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

Meulenbroek, E.M.

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

Elisabeth M. Meulenbroek

Printed by: Gildeprint Drukkerijen, the Netherlands.
Cover Design: E.M.Meulenbroek
ISBN: 978-94-6108-339-5

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

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. P. F. van der Heijden,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 9 oktober 2012
klokke 16.15 uur

door

Elisabeth M. Meulenbroek

geboren te Middelburg, Nederland
in 1985

PROMOTIECOMMISSIE

Promotoren:	Prof. dr. J. P. Abrahams	Universiteit Leiden
	Prof. dr. L.H.F. Mullenders	Leids Universitair Medisch Centrum
Copromotor:	Dr. N.S. Pannu	Universiteit Leiden
Leden:	Prof. dr. J. Brouwer	Universiteit Leiden
	Dr. N. Goosen	Universiteit Leiden
	Prof. dr. J. H. J. Hoeijmakers	Erasmus Medisch Centrum
	Prof. dr. T. K. Sixma	Nederlands Kanker Instituut, Erasmus Medisch Centrum

The work reported in this thesis is financed by the Netherlands Organization for Scientific Research (NWO) through a Toptalent grant (N° 021.002.024).

The research was carried out at the 'Leiden Institute of Chemistry' (LIC).

An electronic version of this dissertation is available at the Leiden University Repository (https://openaccess.leidenuniv.nl).

aan mijn ouders
Bernard en Nicolette

Contents

1 Introduction **1**
- 1.1 DNA under attack . . . 1
- 1.2 Direct reversal . . . 2
- 1.3 Base excision repair . . . 4
- 1.4 Mismatch repair . . . 5
- 1.5 Double-strand break repair . . . 7
- 1.6 Nucleotide excision repair . . . 8
- 1.7 Transcription-coupled repair . . . 10
  - 1.7.1 Discovery and mechanism of transcription-coupled repair . . . 10
  - 1.7.2 Cockayne syndrome proteins A and B . . . 11
  - 1.7.3 Cockayne Syndrome . . . 16
- 1.8 UV damage endonuclease repair . . . 17
- 1.9 Contents of this thesis . . . 20

2 Purification and crystallization of Cockayne Syndrome protein A **23**
- 2.1 Introduction . . . 24
- 2.2 Materials and Methods . . . 24
  - 2.2.1 Cloning and overproduction . . . 24
  - 2.2.2 Purification . . . 25
  - 2.2.3 Crystallization . . . 26
  - 2.2.4 X-ray diffraction analysis . . . 26
- 2.3 Results and discussion . . . 26

3 Crystal structure of Cockayne syndrome protein A **31**
- 3.1 Introduction . . . 32
- 3.2 Materials and methods . . . 32
  - 3.2.1 Protein overproduction, purification and crystallization . . . 32
  - 3.2.2 Data collection and processing . . . 33
  - 3.2.3 Structure solution and refinement . . . 33
  - 3.2.4 Superpositions . . . 34
  - 3.2.5 Complementation assays . . . 35
  - 3.2.6 Protein solubility assays . . . 35
- 3.3 Results . . . 35

3.3.1 Overall structure . . . 35
3.3.2 Interaction of CSA and DDB1 . . . 36
3.3.3 Substrate binding site of CSA . . . 38
3.3.4 Disease-causing mutations . . . 41
3.4 Discussion . . . 45

4 Substrate-binding by Cockayne Syndrome protein A 47
4.1 Introduction . . . 48
4.2 Material and methods . . . 48
4.2.1 Protein overproduction and purification . . . 48
4.2.2 Bandshift assay . . . 49
4.2.3 Construction of CSA mutants . . . 49
4.2.4 CSA interaction studies . . . 49
4.3 Results . . . 50
4.3.1 Verification of CSA's substrate binding site . . . 50
4.3.2 Determination of CSA's binding capacity to DNA . . . 50
4.3.3 CSB is a possible substrate for CSA . . . 53
4.4 Discussion . . . 54

5 Involvement of a carboxylated lysine in UV damage endonuclease 55
5.1 Introduction . . . 56
5.2 Materials and methods . . . 56
5.2.1 Proteins . . . 56
5.2.2 DNA substrates . . . 57
5.2.3 Incision assay . . . 57
5.2.4 Incision of supercoiled plasmid DNA . . . 57
5.2.5 Filter binding assay . . . 57
5.2.6 Mass-spectrometry . . . 58
5.2.7 Crystallization . . . 58
5.2.8 Data collection and processing . . . 58
5.2.9 Structure solution and refinement . . . 58
5.3 Results . . . 59
5.3.1 Identity of the modification . . . 59
5.3.2 Functional studies of Lys229 mutants . . . 62
5.3.3 Structural studies of the Lys229 mutants . . . 64
5.4 Discussion . . . 66

6 Unravelling UVDE's uncanny ability to recognize and incise different types of damaged DNA 71
6.1 Introduction . . . 72
6.2 Materials and methods . . . 72
6.2.1 Cloning . . . 72
6.2.2 Expression and purification . . . 73
6.2.3 DNA substrates . . . 73
6.2.4 Incision assays . . . 74

6.2.5 Bandshift assays . . . 74
6.2.6 Filter-binding assays . . . 74
6.2.7 Crystallization . . . 74
6.2.8 Data collection . . . 75
6.2.9 Structure solution and refinement . . . 75
6.3 Results . . . 76
6.3.1 Overall structure of *Sac*UVDE with and without damaged DNA 76
6.3.2 Recognition of pyrimidine-dimer lesions by UVDE . . . 79
6.3.3 *Sac*UVDE has a preference for 6-4 photoproduct over CPD damaged DNA . . . 81
6.3.4 Structure-based mutagenesis explains *Sac*UVDE's substrate specificity . . . 83
6.4 Discussion . . . 87

7 Crystal structure of a heterodimeric, double-headed Kunitz-type serine protease inhibitor from potato. 91
7.1 Introduction . . . 92
7.2 Methods . . . 92
7.3 Results and Discussion . . . 93
7.3.1 Overall structure . . . 93
7.3.2 Identification of reactive site loops . . . 97
7.3.3 Interfaces for aggregation . . . 99

Bibliography 101

List of publications 115

Summary 117

Samenvatting 121

Curriculum vitae 125

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.

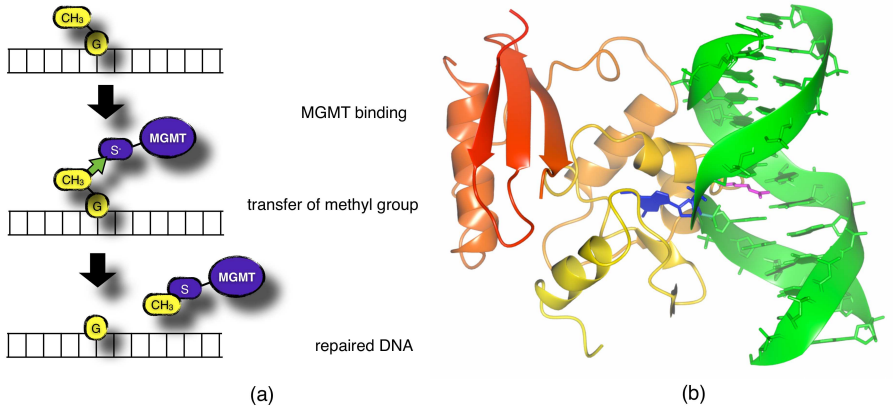

Figure 1.1: Direct reversal
(a) Mechanism of direct reversal as performed by the O^6-meG-DNA methyl transferase (MGMT), based on Eker et al., 2009. See text for detailed explanation.
(b) Crystal structure of MGMT (red-yellow) bound to damaged DNA (green) showing nucleotide flipping with the damaged base 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

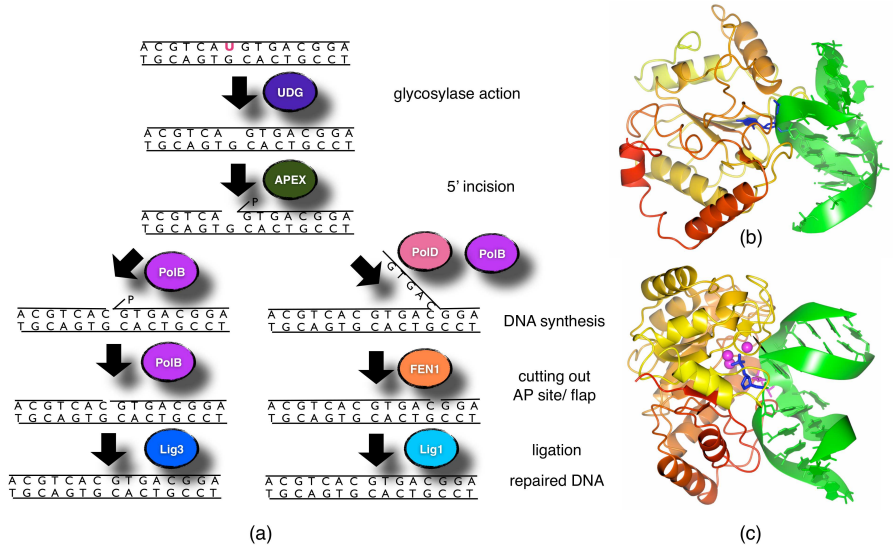

Figure 1.2: Base excision repair
(a) Mechanism of base excision repair based on Robertson et al., 2009 showing UDG as glycosylase and APEX as AP endonuclease. See text for detailed explanation.
(b) Crystal structure of uracil-DNA glycosylase (UDG; red to yellow) bound to DNA (green) showing uracil flipping 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

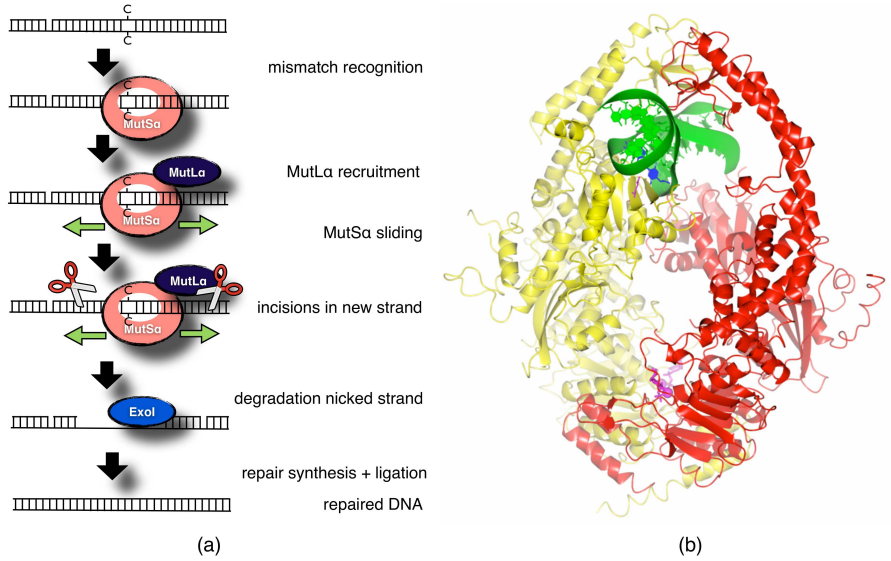

Figure 1.3: Mismatch repair
(a) Mechanism of mismatch repair based on Kunz et al., 2009. Detailed explanation can be found in the text.
(b) Crystal structure of MutSα in complex with DNA (green) with MSH2 in red and MSH6 in yellow, the conserved Glu434 in magenta, ADP in magenta, DNA in green and the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

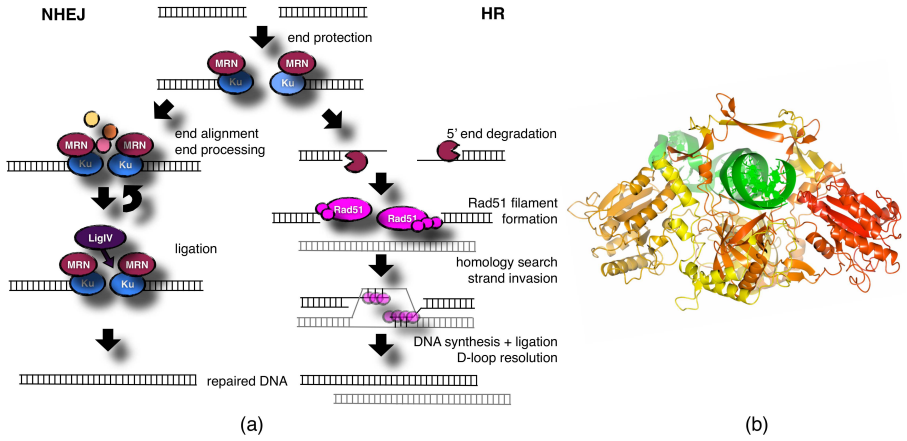

Figure 1.4: Double-strand break repair
(a) Mechanism of double strand break repair based on Pardo et al., 2009. Details can be found in the text.
(b) Crystal structure of the Ku heterodimer (in red to yellow) bound to DNA (green). Structure 1jey was used to make this figure.

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.

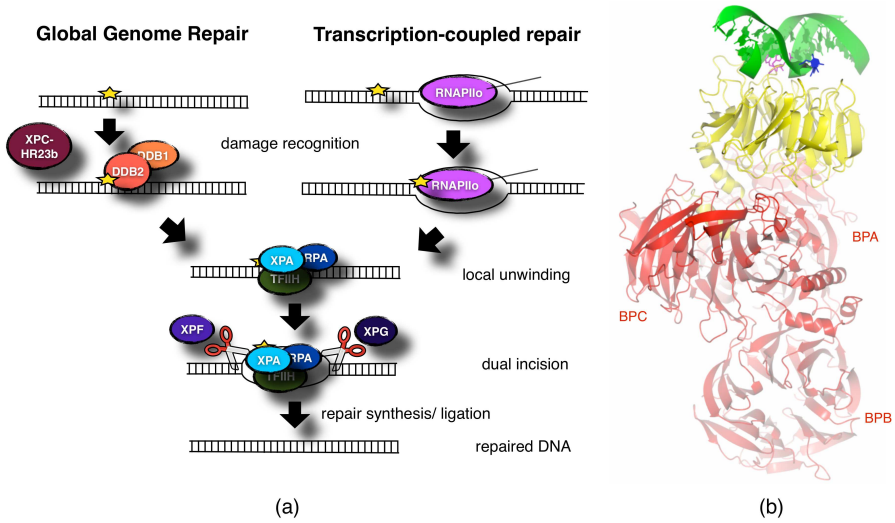

Figure 1.5: Nucleotide excision repair
(a) Mechanism of nucleotide excision repair, showing both the global genome repair branch and the transcription-coupled repair branch based on Nouspikel et al., 2009b (see text for details). (b) Crystal structure of DDB1 (red; the domains are indicated as BPA, BPB 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

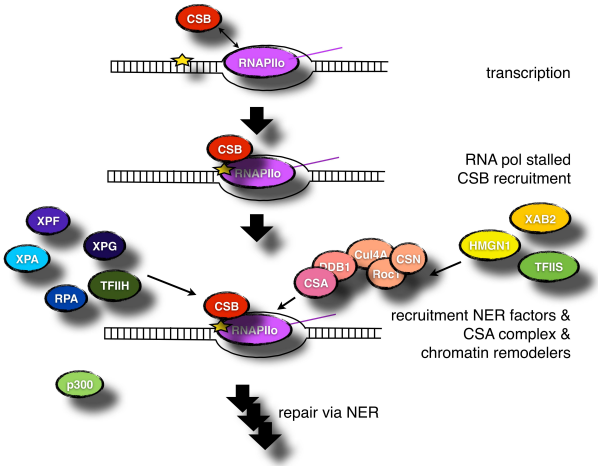

Figure 1.6: Transcription-coupled repair
Mechanism of transcription-coupled repair showing the roles of CSA and CSB (figure based on Fousteri et al., 2008). Details can be found in the text.

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

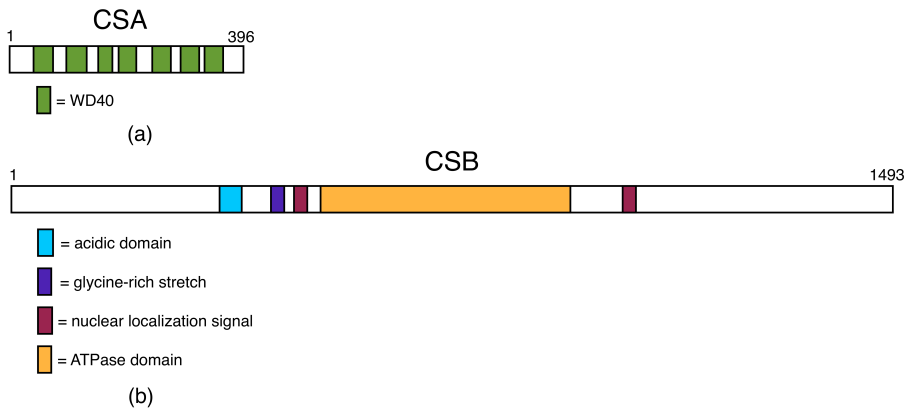

Figure 1.7: Sequence of CSA and CSB.
Schematic representation of the amino acid sequence of CSA (a) and CSB (b), showing the most important domains.

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*,

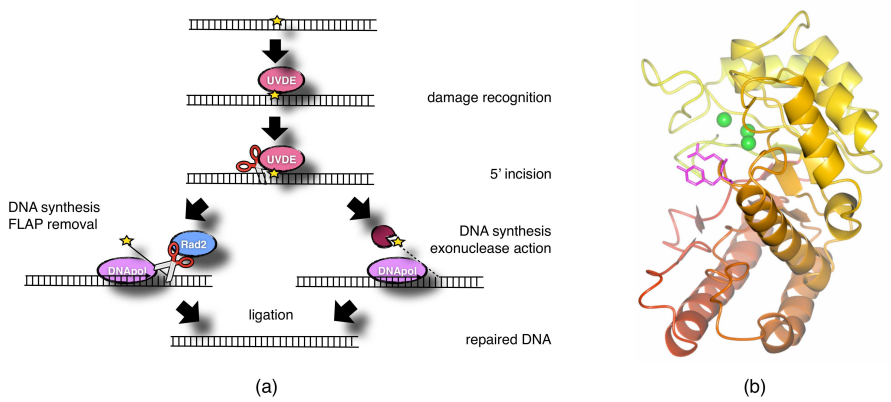

Figure 1.8: UVDE repair
(a) Mechanism of UV damage endonuclease repair based on Kisker, 2007. Details described in the text.
(b) Crystal structure of UV damage endonuclease (red to yellow) with the three metal ions shown in green and residues Gln104 and Tyr105 shown in magenta. Figure based on structure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

2

Purification and crystallization of Cockayne Syndrome protein A

Cockayne Syndrome protein A is one of the main components in mammalian Transcription-Coupled Repair. Here the overproduction, purification, and crystallization of human Cockayne Syndrome protein A in complex with its interacting partner DNA Damage Binding protein 1 are reported. The complex was coproduced in insect cells, copurified and crystallized using sitting drops with PEG3350 and sodium citrate as crystallizing agents. The crystals have unit cell parameters a=b=142.03 Å, c=250.19 Å and diffract to 2.9 Å on beamline ID14-1 at the European Synchrotron Radiation Facility.

published in: E.M. Meulenbroek, N.S. Pannu, *Overproduction, purification, crystallization and preliminary X-ray diffraction analysis of Cockayne syndrome protein A in complex with DNA damage-binding protein 1.*, Acta Cryst. **F68**, 45-48 (2012)

2.1 Introduction

The human genome is constantly subject to damaging agents such as ultra-violet irradiation. One of the major pathways to remove DNA lesions is Nucleotide Excision Repair (NER). It is responsible for removing a wide variety of chemically and structurally diverse lesions. A subpathway of NER is Transcription-Coupled NER (TC-NER), which removes lesions blocking transcription. Most proteins are shared between Global Genome NER (GG-NER) and TC-NER, but two proteins are unique to TC-NER: Cockayne Syndrome Protein A and B (CSA and CSB). As can be read in Chapter 1, these proteins were proposed to have the following functions (Fousteri *et al*., 2006): CSB attracts NER factors and chromatin remodelers to the site of the lesion and it uses its DNA-dependent ATPase activity to remodel the RNA polymerase - DNA interface to prevent RNA polymerase hiding the damaged site from the NER factors. CSA is found in a CSA-DDB1 E3-ubiquitin-ligase/CSN complex, which is inactive for E3-ubiquitin-ligase activity upon recruitment, and might prevent degradation of the stalled RNA polymerase and/ or other TCR factors in early stages of the repair or it might be involved in degradation of CSB at a later stage of the repair process. The degradation of CSB could be important for recovery of transcription after TCR is finished (Groisman *et al*., 2006).

The biological importance of CSA and CSB can be seen since mutations in either protein can cause the recessive human disorder Cockayne Syndrome. This disease, named after the London physician Edward Alfred Cockayne (1880-1956), is characterized by neurologic abnormality, growth retardation, abnormal sensitivity to sunlight and premature aging (Nance & Berry, 1992). Eighteen different mutations in CSA that can cause Cockayne Syndrome (Laugel *et al*., 2010) have been reported.

Cockayne Syndrome protein A is found in the cell in complex with DDB1, which is a multifunctional protein that links several different substrate adaptors like CSA to the Cul4A-Roc1 complex hence constituting the E3-ubiquitin ligase complex (recently reviewed in Iovine *et al*., 2011). Several crystal structures of DDB1 have been reported, the first of which was reported in Li *et al*., 2006. However, to date no structural information is available on CSA.

Determination of the structure of Cockayne Syndrome protein A will provide insight into the important DNA repair mechanism TC-NER and into the molecular basis by which the disease-causing mutations in CSA cause Cockayne Syndrome. To this end, we report the overproduction, purification, and crystallization of CSA in complex with DDB1.

2.2 Materials and Methods

2.2.1 Cloning and overproduction

The open reading frame (ORF; amino acids 1-396) for human CSA (ERCC8; OMIM 609412) was amplified by PCR from a clone in a pFastbac vector kindly made available by Wim Vermeulen (Erasmus Medical Center, The Netherlands) and cloned into the pETM-series of vectors (Dümmler *et al*., 2005) and into pET52b (Invitro-

gen). Overproduction in *E.coli* strains BL21 Rosetta, RIL, RP, pLysS and pLysS star was attempted at 277, 293 and 310 K with 0.1 – 1 mM IPTG.

Overproduction in *E.coli* was concluded to be unsuccessful, after which overproduction of CSA together with its interaction partner DDB1 was attempted in Sf9 insect cells. For this, a vector with the ORF (amino acids 1-1140) for human DDB1 (OMIM 600045) was kindly made available by Andrea Scrima and Nicolas Thomä (Friedrich Miescher Institute, Switzerland). The sequence of a N-terminal 6x-His-tag and a thrombin cleavage site were inserted in front of the ORF and together it was placed behind the p10 promoter in a pFastbac Dual vector (Invitrogen) using *Xma*I and *Acc*65I. The resulting DDB1 molecule had the following residues extra at its N-terminus: MHHHHHHRRLVPRGSGGR.

CSA was amplified from the pET52b construct with a C-terminal 10x His-tag and thrombin cleavage site using the following primers: 5'-TTT CAC GGT CCG GGG ATG CTG GGG TTT TTG TCC GCA CG-3' and 5'-AGT AGT CGA CGT TAA TTA GTG GTG GTG ATG GTG ATG ATG GTG-3' and cloned into the same pFastbac Dual vector as DDB1 behind the PolH promoter using the restriction enzymes *Rsr*II and *Sal*I. The resulting protein misses its C-terminal glycine and has the following residues extra at its C-terminus: LALVPRGSSAHHHHHHHHHH. The construct was verified by sequencing (Baseclear, Netherlands).

The pFastbac Dual with CSA and DDB1 was transformed into the strain DH10EMBacY (Berger *et al*., 2004), kindly provided by Imre Berger (EMBL, France). Recombinant bacmid was isolated and transfected to *Sf9* insect cells using Fugene HD (Roche) following the manufacturer's instructions. Recombinant baculovirus was produced and insect cells were infected at a density of 1.5-2 x 10^6 cells/ml using 2 ml 4^{th} generation virus per 100 ml cell culture in suspension. The cells were harvested 66 hours post-infection.

2.2.2 Purification

All purification steps were executed at 277 K. Harvested cells were resuspended in lysis buffer containing 50 mM Tris pH 8, 20 mM imidazole, 200 mM NaCl, 0,1 % Triton-X-100, 5 mM β-mercaptoethanol and Complete mini EDTA-free protease inhibitor cocktail (Roche). Cells were lysed by sonification and the lysate was centrifuged in an ultracentrifuge at 30.000 rpm (61.700 g) for 30 min at 277 K in a Beckman Coulter 70.1 Ti preparative rotor. The supernatant was loaded on a His-Trap HP column (GE healthcare) using an Äkta Express (GE healthcare) and washed with 20 column volumes of Ni buffer A containing 20 mM Tris pH 8, 30 mM imidazole, 200 mM NaCl and 5 mM β-mercaptoethanol. The CSA-DBB1 complex was eluted with a gradient of 24 column volumes to 30 % Ni buffer B and then 28 column volumes to 100 % Ni buffer B containing 20 mM Tris pH 8, 330 mM imidazole, 200 mM NaCl and 5 mM β-mercaptoethanol.

The Ni column fractions containing the CSA-DBB1 complex were diluted three times with 20 mM HEPES pH 7.2 and loaded on a Hitrap Q HP column (GE healthcare). The column was washed with Q buffer A containing 20 mM HEPES pH 7.2, 100 mM NaCl and 5 mM β-mercaptoethanol. The protein complex was eluted by

a gradient of 25 column volumes to 100 % Q column B containing 20 mM HEPES pH 7.2, 1 M NaCl and 5 mM β-mercaptoethanol. Next, the CSA-DDB1 containing fractions were loaded on a Superdex 200 gel filtration column (GE healthcare) equilibrated with 20 mM HEPES pH 7.2, 200 mM NaCl and 5 mM DTT and the column was run with a flow rate of 0.3 ml/min. Protein purity was assessed with 10 % SDS PAGE stained with Coomassie Blue (Simply Blue Safe stain, Invitrogen) and by silver staining (Silver Stain Plus, Biorad).

2.2.3 Crystallization

CSA-DDB1 was concentrated with a 10 kDa MWCO centrifugal filter unit (Millipore) to a concentration of 5-8 mg/ml. Crystallization conditions were screened using sitting-drop vapour-diffusion using the commercial screens JCSG+ and PACT (Qiagen) at 293 K with a drop size of 0.8 μl. A Genesis RS200 (Tecan) was used for pipetting the reservoir solution (75 μl) and an Oryx6 (Douglas Instruments) was used for pipetting the drops. After two days, small needle-like crystals appeared in conditions 71, 83, and 95 from PACT, which contain 0.2 M sodium citrate, 20% (w/v) PEG 3350 and 0.1 M Bis Tris propane 6.5, 7.5 or 8.5. The crystals were verified to be protein using a fluorescence microscope with filter U-MWU2 (Olympus).

These crystals were optimized at 293 K and the best condition was found to be 0.2 M sodium citrate, 24% PEG 3350, 0.1 M Bis-Tris propane pH 8.0. Based on this condition, an additive screen (Hampton Research) was performed following the manufacturer's instructions. 3% glycerol was found to cause a significant improvement in size and morphology of the crystals. After additional optimization, addition of 5-7% glycerol turned out to be optimal and drops of size between 2 and 4 μl using a protein solution to crystallant solution volume ratio of 2:1 or 3:1.

2.2.4 X-ray diffraction analysis

Crystals were caught in cryoloops and soaked in a solution containing mother liquor and 10–12 % glycerol before flash-cooling them in liquid nitrogen. X-ray diffraction experiments were done on ID14-1 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. 180 images were collected with an oscillation angle of 1.0° and the exposure time was 13 s per frame on ID14-1 at 0.9334 Å at 100 K. Images were processed with *iMosflm* (Leslie, 1999). Scaling and merging was done with *SCALA* from the *CCP4* suite (Winn *et al*., 2011).

2.3 Results and discussion

Overproduction of CSA in *E.coli* was attempted with several tags, but only an MBP-fusion protein was soluble. Purification on amylose resin (New England Biolabs) gave protein of good purity. Cleavage of the MBP-tag with TEV protease, however, led to precipitation of CSA. Gel filtration on the MBP-fusion protein showed that the fusion protein was present as a high-molecular weight entity, presumably a soluble aggregate.

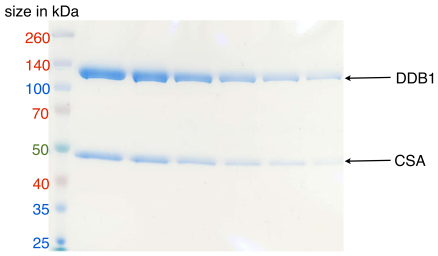

Figure 2.1: Coomassie-stained SDS PAGE gel showing CSA and DDB1 at the end of the purification procedure.

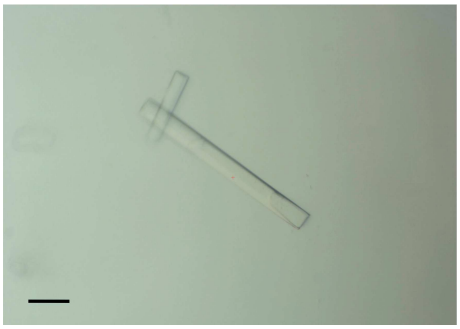

Figure 2.2: Crystals of CSA in complex with DDB1. The scale bar is 100 μm in length.

Coproduction of CSA and its interacting partner DDB1 was then attempted in insect cells, because the structurally and functionally related complex DDB1-DDB2 has been coproduced and purified from insect cells (Scrima *et al.*, 2008). Coproduction of CSA and DDB1 in *Sf9* cells yielded a soluble protein complex and a large excess of DDB1 (more than three times excess). The large excess of DDB1 could be removed by using a customized gradient on the Ni column, since DDB1 alone eluted in the first part of the gradient (0-30 % B), while the CSA-DDB1 complex eluted in the second part of the gradient (30-100 % B). CSA copurified with DDB1 as a 1:1 complex as judged from SDS PAGE (Figure 2.1) throughout the entire purification and no free CSA was obtained. The complex eluted as a single peak from the gel filtration around 170 kDa. This indeed approximates the size of a 1:1 complex of CSA with DDB1, whose theoretical sizes are 46 and 129 kDa in our constructs. Western blotting with antibodies against DDB1 and CSA and mass-spectrometry on proteolytic fragments confirmed the identity of both proteins.

Crystallization conditions were found in a pH anion cation screen (PACT) and optimized by a grid screen around the condition varying PEG concentration and pH. The initial crystals were around 0.09 by 0.01 by 0.01 mm and diffracted to around 8 Å at the ESRF. An additive screen was performed and 3 % glycerol was found to improve crystal size and quality significantly. The optimal glycerol con-

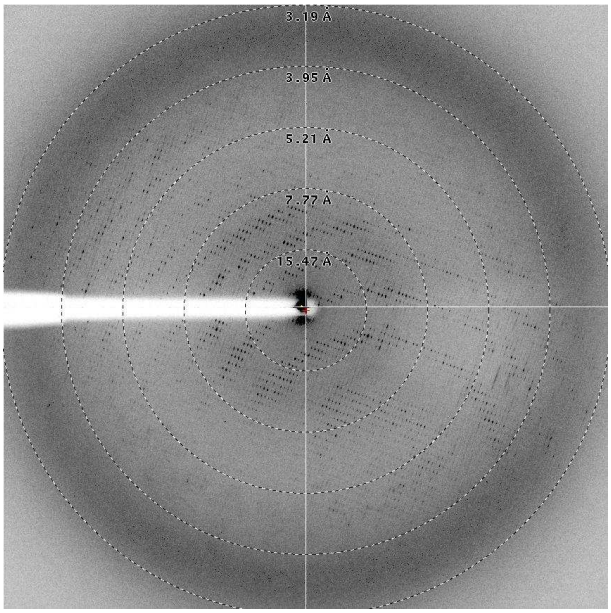

Figure 2.3: A diffraction image of a crystal of the CSA-DDB1 complex, in which the anisotropic nature of the diffraction can be seen.

Table 2.1: Data collection statistics.

Wavelength (Å)	0.9334
Space group	$P3_1$
Resolution (Å)	2.92 (3.08-2.92)[a]
No. measured reflections	610127 (49472)
No. unique reflections	119918 (15498)
Completeness (%)	97.8 (86.4)
Multiplicity	5.1 (3.2)
Mean I/σ(I)	4.9 (1.1)
R_{merge}	0.369 (1.108)
R_{pim}	0.181 (0.708)

[a] Values in parentheses are for the highest resolution shell

centration was found to be 5–7 % and crystals grown in this condition grew to 0.5 by 0.07 by 0.07 mm and diffracted to around 3 Å (Figure 2.2). Silver stained gels of washed crystals confirmed that the crystals contained both DDB1 and CSA.

Processing of the diffraction data showed that the crystals belonged to space group $P6_222$ or $P6_422$ according to Pointless (Evans, 2006) with unit cell parameters a=b=142.03 Å, c= 250.19 Å. However, the L and H twinning tests output by CTRUNCATE (Padilla & Yeates, 2003) indicated that the crystal was twinned. Many crystals were tested and all were perfectly twinned. The correct space group will be determined through molecular replacement and refinement by testing all possible subgroups and enantiomorphs of $P3_1$ (see Table 2.1 for data collection statistics for the crystal in the lowest possible subgroup $P3_1$).

The diffraction was anisotropic (Figure 2.3) and the anisotropic diffraction server (Strong *et al*., 2006) indeed showed that F/σ>3 at 2.9 Å in the direction of c*, but 3.3 Å in a* and b*. We chose a diffraction cutoff of 2.92 Å for our data, because data in this region is still useful though its I/σ is low owing to the anisotropy. It should be noted that the data can be better than the I/σ suggests, because the possible non-crystallographic symmetry (NCS) increases the signal to noise ratio. For example, the I/σ of the data merged in $P6_422$ is 1.9 in the highest resolution shell of 3.08-2.92 Å. The calculated Matthews coefficient is 2.08 $Å^3Da^{-1}$ assuming four DDB1-CSA complexes in the ASU.

3

Crystal structure of Cockayne syndrome protein A

Cockayne syndrome protein A (CSA) is a key component of Transcription-Coupled Nucleotide Excision Repair (TC-NER), a DNA repair pathway that specifically removes transcription-blocking DNA lesions, and a part of an E3-ubiquitin ligase complex. Mutations in CSA lead to the human recessive disorder Cockayne Syndrome. Here we report the crystal structure of CSA in complex with its interacting partner, DNA Damage Binding protein 1 (DDB1). The overall structure of the WD40-repeat domain protein CSA is a seven-bladed β-propeller. Like the related complex of DDB1 with DNA damage binding protein 2 (DDB2), CSA interacts with DDB1 via a helix-loop-helix motif that inserts between the BPA and BPC domain of DDB1. The top of the WD40 domain of CSA contains a positively charged centre, which we propose as a substrate-binding site. The structure also shows the location of the mutations in CSA that cause Cockayne Syndrome and provides explanations for their detrimental effect, namely disrupted local or global folding that probably interferes with proper substrate-binding for most of the mutations.

