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

Meulenbroek, E.M.

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

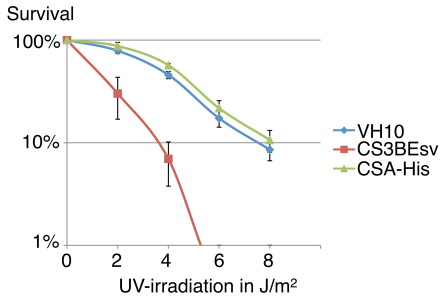

Figure 3.1: Complementation assay with CSA-His
Survival after UV irradiation of VH10-SV cells (blue), CS3BE-SV cells (red) and CS3BE-SV cells transfected with the CSA construct used for crystallization (green), showing that this construct can complement UV-sensitivity of CSA$^-$ cells.

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-

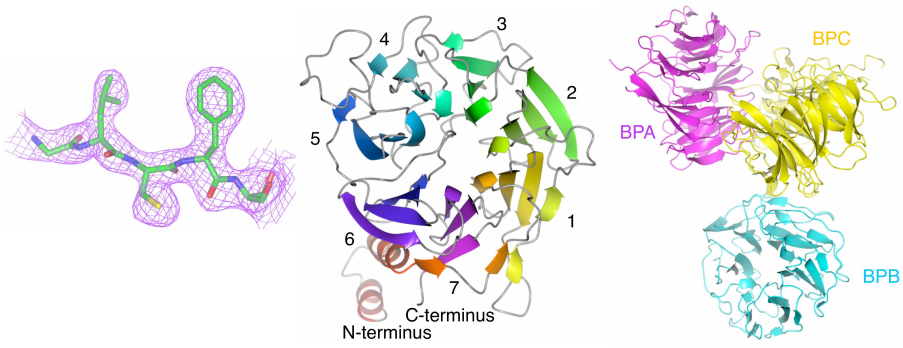

Figure 3.2: Overall fold of CSA and DDB1
(a) Representative part of the structure with electron density map (in CSA).
(b) Overall fold of CSA showing a seven-bladed β-propeller (one molecule of CSA shown; top view) with numbering of the blades of CSA.
(c) Overall fold of DDB1 showing three β-propellers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

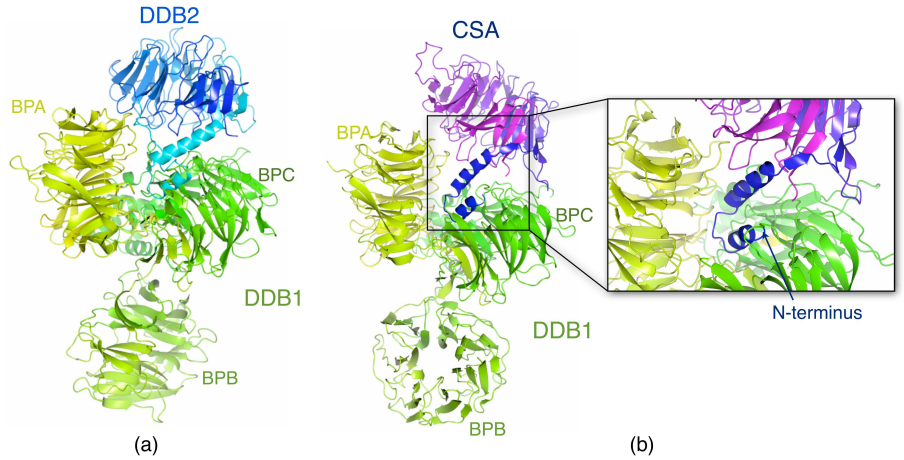

Figure 3.3: Interaction of CSA with DDB1 goes via a general binding mode for DCAFs to DDB1.
(a) Crystal structure of DDB2 (blue) and DDB1 (yellow to green) showing that DDB2 binds DDB1 via a N-terminal helix-loop-helix motif inserted between DDB1's BPA and BPC domain (Scrima et al., 2008).
(b) Complex 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-

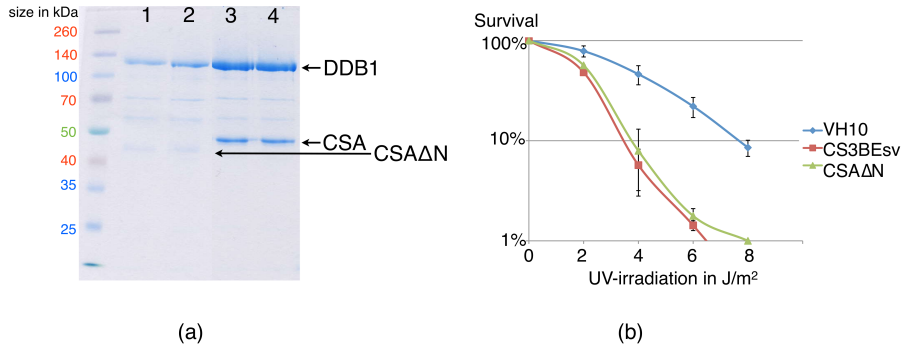

Figure 3.4: Importance of the N-terminal 30 amino acids of CSA.
(a) SDS PAGE gel showing the result of a small-scale Ni column purification with DDB1-CSAΔN (lanes 1 and 2) compared to DDB1-CSAwt (lanes 3 and 4) showing that complex formation is severely reduced in the absence of the N-terminal 30 amino 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

