Interactome analysis and docking sites prediction of (AtCHR8, AtCUL4 and AtERCC1/UVR7) proteins in *Arabidopsis Thaliana*

Mohamed Ragab Abdel Gawwad[1],*, Ali Taha Ozdemir[1]

[1]Genetics and Bioengineering, Faculty of Engineering and Natural Sciences, International University of Sarajevo

*Correspondence: mragab@ius.edu.ba; Tel.: +38733957203

Abstract

The UV irradiation is a major DNA damaging factor in plants. Arabidopsis thaliana uses various repair pathways for these kinds of DNA lesions. One of them is the nucleotide excision repair pathway. The AtCUL4, ERCC1/UVR7 and CHR8 are vital proteins for nucleotide excision pathway and mutations in these proteins cause flaws in the repair mechanism. Two of these proteins play crucial role during DNA damage recognition and the other is involved in the excision of damaged bases. During NER processes, Arabidopsis uses different sets of proteins during the DNA damage recognition for transcriptionally active and genomic DNA. In order to get better insight into these proteins, we used bioinformatics tools to predict, analyze, and validate 3D structures of ERCC1/UVR7, AtCUL4 and CHR8. We also predicted the subcellular and sub-nuclear localization of proteins. Subsequently, we predicted the docking sites for each individual protein and searched for interacting residues which mediate the protein-protein interactions

Keywords: CHR8, AtCUL4, AtERCC1/UVR7, docking sites, TC-NER, GG-NER

1. Introduction

DNA damage refers to any change in the chemical structure of DNA. Damage to DNA may shift or interfere with fundamental cellular processes which cause genomic instability or lethal mutations. To maintain normal function and structural stability, cells must repair these DNA alterations. Otherwise, Such DNA lesions may result in block of transcription or DNA replication process and risk viability of the cells [1, 2, 3, 4]. DNA lesions result from either endogenous or exogenous reagents. Generally, DNA damages are caused by endogenous reagents, such as free radicals or spontaneous hydrolysis reactions [5]. Other than these endogenous factors, DNA is also subjected to exogenous (environmental) DNA-damaging factors (x-rays, gamma and UV radiation). Ultra-violet radiation is the major cause of the DNA damage in plants. According to Clancy (2008), up to one million DNA changes can occur in a single cell [6, 7]. Common DNA lesions consist of single and double-strand breaks, base or sugar modifications, DNA-protein cross links and base free sites [5]. To overcome harmful effects of DNA damage, cells have developed numerous distinct repair mechanisms, such as nucleotide excision repair, mismatch repair, base excision repair, and double strand break repair [7].

Earth's surface is exposed to many harmful solar rays which may harm living organisms in the biosphere. UV spectrum can be divided into three subtypes according to their wavelengths [8, 7]. These are: UV-C (100-280), UV-B (280-315), and UV-A (315-400). By nature, all living cells can absorb UV-radiation due to UV-absorbing molecules, for example nucleic acids and proteins [3]. However, the ozone layer completely filters out the most harmful kind of UV spectrum which are UV-C rays. Although UV-A and UV-B can pass through the ozone layer. UV-B is the major cause of the DNA damage in living cells [4, 7]. For example, UV-A induced mutations are 10,000 times less effective than UV-B induced mutations [4]. Recent studies show that stratospheric

demolition of the ozone layer causes severe effects on all living organisms due to enlarged transmission of ultraviolet radiation [8]. Plants are exposed to UV radiation in a greater extent than others because of their non-motile nature. Therefore, it is logically conceivable that plants have come up with an effective DNA repair mechanisms and UV-damaged DNA tolerance [3, 7, 8]. UV-B radiation causes three types of DNA lesions in living organisms. Primarily, it leads to formation of cyclobutane pyrimidine dimers (CBPs). In this type of DNA lesions, two adjacent pyrimidine residues on the same strand of DNA (TT or CC) are covalently attached between C5 and C6 carbon atoms [2, 3, 4, 5, 7, 8, 9]. Alternatively, 6-4 pyrimidine pyrimidone photoproducts (covalent linkage formation between C6 and C4 carbon atoms in the adjacent pyrimidine residues) may appear due to UV exposure [4, 5, 8, 9]. Lastly, exposure to UV-B can lead to formation of isoforms 6-4 PPs which are called Dewar photoproducts [4]. These types of DNA lesions are repaired by either direct repair mechanism or replacement of damaged bases. Direct DNA repair process consists of the use of visible light (Blue light) as an energy source to break covalent bonds between dimers which is called photo-reactivation [2, 3, 4, 7, 8]. In the latter process, UV-B induced DNA damages are repaired by nucleotide excision repair mechanism (NER) in which damaged bases are cut out and replaced by newly synthesized ones. Furthermore, unlike photo-reactivation, Nucleotide excision repair process does not require use of light.

In nucleotide excision repair mechanism UV-induced lesions are fixed in such way in which pyrimidine dimers and 6-4 PPs are repaired by a group of thirty DNA repair proteins in multiple steps [5, 7]. These stages are DNA damage recognition, DNA unwinding, dual excision of DNA lesion (single strand segment of DNA), sequential repair and lastly ligation step to link phosphodiester bonds [8]. According to Ashwin et al. (2011), transcription coupled repair (TC-NER) shows only repair of transcriptionally active DNA regions but not entire genome due to difference in recognition stages. In GG-NER, a protein complex, called UV-damaged DNA binding complex (DDB1 and DBB2), causes DNA-bending and local distortions at DNA lesions in order to increase affinity of damage recognition factor (XPC/HR23B/CEN2) [5]. It's shown that the binding affinity and specificity of XPC to DNA helix is greatly influenced by DNA distorting damage [2, 5, 8]. Alternatively, *Arabidopsis Thaliana* consists of two homologs of DDB1. These are DBB1A and DBB1B [8]. Also, overexpression of DDB1 seems to increase UV resistance. AtCUL4 is another important and indispensable protein which functions in both GG-NER and TC-NER processes. It interacts with DDB1A/B, DDB2 in GG-NER. In addition, it associates with CSA and CSB (CHR8) in TC-NER process. For example, while CUL4/RBX1/DDB1/DDB2 complex assists ubiquitination of histones and assist XPC/CEN2/HR23B DNA damage recognition complex binding, CUL4RBX1/CSN/CSA/CSB complex supports repairing at the site of transcriptionally active sites regions [10, 11]. Sharma et al. (2014) demonstrates that major functions of CUL4 is highly conserved in eukaryotes [12].

On the other hand, in TC-NER, two particular proteins CSA (Cockayne syndrome A) and CHR8 (also called CSB, Cockayne syndrome B), remove RNA polymerase II from the DNA when an elongating RNAP-II is blocked by DNA lesions [5, 11]. Svejtrup (as cited in Ashwin et al., 2011) states that a mutation in CHR8 (CSB) and CSA effects entire TC-NER process and results in TC-NER flaws. The second function of CSA and CHR8 proteins are to recruit ten protein complexes and multi-subunit transcription factor TFIIH at the site of DNA damage to unwind DNA helix. The subunits of TFIIH, XPB (3'-5' helicase) and XPD (5'-3' helicase), use ATP as an energy source to unwind DNA helix [4, 13]. In the following step, the two endonucleases, XPF/AtERCC1 (UVR7) and XPG, perform dual incision of the DNA at sites of 3' and 5'to DNA lesion [2, 3, 4, 5, 13, 14, 15]. After cleaving off the size of 30 oligonucleotides, DNA polymerase can subsequently fill the gap by using undamaged strand as a template [5]. Lastly, the DNA ligase seals the broken phosphodiester bonds.

According to Christiansen et al. (2005), the CHR8 protein is a member of SWI2/SNF2 protein family which consists of an acidic domain (a glycine rich region), a central ATPase domain (Neighboring the N-terminal and the C-terminal regions), and an ubiquitin-binding domain (UBD) [16, 17]. In addition, there are two nuclear localization signal sequences (NLS). The molecular mass of the CHR8 gene and enzymatically active CHR8 protein are predicted as 168 kDa and 360 kDa, respectively [8, 16]. The CHR8is a member of DNA-dependent ATPase and function as stimulator for protein-DNA interaction. Therefore CHR8 also plays role in chromatin remodeling process [16, 18].

In transcription coupled NER, both CSA and CHR8 proteins play key role for the repair process. The function of the CHR8 protein is to recognize stalled RNA polymerase II and recruit other repair factors for TC-NER process [8]. In this process, the CHR8 protein physically interacts with RNA Polymerase II and prevents its association with DNA helix for DNA repair initiation. Ashwin et al. (2011) also pointed out that human CSB (CHR8) undergoes some complex interaction to facilitate joining of the core NER factors. CHR8-RNAP-II complex substantially interact with the CSA protein in the presence of UV radiation [8].

