ABSTRACT

The role of DNA repair reaction is a critical mechanism in the human body since DNA is constantly threatened and attacked by entities of an endogenous or external nature. One such DNA repair pathway, nucleotide excision repair (NER), deals with bulky DNA damages caused by ultraviolet radiation, toxic chemicals, and anti-cancer medications. These damages include cyclobutane pyrimidine dimers (CPD), 6-4 photoproducts (6-4PP), and helix-distorting platinum (Pt) crosslinks. NER is a multi-step, complex process involving more than 20 proteins. Because it is functional in both transcription-coupled NER (TC-NER) and global genome NER (GG-NER), any change in NER participation causes genetic diseases such as Xeroderma Pigmentosum (XP), Cockayne syndrome (CS), and Trichothiodystrophy (TTD). Additionally, it has been noted that people with cancer who overexpress NER may develop resistance to platinum-based chemotherapy. The majority of NER proteins have been thoroughly investigated, but the Xeroderma pigmentosum complementation Group A (XPA) protein has not been fully understood and is still a mystery.

A very small protein with 273 residues, and being the central member of the NER core member, XPA is typically present in the early stages of NER during the damage recognition, verification, and the formation of the pre-incision complex (PIC) step. Three primary domains of XPA have been distinguished: the central globular domain or DNA binding domain (DBD), the N-terminal domain, and the C-terminal domain. DBD spans residues aa98-219 and has a Zinger-finger motif at its N-terminus and a basic cleft at its C-terminus. This DBD has now been characterized between residues aa98-239. Additionally, despite not having any enzymatic activity, XPA is known as a "Scaffold protein" because it has multiple sites for binding to other NER proteins. By systematizing the assembly of these NER proteins at the lesion site, XPA ensures that the lesion is removed properly. It does not have any enzymatic activity of its own in NER. Through its DBD, XPA is known to bind selectively with DNA that has been damaged, but in which direction is yet to be clear. The other core NER members that XPA interacts with are (i) TFIIH, transcription factor II helicase, which recruits XPA to the lesioned site, and unwinds the DNA forming NER bubble, (ii) RPA, replication protein A, which protects the undamaged stand, (iii) XPE, Xerodemrna pigmentosum complementation group E

protein, involved in damage recognition, and verification, (iv) XPC, Xeroderma pigmentosum complementation group C protein, another DNA damage recognizing protein in GG-NER, and (v) ERCC1, excision repair cross-complementation group 1 protein, via various protein-protein interactions (PPIs) forming protein-protein complexes (PPCs). These proteins collectively form a multiprotein complex known as the PIC. Even though there is a known NER protein malfunction that causes XP disease, it is known that XPA protein mutations cause complete NER failure, which results in the deadliest type of XP known as classical XP. Extreme photosensitivity, developmental delay, and neurodegeneration are the hallmarks of classical XP. This phenotype is also marked by a high incidence of melanomas and carcinomas, occasionally even internal malignancies, pre-eclampsia, and occasionally even Alzheimer's disease.

The DBD of XPA had previously been mapped between aa98-219 by Ikegami et al. in 1998 (PDB ID: 1XPA), and by Buchko et al. in 1999 (PDB ID: 1D4U) through NMR studies, but this has since been changed to between aa98-239 after the studies conducted by Sugitani et al. (2014, and 2017), and by Hilton et al. (2014) since the prior area lacked the fundamental lysine residues necessary to bind DNA correctly. Lian et al in 2019, and 2020 (PDB ID: 6J44, and 6LAE), Barnet et. in 2020 (PDB ID:7AD8), and Kokic et al. in 2019 (PDB ID: 6R04) independently figured out the 3D structure of redefined DBD of XPA using X-ray, and Electron microscopy technique. Most research investigations presented XPA as a monomer, however more recent studies have shown that XPA can also exist as a homodimer. The research by Liu et al. (2005), and Yang et al. (2002) demonstrated that XPA binds to DNA and RPA in a 2:1 ratio and that the dimer form of XPA predominates over the monomer form. The following piece of proof for XPA's dimer status was provided by Gilljam and her team (2012), who noted that XPA in dimer form formed foci with proliferating cell nuclear antigen (PCNA), another NER scaffold protein that is only detected during the ligation phase of NER status, which places another question whether XPA is involved in ligation or not. Hence to understand the property of XPA in monomer and dimer forms, we applied a molecular dynamics (MD) approach to study it. We took 1XPA as our model to represent the XPA monomer and due to the unavailability of the 3D structure of redefined DBD of