E.M. Meulenbroek, M.G. Vrouwe, L.H.F. Mullenders, N.S. Pannu, *Insights into Transcription-Coupled Repair from the crystal structure of Cockayne Syndrome protein A.* (to be submitted)

3.1 Introduction

The human genome is constantly subject to damaging agents, both endogenous factors such as reactive oxygen species from cellular metabolism and environmental factors such as UV-irradiation. To prevent cell death or mutagenesis, several DNA repair pathways have evolved. A major DNA repair pathway is Nucleotide Excision Repair (NER). This highly conserved pathway removes a wide variety of helix-distorting lesions by dual incision and by subsequent excision of a stretch of DNA containing the damage. The remaining gap is filled by DNA repair patch synthesis using the undamaged strand as a template (Nouspikel, 2009).

As described in Chapter 1, the broad substrate recognition of NER is achieved by two distinct sub-pathways of NER: Global Genome NER (GG-NER) and Transcription-Coupled NER (TC-NER), that are triggered by DNA damage mediated structural alterations in the genome. TC-NER specifically repairs DNA lesions that block transcription by elongating RNA polymerase II resulting in a faster removal of some lesions compared to those in the overall genome (Bohr *et al.*, 1985 and Mellon *et al.*, 1987). TC-NER has many factors in common with GG-NER, but two factors are unique for TC-NER: Cockayne syndrome protein A (CSA) and B (CSB) (Fousteri & Mullenders, 2008). CSB is required for recruitment of the core NER factors, chromatin remodelers and CSA to the site of the lesion. CSA resides in a complex with DDB1, Cul4A and Roc1 forming an E3 ubiquitin ligase complex that is inactive for ubiquitin ligase activity upon recruitment. CSA is a WD40 repeat protein that presumably functions as the substrate-recruiting subunit of the E3 ubiquitin ligase complex.

Mutations in CSA or CSB lead to Cockayne syndrome, a recessive human disorder that is characterized by poor growth and neurologic abnormalities (Nance & Berry, 1992). To date, 18 different mutations have been reported in CSA that lead to Cockayne syndrome (Laugel *et al.*, 2009). How these mutations lead to the disease remains elusive. Here, we report the crystal structure of human CSA in complex with its interacting partner human DDB1 solved at 2.9 Å.

3.2 Materials and methods

3.2.1 Protein overproduction, purification and crystallization

We expressed, purified and crystallized the CSA-DDB1 complex as described in chapter 2 of this thesis. Briefly, CSA with a C-terminal 10x His-tag and DDB1 with a N-terminal 6x His-tag were co-expressed in *Sf9* insect cells and purified as a complex with His-trap affinity chromatography, Q-column ion exchange and Superdex200 size exclusion and concentrated to 5-8 mg/ml with a 10kDa MWCO centrifugal filter unit (Millipore) in 20mM HEPES pH 7.2, 200 mM NaCl and 5 mM DTT. The protein was crystallized in 0.1 M Bis-Tris propane pH 8, 0.2 M sodium citrate and 24 % PEG3350 with 5-7 % glycerol.

3.2.2 Data collection and processing

Data was collected on beam-line ID14-1 at the European Synchrotron Radiation Facility at a wavelength of 0.933 Å. Reflections were indexed and integrated with *iMosflm* (Leslie, 1999). Scaling and merging was performed with *SCALA* (Evans, 2006) from the *CCP4* suite (Winn *et al*., 2011). Visual inspection of the images indicated anisotropic diffraction. This observation was confirmed with the anisotropic diffraction server (Strong *et al*., 2006) that showed $F/\sigma > 3$ at 2.9 Å in the direction of c*, but 4 Å in a* and b*. Thus, in one direction of reciprocal space, useful diffraction information was available to 2.9 Å and applying an anisotropic correction in the refinement procedure to down-weight the weaker reflections in the a* and b* directions allowed us to use all data to 2.9 Å. The data were processed with an apparent space group of either $P6_422$ or $P6_222$. However, the L and H twinning tests output by *CTRUNCATE* (Padilla & Yeates, 2003) indicated that the crystal was nearly perfectly twinned and that the true space group should be a subgroup that would allow twinning.

3.2.3 Structure solution and refinement

A combination of twinning, anisotropy and a large solvent content made molecular replacement challenging. Initially, the structure of DDB1 was solved by molecular replacement using the structure of DDB1 in complex with a H-box motif peptide (Li *et al*., 2010) as a starting model in *MOLREP* (Vagin & Teplyakov, 2000) from *CCP4*. A clear solution to the rotation and translation function was obtained in space group $P6_422$. The structure of CSA was solved by molecular replacement using Rack1 as a starting model (Rabl *et al*., 2011) in *MOLREP* and fixing the model of DDB1. This CSA-DDB1 solution was refined in *REFMAC* in $P6_422$ to an R-free of 40.4%.

To determine the space group, the starting model was refined with the diffraction data reprocessed in space groups $P6_4$, $P3_121$, $P3_112$ and $P3_1$. The refinement was performed with strict non-crystallographic symmetry for spacegroups $P3_121$, $P3_112$ and $P3_1$ to insure that the observation to parameter ratio would be the same for all space groups. Significantly better R-frees were obtained with the data processed in $P3_1$, indicating that the crystal might consist of 4 twin domains and 4 non-crystallographically related DDB1-CSA complexes.

The complex structure of DDB1 and CSA was rebuilt with *Buccaneer* (Cowtan, 2006) and manual adjustments to the model were done with *COOT* (Emsley & Cowtan, 2004). Refinement without modeling twinning resulted in an R-free of 34.7% in *REFMAC* (Murshudov *et al*., 1997) using constrained non-crystallographic symmetry (NCS) refinement. After the completion of model building, twinning was modeled in *REFMAC* with constrained NCS leading to a final R-factor of 27.1% and an R-free of 28.2%. This model showed a strong interaction of CSA with the BPB domain of DDB1 (ΔG = -15.8 kcal/mol according to *PISA*, Krissinel & Henrick, 2007).

However, comparison with the structure of Nicolas Thomä (Fischer *et al*., 2011) that came out during preparation of our manuscript, revealed that in our model

Table 3.1: Data collection and refinement statistics.

Data collection	
Space group	$P3_221$
Cell dimensions	
a,b,c (Å)	141.89, 141.89, 250.10
α, β, γ (°)	90.00, 90.00, 120.00
Resolution (Å)	2.92-62.53 (2.92-3.08)[a]
R_{merge}	0.391 (1.219)
R_{pim}	0.131 (0.524)
I/σI	6.9 (1.4)
Completeness (%)	98.6 (90.5)
Redundancy	9.7 (6.0)
Refinement	
Resolution (Å)	2.92 - 62.53
No. reflections	59838
R_{work}/ R_{free}	0.237/ 0.257
No. atoms	
Protein	11414
Ligand/ ion	0
Water	0
B-factors	
Protein	60.1
Ligand/ ion	N/A
Water	N/A
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.16

[a] Values in parentheses are for the highest resolution shell

the DDB1 and CSA models from different twins had been modeled into one model due to an unfortunate combination of twinning and crystallographic symmetry. The structure was then re-refined using a CSA-DDB1 model with CSA binding to the BPA and BPC domains as input (kindly provided by Nicolas Thomä). This final model had an R-factor of 23.7 and an R-free of 25.7 % (statistics shown in Table 3.1) and was virtually identical to our initial model except for the CSA-DDB1 interface.

3.2.4 Superpositions

Most of the superpositions were done with Theseus (Theobald & Wuttke, 2006): superposition of DDB1 with DDB1s from other structures for calculating the rms deviation of the Cα atoms, and of CSA with other WD40 proteins for calculating the rms deviation of the Cα atoms and for making the model of CSA with substrate

peptides. Other superpositions were done with the ssm function in *COOT* (Emsley & Cowtan, 2004). For modeling of the E2 enzyme UbcH5A (2C4P), we superposed c-Cbl of the c-Cbl-UbcH7 structure 1FBV on Roc1 and subsequently superposed UbcH5A onto UbcH7.

3.2.5 Complementation assays

The CSA construct with the C-terminal thrombin cleavage site and 10xHis-tag and the same construct with the sequence for the first 30 amino acids deleted were cloned into the vector pDONR221 (Invitrogen) and verified by sequencing. Subsequently, they were put into the pLenti6.3 vector (Invitrogen) via the Gateway LR recombination reaction. The resulting plasmids were transfected to 293FT cells for virus production and the resulting virus was used to infect CS3BE-SV CS-A cells. After selection with blasticidin, the UV resistance of these cells and of a negative control (untransfected CS3BE-SV cells) and a positive control (VH10-SV) was determined by irradiating the cells with 0, 2, 4, 6 and 8 J/m^2 and determining their colony-forming ability after two weeks.

3.2.6 Protein solubility assays

pFastbac Dual vectors (Invitrogen) were constructed with the sequences for DDB1 and CSA, in which the sequence for the first 30 amino acids of CSA were deleted (CSAdeltaN) or with the mutation W361C (CSA-W361C), but which were otherwise identical to the vector for DDB1-CSAwt expression. Virus was made and amplified following our protocol for DDB1-CSAwt. 20 ml *Sf9* cells were infected with this virus or with DDB1-CSAwt virus and cells were harvested 66 hours after infection. Small-scale native purifications were performed on Ni-NTA material (Qiagen) in conditions in which free DDB1 did not bind to the column material, but CSA-DDB1 complexes do bind. Eluted fractions were run on 10% SDS PAGE gels. Three independent virus productions and Ni column purifications were performed for each test and these gave virtually identical results.

3.3 Results

3.3.1 Overall structure

To ensure that the C-terminal His-tag in our construct of CSA does not interfere with the function of CSA and hence to confirm that we crystallized a biologically relevant complex, we performed a complementation assay with the construct. Our construct could confer UV resistance to a CS-A deficient cell line (Figure 3.1) to the level of cells expressing wildtype CSA, hence it is active and the His-tag does not affect the UV damage response of CSA.

As discussed in the Methods section, the complex structure of CSA-DDB1 was solved via molecular replacement from a twinned crystal that diffracted anisotropically to 2.9 Å – data collection and refinement statistics are summarized in Table

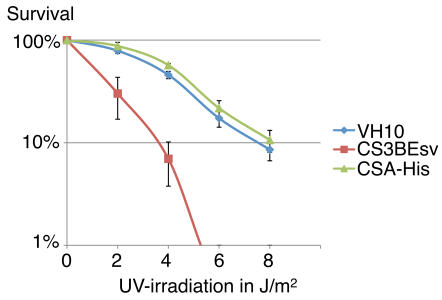

Figure 3.1: Complementation assay with CSA-His
Survival after UV irradiation of VH10-SV cells (blue), CS3BE-SV cells (red) and CS3BE-SV cells transfected with the CSA construct used for crystallization (green), showing that this construct can complement UV-sensitivity of CSA$^-$ cells.

3.1. Figure 3.2(a) shows a representative part of the structure with electron density map, indicating the quality of the map. No density was found for the C-terminal thirty amino acids of CSA nor its His-tag, probably because they are disordered.

As shown in Figure 3.2, the overall structure of CSA in the CSA-DDB1 complex consists of a seven-bladed β-propeller like most other WD40 repeat proteins such as TrCP1 (Wu *et al.*, 2003), DDB2 (Scrima *et al.*, 2008), WDR5 (Schuetz *et al.*, 2006) and Rack1 (Rabl *et al.*, 2011). From these structures, it is most similar to WDR5: the root-mean-square deviation of Cα atoms between CSA and WDR5 is 2.4 Å (using all Cα atoms that are present in both models), while the root-mean-square deviation of Cα atoms with DDB2 is 3.0 Å. The structures of different WD40 domain proteins are seen to differ mostly with CSA and with each other in the loop regions.

A number of structures of DDB1 are already known and these structures are seen to have roughly one of three possible conformations, differing in the rotation of the BPA and BPC domains versus the BPB domain. The structures of 2hye, 2b5m and 3ei4 are representative examples of these three conformations. The overall structure of DDB1 in our CSA-DDB1 complex (Figure 3.2) shows three WD40 β-propellers in a conformation similar to the structure of DDB1 in complex with the Cul4A ubiquitin ligase machinery (2hye; Angers *et al.*, 2006) with a root-mean-square deviation of Cα atoms of only 0.57 Å versus 5.6 Å and 3.3 Å for 2b5m and 3ei4 respectively.

3.3.2 Interaction of CSA and DDB1

DDB1 has been proposed to function as an adaptor protein between the Cul4A-Roc1 ubiquitination machinery and a large range of substrate receptor proteins called DDB1-Cul4A Associated Factors (DCAFs) like CSA (Iovine *et al.*, 2011). One complex structure of DDB1 with a DCAF is already known, namely the structure of DDB1 in complex with DDB2, a substrate receptor protein that functions in GG-NER. The DDB1-DDB2 complex recognizes UV-photolesions and subse-

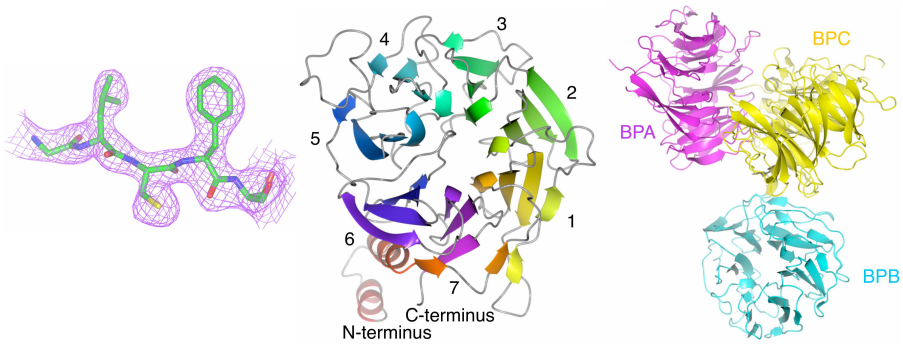

Figure 3.2: Overall fold of CSA and DDB1
(a) Representative part of the structure with electron density map (in CSA).
(b) Overall fold of CSA showing a seven-bladed β-propeller (one molecule of CSA shown; top view) with numbering of the blades of CSA.
(c) Overall fold of DDB1 showing three β-propellers (one molecule of DDB1 shown). The BPA domain is shown in magenta, the BPB domain in cyan and the BPC domain in yellow.

quently ubiquitinates XPC, DDB2 and histones surrounding the damage (Scrima *et al.*, 2011). In the structure of DDB1-DDB2 it was shown that DDB2 binds via an N-terminal helix-loop-helix motif and the bottom side of the WD40 domain of DDB2 to the interface between the BPA and BPC domain of DDB1 (Figure 3.3(a); Scrima *et al.*, 2008). Two years later, it was found that helix-loop-helix motifs from a variety of DCAFs bind to the exact same place as the helix-loop-helix motif of DDB2 in the DDB1-DDB2 structure (Li *et al.*, 2010) and it was then proposed that this is a general mode of binding to DDB1 for DCAFs. CSA, another DCAF, was therefore predicted to adopt the same binding mode.

The CSA-DDB1 complex structure indeed confirms that CSA binds via its bottom face and a 30 amino acid N-terminal helix-loop-helix motif to the interface of the BPA and BPC domain of DDB1 (Figure 3.3(b)). 57 residues in CSA and 69 in DDB1 are involved in this interaction, with 19 hydrogen bonds and 2 salt bridges formed and a buried surface area of 2183.7 Å^2 (which is 13.1% of CSA's surface and 4.5% of DDB1's surface). A large contribution comes from the first thirty amino acids in CSA: 21 of the 57 interfacing residues of CSA involved in complex formation are from this 30 amino acid N-terminal helix-loop-helix (and 7 out of 15 hydrogen bonding residues and 1 out of 2 salt bridge-forming residues). *PISA* (Krissinel & Henrick, 2007) shows that the interaction between CSA and DDB1 is strong (ΔG = -13.2 kcal/mol) and concludes that the interface plays an essential role in complex formation and is stable in solution.

To verify the importance of CSA's N-terminal helix-loop-helix in complex formation, we investigated its role biochemically. Firstly, DDB1 was overexpressed with an N-terminal truncated CSA construct in insect cells in an identical way as wild-

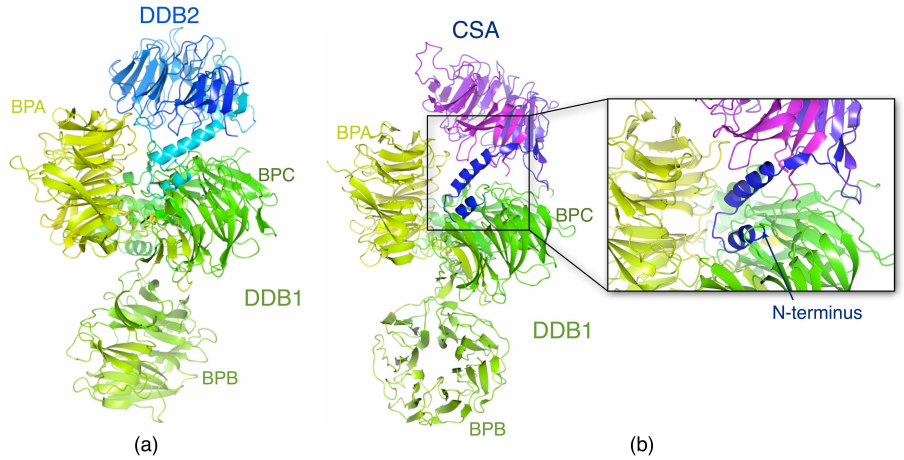

Figure 3.3: Interaction of CSA with DDB1 goes via a general binding mode for DCAFs to DDB1.
(a) Crystal structure of DDB2 (blue) and DDB1 (yellow to green) showing that DDB2 binds DDB1 via a N-terminal helix-loop-helix motif inserted between DDB1's BPA and BPC domain (Scrima et al., 2008).
(b) Complex of CSA and DDB1 as found in our crystal structure with CSA in blue to purple and DDB1 in yellow to green. The interface of CSA with DDB1 with the helix-loop-helix of CSA inserting between the BPA and BPC domain of DDB1 is shown in detail in the box.

type CSA to see whether binding of DDB1 and CSAΔN can occur *in vitro*. Nickel column elutions from this overexpression can be seen in Figure 3.4(a) (using the C-terminal 10xHis-tag on CSA). The amount of DDB1-CSAΔN complex was seen to be severely reduced (around ten times), suggesting that CSA's N-terminal helix-loop-helix indeed plays an important role in stability. To confirm the importance of the N-terminus, we performed a complementation assay with the same construct of CSAΔN30: no complementation of a CS-A cell-line was observed (Figure 3.4(b)), showing that the helix-loop-helix motif of CSA is vital for its function. The reduced stability and lack of complementation of CSAΔN30 can very well be caused by reduced interaction with DDB1.

3.3.3 Substrate binding site of CSA

In TC-NER, CSA gets recruited to the stalled polymerase as part of an E3-ubiquitin ligase complex that is inactive at the time of recruitment. CSA is thought to be the substrate-recognizing protein, but exactly which protein(s) it binds to is not known. One of the probable candidates is Cockayne syndrome protein B (CSB), which has been shown to be ubiquitinated by the Cul4A-Roc1-DDB1-CSA complex *in vitro* (Groisman *et al.*, 2006).

To find a probable substrate-binding site for CSA, its structure was superpositioned on co-crystal structures of WD40 domain proteins bound to substrate pep-

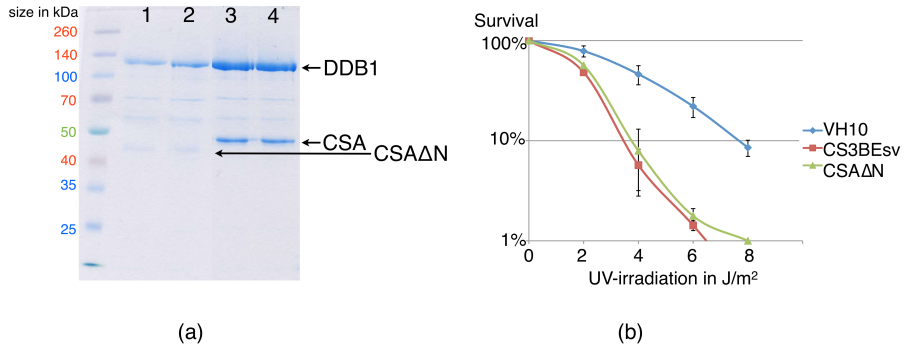

Figure 3.4: Importance of the N-terminal 30 amino acids of CSA.
(a) SDS PAGE gel showing the result of a small-scale Ni column purification with DDB1-CSAΔN (lanes 1 and 2) compared to DDB1-CSAwt (lanes 3 and 4) showing that complex formation is severely reduced in the absence of the N-terminal 30 amino acids, though not completely abolished. Results of two independent virus productions are shown to illustrate the reproducibility.
(b) Survival after UV irradiation of VH10-SV cells (blue), CS3BE-SV cells (red) and CS3BE-SV cells transfected with the CSAΔN construct (green), showing that an N-terminally truncated CSA cannot complement UV-sensitivity of CSA- cells.

tides. Some observed binding sites were clearly not possible for CSA, because the expected shape of the entire E3-ubiquitin ligase complex excludes e.g. peptide binding at or too close to the DDB1-CSA interface (Figure 3.5(a)). Considering the shape of the E3-ubiquiting ligase complex, the peptides bound to the top surface of the WD40 domain gave the most probable outcomes as models for a substrate of CSA. The centre of the top face is also the most frequently occurring site for substrate binding to WD40 proteins (Xu & Min, 2011; Stirnimann *et al.*, 2010).

A charge surface plot of the structure of CSA showed a small, positively charged hole in the centre of the top face that we propose as the most likely substrate binding position (Figure 3.5(b)). This hole is similar to, though slightly smaller than, the one observed in TrCP1 (Wu *et al.*, 2003) where a phosphorylated protein binds. It is unlike DDB2, which has a large positive groove where DNA binds. CSA's small positively charged hole also differs from the binding pocket of WDR5 (Schuetz *et al.*, 2006), which has an extensive negative charge where methylated histone tails bind. Since the substrate-binding interface looks most like the interface of TrCP1, and other WD40 proteins that bind phosphorylated proteins, we propose that the part of the substrate that CSA recognizes is a stretch of a phosphorylated protein, though a negatively charged stretch of a non-phosphorylated protein is also an option.

From the comparison of the CSA structure to structures of WD40 proteins in complex with substrate peptides, we suggest that the residues of CSA involved in substrate-binding include Glu103, Phe120, Lys122, Arg164, Lys247, Lys292 and Arg354 (see Figure 3.5(c)), because these residues stick out into the solvent at the

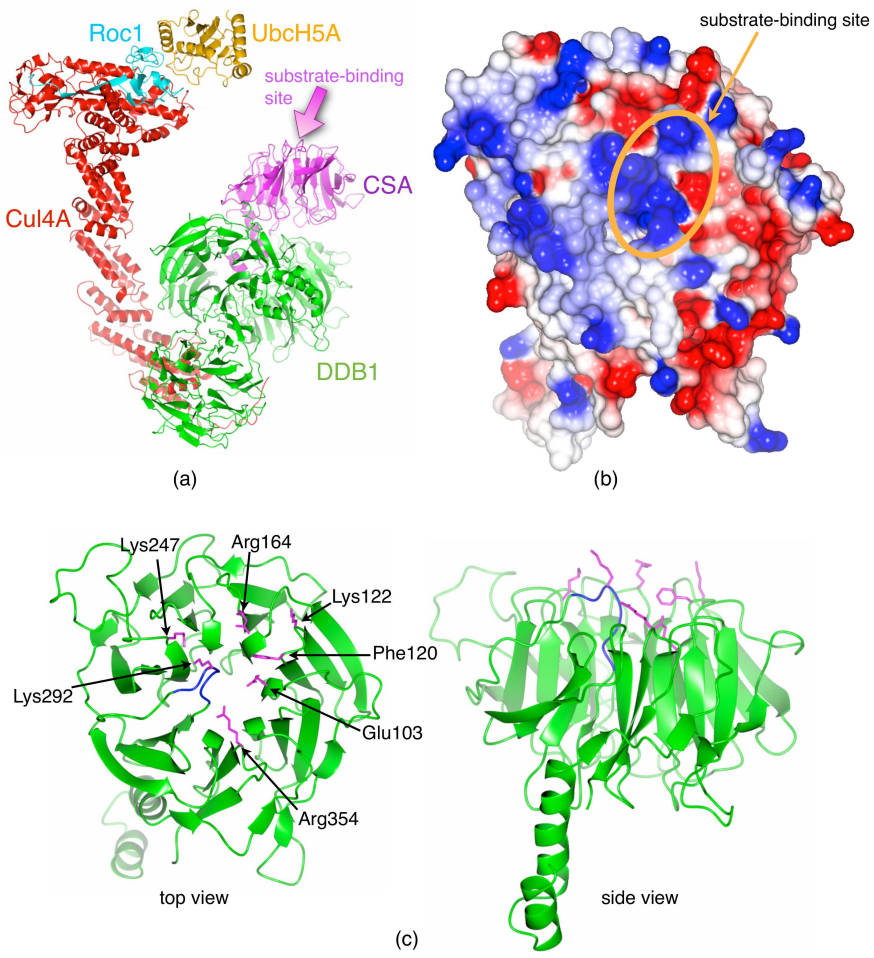

Figure 3.5: The proposed substrate-binding site of CSA.
(a) Model of the CSA-DDB1 structure (CSA in magenta and DDB1 in green) as part of the E3-ubiquitin ligase complex with Cul4A (red), Roc1 (cyan) and UbcH5A (orange). The arrow indicates the proposed substrate-binding site of CSA.
(b) Electrostatic surface representation of CSA with blue representing positive charge and red negative charge. The circle indicates the proposed substrate-binding site.
(c) The structure of CSA showing the proposed substrate-binding residues in magenta and the loop around 290-296 in blue.

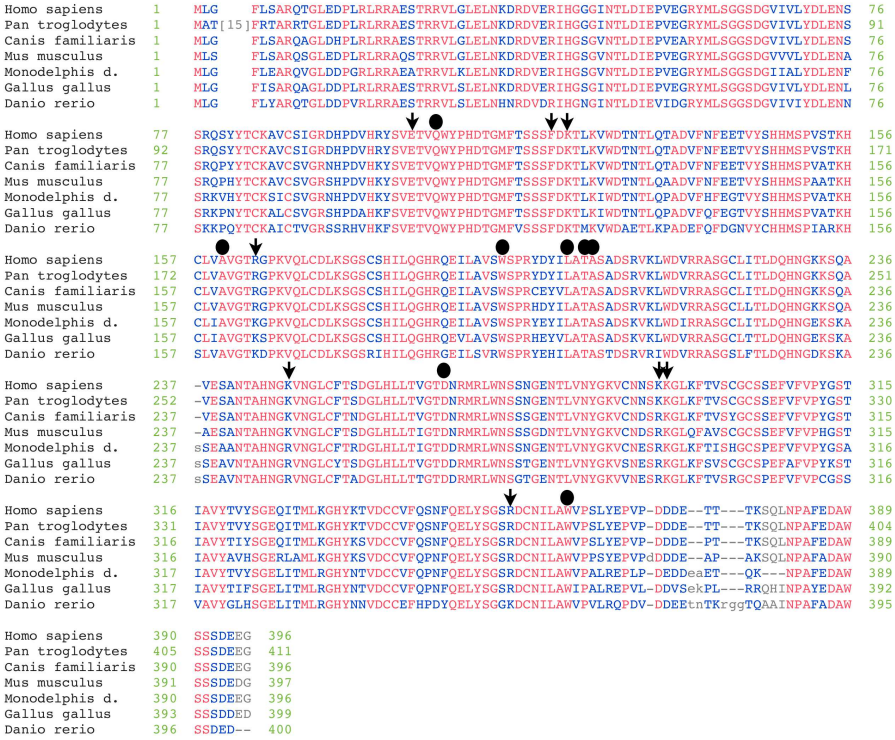

Figure 3.6: Alignment of CSA from different species. Arrows point at proposed substrate-binding residues and dots mark residues mutated in Cockayne syndrome. Red indicates identical residues, blue partially conserved residues and grey unaligned residues.

proposed substrate-binding site and are close to important parts of substrate peptides in the superposition of CSA with WD40-peptide co-crystals. Especially the loop around residues 290-296 is notable, because it is in the middle of the proposed substrate-binding site, it contains four charged residues close to each other (of which three are lysines) and it is notably different from the loops of other WD40 proteins at this location. The proposed substrate-binding residues are also rather conserved in CSA in different species (see Figure 3.6).

3.3.4 Disease-causing mutations

Mutations in CSA can lead to the recessive human disorder Cockayne syndrome (CS). So far 18 different mutations in CSA causing CS are known of which there are 8 distinct missense mutations (Laugel *et al.*, 2009). All eight missense mutations are located in the seven WD40 repeats and consist of changes in residues that are conserved in the orthologs of CSA (Figures 3.6 and 3.7(a)).

The residues Q106 and A205 are located in β-strands in blades 2 and 4 (Figure

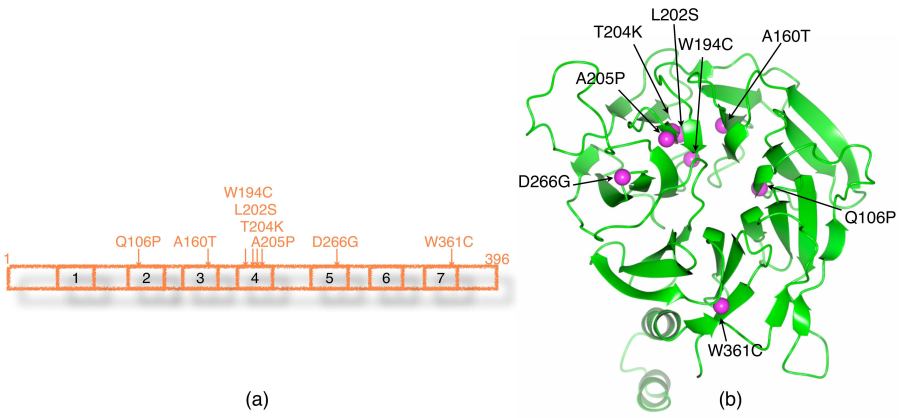

Figure 3.7: Disease-causing mutations in CSA.
(a) Schematic overview of the eight point mutations in CSA that cause Cockayne syndrome, shown on the primary structure. The boxes indicated the seven WD40 repeats of CSA.
(b) Disease-causing mutations shown as magenta balls on a cartoon representation of the structure of CSA.

3.2(b)). The mutation to proline in the disease-causing mutations Q106P and A205P probably disrupts the β-strands' proper conformation due to the rigidity of the proline (see Figure 3.7(b)). Disruption of these β-strands – and potentially their neighboring strand(s) – will put residues near the proposed substrate-binding site (e.g. Glu103 and Lys247) out of their normal position and hence disturb substrate binding. The disruption of the beta-strands may also have a more global, negative influence on the overall folding and thus the stability of CSA.

Mutations A160T, W194C, L202S and T204K, though separated from each other in the primary structure of CSA, were found to be surprisingly close to each other in the three-dimensional structure of CSA (Figure 3.8(a)). These four residues are in β-strands in blades 3 (A160T) and 4 (W194C, L202S and T204K) and besides being conserved in CSA orthologs (see Figure 3.6) they are also fairly conserved in other WD40 repeat proteins. At the corresponding position of A160 usually a small to medium hydrophobic residue is found (G, V, M) in WD40 domain proteins. At the position of W194 usually a medium to large hydrophobic residue can be found (like W, Y, F, I, L). At the position of L202 there is usually a medium to large hydrophobic residue (like L, V, I, F). The position of T204 is usually occupied by a relatively small, uncharged residue (A, S, T, G). It seems that the side-chains of the residues A160, W194 and L202 form a small hydrophobic pocket that is important for the stability of the fold and that a substitution to a hydrophilic residue might disrupt this pocket. Substitution of the small threonine at position 204 for lysine will result in a positive charge very near this pocket – when taking into account the length of a lysine side-chain compared to a threonine – and thus will also result in a disruption around the same point. This may, for example, cause a disruption in the β-strands in these

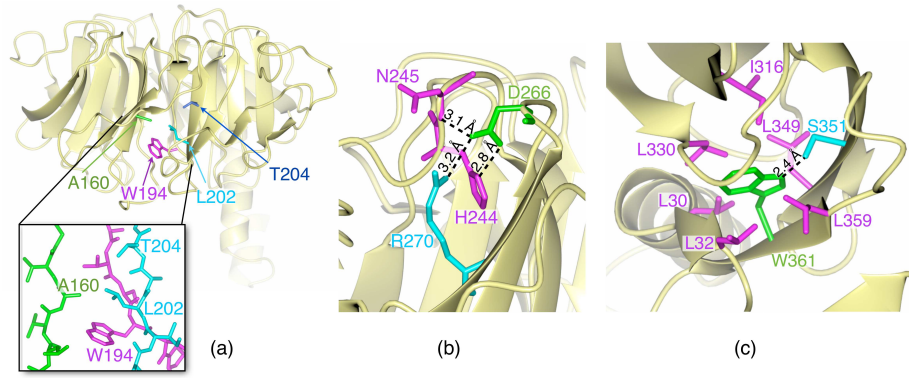

Figure 3.8: Close-up of disease causing mutations.
(a) Side-view of CSA and close-up showing that mutations A160T (green), W194C (magenta), L202S (cyan) and T204K (cyan/ blue) are close together in the three dimensional structure.
(b) Close-up of CSA showing the hydrogen bonding of residue D266 (in green), which is mutated to glycine in some patients.
(c) Side-view of CSA and close-up showing the environment of residue W361 (in green) involved in mutation W361C.

blades and thus make that the proposed substrate-binding residues at the end of these strands (R164 and K247) are positioned incorrectly for binding or this may affect the overall folding and hence reduce the stability of the protein.

Residue D266, which is mutated in some CS patients to a glycine, is located in blade 5 close to CSA's proposed substrate-binding site. This residue is one of the most conserved residues in WD40 motifs and is expected to be important for proper folding via side-chain interaction with H244, which is also conserved in WD40 proteins. Indeed, the side-chain of D266 can be seen in the CSA structure to form a hydrogen bond with the side-chain of H244, as well as with the amide nitrogen of N245 (Figure 3.8(b)). These important interactions will be lost in the mutant D266G. Since this residue is very close to the predicted substrate-binding site, disrupted folding at this position can affect the substrate-binding and/ or the global structure of CSA.