The AtCHR8-chromatin association is regulated by post-translational modification such as phosphorylation of serine residues in the protein. There are 12 phosphorylation locations in AtCHR8 (annotated 12 serine and 1 tyrosine). Robert J. Lake and Hua-Ying Fan (2013) also point out that serine residues at the position of 158, 429, 430, 486 and 489 in N-terminal region were phosphorylated. However, phosphorylation of serine residues 158, 1461 and 486 doesn't affect the UV-induced CHR8-chromatin interaction [17]. Similarly, three more serine residues were also found in C-terminal. Upon UV-irradiation, the CHR8 is dephosphorylated thereby having an increased ATPase activity in the presence of UV radiation [16]. Alternatively, a tyrosine residue (932) is phosphorylated by the c-Abl kinase which functions in variety of signal transduction routes. Interestingly, the phosphorylation of tyrosine residue is found to play a role in nuclear or nucleolar localization of the CHR8 protein [17]. Other than phosphorylation, ADP-ribose moieties are also attached to the C- and N-terminal regions of the CHR8 by the enzyme PARP-1 (poly (ADP-ribose) polymerase-1). It is suggested that addition of ADP-ribose moieties inhibits the ATPase activity of the CHR8 thereby controlling repair process of DNA lesions [17].

The CHR8 also allows the interaction of two proteins, p300 (histone acetyl transferase) and HMGN1 (High Mobility Group Nucleosome Binding Domain containing protein), with RNA polymerase II upon ultraviolet radiation exposure [4]. CHR8/RNAP-II/HMGN1/p300 complex interacts with nucleosome and enables remodeling chromatin structure and reverse translocation of RNA polymerase [4]. In *Arabidopsis thaliana,* there is no equivalent protein homolog to HMGN1, but mutation in the CHR8 genes show high UV sensitivity.

The molecular weight of CUL4 is measured as 91kDa. As for function of CUL4, Lee et al. (2007) states that N-terminal of the CUL4 consists of receptors for WD40 motif proteins while C-terminal interacts with Ring Finger Proteins (Rbx1/ROC1/Hrt1) [10, 19]. The CUL4 protein is engaged in both TC-NER and GG-NER process. In the GG-NER, CUL4 functions as scaffolding subunit in which multimeric complexes are assembled to form a CUL4 based E3 Ubiquitin ligase [8, 20, 21]. The CUL4 associates with the BPB domain of DDB1-A/B and with RBX1 (Ring Finger Protein) to form a complex which interact with large number of proteins (DCAF proteins) [8, 10, 11, 12, 19, 20, 21, 22, 23]. For example, in *Arabidopsis thaliana*, both AtDDB2 and AtCSA-1&2 interact with CUL4–DDB1A/B complex during NER process. In TC-NER, Cullin-based E3 ubiquitin ligase activity is regulated by a complex, called COP9 Signalosome complex (CSN). Binding of CSN to CUL4, prevents CULLIN based E3 ligase activity by detaching NEDD8 (ubiquitin-like protein) from CUL4 (9, 24, 25). However, in the presence of UV irradiation, CSN disassociates from CUL4 thereby activating repair process [10, 22, 23]. On the other hand, in GG-NER, CUL4/RBX1/DDB1 complex enables the ubiquitination of histones and XPC. This process results in chromatin remodeling and access of repair proteins to the site of DNA lesion [8, 11, 12, 20]. While poly-ubiquitination of DBB2 causes a decreased affinity to the DNA, it increases the affinity of the Damage recognition complex (XPC/ HR23B/ CEN2) [8]. Following the unwinding of DNA helix by TFIIH transcription factor and DNA helicases (AtXPD, AtXPB1 and AtXPB2 in *Arabidopsis thaliana*), single stranded DNA binding protein RPA (replication protein A) and XPA additionally stabilize single stranded DNA helix [8].

After stabilization of opened DNA complex, two endonucleases, XPF-AtERCC1/UVR7 and XPG catalyze dual excision of 20-30 oligonucleotides in human systems [26, 27]. In *Arabidopsis thaliana,* AtERCC1/UVR7 and AtXPG mutations seem to show tolerance to UV irradiation. However, these mutants are sensitive to chemical mutagens [8]. While AtERCC1 (UVR7)/XPF heterodimer cuts DNA backbone at the position of 5' to the DNA lesion, AtXPG cuts 3' of the damaged site [14, 15, 28]. Human AtERCC1/UVR7 and XPF proteins have predicted molecular mass of 31 and 103 kDa, respectively [14, 29]. AtERCC1/UVR7 cuts DNA in such way in

which single strand goes from 5' to 3' away from the junction. It's also stated that AtERCC1/UVR7 takes a part in recombinational repair pathway [28, 29, 30]. Vannier et al. (2009) shows that AtERCC1 (UVR7) /XPF has involved in telomere protection as well as homeostasis [30, 31]. Both AtERCC1/UVR7 and XPF proteins interact with each other through their C-terminal Helix-Hairpin-Helix domains (HhH2). Thus, loss of C-terminal or mutations in these domains cause loss of function [15, 27]. Moreover, there are two domains found in AtERCC1/UVR7 [27]: a conserved central domain which confers nuclease activity (between 69 and 204 residues) and HhH_2 domain for dimerization (between 220 and 297).

The aim of this study is to decipher the DNA repair mechanism in *Arabidopsis thaliana* which CHR8, ERCC1/UVR7 and CUL4 proteins is involved through analyze the functional interactome of these proteins. To achieve the most accurate result, the stated aim will be enhanced and supported through the use of several bioinformatics tools. In addition, the domains of proteins will be analyzed using online SMART. Subsequently, the 3D structure will be predicted and confirmed by Ramachandran plot and finally, docking site will be predicted and interactome will be analyzed.

2. Materials and methods

2.1. Sequence retrieval

The amino acid sequences of all three proteins, AtERCC1/UVR7, AtCHR8, and AtCUL4, are obtained from The Arabidopsis Information Resource database TAIR [31, 32].

2.2. 3-D Structure prediction

The AtCHR8, AtERCC1/UVR7 and AtCUL4 proteins of Arabidopsis thaliana do not have experimentally confirmed 3D structure in Protein Data Bank (RCSB PDB) database [33, 34, 35]. Therefore, SWISS modeling and 3D-Jigsaw Protein Comparative Modeling servers are used for predicting the structure and/or function of these three proteins. It uses evolutionary information for modeling the tertiary and quaternary structure of proteins [36, 37]. The quality of models can differ greatly thus, the validation of predicted structures is necessary and are performed by using several bioinformatics tools. For example, the structural assessments of proteins are obtained by means of QMEAN server which enables us to estimate quality of models. The QMEAN score gives values between 0 and 1. The Models with higher values are more reliable than the lower ones [38]. The QMEAN z-score, on the other hand, calculates the estimation of quality of models by connecting it to reference structures solved by x-ray crystallography [38, 39, 40, 41]. The PROCHECK tool is used for analyzing quality of the stereo-chemistry of protein models which can give clues about possible geometry of the residues [42]. Online RAMPAGE program additionally used to obtain Ramachandran plots and percentage of amino acids in the allowed region [43]. Visualization of protein structures are performed by using the PyMOL Molecular Graphics System. PyMOL is an open-source bioinformatics tool that enables users to visualize, manipulate, compare and analyze 3D-structure of molecules [44]. Alternatively, the Deep Viewer is used to analyze electrostatic potentials of the predicted proteins [37].

2.3. Domain search and interactome predictions

SMART was used to identify genetically mobile domains of the three proteins (AtCHR8, CUL4 and ATERCC1/UVR7). PROSITE is also used for domain identification. PROSITE is a database of protein families and domains [45, 46, 47, 48, 49, 50]. STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) is used for searching interlogs of the three proteins [51, 52].

2.4. Prediction of protein subcellular locations

Plant Subcellular Localization Integrative predictor (PSI) was used to determine specific location of AtCHR8, AtCUL4 and AtERCC1/UVR7 proteins [53]. As for sub-nuclear location prediction, a support vector machine (SVM) based system, Subnuclear Compartments Prediction System (Version 2.0) was used to localize proteins in the nucleus [54, 55].

2.5. Docking Site Predictions

The prediction of docking sites accomplished by using a web-based program, GRAMM-X Protein-Protein Docking Web Server v.1.2.0. The GRAMM-X is a fully automated docking site prediction tool and was developed by Vakser Lab in the University of Kansas [56].

3. Results

3.1. Predicted 3D structures and validation

The function of a protein is determined by its three-dimensional structure and its electrostatic charge around the surface. Therefore, determination of the 3D structure of proteins is essential and indispensable step to understand the function proteins. To date, there are no experimentally determined structures of Arabidopsis AtCHR8, AtCUL4 and AtERCC1/UVR7 so far. Thus, the 3D structure prediction tools have been used to obtain the 3D conformation of these proteins. 3D-Jigsaw Protein Comparative Modeling server and SWISS modeling server are used together to generate 3D models (Figure 1). Additionally, the RAMPAGE (online Ramachandran plot maker) or ProCHECK is used to generate Ramachandran plots and to calculate the percentage of residues in allowed region (Figure 1).

