Structure-based insights into the repair of UV-damaged DNA

Meulenbroek, E.M.

1 Introduction

1.1 DNA under attack

Our genetic information is stored in our DNA. From the color of our eyes to the proteins involved in the energy production in our cells, all these things are coded for in this huge macromolecule in a language consisting of only 4 letters (bases): adenine (A), cytosine (C), guanine (G) and thymine (T). This information is faithfully copied by DNA polymerase to a new DNA molecule upon cell division, and in a process called transcription, this information is read by RNA polymerase to form a RNA molecule that can then be translated by the ribosome into proteins, the workhorses in our cells. For both these processes, a characteristic of nucleic acids (such as DNA and RNA) called base pairing is used: an A is complementary to a T (and to an uracil, U, in RNA) and a G to a C. In the double helix of the DNA, an A is hence always found opposite to a T and a C to a G. This conveys an elegant read-out mechanism for the DNA and RNA polymerase: if they encounter a A in the template, they put a T (or U in RNA) opposite to it in the nascent DNA or RNA, and a C opposite to a G and *vice versa*. As can be expected for such an important information holder as DNA, it is a relatively stable macromolecule, helping to preserve our genetic information.

However, there are influences from both outside and inside that cause damage in this stable macromolecule: UV-irradiation, for example. Staying in the bright sun can cause DNA damage, especially due to the UV-B component of sunlight, which ranges from 280-320 nm. The major lesions that this radiation causes are crosslinks between adjacent pyrimidines (C and T bases): cyclobutane pyrimidine dimers (CPDs) and 6-4 photoproducts (6-4PPs) (Yoon *et al.*, 2000). Such lesions interfere with the correct reading of the genetic information, since they, for example, block the transcription of the DNA into RNA, they block replication of the DNA and they can cause mutations.

Other types of radiation also cause damage in the DNA, with ionizing radiation such as γ-radiation being the most dangerous. This type of radiation can cause double-strand breaks in the DNA, which is of course detrimental to the information encoded in the molecule (Rich *et al.*, 2000). Besides radiation there are also other DNA damaging agents in our surroundings. Smoking, for example, causes the intake of polycyclic aromatic hydrocarbons, which lead to the formation of diol epoxide - DNA adducts that are highly tumorigenic (Zhong *et al.*, 2011).

Not only influences from the outside cause damage in our DNA, but metabolic processes also contribute to DNA lesions, since they produce free oxygen radicals and other reactive compounds that can cause oxidative damage to the bases that can lead to the loss of the correct coding information at that spot.

As stated above, DNA is a relatively stable macromolecule, so the percentage of lesions resulting from the above mentioned causes is low. However, the sheer amount of DNA in a cell (approximately 6 billion base pairs in a human cell) makes this low percentage accumulate to form for example about 10.000 base losses per cell per day (Nouspikel, 2009a). This forms a serious problem for a cell, because an unrepaired lesion can give rise to a mutation. For example, if a guanine is methylated at the O^6 position, it looks more like an adenine, hence upon DNA replication a thymine will be erroneously incorporated in the new DNA molecule leading to GC to AT mutations (Warren *et al.*, 2006). In many regions of the DNA a single-base substitution might not represent a major problem, but if the error occurs at the wrong spot in a gene, the information encoded by the gene might be changed in a detrimental way, since even a single base substitution can lead to a severe disease. Cystic fibrosis is just one of the many examples (Lommatzsch *et al.*, 2009).

Another possible consequence of DNA lesions is cell death. The block in replication or transcription caused by unrepaired lesions is a strong signal for apoptosis. In other words, the cell will commit organized suicide. This is actually a safety mechanism for the cell to prevent it from continuing to divide after its genetic information has been changed, but it is also a process that contributes to aging.

To prevent mutation and cell death it is thus crucial that damage in the DNA is repaired. Therefore, several DNA repair pathways have evolved, each with their own specificity for particular lesions. Below, the major DNA repair pathways will be discussed. Most of these pathways are conserved from bacteria to humans, but in this chapter an emphasis will be given on the system in humans. Examples will be shown to demonstrate the contribution to the understanding of these systems that structural biology on macromolecules in the pathways has made.

1.2 Direct reversal

The conceptually simplest pathway for DNA repair is the pathway of direct reversal (DR). As the name suggests, this pathway is responsible for directly reversing a DNA lesion. There are two main types of direct DNA damage reversal systems: the repair of alkylation damage by alkyltransferases and the repair of UV-induced photolesions by photolyases.

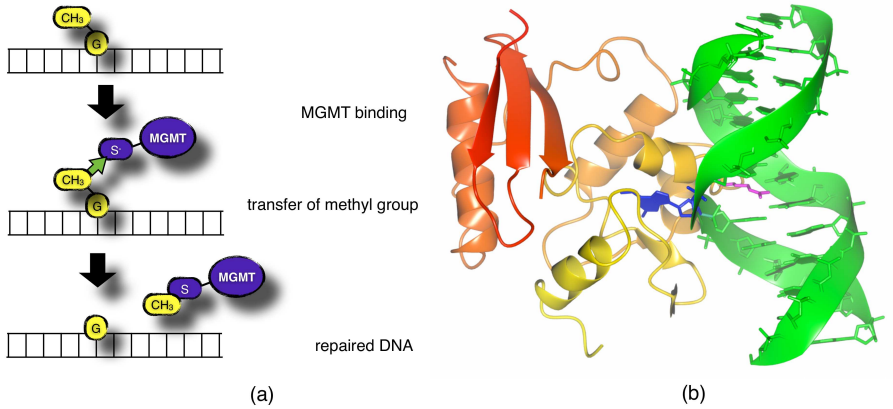

Figure 1.1: Direct reversal
(a) Mechanism of direct reversal as performed by the O^6-meG-DNA methyl transferase (MGMT), based on Eker et al., 2009. See text for detailed explanation.
(b) Crystal structure of MGMT (red-yellow) bound to damaged DNA (green) showing nucleotide flipping with the damaged base shown in blue and the arginine intruding into the helix is shown in magenta. Structure 1t38 was used for this figure.

A representative example of the first type is formed by the enzyme O^6-meG-DNA methyl transferase (MGMT), also referred to as alkylguanine transferase (AGT). This protein repairs an O^6-methylguanine lesion by irreversibly transferring the methyl group of O^6-methylguanine to an active site cysteine, which restores the guanine base to its original form (see Figure 1.1(a)). Methylated MGMT cannot be reused and is rapidly degraded (Eker *et al.*, 2009).

The crystal structure of human MGMT in complex with damaged DNA gives insight into the details of this mechanism (Daniels *et al.*, 2004). In the structure it can be seen that the damaged base is flipped out of the helix into the active site of the enzyme while an arginine of the protein takes the place of the damaged base into the DNA duplex thus stabilizing the extrahelical conformation of the lesion (Figure 1.1(b)). Widening of the minor groove and bending of the DNA upon binding of MGMT's helix-turn-helix motif to the DNA's minor groove facilitates this nucleotide flipping. Also a conserved tyrosine aids in the flipping by stimulating phosphate rotation. The specificity of this enzyme for O^6-alkyl guanines can be explained by the interactions of the protein with the damaged base in the active site.

Another example of direct repair is the repair of pyrimidine dimers by photolyases. These proteins reverse the photolesions CPD and 6-4PP that are caused by the UV-component of sunlight using the visible component of the very same source. A certain photolyase is specific for either a CPD or for a 6-4PP, but either type of photolyase works via a similar mechanism. From crystal structures of photolyases (reviewed in Müller & Carell, 2009) it was found that they first bind to a site of damage, open up the DNA-duplex structure at the damaged site and flip the

dinucleotide lesion out of the duplex into the active site close to a FAD cofactor. The reaction is then predicted to continue as follows (Brettel & Byrdin, 2010): absorption of a photon in the FAD reaction center or in the second cofactor that functions as auxiliary light-harvesting antenna (8-HDF or MTHF) leads to the transfer of an electron to the lesion that destabilizes the damage. The formation of this lesion radical leads to the breaking of the bond between the two pyrimidines and the electron is then transferred back to the cofactor, yielding a repaired DNA substrate and a photolyase that is back in its ground state. Photolyases have been found in bacterial systems, fungi and animals, but not in placental mammals such as humans.