The last point-mutation causing disease, W361C (in blade 7), has an interesting phenotype: this mutant has been reported to confer hypersensitivity to UV radiation but less to inducers of oxidative damage that are usually harmful to cells from CS patients (Nardo *et al.*, 2009). This finding suggests that the mutation does not affect the overall fold substantially, otherwise all the functions of CSA would have been completely lost. It also shows that CSA acts differently in the removal of DNA lesions that are target to TC-NER and in the removal of oxidative lesions (Nardo *et al.*, 2009). This mutation changes tryptophan 361 from the last WD motif and hence it alters a conserved residue that is important for beta-strand stability. In the CSA structure, it is surrounded by medium sized hydrophobic side chains and one serine (S351) to which its side-chain nitrogen makes a hydrogen bond (Figure

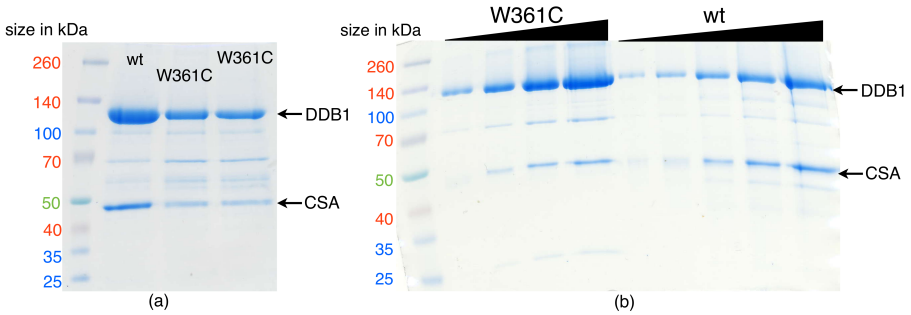

Figure 3.9: Effect of the W361C mutation on CSA in vitro.
(a) SDS PAGE gel of CSA-W361C versus CSA wt after small-scale Ni-columns showing the reduced production of CSA-W361C versus wildtype CSA. Two independent W361C overproductions are shown to illustrate the reproducibility. (b) SDS PAGE gel showing different amounts of (concentrated) CSA-W361C after a full purification scale purification compared to wildtype after a small-scale purification. A difference in the ratio of DDB1 to CSA can be seen between CSAwt and CSA-W361C.

3.8(c)). Mutation of the large and hydrophobic amino acid tryptophan to cysteine might disturb this hydrophobic pocket. W361 is located very close to DDB1 – L30 and L32 in Figure 3.8(c) are located at the end of the helix-loop-helix through which CSA binds to DDB1 – and thus the disturbance might affect the stability of the CSA-DDB1 interaction.

However, the patient phenotype of this mutant shows that it leads to a subtle change that takes out only part of the function of CSA. To investigate this further we overproduced the mutant CSA-W361C in insect cells together with DDB1 and performed small-scale Ni-column purifications of the complex (using the 10xHis-tag on CSA). This showed that the overproduction of DDB1-CSA-W361C was very much reduced compared to DDB1-CSA-wt (Figure 3.9(a)). A large-scale purification was then performed following the same protocol as wildtype (chapter 2). The mutant behaved identical to wildtype in the Ni-column and Q-column, but eluted slightly later on the gel filtration. The yield of DDB1-CSA-W361C complex was again considerably lower than with DDB1-CSAwt (approximately tenfold), though the amount of free DDB1 was slightly higher than wildtype, showing that the virus production had been successful. Moreover, the DDB1 to CSA ratio of the mutant seemed to be higher than wildtype (around 3 to 1 instead of 1 to 1) at the end of the purification procedure (Figure 3.9(b)), though all free DDB1 had been removed during the first stage of the purification, suggesting that CSA is lost during the purification procedure (around 18 hours in total). These results indicate that the CSA-DDB1 interaction in the mutant W361C is partially disturbed and hence some complexes lose CSA over time, which then precipitates since, as described in Chapter 2, CSA is insoluble on its own. The patient phenotype of CSA-W361C might then be explained by assuming a lower stability and hence a lower amount of CSA in the cell, but not complete absence – apparently enough for some oxidative damage re-

pair but not for repair of UV lesions. Since information on CSA protein levels in patient cells is currently unavailable, we cannot verify or falsify this hypothesis. Another possibility is that the interaction of CSA with DDB1 is changed in such a way (e.g. more flexible and hence more likely to dissociate *in vitro*) that CSA cannot position its substrate in TCR correctly, but that it can still function in repair of oxidative damage.

Besides these missense mutations, nonsense mutations and deletion/ insertion mutations have been reported (Laugel *et al*., 2010). All of these lead to a severely truncated protein or lead to a deletion in at least one WD-motif, hence all of these mutations can be expected to have a profound negative influence on the overall structure and therefore on the function of CSA.

3.4 Discussion

Cockayne Syndrome protein A is a crucial player in TC-NER and has a high biological importance: mutations in this protein cause the serious human disease Cockayne Syndrome. The structure has the expected fold, a seven-bladed β-propeller, but now that the crystal structure is available the disease-causing mutations can be mapped onto this structure and it can be explained how they cause the disease. We found that these mutations most probably interfere with the proper folding of the β-strands or interactions between β-strands, and hence disrupt local folding of CSA, interfering with substrate-binding mostly, or with DDB1-interaction, or they affect more global folding of CSA. This type of errors is very difficult to correct for and hence, unfortunately, our structure cannot aid in overcoming the problems caused by these errors that lead to disease.

The structure does, however, yield a more general insight into the function of this protein and the process of TC-NER. One insight is that the interaction of CSA and DDB1 is similar to the complex of DDB1 with the substrate receptor DDB2 that functions in GG-NER and which, like CSA, binds to the interface of the BPA and BPC domain of DDB1 via an N-terminal helix-loop-helix motif and the bottom face of its WD40 domain of DDB2. The biological importance of this binding motif was clear from our *in vitro* and *in vivo* studies. Several other DDB1-CUL4-associated factors (DCAFs) were also proposed to use such a motif to bind to DDB1, because they contain a similar helix-loop-helix N-terminal to the WD40 domain and these helical motifs were shown to bind to DDB1 in a similar position (Li *et al*., 2010). Furthermore, the V protein of simian virus 5 has been shown to interact with DDB1 via the BPC domain and it also uses a similar helical motif to bind to DDB1. This helical motif is not predicted for all DCAFs, hence not all DCAFs necessarily have an identical binding mode, but the fact that our structure shows the exact same binding mode to DDB1 as DDB2, is a strong confirmation that it is a general feature of DCAFs and most probably occurs widely.

Moreover, our structure gives insight into the currently unknown substrate(s) of CSA. Its most probable substrate-binding site has a positive hole that resembles that seen in other WD40 proteins that bind phosphorylated parts of other proteins. Indeed, SCF ubiquitin ligases, which mediate ubiquitination of proteins involved

in transcription, signal transduction and cell cycle progression, use an F-box containing protein as substrate-presenting unit. These proteins use their C-terminal WD40 domain for recognizing phosphorylated substrate proteins and thus mediate their ubiquitination (Xu *et al*., 2011). Phosphorylation of the substrate of CSA could present a way of regulation in order to ensure the proper timing of ubiquination of CSA's substrate(s) in the process of TC-NER.

Based on the structure we cannot predict CSA's substrate, although we can confirm CSB to be a proper candidate. CSA might either recognize a phosphorylated stretch or a negatively charged stretch on CSB. Notably, CSB contains a negatively charged acidic domain in its N-terminal end, but has also been reported to be subject to phosphorylation (Christiansen *et al*., 2003 and Imam *et al*., 2007). However, further research is required to confirm the substrate of CSA and the regulation and role of ubiquitination in transcription-coupled repair.

4

Substrate-binding by Cockayne Syndrome protein A

Cockayne Syndrome protein A (CSA) is one of the main players in mammalian Transcription-Coupled Nucleotide Excision Repair and mutations in this protein lead to the serious human disorder Cockayne Syndrome. CSA is thought to be the substrate adaptor protein of an E3-ubiquitin ligase complex, but its exact substrate is still a matter of debate. To get more insight into the function of CSA in Transcription-Coupled repair, we performed *in vitro* and *in vivo* experiments for determining its substrate. In this chapter we describe the results that show that CSA probably does not bind DNA, while the related protein DDB2 does, but that a more likely substrate is a negative stretch of a protein such as encountered in Cockayne Syndrome protein B.

E.M. Meulenbroek, M.G. Vrouwe, L.H.F. Mullenders, N.S. Pannu, *Insights into Transcription-Coupled Repair from the crystal structure of Cockayne Syndrome protein A.* (to be submitted)

4.1 Introduction

To repair DNA lesions blocking transcription, the DNA repair pathway of Transcription-Coupled Nucleotide Excision Repair (TC-NER) has evolved. In mammals, most of the proteins involved are shared with the Global Genome Nucleotide Excision Repair (GG-NER) pathway, but two factors are unique for TC-NER: Cockayne Syndrome protein A and B (CSA and CSB) (reviewed in a.o. Tornaletti, 2009). Their function in TC-NER is suggested to be as follows (reviewed in Fousteri & Mullenders, 2008). CSB is thought to interact dynamically with RNA polymerase II during transcription. Upon encountering a lesion past which RNA polymerase II cannot transcribe, it gets stalled on the lesion and this stabilizes its interaction with CSB. CSB recruits NER factors and chromatin remodelers to the lesion as well as CSA in an E3 ubiquitin ligase complex (constituted of CSA, DDB1, Cul4A and Roc1) inactive at the time of recruitment. The function of CSB might be to make space for the repair with its DNA-dependent ATPase activity. The CSA complex has been suggested to protect TC-NER factors from degradation in early stages of repair and/ or direct proteins for degradation at later stages of repair to prevent them from obstructing transcription (Groisman *et al.*, 2006). The repair is eventually done by excision of the fragment containing the lesion by the NER factors. The biological importance of CSA and CSB is demonstrated by mutations in either CSA or CSB, that have been reported to cause the serious human recessive disorder Cockayne syndrome, which is characterized by a.o. dwarfism and premature aging (Nance & Berry, 1992).

The crystal structure of CSA has recently been solved (see Chapter 3). In the crystal structure, CSA was seen to form a complex with DDB1 that is similar to the DDB1-DDB2 complex, which is involved in initial detection of UV-lesions in GG-NER (Scrima *et al.*, 2008). Like CSA, DDB2 also forms an E3-ubiquitin ligase complex with Cul4A, Roc1 and DDB1. DDB2 binds to lesions in the DNA and there it promotes ubiquitination of the nearby XPC and itself, leading to the handover of the damage from DDB2 to XPC (Sugasawa *et al.*, 2005).

In the structure of CSA it could be seen that the most likely substrate-binding site is positively charged and small, as is also observed in proteins that bind e.g. phosphorylated stretches of other proteins. This site was considerably smaller than the large positive groove that the DNA-binding protein DDB2 has. Therefore, it was suggested that CSA binds to a phosphorylated or otherwise negative stretch of another protein, but its substrate could not yet be identified from these results. In this chapter we describe experiments to get more insight into the substrate for CSA.

4.2 Material and methods

4.2.1 Protein overproduction and purification

Overproduction and purification of the CSA-DDB1 complex has been reported in Chapter 2 of this thesis. Briefly, CSA with a C-terminal 10x His-tag was co-

expressed in *Sf9* insect cells with N-terminally 6x-His tagged DDB1. The complex was co-purified to near homogeneity with Ni-column affinity chromatography, ion exchange and gel filtration.

4.2.2 Bandshift assay

The oligos containing CPD or 6-4PP were synthesized as described in Iwai, 2006, all the other oligos were bought at Eurogentec, Belgium. The top strands of the DNA substrates were 5' radioactively labelled using polynucleotide kinase as reported previously (Verhoeven *et al.*, 2002). Labelled DNA substrates were incubated for 10 min at 30 °C with up to 0.62 μM CSA-DDB1 in 20 mM Tris pH 6.5 and 100 mM NaCl in a reaction mix of 10 μl. Samples were then loaded on a 3.5 % native Tris-acetate gel at pH 7.2, which was run at 4 °C in 1x TAE pH 7.2. The gel was dried and the result was visualized by irradiation of a photographic film.

4.2.3 Construction of CSA mutants

CSA mutants were created by PCR and cloned into the pDONR221 vector (Invitrogen) with a C-terminal 10x-His-tag as was CSAwt. The presence of the desired mutation was verified by sequencing analysis (Baseclear, the Netherlands). The gene with mutation and tag was then transferred to the pLenti4 vector (Invitrogen) via the Gateway LR recombination reaction. The resulting vector was transfected to 293FT cells for virus production and this virus was then used to infect CS3BE-sv cells. After selection with Zeocin, 250 cells of these mutant cell lines or of the positive control (VH10 cells) or the negative control (untransfected CS3BE-sv) were seeded on a 90 mm petridish and were irradiated with 0, 2, 4, 6, and 8 J/m^2. After two weeks, the colony-forming ability was assessed. All experiments were executed four to eight times.

4.2.4 CSA interaction studies

Around 30 * 10^6 human cells were harvested by trypsination. After inactivation of the trypsin, centrifugation and washing with 1x PBS, the cells were frozen at -80 °C until use. The cell lines used were: VH10 cells or CS3BE-sv cells as negative controls (both gave similar results, hence only one negative control is shown in the remainder of the chapter), and CS3BE-sv cells into which the gene coding for CSA-His was transfected as described above.

Cells were resuspended in 1 ml lysis buffer (50 mM Tris pH 8, 200 mM NaCl, 0.1% Triton-X-100, 5 mM β-mercaptoethanol and Complete mini EDTA-free protease inhibitor cocktail (Roche)) and lysed by sonification. Cells were spun down at 13.000 rpm in an Eppendorf table-top centrifuge for 10 min at 4 °C and the soluble fraction (lysate) was loaded on 200 μl Ni-NTA column material (Qiagen) equilibrated in Ni buffer A (20 mM Tris pH 8, 200 mM NaCl, 20 mM imidazol, 5 mM β-mercaptoethanol). The small column was washed with 1.5 ml Ni buffer A and the CSA-His protein together with interacting proteins were eluted with

300 μl Ni buffer B (20 mM Tris pH 8, 200 mM NaCl, 330 mM imidazol, 5 mM β-mercaptoethanol).

Samples of the lysate and the elution fractions were run on 10% SDS PAGE gels and blotted o/n at 90 mA per gel. The next morning, the blots were washed in 1xTBST and blocked with 5% protifar in TBST for 1 h. The first antibody (GαCSA; W-16 from Santa Cruz biotechnology or GαCSB; E-18 from Santa Cruz biotechnology) was added in 5% protifar in TBST and incubated for 2 h at a 1:200 dilution (αCSA), 1: 250 (αCSB) or 1:10,000 (αHis). After washing 3 times 15 min with 1x TBST, the secondary antibody (RαG and GαM HRP) was added in 5% protifar in TBST and incubated for 1 h. After washing 3 times 15 min with 1x TBST, the blots were developed using a luminol - enhancer - hydrogen peroxide mixture.

4.3 Results

4.3.1 Verification of CSA's substrate binding site

From the crystal structure of CSA, several residues have been identified as potential substrate-interacting residues (see Chapter 3): Glu103, Phe120, Lys122, Arg164, Lys247, Lys292, Lys293, Arg354 (Figure 4.1(a)). These residues were selected using the following three criteria: the side-chains should be probable candidates for interaction (e.g. charged or big hydrophobic side-chains), the side-chains should be interacting with or otherwise be very near substrate peptides in a superposition of CSA with co-crystal structures of WD40 proteins with peptides, and the side-chains should not be involved in any interaction that is likely to be important for the stability of the structure of the protein itself.

To verify the importance of these residues for the function of CSA, complementation studies were performed with these residues changed to alanine. As can be seen in Figure 4.1(b), cells with most of the mutants are only marginally more UV-sensitive than cells with wildtype CSA (VH10 and CSAwt), but one mutant shows a clear UV-sensitivity (Figure 4.1(c)). This is the only double mutant of the set of mutants (K292A + K293A), of which the mutations are located exactly in the centre of the proposed substrate-binding site in a positively charged loop that is unique to CSA (blue in Figure 4.1(a)). The phenotype of this mutant confirms that the top face of WD40 domain of CSA is important for CSA function and hence reinforces the idea that this face is a likely site for substrates to bind to CSA. However, it should be noted that the sensitivity is only partial, though one might have predicted a more severe sensitivity for a double mutant.

4.3.2 Determination of CSA's binding capacity to DNA

Since the structurally related DDB1-DDB2 complex binds strongly and specifically to damaged DNA (Wittschieben *et al.*, 2005), we decided to test whether DDB1-CSA also binds to DNA. To this end bandshift assays were performed in many different conditions. The only conditions in which binding of CSA to DNA was observed, was at pH 6.2 - 7.2 with a low amount of salt (0-100 mM) in gels run in

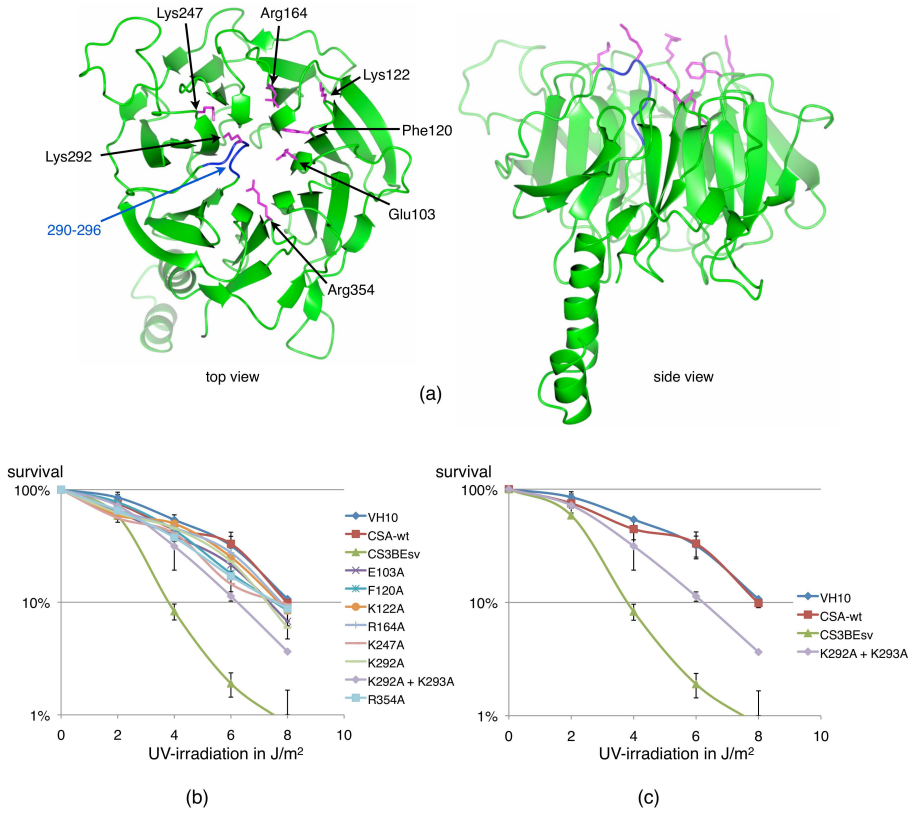

Figure 4.1: Verification of CSA's proposed binding site.
(a) Cartoon representation of the CSA structure with the residues potentially involved in substrate-binding indicated in magenta. The loop 290-296 is indicated in blue.
(b) Results of complementation assays with CSA's substrate-binding mutants, showing that most mutants only have a marginal effect on UV-sensitivity, but that the double mutant K292A + K293A has a larger effect. The positive controls are in red (CSA-His construct) and blue (VH10; cell line with wt CSA) and the negative control is in green (CS3BEsv).
(c) Result of Figure 4.1(b) showing only the controls and the double mutant K292A + K293A to indicate more clearly the disturbed complementation of this mutant.

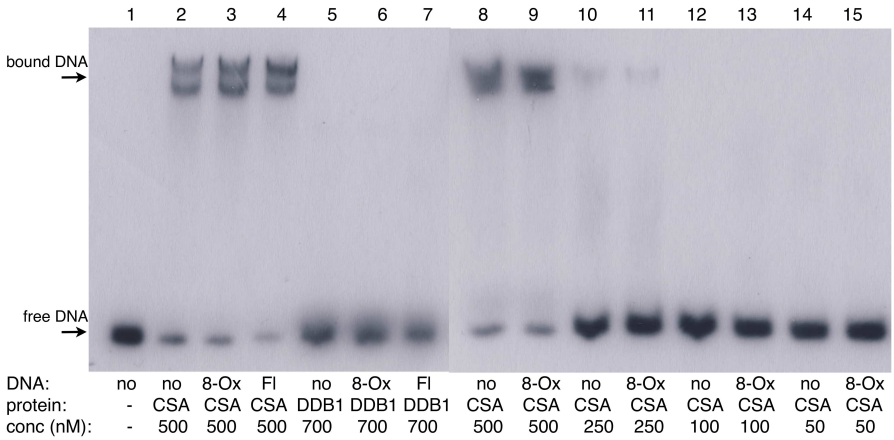

Figure 4.2: DNA-binding by CSA
Representative bandshift assay with CSA showing that CSA only binds DNA at very high protein concentrations in a non-damage-specific way. Below the lanes is indicated what concentration of protein was used, whether the CSA-DDB1 complex (denoted: CSA) was used or only DDB1 (denoted: DDB1) and which DNA damage was used (no: no damage, 8-Ox: 8-oxoguanine, Fl: fluorescein). The bands were verified not to be in the slots of the gel.

TAE at pH 7.2. High concentrations of protein had to be used (0.5 μM) with DNA concentrations up to 1.8 nM to see a significant binding (Figure 4.2). Hardly any binding could be observed at lower protein concentrations, even not at 2x lower protein concentration (Figure 4.2 lanes 8-15). No discrimination between damaged and undamaged DNA was observed (Figure 4.2 lanes 2-4 show some representative examples; a wide range of lesions ranging from 8-oxoguanine to fluorescein to UV lesions as well as different DNA lengths were attempted; not all results shown). The bandshift was CSA-dependent as samples with only DDB1 did not show bandshift (Figure 4.2 lanes 5-7). Filter-binding assays (results not shown) confirmed the above results.

Binding of CSA to DNA could only be observed at high protein excess (250-fold or more molar excess) with no discrimination between different DNA substrates in the large range of conditions tested, while DDB1-DDB2 shows bandshifts at equimolar DNA-protein concentrations and an up to 80x higher affinity for damaged DNA (Wittschieben *et al.*, 2005). Therefore, we are unsure of the biological relevance of the observed binding and should keep in mind that the binding we observed might be an artefact. The fact that we don't see any shift in more biologically relevant ratios while many conditions were tested, shows that it is very unlikely that CSA indeed binds DNA *in vivo*.

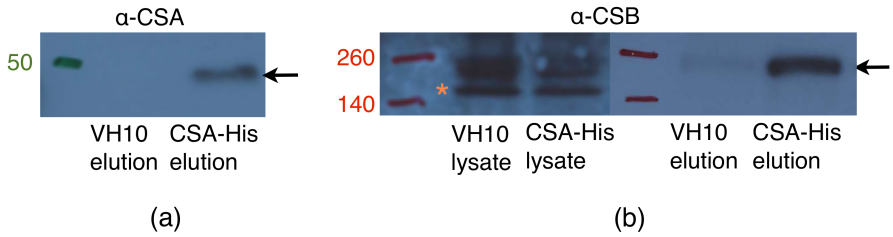

Figure 4.3: CSB as potential substrate for CSA
Western-blot with αCSA (a) or αCSB (b) antibodies on the elutions of small-scale Ni-affinity purifications with cell lines expressing wildtype CSA (VH10) or overexpressed CSA with a C-terminal His-tag (CSA-His). For the CSB Western-blot (b), also the lysates are shown to be able to assess the amount of CSB in the total soluble fraction. The arrows indicate the specific band for CSA or CSB and the orange star indicates the CSB-PiggyBac fusion protein, a splice-variant of CSB. Brightness and contrast of the left panel of (b) have been adapted by computer to improve the image quality and the right panel is taken from the same blot with a shorter exposure time of the film.

4.3.3 CSB is a possible substrate for CSA

The substrate for CSA in TC-NER is still uncertain, but one of the most likely options is CSB, since CSA has been reported to ubiquitinate CSB *in vitro* and a CSA-dependent degradation of CSB after UV-irradation has been observed (Groisman *et al.*, 2006). However, this matter has remained controversial due to contradicting results in literature, for example on the degradation of CSB after UV-irradiation and on the physical interaction between CSA and CSB in cells (see for a discussion Licht *et al.*, 2003). To get more insight into the possibility of CSB as substrate for CSA, we performed pull-down analysis using the 10x His-tag on our CSA construct in complemented CS3BE-sv cells.

In Figure 4.3(a) it can be seen that small-scale Ni-column purification leads to a high enrichment of CSA in the samples from the cell-line with CSA-His overexpressed, as expected, showing that the experimental setup is correct. As can be seen in Figure 4.3(b), less CSB is present in the lysate of the cell line where CSA is overexpressed. This suggests that CSA is indeed involved *in vivo* in ubiquitination and degradation of CSB. Moreover, when these samples are partially purified with Ni column affinity chromatography (using the His-tag of the overexpressed CSA), there is considerably more CSB in the Ni elution fraction of the sample with CSA-His overexpressed than in the negative control (Figure 4.3(b)), in which no tagged CSA is present. This shows that CSB co-purifies with CSA. This indeed suggests that there is a (direct or indirect) interaction between CSA and CSB. Similar experiments using CSA with a mutated top face are currently in progress.

4.4 Discussion

For understanding the function of CSA in TC-NER and how defects in CSA cause the disease phenotype, it is vital to know how CSA recognizes its substrate and what this substrate is. In this chapter site-directed mutagenesis results are reported that suggest that the top face of the WD40 domain of CSA is indeed the substrate-binding site, considering the effect a double mutation in the centre of the proposed site had. Single mutations in this site did not have a clear effect. It is possible that the affinity for the substrate is not strongly affected by mutation of only one residue and that the high expression of CSA in the WD40 top-face mutant cell lines might compensate for slight decreases in efficiency. A double mutation then perhaps has a too strong influence to be compensated for by a larger amount of CSA. Therefore, for the double mutation an effect is observed. Together with the fact that it is the most used substrate-binding site in WD40 proteins and that it is accessible in a model of CSA in the E3-ubiquitin ligase complex, the UV-sensitivity of the double mutant provides evidence pointing toward this being the substrate binding site.

Next, we investigated the identity of the substrate for CSA by looking at the two most likely options. The first option, CSA binds DNA like DDB2 does, is very unlikely based on our results, since binding to DNA could only be observed in one condition of salt and pH with a very large excess of protein (250 times or more) and no discrimination between different substrates was observed. This binding therefore most likely only represents aspecific binding of CSA to the DNA, e.g. to the negatively charged phosphate groups. Moreover, CSA was reported not to co-purify with nucleosomes after UV-irradiation unlike DDB1-DDB2 (Groisman *et al.*, 2003), which agrees with our results.

Our findings indicate that the second option, CSA binds CSB leading to its ubiquitination, is more likely, since less CSB was shown to be present in cell lines over-expressing CSA than in cell lines with no or normal amounts of CSA. Moreover, CSB could be pulled down using the His-tag on CSA. The results in literature on this matter contradict each other. A physical interaction between CSA and CSB was found with *in vitro* translated proteins and with the yeast-two hybrid system (Henning *et al.*, 1995), but not in cells using gel filtration, co-immunoprecipitation or immunofluorescence (discussed in Licht *et al.*, 2003). An explanation for this controversy can be that the amount of CSA in cells normally is so low that very sensitive methods and optimized experimental conditions are required to pick up a CSA-CSB interaction. Different methods have been used in literature and perhaps not all are sensitive enough to allow detection of weakly interacting, rare complexes. Our method, however, has the advantage in this regard that CSA-His is overexpressed, so it allows easier detection of CSA-interactions, though its disadvantage is that overexpression is less close to the normal condition of the cell. Moreover, our method does not prove that the observed interaction is direct or whether it is mediated via another protein, hence further research is needed to identify with certainty the substrate of Cockayne Syndrome protein A.

5

Involvement of a carboxylated lysine in UV damage endonuclease

UV damage endonuclease is a DNA repair enzyme that can both recognize DNA damage such as UV lesions, and introduce a nick directly 5′ to them. Recently, the crystal structure of the enzyme from *Thermus thermophilus* was solved. In the electron density map of this structure, unexplained density near the active site was observed at the tip of Lys229. Based on this finding, it was proposed that Lys229 is post-translationally modified. In this chapter, we give evidence that this modification is a carboxyl group. By combining activity assays and X-ray crystallography on several point mutants, we show that the carboxyl group assists in metal-binding required for catalysis by donating negative charge to the metal-coordinating residue His231. Moreover, functional and structural analysis of the K229R mutant reveals that if His231 shifts away, an increased activity results on both damaged and undamaged DNA. Taken together, the results show that *T. thermophilus* UVDE is carboxylated and the modified lysine is required for proper catalysis and preventing increased incision of undamaged DNA.

published in: E.M. Meulenbroek, K. Paspaleva, E.A. Thomassen, J.P. Abrahams, N. Goosen, N.S. Pannu, *Involvement of a carboxylated lysine in UV damage endonuclease.*, Protein Sci. **18**, 549-558 (2009)

5.1 Introduction

Repairing damage in DNA is essential for maintaining genomic integrity. Therefore, several protein systems have evolved in order to remove DNA lesions. One of them is the ultraviolet damage endonuclease (UVDE) repair system, initially found in the yeast *Schizosaccharomyces pombe*, and described as an alternative repair system for UV-induced lesions (Sidik *et al*., 1992). The UVDE enzyme was shown to introduce a nick 5' to both of the main UV-induced lesions: cyclobutane pyrimidine dimers (CPD) and 6-4 photoproducts (6-4PP) (Bowman *et al*.,1994). Later studies showed, however, that UVDE from *S. pombe* has a much broader substrate specificity than originally thought – recognizing and nicking DNA lesions significantly different from UV induced damage such as abasic sites and small loops (Avery *et al*., 1999). UVDE enzymes have been found in both eukaryotes (e.g. *S. pombe*, *Neurospora crassa*) and in prokaryotes (e.g. *Thermus thermophilus*, *Bacillus subtilis*).

The crystal structure of the *T. thermophilus* UVDE has recently been solved to 1.55 Å resolution (Paspaleva, Thomassen *et al*., 2007). This structure shows UVDE to be a single domain TIM-barrel with extensive positive charges positioned on both sides of a 29 Å groove, which was proposed to be the DNA-binding site. At the bottom of the groove, a cluster of three metal ions was found. Furthermore, a protein pocket was identified in UVDE near Tyr6 and Asn10, into which UVDE might flip damaged base(s).

Unexplained electron density at the tip of Lys229 was observed and it was suggested that Lys229 is post-translationally modified. Here, we investigate the identity of this modification by X-ray crystallography and conclude that it is a carboxyl group. Carboxylated lysines have been observed in several proteins (Abendroth *et al*., 2002; Cha & Mobashery, 2007; Golema-Kotra *et al*., 2003; Golemi *et al*., 2001; Li *et al*. 2005; Thoden *et al*., 2001, 2003). Most often, the carboxyl group has a structural role (e.g. bridging two active site zinc atoms), but a more direct role in reaction mechanisms has also been reported (Cha & Mobashery, 2007). In addition to identifying this modification, we study, by means of X-ray crystallography and activity assays on several point mutants, why the protein uses a carboxylated lysine instead of a standard amino acid.

5.2 Materials and methods

5.2.1 Proteins

The *T.thermophilus* UVDE protein used in this study was expressed and purified as described before (Paspaleva, Thomassen *et al*., 2007). Mutants K229A, K229L and K229R were constructed by site-directed mutagenesis using PCR and purified in the same way as the wild type enzyme.

5.2.2 DNA substrates

DNA substrates used in all activity assays are 30-bp DNA containing either a CPD or a 6-4PP in the following sequence: 5′ CTCGTCAGCATCTTCATCATACAGTCAGTG 3′ with TT representing the position of the UV lesion. In case of the abasic site the same DNA sequence has been used: 5′ CTCGTCAGCATCXTCATCATACAGTCAGTG 3′, with X representing the position of the abasic site. The oligonucleotides containing CPD or 6-4PP lesions were synthesised as described (Iwai, 2006). The 30 bp substrate containing an abasic site lesion has been obtained commercially (Eurogentec, Belgium).

5.2.3 Incision assay

The DNA substrates were labelled at the 5′ side of the top strand using polynucleotide kinase as described (Verhoeven *et al*., 2002). The DNA substrates (0.2 nM) were incubated with 5 nM UVDE in 20 μl reaction mix (20 mM HEPES pH 6.5, 100 mM NaCl, 1 mM $MnCl_2$). After 15 minutes incubation at 55 °C the reaction was terminated by adding 3 μl EDTA/SDS (0.33 M EDTA, 3.3 % SDS) and 2.4 μl glycogen (4 mg/ml) followed by ethanol precipitation. The incision products were loaded on a 15 % denaturing polyacrylamide gel and visualized by irradiation of a photographic film after which they were quantified.

5.2.4 Incision of supercoiled plasmid DNA

Supercoiled plasmids pUC18 (2686 bp; 5 ng/μl; UV-irradiated at 300 J/m^2 or not UV-irradiated) and pNP228 (4686 bp; 5 ng/μl) were incubated with 10 nM UVDE (unless stated otherwise) in a 10 μl reaction mix (20 mM HEPES pH 6.5, 100 mM NaCl, 1 mM $MnCl_2$). After 15 min incubation at 55 °C the reactions were terminated by addition of 3 μl Ficoll/dyes/SDS/EDTA (0.05 M EDTA, 3 % SDS). Samples were loaded on a 0.7 % agarose gel and visualised by staining with ethidiumbromide.

5.2.5 Filter binding assay

The filter binding assays were conducted in 10 μl samples containing 0.1 μM UVDE and 8 nM of ^{32}P-labeled DNA in a reaction buffer containing 20 mM Tris pH 6.5, 100 mM NaCl, 1 mM MnCl2. Samples were incubated for 10 min at 55 °C. At the end of the incubation time 0.5 mL reaction buffer (preheated at 55 °C) was added. The mixture was poured over a nitrocellulose filter and the incubation vial was rinsed with 0.5 mL of the preheated reaction buffer. Finally the filters were washed with 0.5 mL incubation buffer. Each sample was corrected for the amount of DNA retained on a filter in the absence of protein. Binding is expressed as the percentage of the input DNA retained on the filter.

5.2.6 Mass-spectrometry

*Tth*UVDE was precipitated in 10% trichloroacetic acid (TCA) and subsequently washed with acetone. After drying, it was resuspended in 8 M urea and 0.4 M NH_4HCO_3. After 15 min incubation with dithiothreitol at 50 °C and addition of iodoacetamide, trypsin digestion was carried out for 24 h at 37 °C. The trypsinated sample was loaded on an LTQ-orbitrap mass-spectrometer and MS and MSMS was run.