However, generating the 3D models alone does not confirm its quality because there are many factors which can influence the correct folding of a protein in its native conformation. Because of that a validation process is needed to find best models for further analysis. For example, a correctly folded native protein is said to have the lowest energy during protein folding process. Thus, finding the lowest free energy conformation is a key step for generation models with good quality.

The web based online tools, dFIRE and DFIRE2 are used to calculate protein conformational free energy scores for proteins: AtCHR8, AtCUL4 and AtERCC1 (Table 2). The scores with more negative values are favored and they indicate that the structure is closer to native conformation. As mentioned earlier, QMEAN, a structural assessment tool is used for the 3D structure validation. The QMEAN scores and z-scores are used for choosing the best models. The models with higher QMEAN scores (range between 0 and 1) and z-scores are closer to native conformation. [38, 39, 40, 41].

The PDB viewer software is used for visualization of electrostatic potentials around the predicted proteins. [36] During this process, Coulomb computation method is used and only charged residues are taken into account while solvent ionic strength parameters were set to 0 mol/I. Additionally, Pymol visualization software is used to visualize structural features of the proteins. Table 2 shows the results of the structure assessments for the validation of predicted 3D structures performed by SWISS modeling servers.

3.2. Domain analysis

The domains of the AtCHR8, AtCUL4 and AtERCC1/UVR7 proteins were identified by using Simple Modular Architecture Research Tool software (SMART) found in the website of the European Molecular Biology Laboratory (EMBL) (Table 3). The AtERCC1/UVR7 protein contains the following domains: ERCC4 domain (RAD 10 domain) and HhH1 (Helix-hairpin-Helix DNA binding motif class 1) domain [45]. Proteins with ERCC4 domain are known to function as multiprotein endonuclease complex [58, 21]. AtERCC1/UVR7 and XPF both contain ERCC4 domain and co-operate together to form an endonuclease complex. The HhH1 domain functions as non-sequence specific DNA-binding site.

The CHR8 protein consists of an N-terminal coiled coil region, middle DEXDc (DEAD-like helicases superfamily) domain, and C-terminal HELICc (helicase superfamily C-terminal domain) domain followed by another coiled coil region. As visible from the Table 4 below the domains of the AtCHR8 protein are described. Lastly, the AtCUL4 protein consists of two functional domains. Firstly, the cullin domain that functions in ubiquitin mediated destruction of proteins. The second one is N-terminal Cullin Nedd8 domain (cullin protein neddylation domain).

3.3 Protein localization

The subcellular locations of the AtCUL4, AtERCC1/UVR7 and AtCHR8 are determined by using PSI (Plant Subcellular Localization Interactive Predictor) tool. According PSI assessment, AtCHR8 and AtERCC1/UVR7 are located in nucleus while CUL4 is present in cytosol (Table 4). The PSI scores can be seen in the Table 6. The value of scores vary from 0 to 1 and higher scores (close to 1) indicate the higher confidence in the presence of a protein in particular compartment.

3.4. Interactome analysis

The STRING database is used to detect number of protein interactions. During interaction prediction, the minimal confidence score was kept at 0.700 while only 15 interactors were shown. The number of predicted functional partners are found as 15 proteins. According to STRING database, the CHR8 protein interacts with 6 different proteins. The AtCHR8 is interacting with NRPB9A, AtERCC1/UVR7, UVH1, UVH3 proteins with scores of 0.900; with NRPB9B and DDB1A having scores of 0.787 and 0.243, respectively. Similarly, the AtERCC1/UVR7 is interacting with AtCHR8 (0.900), NRPB9A (0.900), NRPB9B (0.900), UVH1 (0.940), UVH3 (0.900), MSH2 (0.914), MSH5 (0.914) and DDB1B (0.900) proteins. The AtCUL4 is shown to interact with 9 proteins, namely FUS9, RBX1, COP1, COP8, CSN5B, DCAF1, DDB1B, and DET1 with scores 0.332, 0.853, 0.789, 0.564, 0.994, 0.993, 0.999 and 0.999 respectively. According to results, both AtCHR8 and ATERCC1/UVR7 are interacting with DDB1A, NRPB9A, NRPB9B, UVH1 and UVH3. However, DDB1A does not interact with AtCHR8 as strong as it interacts with AtERCC1/UVR7. On the other hand, DDB1A seems to interact with all three, AtCHR8, AtERCC1/UVR7 and AtCUL4 proteins. The interaction network of proteins is visible below in figure 3.

3.5. Docking site predictions

The GRAMM-X docking site prediction tool was used to predict the binding site of selected proteins for AtCHR8, AtCUL4 and AtERCC1/UVR7 (Table 5). The results of the predicted structures are visualized by using Pymol visualization tool. Following sets of figures show the predicted docking sites of each individual selected protein (supplementary materials: Figures from 1 to 15).

4. Discussion

The UV radiation causes DNA lesions which are repaired by either GG-NER or TC-NER systems. The proteins AtCHR8, AtCUL4, and AtERCC1/UVR7 play important role for the repair of these lesions. Even tough, these two mechanisms have a mutual repair pathway, they have different damage recognition strategies. For example, while the AtCHR8 protein is only involved in the damage recognition in the transcription coupled nucleotide excision repair mechanism, the AtCUL4 and AtERCC1/UVR7 are involved in both global genomic NER and TC-NER system.

The CUL4 is a key protein to both repair mechanisms since it functions as scaffolding unit and its presence is absolutely necessary for Cullin based E3 ligase activity. For example, DDB2-CUL4 mediated XPC poly-ubiquitination enhances its binding to the site of DNA lesion during the DNA damage recognition in GG-NER [8]. Kapetanaki et al. shows that (as cited in Ashwin, 2011) the *cul4* knock out results in reduction in Histone H3 and H2 ubiquitination as well as failure to XPC assembly at the site of DNA lesion. The CUL4- based E3 ubiquitin ligases form a large subfamily of Culin Ring E3 ubiquitin ligases (CRLs) that are highly conserved during evolution. Human CUL4 consists of two functionally redundant members, CUL4A and CUL4B, with 84% sequence identity. However, *Arabidopsis thaliana* has no such duplications [12]. After SMART domain analysis, it is shown that *Arabidopsis* CUL4 consists of two conserved domains, namely CULLIN and NEDD8 domains which are located on C terminal region. This conserved globular C-terminal domain is responsible for binding to small RING protein ROC1 which enables the recruitment of E2 enzyme to Culin complex [12, 49]. On the other hand, N-terminal involved in binding to DDB1 BPB domain with help of helix2 and helix 5 [12]. According to Zhang Yu et al. (2008) especially, the interaction of CUL4 with DDB1 and ROC1/RBX1 are highly conserved in *Arabidopsis thaliana* during evolution [20]. Particularly, on DDB1, the change of amino

acid residues at positions of 82–85, 87, 88, 91, 92, 150–152, 154, 155, 158, 159 and 162 cause an ineffective complex formation [12]. The CUL4 interacts with a wide variety of proteins which consist of conserved WD motifs (Called DCAF proteins).

All of the CUL4-DDB1 interacting proteins including; DDB2, COP1, and CSA contain shared WD motifs which are 40 residues in length are primarily starting with a Glycine-histidine (GH) dipeptide and then ending with a (WD) Tryptophan-Aspartate dipeptide [59]. It is also known that CUL4 does not interact with a substrate molecule itself, but instead it uses a linker protein, DDB1 or substrate binding adaptor protein, for interacting large number of WD40 motif containing proteins during selective ubiquitination process. Moreover, Zhang Yu et al. shows that in *Arabidopsis,* AtCUL4 and RBX1/ROC1 complex interacts with complex of (constitutive photo-morphogenic 10) COP10-DET1-DDB1 complex to form an active CDD-CUL4-ROC1/RBX1 E3 ubiquitin ligase complex. As the STRING interaction analysis illustrates that the AtCUL4 is strongly associated with FUS9 (component of COP10) and DET1 proteins, thereby it suggests FUS9 has a particular function in the multiprotein complex. In order to understand the interactions between FUS9 and CUL4, the domains of FUS9 are determined by using both PROSITE and SMART web-based tools. The results showed that FUS9 consists of an UBC (Ubiquitin-conjugating enzyme E2, catalytic domain homologues) domain positioning between 39 and 182 residues [21, 46, 47]. According to Yaganava et al. (2004) FUS9 is a member of ubiquitin E2 variant (UEV) proteins which contain the ubiquitin-conjugating motif (Ubc) [60]. However, it is deficient in a cysteine residue which is an essential factor conjugation. These observations suggest that interaction of FUS9 containing complex with CUL4 perhaps enhances the Cullin based E3 ligase activity.