1.3 Base excision repair

Base excision repair (BER) is the main pathway that repairs lesions due to oxidation, alkylation, deamination and depurination/ depyrimidation of bases in the DNA. It functions in three steps (Figure 1.2(a); reviewed in Robertson *et al.*, 2009). First a specialized glycosylase recognizes the damaged base and removes it by cutting the N-glycosidic bond, creating an apurinic/ apyrimidinic (AP) site. One glycosylase typically recognizes only one type of lesion. In mammals up to now eleven different glycosylases have been identified. The second step involves recognition of the newly generated AP site by an AP endonuclease and this enzyme cuts the DNA backbone 5′ to the lesion. Sometimes the first two steps are carried out by a single enzyme, since some glycosylases have an AP-lyase activity. After the 5′ incision, single strand break repair is initiated. There are two possibilities for this third step of BER: a short patch and a long patch option. In the short patch option, DNA polymerase displaces the AP site and adds a nucleotide in the DNA to fill in the gap, and then it catalyzes the removal of the displaced AP site. Finally, a ligase catalyzes the closing of the gap. In the long patch option, DNA polymerase displaces and polymerizes stretches of DNA longer than one base. This produces a flapped substrate, and this flap is removed by FEN1, after which the repair can be finished by DNA ligase (Robertson *et al.*, 2009).

A detailed atomic description of the first step of BER can be seen in the crystal structure of human uracil-DNA glycosylase (UDG) in complex with its DNA substrate before and after cleavage (Parikh *et al.*, 2000). This enzyme recognizes uracil bases that erroneously enter DNA via misincorporation or cytosine deamination and subsequently excises the erroneous base, leaving behind an AP site. In the structure it can be seen that the uracil is flipped out of the helix into the enzyme's active center (Figure 1.2(b)). UDG closes around the substrate and enforces a large distortion in the uracil and the deoxyribose that promotes cleavage of the glycosylic bond, since the distortion is relieved upon cleavage.

Endonuclease IV from *E.coli* forms a nice example for the second step of BER: 5′ incision of the AP site. Its crystal structure in complex with DNA containing an AP site gives insight into its mechanism (Hosfield *et al.*, 1999). The enzyme is seen to bind DNA in a positively charged, crescent shaped groove. At the bottom of this groove, most of the conserved residues are found as well as three Zn^{2+} atoms in the enzyme's active site. Upon binding, the DNA is bent 90° (Figure 1.2(c)). The

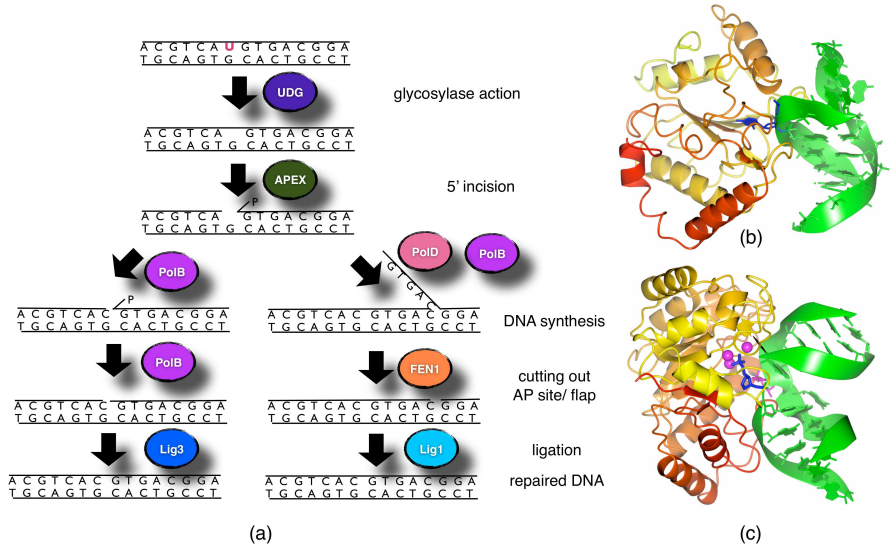

Figure 1.2: Base excision repair
(a) Mechanism of base excision repair based on Robertson et al., 2009 showing UDG as glycosylase and APEX as AP endonuclease. See text for detailed explanation.
(b) Crystal structure of uracil-DNA glycosylase (UDG; red to yellow) bound to DNA (green) showing uracil flipping with the flipped damaged base in blue. Structure 1emh was used for this figure.
(c) Crystal structure of T4 endonuclease (red to yellow) bound to DNA (green) showing the flipping of the abasic site (in blue), the three Zn atoms (in magenta) and the Tyr and Leu sticking into the helix in magenta. The structure 1qum was used for this figure.

insertion of three residues (a tyrosine, a leucine and an arginine) into the double helix assists in this bending of the DNA and helps in flipping both the damage as well as the opposite base out of the duplex. The damage is flipped into a pocket that sterically excludes undamaged bases, hence providing specificity to the enzyme. The target phosphate binds to the active site Zn^{2+} ions and is hence positioned correctly for a nucleophilic attack on the scissile P-O bond by a hydroxide ion that is activated by the Zn^{2+} ions, leading to incision 5′ to the abasic site.

1.4 Mismatch repair

Replication of DNA is done by DNA polymerase. This is a highly accurate enzyme that has a very low error rate (one error per 10^7 bases). A major factor contributing to this accuracy is the proofreading of the enzyme: it checks itself whether it has made a mistake, and it corrects the mistake when it has occurred. However, this low error rate still corresponds to quite some errors, given that in one cell division in a human cell, the polymerase has to copy around six billion bases. Other sources of mismatches are DNA lesions such as the spontaneous addition of a methyl group

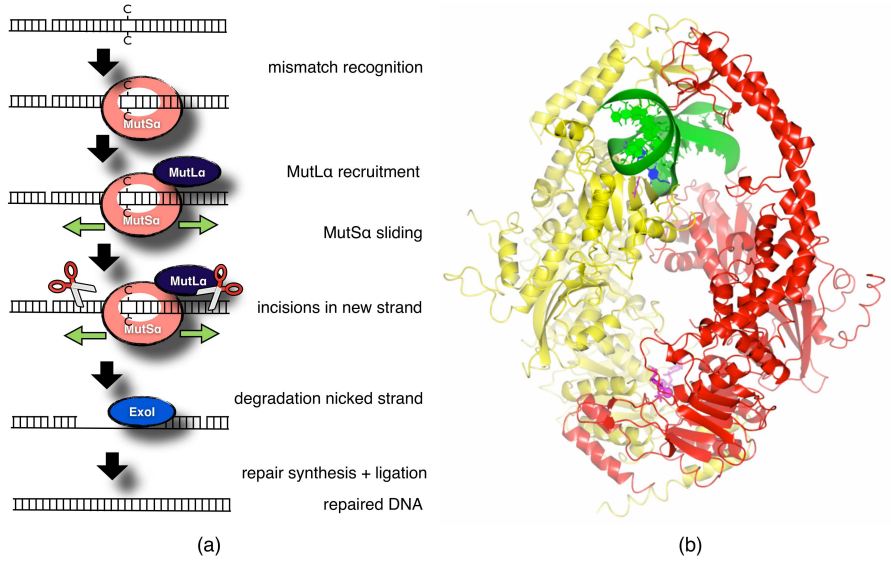

Figure 1.3: Mismatch repair
(a) Mechanism of mismatch repair based on Kunz et al., 2009. Detailed explanation can be found in the text.
(b) Crystal structure of MutSα in complex with DNA (green) with MSH2 in red and MSH6 in yellow, the conserved Glu434 in magenta, ADP in magenta, DNA in green and the mismatch in blue. The structure 2o8b was used for this figure.

to a cytosine followed by deamination, which yields a thymine and hence a G-T mismatch.