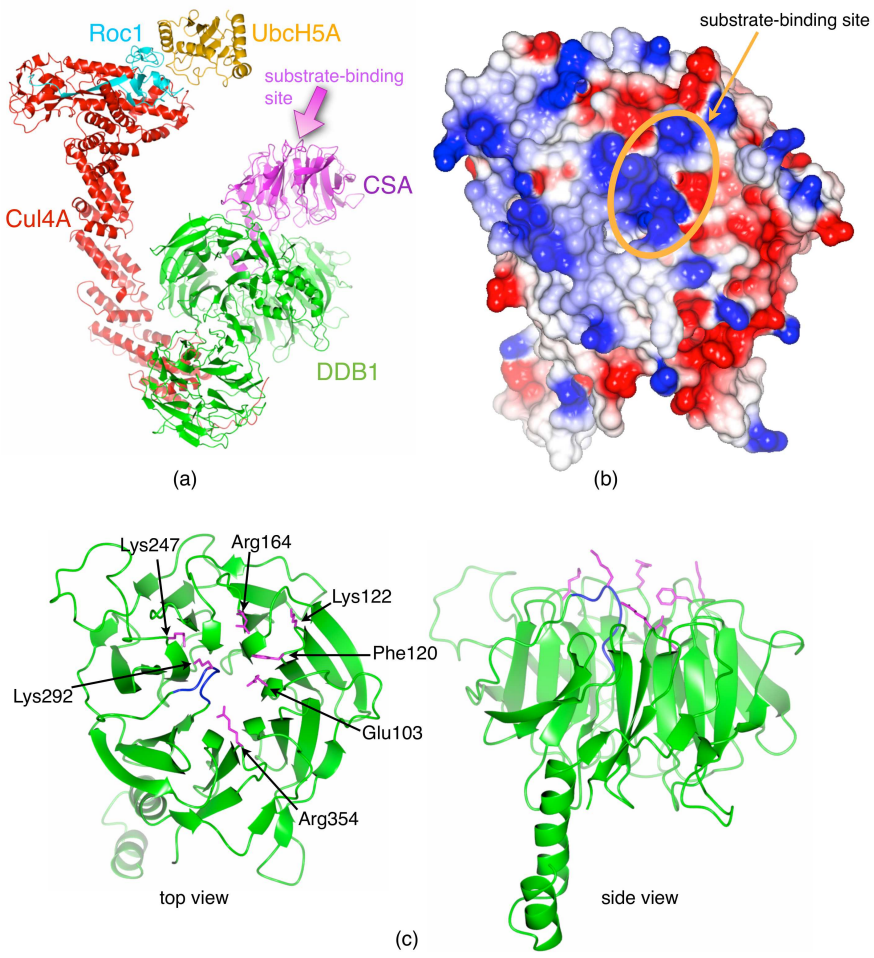

Figure 3.5: The proposed substrate-binding site of CSA.
(a) Model of the CSA-DDB1 structure (CSA in magenta and DDB1 in green) as part of the E3-ubiquitin ligase complex with Cul4A (red), Roc1 (cyan) and UbcH5A (orange). The arrow indicates the proposed substrate-binding site of CSA.
(b) Electrostatic 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.

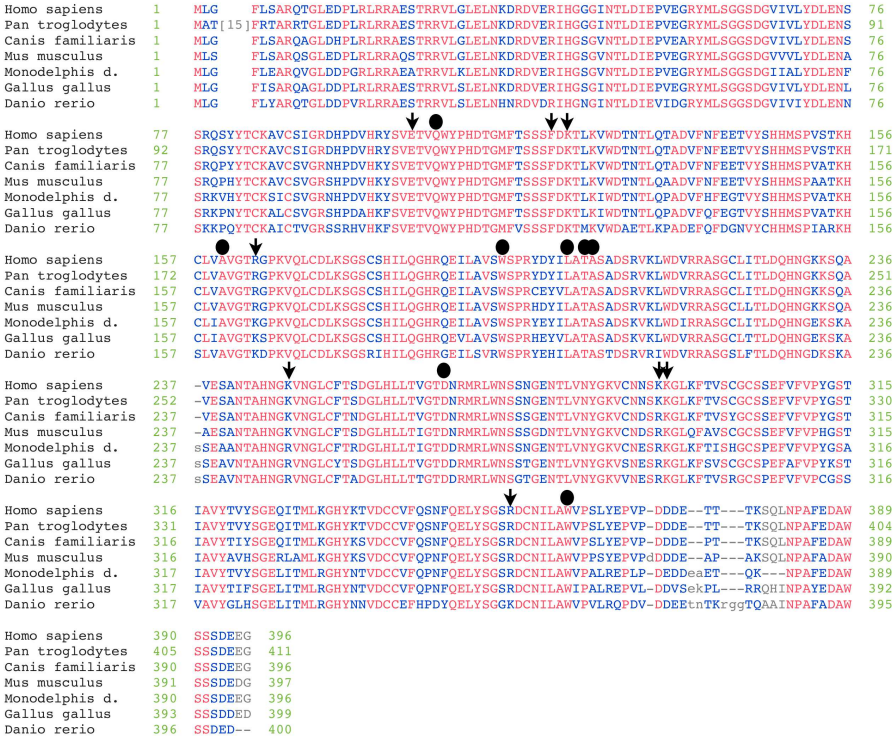

Figure 3.6: Alignment of CSA from different species. Arrows point at proposed substrate-binding residues and dots mark residues mutated in Cockayne syndrome. Red indicates identical residues, blue partially conserved residues and grey unaligned residues.