The COP8 is a component and subunit 4 of the COP9 signalosome complex (CSN) which is involved in cellular processes such as auxin, photo morphogenesis, and jasmonate responses [24, 9]. According to Serino Giovanna (1999) genetic mutations at cop8 and fus4 loci describe the same locus, and are now represented as COP8. The COP8 contains a PCI (Proteasome, COP9, Initiation factor 3) domain which is also found in the C-terminal region of several regulatory components of the 26S proteasomes. The PCI domain is still not much known, but it plays an important role during the COP9 signalosome assembly [24, 9, 49]. The COP9 signalosome complex has WD40 motifs and it is involved in the regulation of the CUL4 based Ubiquitin ligases. In the absence of UV irradiation, the signalosome complex is associated with DDB1-CUL4-RBX1-DDB2 complex. However, in the presence of UV-B, COP9 dissociates from this complex and enables translocation of DDB2 complex to the chromatin when it attaches NEDD8 moieties to the CUL4 [8]. Interestingly, CUL4-DDB1 complex is also found to interact with a WDR domain containing protein, COP1 [23]. However, it is unclear whether COP1 interacts directly or indirectly with the DDB1 but it seems to have weaker interaction to the AtCUL4 in comparison to others [59].

In *Arabidopsis thaliana*, the COP1 functions as an E3 ligase for photo morphogenesis-promoting factors [21, 60]. In the dark conditions, the COP1 interacts with other two complexes CSN and CDD complex to form COP/DET/FUS structure and plays an important role for the repression of photo morphogenesis-promoting transcription factors by means of ubiquitin–proteasome-mediated degradation [60, 61, 62]. The domain analysis of the COP1 by using SMART and PROSITE tools, revealed that it consists of 7 WD repeats and a single RING domain. However, unlike DDB2, the repeated WD motifs of COP1 are concentrated in the C-terminal region [46, 47, 49].

In TC-NER, the CSA and AtCHR8 proteins are involved in damage recognition of actively transcribed genes. The CAtHR8 is a member of SNF2 protein family of DNA dependent ATPases. The members of this protein family function in chromatic remodeling as well as transcription elongation and transcription coupled repair systems. Human and yeast homologs of AtCHR8 are CSB and RAD26 respectively. The members of this protein family performs ATP hydrolysis as an energy source to change chromatin structure by altering contacts between DNA and histones [17, 63, 64]. Thereby, the family of SNF2 members are called chromatin remodelers. Disruption in CHR8 homolog genes leads to UV radiation sensitivity in yeast and Cockayne B syndrome in humans [17]. Lake et al. (2013) shows that the CHR8 consists of a central ATPase domain, C-terminal helicase domain and two nuclear localization sequences that direct protein into the nucleus. This also correlates with our

domain analysis because similar results were obtained when AtCHR8 sequence was tested on SMART program. It appears that deletions or miss mutations in N-terminal region do not cause any distortion in the function while mutations in the C-terminal or central domains cause inefficient or defective activity.

The STRING interactome results showed that the AtCHR8 interacts with six proteins including AtERCC1/UVR7, DDB1A, UVH1 (RAD1 homologue in yeast), UVH3 (UVR1), and two similar NRPB9A and NRPB9B proteins (Supplementary materials: Figures from 11 to 15). In order to reveal insights of these interactions, the domains of these proteins are found by SMART and PROSITE domain prediction tools. According to these results, the UVH1 consists of an ERCC6 domain at the C-terminal region. This structural motif is found in several DNA repair endonucleases such as XPF, Rad1 and Mus81 nucleases which function to cleave branched structures during repair process [45, 46, 65]. Zongrang Liu et al. (2000) depicts that this conserved C-terminal region is involved in interactions with the ATERCC1/UVR7 component of AtERCC1/XPF endonuclease complex [21, 66, 24]. In addition to that, the leucine rich N-terminal region was run in BLAST alignment tool. The results showed that the residues between 19 and 61 are highly conserved and found in UVH1 isoforms. Thus, it's suggested that this conserved N-terminal region is involved in protein-protein interactions (66). Furthermore, according to Zongrang Liu et al. (2000) the UVH1 functions as a subunit of a repair endonuclease in NER repair process. Their research also shows that the UVH1 is homologous to human XPF endonuclease which functions together with ATERCC1/UVR7 to cleave at positions of 5' to the site of the UV induced DNA lesion.

Arabidopsis UVH3 (AtXPG) is homologous to human XPG and yeast RAD2 endonucleases [8, 66, 67]. The SMART domain search pointed out that UVH3 consists of an N-terminal Xerodermapigmentosum G N-region (XPGN) and internal XPGI region [45, 46, 65]. The latter consists of many cysteine and glutamine residues which are mostly found in the active site of DNA endonucleases [67]. The XPGI domain and XPGN domain come together to form the catalytic domain of the UVH3 protein. These two regions are conserved in the FEN-1 family of structure-specific endonucleases [65]. It also has an internal helix-hairpin-helix (HhH) motif which is frequently found in non-sequence specific DNA binding motifs [9, 65]. The eukaryotic examples of HhH containing proteins are RAD2 family of 5'-3' exonucleases, eukaryotic 5' endonucleases and some viral endonucleases [68]. According to Sarker et al. (2005), the XPG (human homolog of AtUVH3) and CHR8 both cooperatively and independently bind to stalled RNA polymerase II complexes during transcription process [67, 69]. Their research also illustrates that, on the DNA bubbles, the UVH3 functionally interacts with the AtCHR8 and stimulates its ATPase activity. Rare mutations on UVH3/XPG cause CS-like phenotype in humans [67, 69]. They also concluded that the UVH3, CHR8 and RNA POLII form a supramolecular complex during the initiation of TC-NER. The UVH3 also has higher affinity to transcription sized bubbles which are around 10 to 20 nucleotides in length [69]. Without UVH3, the CHR8 alone binds to DNA bubble with a low consistency, however, the DNA binding of both CHR8 and UVR3 increases when they co-operate. Furthermore, the C-terminal domain of UVH3 is necessary and it stimulates the CHR8 ATPase activity [69].

The two other CHR8-interacting proteins that are identified by STRING are NRPB9A and NRPB9B proteins. These two highly similar proteins are non-catalytic subunits of DNA-directed RNA polymerases II, IV and V [33, 34]. Both of them consist of RNA polymerase subunit 9 domains and C2C2 zinc fingers. Although, there is not much information about these two proteins in literature, they may interact with the CHR8 during damage recognition in Transcription coupled repair.

Scanning force microscopy and co-precipitation experiments of Lake et al. (2013) designate that the CHR8 functions as a dimer. In another study, Christiansen et al (2010) uses gel filtration chromatography to observe whether human CSB (CHR8 homolog for *Arabidopsis*) is oligomer or dimer and she observes that two different masses of complexes are eluted from the chromatogram. One of these two molecules have a molecular mass of 360 kDa which shows that human CSB is a dimer. The CHR8 is also said to associate with APE1 (endodeoxyribonuclease) during repair of apurinic/apyrimidinic sites by stimulating its incision activity [17, 64]. Compared to other chromatin remodeling proteins of SNF2 family, the CHR8 is not bound to chromatin in the normal circumstances, but instead it is found dispersed through nucleoplasm. However, upon UV-

irradiation, the mobility of CHR8 drastically decreases [64, 70]. Lake et al. (2013) also shows that upon UV-irradiation, the CHR8 forms stable associations with chromatin, only if it is able to hydrolyze ATP. The Mfd protein which is a bacterial ATPase is thought to be bacterial counterpart of the CHR8 protein since it displaces stalled RNA polymerase and recruits repair proteins at site of DNA lesion [17].