5.2.7 Crystallization

The purified protein was dialysed against 1x PBS (25 mM Phosphate, 150 mM NaCl, pH 7.4) and concentrated to 3-5 mg/ml by centrifugation using an Ultrafree Filter Device (Millipore). Crystals were grown by sitting-drop vapour diffusion at 22 °C using as precipitant: 0.1 M sodium acetate buffer pH 4.4, 2 M sodium formate and 1 mM $MnCl_2$ (UVDE pH 4.4; 1 μl protein and 1 μl precipitant) or 0.5 M $(NH_4)_2SO_4$, 0.1 M sodium acetate buffer pH 6.0, 1 M Li_2SO_4 (UVDE K229L and UVDE E175A; 1 μl protein and 0.5 μl precipitant) or 0.1 M sodium acetate buffer pH 5.6, 1 M sodium formate (UVDE K229R; 1 μl protein and 1 μl precipitant). The crystals grew like diamonds (40 to 100 μm) in 1-2 days. Crystals were transferred to precipitant solution with 10 % glycerol prior to data collection.

5.2.8 Data collection and processing

Datasets were collected on beam-line ID14-2 at the European Synchrotron Radiation Facility at a wavelength of 0.933 Å. The crystals were flash-frozen and kept at 100 K during data collection. Data collection strategy was determined with the program *BEST* (Popov & Bourenkov, 2003). Reflections were indexed and integrated with *iMosflm* (Leslie, 1999). Scaling and merging was performed with *SCALA* from the *CCP4* (CCP4, 1994) suite. For data statistics, see Table 5.1.

5.2.9 Structure solution and refinement

Structure solution was performed by molecular replacement with as search model the structure of *Tth*UVDE wild type (pdb-code: 2j6v) with an unmodified lysine using *MOLREP* (Vagin & Teplyakov, 1997) from the *CCP4* suite, though for K229R, the structure of UVDE pH 4.4 was used as input directly into refinement. Manual adjustments to the model were done with *COOT* (Emsley & Cowtan, 2004). Refinement was done using *REFMAC* (Murshudov *et al.*, 1997) and water molecules were added using *ARP/wARP* (Perrakis *et al.*, 1999) and *COOT*. TLS refinement was used in refinement for structures UVDE pH 4.4, UVDE K229L and UVDE E175A. Illustrations were prepared using *ccp4mg* (Potterton *et al.*, 2004).

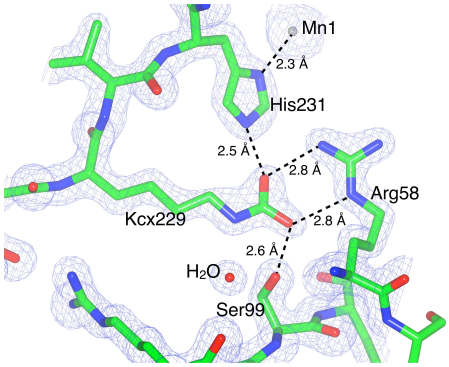

Figure 5.1: Detail of the original structure of TthUVDE wild type with electron density map (contoured at 1.5 σ) showing a carboxyl group modeled into the additional density at the tip of Lys229 (Kcx229). Carbons are depicted in green, oxygens in red, nitrogens in blue and the metal ion in grey. Distances between the oxygen atoms of the carboxyl group and neighboring residues, and the distance of His231 to Mn1 are indicated.

5.3 Results

5.3.1 Identity of the modification

The unexplained electron density observed at the tip of Lys229 is situated between Arg58, Ser99 and His231. His231 was shown to coordinate one of the metal ions (Mn1). By comparing the metal ion cluster of TthUVDE to that of endonuclease IV (Hosfield *et al.*, 1999), Mn1 is probably one of the two metal ions that together activate the water molecule (bridging Mn1 and Mn2) that serves as a nucleophile to attack the DNA phosphodiester backbone.

Several options are possible for the unexplained electron density at the tip of Lys229 among which are an acetyl group or a carboxyl group. Considering the potential hydrogen bonding to the neighboring residues (Arg58, Ser99 and His231), a carboxyl group is the more likely option (Figure 5.1). Moreover, if an acetyl group is modeled into the density, its methyl group's B-factor is 13.94 Å^2 while the Nζ and Cϵ of the lysine and the carbonyl of the acetyl group all have B-factors between 20 and 22 Å^2. If a carboxyl group is modeled into the density, all atoms of the carboxyl group and the Nζ and Cϵ of the lysine have B-factors between 18 and 21 Å^2.

To begin the investigation on the modification's identity we performed mass-spectrometry on trypsinated UVDE. In the resulting mass-spectrum, however, only a peptide fragment with an unmodified Lys229 was found. Considering the preparation of the sample (TCA precipitation and trypsin digestion), an acetyl group probably would have been observed if it had been present. In contrast, a carboxyl group on a lysine residue can be expected to fall off during the sample preparation, since at low pH the ζ-nitrogen of the lysine is protonated, which results in a loss of the carboxyl group as carbon dioxide.

To confirm the above result, we obtained UVDE crystals at low pH (expecting

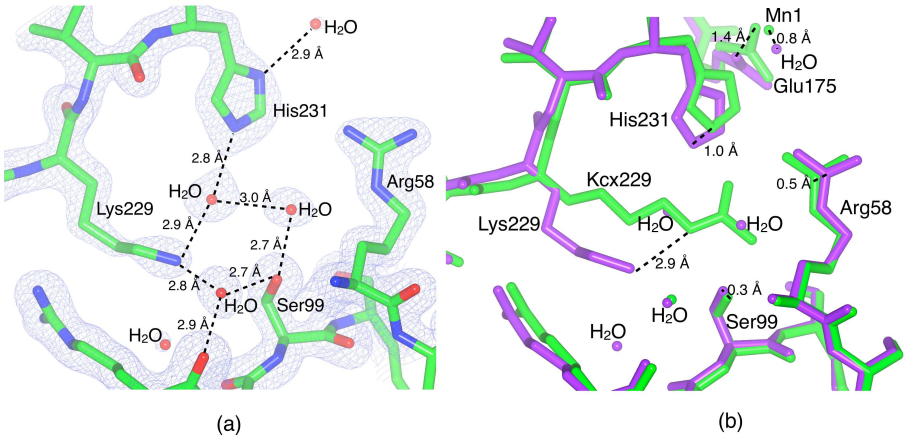

Figure 5.2: Structure of UVDE pH 4.4.
(a) Model and map (contoured at 1.7 σ) of UVDE pH 4.4, showing an unmodified Lys229. Carbons are depicted in green, oxygens in red and nitrogens in blue. Hydrogen network with Ser99, Lys229 and His231 is indicated.
(b) Detail of the superposition of the original structure of UVDE (green) and UVDE pH 4.4 (purple). Shifts of residues Arg58, Ser99, Glu175, Lys229 and His231 and of Mn1 compared to the water in UVDE pH 4.4 are indicated at the atoms where the shift was measured.

that the carboxylation will not be present, due to the acid liability of the carboxylation). For this, *Tth*UVDE was crystallized in 0.1 M acetate buffer pH 4.4, 2 M sodium formate and 1 mM $MnCl_2$ (UVDE pH 4.4) and crystallized in a different space group ($P6_122$) with different packing than the wild type structure (P1) (See Table 5.1 for the crystallographic statistics). In this structure Lys229 is indeed not modified (Figure 5.2). Two waters are at the place where the modified lysine originally was. His231 seems to have moved inwards (over 0.96 Å) in order to form hydrogen bonds with one of these waters. This shift seems to have cause a 1.36 Å shift to Glu175. In the structure of UVDE pH 4.4, no metal ions were found, probably because of protonation of the metal-coordinating histidines due to the low pH. A water molecule was found near the site (0.81 Å distance) where originally Mn1 was located.

In order to check that the above result is caused by the low pH of the crystallization condition and not due to the different space group ($P6_122$) of the previously published structure (P1), we also solved a structure of *Tth*UVDE in $P6_122$ that was crystallized at a higher pH (0.5 M $(NH_4)_2SO_4$, 0.1 M sodium acetate buffer pH 5.6, 1 M Li_2SO_4, 1 mM $MnCl_2$; results not shown). In this crystal structure, Lys229 was seen to be modified, showing that the modification can be accommodated in $P6_122$.

	UVDE pH 4.4	UVDE K229L	UVDE K229R	UVDE E175A
Data collection				
Beam line	ESRF ID14-2	ESRF ID14-2	ESRF ID14-2	ESRF ID14-2
Detector	MAR225 CCD	MAR225 CCD	MAR225 CCD	MAR225 CCD
Resolution range (Å)	28.78-1.91 (2.01-1.91)[a]	37.11-2.30 (2.42-2.30)	46.37-3.15 (3.32-3.15)	40.46-2.74 (2.89-2.74)
Multiplicity	6.4 (6.5)	11.3 (11.5)	10.7 (11.0)	4.4 (4.4)
Completeness (%)	99.2 (100)	99.2 (99.8)	95.3 (96.2)	99.3 (99.8)
R_{merge}[b]	0.048 (0.282)	0.078 (0.367)	0.134 (0.366)	0.087 (0.364)
R_{pim}[c]	0.020 (0.113)	0.025 (0.122)	0.043 (0.117)	0.048 (0.203)
I/σ	10.2 (2.0)	8.3 (2.0)	5.4 (2.0)	8.0 (2.0)
Space group	$P6_122$	$P6_122$	$P6_122$	$P6_122$
N° of molecules in ASU	1	1	1	1
Unit cell parameters a, b, c (Å)	107.14x107.14x90.83	113.39x113.39x89.39	107.11x107.11x91.12	113.38x113.38x89.08
Crystallization conditions	0.1 M Ac pH 4.4, 2 M Naformate, 1 mM $MnCl_2$	0.5 M $(NH_4)_2SO_4$, 0.1 M Ac pH 6, 1 M Li_2SO_4	0.1 M Ac pH 5.6, 1 M Naformate	0.5 M $(NH_4)_2SO_4$, 0.1 M Ac pH 6, 1 M Li_2SO_4
Refinement				
N° reflections used	22835	14615	5122	8773
Resolution range (Å)	28.71-1.91 (1.96-1.91)	34.28-2.30 (2.36-2.30)	46.18-3.15 (3.23-3.15)	35.03-2.74 (2.81-2.74)
R_{factor}	0.196 (0.286)	0.188 (0.198)	0.195 (0.232)	0.198 (0.295)
R_{free}[d]	0.244 (0.363)	0.244 (0.278)	0.249 (0.360)	0.245 (0.363)
Ramachandran statistics[e] (%)	91.1/7.6/0.8/0.4	89.8/8.9/0.4/0.8	91.5/7.7/0.0/0.9	91.9/6.8/0.9/0.4
R.m.s. dev. (bonds, Å/ angle, °)	0.019/ 1.689	0.014/ 1.545	0.006/ 0.945	0.009/ 1.320
Average atomic B-factor				
protein	25.4	30.6	32.6	22.8
Mn	NA	65.2	NA	74.3
solvent atoms	33.0	30.6	32.4	11.0
Wilson plot B-factor	25.0	35.9	54.4	47.3

[a] Values in parentheses are for the highest resolution shell

[b] $R_{merge}=\Sigma|I-<I>|/\Sigma|I|$

[c] Weiss & Hilgenfeld, 1997; Diederichs & Karplus, 1997

[d] About 5% of the reflections were used for the cross-validation set. These reflections were randomly chosen.

[e] According to the program *PROCHECK* (Laskowski *et al.*, 1993). The percentages are indicated of residues in the most favored, additionally allowed, generously allowed and disallowed regions of the Ramachandran plot, respectively.

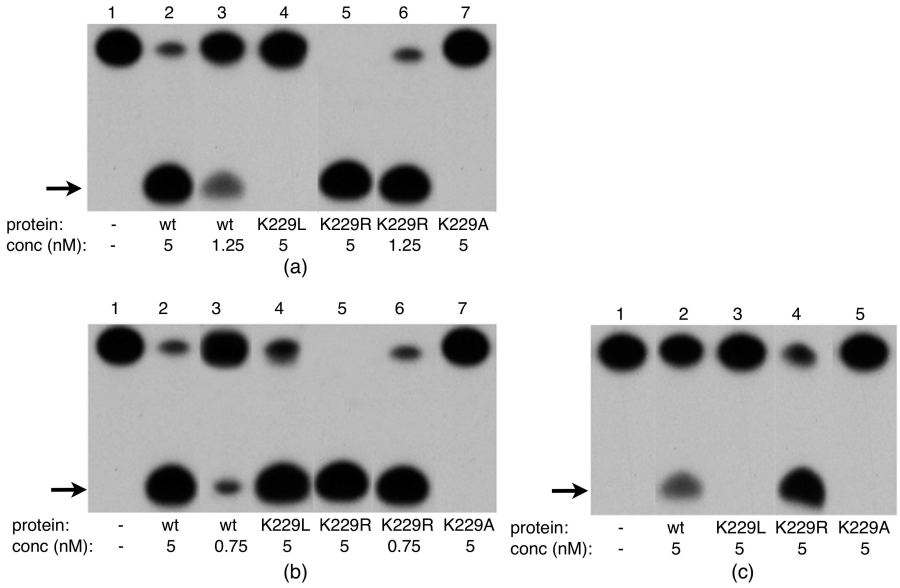

Figure 5.3: Activity of wild type and mutant UVDE proteins.
Activity assay with terminally labelled 30 bp DNA substrates containing a CPD (a), 6-4PP lesion (b), abasic site (c). The incision product is indicated with an arrow. Below the lanes is indicated which protein is used and the concentration of protein used (in nM).

5.3.2 Functional studies of Lys229 mutants

To investigate the role of the carboxylated lysine in the activity of UVDE, we mutated this residue into an alanine (UVDE K229A) and a leucine (UVDE K229L). UVDE K229A showed an extremely reduced catalytic activity on all tested DNA substrates (Figure 5.3). The K229L mutant showed severely reduced catalytic activity on the CPD (10%) and the abasic site lesion (less than 1%). On the 6-4PP the incision efficiency of this mutant is slightly reduced compared to that of the wild type (Figure 5.3). These results indicate that Lys229 has an important role in the function of *Tth*UVDE. Experiments attempting to decarboxylate UVDE (by lowering the pH) and then carboxylate it again by addition of HCO_3^- were unsuccessful, probably due to loss of activity of the protein (also in the neutral pH controls) in the many dialysis steps that were necessary in the procedure (amongst others to remove the HCO_3^- to prevent interference in the activity assay).

To investigate if the observed reduction in the incision efficiency is due to an effect on the formation of protein - DNA complexes, we tested the binding properties of UVDE K229A and UVDE K229L in a filter binding assay. As can be seen in Table 5.2 the binding capacity of both mutants is similar to the wild type. This indicates that the reduced incision activity is not due to changes in the enzymes ability to bind DNA, but an impairment in catalysis. Moreover, DNA-binding is similar either in absence and presence of Mn^{2+} (for wild type and mutants) showing that

Table 5.2: DNA binding by TthUVDE wildtype and mutants.

DNA lesion	**UVDE**	**% binding, no Mn^{2+}**	**% binding, 1 mM Mn^{2+}**
CPD	Wildtype	34 ± 2	33 ± 1.7
CPD	K229L	31 ± 0.7	31 ± 0.7
CPD	K229R	31 ± 0.5	32 ± 0.2
6-4PP	Wildtype	19 ± 3.7	20 ± 1
6-4PP	K229L	20 ± 0.5	21 ± 1.8
6-4PP	K229R	17 ± 2.8	19 ± 0.7
Abasic site	Wildtype	20 ± 2	23 ± 0.7
Abasic site	K229L	21 ± 0.7	21 ± 1.4
Abasic site	K229R	21 ± 2	21 ± 0.7
No damage	Wildtype	2.5 ± 0.1	2.8 ± 0.4
No damage	K229L	3.7 ± 0.2	3.8 ± 0.4
No damage	K229R	7.6 ± 0.7	5.8 ± 1

Mn^{2+} does not influence DNA binding.

To gain a more detailed insight into the function of the carboxylated lysine, a mutant was constructed in which Lys229 was changed to an arginine (UVDE K229R). An arginine has a positive charge like a lysine, but an arginine cannot be carboxylated. The positive side-chain of an arginine can therefore not be turned into a negative group by carboxylation.

Surprisingly, the K229R mutant was found to be active and shows very high activity on CPD (Figure 5.3(a)) and 6-4PP (Figure 5.3(b)) lesions, even higher than that of wild type. Also abasic site, which is not an optimal substrate for the *Tth*UVDE wild type (20 % incision), is cleaved by the K229R mutant with high efficiency (80 %, Figure 5.3(c)).

We also tested the activity of K229R on UV-damaged DNA in a supercoiled incision assay (Figure 5.4(a); plasmid I). In the lanes with UVDE K229R (lane 6-8), much more relaxed and even linear plasmid DNA as a result of damage-specific incision can be seen compared to the lanes with UVDE wild type (lanes 2-4). This confirms that UVDE K229R is much more active on UV damage sites than the wild type protein. UVDE K229R also has incision activity on the undamaged supercoiled plasmid DNA present in the assay (Figure 5.4(a), plasmid II, lane 8), though this activity is very small compared to the activity on UV-damaged supercoiled plasmid DNA. To investigate this activity on undamaged DNA more closely, we performed an assay with only undamaged supercoiled plasmid DNA (Figure 5.4(b)). In this assay we could see that, surprisingly, UVDE wild type has some activity on undamaged plasmid DNA (lane 2). The activity of K229R on undamaged DNA, however, was again significantly higher (lane 4).

Filter-binding assays (Table 5.2) showed that K229R has similar DNA binding properties to wild type UVDE. Thus, the higher activity of this mutant on both damaged and undamaged DNA is not caused by a change to DNA binding, but by an increased efficiency of the incision reaction.

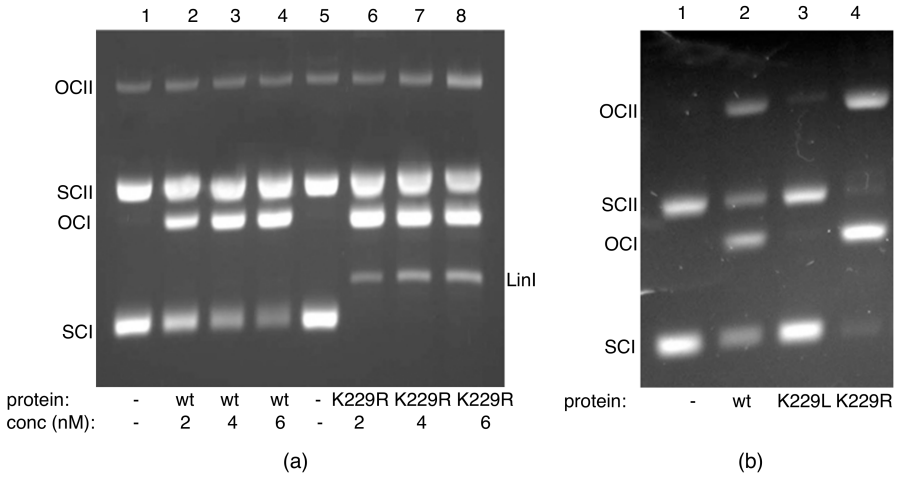

Figure 5.4: Activity of UVDE wild type and K229R on UV-irradiated and non-irradiated plasmid DNA.
(a) A mixture of UV-irradiated supercoiled DNA of pUC18 (I, 2686 bp) and non-irradiated pNP228 (II, 4686 bp) was incubated with different concentrations (as indicated) UVDE wild type and K229R mutant for 15 min. The positions of the supercoiled (sc), open circle (oc) and linear (lin) forms of the plasmids are indicated.
(b) A mixture of undamaged supercoiled DNA of pUC18 (I, 2686 bp) and pNP228 (II, 4686 bp) was incubated with 10 nM UVDE wild type and mutant proteins as indicated. The positions of the supercoiled (sc) and open circle (oc) forms of the plasmids are indicated.

5.3.3 Structural studies of the Lys229 mutants

To obtain a structural basis for the above results, the mutants K229L, K229R and E175A were crystallized and their structures were solved by molecular replacement (see Table 5.1 for statistics). In all three mutants, the overall structure is the same as wild type.

In the structure of K229L, a water molecule is present at the place of the carboxylated lysine (Figure 5.5). Only one of the three metal ions is present: Mn3. This is the metal ion near residues His244 and His203 and it corresponds to Zn3 in endonuclease IV (Hosfield *et al.*, 1999). The two metal ions (Mn1 and Mn2) that are proposed to activate the catalytic water molecule (based on comparison to EndoIV) are absent. These results suggest a correlation between the presence of the metal ions and the carboxyl group.

To study whether the carboxyl group can still be present if there are no metal ions, the previously constructed mutant (Paspaleva *et al.*, 2007) E175A was studied in detail. Glu175 bridges Mn1 and Mn2. The mutant E175A was previously found to have no observable activity on DNA containing CPD and a severely reduced activity on DNA containing 6-4PP (5 % activity left). In the structure of this mutant, Mn1 and Mn2 were not observed, as expected. Instead of these ions, there is one

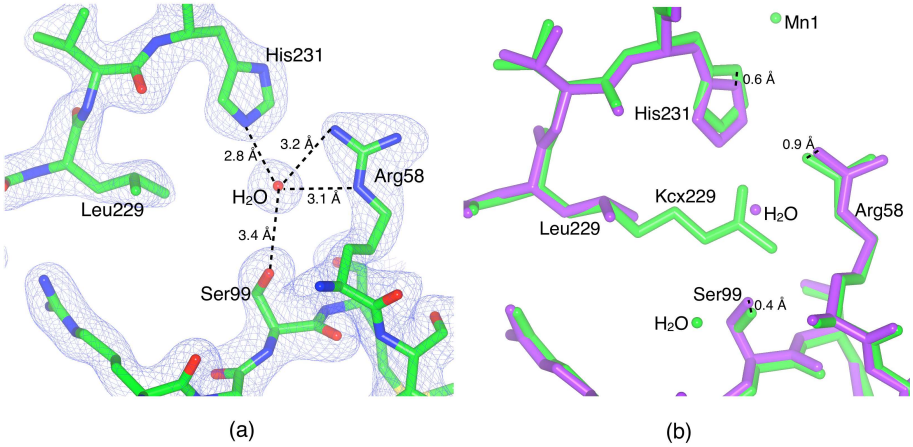

Figure 5.5: Structure of UVDE K229L.
(a) Detail of the model and the map of UVDE K229L showing the environment of Leu229. Carbons are depicted in green, oxygens in red and nitrogens in blue. Distances of the water molecule to neighboring residues are indicated.
(b) Detail of a superposition of UVDE K229L (in purple) with the original structure of UVDE (in green). The slight shifts of residues Arg58, Ser99 and His231 are indicated.

water molecule (Figure 5.6). The metal ion near His244 and His203, Mn3, is present. Most importantly, the modification on Lys229 was still observed. This result shows that the modification can be present in the absence of the Mn1 and Mn2.

Together, the above results suggest that the carboxyl group can be present on Lys229 in absence of Mn1 (as in E175A), but that perhaps this metal ion cannot bind stably in the absence of the carboxyl group (as in structure K229L).

In K229R, Arg229 has moved compared to the carboxylated lysine (Figure 5.7(a) and (b)). A water molecule was found at the place where the tip of the carboxylated lysine originally was. Significant shifts are seen for the side-chains of His231 (2.5 Å) and Glu269 (3.2 Å), which have moved outwards away from Arg229. Indeed, His231 moves away from the positive charge of Arg229 and Glu269 moves accordingly. It should be noted that UVDE K229R is actually a double mutant (K229R L133I) due to a misincorporation during the PCR reaction. No influence of the second mutation on the structure of the protein could be seen (The shift of the Cα of residue 133 is only 0.34 Å). Furthermore, Leu133 is a non-conserved residue far away from the DNA binding site, leucine is very similar to isoleucine, and, in fact, some bacterial UVDE have isoleucine in this position. Thus, the mutation L133I is likely to have no impact on the behavior of K229R.

Metal ions were not observed in the K229R structure, which is probably caused by the low pH of the crystallization condition (below the pKa of histidine side-chains). Although two of the metal-coordinating residues have shifted considerably compared to the wild type structure (His231, Glu269), the coordination en-

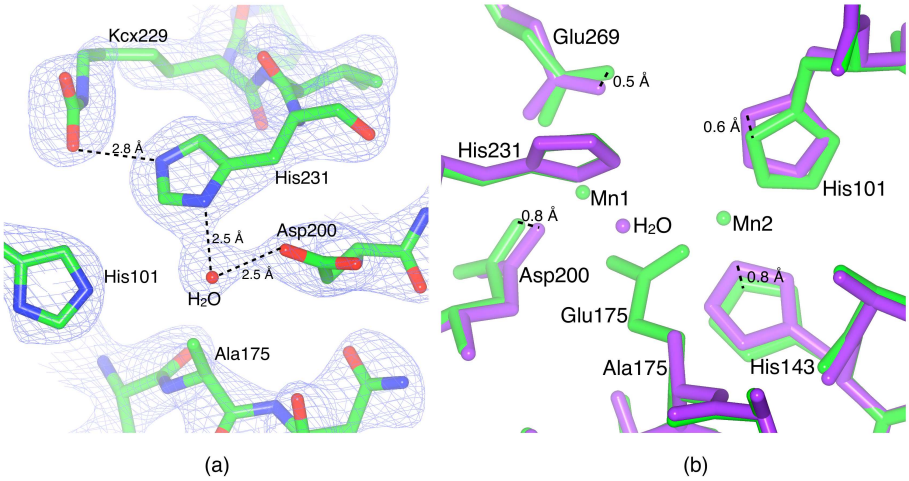

Figure 5.6: Structure of UVDE E175A.
(a) Detail of the map (contoured at 1.25 σ) and the model of UVDE E175A showing the environment of Ala175 and the presence of a carboxylated lysine. Carbons are depicted in green, oxygens in red and nitrogens in blue. Distance between an oxygen atom of Kcx229 and His231 is indicated as well as the distances to the water taking the place of Mn1 and Mn2 to neighboring residues.
(b) Detail of the superposition of the original structure of UVDE (in green) and UVDE E175A (in purple) showing the surroundings of residue Glu175/ Ala175.

vironments for the three metal ions seem to be intact. Thus, the metals probably can bind stably in all three sites (as the activity also suggests) if the pH permits it, although Mn1 would have to shift over about 1 Å to get into a proper coordination environment.

To get insight into the possible influence of the shifts of His231 and Glu269 on the activity of UVDE K229R, DNA was modeled into K229R based on the crystal structure of endonuclease IV with DNA (Hosfield *et al.*, 1999). In this model, it can be seen that the observed shifts are near the part of the DNA where the damage is expected to be located. In contrast to wild type, in K229R the shifted Glu269 clashes with the abasic site (damaged site) of this model (see Figure 5.7(c)), suggesting that the part of DNA near the damage might have to bind slightly different in UVDE K229R compared to wild type.

5.4 Discussion

The results from mass-spectrometry and crystallization at low pH showed that the modification we initially observed on Lys229 is (acid) labile. Together with the observed hydrogen bonding pattern of the modification in the original structure, we conclude that the modification is a carboxyl group. Carboxylated lysines have

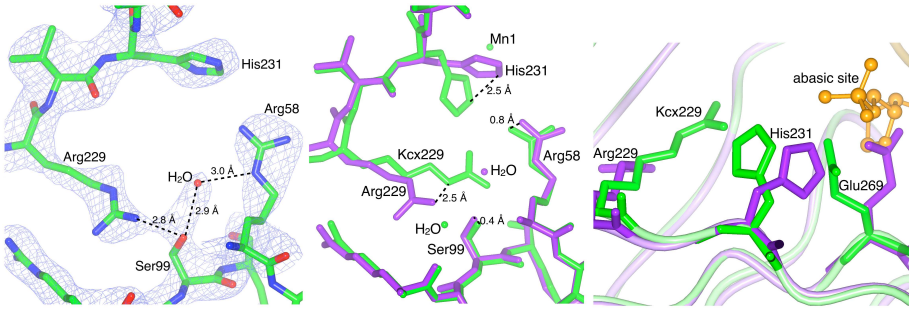

Figure 5.7: Structure of UVDE K229R and model with DNA.
(a) Model and map (contoured at 1.5 σ) of UVDE K229R. Carbons are depicted in green, oxygens in red and nitrogens in blue. Distances between a water molecule and Arg58 and Ser99 are indicated.
(b) Detail of the superposition of the original structure of UVDE (in green) and UVDE K229R (in purple) showing the position of Arg229 compared to the carboxylated lysine and the large shift in the position of His231. Also the shifts in the position of Arg58 and Ser99 are indicated.
(c) Detail of superposition of the original structure of UVDE (in green) and the structure of UVDE K229R (in purple) in which DNA with an abasic site was modeled (orange) based on a comparison to the structure of endonuclease IV with damaged DNA. Arg229/ Kcx229, His231 and Glu269 are depicted in cylinder representation, the abasic site of the DNA is depicted in ball-and-stick representation while the rest of the protein and the DNA is depicted in ribbon representation.

been observed before in prokaryotic proteins and carboxylation of a lysine does not require an enzyme, but can happen in solution at a suitable position in a protein due to dissolved carbon dioxide (Li *et al.*, 2005). Therefore, it is a likely possibility for a protein (*Tth*UVDE) overexpressed in a foreign host (*E. coli*). With activity assays and structural analysis on mutants, we conclude the carboxyl group might be involved in the stable binding of Mn1 by donating some of its negative charge to the His231. Reduced stability of metal binding explains the reduced activity of K229A and K229L. For K229L, this reduction in activity is stronger on CPD than on 6-4PP, possibly because the presence of DNA with a 6-4PP might facilitate metal-binding and thus allow activity on 6-4PP. Notably, also UVDE E175A (in which also reduced binding of Mn1 is involved) has a more severe phenotype on CPD than on 6-4PP (Paspaleva *et al.*, 2007).

The mutant K229R, however, does not have a negative charge near His231. Quite the contrary: the protein has a positive charge at position 229 (an arginine). Still, the mutant is active on CPD, 6-4PP and abasic sites with a higher incision efficiency than wild type on these substrates. This increased activity of K229R might be explained by considering a potential mechanism for UVDE where, the protein first recognizes a general distortion in the DNA and binds to it (proposed previously based on the broad substrate specificity of SpUVDE (Avery *et al.*, 1999)). K229R

performs this step similar to UVDE wild type, since overall DNA binding was seen not to be affected by the mutation. Then, the DNA gets into the right conformation for incision and the incision takes place. This step is performed more efficiently in K229R than wild type. Structural changes were observed in the active site of K229R and the induced changes block the previously proposed damage-binding pocket. However, the structure of UVDE in complex with DNA is needed to clarify the relevance of the latter observation. The structural changes in the active site of K229R may cause the portion of the DNA near the damage to bind slightly different in the active site, perhaps in a conformation more favorable for incision.

The results on mutant UVDE K229R raise an interesting question. Why did *T. thermophilus* not use an arginine at position 229 if this protein is more efficient than wild type? An answer to this may be found in the activity of UVDE K229R on undamaged DNA. Both UVDE wild type and K229R were found to incise undamaged DNA, which has not been reported for UVDE from *S. pombe*. Apparently, *Tth*UVDE has more difficulties in discriminating damaged and undamaged DNA. The high temperature used for incision assays with this thermophile, might play a role in the incision of undamaged DNA, since the DNA can be more readily distorted at higher temperature. Interestingly, we found that the activity on undamaged DNA was much higher for K229R than for UVDE wild type. Perhaps Arg229 is not favorable for *T. thermophilus* because the incision on undamaged DNA in the cell might be too high.

UVDE is present in several organisms, both prokaryotic and eukaryotic. Is carboxylation a general phenomenon in all these proteins? Two of the residues near the carboxylated lysine, Arg58 and His231, are fully conserved in all UVDE proteins (Golemi *et al*., 2001). In eukaryotic UVDEs, Lys229 itself is also conserved and these proteins have a threonine instead of a serine at the position corresponding to Ser99. Thus, these proteins might have a similar site at this position and therefore, they might be carboxylated like *Tth*UVDE.

In the prokaryotic UVDEs, however, Lys229 is only partially conserved. Most UVDE proteins have a lysine at this position as well as a serine at the position corresponding to Ser99 and thus might be carboxylated like *Tth*UVDE. A few UVDEs, however, have a leucine, isoleucine, methionine, glutamic acid, threonine or a valine at the position corresponding to Lys229. Thus, these UVDE homologues cannot have the same modification.

We have shown here that a leucine at position 229 results in an inactive enzyme for *Tth*UVDE while some of the other UVDEs do have a leucine at that position, such as *Desulfotelea psychrophila* UVDE (Goosen & Moolenaar, 2008). Perhaps in these UVDEs the environment near His231 is such that a negative charge is provided to His231 by other residues nearby so that stable metal-binding is assured in these proteins as well. Indeed, the UVDEs with a leucine at position 229 all have a glutamic acid or an aspartic acid at the position of Met267, which is spatially close to His231 (the side chain of glutamic acid is 2.5 Å away from the side-chain of His231 if glutamic acid is modeled at this position in the structure of *Tth*UVDE). Such a negatively charged residue is not present in the UVDEs that do have a lysine at position 229; those all have a methionine at that position. Moreover, most of the other

UVDEs that do not have a lysine at the position of Lys229, also have a glutamic acid near or at position 229. Thus, also these UVDEs might have a negative charge near His231 for assuring proper metal binding, though no definite conclusions can be made in absence of structural data on these proteins.

In conclusion, we think that UVDE from *T.thermophilus* is carboxylated at Lys229. The carboxyl group might be required for stable metal-binding and perhaps also for preventing an unfavorably high incision of undamaged DNA.

Note

Coordinates and structure factor amplitudes have been deposited in the Protein Data Bank under ID code 3bzg for UVDE pH 4.4, 3bzj for UVDE K229L, 3c0l for UVDE K229R and 3c0q for UVDE E175A.

6

Unravelling UVDE's uncanny ability to recognize and incise different types of damaged DNA

UV damage endonuclease (UVDE) is a DNA repair enzyme that can recognize and incise a diverse set of DNA lesions including UV-induced DNA damage and DNA containing abasic sites. In this chapter, we present the structure of UVDE from *Sulfolobus acidocaldarius* (*Sac*UVDE) and a pre-catalytic structure of *Sac*UVDE in complex with DNA containing a 6-4 photoproduct. These structures show that UVDE has an intriguing "dual flip" mechanism: the two purines opposite to the damaged pyrimidine bases are flipped into a dipurine-specific pocket, whilst the damaged bases are also flipped into a pocket. In contrast to UVDEs from other species, *Sac*UVDE shows a marked preference for DNA substrates containing 6-4 photoproducts compared to cyclobutane pyrimidine dimers: biochemical assays and mutagenic studies show that the flexibility of the damage binding pocket and positive charges on both sides of the substrate binding groove are main contributors to the broad substrate specificity for UVDE in general.