To correct errors made in DNA replication, the mismatch repair system (MMR) has evolved (see Figure 1.3(a)). In this pathway, first the mismatch is recognized. In humans this is done by the heterodimeric complex MutSα, which consists of MSH2 and MSH6. After mismatch recognition, the MLH1-PMS2 dimer called MutLα is recruited. Cycles of ATP binding, hydrolysis and ADP-ATP exchange in MutSα then leads to a sequence of conformational changes in MutSα that enables it to slide along the DNA away from the mismatch. It has been proposed that it looks for the discontinuous strand, which is the newly synthesized strand and hence the strand containing the erroneous base. Latent endonuclease activity in the PMS2 subunit of MutLα is activated and a nick is introduced in the strand containing the discontinuity. Subsequently, a stretch of DNA in the nicked strand is degraded by EXO1, after which the gap is filled by DNA polymerase δ and then sealed by DNA Ligase I (Kunz *et al.*, 2009). Mismatches that arise from DNA damage, such as the above-mentioned G-T mismatch, are corrected by BER (section 1.3), which has a special glycosylase for excision of T from G-T mismatches.

Clues as to how exactly a mismatch is recognized by the MMR system come

from the crystal structure of MutSα in complex with G-T mismatch containing DNA (Warren *et al.*, 2007). This structure shows that MutSα forms an asymmetric oval disc with two channels in the center. The ATPase sites are found at one end of the oval and the DNA containing the mismatch in the channel at the other end. The DNA is bent about 45° (Figure 1.3(b)). Only MSH6 interacts with the mismatch and a glutamate of a conserved Phe-X-Glu motif of MSH6 hydrogen bonds to the mispaired thymine, which is also sandwiched between two hydrophobic residues, while a backbone carbonyl interacts with the mispaired G. Many non-specific protein-DNA interactions are also seen that by itself are probably sufficient for binding to deformable DNA substrates, since MutSα can also bind C-C mismatches that cannot make all the above interactions. Besides this, co-crystal structures with a number of very different DNA substrates show virtually identical complexes, suggesting that MutSα recognizes a global feature such as helix instability (Warren *et al.*, 2007).

1.5 Double-strand break repair

Probably the most detrimental lesion for a cell is a double-strand break (DSB): the DNA has been broken in both the strands of the double helix. Even a single unrepaired double strand break can lead to cell death (Rich *et al.*, 2000). At least two repair pathways have evolved to deal with this deleterious lesion (see Figure 1.4(a)): Non-Homologous End-Joining (NHEJ) and Homologous Recombination (HR) (reviewed in Pardo *et al.*, 2009). The choice between these two pathways depends on, amongst others, the stage of the cell cycle.

NHEJ is the most general applicable pathway of the two, because it can religate any two ends together without the requirement of a homologous sequence (Pardo *et al.*, 2009). In this pathway, the MRN (Mre11-Rad50-Nbs1) and the DNA-PK/Ku (DNA-dependent protein kinase/ Ku) complexes bind to the ends of the double strand break shortly after formation of the break. These protein complexes bridge together the ends of the break and prevent their degradation. Different alignments of the ends are attempted. Subsequently, DNA ligase IV is recruited to perform the ligation. If ligation is not immediately possible, the ligase and the DNA-Pk/Ku complexes recruit DNA-processing enzymes that for example correct for non-ligatable ends of the DNA molecules (missing the 5′ phosphate etc.) or fill in missing parts of complementary strands and then the two ends are ligated together.

The second pathway for repair of DSBs is the Homologous Recombination (HR) pathway in which an intact homologous DNA molecule is used as a template for repair. Also in this pathway the ends are stabilized by MRN and DNA-Pk/Ku, but then nucleolytic degradation of the 5′ ends takes place to generate a long 3′ single-stranded end, to which the ssDNA binding protein RPA binds. RPA is then displaced by a filament of Rad51 (also involving Rad52, Rad54 and Rad55-57), which is responsible for the homology search (to find a template for the repair, such as the sister chromatid during the S and G2 stages of the cell cycle), and for invasion of the 3′ single stranded end into an intact template DNA molecule. This forms a

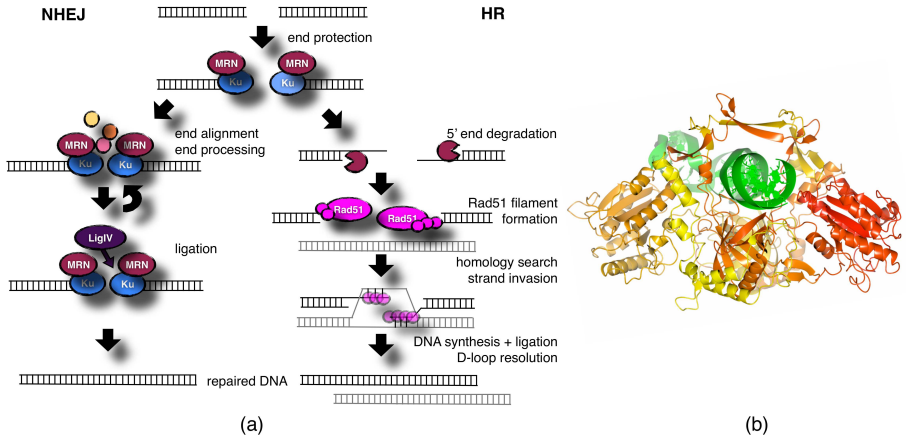

Figure 1.4: Double-strand break repair
(a) Mechanism of double strand break repair based on Pardo et al., 2009. Details can be found in the text.
(b) Crystal structure of the Ku heterodimer (in red to yellow) bound to DNA (green). Structure 1jey was used to make this figure.

D-loop, in which DNA polymerase fills in the missing DNA stretches. After this, the ends are ligated and the D-loop is resolved, giving a repaired DNA molecule.

Protein crystallography gives some insights into the mechanism of DSB repair, in particular, on the function of the Ku70-Ku80 dimer (Ku). This complex binds double strand breaks shortly after their formation and protects them, but it still allows end-processing to occur, so apparently it does not hide the entire DNA end. The crystal structure of the Ku heterodimer bound to DNA (Walker *et al.*, 2001) provides an explanation: it was shown that Ku forms a positively charged cradle into which DNA lies and also bridges around the DNA (Figure 1.4(b)). One part of the DNA is covered with Ku (the bottom part in Figure 1.4(b)), but the rest of the DNA is accessible for end-processing and end-joining reactions. The observation that the Ku heterodimer forms a ring around the DNA immediately explains why it only can bind to DSBs and not to a DNA molecule without end.

1.6 Nucleotide excision repair

The array of DNA lesions is very large and diverse. The pathway of Nucleotide Excision Repair (NER) solves this problem (reviewed in Nouspikel, 2009b): this pathway can recognize a large range of lesions and excise the stretch of DNA containing the damage. The main feature of the damage to be recognized seems to be that it has to disturb the double helix to a certain extent (e.g. by disrupted base-pairing and/ or stacking): it is more this distortion that is recognized than the exact identity of the lesion.