XPA, we modeled it computationally by joining the 1XPA (aa-98-210) structure with the RaptorX server designed XPA residues using the copy/combine tool of UCSF Chimera. Then we created the homodimers of $XPA_{98-210/98-239}$ using the SymmDock server. The MD simulation showed the stability of redefined XPA_{98-39} over XPA_{98-210} in both monomer and dimer forms. We also saw the presence of the fourth helix as predicted by Hilton et al. Additionally, it was observed that XPA_{98-239} contains more hydrogen bonds, salt bridges, and hydrophobic interactions than the homodimer XPA_{98-210}. We also discovered that in XPA homodimers, Lys, Glu, Gln, Asn, and Arg residues contributed significantly to the intermolecular interactions. The distance between monomers of XPA dimer was observed to be ~20-22 Å apart, which may be utilized to bind DNA during the NER process. The binding free energy (BFE) analysis using MMPBSA/GBSA for both XPA homodimers supported the finding that the XPA_{98-239} homodimer is stable.

The ssDNA and dsDNA junctions as well as the RPA70 domain are all recognized by XPA, although it favors ssDNA-dsDNA or Y-junctions over ssDNA or duplex at Y-junction, binding in the 5| site of the lesion. The research by Buchko et al. (2001) and Missura et al. (2001) demonstrated that XPA can detect kinked DNA distortions more quickly than other kinds of DNA aberrations. Hence, we attempted to study the structural dynamics of XPA in both monomer and dimer forms with DNA next. Comparing XPA_{98-210} in its apo form to the presence of the damaged DNA, greater structural changes were visible. Analysis of the trajectory of the B-factor, SASA, RMSD, and Rg also supported the same result. Due to the creation of distinctive longer antiparallel β-sheets, primarily made up of lysine residues, when XPA is in contact with DNA. DNA was kinked during the simulation process, which was experimentally proved by Koch et al. (2015). Based on the surface structure analysis and the data from the DNAproDB server, we noticed helices were located close to the DNA. For the study of XPA homodimer (XPA_{98-239}) with the DNA, we created the DNA-protein complex (DPC) using the HexDock software. The XPA homodimer's ability to bind DNA was then investigated using the MD study. Additionally, we characterized the DNA-protein interaction (DPIs) between the two DPCs. The ASA values for the XPA homodimer in B and U values, with the probe radius of 1.4, were calculated using the PISA server to find out

whether this DPI is a partner attraction effect or partner accommodation effect. We observed the stability of this DPC throughout the simulation time period. Higher intermolecular H-bonds and hydrophobic interactions were detected during DPI profiling from the DNAproDB server. BFE analysis using the MM-PBSA technique showed a higher binding affinity between the DNA and protein (-62.33 kcal mol^{-1}). According to the per-residue (PRED) study, the C-terminal end of the XPA homodimer's residues K213, K217, K221, K222, K224, K236, E225, R228, and R237 were involved in the DPI, which was experimentally shown by Sugitani et al. (2017). Upon comparing XPA homodimer in both the DNA-bound (B) and DNA-unbound (U) states, we saw a rise in the size of ASA, the number of interface residues, and the region for the XPA homodimer residues in the B state. Calculating the ASA values for B and U states revealed that the partner attraction effect was responsible for these changes.

