Helix. *Chem. Phys. Chem*., **11,** 673-677.

Guidice, E.; Va´rnai, P.; Lavery, R. (2003) Base pair opening within B-DNA: free energy pathways for GC and AT pairs from umbrella sampling simulations. *Nucl. Acids Res.,* **31,** 1434-1443.

Hamm-Alvarez, S., Sancar, A. and Rajagopalan, K.V. (1989) Role of enzyme-bound 5,10-methenyltetrahydropteroylpolyglutamate in catalysis by Escherichia coli DNA photolyase. *J. Biol. Chem.,* **264,** 9649-9656.

Hey, T., Lipps, G., Sugasawa, K., Iwai, S., Hanaoka, F. and Krauss, G. (2002) The XPC-HR23B complex displays high affinity and specificity for damaged DNA in a true-equilibrium fluorescence assay. *Biochemistry,* **41,** 6583-6587.

Higgins, C.F. and Linton, K.J. (2004) The ATP switch model for ABC transporters. *Nat. Struct. Mol. Biol.,* **11,** 918-926.

Hildebrand, E.L. and Grossman, L. (1999) Oligomerisation of the UvrB nucleotide excision repair protein of *Escherichia coli*. *J. Biol. Chem.,* **274,** 27885-27890.

Hitomi, K., Okamoto, K., Daiyasu, H., Miyashita, H., Iwai, S., Toh, H., Ishiura, M. and Todo, T. (2000) Bacterial cryptochrome and photolyase: characterization of two photolyase-like genes of *Synechocystis* sp. PCC6803. *Nucleic Acids Res.,* **28,** 2353-2362.

Hollis, T., Ichikawa, Y. and Ellenberger, T. (2000) DNA bending and a flip-out mechanism for base excision by the helix-hairpin-helix DNA glycosylase, Escherichia coli AlkA. *EMBO J.,* **19,** 758-766.

Hosfield, D.J., Guan, Y., Haas, B.J., Cunningham, R.P. and Tainer, J.A. (1999) Structure of the DNA repair enzyme endonuclease IV and its DNA complex: double-nucleotide flipping at abasic sites and three-metal-ion catalysis. *Cell,* **98,** 397-408.

Howard-Flanders, P., Boyce, R.P. and Theriot, L. (1966) Three loci in *Escherichia coli* K-12 that control the excision of pyrimidine dimers and certain other mutagen products from DNA. *Genetics,* **53,** 1119-1136.

Hsu, D.S., Kim, S.T., Sun, Q. and Sancar, A. (1995) Structure and function of the UvrB protein. *J. Biol. Chem.,* **270,** 8319-8327.

Hsu, D.S., Zhao, X., Zhao, S., Kazantsev, A., Wang, R.P., Todo, T., Wei, Y.F. and Sancar, A. (1996) Putative human blue-light photoreceptors hCRY1 and hCRY2 are flavoproteins. *Biochemistry,* **35,** 13871-13877.

Husain, I., Griffith, J. and Sancar, A. (1988) Thymine dimers bend DNA. *Proc. Natl. Acad. Sci. USA,* **85,** 2558-2562.

Hwang, B.J., Toering, S., Francke, U. and Chu, G. (1998) p48 Activates a UV-damaged-DNA binding factor and is defective in xeroderma pigmentosum group E cells that lack binding activity. *Mol. Cell. Biol.,* **18,** 4391-4399.

Izumi, T., Wiederhold, L.R., Roy, G., Roy, R., Jaiswal, A., Bhakat, K.K., Mitra, S. and Hazra, T.K. (2003) Mammalian DNA base excision repair proteins: their interactions and role in repair of oxidative DNA damage. *Toxicology,* **193,** 43-65.

Jones, C.J. and Wood, R.D. (1993) Preferential binding of the xeroderma pigmentosum group A complementing protein to damaged DNA. *Biochemistry,* **32,** 12096-12104.

Karakas, E., Truglio, J.J., Croteau, D., Rhau, B., Wang, L., Van Houten, B. and Kisker, C. (2007) Structure of the C-terminal half of UvrC reveals an RNase H endonuclease domain with an Argonaute-like catalytic triad. *EMBO J.,* **26,** 613-622.

Kaur, B., Fraser, J.L., Freyer, G.A., Davey, S. and Doetsch, P.W. (1999) A Uve1p-mediated mismatch repair pathway in Schizosaccharomyces pombe. *Mol. Cell. Biol.,* **19,** 4703-4710.

Keeney, S., Eker, A.P., Brody, T., Vermeulen, W., Bootsma, D., Hoeijmakers, J.H. and Linn, S. (1994) Correction of the DNA repair defect in xeroderma pigmentosum group E by injection of a DNA damage-binding protein. *Proc. Natl. Acad. Sci. USA,* **91,** 4053-4056.