E.M. Meulenbroek, C. Peron Cane, I. Jala, S. Iwai, G.F. Moolenaar, N. Goosen, N.S. Pannu, *UV damage endonuclease employs a novel dual-dinucleotide flipping mechanism to recognize and incise different types of damaged DNA* (to be submitted)

6.1 Introduction

The UV damage endonuclease (UVDE) repair pathway, present in some prokaryotes and lower eukaryotes, was first identified in *S. pombe* where Nucleotide Excision Repair deletion mutants were seen to be only moderately UV-sensitive (McCready *et al*., 1993). Its central enzyme, UVDE, was found to be a DNA endonuclease that not only recognizes and incises DNA 5′ to cyclobutane pyrimidine dimers (CPDs) and 6-4 photoproducts (6-4PPs) (Bowman *et al*., 1994), but also non-UV-induced DNA damage such as abasic sites, nicks and gaps (Avery *et al*., 1999). The activity on abasic sites, nicks and gaps was seen to depend on the presence of neighboring pyrimidines, suggesting that UVDE's active site is most optimal for binding distorted pyrimidines (Paspaleva *et al*., 2009). Later, UVDE homologues were discovered in *Neurospora crassa* and from *Bacillus subtilis* that incise at least CPD, 6-4PP and abasic site very efficiently (Kanno *et al*., 1999; unpublished results) and UVDE from *Thermus thermophilus* that efficiently cleaves CPDs and 6-4PPs while incising abasic sites only with moderate efficiency (Paspaleva, Thomassen *et al*., 2007).

Insights into the mechanism of UVDE came from the crystal structure of UVDE from *Thermus thermophilus* (Paspaleva, Thomassen *et al*., 2007). *Tth*UVDE has a TIM-barrel fold with a large groove with positive charges on the edges where DNA was proposed to bind based on the structural similarity to Endonuclease IV (Hosfield *et al*., 1999). The active site containing three metal ions is located on the bottom of this groove and two conserved residues (Gln and Tyr; together called the 'probing finger') were proposed to aid in flipping out the damaged bases from the DNA helix. Thus, the structure suggested plausible mechanisms for DNA binding and three-metal-ion catalysis, but could not explain UVDE's wide substrate specificity.

To explain the broad specificity, we have determined the structures of UVDE *Sulfolobus acidocaldarius* (*Sac*UVDE) on its own to 1.5 Å resolution and in a pre-catalytic complex with DNA containing a 6-4 photoproduct determined to 2.7 Å. The structures show that UVDE recognizes damaged pyrimidine dimers by flipping the two bases opposite to the damage into a dipurine-specific pocket and by flipping the damage itself into a flexible damage-binding pocket. Positive charges on both sides of the DNA binding-groove probably aid in binding different damaged DNA substrates.

6.2 Materials and methods

6.2.1 Cloning

Genomic DNA was isolated from *S. acidocaldarius* by resuspension of cells in TEN-buffer (20 mM Tris pH 8, 1 mM EDTA, 100 mM NaCl), followed by lysis in TENST-buffer (20 mM Tris pH 8, 1 mM EDTA, 100 mM NaCl, 1.6 % sarcosyl, 0.12 % Triton) and phenol/ chloroform extraction. The gene for UVDE was amplified using the primers 5′ ATTAATAACATATGAGAGTAGGTTACGTATCCAC 3′ and 5′ TAGGATCCATTAATCCAGTTTGTTTAACTCCTTTAAC 3′. Subsequently, it was cloned into the pETUVDEΔ228 vector (Paspaleva *et al*., 2009) using *Nde*I and *Bam*HI, re-

Table 6.1: 30-mer DNA substrates used in this study.

No damage	5′ CTCGTCAGCATCTTCATCATACAGTCAGTG 3′ 3′ GAGCAGTCGTAGAAGTAGTATGTCAGTCAC 5′
CPD and 6-4PP	5′ CTCGTCAGCATC**TT**CATCATACAGTCAGTG 3′ 3′ GAGCAGTCGTAGAAGTAGTATGTCAGTCAC 5′
Abasic site	5′ CTCGTCAGCATC**X**TCATCATACAGTCAGTG 3′ 3′ GAGCAGTCGTAGAAGTAGTATGTCAGTCAC 5′

The positions of the CPD (**TT**), (6-4)PP (**TT**) and AP site (**X**) are indicated in bold.

sulting in the gene for SacUVDE with a N-terminal 10x His-tag and factor Xa cleavage site. Mutants of *Sac*UVDE were created by PCR and cloned into the same vector with *Nde*I and *Bam*HI. All constructs were verified by sequencing.

6.2.2 Expression and purification

The plasmid with the gene for *Sac*UVDE was transformed to *E.coli* BL21(DE3)-codon+ and overexpressed for 2 hours at 37°C after induction by 0.5 mM IPTG. After harvesting, the pellet was resuspended in Ni buffer A (20 mM Tris pH 7.5, 500 mM NaCl, 50 mM imidazole, 8 mM β-mercaptoethanol and 10 % glycerol) and the cells were lysed by sonication. The lysate was spun down at 37.000 rpm (100.000 g) for 30 min and the soluble fraction was loaded on a His-trap column (GE healthcare) equilibrated with Ni buffer A. The column was washed with 20 column volumes Ni buffer A and the protein was then eluted with a 60 column volumes gradient to Ni buffer B (20 mM Tris pH 7.5, 500 mM NaCl, 500 mM imidazole, 8 mM β-mercaptoethanol and 10 % glycerol). Fractions containing *Sac*UVDE were dialyzed to 20 mM Tris pH 8 and were then loaded on a HiTrap Q column (GE healthcare) equilibrated with Q buffer A (20 mM Tris pH 8 and 10 % glycerol). The column was washed with 10 column volumes Q buffer A and eluted with a 60 column volumes gradient to Q buffer B (20 mM Tris pH 8, 1 M NaCl and 10 % glycerol).

For crystallization, the purification protocol was adapted: size exclusion (Superdex 200, GE healthcare) was performed in GF buffer (20 mM HEPES 7.2, 200 mM NaCl, 5 mM DTT) after the Ni purification instead of ion exchange. The protein from either protocol was found to be more than 95 % pure as judged from SDS PAGE. All purification steps were performed at 4 °C.

6.2.3 DNA substrates

The DNA substrates used in this study can be seen in Table 6.1. The oligos containing CPD or 6-4PP were synthesized as described in Iwai, 2006. The top strands of the DNA substrates were 5′ radioactively labelled using polynucleotide kinase as reported previously (Verhoeven *et al.*, 2002).

6.2.4 Incision assays

Labelled DNA substrates (1 nM) were incubated for 15 min at 55 °C with the indicated amount of UVDE (in the range 0.05 to 50 nM) in 20 mM HEPES pH 6.5, 100 mM NaCl and 1 mM $MnCl_2$ in a reaction mix of 20 μl. The reaction was then stopped by adding 3 μl stop mix (0.33 M EDTA, 3.3 % SDS), after which 2.4 μl 4 mg/ml glycogen was added and the DNA was precipitated by ethanol. Samples were loaded on a 15 % denaturing polyacrylamide gel and visualized by autoradiography. For kinetics incision assays, a mix was prepared of buffer, cofactor, protein (25 nM) and DNA and put at 55 °C. At the indicated time points, samples were taken out and the reaction was stopped in these samples.

6.2.5 Bandshift assays

Labelled DNA substrates (0.1 nM) were incubated for 10 min on ice with 660 nM UVDE in 20 mM Tris pH 6.5, 100 mM NaCl and 1 mM $MnCl_2$ in a reaction mix of 10 μl. Samples were loaded on a 6 % native gel, which was run at 4 °C in 1x TBE. The gel was dried and the result was visualized by autoradiography.

6.2.6 Filter-binding assays

Labelled 50-mer DNA substrates (4 nM) were incubated for 7 min with 50 nM UVDE at 55 °C in a reaction buffer containing 20 mM Tris pH 6.5, 100 mM NaCl and 1 mM $MnCl_2$ in a reaction mix of 20 μl. After incubation, 0.1 ml reaction buffer was added and the mixture was poured over a nitrocellulose filter. The filters were washed with 0.2 ml reaction buffer. The amount of DNA retained on the filter (by binding to UVDE) was determined using a scintillation counter. Each sample was corrected for the amount of DNA retained on a filter in the absence of protein. Binding is expressed as the percentage of the DNA retained on the filter divided by the input DNA.

6.2.7 Crystallization

*Sac*UVDE was concentrated to 3-5 mg/ml with a 3 kDa MWCO centrifugal filter unit (Millipore). Crystallization trials were performed using the sitting-drop vapour diffusion method and the JCSG+ and PACT (Qiagen) screens. *Sac*UVDE crystals were obtained in 20 % PEG3350 with 0.2 M NH_4Cl or 0.2 M NaI. The conditions were optimized by a systematic screen around these conditions and the largest crystals were grown in 14 to 28 % PEG3350 with 0.15 to 0.3 M NH_4Cl.

The damaged strand of the oligo containing 6-4PP for crystallization was synthesized as previously described (Iwai, 2006) while the undamaged strand was purchased from Eurogentec, Belgium. The sequence of the oligo was: 5' GCGTCC**TT**GACGACG 3' with the site of the damage printed in bold. The two strands were hybridized by heating to 80 °C for two minutes in 20 mM Tris pH 7 and then allowed to slowly cool down to room temperature. For co-crystallization, protein (at 0.11 mM) and DNA (at 0.21 mM) were incubated on ice for 15 minutes,

after which sitting-drop vapour diffusion experiments were set up in the NucPro screen (Jena Biosciences). Damaged DNA:protein complex crystals appeared after several days in 30 % PEG2000-MME, 100 mM acetate buffer pH 4.6, and 200 mM $(NH_4)_2SO_4$.

6.2.8 Data collection

Crystals were caught with SPINE sample loops and put in cryoprotectant solution (precipitant solution with 10-15 % glycerol) and flash-frozen. Data were collected at the European Synchrotron Radiation Facility (Grenoble, France). 180 images were collected with an oscillation angle of 1.0° with transmission of 13 % and exposure time of 0.5 s per frame at 0.9393 Å at 100 K on beamline ID14-4 for the apoprotein crystals. For the DNA:protein complex crystals, 150 images were collected with an oscillation angle of 1.0° and an exposure time of 25 s per frame at 0.934 Å at 100 K on beamline ID14-1. The images were processed with *iMosflm* (Leslie, 2011). Scaling and merging were done with *SCALA* (Evans, 2006) from the *CCP4* suite (Winn *et al*., 2011). For the apoprotein structure, two datasets (from two different crystals) were merged to yield the final dataset used for determining the structure. Data collection statistics are shown in Table 6.2.

6.2.9 Structure solution and refinement

The phase problem for the apoprotein was solved by molecular replacement using the structure of *Tth*UVDE (PDB entry 2j6v) as a search model. The model was automatically rebuilt using *ArpWarp* (Perrakis *et al*., 1999) and refined with *Refmac* (Murshudov *et al*., 2011). Manual fitting was performed using *Coot* (Emsley *et al*., 2010). For the DNA:protein complex, the phase problem was solved by molecular replacement using the structure of *Sac*UVDE as a search model. Clear difference density was visible for the DNA (Figure 6.1(e)) and the DNA was built in manually in *Coot*. The model was refined with *Refmac* and further manual fitting was also performed using *Coot*. The final R-factor and R_{free} for uncomplexed *Sac*UVDE were 0.177 and 0.214 respectively and 0.201 and 0.271 for *Sac*UVDE in complex with DNA. Refinement statistics are shown in Table 6.2.

Superpositions were done with the ssm function in *Coot*. Root-mean-square deviation calculations were done using *Theseus* (Theobald & Wuttke, 2006). Structure-based sequence alignment was performed using the program *VAST* (Thompson *et al*., 2009). Atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank (accession code 3tc3 for the apoprotein and 4gle for the cocrystal structure). All figures were made with *CCP4MG* (Potterton *et al*., 2004).

6.3 Results

6.3.1 Overall structure of *Sac*UVDE with and without damaged DNA

To gain insight into UVDE's damage recognition and incision mechanism, a co-crystal structure of UVDE with damaged DNA is needed. Crystallization of *Tth*UVDE and *Bsu*UVDE with different damage-containing oligos (abasic site, CPD, different lengths and ends of oligos) was unsuccessful: only crystals of either protein or DNA alone were obtained. This is probably due to the relatively low affinity of UVDE for damaged DNA and the high propensity of *Tth*UVDE to crystallize on its own. Luckily, we found a UVDE homologue that more strongly interacts with DNA: UVDE from *Sulfolobus acidocaldarius*, an aerobic thermoacidophilic crenarchaeon. This homologue is more specific for 6-4PPs and also binds slightly stronger to this damage than *Tth*UVDE (see section 6.3.3). Crystallization of this protein with and without DNA were successful: a crystal of *Sac*UVDE diffracted to 1.5 Å and a crystal of *Sac*UVDE in complex with DNA containing 6-4 photoproduct, diffracted to 2.7 Å (statistics are shown in Table 6.2). The quality of the map of both structures is shown in Figure 6.1(a) and (c). Clear difference density for DNA was seen in the complex after molecular replacement with the apoprotein, into which the damaged DNA could be built manually. An OMIT map for the DNA can be seen in Figure 6.1(e). For uncomplexed *Sac*UVDE, two very similar chains are present in the asymmetric unit (root mean square deviation of the Cα atoms is 0.064 Å), while the DNA:protein complex crystals contain only one chain in the asymmetric unit.

*Sac*UVDE has a TIM-barrel fold (Figure 6.1(b)) and the *Sac*UVDE backbone structure is very similar to *Tth*UVDE: the root mean square deviation of the Cα atoms is only 0.651 Å. The sequence identity between these two proteins is 38 %. Figure 6.2 shows a structure-based sequence alignment. Density is only present for one metal in the *Sac*UVDE structure; this ion is located near (0.42 Å) the position of metal ion Mn1 in *Tth*UVDE (the most stably bound metal ion). During refinement, the occupancy for this metal ion was set to unity, but the B-factor of this ion refined to a relatively high value (50.78), suggesting either high mobility or low occupancy. The metal coordinating residues in *Sac*UVDE are at similar positions as those of all three metal ions in *Tth*UVDE. Addition of manganese is needed for activity in incision assays and therefore it is very likely that active *Sac*UVDE also uses a three-metal ion catalysis. Another notable difference between the structures of *Sac*UVDE and *Tth*UVDE, is that *Sac*UVDE has one extra α-helix (α8; Figure 6.1(b)) at its C-terminus. Thus, *Sac*UVDE does have a complete $\alpha 8\beta 8$ TIM-barrel, like the majority of UVDEs, in contrast to *Tth*UVDE.

The overall fold of *Sac*UVDE in the complex structure with DNA containing a 6-4 photoproduct is very similar to that of the apoprotein; the root mean square deviation of the Cα atoms is only 0.11 Å and thus hardly any changes occur upon DNA binding (Figure 6.1(d)). In the DNA:protein complex structure, the DNA is bound in the previously predicted DNA-binding groove. The DNA makes a bend of around 90 ° (Figure 6.1(f)), like in the related structure of Endonuclease IV with DNA containing an abasic site (Hosfield *et al.*, 1999). Numerous interactions of side-

Table 6.2: Data collection and refinement statistics.

Data collection		
	*Sac*UVDE	*Sac*UVDE with 6-4PP
Space group	P1	$C222_1$
Cell dimensions		
a, b, c (Å)	42.08 x 53.59 x 77.39	57.20 x 112.51 x 153.85
α, β, γ (°)	102.09, 93.02, 111.76	90.00, 90.00, 90.00
Resolution (Å)	46.05-1.50 (1.58-1.50)[a]	52.83-2.70 (2.85-2.70)[a]
Wilson plot B-factor	17.5	64.7
R_{merge}	0.099 (0.520)	0.121 (0.832)
I/σI	9.3 (1.9)	11.1 (1.9)
Completeness (%)	97.3 (94.0)	99.9 (99.2)
Redundancy	3.4 (2.4)	5.9 (5.2)
Total n° observations	317749	82168
N° unique reflections	94666	14041
Refinement		
Resolution (Å)	46.05-1.50	52.89-2.70
N° reflections	89933	13316
N° molecules in ASU	2	1
R_{work}/ R_{free}	0.177/ 0.214	0.191/ 0.262
N° atoms		
Protein	4706	2353
Metal ions	2	0
Water	579	4
DNA	0	609
Other ions	0	10
B-factors		
Protein	20.37	55.28
Metal ions	50.78	NA
Water	31.93	34.14
DNA	NA	67.83
Other ions	NA	60.77
R.m.s. deviations		
Bond lengths (Å)	0.026	0.011
Bond angles (°)	2.21	1.57
N° TLS bodies	2	NA
Ramachandran favored[b]	96.87 %	92.8 %
Ramachandran outliers[b]	0.00 %	0.69 %
Rotamer outliers[b]	1.48 %	2.95 %

[a] Values in parentheses are for the highest resolution shell.
[b] As determined by *Molprobity* (Davis *et al*, 2007).

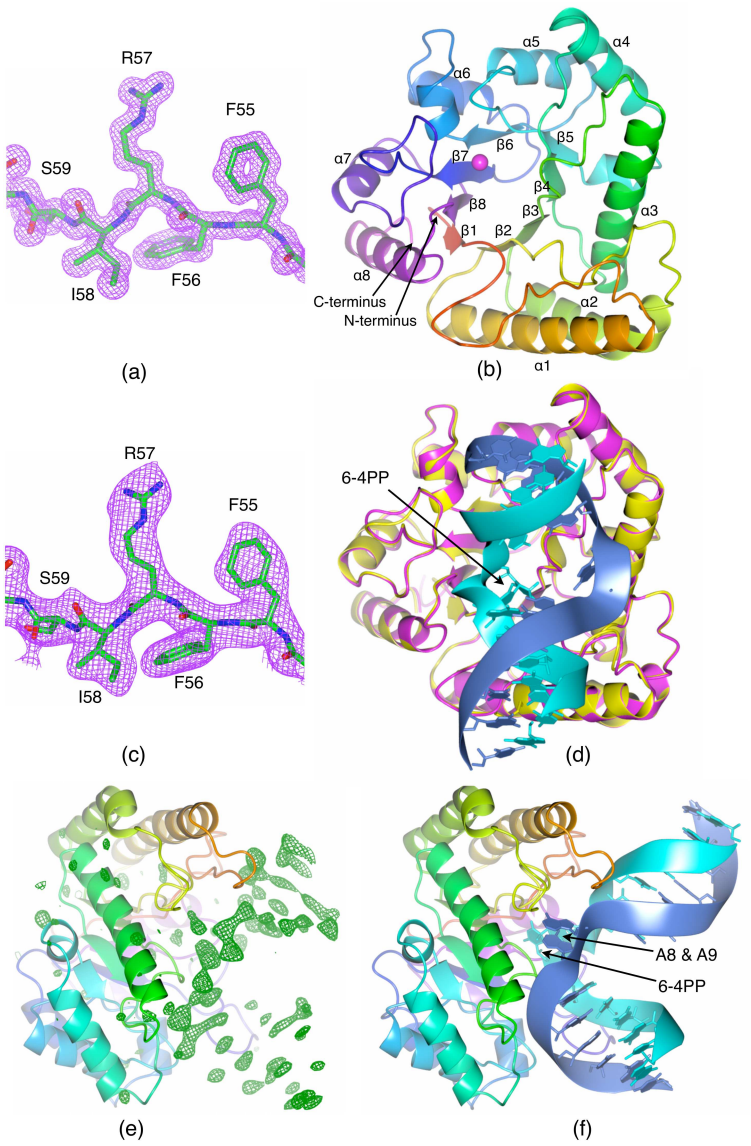

Figure 6.1: Overall structure of SacUVDE with and without DNA.
(a) Representative part of the electron density map of SacUVDE.
(b) Overall fold of SacUVDE, which has a TIM-barrel fold. The metal ion is shown in magenta.
(c) Representative part of the electron density map of SacUVDE in complex with DNA.
(d) Superposition of SacUVDE with (magenta) and without (yellow) DNA (blue and cyan), showing that the two structures are very similar.
(e) Omit map of the SacUVDE cocrystal with DNA showing clear positive difference density for the DNA (contoured at 3 σ).
(f) Overall fold of the SacUVDE-DNA complex, showing the 90 $^\circ$ bend in the DNA.

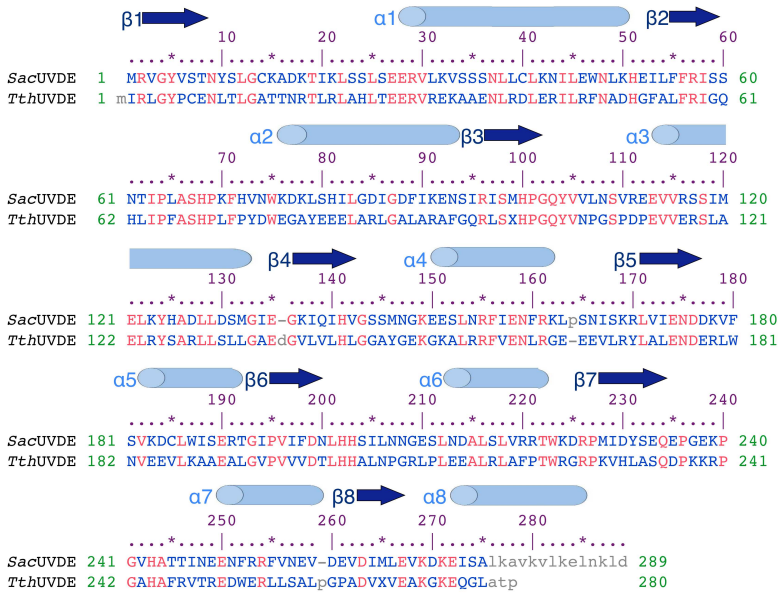

Figure 6.2: Structure-based sequence alignment of SacUVDE and TthUVDE. Secondary structure elements are indicated with cylinders and arrows for α-helices and β-strands respectively. Identical residues are in red, variable residues in blue, and unaligned residues in grey.

chains or backbone amides of the protein mostly with the DNA phosphates (and some with the base or the deoxyribose ring) hold the DNA in the DNA binding groove (Figure 6.3). The residues responsible for these interactions are partially or fully conserved in UVDE from different organisms, hence our structure is of general relevance for the understanding of the mode of action of the entire UVDE family. In the structure of *Sac*UVDE in complex with damaged DNA, no metal ions can be seen and thus the DNA is not incised and the complex is in a pre-incision state. The absence of metals in the structure is probably an effect of the crystallization condition. The position of the metal ions can be estimated by superposition with the crystal structure of *Tth*UVDE that contains all three metals. In the DNA:protein complex structure, the scissile P-O bond is still several Å away from the correct position for cleavage, but the positive charge of the metal ions is likely required to draw the scissile bond inwards to the correct position, allowing incision to take place.

6.3.2 Recognition of pyrimidine-dimer lesions by UVDE

Strikingly, the two bases opposite to the damage (nucleotides A8 and A9) are flipped from the DNA helix into a "undamaged-bases-binding" pocket of UVDE (Figure 6.4(a)). In this pocket, the conserved Tyr104 residue of the probing fin-

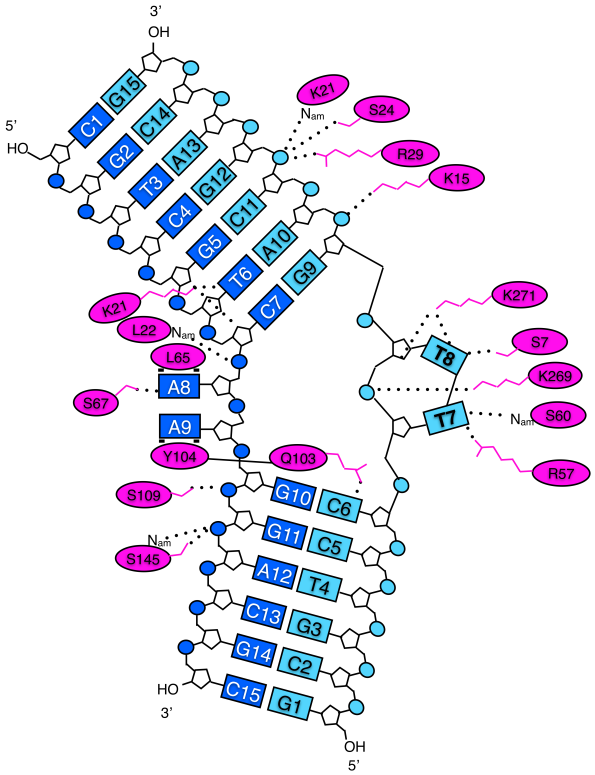

Figure 6.3: Schematic representation of UVDE-DNA interactions with the undamaged DNA strand in blue, the damaged strand in cyan and the protein in magenta. In three dimensions, the damaged bases T7 and T8 and the undamaged bases A8 and A9 are actually below the plane of the figure and the residues Q103 and Y104 are inserting in the helix.

ger forms stacking interactions with the base of A9. Leu65 aids in creating a hydrophobic environment on the other side of this pocket near A8. In other UVDEs, usually a leucine or phenylalanine is found at this position and in a superposition of *Tth*UVDE with the DNA:protein complex structure, this phenylalanine indeed forms stacking interaction with the base of A8. A hydrogen bond between the conserved Ser67 and the N1 of the A8 base is present that would be lost if a pyrimidine was present in this position due to its smaller size. The size of the pocket together with these stacking and hydrogen bond interactions make that the pocket is customized to fit two purines. Such a specific pocket for two bases of the undamaged DNA strand seems to be a novel property for a DNA repair enzyme. The existence of this pocket is a ingenious, useful feature for an enzyme that needs to recognize UV-damaged DNA, since dipyrimidines always have two purines opposite the lesion. It also gives an explanation for the strong preference of UVDE for incising abasic sites flanked by a pyrimidine over those flanked by a purine (Pas-

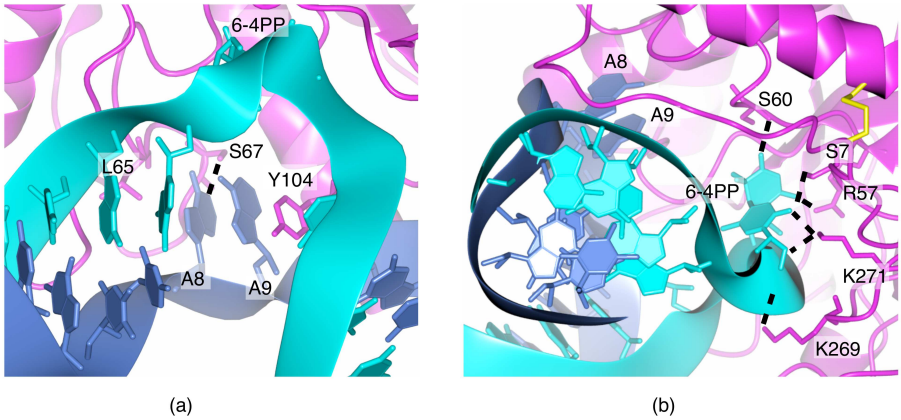

Figure 6.4: Detailed view of the pockets for the undamaged bases (a) and the damaged bases (b), showing their environment in the pockets. Hydrogen bonds are indicated with dashed lines.

paleva *et al.*, 2009), since the abasic sites flanked by a pyrimidine used in (Paspaleva *et al.*, 2009) had two purines opposite the damage. The importance of the pocket for the enzyme's function has been previously shown, since the mutation of the Tyr104 residue to alanine in *Tth*UVDE abolishes its incision activity (Paspaleva, Thomassen *et al.*, 2007). Moreover, flipping of the bases opposite to the damage into a pocket concurs with the previously reported fluorescence studies (Paspaleva *et al.*, 2009). In that study, it was noticed that the bases opposite to the damage are flipped from the helix into a partially to not solvent-exposed area, which agrees with bases flipped into a pocket.

Not only are the undamaged bases flipped, but the damage itself is also flipped into a protein pocket. A hydrogen bond by the probing finger residue Gln103 to the base just 5′ to the damage probably helps in stabilizing this flipped conformation (Figure 6.3). The damage pocket is lined with several residues capable of making hydrogen bonds to the damage (shown in Figure 6.3 and Figure 6.4(b)) and is close to the active site responsible for the enzyme's catalytic function. The pocket is tight around the 6-4PP, but the shape does allow a CPD to fit and be stabilized, while purines would be excluded from the pocket due to their larger size. To enter the pocket, the DNA backbone must deform substantially. The loss of base-pairing and amount of deformation that an undamaged DNA substrate would have to undergo is probably too energetically unfavorable to allow an undamaged DNA substrate to enter the pocket. This explains the broad specificity of the enzyme for many damaged types of DNA, while not cutting undamaged DNA.

6.3.3 *Sac*UVDE has a preference for 6-4 photoproduct over CPD damaged DNA

Although the structure of the apoprotein shows a high structural similarity to *Tth*UVDE, *in vitro* assays surprisingly showed that *Sac*UVDE has a very high in-

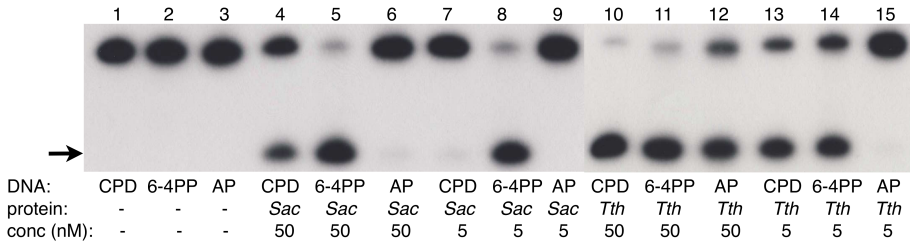

Figure 6.5: Activity of SacUVDE versus TthUVDE.
Incision assay with SacUVDE or TthUVDE (indicated below the lanes) showing that SacUVDE has a strong preference for incising 6-4PP compared to CPD in contrast to TthUVDE. The incision product is indicated with an arrow. The assay was carried out at 55 °C for both proteins.

Table 6.3: DNA binding by SacUVDE wildtype.

DNA lesion	**% binding**
CPD	8.6 ± 4
6-4PP	42.9 ± 2
Abasic site	8.1 ± 1
No damage	8.3 ± 7

cision activity for the 6-4PP, even higher than that of *Tth*UVDE, but it is low for CPD and virtually absent for abasic sites (Figure 6.5 lanes 4 to 9). This is in contrast to *Tth*UVDE that has a high incision activity for both CPD and 6-4PP and a lower incision activity for abasic sites (Figure 6.5 lanes 10 to 15). We set out to compare the structure and function of these two homologues to explain this difference and, by extending this explanation, to get insight into the generally broad substrate specificity of UVDE.

We tested whether the different substrate specificity is caused by experimental conditions, such as DNA length, pH, metal cofactor and temperature (results not shown). Much higher incision on 6-4PP than on CPD has been confirmed on 50-mer DNA substrates at both 55 °C (the standard temperature for our thermophilic incision assays) and 80 °C. The optimum activity is at pH 5.5 to 6.5 (range of 3.5 to 8.5 attempted), which is not surprising since the cellular pH is 6.5 in *S. acidocaldarius* (Moll & Schaefer, 1988). The optimal cofactor is manganese or cobalt; absence of metal ions, calcium, zinc and nickel gives little to no activity, and the addition of magnesium lowers the activity slightly. At all pHs and with all cofactors tested, the activity is always much higher for the 6-4PP than for the CPD, confirming that our result is a genuine characteristic of the protein and is not an artifact caused by experimental conditions.

Bandshift assays were performed on 6-4PP and CPD substrates to determine if the difference in activity is caused by binding or by catalysis. As can be seen in

Figure 6.8 lanes 3 and 4, binding to 6-4PP seems stronger than to the CPD. Both for CPD and 6-4PP the amount of free DNA can be seen to decrease, so the protein binds to both, but only on 6-4PP it gives a complex that is stable and specific enough to see a tight band on the gel. To confirm the results of the bandshift (stronger binding to 6-4PP than CPD), filter-binding assays were performed. As can be seen in Table 6.3, the filter-binding assays show that *Sac*UVDE indeed binds stronger to 6-4PP than to CPD. We therefore conclude that the difference in activity is caused by a difference in binding to the damage.

6.3.4 Structure-based mutagenesis explains *Sac*UVDE's substrate specificity

To explain *Sac*UVDE's preference for binding and incising 6-4PP, we took a closer look at the structural differences between *Sac*UVDE and *Tth*UVDE. In contrast to *Tth*UVDE, *Sac*UVDE has an extra helix (α8) at its C-terminus. There is a shift of around 2 Å in the neighboring α-helix of residues 248 to 260 (α7) and a shift of up to 3.3 Å in the loop of residues 235 to 242 (see Figure 6.6(a)). This latter loop is positively charged in *Tth*UVDE and is close to the DNA groove in the UVDE-DNA structure (Figure 6.9(b)), hence a shift of this loop might cause differences in DNA binding. To test the function of the α8-helix, we made a C-terminal truncation (*Sac*UVDE 1-275). Hardly any overexpression of soluble protein was detected for this construct, indicating that the α8-helix is important for stability in *Sac*UVDE. Moreover, other UVDEs like *Sp*UVDE and *Bs*UVDE have this additional helix as well, yet they can recognize both CPD and 6-4PP. It is therefore unlikely that this helix is the main reason for the difference.

Another reason for the different phenotypes could be that the the damage-binding pockets differ (see Figure 6.6(b)). However, the damage-binding pockets of both enzymes are very similar. The most prominent side-chain difference in the pocket is the Tyr10 in *Sac*UVDE (versus leucine in *Tth*UVDE and valine, methionine or threonine in most other UVDEs). A tyrosine at this position might favor binding of 6-4PP over CPD, because it might have some specific interactions with 6-4PP (the oxygen atom in the deoxyribose ring of base 8 will come into hydrogen bonding distance of the hydroxyl group of Tyr10 when metal ions pull down the lesions deeper into pocket). To test this, we mutated Tyr10 to alanine and tested its activity. As can be seen in Figure 6.7, the activity of the *Sac*UVDE Y10A mutant is very similar to wildtype, with high incision at the 6-4PP and hardly any incision at the CPD and abasic site (compare lanes 4-6 to lanes 11-13). The binding to either substrate is also comparable to wildtype (Figure 6.8). Hence we conclude that the different phenotype is most probably not due to side-chain differences in the damage-binding pocket.