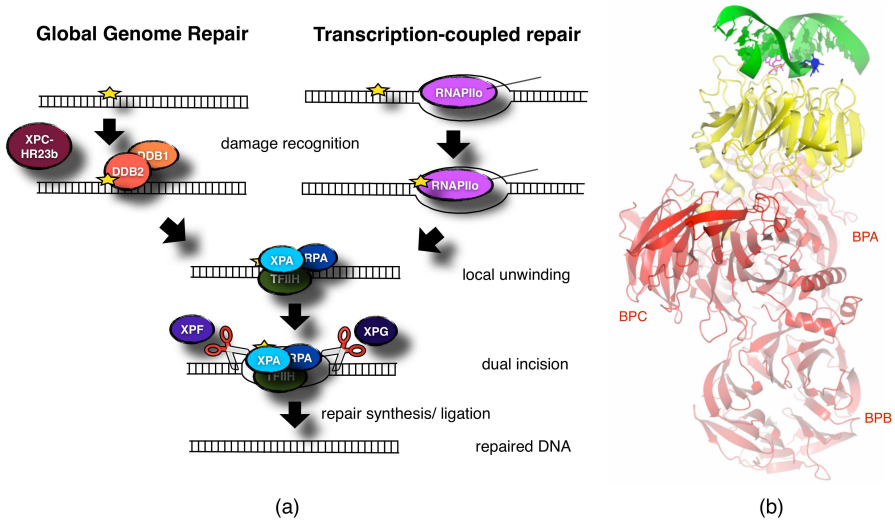

Figure 1.5: Nucleotide excision repair
(a) Mechanism of nucleotide excision repair, showing both the global genome repair branch and the transcription-coupled repair branch based on Nouspikel et al., 2009b (see text for details). (b) Crystal structure of DDB1 (red; the domains are indicated as BPA, BPB and BPC) and DDB2 (yellow) bound to DNA (green) containing a 6-4PP lesion (blue). The DDB2 hairpin inserting into the helix is shown in magenta. For this figure structure 3ei1 was used.

In humans it is the XPC-HR23B-Cen2 complex that most often does the initial recognition (see Figure 1.5(a)). It flips out the damaged bases and the bases opposite to it and it then binds to the bases opposite to the lesion in a non-sequence dependent manner (Min *et al.*, 2007), showing that it recognizes the helix-distorting properties of the damage and not the damage itself. Not all lesions distort the helix to the same extent and therefore, there is the DDB1-DDB2 complex for the recognition of smaller lesions (such as UV-photoproducts). DDB1-DDB2 is part of an E3-ubiquitin ligase complex. For such smaller damages, it is the DDB1-DDB2 complex that first recognizes the lesion and binds to it, with the lesion being flipped into a shallow pocket that determines DDB2's specificity (Scrima *et al*, 2009; Figure 1.5(b)). After this, it recruits XPC-HR23B-Cen2 to the damage. Then both XPC and DDB2 get ubiquitinated. This leads to the degradation of DDB2, while XPC binds stronger to the DNA after ubiquitination and thus the lesion is transferred to XPC (Sugasawa *et al.*, 2005).

Subsequently, Transcription Factor II H (TFIIH) is recruited to the site of damage, which opens up a bubble of around 30 nucleotides using the ATPase activity of its subunit XPB to anchor TFIIH to the DNA and the 5' to 3' helicase activity of its subunit XPD to open up the DNA (Egly & Coin, 2011). The blockage of this translocation of XPD by a genuine damage has been proposed to be involved in verification of the presence of a damage (Sugasawa *et al.*, 2009). XPA then replaces

XPC and might be involved in additional damage verification by probing for appropriately distorted DNA (Gillet *et al.*, 2006). RPA is recruited to the site as well and binds the single stranded DNA in the repair bubble. XPF-ERCC1 then incises the DNA strand containing the lesion 5′ to the damage (Matsunaga *et al.*, 1995) and XPG 3′ to the damage (O'Donovan *et al.*, 1994). A stretch of DNA of around 25-30 nucleotides is then taken out and is later filled in by DNA polymerase δ, ϵ or κ and ligated, yielding a corrected DNA (Nouspikel, 2009b).

How the DDB1-DDB2 complex recognizes photolesions (CPD and 6-4PP) is shown from its crystal structure in complex with DNA containing a 6-4PP (Scrima *et al.*, 2008). DDB1's three WD40 domains fold into three β-propeller domains (BPA, BPB and BPC). DDB2 also has a β-propeller fold and binds to DDB1 using a N-terminal helix-loop-helix motif inserted between the BPA and BPC domains and a large hydrophobic surface patch on the bottom face of DDB2. In the crystal structure it can be seen that the lesion is held exclusively by the WD40 domain of DDB2, and more precisely its top face (Figure 1.5(b)), which has extensive positive charges. A small hairpin of DDB2 intrudes in the DNA's minor groove and flips the damage into a shallow pocket on the protein's surface, which would exclude larger lesions, hence explaining DDB1-DDB2's specificity. The DNA duplex is kinked by about 40° upon DDB2 binding.

1.7 Transcription-coupled repair

1.7.1 Discovery and mechanism of transcription-coupled repair

In 1985 it was noted that some lesions in the transcribed regions of the genome are repaired faster than those in the overall genome (Bohr *et al.*, 1985). Two years later it was reported that the lesions in the transcribed strand of the transcribed genes are repaired preferentially (Mellon *et al.*, 1987). This led to the establishment of the pathway of Transcription-Coupled Nucleotide Excision Repair (TC-NER), which was found to be a subpathway of NER.

This subpathway is thought to act in mammalians as in Figure 1.6 (reviewed in Fousteri & Mullenders, 2008). When RNA polymerase is transcribing the genome, it is dynamically interacting with Cockayne Syndrome protein B (CSB). Upon encountering a lesion past which it cannot transcribe, this interaction is stabilized and CSB recruits chromatin remodelers and Cockayne Syndrome protein A (CSA) in complex with DNA Damage Binding protein 1 (DDB1) as part of an E3 ubiquitin ligase complex inactive at the time of recruitment. Also the NER factors are recruited that repair the lesion by excising a short stretch of DNA containing the lesion (as described in section 1.6). Once the lesion is repaired, RNA polymerase can resume transcription (Fousteri *et al.*, 2006).

In bacteria such as *E. coli*, it is the product of the *mfd* gene, transcription-repair coupling factor (TRCF), that starts repair upon stalling of RNA polymerase (Selby *et al.*, 1991). This protein detects stalled RNA polymerase and then recruits NER factors (UvrA and UvrB in *E. coli*). In contrast to the mammalian system, TRCF releases stalled RNA polymerase from the DNA in an ATP-dependent fashion to

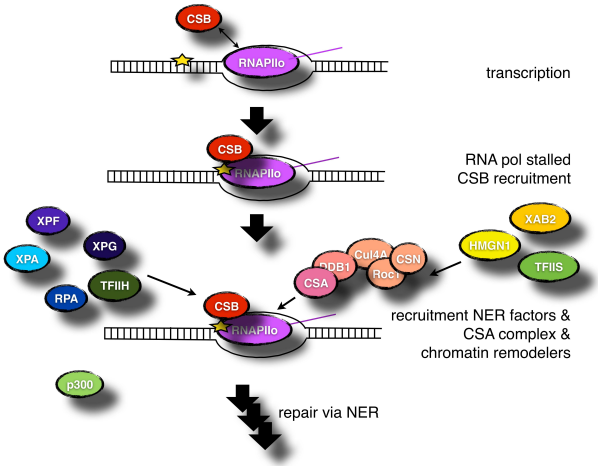

Figure 1.6: Transcription-coupled repair
Mechanism of transcription-coupled repair showing the roles of CSA and CSB (figure based on Fousteri et al., 2008). Details can be found in the text.

allow room for repair (Selby & Sancar, 1993). As in mammalians, Nucleotide Excision Repair then finishes the repair by excision and filling up the stretch of DNA containing the lesion. No factor equivalent to CSA in humans has been found in bacteria.