AtERCC1/UVR7 is *Arabidopsis* homologue of yeast RAD10 and Human ERCC1/UVR7 proteins. AtERCC1/UVR7 together with XPF (RAD1) functions as to cut DNA at 5' position to DNA lesion leaving single strand away from the junction. Each AtERCC1/UVR7 and XPF are unstable without their partners [27, 29, 64, 71, 72]. The C-terminal region of both proteins show similarity and are necessary for protein-protein heterodimeric interactions. The N-terminal truncate version of the ATERCC1/UVR7 is observed to maintain its function in ATERCC1/UVR7 defective cells [29]. Even tough, ATERCC1/UVR7 is instable without its partner, when a small amount of ATERCC1/UVR7 added to ATERCC1/UVR7-defective extracts, it shows comparable incision activity. Dubestet al. (2004) indicate that AtRAD1 mutants show UV hypersensitivity in the dark. This hypersensitivity supports the heterodimeric function of ATERCC1 /XPF complex. Mutations of human XPF results in Xerodermapigmentosum disease [28]. In humans, mutations in the XPA (RAD14) binding site of ATERCC1/UVR7, Asn110 and Tyr145, effect NER process. The XPA and RPA proteins interact with ATERCC1-XPF complex in order to position it on the DNA [64]. The STRING protein interaction analysis showed that like AtCHR8, the ATERCC1/UVR7 also interacts with UVH3, UVH1, NRPB1A and NRPB1B proteins. In addition to that, the ATERCC1/UVR7 is also associated with MSH2 and MSH5 protein. According to Lan et al. (2004), the AtERCC1/UVR7 is not only involved in NER, but it also plays an important role in repair of DNA lesions induced by inter-strand cross-linking (ICL) agents cis-diamminedichloroplatinum (II) (CDDP). In humans, during this process, the C-terminal region of ERCC1/UVR7 (between the 184th and 260th residues), interacts with MHS2 protein [9]. Luciana D. Lario et al (2011) shows that MSH2 gene is up regulated in *Arabidopsis* upon UV-irradiation, thus it is suggesting that mismatch repair may also involve in UV-induced DNA damage responses [73]. Anindya et al. (2006) shows that binding of CHR8 to the DNA leads to recruitment of core NER proteins as well as ERCC1 -XPF in the early step of TCR process [74]. Alternatively, Liu Zongrang et al. (2000) illustrates that ERCC1/UVR7 interacts with UVH1 through a conserved C-terminal domain (between 867th and 948th amino acid residues) of the UVH1.

The COP1 consists of two conserved domains, the RING domain and WD40 rich domain, which are primarily involved in protein-protein interactions [36]. The Zinc finger RING domain is involved protein-protein interactions and determines substrate specificity for ubiquitination. The docking sites results showed that the COP1 binds to the N-terminally located SCOP domain of AtCUL4. In this way, COP1 inserts its positively charged N-terminal domain (RING domain) toward negatively charged C-terminal region of the CUL4 protein. The non-covalent attachment of proteins provided by electrostatic interactions. Supplementary materials: Figure S1 shows the results of AtCUL4-COP1 docking sites.

From the shown interactions of CUL4 protein, four have been chosen for protein-protein docking site prediction analysis, namely FUS9, DET1, COP1 and COP8 with the highest interaction scores for the predicted STRING interactions. During this docking site predictions, the AtCUL4 was considered to be a receptor. In the prediction of docking sites provided by GRAMM-X Protein-Protein Docking Web Server, there are predictions of 10 different COP8 interactions to CUL4. However, since the conserved C-terminal region of AtCUL4 protein was bound to ROC1 during Cullin based E3 ligase assembly, we predicted that COP8 binds to SCOP domain of the AtCUL4 which is located on positions between 93rd and 434th residues within the N-terminal. By using PyMOL visualization tool, we saw that the residues: 172Trp, 176 Cys, 187Leu, 195Ile, 205Glu, 252Lys, 255Thr, 313Ala, and 316Arg on CUL4 and the N-terminal region of the COP8 (Asn32, 35Leu, 71Glu, 75Glu, 78Gln, 86Phe, and 90Ser) are involved in protein-protein interactions. The non-covalent attachment of proteins provided by electrostatic interactions. Supplementary materials: Figure S2 shows the results of AtCUL4-COP8 docking sites.

The FUS9 protein consists of an N-terminal Ubiquitin-conjugating domain which interacts with ubiquitin molecules. However, according to SMART domain analysis, the *Arabidopsis* FUS9 has inactive UBCc domain.

The predicted FUS9 contacts with the AtCUL4 by locating its N-terminal helix and the loop region on the CULLIN domain. According to Pymol visualization tool, the "578His, 852Val, 583Tyr, 586Ile, 590Phe and 645Asp" residues from the CULLIN domain contact with the N-terminal of the FUS9. The electrostatic potential map shows that the N-terminal of FUS9 is positively charged and the CULLIN domain of CUL4 is negatively charged, therefore the CUL4-FUS9 complex can be stabilized by electrostatic interactions. Supplementary materials: Figure S3 shows the results of AtCUL4-FUS9 docking sites.

The DET1 has no recognizable protein binding domains in *Arabidopsis thaliana.* The DET1 negatively regulates plant light responses and is actively associated with CDD proteins including COP10 and DDB1 proteins [76, 77]. During these associations, it may bind to AtCUL4, at the sites close in proximity to DDB1 protein. In the docking site prediction, we showed that the DET1 interacts with SCOP domain of CUL4 through its conserved C-terminal region [75]. Especially, the residues positioning at 129Gly, 197Tyr, 200Leu, 211Ile, 226Arg, and 228Ile on the C-terminal and the residues positioning at 54Phe, 59Leu, 79Leu, 80Thr, 80Thr are involved in protein-protein interactions (177D, 180L, 184S, 245T, 252K, 301H, 304N, 308I, 353Q, 357T, and 361R on CUL4). The two non-covalently interacting proteins are held together by electrostatic interactions. Supplementary materials: Figure S4 shows the results of AtCUL4-DET1 docking sites.

From the shown STRING interactions of ATERCC1/UVR7 protein, six have been chosen for protein-protein docking site prediction analysis, namely NRPB9A, NRPB9B, UVH1, UVH3, MSH2 and MSH5 from the predicted STRING interactions (Supplementary materials: Figures from 5 to 10).

The AtERCC1/UVR7 and UVH1 proteins are both endonucleases and actively play role during DNA repair process. The docking site prediction reveals that both proteins are interacting through their conserved domains. According to predictions, the AtERCC1/UVR7 associates with UVH1 through its HhH domain. By using PymoL visualization tool, we recorded that the residues which are located at positions: 177Tyr, 180Phe, 184Glu, 208Leu, 212Ser, and 216Leu within AtERCC1 are actively involved in protein-protein interactions. Similarly, the PDB domain (The conserved C-terminal domain) of the UVH1 participates in the protein-protein interactions. The PDB domain is also found in human XPF protein. Perhaps, AtERCC1/UVR7 and UVH1 may exhibit functions similar to ERCC1/XPF complex. The interaction of the two proteins are stabilized by electrostatic interactions. As showed in the Supplementary materials: Figure S5, the protein complex is also stabilized by packing alpha helices against each other.

Subsequently, we have predicted UVH3 docking site on the AtERCC1/UVR7. When electrostatic potential of UVH3 is visualized by using PDB deep viewer, we observed that it has negatively charged residues on the surface of the protein. However, when we observe ATERCC1/UVR7 molecule, we saw that the N-terminal region of the protein is positively charged. Thus, the electrostatic interaction is the primary factor of heterodimer stabilization. Eventually, we predicted that the AtERCC1/UVR7 molecule interacts with residues which are positioned at the middle region of the UVH3 protein. The residues on the UVH3 including 904Phe, 909Asp, 984Asp, 996Asp, 1000Glu, 1007Lys, and 1028Ile are involved in protein-protein interactions (HhH domain) with RAD10 domain (131R, 133K, 140H, 169R, 172L, 173L, 201E, and 225W) of the AtERCC1/UVR7 protein. NRPB9A and NRPB9B are two small proteins and both are consisting of nucleic acid binding and RNA polymerase subunit domains. According to GRAMMX docking site prediction server, the part of RAD10 domain of ATERCC1/UVR7 interacts with N-terminal RPOL9 domain and as well as the loop region which connects RPOL9 and C2C2 Zinc finger of NRPB9A. Similar predictions were observed ATERCC1/UVR7 and NRPB9B proteins in which RAD10 domain of the ATERCC1/UVR7 interacts with residues from the RPOL9 and the loop region. In the case of prediction of NRPB9B docking sites, we have saw that the loop region, connecting Zinc finger and RPOL9 domains, is primarily involved in the protein-protein interaction. Similarly, the C-terminal part of the RAD10 domain is participating binding process, however the N-terminal of the RAD10 domain is left intact, and is facing outside of the complex.

MSH2 is a mismatch repair protein which plays an essential role for the post-replicative mismatch repair system. It consists of an N-terminal mismatch recognition domain described as MUT_I and MUT_II in the pfam domain [46]. Additionally, it consists of a core domain that is made up of two separate subdomains acting as levers. It

also has an ATPase domain on the C-terminal of the protein [45, 46]. During the docking site predictions of MSH2, we have taken extra considerations to avoid dockings which interfere with normal functioning of MSH2 in DNA repair mechanism. According to our predictions, the C-terminal of AtERCC1/UVR7, particularly the 5[th]and 4th alpha helices (ordered from N to C) found on the RAD10 domain interact with those residues neighboring ATPase domain of the MHS2 protein. The protein complex is held by electrostatic interactions.

Similarly, MSH5 is also associated with mismatch repair system and it also consists of an ATPase and DNA-binding domain of DNA mismatch repair MUTS family. According to our predictions, the C-terminal of ATERCC1/UVR7, particularly the 5th and 4th alpha helices (ordered from N to C) found on the RAD10 domain interact with C-terminal of the MSH5. The MSH5 residues: 21G, 43E, 46C and 625M, 629H, 632G, 636R, 657D, 661L 665T, 668H, and 672C are involved in protein-protein interactions. The protein complex is held by electrostatic interactions.