Kelley, S.O. and Barton, J.K. (1999) Electron transfer between bases in double helical DNA. *Science,* **283,** 375-381.

Kelman, Z. (1997) PCNA: structure, functions and interactions. *Oncogene,* **14,** 629-640.

Kim, J.L., Morgenstern, K.A., Griffith, J.P., Dwyer, M.D., Thomson, J.A., Murcko, M.A., Lin, C. and Caron, P.R. (1998) Hepatitus C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. *Structure,* **6,** 89-100.

Korolev, S., Hsieh, J., Gauss, G.H., Lohmann, T.M. and Waksman, G. (1997) Major domain swiveling revealed by the crystal structures of complexes of *E. coli* Rep Helicase bound to single-stranded DNA and ADP. *Cell,* **90,** 635-647.

Krokan, H.E., Standal, R. and Slupphaug, G. (1997) DNA glycosylases in the base excision repair of DNA. *Biochem. J.,* **325,** 1-16.

Krosky, D.J., Song, F. and Stivers, J.T. (2005) The origins of high-affinity enzyme binding to an extrahelical DNA base. *Biochemistry,* **44,** 5949–5959.

Lau, A.Y., Schärer, O.D., Samson, L., Verdine, G.L. and Ellenberger, T. (1998) Crystal structure of a human

alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. *Cell,* **95,** 249-258.

Lau, A.Y., Wyatt, M.D., Glassner, B.J., Samson, L.D. and Ellenberger, T. (2000) Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG. *Proc. Natl. Acad. Sci. USA,* **97,** 13573-13578.

Lin, J.J. and Sancar, A. (1992) Active site of (A)BC excinuclease. I. Evidence for 5' incision by UvrC through a catalytic site involving Asp399, Asp438, Asp466, and His538 residues. *J. Biol. Chem.,* **267,** 17688-17692.

Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature.*, **362,** 709-715.

Linton, K.J. (2007) Structure and function of ABC transporters. *Physiology (Bethesda),* **22,** 122-130.

Lipscomb, L.A., Peek, M.E., Morningstar, M.L., Verghis, S.M., Miller, E.M., Rich, A., Essigmann, J.M. and Williams, L.D. (1995) X-ray structure of a DNA decamer containing 7,8-dihydro-8-oxoguanine. *Proc. Natl. Acad. Sci.,* **92,** 719-723.

Liu, H., Rudolf, J., Johnson, K.A., McMahon, S.A., Oke, M., Carter, L., McRobbie, A-M., Brown, S.E., Naismith, J.H. and White, M.F. (2008) Structure of the DNA repair helicase XPD. *Cell,* **133,** 801-812.

Lovell, S., Goryshin, I.Y., Reznikoff, W.R. and Rayment, I. (2002) Two-metal active site binding of a Tn5 transposase synaptic complex. *Nat. Struct. Biol.,* **9,** 278-281.

Machius, M., Henry, L., Palnitkar, J. and Deisenhofer, J. (1999) Crystal structure of the DNA nucleotide excision repair enzyme UvrB from *Thermus thermophilus*. *Proc. Natl. Acad. Sci. USA,* **86,** 11717-11722.

Masutani, C., Sugasawa, K., Yanagisawa, J., Sonoyama, T., Ui, M., Enomoto, T., Takio, K., Tanaka, K., van der Spek, P.J., Bootsma, D. et al. (1994) Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast RAD23. *EMBO J.,* **13,** 1831-1843.

Mees, A., Klar, T., Gnau, P., Hennecke, U., Eker, A.P., Carell, T. and Essen, L.O. (2004) Crystal structure of a photolyase bound to a CPD-like DNA lesion after in situ repair. *Science*, **306,** 1789-1793.

Mellon, I. and Hanawalt, P.C. (1989) Induction of the Escherichia coli lactose operon selectively increases repair of its transcribed DNA strand. *Nature,* **342,** 95-98.

Michaels, M.L. and Miller, J.H. (1992) The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). *J. Bacteriol.*, **174,** 6321-6325.

Min, J.-H. and Pavletich, N.P. (2007) Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature,* **449,** 570-575

Minko, I.G., Zou, Y. and Lloyd, R.S. (2002) Incision of DNA-protein crosslinks by UvrABC nuclease suggests a potential repair pathway involving nucleotide excision repair. *Proc. Natl. Acad. Sci. USA,* **99,** 1505-1509.

Missura, M., Buterin, T., Hindges, R., Hübscher, U., Kasparkova, J., Brabec, V. and Naegeli, H. (2001) Double-check probing of DNA bending and unwinding by XPA-RPA: an architectural function in DNA repair. *EMBO J.,* **20,** 3554-3564.

Mol, C.D., Izumi, T., Mitra, S. and Tainer, J.A. (2000) DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination. *Nature,* **403,** 451-456.