TC-NER repairs lesions that stall transcription such as UV-induced lesions and bulky chemical adducts such as polycyclic aromatic hydrocarbons. Most studies on TC-NER are conducted using UV-lesions. Cellular studies have shown that CSA- or CSB-deficient cells are not only sensitive to UV-irradiation, but are also hypersensitive to oxidative stress (D'Errico *et al.*, 2007; Tuo *et al.*, 2003). However, it is generally thought that RNA polymerase does not stall at oxidative lesions (Kathe *et al.*, 2004). Normally oxidative lesions are repaired by BER (section 1.3). It has been hypothesized that either the CS proteins are also involved in BER or that oxidative lesions are nonetheless repaired by TC-NER. However, repair of oxidative lesions in TC-NER is currently an area filled with controversies due to the retraction of several key papers in this area (discussed in Stevsner *et al.*, 2008). In bacteria, it has been reported that TRCF is not involved in the repair of oxidative lesions (Schalow *et al.*, 2012).

1.7.2 Cockayne syndrome proteins A and B

Most of the pathway of TC-NER is identical to Global Genome Nucleotide Excision Repair (GG-NER). The major difference is the initial step of damage recognition, which is done in TC-NER by a stalled RNA polymerase. The two factors unique to TC-NER, besides RNA polymerase, are CSA and CSB, which will be described below in more detail.

The gene for CSB is located on chromosome 10.q11-21 and encodes a protein of 1493 amino acids (168 kDa)(Troelstra *et al*., 1992). As shown Figure 1.7(b), several motifs can be recognized in the CSB sequence such as an acidic region of CSB (with a currently unknown function), a glycine-rich stretch (presumably needed to yield conformational flexibility to the N-terminal part of the protein), a bipartite nuclear localization signal and an ATPase domain (Licht *et al*., 2003). The latter domain consists of seven ATPase motifs and places CSB in the SWI2/SNF2 family of DNA-dependent ATPase. Proteins of this family usually are molecular motors that use energy of ATP hydrolysis to remodel protein-DNA interfaces (Hauk *et al*., 2011). *In vitro* studies on CSB have shown that CSB indeed has a DNA-stimulated ATPase activity, but no helicase activity (Citterio *et al*., 1998). Dephosphorylation of CSB, which occurs in response to UV-irradiation, leads to increased ATPase activity and hence provides a regulatory mechanism (Christiansen *et al*., 2003). Another interesting feature of CSB is that it contains a ubiquitin-binding domain (UBD) that is crucial for repair: without it, the TC-NER complexes fully assemble, but no repair is initiated, and the repair can be restored by a heterologous UBD on CSB (Anindya *et al*., 2010).

The best-known function of CSB is in TC-NER. CSB has been shown to transiently interact with RNA polymerase during transcription and this interaction is prolonged when there is transcription arrest (Van den Boom *et al*., 2004). CSB then initiates TC-NER by recruiting other TC-NER factors to the site of the stalled RNA polymerase: histone acetyltransferase p300, NER proteins, and CSA-DDB1 E3-ubiquitin ligase complex with the COP9 signalosome (Fousteri *et al*., 2006). It has then been proposed to use its DNA-dependent ATPase activity for modifying the RNA polymerase - DNA interface to make space for the NER factors to repair the lesions, especially for the 3' incision by XPG. This is necessary in TCR because the large RNA polymerase would otherwise cover the site of the lesion (Licht *et al*., 2003).

CSB not only functions in TC-NER, but also in transcription in general. The above-mentioned transient interaction of CSB and RNA polymerase has been shown to help elongation by RNA pol I, II and possibly III (Licht *et al*., 2003), both in nuclear and mitochondrial transcription (Berquist *et al*., 2012). CSB stimulates elongation by RNA polymerase at natural pause sites or by strong RNA secondary structure and it helps the polymerase to bypass (small) lesions such as thymine glycol and 8-oxoguanine (Charlet-Berguerand *et al*., 2006). Moreover, CSB has been reported to be involved in regulation of transcription after UV or oxidative stress (Proietti-De-Santis *et al*., 2006; Kyng *et al*., 2003 respectively). A function for CSB in the transcription for some but not all promoters was found. In another study, transcription of genes involved in chromatin maintenance and remodeling was seen to be influenced by CSB (Newman *et al*., 2006).

A third function of CSB seems to be in oxidative damage repair, since CSB-deficient cells are hypersensitive to oxidative stress. CSB has therefore been suggested to have a function in BER, which is the main pathway for repair of these lesions. CSB-deficient cells were seen to have reduced repair of the oxidative damage 8-oxoG due to reduced amount of the OGG1 protein (Dianov *et al*., 1999). This

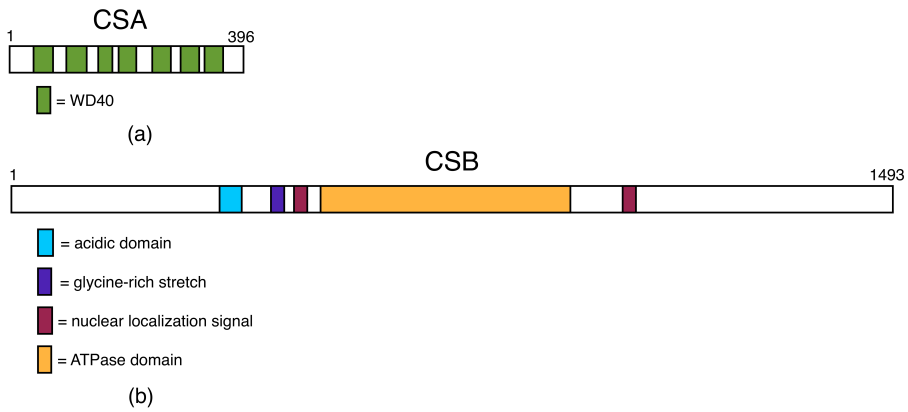

Figure 1.7: Sequence of CSA and CSB.
Schematic representation of the amino acid sequence of CSA (a) and CSB (b), showing the most important domains.

deficient repair could be corrected for with the CSB gene. The precise mechanism via which CSB functions in BER is currently not known; CSB has been reported to stimulate expression of OGG1, and the two proteins have been observed to reside in the same complex, but no direct physical interaction between them has been established. *In vivo* CSB also interacts with other BER proteins: poly(ADP-ribose) polymerase 1 (PARP-1) and apurinic/apyrimidinic endonuclease 1 (APE1) (Thorslund *et al.*, 2005, Wong *et al.*, 2007), but exactly how this influences BER is unknown.

The gene for CSA is located on chromosome 5q12-q31 (Henning *et al.*, 1995). It encodes a 396 amino acid protein (44 kDa) and it has been predicted to fold into a WD40 domain, since it contains seven WD40 repeats (see Figure 1.7(a)). Proteins with this fold are often involved in protein-protein interactions (Stirnimann *et al.*, 2010 and Xu *et al.*, 2011). In the cell CSA is found in an E3 ubiquitin ligase complex that consists of DDB1, Cul4A and Roc1 (reviewed in Scrima *et al.*, 2011), which is inactive in initial stages of repair due to binding of COP9 signalosome (Groisman *et al.*, 2003).