Another remarkable difference between the structures *Sac*UVDE and *Tth*UVDE is a backbone shift (up to 2.8 Å) of residues 10-18 in *Sac*UVDE (Figure 6.6(c)), which is located near the damage-binding pocket. This shift is most probably caused by the disulfide bridge between C14 and C40 in *Sac*UVDE that is not present in *Tth*UVDE nor most other UVDEs. To study the importance of the disulfide bridge, we performed activity assays in the presence of reducing agents (5 - 250 mM β-

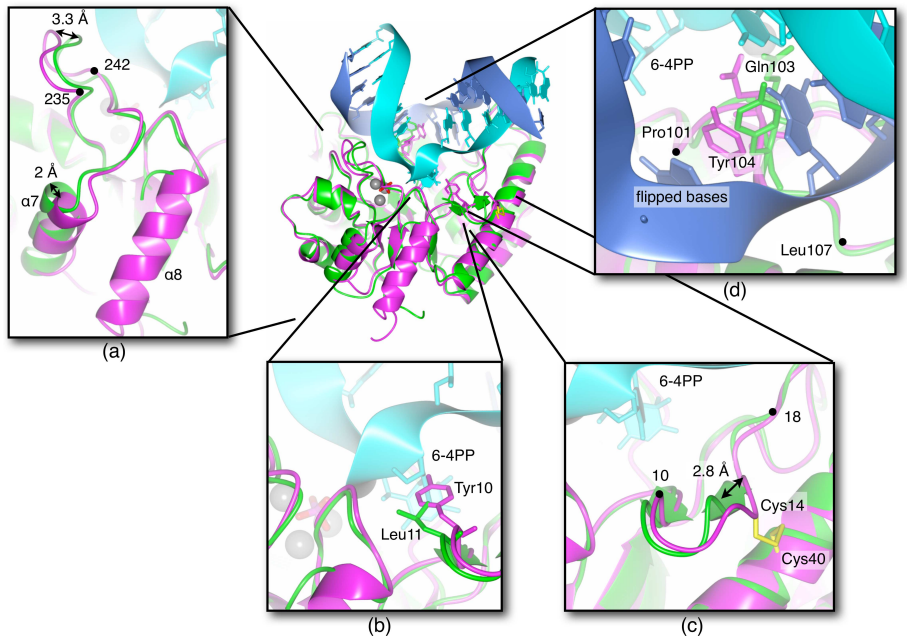

Figure 6.6: Most notable differences between structures of SacUVDE and TthUVDE. Superposition is shown with SacUVDE depicted in magenta, TthUVDE in green, the metal ions in grey, undamaged DNA in blue and damaged DNA in cyan. The following features can be seen. (a) The presence of α8-helix in SacUVDE causes a shift of the α7-helix and the 235-242 loop. (b) The substrate-binding pocket with residue Tyr10 in SacUVDE and Leu11 in TthUVDE. (c) The disulfide bridge C14-C40 causes a shift of residues 10-18 near the damage binding pocket. (d)The probing finger loop of UVDE starting at Pro101 and ending at Leu107. The probing finger itself (Gln103 and Tyr104) is also indicated. The numbering of SacUVDE is used.

mercaptoethanol or DTT; results not shown). No difference in activity could be observed even at the highest reducing agent concentration. Since there is 5 mM DTT present in the crystallization buffer and still the disulfide bridge was seen in the crystal structure, it is likely that the disulfide bridge is not easily broken. During the assay in presence of reducing agents this might also have been the case.

To disrupt the disulfide bridge with certainty and hence be able to verify its importance, we mutated C14 to alanine and tested its activity. Figure 6.7 shows that *Sac*UVDE C14A (lanes 14-16) has a strongly reduced activity compared to *Sac*UVDE wt (lanes 4-6). This can be seen more clearly in an assay with reduced concentrations of protein, which shows a much lower incision activity of *Sac*UVDE C14A (lanes 20-26) compared to wildtype (lanes 4-10). At all temperatures tested (35 °C, 45 °C and 55 °C), the relative difference between mutant and wildtype stays the same (results not shown). The solubility of the mutant during purification is comparable to wildtype. Based on these two observations, we conclude that the re-

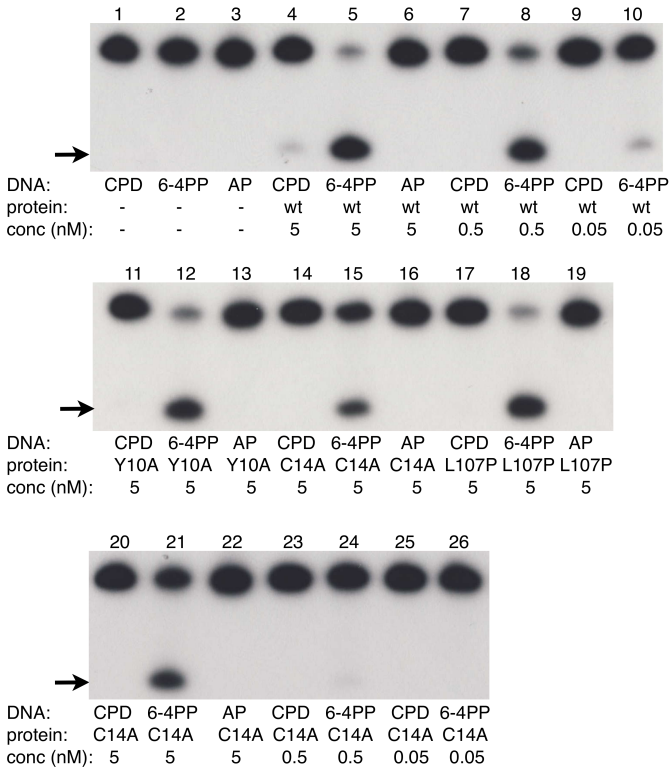

Figure 6.7: Activity of mutants of SacUVDE.
Lanes 1-10: Incision activity of SacUVDE wt at concentrations of 0.05 to 5 nM showing a strong preference for incision on 6-4PP. The incision product is indicated with an arrow.
Lanes 11-19: Incision activity of SacUVDE mutants showing that mutants Y10A and L107P have a similar activity to SacUVDE wt and that SacUVDE C14A has a strongly reduced activity.
Lanes 20-26: Incision activity of SacUVDE C14A in more detail confirming the strongly reduced activity of this mutant.

duced activity is not caused by reduced stability during the time of the assay, but by a genuine defect in the protein's function. We tested DNA binding of this mutant by a bandshift assay (Figure 6.8), which also shows a reduced affinity, suggesting that mutant C14A has a defect in binding to damaged DNA. This result was confirmed by filter-binding assays (results not shown). It is possible that the rigidity caused by the disulfide bridge prevents *Sac*UVDE to incise efficiently both CPD and 6-4PP, but that removing this disulfide bridge leads to a pocket that is too flexible to cut anything efficiently.

An important factor for DNA recognition and binding is the probing finger of the residues Gln103 and Tyr104 in *Sac*UVDE (Gln104 and Tyr105 in *Tth*UVDE; Figure 6.6(d)), which helps in flipping out the damaged bases and the opposite bases from the double DNA helix. In *Tth*UVDE, the loop towards this finger is more

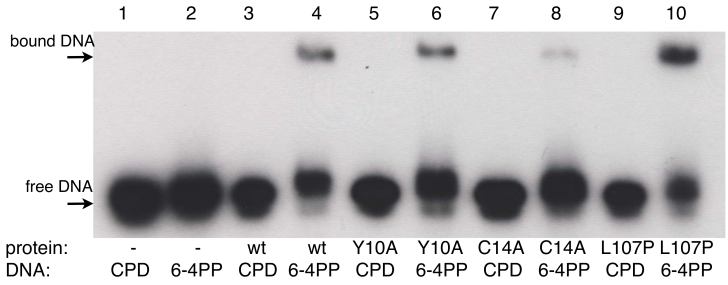

Figure 6.8: Bandshift assay of SacUVDE wt and mutants Y10A, C14A and L107P, showing all proteins have a strong preference of binding 6-4PP to CPD and reduced binding of mutant C14A. Protein and DNA substrates used are indicated below the lane.

rigid, since it has a proline at the beginning and the end of it (Pro102 and Pro108), while *Sac*UVDE only has one proline at the corresponding positions (Pro101 and Leu107). A more rigid finger loop might be more suited to flip out less distorted DNA substrates and hence a protein with such a rigid finger might have a broader substrate range. To test this hypothesis, we mutated Leu107 in *Sac*UVDE to a proline. The resulting protein was poorly soluble (most of the protein was in the pellet after lysis and centrifugation), but the soluble part could still be purified to near homogeneity. As can be seen in Figure 6.7 the mutant protein (lanes 17-19) has an incision activity and substrate specificity comparable to wildtype (lanes 4-6): it only incises 6-4PP efficiently. Also binding studies show the same pattern (see Figure 6.8), with a large preference for 6-4PP over CPD. The binding efficiency of L107P looks slightly higher than wildtype, but this is probably due to inaccurate concentration determination of the mutant due to its low concentration caused by problems in its solubility. Taken together the above results, we deem it unlikely that the difference in rigidity of the probing finger causes the different phenotypes of *Sac*UVDE and other UVDEs.

A charge surface plot of both *Sac*UVDE and *Tth*UVDE (Figure 6.9) shows a difference in the positive charges on the end of the DNA-binding groove that insert into the DNA helix, which have previously been hypothesized to be involved in DNA binding (Paspaleva *et al.*, 2007). For *Tth*UVDE there is a stretch of three positively charged residues (Lys238, Lys239 and Arg240) and another lysine close by (Lys179) on one side of the groove, while *Sac*UVDE only has two (Lys178, Lys239). On the other side of the groove, *Tth*UVDE also has more positive charges: four (Arg22, His25, Arg30, Lys34), while *Sac*UVDE has only two (Lys21, Arg29). Having a prominent positive charge on both sides of the groove likely helps in fixing and bending the DNA and might explain the different phenotype of *Sac*UVDE and *Tth*UVDE. To verify the importance of this positive charge at both sides of the DNA-binding groove, we mutated residues Gly237 and Glu238 to lysine in *Sac*UVDE. To test its activity in detail, we performed a kinetics incision assay (shown in Figure 6.10). In this assay it can be seen that the difference in incision activity at CPD and

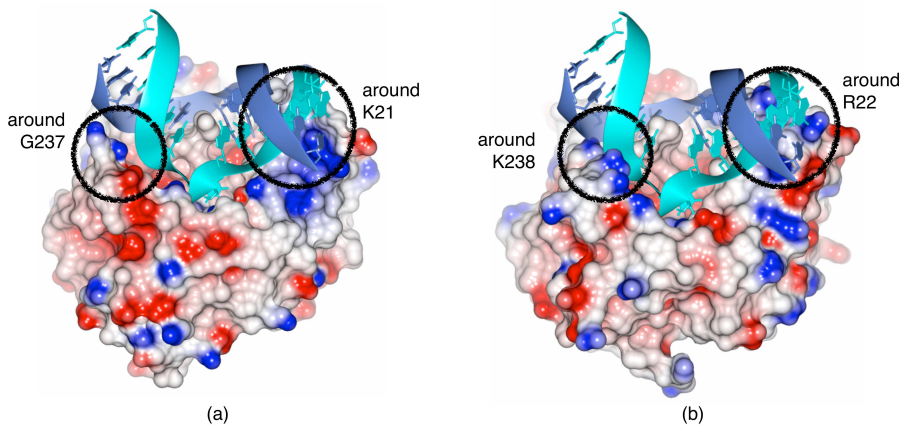

Figure 6.9: Charge surface representations of SacUVDE and TthUVDE with positive charge in blue and negative charge in red. DNA is depicted in cyan (damaged strand) and blue (undamaged strand).
(a) SacUVDE; regions corresponding to that with more positive charge in TthUVDE are shown in circles with one residue number for orientation.
(b) TthUDVE; regions which have more positive charge shown in circles.

6-4PP has become smaller in the G237K + E238K mutant. This can be explained by assuming a combination of the two following effects. First, the overall activity of the protein goes down, as can be seen with the 6-4PP, due to addition of non-natural charge for this homologue that might put the DNA slightly in the wrong position. Secondly, the CPD goes up relatively to the 6-4PP, potentially due to the easier bending of the DNA. Together, these two effects can cause the result we observed: a smaller difference in the activity between the two substrates. Thus, the positive charge on both sides of the groove can be indeed an important factor in *Sac*UVDE substrate preference and explains at least in part its difference with UVDEs from other organisms.

6.4 Discussion

The pre-incision complex structure has provided confirmation of previously proposed elements of the mechanism of UVDE, such as the DNA binding groove and the 90° bend in the DNA. It also presents a novel mechanism for recognizing damaged dipyrimidines by a dual flipping mechanism. UVDE flips the undamaged bases opposite to the damage (always two purines, since the damage consists of a dipyrimidine) in a dipurine-specific pocket. Not only the undamaged bases are flipped, but also the damaged ones are flipped in a protein pocket. This combination represents an elegant mechanism for recognizing and verifying the presence of the lesion. First, the flipping probably can only occur in damaged DNA, since the presence of the lesion causes a distortion in the DNA (such as the loss of hydro-

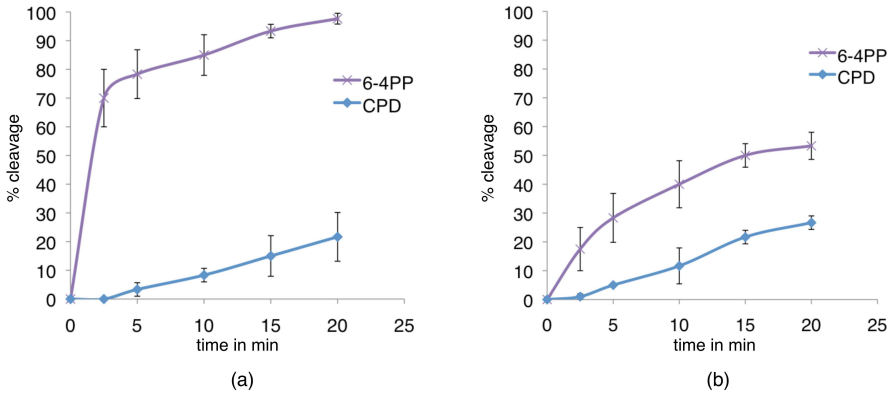

Figure 6.10: Activity of SacUVDE wt and SacUVDE G237K E238K.
(a) Kinetics incision assay with SacUVDE wildtype, showing a strong difference between incision on CPD and 6-4PP.
(b) Kinetics incision assay with SacUVDE G237K E238K, showing a smaller difference between incision on CPD and 6-4PP.

gen bonding). It is therefore highly likely that flipping is easier in damaged DNA. Second, dipurines are stabilized in the undamaged-DNA-binding pocket, providing a mechanism for preferentially repairing distorted dipyrimidines. Thirdly, the damaged bases are flipped into a pocket that can accommodate damaged dipyrimidines very well, but will exclude larger lesions. Smaller lesions such as abasic sites can be incised by UVDE, albeit with a lower efficiency, perhaps because the lesion is too small to be held stably in the damage pocket. And last, the DNA backbone is rather distorted at the point where the damaged bases are flipped, which likely helps in discriminating between damaged and undamaged DNA: for undamaged DNA such a distortion would be energetically quite unfavorable.

To gain more insight into the broad specificity of UVDE, we have compared the phenotype and structure of the more general *Tth*UVDE to the more specific *Sac*UVDE. The latter only efficiently incises 6-4PPs, the UV-lesion that causes severe distortion in the DNA backbone (a 44° kink; Kim & Choi, 1995). Surprisingly, *Sac*UVDE's structure was found to be very similar to the structure of *Tth*UVDE, which also incises CPDs well and abasic sites moderately well. We demonstrated that the different phenotype is not caused by a more specific damage-binding pocket nor by less rigidity of the probing finger. However, the disulfide bridge in *Sac*UVDE might have a role in this. Removing this disulfide bridge in *Sac*UVDE has a detrimental effect on the protein's ability to bind and incise damaged DNA. This could be explained as follows. Rigidity around the binding pocket potentially caused by the disulfide bridge in *Sac*UVDE might prevent the protein from binding to both CPD and 6-4PP. This non-flexible pocket might be shaped to only fit well, selectively, 6-4PP. Removing the disulfide bridge might then result in a protein with a binding pocket too flexible to bind anything stably, as we observed in

our biochemical assays. The disulfide bridge might have arisen in evolution to yield a more thermoresistant protein for this thermophilic organism, which has a 10 to 15 °C higher optimum growth temperature than *Thermus thermophilus*). Perhaps as trade-off for the extra stability, it had to give up the broader substrate specificity typical for UVDE. Potentially it could do this, because the organism already has a photolyase for repairing CPDs. This scenario is feasible, since a homologous protein to the potential ORF for *S. acidocaldarius* photolyase has been shown to reverse CPDs *in vitro* (Fujihashi *et al.*, 2007).

Less positive charge on both sides of the DNA binding groove might also make binding a broad range of DNA substrates more difficult for *Sac*UVDE, since this charge likely assists in bending the DNA; a 90° bend in the DNA is needed for optimal binding. The 6-4PP is already distorted more than the CPD in solution, explaining why the protein can bind 6-4PP without the charge on both sides of the groove. Indeed, addition of this charge in *Sac*UVDE decreases the difference in incision efficiency between CPD and 6-4PP. This could be taken as a confirmation of this hypothesis. In conclusion, we think that the difference in substrate specificity between *Tth*UVDE and *Sac*UVDE is caused by a combination of effects, such as the rigidity caused by the disulfide bridge and more difficulty in bending DNA caused by having less positive charges on both sides of the groove.

Comparison of the very specific *Sac*UVDE to the more broad *Tth*UVDE gives insight into the relatively broad substrate specificity of UVDE in general. It seems that flexibility around the binding pocket and capacity of bending the DNA are important factors for UVDE's capacity to recognize and incise different DNA lesions. *Sac*UVDE might have given up these features to obtain a more stable protein for living at extraordinarily high temperatures. Though *Sac*UVDE perhaps plays a lesser role in DNA repair in *S. acidocaldarius*, it may be useful for biochemical assays to distinguish between CPD and 6-4PP, since it is, to the authors' knowledge, the most specific enzyme for incision only next to 6-4PPs.

Taking together the above results, we can make the following model for UVDE activity. First, UVDE recognizes a distortion in the DNA, such as a kink and/ or loss of hydrogen bonding, and binds to it. The residues of the probing finger (Gln103 and Tyr104) flip the damaged bases as well as the opposite bases out of the helix into their respective pockets. Only if the opposite bases are two purines (so they fit well in their pocket) and if the damaged bases fit in the damage pocket, does the enzyme bind stably to the damage. The positive charge of the metal ions then draws in the scissile phosphodiester bond and incision by a hydroxyl ion takes place using three metal-ion mediated catalysis.

7

Crystal structure of a heterodimeric, double-headed Kunitz-type serine protease inhibitor from potato.

The Potato Serine Protease Inhibitor (PSPI) constitutes about 22% of the total amount of proteins in potato tubers (cv. Elkana) making it the plant's most abundant protease inhibitor. PSPI is a heterodimeric, double-headed Kunitz-type serine protease inhibitor that can tightly and simultaneously bind two serine proteases by mimicking the enzyme's substrate with its reactive site loops. We report the crystal structure of PSPI, representing the first heterodimeric double-headed Kunitz-type serine protease inhibitor structure determined. PSPI has a β-trefoil fold and based on the structure, we have identified two reactive site loops bearing residues Phe75 and Lys95.

published in: E.M. Meulenbroek, E.A.J. Thomassen, L. Pouvreau, J.P. Abrahams, H. Gruppen, N.S. Pannu, *Structure of a post-translationally processed heterodimeric double-headed Kunitz-type serine protease inhibitor from potato.*, Acta Cryst. **D68**, 794-799 (2012)

7.1 Introduction

Protease inhibitors are abundantly present in plant seeds and tubers. In the tubers of the potato cultivar Elkana, for example, circa 50 % (w/w) of the total amount of soluble proteins are protease inhibitors (Pouvreau *et al*., 2001). Plant protease inhibitors prevent pathogens or predators feeding on the plant by inhibiting their proteases. Protease inhibitors are, therefore, also expressed in response to wounding of the plant (reviewed in Valueva & Mosolov, 2004). In addition to their importance in plant physiology, protease inhibitors are also reported to show anti-carcinogenic effects: they were shown to prevent carcinogenesis in many different model systems and are effective at very low concentrations (reviewed in Kennedy, 1998).

Kunitz-type protease inhibitors can bind and inhibit serine, cysteine and aspartic proteases (Oliva *et al*., 2010). Plant Kunitz-type protease inhibitors are small 20 kDa proteins that usually have two disulfide bonds and a single reactive site (single-headed), though two reactive site (double-headed) Kunitz-type inhibitors have also been described (e.g. Dattagupta *et al*., 1999 and Azarkan *et al*., 2011). The proteins inhibit proteases by binding tightly to the enzyme's active site via their reactive site loops in a substrate-like manner.

Potato serine protease inhibitor (PSPI) is the most abundant protease inhibitor group in potato tubers and constitutes 42 % of the protease inhibitors present in potato juice (Pouvreau *et al*., 2003). This group of Kunitz-type serine protease inhibitors is composed of seven different isoforms, differing slightly in their pI's. PSPI is reported to inhibit the serine proteases trypsin, chymotrypsin and human leukocyte elastase (Valueva *et al*., 2000). It is expressed as a single polypeptide, but six amino acids are deleted during post-translational processing to yield a protein consisting of a large (16.2 kDa) and small (4.2 kDa) subunit that are held together by a disulfide bridge and non-covalent interactions (Pouvreau *et al*., 2003); literature on Kunitz-type protease inhibitors refers to this multimeric complex as a "dimer", but for clarity, we refer to the complex as a heterodimer. PSPI is proposed to be a double-headed protease inhibitor, i.e. having two independent reactive site loops, since it can form a ternary complex with both trypsin and chymotrypsin (Valueva *et al*., 2000). These two loops, however, have never been identified. To be able to determine the location and residues involved in the reactive site loops and thus gain insight into the mechanism of PSPI and since no three-dimensional atomic structure of a heterodimeric double-headed Kunitz-type protease inhibitor is known, we set out to determine the crystal structure of this protein. Here we present the high resolution structure of PSPI. The structure shows a β-trefoil fold with large protruding loops and based on the structure, we identified the two reactive loops as being centered on Phe75 and on Lys95.

7.2 Methods

Protein purification, crystallization and data collection have been reported previously (Thomassen *et al*., 2004). The crystal diffracted to 1.60 Å, but the data com-

pleteness at this resolution is very low, with only 18.4 % completeness in the highest resolution shell (1.69 - 1.60 Å). Although incomplete, we included this high-resolution data in refinement. The resolution of the data corresponding to a 100 % complete dataset, as calculated by *SFTOOLS* (Bart Hazes, unpublished) is 1.80 Å. The phase problem was solved by molecular replacement using *BALBES* (Long *et al.*, 2008) with the structure of soybean trypsin inhibitor (Song & Suh, 1998) as a search model. The molecular replacement solution was then rebuilt with *ARP/wARP* (Perrakis *et al.*, 1999) resulting in a model that was 87 % complete. Refinement was performed with *REFMAC* using TLS (Murshudov *et al.*, 2011) and manual fitting was performed using *COOT* (Emsley *et al.*, 2010). Data and refinement statistics are shown in Table 7.1. In the analysis of the structure, superpositions with other Kunitz-type serine proteases and root-mean-square deviation calculations were done by *THESEUS* (Theobald & Wuttke, 2006). Structure-based sequence alignment was done with *VAST* (Thompson *et al.*, 2009). Atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank (accession code 3tc2). Figures were made with *CCP4mg* (Potterton *et al.*, 2004).

7.3 Results and Discussion

7.3.1 Overall structure

The structure of PSPI was determined to high resolution and contains three molecules in the asymmetric unit. These three molecules are very similar with the root-mean-square deviations of the Cα-backbone being 0.23 Å between chains A and B, 0.33 Å between A and C and 0.28 Å between B and C.

The quality of the electron density map is excellent for nearly all of the structure (Figure 7.1(a)). The map is poor around the loop of amino acids 42-46 in chains B and C and around amino acids 72-77 in chains A and B. Nevertheless, for each of these loops, an NCS-related chain with moderate to good density (chain A for loop 42-46 and chain C for loop 72-77) was available to build a complete model. The electron density being well defined only in A for 42-46 and only in C for 72-77 could not be attributed to any obvious interface interactions in the crystal.

The overall structure of PSPI shows a β-trefoil fold (Figure 7.1(b)). It consists of twelve anti-parallel β-strands, which form six two-stranded β-hairpins. Three of these hairpins form a small β-barrel, while the other three hairpins close off the barrel in a triangular fashion. The β-strands are connected by long, protruding loops. Two disulfide bridges are present: between Cys48 and Cys97 and between Cys146 and Cys157. This fold is usually observed for Kunitz-type serine protease inhibitors (e.g. Soybean Trypsin Inhibitor, (Song & Suh, 1998)) and the reactive sites are commonly found in the protruding loops. The β-sheets provide a stable core that can hold together the structure even if incision by the protease takes place, both by the covalent interaction of the disulfide bridges and the numerous non-covalent interactions.

Several structures of Kunitz-type serine protease inhibitors have been deter-

Table 7.1: Refinement and validation statistics

Data collection	
Space group	$P12_11$
Cell dimensions	
a, *b*, *c* (Å)	54.82 x 93.92 x 55.44
α, β, γ (°)	90.00, 100.69, 90.00
Resolution (Å)	54.23–1.60 (1.69-1.60)[a]
Resolution of a 100 % complete data set (Å)	1.80[b]
Wilson plot B-factor	20.1
R_{merge}[c]	0.044 (0.370)
I/σI	14.0 (1.8)
Completeness (%)[b]	70.0 (18.4)
Redundancy	2.1 (1.8)
Total n°. observations	104501
N° unique reflections	50787
N° reflections in R_{free} set	2533
Refinement	
Resolution (Å)	54.23–1.60
N° reflections	48229
N° molecules in ASU	3
R_{work} / R_{free}	16.8/ 22.1
N° atoms	
Protein	4141
Ligand/ ion	0
Water	504
B-factors	
Protein	24.2
Ligand/ ion	N/A
Water	32.2
R.m.s. deviations	
Bond lengths (Å)	0.030
Bond angles (°)	2.25
No. TLS bodies	3
Ramachandran favored[d]	95.23 %
Ramachandran outliers[d]	0.38 %
Rotamer outliers[d]	0.88 %

[a] Values in parentheses are for highest-resolution shell.
[b] Data is complete to 1.80 Å.
[c] Multiplicity weighted R-merge, (Diederichs & Karplus, 1997; Weiss & Hilgenfeld, 1997)
[d] As determined by *Molprobity* (Davis *et al.*, 2007)

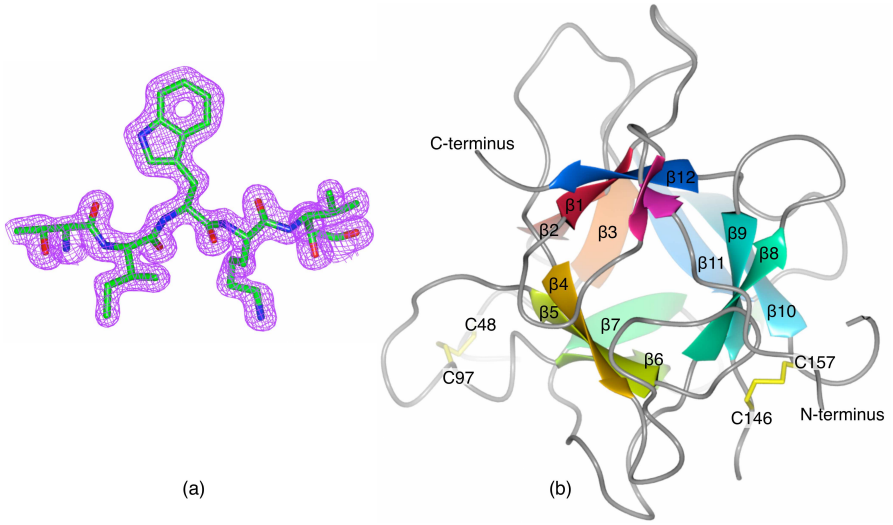

Figure 7.1: Overall structure of PSPI.
(a) Representative part of the electron density map of PSPI with a short stretch of the model (map clipped to show only this part of the model), demonstrating excellent electron density to a high resolution. Map is contoured at 1.5 σ.
(b) Ribbon diagram of one molecule of PSPI with β-sheets indicated, showing a β-trefoil fold. The N- and C-terminus are labelled as well as the two disulfide bridges (in yellow).

mined previously. The structures most similar to PSPI in terms of amino acid sequence are a miraculin-like protein with 32.7% identity (PDB entry 3iir, Gahloth *et al.*, 2010), a Kunitz-type inhibitor from *Delonix Regia* seeds with 31.7% identity (PDB entry 1r8n, Krauchenco *et al.*, 2003), a winged-bean chymotrypsin inhibitor with 27.9% identity (PDB entry 2wbc, Dattagupta *et al.*, 1999) and a soybean trypsin inhibitor with 27.4% identity (PDB entry 1avx, Song & Suh, 1998). A structure-based sequence alignment of PSPI with other similar proteins (Figure 7.2) shows that the sequences are most similar in the β-strand regions, which form the scaffold of the structure, compared with that of protruding loops, which determine the specificity of the inhibitors.

Structural superposition gives a similar pattern. The overall root-mean-square deviations of the Cα-backbone of PSPI with these four structures are (calculated with subunit A of PSPI): 1.47 Å for the miraculin-like protein, 1.22 Å for the Kunitz-type inhibitor from *Delonix regia* seeds, 1.21 Å for a winged-bean chymotrypsin inhibitor, and 1.35 Å for a soybean trypsin inhibitor. This indicates that the structures are indeed similar, but as expected, the similarity is mostly in the β-sheet core, whilst the protruding loops differ substantially.

An important difference between PSPI and the previously determined structures of Kunitz-type serine protease inhibitors is that, to the best of our knowledge, PSPI is the first structure of a post-translationally modified heterodimeric

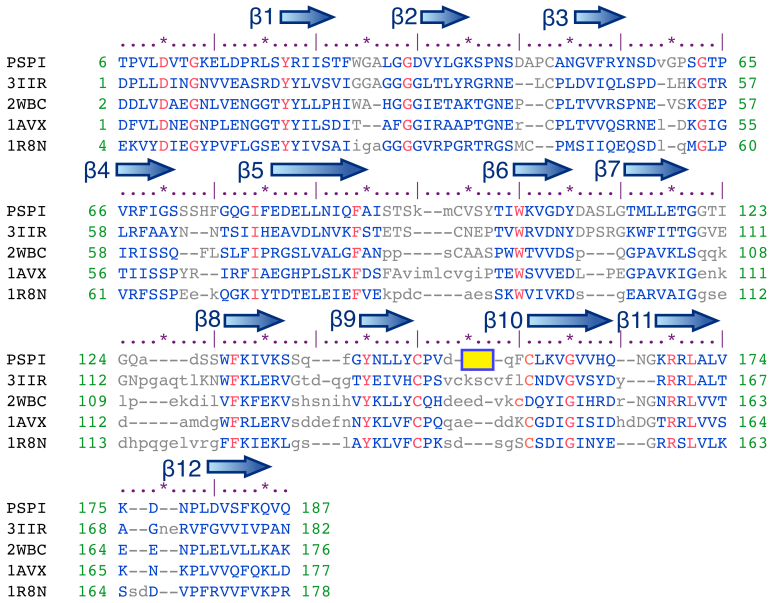

Figure 7.2: Structure-based sequence alignment.
Alignment of PSPI with other previously structurally characterized Kunitz-type serine protease inhibitors. β-strands are indicated with blue arrows and the yellow box shows where six residues are post-translationally deleted from PSPI.

double-headed protease inhibitor. Only crystal structures of *monomeric* Kunitz-type double-headed serine protease inhibitors are known, such as winged bean α-chymotrypsin inhibitor (Dattagupta *et al.*, 1999).

PSPI is expressed as a single polypeptide chain, but in a post-translational process around six residues after Thr150 are removed. Thus, there is no loop between Thr150 and Ser151 and despite the consecutive numbering of the amino acids, there is also no peptide bond between Thr150 and Ser151, because of the deletion. The residues just before the site of the deletion (Thr149 and Thr150) and just after (Ser151, Ser152 and Asp153) are not modelled in our structure due to poor density probably caused by structural flexibility. Despite being (hetero)dimeric, the conformation of the whole molecule including the region close to the deleted loop is very similar to other, monomeric, Kunitz-type serine protease inhibitors. The structure seems to be held together in this very similar conformation by the disulfide bridge Cys146-Cys157 close to the beginning and the end of the deleted loop, since only the residues between these two cysteines show a different backbone direction from monomeric Kunitz-type serine protease inhibitors. Apparently, the presence of the extra loop and the post-translational deletion of the loop has not affected the overall structure of the inhibitor. It is possible that the loop functions as a shield for an important interface of the protein leading to a (partially) inactive form, which gets

activated upon cleavage. The range in which the excised loop can be is indicated in Figure 7.3(a).

7.3.2 Identification of reactive site loops

The reactive site loops for PSPI have not yet been identified. Previous predictions based on sequence alignment with other Kunitz-type inhibitors located the reactive site loops around Arg67-Phe68 (with Arg67 being the P1 position) and around Met115-Leu116 (with Met115 occupying the P1 position; Valueva *et al*., 2000). However, both these proposed sites are located on the inside of the structure and not in protruding loops. Indeed, Arg67 is in the middle of a β-strand in the protein's core and hence cannot be the P1 residue of the reactive site loop. Similarly, there would also be a considerable amount of steric hindrance for a protease to bind if Met115 was a P1 amino acid.

To find more likely reactive site loops, we compared the PSPI structure to other Kunitz-type serine protease inhibitors with previously determined and verified reactive site loops. Five possibilities for reactive side loops arose. Based on comparison with double-headed arrowhead protease inhibitor, the reactive loops could be located around Asn50 and around Leu84 of PSPI (Xie *et al*., 1997). However, in the PSPI structure these residues are also not in protruding loops, but at sterically more inaccessible positions, and thus ruling out these sites.

Another possibility for a reactive site loop is near Asp45-Asn50 in PSPI (see Figure 7.3(a) and (b)). This loop corresponds to the Asn38-Leu43 loop that has been proposed to be one of the reactive site loops for winged-bean α-chymotrypsin inhibitor (Dattagupta *et al*., 1999) based on structural similarity to its other verified reactive site. The Asp45-Asn50 loop in PSPI looks similar in structure and partially similar in sequence to the winged-bean α-chymotrypsin inhibitor reactive loop. The P1 residue in this loop is proline (Pro47 in PSPI) for both proteins. In PSPI this loop is partially disordered in chains B and C and has higher than average B-factors for chain A. This is a typical property of substrate loops that probably aids in binding the protease. The reactive site loops adopt a more rigid conformation upon protease binding (Song & Suh, 1998).

The most common reactive site loop for monomeric Kunitz-type serine protease inhibitors is the loop starting at Ser71 to Phe80 with Phe75 being the P1 residue in PSPI. This sequence was proposed to be the reactive site loop of other serine protease inhibitors including miraculin-like protein (Gahloth *et al*., 2010), for a mutant trypsin inhibitor based on a co-crystal with trypsin (PDB entry 3i29), for double-headed winged bean chymotrypsin inhibitor based on biochemical evidence (Shibata *et al*., 1988) and, finally, for soybean trypsin inhibitor based on a co-crystal with trypsin (Song & Suh, 1998). In PSPI this loop has poor density in chains A and B and has high B-factors in chain C, indicating flexibility. As can be seen in Figure 7.3(a), Phe75 sticks out into the solvent in chain C (the only NCS-related molecule for which there is moderately good electron density; see Figure 7.3(c) for an OMIT map), which would facilitate its recognition by a target protease.