Moolenaar, G.F., Visse, R., Ortiz-Buysse, M., Goosen, N. and van de Putte, P. (1994) Helicase motifs V and VI of the Escherichia coli UvrB protein of the UvrABC endonuclease are essential for the formation of the preincision complex. *J. Mol. Biol.,* **240,** 294-307.

Moolenaar, G.F., Franken, K.L., Dijkstra, D.M., Thomas-Oates, J.E., Visse, R., van de Putte, P. and Goosen, N. (1995) The C-terminal region of the UvrB protein of Escherichia coli contains an important determinant for UvrC binding to the preincision complex but not the catalytic site for 3'-incision. *J. Biol. Chem.,* **270,** 30508-30515.

Moolenaar, G.F., Franken, K.L., van de Putte, P. and Goosen, N. (1997) Function of the homologous regions of the *Escherichia coli* DNA excision repair proteins UvrB and UvrC in stabilization of the UvrBC-DNA complex

and in 3'-incision. *Mutat. Res.,* **385,** 195-203.
Moolenaar, G.F., Bazuine, M., van Knippenberg, I.C., Visse, R. and Goosen, N. (1998a) Characterization of the Escherichia coli damage-independent UvrBC endonuclease activity. *J. Biol. Chem.,* **273,** 34896-34903.
Moolenaar, G.F., Uiterkamp, R.S., Zwijnenburg, D.A. and Goosen, N. (1998b) The C-terminal region of the Escherichia coli UvrC protein, which is homologous to the C-terminal region of the human ERCC1 protein, is involved in DNA binding and 5'-incision. *Nucleic Acids Res.,* **26,** 462-468.
Moolenaar, G.F., Herron, M.F., Monaco, V., van der Marel, G.A., van Boom, J.H., Visse, R. and Goosen, N. (2000) The role of ATP binding and hydrolysis by UvrB during nucleotide excision repair. *J. Biol. Chem.,* **275,** 8044-8050.
Moolenaar, G.F., Höglund, L. and Goosen, N. (2001) Clue to damage recognition by UvrB: residues in the β-hairpin structure prevent binding to non-damaged DNA. *EMBO J.,* **20,** 6140–6149.
Moolenaar, G.F., van Rossum, S., van Kesteren, M. and Goosen, N. (2002) Cho, a second endonuclease involved in *Escherichia coli* nucleotide excision repair. *Proc. Natl. Acad. Sci. USA,* **99,** 1467-1472.
Moolenaar, G.F., Schut, M. and Goosen, N. (2005) Binding of the UvrB dimer to non-damaged and damaged DNA: residues Y92 and Y93 influence the stability of both subunits. *DNA repair,* **4,** 699-713.
Moser, J., Volker, M., Kool, H., Alekseev, S., Vrieling, H., Yasui, A., van Zeeland, A.A. and Mullenders, L.H. (2005) The UV-damaged DNA binding protein mediates efficient targeting of the nucleotide excision repair complex to UV-induced photo lesions. *DNA repair,* **4,** 571-582.
Mu, D., Hsu, D.S. and Sancar, A. (1996) Reaction mechanism of human DNA repair excision nuclease. *J. Biol. Chem.,* **271,** 8285-8294.
Mu, D. and Sancar, A. (1997) Model for XPC-independent transcription-coupled repair of pyrimidine dimers in humans. *J. Biol. Chem.,* **272,** 7570-7573.
Myles, G.M., Hearst, J.E. and Sancar, A. (1991) Site-specific mutagenesis of conserved residues within Walker A and B sequences of Escherichia coli UvrA protein. *Biochemistry,* **30,** 3824-3834.
Nakagawa, N., Sugahara, M., Masui, R., Kato, R., Fukuyama, K. and Kuramitsu, S. (1999) Crystal structure of *Thermus thermophilus* HB8 UvrB protein, a key enzyme of nucleotide excision repair. *J. Biochem.,* **126,** 986-990.
Nash, H.M., Bruner, S.D., Schärer, O.D., Kawate, T., Addona, T.A., Spooner, E., Lane, W.S. and Verdine, G.L. (1996) Cloning of a yeast 8-oxoguanine DNA glycosylase reveals the existence of a base-excision DNA-repair protein superfamily. *Curr. Biol.,* **6,** 968-980.
Navaratnam, S., Myles, G.M., Strange, R.W. and Sancar, A. (1989) Evidence from extended X-ray absorption fine structure and site-specific mutagenesis for zinc fingers in UvrA protein of *Escherichia coli*. *J. Biol. Chem.,* **264,** 16067-16071.
Ng, J.M., Vermeulen, W., van der Horst, G.T., Bergink, S., Sugasawa, K., Vrieling, H. And Hoeijmakers, J.H. (2003) A novel regulation mechanism of DNA repair by damage-induced and RAD23-dependent stabilization of xeroderma pigmentosum group C protein. *Genes Dev.,* **17,** 1630-1645.
Nishi, R., Okuda, Y., Watanabe, E., Mori, T., Iwai, S., Masutani, C., Sugasawa, K. and Hanaoka, F. (2005) Centrin 2 stimulates nucleotide excision repair by interacting with xeroderma pigmentosum group C protein. *Mol. Cell. Biol.,* **25,** 5664-5674.
Nowotny, M., Gaidamakov, S.A., Crouch, R.J. and Yang, W. (2005) Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell,* **121,** 1005-1016.
Nunez, M.E., Hall, D.B. and Barton, J.K. (1999) Long-range oxidative damage to DNA: effects of distance and sequence. *Chem. Biol.,* **6,** 85-97.
Nunez, M.E., Holmquist, G.P. and Barton, J.K. (2001) Evidence for DNA charge transport in the nucleus. *Biochemistry,* **40,** 12465-12471.
Nunez, M.E., Noyes, K.T. and Barton, J.K. (2002) Oxidative charge transport through DNA in nucleosome core particles. *Chem. Biol.,* **9,** 403-415.