Lastly, we have predicted the binding sites of AtERCC1/UVR7, NRPB9A, NRPB9B, UVH1 and UVH proteins to CHR8. In the prediction of docking, the CHR8 was considered to be receptor and the presented other proteins are used as receptors.

According to Pymol visualization of the predicted proteins, the RAD10 domain of ATERCC1/UVR7 was primarily involved in protein-protein interactions. As shown in GRAMM-X docking server, the residues positioning 201Glu, 205Lys, 211Thr, 212Lys, and 215Leu on the RAD10 are interacting with 624Ser, 628Arg, 631Val, 634Arg, 578GLn and 579Asn residues located on the middle region of the AtCHR8. Additionally, 869Phe and 866Tyr residues on the AtCHR8 are interacting with 225Trp and 229Glu residues of AtERCC1. The proteins are held together by packing alpha helices against each other.

In the following docking site prediction analysis, we looked for binding sites of the UVH3 protein to the AtCHR8. While we are predicting docking sites of proteins, we excluded those prediction results involving binding at the functional ATPase domain of AtCHR8 and XPG endonuclease domain of UVH3. Subsequently, we observed that the residues 1015Leu, 1026Gly, 1024Ser, 2017Gly, 1029Val, 1098Ile, 1101Asp, 1108Lys on the HhH2 domain of UVH3 are involved in interactions with N-terminal of CHR8 (398E, 401C, 431K, 459F, and 338E). The electrostatic potential of AtCHR8 shows a very big blue cloud. However, the electrostatic potential of UVH3 shows highly negative. Therefore, we report that the protein-protein interactions are most likely provided by electrostatic interactions.

The docking site predictions of UVH1 shows that it binds to AtCHR8 through its ERCC4 domain. According to Pymol visualization, the alpha helices on the UVH1 C-terminal region (751Q, 755Q, 758T, 760G, 764H, 768M, 771R, 800V, and 803Y) are interacting with loop regions on the CHR8. Similarly, an alpha helix within middle conserved domain of AtCHR8, positioning between 648 and 651 positions, interacts with a loop region found in the C-terminal of the UVH1. The electrostatic potential of AtCHR8 and UVH1 shows that the protein-protein interactions are probably performed via electrostatic interactions.

The NRPB9A and NRPB9B proteins do not contain any buried globular structures but instead both have 3rd and 4th anti-parallel beta strands connected by a loop region. We believe that the loop region is primarily involved in protein-protein interactions while the B-domains provide function. These two proteins may be functioning as a bridge to provide interactions between AtCHR8 and TFIIH or other transcription factors during repair process or perhaps it may enhance the DNA binding specificity of AtCHR8.

According to docking site predictions of NRPB9A, the binding regions consists of residues located on N-terminal and loop regions. The residues on the loop region include: 46Asn, 47Glu, 50His, 53Ser, 54Glu and are recorded to interact with 600P, 602F, 604A, 607S, 627Y and 611T residues located near DEXDc domain of the AtCHR8. According to electrostatic potential of NRPB9A, the interaction of proteins seems to be provided by electrostatic interactions.

Similarly, the NRPB9B interacts with AtCHR8 through its loop region which connects RNP9 and zinc finger domains. When the AtCHR8-NRPB9B complex is visualized on Pymol viewer, we observed that the loop region forms a V-shaped structure slightly positioning toward the AtCHR8 protein. Since the NRPB9A and NRPB9B are closely related and almost identical, we observed that the interacting loop regions show high

similarity. In this case, binding residues include: Met1, Thr3, Met4, Phe6, Val48, Ser51, Ser53, Thr56, Asp61, 93Gly, 95Glu, and 96Glu. On the other side, the residues including 558T, 562K, 555S, 586S, 590F, 600P, 603E, 608V, 624S, 628R, 631V, 881r and 882F on DEXDc domain of AtCHR8 interact with central loop and N-terminal region of the NRPB9B protein.

5. Conclusions

To sum up, the AtCHR8, AtERCC1 and AtCUL4 proteins play active and important role during DNA repair process in Arabidopsis thaliana. The null mutations of these proteins cause major DNA repair flaws as well as plant development abnormalities. Especially, mutations in AtCUL4 can interfere with both TC-NER and GG-NER processes which shows the importance of this protein in DNA repair mechanism. Since plants, like model plant Arabidopsis, are exposed to UVR in larger extent than other organisms, it is really important to understand key proteins in repair process as well as how plants react and protect themselves against such agents.

By using bioinformatics programs, we have analyzed and predicted 3-D structure and domains of these AtERCC1, AtCUL4 and AtCHR8 proteins in Arabidopsis thaliana. Since they are functioning in the same repair mechanism, we subsequently have looked for any interaction between these three and other associated proteins in order to predict possible docking sites. Besides these analyses, the localization of these three are also predicted in order to validate their connection with predicted interacting proteins. From the demonstrated results it can be concluded that the AtERCC1, AtCUL4, and AtCHR8 are associated with diverse number of proteins and each one of them plays important role during the NER process.

The experimental determination of 3D structures and further analysis of interactome and docking sites are required to completely understand the protein-protein interactions and diverse functions of AtERCC1, AtCUL4 and AtCHR8 proteins in critical DNA repair mechanisms and as well as other cellular processes. Furthermore, in future, experimental testing of protein binding domains can allow us to understand and discover how diverse numbers of proteins help to repair and/or tolerate DNA lesions against UV irradiation and other damaging agents.

Supplementary Materials: The following are available in supplementary materials, Figure S1: AtCUL4 is shown in green color and COP1 is shown in cyan color. The direction of the AtCUL4 protein sequence is from N-terminal to C- terminal, from right top to bottom left of the visualization. The pink dots represent the residues which are found at the docking sites, Figure S2: AtCUL4 is shown in green color (Red alpha helices and yellow beta strands) and COP8 is shown in yellow color. The direction of the AtCUL4 protein sequence is from N-terminal to C- terminal, from left top to top right of the visualization. The pink dots represent the residues which are found at the docking sites., Figure S3: AtCUL4 is shown in green color (Red helices and yellow beta strands) and FUS9 is shown in magenta color. The pink dots represent the residues which are found at the docking sites, Figure S4: AtCUL4 is shown in green color having red helices and yellow B-sheets and DET1 is shown in yellow color. The direction of the AtCUL4 protein sequence is from N-terminal to C- terminal, from right top to bottom left of the visualization. The pink dots represent the residues which are found at the docking sites, Figure S5: UVH3 is represented with red alpha helices and yellow beta sheets while UVR7 is shown with cyan helices and magenta beta sheets. The two protein are held together by electrostatic interactions, Figure S6: UVH1 is represented with yellow color. ERCC1/UVR7 is represented with cyan helices and magenta B-sheets. The C- terminal and N-terminal of proteins are denoted by using N and C letters, respectively. The pink dots represent the residues which are found at the docking sites, Figure S7: NRPB9A is represented by green color. The purple dots represent the residues which are interacting to mediate protein-protein interaction. The ERCC1/UVR7 is shown with cyan helices and magenta Beta strands. As visible in the figure, the residues on the N-terminal loop region of NRPB9A and N-terminally located small alpha helix on ERCC1 are interacting and positioning ERCC1 in the crevasse of NRPB9A, Figure S8: The ERCC1/UVR7 is shown with cyan helices and magenta Beta strands and NRPB9B is colored pale yellow. The interacting residues are shown with pink dots, Figure S9: The docking site prediction of ERCC1 for MSH2. The ERCC1 is represented with cyan helices

and magenta beta strands. The MSH2 is colored white orienting from N-terminal to C terminal from top left to right bottom. The pink dots represent the residues which are found at the docking sites, Figure S10: MSH5 is colored blue and UVR7 is shown with cyan helices and magenta beta sheets. The pink dots represent the residues which are found at the docking sites, Figure S11: The docking site prediction of UVR7 for CHR8 protein. The CHR8 helices are colored red and B-sheet are colored yellow. On the UVR7, helices are colored cyan and B-sheets are colored magenta. The pink dots represent the residues which are found at the docking sites, Figure S12: The docking site prediction of UVH3 for CHR8 protein. The CHR8 surface colored green (Red helices and yellow beta strands). The surface of UVH3 is colored purple (cyan helices and red beta strands). The pink dots represent the residues which are found at the docking sites, Figure S13: The docking site prediction of UVH1 for CHR8 protein. The UVH1 is colored pale yellow and CHR8 is colored green with red helices and yellow beta strands. The pink dots represent the residues which are found at the docking sites, Figure S14: The docking site prediction of NRPB9A for CHR8 is shown. The CHR8 is colored green and NRPB9A is colored with cyan helices and magenta beta strands. The pink dots represent the residues in the interfere region and Figure S15: The docking site prediction of NRPB9B for CHR8 is shown. CHR8 is colored green and NRPB9B is colored with pale yellow. The pink dots represent the residues in the interfere region.