E3 ubiquitin ligases such as the one CSA is involved in, mediate the process of ubiquitination, which means covalently attaching one ubiquitin - a 76 amino acid, very conserved protein - or a polyubiquitin chain to a target protein. This is often a signal for proteasome-mediated destruction, but it can also alter the function of the the target protein non-proteolytically. The ubiquitination pathway starts with the ATP-dependent activation of ubiquitin by the E1 ubiquitin-activating enzyme. The human genome contains only one type of E1 enzyme. Next, the activated ubiquitin is transferred to the E2 ubiquitin-conjugating enzyme. 25 different E2 enzymes are present in humans. In the final step, a E3 ubiquitin ligase recognizes the substrate and promotes the transfer of the ubiquitin from the E2 enzyme to the substrate. The E3 enzymes thus selects the target for ubiquitination and hence determines

the specificity. There are assumed to be more than 1000 different E3 enzymes in humans with different specificities so that proteins can be selectively targeted for ubiquitination. The largest family of E3 ubiqituin ligases are the Cullin-RING ubiquitin ligases (reviewed in amongst others Lee & Zhou, 2007 and Jackson & Xiong, 2009). Seven different Cullin proteins exist in humans. With their C-terminus, Cullins bind to ROC1 (Regulator Of Cullins), which contains a RING domain (Really Interesting New Gene) for binding to the E2 enzyme. Binding of (a linker and) a substrate receptor protein that determines the specificity occurs on the N-terminus of the Cullin.

One of the linker proteins involved in such a Cullin-RING E3 ubiquitin ligase complex is DDB1 (reviewed in Iovine *et al*., 2011). It has three WD40 domains that form a triple β-propeller (Li *et al*., 2006). With its second propeller, BPB, it binds to the N-terminus of Cul4A (Angers *et al*., 2006). In a cavity between its first and third β-propeller (BPA and BPC), it binds to one of many different substrate receptors. These are called DDB1-Cul4 associated factors (DCAFs). Through proteomic studies, many DCAFs have been identified (amongst others Higa *et al*., 2006; Jin *et al*., 2006; Olma *et al*., 2009) that function in a variety of biological processes (reviewed in Scrima *et al*., 2011). They are thought to usually bind to DDB1 by inserting a helical motif between the BPA and BPC propeller of DDB1, though not all known substrate receptors for DDB1 have a sequence for such a motif, so these have either a different binding mode or the motif has passed unnoticed in their sequence (Li *et al*., 2010). The majority of DCAFs have a WD40 propeller domain, but other types of DCAFs also exist. Most of them are thought to recognize protein epitopes, or post-translational modifications of proteins, which they are assumed to usually do with the top face of their WD40 domain. This target protein is not necessarily the protein that undergoes ubiquitination, since the E3 ubquitin ligase is able to span up to 100 Å due to the rotational flexibility of the BPB domain of DDB1 versus the BPA-BPC domain (Scrima *et al*., 2008).

The most-studied DCAFs besides CSA are DDB2 and CDT2. DDB2 functions in NER, as described in section 1.6, and in contrast to most DCAFs, it binds to damaged DNA and then promotes ubiquitination of proteins in the surroundings. The Cul4A-DDB1-CDT2 E3 ligase functions in cell cycle regulation, since it is in control of S/G2 cell cycle transitions through degradation of the replication licensing factor CDT1 (Zhong *et al*., 2003). Other targets of this DCAF include cell cycle regulator p21 (Abbas *et al*., 2008) and the histone methyltransferase SET8 (Centore *et al*., 2010). The variety of cellular processes that are regulated by Cul4-DDB1 E3 ubiquitin ligases make them a nice target for pathogens. These ligases are thus sometimes hijacked by viruses in order to target specific proteins for degradation to create a nice host cell environment for virus replication and dissemination (Barry & Früh, 2006). For example, interferon-activated transcription factor STAT1 is ubiquitinated after infection with simian virus 5 due to V protein of this virus, which functions as a DCAF in its host cell and redirects the Cul4-DDB1 E3 ubiquitin ligase towards this target. The STAT family of transcription factors controls the expression of over a hundred genes with diverse antiviral functions, hence degrading these has a clear benefit for the virus (Li *et al*., 2006). Interestingly, this protein also uses helical motif

to bind to DDB1 even though the protein does not have a WD40 domain.

Cullin-RING E3 ubiquitin ligases are usually regulated by the COP9 signalosome (CSN) (reviewed in Schwechheimer, 2004). The E3 ligase activity is activated through attachment of NEDD8 (neddylation), a small-ubiquitin-like modifier. Removal of NEDD8 is done by the metallo-isopeptidase activity of the fifth subunit of the COP9 signalosome (CSN5) and thus the CSN can regulate the E3 ubiquitin ligase. The timing of the regulation of different E3 ubiquitin ligases differs. For example, for the Cul4A-DDB1-DDB2 E3 ligase, CSN dissociates from the complex upon UV-damage, after which the complex gets neddylated and activated. However, with the Cul4A-DDB1-CSA complex, CSN stays in the complex after UV-irradiation and only dissociates much later; this means that the CSA-containing E3 ubiquitin ligase only gets active at later stages of the repair.

As described above, it is clear that CSA is present in the cell in a E3 ubiquitin ligase complex and what this type of complexes does in general. However, the exact function of CSA has not yet been determined. CSA is known to recruit nucleosomal binding protein HMGN1, the transcription cleavage factor TFIIS and the pre-mRNA splicing factor XAB2 to the vicinity of damage once there are stalled RNA polymerase-CSB complexes (Fousteri *et al*., 2006). In contrast to CSB, CSA is not required for the assembly of NER pre-incision factors TFIIH, XPA, RPA and XPF-ERCC1. As other possible functions for CSA it has been proposed that it protects TC-NER factors from degradation at the initial stages of repair and/ or that it causes the removal of some of these factors in later stages of repair so that RNA polymerase can resume transcription.

It has been proposed that CSA is involved in degradation of the stalled RNA polymerase to make room for repair, since CSA co-localizes with RNA polymerase upon UV-irradiation (Kamiuchi *et al*., 2002) and RNA polymerase was reported to be a target for ubiquitination in UV-irradiated cells depending on CSA and CSB (Bregman *et al*., 1996). Cul4A-DDB1-CSA would then be a logical choice for the responsible E3 ligase. Later, however, it was reported that ubiquitinated RNA polymerase was only present in the soluble fraction, but not in the chromatin fraction and it was noted that RNA polymerase does not need to dissociate for repair to take place (Fousteri *et al*., 2006). It was then proposed that CSA perhaps protects RNA polymerase from degradation and that ubiquitination is only an emergency strategy in case the cell can't cope with the damage (Fousteri *et al*., 2006), though this has not been proven conclusively.

CSB has also been proposed as a potential substrate of CSA, linking the two main TC-NER factors together. However, the results on this matter are controversial. CSB-CSA interaction has been found with *in vitro* translated proteins and in a yeast two-hybrid assay (Henning *et al*., 1995). Moreover, translocation of CSA to the nuclear matrix upon UV-radiation was found to be dependent on the presence of CSB (Kamiuchi *et al*., 2002). No interaction was found, though, in gel filtration experiments of whole-cell extracts (Van Gool *et al*., 1997) nor in co-immunoprecipitation and immunofluorescence experiments (Bradsher *et al*., 2002). An argument in favor for CSB being the substrate of CSA is that the CSA ubiquitin ligase complex can ubiquitinate CSB *in vitro* (Groisman *et al*., 2006). Also, CSB

has been reported to be ubiquitinated and degraded in a CSA-dependent matter several hours after UV-irradiation (Groisman *et al*., 2006). However, this doesn't fit with the observation that UV-lesions are removed by TC-NER at an approximately linear rate during at least 30 hours (Van Hoffen *et al*., 1995) and other groups were unable to obtain the same result (Mullenders, personal communication). It is possible that use of different antibodies by different groups leads to biased protein detection after post-translational modifications of CSB in the process of TC-NER. In short, the results concerning the substrate(s) for CSA are contradictory in the literature and hence the substrate(s) for CSA have not been established conclusively.