The last possibility for the reactive site loop is around Lys95 in PSPI. The

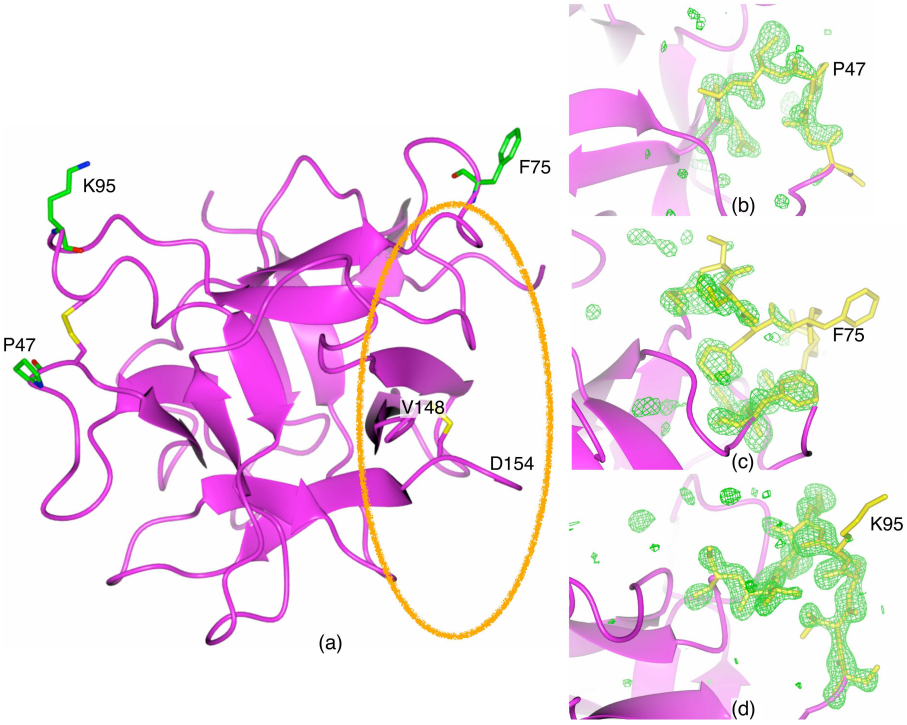

Figure 7.3: Reactive site loops of PSPI.
(a) P1 residues of the three most probable candidates for reactive site loops of PSPI highlighted on the structure of PSPI, showing their position to be in loops and showing how they stick out into the solvent. Disulfide bridges are indicated in yellow. The orange circle shows the range in which the excised loop can be (on the right side of the protein in this view) and the last and first ordered residue in the structure at the beginning and end of the excised loop are indicated (V148 and D154).
(b) Omit map of putative reactive site loop of Asp45-Asn50 (yellow).
(c) Omit map of putative reactive site loop of Ser71-Phe80 (yellow).
(d) Omit map of putative reactive site loop of Ser92-Ser99 (yellow).
The electron difference density maps are calculated based on phases from the model without the residues shown here in yellow and are contoured at 3σ.

corresponding loop was identified to be a reactive site loop in both double-headed arrowhead protease inhibitor (PDB entry 3e8l) based on mutagenesis studies and a co-crystal with two trypsins (Bao *et al*., 2009) and also in barley α-amylase/subtilisin inhibitor based on a co-crystal with subtilisin savinase (Micheelsen *et al*., 2008). The latter inhibitor is identified as fifth structurally most related protein to PSPI by the *VAST* program. The loop around Lys95 has a similar conformation to the loop from the double-headed arrowhead protease inhibitor and protrudes from the structure with Lys95 sticking out to the solvent as can be expected for a P1 reactive site residue (Figure 7.3(a) and (d)).

As stated above, the loops around Pro47, Phe75 and Lys95 are the three most likely candidates for reactive site loops. Since in the trypsin active site a large positively charged amino acid fits best (to make contacts with Asp189), we propose that the Lys95 is the trypsin interacting loop. Chymotrypsin is specific for sequences with bulky hydrophobic amino acids; hence it is plausible that chymotrypsin binds to the Phe75 loop. Moreover, these two reactive site loops have been confirmed by co-crystallization in other Kunitz-type serine protease inhibitors providing stronger evidence for its potential biological function than structural similarity of a loop in an uncomplexed structure, as is the case with Pro47. The Phe75 reactive site loop might be hidden before the post-translational processing of the loop between Thr150 and Ser151 (Figure 7.3(a)).

It has been reported that one molecule of PSPI can bind a trypsin and a chymotrypsin molecule at the same time (Valueva *et al*., 2000). To verify whether this is indeed possible with the suggested reactive site loops, we modelled two proteases on the reactive site loops based on co-crystal structures in which a trypsin is bound to one of these loops (Song & Suh, 1998 and Bao *et al*., 2009) using the secondary-structure matching (SSM) function in *COOT*. As can be seen in Figure 7.4, it is possible for the two proteases to simultaneously bind PSPI.

It has also been reported that PSPI inhibits human leukocyte elastase (Valueva *et al*., 2000). Since elastase is specific for a small neutral amino acid, we are unsure whether it can act on either of the loops proposed for trypsin or chymotrypsin. Thus, it might need another binding site such as the Pro47 loop. Unfortunately, no elastase structure with bound Kunitz-type serine protease inhibitor is available for comparison.

7.3.3 Interfaces for aggregation

The *PISA* (Protein Interfaces, Surfaces and Assemblies) server (Krissinel & Henrick, 2007) was used to investigate potential interfaces for interaction in the PSPI crystal structure. The surface including residues 107, 127 and 148 with the plane formed by residues 30, 58 and 169 was predicted to be a strong interface and have a high probability to occur in solution. Since PSPI was reported to exist as a single molecule in solution (Pouvreau *et al*., 2005), this interface is apparently not seen under ambient conditions. However, it has been reported that PSPI aggregates upon heating into a product with four molecules of PSPI (Pouvreau *et al*., 2005). The interface that is detected by *PISA* could be the starting interface of such aggrega-

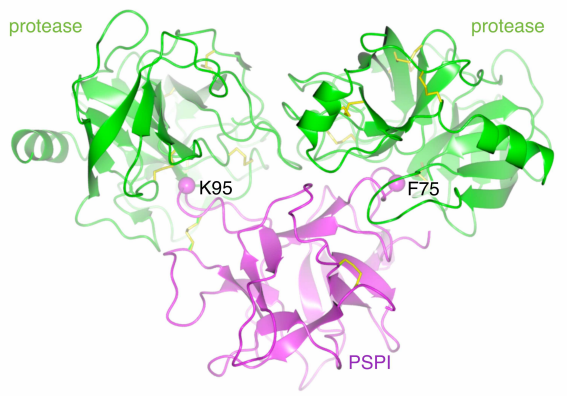

Figure 7.4: Model of PSPI with two proteases.
Model of PSPI (magenta) with two proteases (green) on it, showing that it is possible for two proteases to bind simultaneously at the reactive sites F75 and K95 to one molecule of PSPI. In this picture two trypsins were modelled, but the picture will be equivalent for one chymotrypsin and one trypsin, due to the structural similarity of trypsin and chymotrypsin.

tion when the conditions for solubility are changed, as happens in a crystallization experiment with increasingly high concentrations, after which two seeding two-molecule-PSPI-complexes join to form an aggregate containing four molecules.

Bibliography

Abbas, T., Sivaprasad, U., Terai, K., Amador, V., Pagano, M., Dutta, A., PCNA-dependent regulation of p21 ubiquitylation and degradation via the CRL4(Cdt2) ubiquitin ligase complex. *Genes Dev.* **22**, 2496-2506 (2008)

Abendroth, J., Niefind, K., and Schomburg, D., X-ray structure of a dihydropyrimidinase from *Thermus sp* at 1.3 Å resolution. *J. Mol. Biol.* **320** 143-156 (2002)

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

Anindya, R., Mari, P.O., Kristensen, U., Kool, H., Giglia-Mari, G., Mullenders, L.H., Fousteri, M., Vermeulen, W., Egly, J.M., Svejstrup, J.Q., A ubiquitin-binding domain in Cockayne syndrome B required for transcription-coupled nucleotide excision repair. *Mol. Cell* **38**, 637-648 (2010)

Angers, S., Li, T., Yi, X., MacCoss, M.J., Moon, R.T., Zheng, N., Molecular architecture and assembly of the DDB1-Cul4A ubiquitin ligase machinery. *Nature* **443**, 590-593 (2006)

Avery, A.M., Kaur, B., Taylor, J., Mello, J.A., Essigmann, J.M., and Doetsch, P.W., Substrate specificity of ultraviolet DNA endonuclease (UVDE/Uve1p) from *Schizosaccharomyces pombe*. *Nucleic Acids Res.* **27**, 2256-2264 (1999)

Azarkan, M., Martinez-Rodriguez, S., Buts, L., Baeyens-Volant, D., Garcia-Pino, A., The plasticity of the β-trefoil fold constitutes an evolutionary platform for protease inhibition. *J. Biol. Chem.* **286**, 43726-43734 (2011)

Barry, M. & Früh, Viral modulators of Cullin RING ubiquitin ligases: culling the host defense. *Sci. STKE* **335**, pe21(2006)

Battye, T.G., Kontogiannis, L., Johnson, O., Powell, H.R., Leslie, A.G., iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Cryst.* **D67**, 271-281 (2011)

Bao, R., Zhou, C.Z., Jiang, C., Lin, S.X., Chi, C.W., Chen, Y., The ternary structure of the double-headed arrowhead protease inhibitor API-A complexed with two

trypsins reveals a novel reactive site conformation. *J. Biol. Chem.* **284**, 26676-26684 (2009)

Berger, I., Fitzgerald, D.J., Richmond, T.J., Baculovirus expression system for heterologous multiprotein complexes. *Nat. Biotechnol.* **12**, 1583-1587 (2004)

Berquist, B.R., Canugovi, C., Sykora, P., Wilson III, D.M., Bohr, V.A., Human Cockayne syndrome B protein reciprocally communicates with mitochondrial proteins and promotes transcriptional elongation. *Nucleic Acids Res.* epub ahead of print (2012)

Bhatia, P.K., Verhage, R.A., Brouwer, J., Friedberg, E.C., Molecular cloning and characterization of *Saccharomyces cerevisiae* RAD28, the yeast homolog of the human Cockayne syndrome A (CSA) gene. *J. Bacteriol.* **178**, 5977-5988 (1996)

Birnboim, H.C., Nasim, A., Excision of pyridmidine dimers by several UV-sensitive mutants of *S. pombe*. *Mol. Gen. Genet*. **136**, 1-8 (1975)

Bohr, V.A., Smith, C.A., Okumoto, D.A., Hanawalt, P.C., 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 (1985)

Bowman, K.K., Sidik, K., Smith, C.A., Taylor, J.S., Doetsch, P.W., Freyer, G.A., A new ATP-independent DNA endonuclease from Schizosaccharomyces pombe that recognizes cyclobutane pyrimidine dimers and 6-4 photoproducts. *Nucleic Acids Res.* **22**, 3026-3032 (1994)

Bradsher, J., Auriol, J., Proietti de Santis, L., Iben, S., Vonesch, J.L., Grummt, I., Egly, J.M., CSB is a component of RNA pol I transcription. *Mol. Cell* **10**, 819-829 (2002)

Bregman, D.B., Halaban, R., Van Gool, A.J., Henning, K.A., Friedberg, E.C., Warren, S.L., UV-induced ubiquitination of RNA polymerase II: a novel modification deficient in Cockayne syndrome cells. *PNAS* **93**, 11586-11590 (1996)

Brettel, K., Byrdin, M., Reaction mechanisms of DNA photolyase. *Curr. Opin. Struct. Biol.* **20**, 693-701 (2010)

Centore, R.C., Havens, C.G., Manning, A.L., Li, J.M., Litman Flynn, R., Tse, A., Jin, J., Dyson, N.J., Walter, J.C., Zou, L., CRL4(Cdt2)-mediated destruction of the histone methyltransferase Set8 prevents premature chromatin compaction in S phase. *Mol. Cell* **40**, 22-33 (2010)

Cha, J., and Mobashery, S., Lysine N-ζ-decarboxylation in the BlaR1 protein from *Staphylococcus aureus* at the root of its function as an antibiotic sensor. *J. Am. Chem. Soc.* **129**, 3834-3835 (2007)

Charlet-Berguerand, N., Feuerhahn, S., Kong, S.E., Ziserman, H., Conaway, J.W., Conaway, R., Egly, J.M., RNA polymerase II bypass of oxidative DNA damage is regulated by transcription elongation factors. *EMBO J.* **25**, 5481-5491 (2006)

Christiansen, M., Stevnsner, T., Modin, C., Martensen P.M., Brosh, R.M., Bohr, V.A., Functional consequences of mutations in the conserved SF2 motifs and post-translational phosphorylation of the CSB protein. *Nucleic Acids Res.* **31**, 963-973 (2003)

Citterio, E., Rademakers, S., Van der Horst, G.T.J., Van Gool, A.J., Hoeijmakers, J.H.J., Vermeulen, W., Biochemical and biological characterization of wild-type and ATPase-deficient Cockayne syndrome B repair protein. *J. Biol. Chem.* **273**, 11844-11851 (1998)

Collaborative Computational Project 4, The CCP4 suite: Programs for protein crystallography. *Acta Cryst.* **D50**, 760-763 (1994)

Cowtan, K., The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Cryst.* **D62**, 1002-1011 (2006)

Daniels, D.S., Woo, T.T., Luu, K.X., Noll, D.M., Clarke, N.D., Pegg, A.E., Tainer, J.A., DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nat. Struct. Mol. Biol.* **11**, 71-720 (2004)

Dattagupta, J.K., Podder, A., Chakrabarti, C., Sen, U., Mukhopadhyay, D., Dutta, S.K., Singh, M., Refined crystal structure (2.3 Å) of a double-headed winged bean α-chymotrypsin inhibitor and location of its second reactive site. *Proteins* **35**, 321-331 (1999)

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., Snoeyink, J., Richardson, J.S., Richardson, D.C., MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, 375-383 (2007)

Dianov, G., Bischoff, C., Sunesen, M., Bohr, V.A., Repair of 8-oxoguanine in DNA is deficient in Cockayne syndrome group B cells. *Nucleic Acids Res.* **27**, 1365-1368 (1999)

D'Errico, M., Parlanti, E., Teson, M., Degan, P., Lemma, T., Calcagnile, A., Iavarone, I., Jaruga, P., Ropolo, M., Pedrini, A.M., Orioli, D., Zambruno, G., Dizdaroglu, M., Stefanini, M., Dogliotti. E., The role of CSA in the response to oxidative DNA damage in human cells. *Oncogene* **26**, 4336-4343 (2007)

Diederichs, K., and Karplus, P.A., Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* **4**, 269-275 (1997)

Dümmler, A., Lawrence, A.M., de Marco, A., Simplified screening for the detection of soluble fusion constructs expressed in *E. coli* using a modular set of vectors. *Microb. Cell Fact.* **4**, 34 (2005)

Egly, J.M., Coin, F., A history of TFIIH: Two decades of molecular biology on a pivotal transcription/ repair factor. *DNA repair* **10**, 714-721 (2011)

Eker, A.P.M., Quayle, C., Chaves, I., Van der Horst, G.T.J., Direct DNA damage reversal: elegant solutions for nasty problems. *Cell. Mol. Life Sci.* **66**, 969-980 (2009)

Emsley, P., Cowtan, K., Coot: Model-building tools for molecular graphics. *Acta Cryst.* **D60**, 2126-2132 (2004)

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., Features and development of Coot. *Acta Cryst.* **D66**, 486-501 (2010)

Evans, P., Scaling and assessment of data quality. *Acta Cryst.* **D62**, 72-82 (2006)

Fischer, E.S., Scrima, A., Böhm, K., Matsumoto, S., Lingaraju, G.M., Faty M., Yasuda, T., Cavadini, S., Wakasugi, M., Hanaoka, F., Iwai, S., Gut, H., Sugasawa, K., Thomä, N., The molecular basis of CRL4-DDB2/CSA ubiquitin ligase architecture, targeting, and activation, *Cell* **147**, 1024-1039 (2011)

Fousteri, M., Vermeulen, W., van Zeeland, A.A., Mullenders, L.H.F., Cockayne syndrome A and B proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled RNA polymerase II *in vivo*. *Mol. Cell* **23**, 471-482 (2006)

Fousteri, M., Mullenders, L.H.F., Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. *Cell Res.* **18**, 73-84 (2008)

Freyer, G.A., Davey, S., Ferrer, J.V., Martin, A.M., Beach, D., Doetsch, P.W., An alternative eurkaryotic DNA excision repair pathway. *Mol. Cell. Biol.* **15**, 4572-4577 (1995)

Fujihashi, M., Numoto, N., Kobayashi, Y., Mizushima, A., Tsujimura, M., Nakamura, A., Kawarabayasi, Y., Miki, K., Crystal structure of archaeal photolyase from *Sulfolobus tokodaii* with two FAD molecules: implication of a novel light-harvesting cofactor. *J. Mol. Biol.* **365**, 903-910 (2007)

Gahloth, D., Selvakumar, P., Shee, C., Kumar, P., Sharma, A.K., Cloning, sequence analysis and crystal structure determination of a miraculin-like protein from Murraya koenigii. *Arch. Biochem. Biophys.* **494**, 15-22 (2010)

Gillet, L.C.J., Schärer, O.D., Mechanisms of mammalian global genome nucleotide excision repair. *Chem. Rev.* **106**, 253-276 (2006)

Golema-Kotra, D., Cha, J.Y., Meroueh, S.O., Vakulenko, S.B., and Mobashery, S., Resistance of β-lactam antibiotics and its mediation by the sensor domain of the transmembrane BlaR signalling pathway in *Staphylococcus aureus*. *J. Biol. Chem.* **278**, 18419-18425 (2003)

Golemi, D., Maveyraud, L., Vakulenko, S., Samama, J., and Mobashery, S., Critical involvement of a carbamylated lysine in catalytic function of class D β-lactamases. *PNAS* **98**, 14280-14285 (2001)

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

Groisman, R., Polanowska, J., Kuraoka, I., Sawada, J., Saijo, M., Drapkin, R., Kisselev, A.F., Tanaka, K., Nakatani, Y., The ubiquitin ligase activity in the DDB2 and CSA complexes is differentially regulated by the COP9 signalosome in response to DNA damage. *Cell* **113**, 357-367 (2003)

Groisman, R., Kuraoka, I., Chevallier, O., CSA-dependent degradation of CSB by the ubiquitin-proteasome pathway establishes a link between complementation factors of the Cockayne syndrome. *Genes dev.* **20**, 1429-1434 (2006)

Hauk, G., Bowman, G.D., Structural insights into regulation and action of SWI2/SNF2 ATPases. *Curr. Opin. Struct. Biol.* **21**, 719-727 (2011)

Hashimoto, S., Suga, T., Kudo, E., Ihn, H., Uchina, M., Tateishi, S., Adult-Onset Neurological Degeneration in a Patient with Cockayne Syndrome and a Null Mutation in the *CSB* Gene. *J. Invest. Dermatol.* **128**, 1597-1599 (2008)

Henning, K.A., Li, L., Iyer, N., McDaniel, L.D., Reagan, M.S., Legerski, R., Schultz, R.A., Stefanini, M., Lehmann, A.R., Mayne, L.V., Friedberg, E.C., The Cockayne Syndrome Group A gene encodes a WD repeat protein that interacts with CSB protein and a subunit of RNA Polymerase II TFIIH. *Cell* **82**, 555-564 (1995)

Higa, L.A., Wu, M., Ye, T., Kobayashi, R., Sun, H., Zhang, H., CUL4-DDB1 ubiquitin ligase interacts with multiple WD40-repeat proteins and regulates histone methylation. *Nat. Cell Biol.* **8**, 1277-1283 (2006)

Horibata, K., Iwamoto, Y., Kuraoka, Jaspers, N.G.J., Kurimasa, A., Oshimura, M., Ichihashi, M., Tanaka, K., Complete absence of Cockayne syndrome group B gene product gives rise to UV-sensitive syndrome but not Cockayne syndrome. *PNAS* **101**, 15410-15415 (2004)

Hosfield, D.J., Guan, Y., Haas, B.J., Cunningham, R.P., Tainer, J.A., 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 (1999)

Imam, S.Z., Indig, F.E., Cheng, W.H., Saxena, S.P., Stevnsner, T., Kufe, D., Bohr, V.A., Cockayne syndrome protein B interacts with and is phosphorylated by c-Abl tyrosine kinase. *Nucleic Acids Res.* **35**, 4941-4951 (2007)

Iovine, B., Ianella, M.L., Bevilacqua, M.A., Damage-specific DNA binding protein 1 (DDB1): a protein with a wide range of functions. *Int. J. Biochem. Cell Biol.* **43**, 1664-1667 (2011)

Iwai, S., Chemical synthesis of oligonucleotides containing damaged bases for biological studies. *Nucleosides Nucleotides Nucleic Acids* **25**, 561-582. (2006)

Jackson, S., Xiong, Y., CRL4s: the CUL4-RING E3 ubiquitin ligases. *Trends Biochem. Sci.* **34**, 562-570 (2009)

Jin, J., Arias, E.E., Chen, J., Harper, J.W., Walter, J.C., A family of diverse Cul4-Ddb1-interacting proteins includes Cdt2, which is required for S phase destruction of the replication factor Cdt1. *Mol. Cell* **23**, 709-721 (2006)

Kamiuchi, S., Saijo, M., Citterio, E., De Jager, M., Hoeijmakers, J.H.J., Tanaka, K., Translocation of Cockayne syndrome group A protein to the nuclear matrix: possible relevance to transcription-coupled DNA repair. *PNAS* **99**, 201-206 (2002)

Kanno, S., Iwai, S., Takao, M., Yasui, A., Repair of apurnic/ apyrimidinic sites by UV damage endonuclease; a repair protein for UV and oxidative damage. *Nucleic Acids Res.* **27**, 3096-3103 (1999)

Kathe, S.D., Shen, G.P., Wallace, S.S., Single-stranded breaks in DNA but not oxidative DNA base damages block transcriptional elongation by RNA polymerase II in HeLa cell nuclear extracts. *J. Biol. Chem.* **279**, 18511-18520 (2004)

Kaur, B., Avery, A.M., Doetsch, P.W., Expression, purification and characterization of ultraviolet DNA endonuclease from *Schizosaccharomyces pombe*. *Biochemistry* **37**, 11599-11604 (1998)

Kaur, B., Doetsch, P.W., Ultraviolet damage endonuclease (Uve1p): a structure and strand-specific DNA endonuclease. *Biochemistry* **39**, 5788-5796 (2000)

Kennedy, A.R., Chemopreventive agents: protease inhibitors. *Pharmacol. Ther.* **78**, 167-209 (1998)

Kim, J.K., Choi, B.S., The solution structure of DNA duplex-decamer containing the (6-4) photoproduct of thymidylyl (3′→5′) thymidine by NMR and relaxation matrix refinement. *Eur. J. Biochem.* **228**, 849-854 (1995)

Kisker, C., When one protein does the job of many. *Structure* **15**, 1163-1165 (2007)

Krauchenco, S., Pando, S.C., Marangoni, S., Polikarpov, I., Crystal structure of the Kunitz (STI)-type inhibitor from Delonix regia seeds. *Biochem. Biophys. Res. Commun.* **312**, 1303-1308 (2003)

Krissinel, E., Henrick, K., Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Cryst.* **D60**, 2256-2268 (2004)

Krissinel, E., Henrick, K., Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774-797 (2007)

Kunz, C., Saito, Y., Schär, P., Mismatched repair: variations on a theme. *Cell. Mol. Life Sci.* **66**, 1021-1038 (2009)

Kyng, K.J., May, A., Brosh Jr, R.M., Cheng, W.H., Chen, C., Becker, K.G., Bohr, V.A., The transcriptional response after oxidative stress is defective in Cockayne syndrome group B cells. *Oncogene* **22**, 1135-1149 (2003)

Laskowski, R.A., MacAuthor, M.W., Moss, D.S., and Thornton, J.M., PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291 (1993)

Laugel, V., Dalloz, C., Stary, A., Cormier-Daire, V., Desguerre, I. Renouil, M., Fourmaintraux, A., Velez-Cruz, R., Egly, J.M., Sarasin, A., Dollfus, H., Deletion of 5′ sequences of the CSB gene provides insight into the pathophysiology of Cockayne syndrome. *Eur. J. Hum. Genet.* **16**, 320-327 (2008)

Laugel, V., Dalloz, C., Durand, M., Sauvanaud, F., Kristensen, U., Vincent, M.C., Pasquier, L., Odent, S., Cormier-Daire, V., Gener, B., Tobias, E.S., Tolmie, J.L., Martin-Coignard, D., Drouin-Garraud, V., Heron, D., Journel, H., Raffo, E., Vigneron, J., Lyonnet, S., Murday, V., Gubser-Mercati, D., Funalot, B., Brueton, L., Sanchez Del Pozo, J., Muñoz, E., Gennery, A.R., Salih, M., Noruzinia, M., Prescott, K., Ramos, L., Stark, Z., Fieggen, K., Chabrol, B., Sarda, P., Edery, P., Bloch-Zupan, A., Fawcett, H., Pham, D., Egly, J.M., Lehmann, A.R., Sarasin, A., Dollfus, H. Mutation update for the CSB/ERCC6 and CSA/ERCC8 genes involved in Cockayne syndrome. *Human mutation* **31**, 113-126 (2010)

Lee, J., Zhou, P., DCAFs, the missing link of the CUL4-DDB1 ubiquitin ligase. *Mol. Cell* **26**, 775-780 (2007)

Lehmann, A.R., DNA repair-deficient diseases, xeroderma pigmentosum, Cockayne syndrome and trichothidystrophy. *Biochimie* **85**, 1101-1111 (2003)

Leslie, A.G., Integration of macromolecular diffraction data. *Acta Cryst.* **D55**, 1696-1702 (1999)

Li, J., Cross, J.B., Vreven, T., Meroueh, S.O., and Schlegel, H.B., Lysine carboxylation in proteins: OXA-10 β-lactamase. *Proteins* **61** 246-257 (2005)

Li, T., Chen, X., Garbutt, K.C., Zhou, P., Zheng, N., Structure of DDB1 in complex with a paramyxovirus V protein: viral hijack of a propeller cluster in ubiquitin ligase. *Cell* **124**, 105-117 (2006)

Li, T., Robert, E.I., van Breugel, P.C., Strubin, M., Zheng, N. A promiscuous α-helical motif anchors viral hijackers and substrate receptors to the CUL4-DDB1 ubiquitin ligase machinery. *Nat. Struct. Mol. Biol.* **17**, 105-111 (2010)

Licht, C.L., Stevsner, T., Bohr, V.A., Cockayne syndrome group B cellular and biochemical functions. *Am. J. Hum. Genet.* **73**, 1217-1239 (2003)

Lommatzsch, S.T., Aris, R., Genetics of cystic fibrosis. *Semin. Respir. Crit. Care Med.* **30**, 531-538 (2009)

Long F., Vagin A.A., Young P., Murshudov G.N., BALBES: a molecular-replacement pipeline. *Acta Cryst* **D64**, 125-132 (2008)

Matsunaga, T., Mu, D., Park, C.H., Reardon, J.T., Sancar, A., Human DNA repair excision nuclease. *J. Biol. Chem.* **270**, 20862-20869 (1995)

Mayne, L.V., Lehmann, A.R., Failure of RNA synthesis to recover after UV-irradiation: an early defect in cells from individuals with Cockayne's syndrome and xeroderma pigmentosum. *Cancer Res.* **42**, 1473-1478 (1982)

McCready, S., Carr, A.M., Lehmann, A.R., Repair of cyclobutane dimers and 6-4 photoproucts in the fission yeast *Schizosaccharomyces pombe*. *Mol. Microbiol.* **10**, 885-890 (1993)

Mellon, I., Spivak, G., Hanawalt, P.C., Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell* **51**, 241-249 (1987)

Micheelsen, P.O., Vevodova, J., De Maria, L., Ostergaard, P.R., Friis, E.P., Wilson, K., Skjot, M., Structural and mutational analyses of the interaction between the barley alpha-amylase/subtilisin inhibitor and the subtilisin savinase reveal a novel mode of inhibition. *J. Mol. Biol.* **380**, 681-690 (2008)

Min, J.H., Pavletich, N.P., Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* **449**, 570-576 (2007)

Moll, R., Schäfer, G., Chemiosmotic H^+ cycling across the plasma membrane of the thermoacidophilic archaebacterium *Sulfolobus acidocaldarius*. *FEBS lett.* **232**, 359-363 (1988)

Müller, M., Carell, T., Structural biology of DNA photolysases and cryptochromes. *Curr. Opin. Struct. Biol.* **19**, 277-185 (2009)

Murshudov, G.N., Vagin, A.A., Dodson, E.J., Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.* **D53**, 240-255 (1997)

Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., Vagin, A.A., REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst.* **D67**, 355-367 (2011)

Nance, M.A., Berry, S.A., Cockayne syndrome: review of 140 cases. *Am. J. Med. Genet.* **42**, 68-84 (1992)

Nardo, T., Oneda, R., Spivak, G., Vaz, B., Mortier, L., Thomas, P., Orioli, D., Laugel, V., Stary, A., Hanawalt, P.C., Sarasin, A., Stefanini, M., A UV-sensitive syndrome patient with a specific CSA mutation reveals separable roles for CSA in response to UV and oxidative DNA damage. *PNAS* **106**, 6209-6214 (2009)

Newman, J.C., Bailey, A.D., Weiner, A.M., Cockayne syndrome group B protein (CSB) plays a general role in chromatin maintenance and remodeling. *PNAS* **103** 9613-9618 (2006)

Newman, J.C., Bailey, A.D., Fan, H.Y., Pavelitz, T., Weiner, A.M., An abundant evolutionarily conserved CSB-PiggyBac Fusion protein expressed in Cockayne Syndrome. *PLoS Genet.* **4**, e1000031 (2008)

Nouspikel, T. So DNA repair really is that important? *Cell. Mol. Life Sci.* **66**, 965-967 (2009a)

Nouspikel, T., Nucleotide excision repair: variations on versatility. *Cell Mol. Life Sci.* **66**, 994-1009 (2009b)

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

Oliva, M.L.V., Silva, M.C.C., Sallai, R.C., Brito, M.V., Sampaio, M.U., A novel subclassification for Kunitz proteinase inhibitors from leguminous seeds. *Biochimie* **92**, 1667-1673 (2010)

Olma, M.H., Roy, M., Le Bihan, T., Sumara, I., Maerki, S., Larsen, B., Quadroni, M., Peter, M., Tyers, M., Pintard, L., An interaction network of the mammalian COP9 signalasome identifies Dda1 as a core subunit of multiple Cul4-based E3 ligases., *J. Cell Sci.* **122**, 1035-1044 (2009)

Padilla, J.E., Yeates, T.O., A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning. *Acta Cryst.* **D59**, 1124-1130 (2003)

Pardo, B., Gómez-González, B., Aguilera, A., DNA double-strand break repair: how to fix a broken relationship. *Cell. Mol. Life Sci.* **66**, 1039-1056 (2009)

Parikh, S.S., Walcher, G., Jones, G.D., Slupphaug, G., Krokan, H.E., Blackburn, G.M., Tainer, J.A., Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *PNAS* **97**, 5083-5088 (2000)

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

Paspaleva, K., Moolenaar, G.F., Goosen, N., Damage recognition by UV damage endonuclease from *Schizosaccharomyces pombe*. *DNA repair* **8**, 600-611 (2009)

Perrakis, A., Morris, R., and Lamzin, V.S., Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458-463 (1999)

Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P., Murshudov, G.N, Cohen, S., Perrakis, A., and Noble, M., Developments in the CCP4 molecular graphics project. *Acta Cryst.* **D60**, 2288-2294 (2004)

Popov, A.N., and Bourenkov, G.P., Choice of data-collection parameters based on statistical modeling. *Acta Cryst.* **D59**, 1145-1153 (2003)

Pouvreau, L., Gruppen, H., Piersma, S.R., Van den Broek, L.A.M., Van Koningsveld, G.A., Voragen, A.G.J., Relative abundance and inhibitory distribution of protease inhibitors in potato juice from cv. Elkana. *J. Agric. Food Chem.* **49**, 2864-2874 (2001)

Pouvreau, L., Gruppen, H., Van Koningsveld, G.A., Van den Broek, L.A.M., Voragen, A.G.J., The most abundant protease inhibitor in potato tuber (cv. Elkana) is a serine protease inhibitor from the Kunitz family. *J. Agric. Food Chem.* **51**, 5001-5005 (2003)

Pouvreau, L., Gruppen, H., Van Koningsveld, G.A., Voragen, A.G.J., Conformational stability of the potato serine protease inhibitor group. *J. Agric. Food Chem.* **53**, 3191-3196 (2005)

Proietti-De-Santis, L., Drané, P., Egly, J.M., Cockayne syndrome B protein regulates the transcriptional program after UV irradiation. *EMBO J.* **25**, 1915-1923 (2006)

Rabl, J., Leibundgut, M., Ataide, S.F., Haag, A., Ban, N., Crystal structure of the eukaryotic 40 S subunit in complex with initiation factor 1. *Science* **331**, 730-736 (2011)

Rich, T., Allen, R.L., Wyllie, A.H. Defying death after DNA damage. *Nature* **407**, 777-783 (2000)

Robertson, A.B., Klungland, A., Rognes, T., Leiros, I., Base excision repair: the long and the short of it. *Cell. Mol. Life Sci.* **66**, 981-993 (2009)

Schalow, B.J., Courcelle, C.T., Courcelle, J., Mfd is required for rapid recovery following UV-induced DNA damage but not oxidative DNA damage in *Escherichia coli*. *J. Bacteriol.* **194**, 2637-2645 (2012)

Schuetz, A., Allali-Hassani, A., Martin, F., Loppnau, P., Vedadi, M., Bochkarev, A., Plotnikov, A.N., Arrowsmith, C.H., Min, J., Structural basis for molecular recognition and presentation of histone H3 by WDR5. *EMBO J.* **25**, 4245-4252 (2006)

Schwechheimer, C., The COP9 signalosome (CSN): an evolutionary conserved proteolysis regulator in eukaryotic development. *Biochim. Bhiophys. Acta* **1695** 45-54 (2004)

Scrima, A., Konickova, R., Czyzewski, B.K., Kawasaki, Y., Jeffrey, P.D., Groisman, R., Nakatani, Y., Iwai, S., Pavletich, N.P., Thomä, N.H., Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex. *Cell* **135**, 1213-1223 (2008)

Scrima, A., Fischer, E.S., Lingaraju, G.M., Böhm, K., Cavadini, S., Thomä, N.H., Detecting UV-lesions in the genome: the modular CRL4 ubiquitin ligase does it best. *FEBS lett.* **585**, 2818-2825 (2011)