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

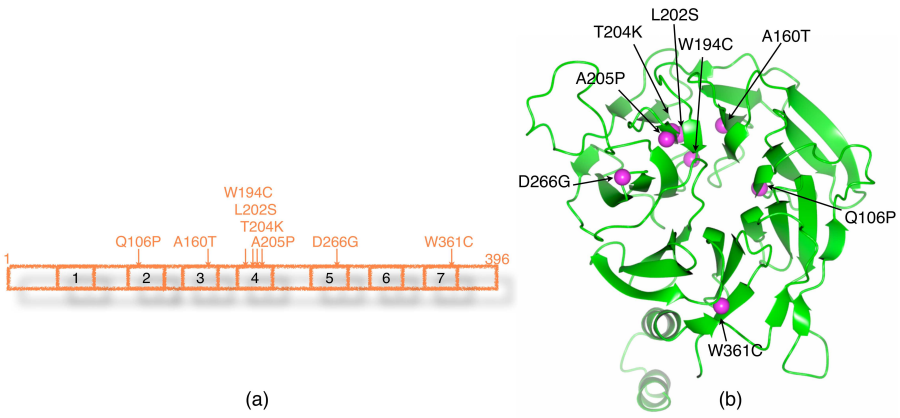

Figure 3.7: Disease-causing mutations in CSA.
(a) Schematic overview of the eight point mutations in CSA that cause Cockayne syndrome, shown on the primary structure. The boxes indicated the seven WD40 repeats of CSA.
(b) Disease-causing mutations shown as magenta balls on a cartoon representation of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

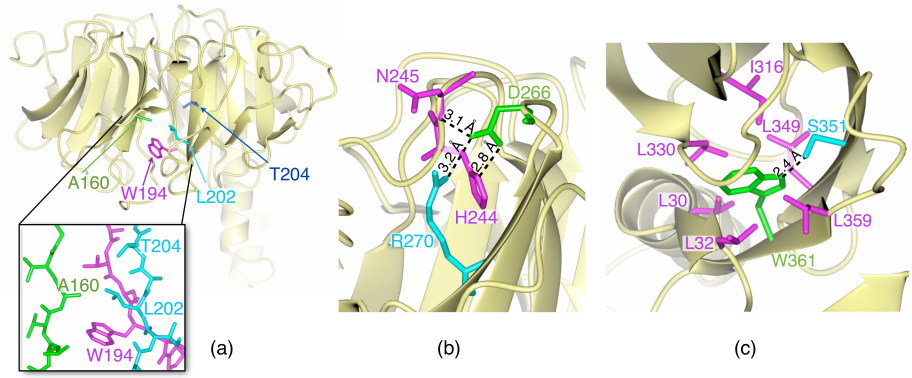

Figure 3.8: Close-up of disease causing mutations.
(a) Side-view of CSA and close-up showing that mutations A160T (green), W194C (magenta), L202S (cyan) and T204K (cyan/ blue) are close together in the three dimensional structure.
(b) Close-up of CSA showing the hydrogen bonding of residue D266 (in green), 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

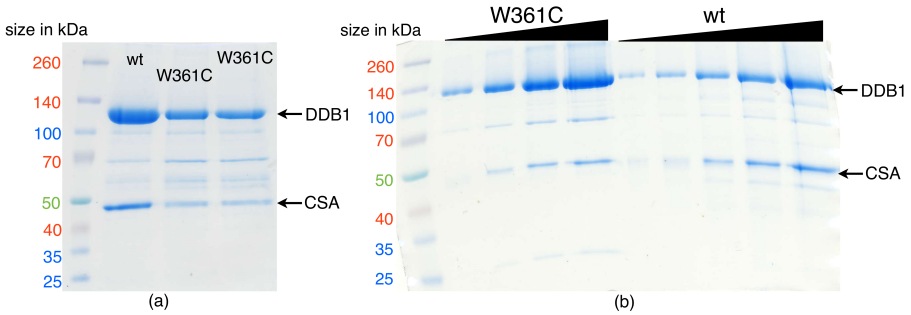

Figure 3.9: Effect of the W361C mutation on CSA in vitro.
(a) SDS PAGE gel of CSA-W361C versus CSA wt after small-scale Ni-columns showing the reduced production of CSA-W361C versus wildtype CSA. Two independent W361C overproductions are shown to illustrate the reproducibility. (b) SDS PAGE gel showing 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.