Another possible way of linking CSA and CSB together is via the UBD of CSB: perhaps this domain recognizes ubiquitin on either CSB itself or on another, yet unidentified, protein ubiquitinated by the CSA complex, after which repair can be initiated, and hence CSA and CSB might be coupled in this way. This hypothesis is supported by the observation that the yeast homologue of CSB (Rad26) does not have a UBD and yeast also does not have the requirement for a homologue of CSA in TC-NER, since the most likely CSA homologue (Rad28) is not required for TC-NER in yeast (Bhatia *et al*., 1996; Anindya *et al*., 2010).

1.7.3 Cockayne Syndrome

Mutations in the genes for either CSA or CSB (ERCC8 and ERCC6 respectively) lead to the severe human disorder Cockayne Syndrome (CS). This rare disease, named after the London physician Edward Alfred Cockayne (1880-1956), is characterized by UV sensitivity, premature aging, growth failure and progressive neurodevelopmental abnormality. The mean age of death is around 12 years (Nance & Berry, 1992). The cellular phenotype of CS is failure of recovery of transcription after UV-exposure and increased sensitivity to oxidative stress (Mayne & Lehmann, 1982). The severity of the disease ranges widely amongst patients. In the classical CS type I the symptoms begin to appear in the first two years of life and the disease progresses until death in the first or second decade of life. In CS type II the onset of the disease is earlier, shortly after birth, and death occurs around seven years, while CS type III has a much milder phenotype with a later onset.

Mutations in either ERCC6 and ERCC8 can cause Cockayne Syndrome, with no obvious clinical or cellular differences in patients with mutations in the gene coding for CSA or CSB (Lehmann, 2003). In 62 % of the cases, the mutated gene is the ERCC6 (Laugel *et al*., 2010). In total 37 distinct mutations have been identified in this gene. Missense mutations in this gene are partially clustered in ATPase motif III (four out of ten) and the six other missense mutations are also found in the approximate central 1/3 of the protein containing the ATPase motifs. Nonsense or frame-shift mutations lead to truncations of at least 330 amino acids (> 20 % of the protein length). Besides mutations in ERCC6, 18 distinct mutations in ERCC8 leading to CS are reported. All disease-causing point mutations ERCC8 are found in WD40 repeats. All reported truncations take out at least one WD40 repeat and can be expected to have a detrimental influence on the overall fold and hence the stability of the protein (Laugel *et al*., 2010).

It has been attempted to link the type of mutation (truncation, missense mutation) to the severity of the disease. A success in this seemed to be booked in 2004, when it was reported that the complete absence of CSB in a patient due to a homozygous null mutation causes a less severe phenotype than other mutations such as missense mutations: it was seen to cause a very mild form of CS (Hashimoto *et al*., 2008) or the much milder disease UV-sensitivity syndrome (UVSS) (Horibata *et al*., 2004). This latter disease is characterized by cutaneous photosensitivity, but no neurological impairment and a normal life expectancy. At cellular level, fibroblasts from UVSS patients show a defect in recovery of RNA synthesis after UV exposure, but there is no defect in oxidative damage repair. It was then proposed that in some patients with CS, CSB forms a block for transcription after repair has completed due to a defect in removing it (e.g. incapacity for it to be ubiquitinated), while this problem does not occur in complete absence of CSB and thus the complete absence leads to the milder disease UVSS. This hypothesis was adapted some years later, when the existence of a fusion protein was reported that consisted of the first five exons of CSB and of the PGBD3 transposon (Newman *et al*., 2008). This PiggyBac fusion protein is just as abundant in normal cells as CSB and it is conserved in primates since just prior to divergence of marmosets, therefore it is likely to have some beneficial function. It was hypothesized that perhaps this fusion protein may cause CS in absence of functional CSB. Mutations downstream of intron 6 hence should cause CS and upstream only UVSS, since in that case both CSB and the PiggyBac-protein are absent. However, later two other patients with severe CS were described that have neither CSB nor the PiggyBac fusion protein due to a homozygous deletion involving non-coding exon 1 and upstream regulatory sequences, putting this hypothesis in doubt (Laugel *et al*., 2008). Therefore, it can be concluded that up to now, no clear correlation has been found in the type of mutation and the severity of the disease. It is hence difficult to link the genetic cause of the disease to its actual pathogenesis.

1.8 UV damage endonuclease repair

Besides the pathways for the repair of UV damages described above, there is also another DNA repair pathway present in some bacteria and lower eukaryotes. It was discovered when NER deletion mutants of *S. pombe* were observed to have some residual activity for removing pyrimidine dimers (Birnboim & Nasim, 1975 and McCready *et al*., 1993). This was later attributed to an ATP-independent incision activity on CPDs and 6-4PPs that was found in cell extracts of *S. pombe*. The protein responsible was initially called *S. pombe* DNA endonuclease, but later renamed as UV damage endonuclease (UVDE; Bowman *et al*., 1994). It was found to incise the DNA directly 5′ to UV-photoproducts leaving 3′ hydroxyl and 5′ phosphoryl groups as termini (Bowman *et al*., 1994). Around the same time, a gene was discovered *N. Crassa* that could complement for UV-sensitivity in a *E.coli* repair-deficient strain (Yajima *et al*., 1995). Since the resulting protein was seen to cut CPD and 6-4PP 5′ next to damage and had sequence similarity to *S. pombe* UVDE (*Sp*UVDE; 36.6 % sequence identity), it was concluded to be a UVDE homologue. In *S. pombe*,

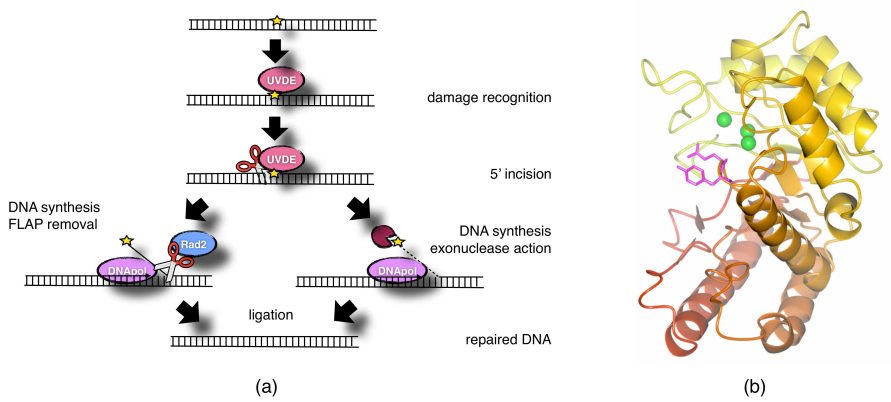

Figure 1.8: UVDE repair
(a) Mechanism of UV damage endonuclease repair based on Kisker, 2007. Details described in the text.
(b) Crystal structure of UV damage endonuclease (red to yellow) with the three metal ions shown in green and residues Gln104 and Tyr105 shown in magenta. Figure based on structure 2j6v.

double mutants in both NER and UVDE were found to be more sensitive than either single mutant, confirming that this pathway was indeed an alternative to NER (Freyer *et al.*, 1995).

The general mechanism of UVDE is now thought to be as follows. First, UVDE recognizes a damage in the DNA, presumably because of flexibility of the DNA due to the damage and/ or loss of hydrogen bonding. It then incises the phosphodiester backbone directly 5′ to the damage, leaving a 5′ phosphate and a 3′ hydroxyl group. The repair can be terminated via one of two possible pathways (see figure 1.8(a)). One option is that DNA polymerase fills in a stretch of DNA, after which the overhanging flap is processed by a FLAP-endonuclease such as rad2 and DNA ligase seals the nick. The other option is that the 5′ end is digested by an exonuclease, after which the gap is filled by DNA polymerase and the nick is closed by DNA ligase (Yoon *et al.*, 1999; Alleva *et al.*, 2000).