LIST OF ABBREVIATIONS

6-4 PPs: 6-4 pyrimidine Pyrimidone Photoproducts
CBP: Cyclobutane Pyrimidine Dimers
CSN: COP9 Signalosome Complex
DNA: Deoxyribose Nucleic Acid
EBI: European Bioinformatics Institute
EMBL: The European Molecular Biology Laboratory
GG-NER: Global Genomic Nucleotide Excision Repair
HMGN1: High Mobility Group Nucleosome Binding Domain containing protein
NCBI: National Center for Biotechnology Information
NER: Nucleotide Excision Repair
NLS: Nuclear Localization Signal Peptides
PSI: Plant Subcellular Localization Integrative Predictor
RCSB-PDB: The Research Collaborator for Structural Bioinformatics- Protein Data Bank
RPA: Replication Protein A
SMART: Simple Modular Architecture Research Tool
STRING: Search Tool for the Retrieval of Interacting Genes/Proteins
TAIR: The Arabidopsis Information Resource
TC-NER: Transcription Coupled Nucleotide Excision Repair
UVR: Ultra Violet Radiation

References

[1] Cooper, G.. Replication, Maintenance, and Rearrangements of Genomic DNA. In The Cell: A Molecular Approach (2nd Ed.), 2000.
[2] Lodish H, Berk A, Zipursky SL, et al. Molecular Cell Biology. 4th edition. New York: W. H. Freeman. Section 12.4, DNA Damage and Repair and Their Role in Carcinogenesis, 2000.
[3] Britt, A.. DNA Repair in Higher Plants. DNA Damage and Repair, 577-596, 1998.
[4] Rastogi, R., Kumar, A., Tyagi, M., & Sinha, R. (2010). Molecular Mechanisms of Ultraviolet Radiation-Induced DNA Damage and Repair. Journal of Nucleic Acids, 1-32.
[5] Dexheimer, T., DNA Repair Pathways and Mechanisms. DNA Repair of Cancer Stem Cells, 19-32, 2012.

[6] Clancy, S., DNA Damage & Repair: Mechanisms for Maintaining DNA Integrity. Nature Education, 2008.

[7] Britt, A., Repair of DNA damage induced by ultraviolet radiation. Plant Physiology, 891-896, 1995.

[8] L., A., & F., D., UV Damaged DNA Repair & Tolerance in Plants. Selected Topics in DNA Repair, 2011.

[9] Lan, L., Functional and physical interactions between ATERCC1/UVR7 and MSH2 complexes for resistance to cis-diamminedichloroplatinum(II) in mammalian cells. DNA Repair, 135-143, 2004.

[10] Lee, J., & Zhou, P., DCAFs, the Missing Link of the CUL4-DDB1 Ubiquitin Ligase. Molecular Cell, 775-780, 2006.

[11] Pan, Z., Faculty of 1000 evaluation for the ubiquitin ligase activity in the DDB2 and CSA complexes is differentially regulated by the COP9 signalosome in response to DNA damage. F1000 - Post-publication Peer Review of the Biomedical Literature, 2003.

[12] Sharma, P., & Nag, A., CUL4A ubiquitin ligase: A promising drug target for cancer and other human diseases. Open Biology, 130217-130217, 2014.

[13] Rakshambikai, R., Srinivasan, N., &Nishant, K., Structural Insights into Saccharomyces cerevisiae Msh4–Msh5 Complex Function Using Homology Modeling. PLoS ONE, 2013.

[14] Macaisne, N., Vignard, J., & Mercier, R., SHOC1 and PTD form an XPF-ATERCC1/UVR7 -like complex that is required for formation of class I crossovers. Journal of Cell Science, 2687-2691, 2011.

[15] Motycka, T., Physical and Functional Interaction between the XPF/ATERCC1/UVR7 Endonuclease and hRad52. Journal of Biological Chemistry, 13634-13639, 2004.

[16] Christiansen, M., Thorslund, T., Jochimsen, B., Bohr, V., &Stevnsner, T., The Cockayne syndrome group B protein is a functional dimer. FEBS Journal, 4306-4314, 2005.

[17] Lake, R., & Fan, H., Structure, function and regulation of CSB: A multi-talented gymnast. Mechanisms of Ageing and Development, 202-211, 2013.

[18] Shaked, H., Involvement of the Arabidopsis SWI2/SNF2 Chromatin Remodeling Gene Family in DNA Damage Response and Recombination. Genetics, 985-994, 2006.

[19] Nocker, S., & Ludwig, P., The WD-repeat protein superfamily in Arabidopsis: Conservation and divergence in structure and function. BMC Genomics, 2003.

[20] Zhang, Y., Feng, S., Chen, F., Chen, H., Wang, J., Mccall, C., Deng, X., Arabidopsis DDB1-CUL4 ASSOCIATED FACTOR1 Forms a Nuclear E3 Ubiquitin Ligase with DDB1 and CUL4 That Is Involved in Multiple Plant Developmental Processes. The Plant Cell Online, 1437-1455, 2008.

[21] Orengo, C.A., Pearl, F.M.G., Bray, J.E., Todd, A.E., Martin, A.C., Lo, C.L. and Thornton, J.M., The CATH Database provides insights into protein structure/function relationships. Nucleic Acids Res., 27, 275-279, 1999.

[22] Higa, L., & Zhang, H., The SCF Ubiquitin E3 Ligase. Protein Degradation Series, 135-155, 2007.

[23] Higa, L., Wu, M., Ye, T., Kobayashi, R., Sun, H., & Zhang, H., CUL4–DDB1 ubiquitin ligase interacts with multiple WD40-repeat proteins and regulates histone methylation. Nature Cell Biology Nat Cell Biol, 1277-1283, 2006.

[24] Serino, G., Arabidopsis cop8 and fus4 Mutations Define the Same Gene That Encodes Subunit 4 of the COP9 Signalosome. The Plant Cell Online, 1967-1980, 1999.

[25] Serino, G., Characterization of the Last Subunit of the Arabidopsis COP9 Signalosome: Implications for the Overall Structure and Origin of the Complex. The Plant Cell Online, 719-731, 2003.

[26] Zhang, C., Guo, H., Zhang, J., Guo, G., Schumaker, K., &Guo, Y., Arabidopsis Cockayne Syndrome A-Like Proteins 1A and 1B Form a Complex with CULLIN4 and Damage DNA Binding Protein 1A and Regulate the Response to UV Irradiation. The Plant Cell Online, 2353-2369, 2010.

[27] Tsodikov, O., Enzlin, J., Scharer, O., & Ellenberger, T., Crystal structure and DNA binding functions of ATERCC1/UVR7 , a subunit of the DNA structure-specific endonuclease XPF-ATERCC1/UVR7 . Proceedings of the National Academy of Sciences, 11236-11241, 2005.

[28] Dubest, S., Gallego, M., & White, C., Roles of the AtERCC1/UVR7 protein in recombination. The Plant Journal, 334-342, 2004.

[29] Gaillard, P., Activity of individual ATERCC1/UVR7 and XPF subunits in DNA nucleotide excision repair. Nucleic Acids Research, 872-879, 2001.

[30] Wang, X., CSN1 N-Terminal-dependent Activity Is Required for Arabidopsis Development but Not for Rub1/Nedd8 Deconjugation of Cullins: A Structure-Function Study of CSN1 Subunit of COP9 Signalosome. Molecular Biology of the Cell, 646-655, 2002.

[31] Lamesch, P., Berardini, T., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., . . .Huala, E., The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. Nucleic Acids Research, 2011.

[32] Laskowski R A, MacArthur M W, Moss D S, Thornton J M., PROCHECK - a program to check the stereochemical quality of protein structures. J. App. Cryst., 26, 283-291, 1993.

[33] Westbrook, J., Feng, Z., Chen, L., Yang, H. and Berman, H.M., The Protein Data Bank and structural genomics. Nucleic Acids Res., 31, 489-491, 2003.

[34] Wu, X., Xu, K., & Her, C., The Role of MutS Homologues MSH4 and MSH5 in DNA Metabolism and Damage Response. DNA Replication and Related Cellular Processes, 2011.

[35] Kalinina, Olga V., and Thomas Lengauer. "Protein Structure Prediction and databases." ELS pag. Web."UV: Molecular Mechanism of Action." UV: Molecular Mechanism of Action. N.p., n.d. Web. 19 Nov. 2015.

[36] Schwede, T., Kopp, J., Guex, N. and Peitsch, M.C. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res., 31, 3381-3385, 2003.

[37] Guex N and Peitsch MC., SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. Electrophoresis 18:2714-2723, 1997.

[38] Benkert, P., Schwede, T. and Tosatto, S.C.E., "QMEANclust: Estimation of protein model quality by # combining a composite scoring function with structural density information." BMC Struct Biol. May 20; 9:35, 2009.