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

O'Brien, P.J. and Ellenberger, T. (2004) Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase. *J. Biol. Chem.,* **279,** 9750-9757.

O'Donovan, A., Davies, A.A., Moggs, J.G., West, S.C. and Wood, R.D. (1994) XPG endonucleases makes the 3' incision in human DNA nucleotide excision repair. *Nature,* **371,** 432-435.

Oh, E.Y. and Grossman, L. (1987) Helicase properties of the Escherichia coli UvrAB protein complex. Proc. Natl. Acad. Sci. USA, **84,** 3638-3642.

Oh, E.Y., Claassen, L., Thiagalingam, S., Mazur, S. and Grossman, L. (1989) ATPase activity of the UvrA and UvrAB protein complexes of the Escherichia coli UvrABC endonuclease. *Nucleic Acids Res.,* **17,** 4145-4159.

Okamura, H., Miyake, S., Sumi, Y., Yamaguchi, S., Yasui, A., Muijtjens, M., Hoeijmakers, J.H. and van der Horst, G.T. (1999) Photic induction of mPer1 and mPer2 in cry-deficient mice lacking a biological clock. *Science,* **286,** 2531-2534.

O'Neil, L.L., Grossfield, A. and Wiest, O. (2007) Base flipping of the thymine dimer in duplex DNA. *J Phys Chem B.,* **111,** 11843-11849.

Orren, D.K. and Sancar, A. (1989) The (A)BC excinuclease of Escherichia coli has only the UvrB and UvrC subunits in the incision complex. *Proc. Natl. Acad. Sci. USA,* **86,** 5237-5241.

Orren, D.K., Selby, C.P., Hearst, J.E. and Sancar, A. (1992) Post-incision steps of nucleotide excision repair in *Escherichia coli*. Disassembly of the UvrBC-DNA complex by helicase II and DNA polymerase I. *J. Biol. Chem.,* **267,** 780-788.

Ortolan, T.G., Tongaonkar, P., Lambertson, D., Chen, L., Schauber, C. and Madura, K. (2000) The DNA repair protein rad23 is a negative regulator of multi-ubiquitin chain assembly. *Nat. Cell. Biol.,* **2,** 601-608.

Pakotiprapha, D., Inuzuka, Y., Bowman, B.R., Moolenaar, G.F., Goosen, N., Jeruzalmi, D. and Verdine, G.L. (2008) Crystal structure of *Bacillus stearothermophilus* UvrA provides inside into ATP-modulated dimerization, UvrB interaction, and DNA binding. *Mol. Cell,* **29,** 122-133.

Parikh, S.S., Mol, C.D., Slupphaug, G., Bharati, S., Krokan, H.E. and Tainer, J.A. (1998) Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J.,* **17,** 5214-5226.

Park, H.W., Kim, S.T., Sancar, A. and Deisenhofer, J. (1995) Crystal structure of DNA photolyase from Escherichia coli. *Science,* **268,** 1866-1872.

Park, J.-S., Marr, M.T. and Roberts, J.W. (2002) E. coli transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. *Cell,* **109,** 757-767.

Parvin, J.D., McCormick, R.J., Sharp, P.A. and Fisher, D.E. (1995) Pre-bending of a promoter sequence enhances affinity for the TATA-binding factor. *Nature,* **373,** 724-727.

Paspaleva, K., Thomassen, E., Pannu, N.S., Iwai, S., Moolenaar, G.F., Goosen, N. and Abrahams, J.P. (2007) Crystal structure of the DNA repair enzyme Ultraviolet Damage Endonuclease. *Structure,* **15,** 1316-1324.