Selby, C.P., Witkin, E.M., Sancar, A., *Escherichia coli mfd* mutant deficient in "mutation frequency decline" lacks strand-specific repair: *in vitro* complementation with purified coupling factor. *PNAS* **88**, 11547-11578 (1991)

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

Shibata, H., Hara, S., Ikenaka, T., Amino acid sequence of winged bean (Psophocarpus tetragonolobus (L.) DC.) chymotrypsin inhibitor, WCI-3. *J. Biochem.* **104**, 537-543 (1998)

Sidik, K., Lieberman, H.B., and Freyer, G.A., Repair of DNA damaged by UV-light and ionizing-radiation by cell-free-extracts prepared from *Schizosaccharomyces pombe*. *PNAS* **89**, 1212-1216 (1992)

Song, H.K., Suh, S.W., Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from *Erythrina caffra* and tissue-type plasminogen activator. *J. Mol. Biol.* **275**, 347-363 (1998)

Stevnsner, T., Muftuoglu, M., Aamann, M.D., Bohr, V.A., The role of Cockayne Syndrome group B (CSB) protein in base excision repair and aging. *Mech. Ageing Dev.* **129**, 441-448 (2008)

Stirnimann, C.U., Petsalaki, E., Russell, R.B., Müller, C.W., WD40 proteins propel cellular networks. *Trends Biochem. Sci.* **35**, 565-574 (2010)

Strong, M., Sawaya, M.R., Wang, S., Philips, M., Cascio, D., Eisenberg, D., Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *PNAS* **103**, 8060-8065 (2006)

Sugasawa, K., Okuda, Y., Saijo, M., Nishi, R., Matsuda, N., Chu, G., Mori, T., Iwai, S., Tanaka, K., Tanaka, K., Hanaoka, F., UV-induced ubiquitylation of XPC protein mediated by UV-DDB-ubiquitin ligase complex. *Cell* **121**, 387-400 (2005)

Sugasawa, K., Akagi, J., Nishi, R., Iwai, S., Hanaoka, F., Two-step recognition of DNA damage for mammalian nucleotide excision repair: directional binding of the XPC complex and DNA strand scanning. *Mol. Cell* **36**, 642-653 (2009)

Takao, M., Yonemasu, R., Yamamoto, K., Yasui, A., Characterization of a UV endonuclease gene from the fission yeast *Schizosaccharomyces pombe* and its bacterial homolog. *Nucleic Acids Res.* **24**, 1267-1271 (1996)

Theobald, D.L., Wuttke, D.S., THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures. *Bioinformatics* **22**, 2171-2172 (2006)

Thoden, J.B., Philips, G.N., Neal, T.M., Raushel, F.M., and Holden, H.M., Molecular structure of dihydroorotase: A paradigm for catalysis through the use of a binuclear metal center. *Biochem.* **40**, 6989-6997 (2001)

Thoden, J.B., Marti-Arbona, R., Raushel, F.M., and Holden, H.M., High-resolution X-ray structure of isoapartyl dipeptidase from *Escherichia coli*. *Biochem.* **42**, 4874-4882 (2003)

Thomassen, E.A.J., Pouvreau, L., Gruppen, H., Abrahams, J.P., Crystallization and preliminary X-ray crystallographic studies on a Kunitz-type potato serine protease inhibitor. *Acta Cryst.* **D60**, 1464-1466 (2004)

Thompson, K.E., Wang, Y., Madej, T., Bryant, S.H., Improving protein structure similarity searches using domain boundaries based on conserved sequence information. *BMC Struct. Biol.* **19**, 33 (2009)

Thorslund, T., Von Kobbe, C., Harrigan, J.A., Indig, F.E., Christiansen, M., Stevnsner, T., Bohr, V.A., Cooperation of the Cockayne Syndrome group B protein and poly(ADP-ribose) polymerase 1 in the response to oxidative stress. *Mol. Cell. Biol.* **25**, 7625-7636 (2005)

Tornaletti, S., Transcription-coupled DNA repair: directing your effort where it's most needed. *Cell. Mol. Life Sci.* **66**, 1010-1020 (2009)

Troelstra, C., Landsvater, R.M., Wiegant, J., Van der Ploeg, M., Viel, G., Buys, C.H.C.M., Hoeijmakers, J.H.J., Localization of the nucleotide excision repair gene *ERCC6* to human chromosome 10q11-q21. *Genomics* **12**, 745-749 (1992)

Tuo, J., Jaruga, P., Rodriguez, H., Bohr, V.A., Dizdaroglu, M., Primary fibroblasts of Cockayne syndrome patients are defective in cellular repair of 8-hydroxyguanine and 8-hydroxyadenine resulting from oxidative stress. *FASEB J.* **17**, 668-674 (2003)

Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022-1025 (1997)

Vagin, A., Teplyakov, A., An approach to multi-copy search in molecular replacement. *Acta Cryst.* **D56**, 1622-1624 (2000)

Valueva, T.A., Revina, T.A., Mosolov ,V.V., Mentele, R., Primary structure of potato kunitz-type serine proteinase inhibitor. *Biol. Chem.* **381**, 1215-1221 (2000)

Valueva, T.A. & Mosolov, V.V., Role of inhibitors of proteolytic enzymes in plant defense against phytopathogenic microorganisms. *Biochemistry* **69**, 1305-1309 (2004)

Van den Boom, V., Citterio, E., Hoogstraten, D., Zotter, A., Egly, J.M., Van Cappellen, W.A., Hoeijmakers, J.H.J., Houtsmuller, A.B., Vermeulen, W., DNA damage stabilizes interaction of CSB with the transcription elongation machinery. *J. Cell Biol.* **166** 27-36 (2004)

Van Gool, A.J., Citterio, E., Rademakers, S., Van Os, R., Vermeulen, W., Constantinou, A., Egly, J.M., Bootsma, D., Hoeijmakers, J.H.J., The Cockayne syndrome B protein, involved in transcription-coupled DNA repair, resides in an RNA polymerase II-containing complex. *EMBO J.* **16**, 5955-5965 (1997)

Van Hoffen, A., Venema, J., Meschini, R.,Van Zeeland, A.A., Mullenders, L.H.F., Transcription-coupled repair removes both cyclobutane pyrimidine dimers and 6-4 photoproducts with equal efficiency and in a sequential way from transcribed DNA in xeroderma pigmentosum group C fibroblasts. *EMBO J.* **14**, 360-367 (1995)

Verhoeven, E.E., Van Kesteren, M., Turner, J.J., Van der Maarel, G.A., Van Boom, J.H., Moolenaar, G.F. and Goosen, N., 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 (2002)

Walker, J.R., Corpina, R.A., Goldberg, J., Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* **412**, 607-614 (2001)

Warren, J.J., Forsberg, L.J., Beese, L.S., The structural basis for the mutagenicity of O^6-methyl-guanine lesions. *PNAS* **103**, 19701-19706 (2006)

Warren, J.J., Pohlhaus, T.J., Changela, A., Iyer, R.R., Modrich, P.L., Beese, L.S., Structure of the human MutSα DNA lesion recognition complex. *Mol. Cell* **26**, 579-592 (2007)

Weiss, M.S., and Hilgenfeld, R. On the use of the merging R factor as a quality indicator for X-ray data. *J. Appl. Cryst.* **30**, 203-205 (1997)

Winn, M.D, Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., McNicholas, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A., Wilson, K.S., Overview of the CCP4 suite and current developments. *Acta Cryst* **D67**, 235-242 (2011)

Wittschieben, B.O., Iwai, S., Wood, R.D., DDB1-DDB2 (Xeroderma Pigmentosum Group E) protein complex recognizes a cyclobutane pyrimidine dimer, mismatches, apurinic/ apyrimidinic sites, and compound lesions in DNA. *J. Biol. Chem.* **280**, 39982-39989 (2005)

Wong, H.K., Muftuoglu, M., Beck, G., Imam, S.Z., Bohr, V.A., Wilson III, D.M., Cockayne syndrome B protein stimulates apurinic endonuclease 1 activity and protects against agents that introduce base excision repair intermediates. *Nucleic Acids Res.* **35**, 4103-4113 (2007)

Wu, G., Xu, G., Schulman, B.A., Jeffrey, P.D., Harper, J.W., Pavletich, N.P., Structure of a β-TrCP1-Skp1-β-catenin complex: destruction motif binding and lysine specificity of the SCF(β-TrCP1) ubiquitin ligase. *Mol. Cell* **11**, 1445-1456 (2003)

Xie, Z.W., Luo, M.J., Xu, W.F., Chi, C.W., Two reactive site locations and structure-function study of the arrowhead proteinase inhibitors, A and B, using mutagenesis. *Biochemistry* **36**, 5846-5852 (1997)

Xu, C., Min, J., Structure and function of WD40 proteins. *Protein Cell* **2**, 202-214 (2011)

Yajima, H., Takao, M., Yasuhira, S., Zhao, J.H., Ishii, C., Inoue, H., Yasui, A., A eukaryotic gene encoding an endonuclease that specifically repairs DNA damaged by ultraviolet light. *EMBO J.* **14**, 2393-2399 (1995)

Yasui, A., McCready, S.J., Alternative repair pathways for UV-induced DNA damage. *Bioessays* **20**, 291-297 (1998)

Yonemasu, R., McCready, S.J., Murray, J.M., Osman, F., Takao, M., Yamamoto, K., Lehmann, A.R., Yasui, A., Characterization of the alternative repair pathway of UV-damaged DNA in *Schizosaccharomyces pombe*. *Nucleic Acids Res.* **25**, 1553-1558 (1997)

Yoon, J.H., Swiderski, P.M., Kaplan, B.E., Takao, M., Yasui, A., Shen, B., Pfeifer, G.P., Processing of UV damage in vitro by FEN-1 proteins as part of an alternative DNA excision repair pathway. *Biochemistry* **38**, 4809-4817 (1999)

Yoon, J.H., Lee, C.S., O'Connor, T.R., Yasui, A., Pfeifer, G.P., The DNA damage spectrum produced by stimulated sunlight. *J. Mol. Biol.* **299**, 681-693 (2000)

Zhong, W., Feng, H., Santiago, F.E., Kipreos, E.T., CUL-4 ubiquitin ligase maintains genome stability by restraining DNA-replication licensing. *Nature* **423**, 885-889 (2003)

Zhong, Y., Carmella, S.G., Upadhyaya, P., Hochalter, J.B., Rauch, D., Oliver, A., Jensen, J., Hatsukami, D., Wang, J., Zimmerman, C., Hecht, S.S. Immediate consequences of cigarette smoking: rapid formation of polycyclic aromatic hydrocarbon diol epoxides. *Chem. Res. Toxicol.* **24**, 246-252 (2011)

List of publications

- Meulenbroek, E.M., Paspaleva, K., Thomassen, E.A., Abrahams, J.P., Goosen, N., Pannu, N.S., Involvement of a carboxylated lysine in UV damage endonuclease, *Protein Sci.* **18**, 549-558 (2009)
- Meulenbroek, E.M., Pannu, N.S., Overproduction, purification, crystallization and preliminary X-ray diffraction analysis of Cockayne syndrome protein A in complex with DNA damage binding protein 1, *Acta Cryst.* **F68**, 45-48 (2012)
- Liu, Z., Galli, F., Waterreus, W.J., Meulenbroek, E.M., Koning, R.I., Lamers, G.E., Olsthoorn, R.C., Pannu, N.S., Oosterkamp, T.H., Koster, A.J., Dame, R.T., Abrahams, J.P., Single-walled carbon nanotubes as scaffolds to concentrate DNA for the study of DNA-protein interactions, *Chemphyschem* **13**, 1569-1575 (2012)
- Meulenbroek, E.M., Thomassen, E.A.J., Pouvreau, L., Abrahams, J.P., Gruppen, H., Pannu, N.S., Structure of a post-translationally processed heterodimeric double-headed Kunitz-type serine protease inhibitor from potato, *Acta Cryst.* **D68**, 794-799 (2012)
- Bashir, Q., Meulenbroek, E.M., Volkov, A.N., Pannu, N.S., Ullmann, G.M., Ubbink, M., Engineering specificity in the non-physiological complex of horse cytochrome c and yeast cytochrome c peroxidase by a single conserved mutation. (manuscript in preparation)
- Meulenbroek, E.M., Vrouwe, M.G., Mullenders, L.H.F., Pannu, N.S., Insights into Transcription-Coupled Repair from the crystal structure of Cockayne Syndrome protein A. (manuscript in preparation)
- Meulenbroek, E.M., Peron Cane, C., Jala, I., Iwai, S., Moolenaar, G.F., Goosen, N., Pannu, N.S., UV damage endonuclease employs a novel dual-dinucleotide flipping mechanism to recognize and incise different types of damaged DNA. (manuscript in preparation)

Oral presentations

1. Lunteren, the Netherlands, *NWO study group crystallography meeting*, 2009.
2. Rio de Janeiro, Brazil, *Brazil school for single particle cryo-electron microscopy*, 2010.
3. Amsterdam, the Netherlands, *Structural bioinformatics meets structural biology meeting*, 2011.
4. Maarssen, the Netherlands, *CHAINS meeting*, 2011.

Poster presentations

1. Lunteren, the Netherlands, *NWO study group proteins meeting*, 2008.
2. Hamburg, Germany, *Protein Expression, Purification and Crystallization course*, 2008.
3. Leiden, the Netherlands, *Reedijk symposium*, 2011, best poster award.

Courses

1. Hamburg, Germany, *Protein Expression, Purification and Crystallization course*, 2008.
2. Granada, Spain, *International School on Biological Crystallization*, 2009.
3. Erice, Italy, *Structure and function from macromolecular crystallography*, 2010.
4. Rio de Janeiro, Brazil, *Brazil school for single particle cryo-electron microscopy in Rio de Janeiro*, 2010.

Summary

DNA carries our genetic information and hence its integrity is of the utmost importance. However, there are many influences from inside the cell, like free oxygen radicals produced during metabolism, and from outside the cell, like UV-irradiation, that result in damage in the DNA. If left unrepaired, these lesions can lead to cell death and mutations, the latter potentially resulting in cancer. To prevent these deleterious effects of DNA damage, several DNA repair pathways have evolved. The major pathways are described in **Chapter 1**, where insights into these pathways gained from structural biology are emphasized. In this thesis, research into components of Transcription-Coupled Nucleotide Excision Repair (TC-NER) and UV Damage Endonuclease repair (UVDE) are described, with the most important findings listed below.

TC-NER is a conserved DNA repair pathway that removes DNA damages from the transcribed strand of actively transcribed genes. It is a subpathway of Nucleotide Excision Repair (NER) that removes DNA lesions by excising a 25-30 nucleotide DNA fragment containing the lesion and filling up the gap. Most proteins involved in TC-NER come from this more global pathway of NER, but two proteins are unique in humans: Cockayne Syndrome protein A and B (CSA and CSB). CSA is the substrate-adaptor protein in an E3-ubiquitin ligase complex and CSB is a protein with DNA-dependent ATPase activity that probably assists in making space for repair around the site of the damage. Mutations in either CSA and CSB lead to the severe human disorder Cockayne Syndrome. This disease is characterized by, amongst others, premature aging and neurodevelopmental abnormalities. This shows that CSA and CSB are biologically important, but their exact function is not clear.

To gain more insight into the function of CSA and into how mutations cause disease, a detailed picture of the protein is useful. Structural prediction tools indicated that CSA probably has a seven-bladed β-propeller fold, since it has seven WD40 repeats. Hence the general shape was known, but what was not known was the exact structure with the conformation of all the loops and the position of all the side-chains that makes CSA different from other WD40 proteins. Exactly that information is needed to be able to explain the disease. Therefore, we decided to solve its crystal structure by X-ray crystallography, which is a technique that yields a picture of a protein to atomic detail.

The first steps for obtaining a crystal structure are getting the protein and then

crystallizing it. In **Chapter 2** we describe the expression of CSA in complex with its interaction partner DNA Damage Binding protein 1 (DDB1), since CSA was not soluble on its own. The complex was then purified in a three-step procedure and crystallized. The crystals diffracted anisotropically to 2.9 Å.

Chapter 3 describes the crystal structure of CSA in complex with DDB1. As expected, the overall structure of CSA showed a β-propeller fold. CSA was found to interact with DDB1 using its N-terminal helix-loop-helix motif. The disease-causing mutations could be mapped onto the structure. It was found that the mutations mostly disrupt local or global folding and probably lead to problems in either binding to CSA's substrate or binding to CSA's interaction partner DDB1. This kind of errors are, unfortunately, difficult to counteract. The most likely substrate-binding site for CSA was found to be the centre of the top face of its WD40 domain. This was seen to have a small, positively charged centre, much like substrate-binding sites for phosphorylated proteins. This suggested that CSA might recognize a phosphorylated protein or an otherwise negatively charged stretch on another protein.

The crystal structure of CSA suggested a potential substrate-binding site and residues involved in substrate binding, but this had to be verified biochemically. Experiments performed to this end are described in **Chapter 4**. Mutation of two positively charged residues (K292 and K293) in the centre of the proposed substrate-binding site resulted in a protein that could not complement for UV-sensitivity in a CS-A cell line, showing that this site is indeed important for CSA's function. The two most likely options for CSA to bind to, DNA and CSB, were then tested. CSA was found not to bind DNA at biologically significant protein-DNA ratios, virtually excluding DNA as binding partner of CSA. Pull-down experiments using a His-tagged CSA overexpressed in a CS-A cell line showed CSB being pulled-down by CSA. Moreover, in cell lines overexpressing CSA, less CSB was detected compared to cell lines without CSA. This suggests that CSB might be the substrate that CSA targets for ubiquitination.

Bacteria and lower eukaryotes have another pathway for the repair of UV-lesions: the UV Damage Endonuclease (UVDE) repair pathway. UVDE is an enzyme that can both recognize and make a 5' incision in a number of different DNA lesions, such as the UV-photoproducts Cyclobutane Pyrimidine Dimer (CPD) and 6-4 Photoproduct (6-4PP), but also abasic sites and other lesions. After the 5' incision, other enzymes complete the repair. The fact that one enzyme both recognizes different damages and incises them is rather unique, and it makes UVDE a nice, simplified model system for DNA repair.

The crystal structure of UVDE from *Thermus thermophilus* was solved in Leiden some years ago. In this structure, a post-translational modification was seen on a lysine residue near the active site. This is uncommon for a bacterial protein expressed in a foreign host and therefore we investigate the identity of the modification and its function in **Chapter 5**. With mass-spectrometry and X-ray crystallography, the modification was shown to be a carboxylation of a lysine residue. Site-directed mutagenesis in combination with activity assays and crystallography was used to show that the carboxylation was necessary for donating a negative charge for bind-

ing the metal ions required for the activity of UVDE. Moreover, it was seen that a positive charge at this position in the protein is actually dangerous, since it leads to rearrangements in the structure near the site where the damage in the DNA has to bind and this leads to increased incision of undamaged DNA.

Another open question was how can UVDE, as a single enzyme, recognize different DNA damages. To answer this question, we cocrystallized UVDE with 6-4PP containing DNA. In the crystal structure (described in **Chapter 6**), we saw that UVDE has a unique mechanism for recognizing UV-damaged DNA: double two-nucleotide flipping. Not only the damage flipped is into a pocket that excludes larger damages, but also the two opposite bases are flipped into a dipurine-specific pocket, hence providing extra specificity for distorted dipyrimidines. The DNA has to be bent around 90 ° to fit into the protein; this is probably energetically rather unfavourable for undamaged DNA, hence providing extra specificity for damaged DNA where e.g. already part of the hydrogen bonding is lost. To investigate further why most UVDEs can both incise CPD and 6-4PP efficiently, we compared the UVDE homologues *Tth*UVDE (incises CPD, 6-4PP and abasic site) and *Sac*UVDE (only incises 6-4PP efficiently) both in activity and it structure. Rigidity around the damage-binding pocket in *Sac*UVDE due to a disulfide bridge, combined with the lack of positive charges on opposite sides of the DNA-binding groove that are needed to bend the DNA, probably are the causes for the narrow substrate specificity of this homologue. These hypotheses were confirmed by site-directed mutagenesis and activity assays. These results suggest that the relatively broad substrate specificity of the other UVDEs is caused by their ability to bend the damaged (hence flexible) DNA correctly in its DNA binding groove combined with a moderately flexible DNA binding pocket.

The last chapter of this thesis (**Chapter 7**) describes the crystal structure of an unrelated protein, Potato Serine Protease Inhibitor (PSPI), to show the general applicability of X-ray crystallography for gaining insight into biological systems. PSPI is a protein involved in the defense mechanism of the plant. It inhibits serine proteases such as trypsin and chymotrypsin and thus prevents predators and pathogens to feed on the plant. It was found to have a β-trefoil fold, which has a very stable core with protruding loops. The reactive site loops are commonly found in these protruding loops. The protein inhibits proteases by trapping them on these loops, because the core of the protein is too stable to fall apart. With our structure, we could determine the reactive site loops of PSPI based on comparison to other protease inhibitor structures. PSPI is special in that it is a heterodimeric, double-headed protease inhibitor: in a post-translational process six amino acids in a loop are deleted. Our structure provided a possible reason for this: it might be involved in activating the inhibitor, since taking away of this loop uncovers one of the reactive site loops.

Concluding remark

This thesis gives some examples of how X-ray crystallography can give insight into protein function. The three-dimensional picture of a protein at atomic resolution

helps in explaining the function of the protein from its overall fold and from what residues are found at what positions. With the wealth of structural information already present in the Protein Data Bank, any new structure can be interpreted by comparing it to other known structures of proteins with known function. However, the interesting bits and pieces of a protein are usually found exactly in the parts of the protein that differ from the previous structures. Interpretation of those differences is subjective and needs verification. Biochemical techniques like site-directed mutagenesis and activity assays are very well suited for this: they form a complementary source of information to structural information. As is also demonstrated in the chapters on CSA and UVDE in this thesis, not only structural biology, but especially its combination with biochemistry makes a powerful tool to investigate biological systems.

Samenvatting

DNA is de drager van ons genetisch materiaal. Het is daarom belangrijk dat de informatie in het DNA correct blijft. Door allerlei oorzaken kan er echter schade ontstaan in het DNA. Dat gebeurt zowel door processen in onze cellen, zoals door vrije zuurstofradicalen die ontstaan tijdens het metabolisme in de cel, als door invloeden van buitenaf, zoals UV-straling. Als de schade niet wordt hersteld, kan dat tot celdood en mutaties leiden. Dat laatste kan kanker tot gevolg hebben. Om deze negatieve gevolgen van DNA-schade te voorkomen, zijn er in de evolutie verschillende DNA-herstelmechanismen ontstaan. In **Hoofdstuk 1** van dit proefschrift staan de belangrijkste DNA-herstelmechanismen samengevat. De nadruk wordt hier gelegd op inzichten die zijn verkregen door middel van structuurbiologie. Het onderzoek beschreven in dit proefschrift, gaat over de herstelmechanismen Transcriptie-gekoppelde Nucleotide Excisie Reparatie (TC-NER) en UV-schade endonuclease herstel (UVDE). De belangrijkste resultaten van dit promotieonderzoek staan hieronder samengevat.

TC-NER is een geconserveerd DNA-herstelmechanisme dat schades herstelt in de DNA-regio's waar actieve transcriptie plaatsvindt. Het is een submechanisme van Nucleotide Excisie Reparatie (NER), waarbij DNA-schade wordt hersteld door het uitknippen van een stuk DNA van 25-30 nucleotiden met de schade erin en door het vervolgens opvullen van het gat. De meeste eiwitten die zijn betrokken bij TC-NER komen uit het meer algemene mechanisme van NER, maar er zijn twee eiwitten uniek voor TC-NER: Cockayne Syndroom eiwit A en B (CSA en CSB). CSA is het substraat-bindende eiwit in een E3-ubiquitine ligase complex en CSB is een eiwit dat waarschijnlijk met zijn ATPase activiteit ruimte maakt rond de schade om herstel mogelijk te maken. Mutaties in CSA of CSB leiden tot de ernstige ziekte Cockayne Syndroom. Patiënten met deze ziekte vertonen o.a. versnelde veroudering. Dit laat zien dat CSA en CSB belangrijk zijn voor de mens, maar hun exacte functie is nog niet duidelijk.

Om meer inzicht te krijgen in de functie van CSA en om te kunnen verklaren hoe mutaties erin Cockayne Syndroom veroorzaken, zou een gedetailleerd plaatje van het eiwit nuttig zijn. Met bioinformatica is de algemene vorm van CSA te voorspellen. Het heeft een WD40 domein en dergelijke eiwitten hebben meestal een β-propeller vouwing. Wat niet voorspeld kan worden, is hoe alle lussen zijn gevouwen en welke zijketens waar zitten. Dat is namelijk het aspect waarin CSA anders is dan andere WD40 eiwitten. En het is juist die informatie die nodig is

om de ziekte te kunnen verklaren. Die informatie is alleen te verkrijgen door de structuur van het eiwit op te lossen. Dat hebben we dan ook gedaan en wel met de techniek eiwitkristallografie, waarmee een driedimensionaal plaatje van het eiwit met atomaire resolutie gemaakt kan worden.

De eerste stappen voor het oplossen van een eiwitstructuur bestaan uit het verkrijgen van het eiwit en de kristallisatie ervan. In **Hoofdstuk 2** staat beschreven hoe we CSA overgeproduceerd hebben in een complex met zijn interactiepartner DNA-schade Bindend eiwit 1 (DDB1), omdat CSA afzonderlijk niet oplosbaar was. We hebben het complex gezuiverd met drie zuiveringsstappen en vervolgens gekristalliseerd. De kristallen diffracteerden tot 2.9 Å.

Hoofdstuk 3 gaat verder met de beschrijving van de kristalstructuur van CSA in een complex met DDB1. Zoals verwacht vormt CSA een β-propeller. CSA bindt aan DDB1 via een N-terminaal helix-lus-helix motief. In de structuur kunnen we zien waar de mutaties zitten die Cockayne Syndroom veroorzaken. Deze mutaties leiden waarschijnlijk vooral tot problemen, doordat ze de lokale of globale vouwing verstoren. Dit zorgt er dan voor dat CSA niet meer kan binden aan zijn substraat of aan zijn interactiepartner DDB1. Helaas zijn zulke fouten in een eiwit moeilijk om te corrigeren met een klassiek medicijn. Verder is gekeken naar wat de meest waarschijnlijke plek van CSA is waarmee het aan zijn substraat bindt. Dit bleek het midden van de bovenkant van het WD40 domein te zijn. Daar zit een klein, positief geladen gat dat lijkt op de substraatbindende plekken voor gefosforyleerde eiwitten. Dit suggereert dat CSA een gefosforyleerd eiwit herkent of een anderszins negatief geladen deel van een ander eiwit.

De kristalstructuur van CSA liet zien welke aminozuren betrokken kunnen zijn bij binding van het substraat. Dit moest wel nog gecontroleerd worden met biochemische experimenten. Dergelijke experimenten staan beschreven in **Hoofdstuk 4**. Een variant van CSA met mutaties middenin de mogelijke substraatbindende plek, bleek niet de UV-gevoeligheid van een CS-A cellijn te kunnen complementeren. Dit wijst erop dat het inderdaad een belangrijke plek voor de functie van het eiwit is. Vervolgens keken we naar de twee meest waarschijnlijke substraten voor binding door CSA: DNA en CSB. CSA bleek niet aan DNA te binden bij biologisch relevante eiwit-DNA verhoudingen. Daardoor is het onwaarschijnlijk dat CSA aan DNA bindt in de cel. In andere experimenten bleek dat CSB meegetrokken wordt met CSA als CSA met zijn C-terminale His-tag uit menselijke cellen wordt gehaald. Bovendien zat er minder CSB in cellijnen waar CSA tot overexpressie werd gebracht. Dit wijst erop dat CSB het daadwerkelijke substraat voor CSA kan zijn.

Bacteriën en lagere eukaryoten hebben nog een andere manier om DNA-schades te herstellen, namelijk die van UV-schade endonuclease herstel (UVDE). UVDE is een eiwit dat verschillende schades kan herkennen, waaronder UV-schades als Cyclobutaan Pyrimidine Dimeer (CPD) en 6-4 fotoproduct (6-4PP). Vervolgens kan UVDE ook een knip in het DNA maken aan de 5′ kant van de schade. Andere enzymen maken het herstel daarna af. Het is bijzonder dat één enkel eiwit zowel meerdere schades kan herkennen als deze kan knippen. Normaal gesproken doen meer eiwitten dat samen. Daarom is UVDE een mooi, simpel modelsysteem voor DNA-schade herstel.

Een aantal jaren geleden is in Leiden de kristalstructuur van UVDE uit *Thermus thermophilus* opgelost. Daarin kon gezien worden dat een lysine in de buurt van de actieve plek van UVDE een post-translationele modificatie heeft. Dit is ongebruikelijk voor een bacterieel eiwit dat in een gastheer tot overexpressie wordt gebracht en daarom hebben we de identiteit en functie van deze modificatie onderzocht in **Hoofdstuk 5**. Met massa-spectrometrie en eiwitkristallografie hebben we aangetoond dat het om een carboxylering gaat. Met functionele en structurele testen van mutanten van UVDE hebben we de functie van deze carboxylering kunnen aantonen. Het bleek nodig om negatieve lading te geven die helpt bij de binding van de metaalionen die UVDE gebruikt voor katalyse. Bovendien bleek dat positieve lading op deze plek erg gevaarlijk is voor het eiwit, omdat dat voor veranderingen in de structuur zorgt dichtbij de plek waar de schade moet binden en daardoor gaat het eiwit ook onbeschadigd DNA knippen.

Verder was nog niet duidelijk uit de kristalstructuur van *Tth*UVDE hoe dit eiwit verschillende DNA schades kan herkennen en knippen. Om deze vraag te beantwoorden hebben we de cokristalstructuur van UVDE met beschadigd DNA (6-4PP) bepaald. In deze structuur (beschreven in **Hoofdstuk 6**) konden we zien dat UVDE een uniek mechanisme heeft om UV-schade in het DNA te herkennen. Niet alleen wordt de schade in een bindingsplek geduwd, ook de basen tegenover de schade gaan in een inkeping die specifiek is voor twee purines, wat leidt tot extra specificiteit voor beschadigde dipyrimidines. Het DNA moet over 90° buigen om in de bindingsgroef te passen; dit is waarschijnlijk energetisch ongunstig voor onbeschadigd DNA, wat leidt tot specificiteit voor beschadigd DNA waar bijvoorbeeld een deel van de waterstofbruggen al zijn verstoord. Om verder te onderzoeken waarom de meeste UVDEs meerdere schades kunnen knippen, hebben we de UVDE homologen *Tth*UVDE (die CPD, 6-4PP en AP kan knippen) en *Sac*UVDE (die alleen 6-4PP kan knippen) vergeleken in activiteit en structuur. De plek waar de schade bindt in het eiwit bleek minder flexibel te zijn in *Sac*UVDE door een zwavelbrug. Ook bleek *Sac*UVDE geen positieve ladingen te hebben aan beide kanten van de DNA-bindingsgroef - wat *Tth*UVDE wel heeft - en die ladingen zijn waarschijnlijk nodig om meerdere DNA substraten correct te kunnen buigen, zodat ze in de groef kunnen binden. Deze hypotheses hebben we bevestigd met activiteitstesten op mutanten van UVDE. De resultaten tonen aan dat de meeste UVDEs juist veel verschillende substraten kunnen binden, doordat ze goed DNA kunnen buigen met positieve ladingen aan beide kanten van de bindingsgroef en doordat ze een relatief flexibele bindingsplek voor schade hebben.

Het laatste hoofdstuk van dit proefschrift (**Hoofdstuk 7**) gaat over de kristalstructuur van een ongerelateerd eiwit, Aardappel Serine Protease Remmer (PSPI). PSPI is een onderdeel van het verdedigingsmechanisme van de plant. Het remt serine proteases zoals trypsine en chymotrypsine en kan daardoor voorkomen dat dieren of ziekteverwekkers zich voeden met de plant. PSPI bleek een β-trefoil vouwing te hebben. Karakteristiek hiervoor is een zeer stabiele kern met lussen die van daaruit naar buiten steken. De reactieve plekken van het eiwit zitten meestal in deze lussen. Het eiwit kan proteases remmen door ze vast te laten lopen op deze lussen: de proteases binden eraan met het doel de remmer kapot te knippen.

Echter, door de stabiele kern valt de remmer niet uit elkaar na de knip en blijft de protease op de remmer vastzitten doordat het blijft proberen die kapot te knippen. Door onze structuur te vergelijken met andere structuren van protease-remmers konden we bepalen wat de reactieve plekken zijn van PSPI. PSPI is bijzonder, omdat het een heterodimere protease-remmer is met twee reactieve plekken. In een post-translationeel proces zijn er namelijk zes aminozuren uit een lus weggeknipt. Uit onze structuur bleek een mogelijke functie hiervoor: het is misschien betrokken bij de activatie van de remmer, want het wegknippen van de lus legt één van de reactieve plekken bloot.

Afsluitende opmerking

Dit proefschrift laat zien hoe eiwitkristallografie inzicht kan geven in de functie van eiwitten. Het geeft voorbeelden van hoe een drie-dimensionaal plaatje van een eiwit met atomaire resolutie kan helpen in het verklaren hoe een eiwit werkt; zowel de globale vouwing als de exacte locaties van bepaalde aminozuren zijn hierbij van belang. Met de overvloed aan informatie die al beschikbaar is in de Eiwit Data Bank (PDB) kan bijna elke nieuwe structuur geïnterpreteerd worden door de structuur te vergelijken met eerder bekende structuren. Echter, de interessantste onderdelen van een eiwit zijn juist de delen die anders zijn dan in eerdere structuren. De interpretatie van die verschillen is vaak speculatief en daarom is biochemische verificatie ervan nodig. Technieken als mutagenese en activiteitstesten zijn hier zeer geschikt voor: ze vormen een complementaire bron van informatie. Zoals te zien is in de hoofdstukken over CSA en UVDE in dit proefschrift, is vooral de combinatie van structurele biologie met biochemie een erg sterke methode om biologische systemen te onderzoeken.

Curriculum vitae

Elisabeth M. Meulenbroek

18 February 1985	Born in Middelburg, The Netherlands
1997-2003	Secondary School, Stedelijke Scholengemeenschap Nehalennia, Middelburg
2003 - 2006	B.Sc. in Chemistry at Leiden University, *cum laude* *Emphasis on Organic Chemistry and Biochemistry*
2006 - 2008	M.Sc. in Chemistry at Leiden University, *cum laude* *Emphasis on Biochemistry and Science Communication* *Master thesis: Functional and structural studies of UVDE.* *Supervisor: Dr. N. Goosen*
2008 - 2012	Ph.D. Research at Leiden University *Thesis: Structure-based insights into the repair of UV-damaged DNA* *Supervisor: Dr. N.S. Pannu* *Promotors: Prof. dr. J.P. Abrahams and Prof. dr. L.H.F. Mullenders*