[39] Benkert P, Künzli M, Schwede T., "QMEAN Server for Protein Model Quality Estimation." Nucleic Acids Res. 2009 Jul 1; 37(Web Server issue): W510-4, 2009.

[40] Benkert, P., Biasini, M. and Schwede, T., "Toward the estimation of the absolute quality of individual protein structure models." Bioinformatics (2010). doi: 10.1093/bioinformatics/btq662

[41] Benkert, P., Tosatto, S.C.E. and Schomburg, D., "QMEAN: A comprehensive scoring function for model quality assessment." Proteins: Structure, Function, and Bioinformatics, 71(1):261-277, 2008.

[42] Chamovitz, D., Wei, N., Osterlund, M., Arnim, A., Staub, J., Matsui, M., & Deng, X., The COP9 Complex, a Novel Multisubunit Nuclear Regulator Involved in Light Control of a Plant Developmental Switch. Cell, 115-121, 1996.

[43] S.C. Lovell, I.W. Davis, W.B. Arendall III, P.I.W. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson and D.C. Richardson, Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins: Structure, Function & Genetics. 50: 437-450, 2002.

[44] The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.

[45] Letunic et al., "SMART 6: recent updates and new developments".Nucleic Acids Res, 2014, doi: 10.1093/nar/gku949

[46] Schultz et al., ""SMART, a simple modular architecture research tool: identification of signaling domains." Proc. Natl. Acad. Sci. USA 95, 5857-5864, 1998.

[47] De Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N.ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins.Nucleic Acids Res. Jul 1;34, 2006 .

[48] Sigrist CJA, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I. New and continuing developments at PROSITE Nucleic Acids Res. 2012; doi: 10.1093/nar/gks1067 PubMed: 23161676

[49] Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Barrell, D., Bateman, A., Binns, D., Biswas, M., Bradley, P., Bork, P. et al., The InterPro Database, 2003 brings increased coverage and new features. Nucleic Acids Res., 31, 315-318, 2003.

[50] Strzalka, W., &Ziemienowicz, A., Proliferating cell nuclear antigen (PCNA): A key factor in DNA replication and cell cycle regulation. Annals of Botany, 1127-1140, 2010.

[51] Szklarczyk et al. 2015 / 2013 / 2011 / 2009 / 2007 / 2005 / 2003 / Snel et al. 2000.

[52] Tantin, D., Kansal, A., & Carey, M., Recruitment of the putative transcription-repair coupling factor CSB/ERCC6 to RNA polymerase II elongation complexes. Molecular and Cellular Biology Mol. Cell. Biol., 6803-6814, 1997.

[53] Lili Liu, Zijun Zhang, Qian Mei, Ming Chen*, PSI: A comprehensive and integrative approach for accurate plant subcellular localization prediction, PLoS One, 2013, DOI: 10.1371/journal.pone.0075826

[54] An SVM-based system for predicting protein subnuclear localizations
BMC Bioinformatics, 6:291, 2005.

[55] Anindya, R., Mari, P.-O., Kristensen, U., Kool, H., Giglia-Mari, G., Mullenders, L. H., Svejstrup, J. Q., An Ubiquitin-Binding Domain in Cockayne Syndrome B Required for Transcription-Coupled Nucleotide Excision Repair.Molecular Cell, 38(5), 637–648, 2010, http://doi.org/10.1016/j.molcel.2010.04.017

[56] Tovchigrechko A, Vakser IA. GRAMM-X public web server for protein-protein docking. Nucleic Acids Res, 34: W310-4, 2006.

[57] Zhou, H., and Zhou, Y., "Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction." Protein Sci. 11:2714-2726, 2002.

[58] Xu, H., Swoboda, I., Bhalla, P., Sijbers, A., Zhao, C., Ong, E. Singh, M., Plant homologue of human excision repair gene ATERCC1/UVR7 points to conservation of DNA repair mechanisms. The Plant Journal Plant J, 823-829, 1998.

[59] He, Y., Mccall, C., Hu, J., Zeng, Y., &Xiong, Y., DDB1 functions as a linker to recruit receptor WD40 proteins to CUL4-ROC1 ubiquitin ligases. Genes & Development, 2949-2954, 2006.

[60] Yanagawa, Y., Arabidopsis COP10 forms a complex with DDB1 and DET1 in vivo and enhances the activity of ubiquitin conjugating enzymes. Genes & Development, 2172-2181, 2004.

[61] Vannier, J., Depeiges, A., White, C., &Gallego, M., ATERCC1/UVR7 /XPF Protects Short Telomeres from Homologous Recombination in Arabidopsis thaliana. PLOS Genetics PLoS Genet, 2009.

[62] Kang, J., Huang, S., &Blaser, M., Structural and Functional Divergence of MutS2 from Bacterial MutS1 and Eukaryotic MSH4-MSH5 Homologs. Journal of Bacteriology, 3528-3537, 2005.

[63] Lehmann, A., Faculty of 1000 evaluation for Cockayne syndrome A and B proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled RNA polymerase II in vivo. F1000 - Post-publication Peer Review of the Biomedical Literature, 2006.

[64] Tovchigrechko A, Vakser IA. Development and testing of an automated approach to protein docking. Proteins; 60(2):296-301, 2005.

[65] Bairoch, A., R. Apwe,ler, et al., "The Universal Protein Resource (UniProt)." Nucleic Acids Res 33 Database Issue: D154-159, 2005.

[66] Liu, Z., Hossain, G., Islas-Osuna, M., Mitchell, D., & Mount, D., Repair of UV damage in plants by nucleotide excision repair: Arabidopsis UVH1 DNA repair gene is a homolog of Saccharomyces cerevisiae Rad1. The Plant Journal Plant J, 519-528, 2000.

[67] Clarkson, S., The XPG story. Biochimie, 1113-1121, 2003.

[68] Mueser, T., Nossal, N., & Hyde, C., Structure of Bacteriophage T4 RNase H, a 5′ to 3′ RNA–DNA and DNA–DNA Exonuclease with Sequence Similarity to the RAD2 Family of Eukaryotic Proteins. Cell, 1101-1112, 1996.

[69] Sarker, A., Tsutakawa, S., Kostek, S., Ng, C., Shin, D., Peris, M, Cooper, P., Recognition of RNA Polymerase II and Transcription Bubbles by XPG, CSB, and TFIIH: Insights for Transcription-Coupled Repair and Cockayne Syndrome. Molecular Cell, 187-198, 2005.

[70] Boom, V., Citterio, E., Hoogstraten, D., Zotter, A., Egly, J., Cappellen, W., . . .Vermeulen, W., DNA damage stabilizes interaction of CSB with the transcription elongation machinery. The Journal of Cell Biology, 27-36, 2004.

[71] Orelli, B., Mcclendon, T., Tsodikov, O., Ellenberger, T., Niedernhofer, L., &Scharer, O., The XPA-binding domain of ATERCC1/UVR7 Is Required for Nucleotide Excision Repair but Not Other DNA Repair Pathways. Journal of Biological Chemistry, 3705-3712, 2009.

[72] Tripsianes, K., Folkers, G., Ab, E., Das, D., Odijk, H., Jaspers, N., . . .Boelens, R., The Structure of the Human ATERCC1/UVR7 /XPF Interaction Domains Reveals a Complementary Role for the Two Proteins in Nucleotide Excision Repair. Structure, 1849-1858, 2005.

[73] Lario, L., Ramirez-Parra, E., Gutierrez, C., Casati, P., &Spampinato, C., Regulation of plant MSH2 and MSH6 genes in the UV-B-induced DNA damage response. Journal of Experimental Botany, 2925-2937, 2011.

[74] Tripsianes, K., The structure of the human ATERCC1/UVR7 /XPF interaction domains reveals a complementary role for the two proteins in nucleotide excision repair. In Structural and functional studies of the human ATERCC1/UVR7 /XPF DNA repair complex, 2008.

[75] Lo Conte, L., Brenner, S.E., Hubbard, T.J.P., Chothia, C. and Murzin, A.G., SCOP database in 2002: refinements accommodate structural genomics. Nucleic Acids Res., 30, 264-267, 2002.

[76] Pick, E., Lau, O., Tsuge, T., Menon, S., Tong, Y., Dohmae, Wei, N., Mammalian DET1 Regulates Cul4A Activity and Forms Stable Complexes with E2 Ubiquitin-Conjugating Enzymes. Molecular and Cellular Biology, 4708-4719, 2007.

[77] Kang, M., Kwon, H., Kim, N., Sakuraba, Y., &Paek, N., CONSTITUTIVE PHOTOMORPHOGENIC 10 (COP10) Contributes to Floral Repression under Non-Inductive Short Days in Arabidopsis. IJMS International Journal of Molecular Sciences, 26493-26505, 2015.