Pearlman, D.A., Holbrook, S.R., Pirkle, D.H. and Kim, S.H. (1985) Molecular models for DNA damaged by photoreaction. *Science,* **227,** 1304-1308.

Potter, P.M., Brennand, J. and Margison, G.P. (1986) Lack of sequence homology between a fragment of *E. coli* DNA encoding an O^6-methylguanine methyltransferase and the ada gene. *Br. J. Cancer,* **54,** 366-367.

Potter, P.M., Wilkinson, M.C., Fitton J., Carr, F.J., Brennand, J., Cooper, D.P. and Margison, G.P. (1987) Characterization and nucleotide sequence of *ogt,* the O^6-alkylguanine-DNA-alkyltransferase gene of *E. coli*. *Nucleic Acids Res.,* **15,** 9177-9193.

Pugh, R.A., Honda, M., Leesley, H., Thomas, A., Lin, Y., Nilges, M.J., Cann, I.K. and Spies, M. (2008) The iron-containing domain is essential in Rad3 helicases for coupling of ATP hydrolysis to DNA translocation and for targeting the helicase to the single-stranded DNA-double-stranded DNA junction. *J. Biol. Chem.,* **283,** 1732-1743.

Rapić-Otrin V, Navazza V, Nardo T, Botta E, McLenigan M, Bisi DC, Levine AS, Stefanini M. (2003) True XP group E patients have a defective UV-damaged DNA binding protein complex and mutations in DDB2 which reveal the functional domains of its p48 product. *Hum. Mol. Genet.,* **12,** 1507-1522.

Rasimas, J.J., Pegg, A.E. and Fried, M.G. (2003) DNA-binding mechanism of O6-alkylguanine-DNA alkyltransferase. Effects of protein and DNA alkylation on complex stability. *J. Biol. Chem.,* **278,** 7973-7980.

Reardon, J.T., Nichols, A.F., Keeney, S., Smith, C.A., Taylor, J.S., Linn, S. and Sancar, A. (1993) Comparative analysis of binding of human damaged DNA-binding protein (XPE) and Escherichia coli damage recognition protein (UvrA) to the major ultraviolet photoproducts: T[c,s]T, T[t,s]T, T[6-4]T, and T[Dewar]T. *J. Biol. Chem.,* **268,** 21301-21308.

Reardon, J.T., Mu, D. and Sancar, A. (1996) Overproduction, purification, and characterization of the XPC subunit of the human DNA repair excision nuclease. *J. Biol. Chem.,* **271,** 19451-19456.

Reardon, J.T. and Sancar, A. (2002) Molecular anatomy of the human excision nuclease assembled at sites of DNA damage. *Mol. Cell. Biol.,* **22,** 5938-5945.

Riedl, T., Hanaoka, F. and Egly, J.M. (2003) The comings and goings of nucleotide excision repair factors on damaged DNA. *EMBO J.,* **22,** 5293-5303.

Rivas, F.V., Tolia, N.H., Song, J.J., Aragon, J.P., Liu, J., Hannon, G.J. and Joshua-Tor, L. (2005) Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat. Struct. Mol. Biol.,* **12,** 340-349.

Rudolf, J., Makrantoni, V., Ingledew, W.J., Stark, M.J. and White, M.F. (2006) The DNA repair helicases XPD and FancJ have essential iron-sulfur domains. *Mol. Cell,* **23,** 801-808.

Rupert, C.S., Goodgal, S.H. and Herriott, R.M. (1958) Photoreactivation in vitro of ultraviolet-inactivated Hemophilus influenzae transforming factor. *J. Gen. Physiol.,* **41,** 451-471.

Sancar, A. and Rupp, D. (1983) A novel repair enzyme: UVRABC excision nuclease of Escherichia coli cuts a DNA strand on both sides of the damaged region. *Cell,* **33,** 249-260.

Sancar, A., Franklin, K.A., Sancar, G. and Tang, M.S. (1985) Repair of psoralen and acetylaminofluorene DNA adducts by ABC excinuclease. *J. Mol. Biol.,* **184,** 725-734.

Saparbaev, M. and Laval, J. (1994) Excision of hypoxanthine from DNA containing dIMP residues by the Escherichia coli, yeast, rat, and human alkylpurine DNA glycosylases. *Proc. Natl. Acad. Sci. USA,* **91,** 5873-5877.

Saparbaev, M., Kleibl, K. and Laval, J. (1995) Escherichia coli, Saccharomyces cerevisiae, rat and human 3-methyladenine DNA glycosylases repair 1,N6-ethenoadenine when present in DNA. *Nucleic Acids Res.,* **23,** 3750-3755.