XPA has been known to interact with XPE part of DNA damage binding 2 protein and XPE (DDB2/XPE) complex with residues aa185-236, however, the residues of XPE involved in this PPI were not known. They both are responsible for DNA damage recognition in GG-NER. Moreover, it has been observed the R207G mutation of XPA resulted in its failure to form PPC with XPE, causing the collapse of the entire NER process. Thus, we strived to understand these questions computationally. Three different docked models were used to examine the interaction between the $XPA_{185-226}$-XPE complex. We observed that Model and Model 3 had greater values than Model 2 during the simulation study of RMSD, SASA, and hydrogen bond analyses. Additionally, it was observed that the orientation of XPA varied by about 180 degrees in Models 1 and 3, while remaining unchanged in Model 2. This indicates that XPA interacts with XPE with its N-terminal end facing below and C-terminal end facing upward. When XPA's DNA binding domain (DBD) region (aa98-239) was bound to XPE, the same thing happened, we obtained the same result. Major residues from XPE, including Arg20, Arg47, Asp51, and Leu57, as well as Leu191, Gln192, Val193, Trp194, Glu198, Glu202, Glu205, Arg207, Glu209, Gln216, and Phe219, were observed in all models as being in charge of this complex's PPI. To fit XPA, the N-terminal (cleft) of XPE was observed to have stretched. The binding affinity between these proteins was discovered to be reliant on

the residues R20, R47, and L57 of XPE and the residues L191, V193, W194, E202, E205, R207, and F219 of XPA, respectively. BFE analyses showed higher binding between XPA-XPE complex. The PPI between the PPC of XPA-XPE was observed to be transient. In comparison to WT XPA, it was discovered that the atomic fluctuations in R207G XPA were significantly higher, and were unstable. We also observed the disassociation of XPA from XPE when mutated to the R207G position *in silico*. XPA had higher binding towards XPA in WT compared to its R207G mutant form. Overall, we discovered that R207 plays a critical role in controlling the protein function of XPA that can be impacted by the R207G mutation, having a significantly greater effect on its ability to connect to XPE as well other NER proteins.

Even though numerous research groups have established the solution structure of XPA (aa98-219/239). The full-length 3D structure of XPA is not yet produced because of the disordered nature (aa1-97 and aa240-273) on either side of the DBD region. Moreover, the majority of studies on XPA so far have been done taking solely the DBD region of XPA. This makes it even more challenging to analyze the PPCs and the location of PPIs between the full-length XPA and its fellow NER members. Although Brosey et al. (2013) provided a theoretical model for the PIC that included XPA and other fellow NER core members, it has not yet been achieved experimentally. Thus, we sought to explore the nature of XPA with its fellow members of PIC in the NER pathway computationally. We used the FASTA sequence of XPA from the UniProt database (ID number P23025) and then submitted it to the I-TASSER server, which allowed us to computationally compute the full-length 3D structure of XPA. 3D structure of XPA validated with various online servers. The ClusPro server was used to create the full-length XPA homodimer, followed by docking XPA to the members of PIC in both monomer and dimer forms. Docking was conducted in the sequence in which they appear in NER to form PIC. We then described its PPIs together with those of other PIC members. We saw that RPA70AB and ERCC1 were tethered to either side of the DBD region of XPA. The C-terminal region of XPA, which is largely made up of acidic sections of XPA, was where the P52 and p8 subunits of the TFIIH complex were associated. On the other hand, XPA's N-terminal region was linked to RPA32C. We also noticed PPIs

between ERCC1 and RPA70AB, as well as p8 of the TFIIH complex and RPA70AB, which were mostly maintained by a high number of hydrogen bonds, salt bridges, and hydrophobic contacts. XPA homodimer shared similar characteristics with XPA monomer while interacting with PIC members. Particularly, XPA2 was found to bind to the p5 protein of TFIIH, indicating that it helps TFIIH function as a helicase to start the NER process.