To date, UVDE has been found in 4 archaebacteria, 30 eubacteria and a number of lower eukaryotes such as *S. pombe* and *N. crassa*, but not in humans (Goosen & Moolenaar, 2008). The eukaryotic UVDEs are in general larger than the bacterial ones: they have an extra domain on their N-terminus, which is potentially important for interaction with other proteins. The most-studied UVDE, *Sp*UVDE, for example, is 68 kDa full-length (599 amino acids), with amino acids 250 to 527 being the most conserved region. Truncations of the N-terminus until the 232th residue were reported not to influence *Sp*UVDE's catalytic function, but deletion till the 273th residue yielded a soluble but inactive protein (Takao *et al.*, 1996), confirming the region 250-527 as the catalytic domain. Since full-length *Sp*UVDE is a very unstable protein, a truncation of the first 228 residues is mostly used in liter-

ature, which has been shown to be fully functional (Kaur *et al*., 1998) and able to complement repair deficient *E.coli* (Takao *et al*., 1996). The bacterial UVDE homologues only span this catalytic region with around 30 % sequence identity with the C-terminal half of *Sp*UVDE. At the C-terminus, there is more variation amongst eubacteria. Most eukaryotes, like *Sp*UVDE, have a stretch relatively rich in charged amino acids at their C-terminus which is indispensable for UVDE's function, since already a 35 amino acid C-terminal truncation abrogates *Sp*UVDE's catalytic function (Takao *et al*., 1996). Some bacterial UVDEs also have this region, though many do not have it and still are active (for example, *T. thermophilus* UVDE).

With purified (N-terminally truncated) *Sp*UVDE, its substrate specificity has been explored. *Sp*UVDE was found to have a remarkably broad substrate specificity. It not only cuts efficiently the UV-photoproducts CPD and 6-4PP, but also their Dewar isomers and abasic sites and, to a smaller extent, mismatches, small loops, single-strand nicks and gaps. Very small incision activity has been reported on uracil, dihydrouracil and platinum-DNA GG diadducts (Kaur *et al*., 2000; Paspaleva *et al*., 2009). In the case of the platinum-DNA GG diadducts and the small loops, the incision site was found to be several nucleotides away from the damage (Avery *et al*., 1999). The activity on abasic sites was highest if the site was flanked by a pyrimidine, suggesting that UVDE is actually optimized for incising distorted dipyrimidines (Paspaleva *et al*., 2009).

The substrate range for UVDE is quite diverse, including both lesions causing a strong distortion in the DNA (such as 6-4PP; Kim & Choi, 1995) as well of those that distort the DNA less (such as CPD), hence it cannot be only the degree of the kink that determines activity by UVDE. Loss of hydrogen bonding has also been suggested as factor for recognition, since mismatches are cleaved, but some substrates with distorted hydrogen bonding (such as xanthine opposite T or C) are not incised by UVDE (Avery *et al*., 1999). Hence it is probably a combination of factors that is important for substrate recognition.

The substrate specificity of other characterized UVDEs has been reported to be slightly less broad: *Nc*UVDE was reported to incise UV-photoproducts, Dewar isomers and abasic sites (Kanno *et al*., 1999), *B. subtilis* UVDE (*Bs*UVDE) is known to incise UV-lesions and abasic sites (Moolenaar, personal communication), while *Tth*UVDE incises only CPD and 6-4PP efficiently and it incises abasic sites less well (Paspaleva, Thomassen *et al*., 2007).

The speed of the repair by UVDE has been measured in *S. pombe* using antibodies against the lesions (Yonemasu *et al*., 1997). Removal of CPD and 6-4PP was found to be still very fast in Δrad13 cells (only UVDE, no NER), while this removal is much slower in ΔUVDE cells (no UVDE, only NER) and no repair takes place in absence of both UVDE and NER. This shows that UVDE repair is faster than NER. However, Δrad13 cells are considerably more UV-sensitive than ΔUVDE cells, so apparently NER removal of UV-lesions is more important for the organism than UVDE. A potential reason for this is the existence of TC-NER (section 1.7), which directs NER first towards the most important regions of the genome: the actively transcribed genes (Yasui *et al*., 1998).

The crystal structure of UVDE from *T. thermophilus* gives insight into the mecha-

nism of UVDE (Paspaleva, Thomassen *et al.*, 2007). It was found that *Tth*UVDE has a TIM-barrel fold, though missing the α8-helix (Figure 1.8(b)). It has a large groove with positive charges on the ends, which is proposed to be the DNA binding groove based on structural similarity to endonuclease IV (see section 1.3). At the bottom of this groove, three metal ions are found, which are postulated to be manganese based on the requirement of this metal cofactor for the enzyme's activity. Just like endonuclease IV there are two residues sticking out into the solvent in the structure of UVDE (Gln104 and Tyr105; together called the 'probing finger'), which probably are involved in flipping the damaged base out of the helix and stabilizing the kink in the DNA. Additional evidence for this flipping comes from biochemical experiments that show that the two bases opposite the damage are flipped out of the DNA helix (Paspaleva *et al.*, 2009).

The relatively broad substrate specificity is a special feature of UVDE. Usually in DNA repair, single proteins have a very narrow specificity for only one damage, which is often achieved by flipping this damaged base in a specific pocket. An example is photolyase: it only recognizes or CPD or 6-4PP (section 1.2). When there is a broad specificity, like in NER (section 1.6), there are usually many proteins involved that recognize a more global feature of the DNA that indicates that it is distorted (bendability of the DNA, ease to insert a hairpin between the two strands). In the case of NER, there are at least 30 proteins involved in eukaryotes and 6 in bacteria. UVDE, however, is only a single protein that has a relatively broad substrate specificity, which makes it an interesting model to study DNA repair. It has been proposed that UVDE first senses bendability of DNA (explaining why it incises single-strand nicks and gaps) and then it flips the damage into a pocket that is most suited for a damaged dipyrimidine (Paspaleva *et al.*, 2009). However, in absence of a co-crystal of UVDE with DNA, it is difficult to give a good explanation for the broad substrate specificity of UVDE.

1.9 Contents of this thesis

As is clear from this introduction, DNA repair is essential to maintain genome integrity. Now core proteins in excision repair have been identified, further understanding of their mode of action requires high resolution insights in their interaction with damaged DNA, which can be obtained via protein crystallography. This approach allows one to elucidate structures of macromolecules at atomic resolution. Such a detailed picture of the molecule of interest clarifies the protein's mechanism and it is therefore widely used. The examples given in this introduction form just a small excerpt of the wealth of information that has been generated in the last decades using this technique, but they are representative of the type of information crystal structures can yield.

The principal focus of this thesis is to get detailed insight into how UV-damage can be removed from DNA. To reach this goal we focused on two different repair systems involved in the removal of UV lesions from DNA, namely the human TC-NER system (section 1.7) and the bacterial UVDE system (section 1.8). In this thesis new insights into these DNA repair pathways are described, which were primarily

gained from protein crystallography. In the first part of the thesis (Chapters 2-4) the question of CSA's role in TC-NER is addressed, in particular considering the interaction of CSA with DDB1 and other putative partners and how mutations in this protein lead to the human disease Cockayne Syndrome. The second part of this thesis (Chapters 5-6) is about UVDE and addresses the question of how this enzyme can repair different UV-lesions.

- **Chapter 2** describes the first steps necessary to obtain a crystal structure: protein expression, purification and crystallization, in this case applied to the human protein CSA (see section 1.7.2).
- **Chapters 3 and 4** then continue to describe the crystal structure of CSA and insights gained from it.
- **Chapters 5 and 6** deal with the recognition and incision of damaged DNA by the enzyme UVDE (section 1.8). First, the identity and role of a post-translational modification is explored. A study addressing UVDE's mechanism of damage recognition and the basis of its broad substrate specificity follows.
- **Chapter 7** is devoted to the crystal structure of an unrelated protein, potato serine protease inhibitor, to show the more general applicability and strength of the technique of protein crystallography.