Schultz, P., Fribourg, S., Poterszman, A., Mallouh, V., Moras, D. and Egly, J.M. (2000) Molecular structure of human TFIIH. *Cell,* **102,** 599-607.

Sedgwick, B. (2004) Repairing DNA-methylation damage. *Nat. Rev. Mol. Cell. Biol.,* **5,** 148-157.

Seeberg, E. and Steinum, A. (1982) Purification and properties of the UvrA protein from *Escherichia coli. Proc. Natl. Acad. Sci. USA,* **79,** 988-992.

Selby, C.P. and Sancar, A. (1990) Transcription preferentially inhibits nucleotide excision repair of the template DNA strand in vitro. *J. Biol. Chem.,* **265,** 21330-21336.

Selby, C.P. and Sancar, A. (1993) Molecular mechanisms of transcription-repair coupling. *Science,* **260,** 53-58.

Selby, C.P. and Sancar, A. (2006) A cryptochrome/photolyase class of enzymes with single-stranded DNA-specific photolyase activity. *Proc. Natl. Acad. Sci. USA,* **103,** 17696-17700.

Shibutani, S., Takeshita, M. and Grollman, A.P. (1991*)* Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *Nature,* **349,** 431-434.

Sijbers, A.M., de Laat, W.L., Ariza, R.R., Biggerstaff, M., Wei, Y.F., Moggs, J.G., Carter, K.C., Shell, B.K., Evans, E., de Jong, M.C., Rademakers, S., De Rooij, J., Jaspers, N.G., Hoeijmakers, J.H. and Wood, R.D. (1996) Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonucleases. *Cell,* **86,** 811-822.

Singh, S., Folkers, G.E., Bonvin, A.M., Boelens, R., Wechselberger, R., Niztayev, A. and Kaptein, R. (2002) Solution structure and DNA-binding properties of the C-terminal domain of UvrC from *E. coli*. *EMBO J.,* **21,** 6257-6266.

Skorvaga, M., Theis, K., Mandavilli, B.S., Kisker, C. and Van Houten, B. (2002) The beta-hairpin motif of UvrB is essential for DNA binding, damage processing and UvrC-mediated incisions. *J. Biol. Chem.,* **277,** 1553-1559.

Skorvaga, M., DellaVecchia, M.J., Croteau, D.L., Theis, K., Truglio, J.J., Mandavilli, B.S., Kisker, C. and Van Houten, B. (2004) Identification of residues within UvrB that are important for efficient DNA binding and damage processing. *J. Biol. Chem.,* **279,** 51574-51580.

Smith, A.J., Szczelkun, M.D. and Savery, N.J. (2007) Controlling the motor activity of a transcription-repair coupling factor: autoinhibition and the role of RNA polymerase. *Nucleic Acids Res.,* **35,** 1802-1811.

Snowden, A., Kow, Y.W. and van Houten, B. (1990) Damage repertoire of the Escherichia coli UvrABC nuclease complex includes abasic sites, base-damage analogues, and lesions containing adjacent 5' or 3' nicks. *Biochemistry,* **29,** 7251-7259.

Sohi, M., Alexandrovich, A., Moolenaar, G.F., Visse, R., Goosen, N., Vernede, X., Fontecilla-Camps, J., Champness, J. and Sanderson, M.R. (2000) Crystal structure of Escherichia coli UvrB C-terminal domain, and a model for UvrB-UvrC interaction. *FEBS Lett.,* **465,** 161-164.

Stanewsky, R., Kaneko, M., Emery, P., Beretta, B., Wager-Smith, K., Kay, S.A., Rosbash, M. and Hall, J.C. (1998) The cryb mutation identifies cryptochrome as a circadian photoreceptor in Drosophila. *Cell,* **95,** 681-692.

Steitz, T.A. and Steitz, J.A. (1993) A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad Sci. USA,* **90,** 6498-6502.

Sugasawa, K., Ng, J.M., Masutani, C., Iwai, S., van der Spek, P.J., Eker, A.P., Hanaoka, F., Bootsma, D. and Hoeijmakers, J.H. (1998) Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Mol. Cell,* **2,** 223-232.

Sugasawa, K., Okamoto, T., Shimizu, Y., Masutani, C., Iwai, S. and Hanaoka, F. (2001) A multistep damage recognition mechanism for global genomic nucleotide excision repair. *Genes Dev.,* **15,** 507-521.

Tamada, T., Kitadokoro, K., Higuchi, Y., Inaka, K., Yasui, A., de Ruiter, P.E., Eker, A.P. and Miki, K. (1997) Crystal structure of DNA photolyase from Anacystis nidulans. *Nat. Struct. Biol.,* **4,** 887-891.

Tang, Ms., Nazimiec, M., Ye, X., Iyer, G.H., Eveleigh, J., Zheng, Y., Zhou, W. and Tang, Y.Y. (2001) Two forms of UvrC protein with different double-stranded DNA binding affinities. *J. Biol. Chem.,* **276,** 3904-3910.

