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

Meulenbroek, E.M.

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.