Theis, K., Chen, P.J., Skorvaga, M., Van Houten, B. and Kisker, C. (1999) Crystal structure of UvrB, a DNA helicase adapted for nucleotide excision repair. *EMBO J.,* **18,** 6899-6907.

Thiagalingam, S. and Grossman, L. (1991) Both ATPase sites of Escherichia coli UvrA have functional roles in nucleotide excision repair. *J. Biol. Chem.,* **266,** 11395-11403.

Trewick, S.C., Henshaw, T.F., Hausinger, R.P., Lindahl, T. and Sedgwick, B. (2002) Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. *Nature,* **419,** 174-178.

Tripsianes, K., Folkers, G., Ab, E., Das, D., Odijk, H., Jaspers, N.G., Hoeijmakers, J.H., Kaptein, R. and Boelens, R. (2005) The structure of the human ERCC1/XPF interaction domains reveals a complementary role for the two proteins in nucleotide excision repair. *Structure,* **13,** 1849-1858.

Truglio, J.J., Croteau, D.L., Skorvaga, M., DellaVecchia, M.J., Theis, K., Mandavilli, B.S., Van Houten, B. and Kisker, C. (2004) Interactions between UvrA and UvrB: the role of UvrB's domain 2 in nucleotide excision repair. *EMBO J.,* **23,** 2498-2509.

Truglio, J.J., Rhau, B., Croteau, D.L., Wang, L., Skorvaga, M., Karakas, E., DellaVecchia, M.J., Wang, H., Van Houten, B. and Kisker, C. (2005) Structural insights into the first incision reaction during nucleotide excision repair. *EMBO J.,* **24,** 885-894.

Truglio, J.J., Croteau, D.L., Van Houten, B. and Kisker, C. (2006a) Prokaryotic nucleotide excision repair: the UvrABC system. *Chem. Rev.,* **106,** 233-252.

Truglio, J.J., Karakas, E. Rhau, B., Wang, H., DellaVecchia, M.J., Van Houten, B. and Kisker, C. (2006b) Structural basis for DNA recognition and processing by UvrB. *Nat. Struct. Mol. Biol.,* **13,** 360-364.

Tsodikov, O.V., Enzlin, J.H., Schärer, O.D. and Ellenberger, T. (2005) Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF-ERCC1. *Proc. Natl. Acad. Sci. USA,* **102,** 11236-11241.

Tsodikov, O.V., Ivanov, D., Orelli, B., Staresincic, L., Shoshani, I., Oberman, R., Schärer, O.D., Wagner, G. and Ellenberger, T. (2007) Structural basis for the recruitment of ERCC1-XPF to nucleotide excision repair complexes by XPA. *EMBO J.,* **26,** 4768-4776.

van der Horst, G.T., Muijtjens, M., Kobayashi, K., Takano, R., Kanno, S., Takao, M., de Wit, J., Verkerk, A., Eker, A.P., van Leenen, D., Buijs, R., Bootsma, D., Hoeijmakers, J.H. and Yasui, A. (1999) Mammalian Cry1 and Cry2 are essential for maintenance of circadian rhythms. *Nature,* **398,** 627-630.

van der Spek, P.J., Kobayashi, K., Bootsma, D., Takao, M., Eker, A.P. and Yasui, A. (1996) Cloning, tissue expression, and mapping of a human photolyase homolog with similarity to plant blue-light receptors. *Genomics,* **37,** 177-182.

Van Houten, B., Gamper, H., Sancar, A. and Hearst, J.E. (1987) DNase I footprint of ABC excinuclease. *J. Biol. Chem.,* **262,** 13180-13187.

Van Houten, B., Croteau, D.L., DellaVecchia, M.J., Wang, H. and Kisker, C. (2005) 'Close-fitting sleeves': DNA damage recognition by the UvrABC nuclease system. *Mutat. Res.,* **577,** 92-117.

Velankar, S.S., Soultanas, P., Dillingham, M.S., Subramanya, H.S. and Wigley, D.B. (1999) Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell,* **97,** 75-84.

Venema, J., van Hoffen, A., Natarajan, A.T., van Zeeland, A.A. and Mullenders, L.H. (1990) The residual repair capacity of xeroderma pigmentosum complementation group C fibroblasts is highly specific for transcriptionally active DNA. *Nucleic Acids Res.,* **18,** 443-448.

Verdine, G.L. and Bruner, S.D. (1997) How do DNA repair proteins locate damaged bases in the genome? *Chem. Biol.,* **4,** 329-334.

Verhoeven, E.E.A., van Kesteren, M., Moolenaar, G.F., Visse, R. and Goosen, N. (2000) Catalytic sites for 3' and 5' incision of Escherichia coli nucleotide excision repair are both located in UvrC. *J. Biol. Chem.,* **275,** 5120-5123.

Verhoeven, E.E., Wyman, C., Moolenaar, G.F., Hoeijmakers, J.H. and Goosen, N. (2001) Architecture of nucleotide excision repair complexes: DNA is wrapped by UvrB before and after damage recognition. *EMBO J.,* **20,** 601-611.

Verhoeven, E.E.A., Wyman, C., Moolenaar, G.F. and Goosen, N. (2002a) The presence of two UvrB subunits in the UvrAB complex ensures damage detection in both DNA strands. *EMBO J.,* **21,** 4196-4205.

Verhoeven, E.E.A., van Kesteren, M., Turner, J.J., van der Marel, G.A., van Boom, J.H., Moolenaar, G.F. and Goosen, N. (2002b) The C-terminal region of *Escherichia coli* UvrC contributes to the flexibility of the UvrABC nucleotide excision repair system. *Nucleic Acids Res.,* **30,** 2492-2500.

Visse, R., de Ruijter, M., Ubbink, M., Brandsma, J.A. and van de Putte, P. (1993) The first zinc-binding domain of UvrA is not essential for UvrABC-mediated DNA excision repair. *Mutat. Res.,* **294,** 263-274.

Volker, M., Moné, M.J., Karmakar, P., van Hoffen, A., Schul, W., Vermeulen, W., Hoeijmakers, J.H., van Driel, R., van Zeeland, A.A. and Mullenders, L.H. (2001) Sequential assembly of the nucleotide excision repair factors in vivo. *Mol. Cell,* **8,** 213-224.

Wagner, R. and Meselson, M. (1976) Repair tracts in mismatched DNA heteroduplexes. *Proc. Natl. Acad. Sci. USA,* **73,** 4135-4139.

Wakasugi, M. and Sancar, A. (1999) Order of assembly of human DNA repair excision nuclease. *J. Biol. Chem.,* **274,** 18759-18768.

Wang, J., Mueller, K.L. and Grossman, L. (1994) A mutational study of the C-terminal zinc-finger motif of the

Escherichia coli UvrA protein. *J. Biol. Chem.,* **269,** 10771-10775.

Wang, H., DellaVecchia, M.J., Skorvaga, M., Croteau, D.L., Erie, D.A. and Van Houten, B. (2006) UvrB domain 4, an autoinhibitory gate for regulation of DNA binding and ATPase activity. *J. Biol. Chem.,* **281,** 15227-15237.

Waters, T.R., Eryilmaz, J., Geddes, S. and Barrett, T.E. (2006) Damage detection by the UvrABC pathway: Crystal structure of UvrB bound to fluorescein-adducted DNA. *FEBS Lett.,* **580,** 6423-6427.

Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature,* **171,** 737-738.

Winkler, G.S., Araujo, S.J., Fiedler, U., Vermeulen, W., Coin, F., Egly, J.M., Hoeijmakers, J.H., Wood, R.D., Timmers, H.T. and Weeda, G. (2000) TFIIH with inactive XPD helicase functions in transcription initiation but is defective in DNA repair. *J. Biol. Chem.,* **275,** 4258-4266.

Witkin, E.M. (1966) Radiation-induced mutations and their repair. *Science,* **152,** 1345-1353.

Wolski, S.C., Kuper, J., Hänzelmann, P., Truglio, J.J., Croteau, D.L., Van Houten, B. and Kisker, C. (2008) Crystal structure of the FeS cluster-containing nucleotide excision repair helicase XPD. *PLoS Biol.,* **6,** 1332-1342.

Worthington, E.N., Kavakli, I.H., Berrocal-Tito, G., Bondo, B.E. and Sancar, A. (2003) Purification and characterization of three members of the photolyase/cryptochrome family glue-light photoreceptors from *Vibrio cholerae*. *J. Biol. Chem.,* **278,** 39143-39154.

Yang, W. (2006) Poor base stacking at DNA lesions may initiate recognition by many repair proteins. *DNA repair,* **5,** 2006.

Yu, B., Edstrom, W.C., Benach, J., Hamuro, Y., Weber, P.C., Gibney, B.R. and Hunt, J.F. (2006) Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature,* **439,** 879-885.

Zhao, X., Liu, J., Hsu, D.S., Zhao, S., Taylor, J.S. and Sancar A. (1997) Reaction mechanism of (6-4) photolyase. *J. Biol. Chem.,* **272,** 32580-32590.

Zharkov, D.O., Ishchenko, A.A., Douglas, K.T. and Nevinsky, G.A. (2003) Recognition of damaged DNA by Escherichia coli Fpg protein: insights from structural and kinetic data. *Mutat. Res.,* **531,** 141-156.