CHAPTER 6

To study the protein-protein interaction between the members of the pre-incision complex (PIC)

To study the protein-protein interaction between the members of the pre-incision complex (PIC)

6.1 Abstract

Protein-protein interactions (PPIs) are integral to the functioning of cellular mechanisms. Most of these regulating and signaling pathways are often dependent on disordered proteins. One such protein is XPA (*XP* complementation group A) protein, which is monumental in the initiation of the nucleotide excision repair (NER) pathway, acting as a scaffold protein. Most of the symptoms associated with *XP* diseases are majorly associated with mutations in XPA, resulting in neurological damage, and skin cancers. Due to its disordered nature, and lack of full-length structure, the mechanistic regulations of XPA in NER are still a big debate. Here, we have computationally determined the full-length 3D structure of XPA, validated the structure using various tools, and then docked the proteins in the sequence in which they appear in NER forming PIC using XPA in both monomer and dimer forms. We further characterized its PPIs with fellow members of the pre-incision complex (PIC) with XPA in both forms. We observed that RPA70AB and ERCC1 were bound to the DNA binding domain (DBD) of XPA. The p52 and p8 subunits of the TFIIH complex interacted with the C-terminal region of XPA, while the RPA32C was bound to the N-terminal region of XPA. The second unit of XPA homodimer, in particular, was bound to the p5 protein of TFIIH, indicating that it aids TFIIH in helicase activity to initiate the NER process, as mentioned in previous works of literature. Both the PICs, whether in monomer and homodimer form of XPA, were stabilized by hydrogen bonding, salt bridges, and numerous hydrophobic interactions. All the interactions between the members of PIC were observed to be transient.

6.2 Introduction

In biological systems, many biomolecules come and interact with each other forming various protein-protein complexes (PPCs), DNA-protein complexes (DPCs), protein-lipid complexes (PLCs), and protein-carbohydrate complexes (PCCs). This intricate molecular fashion in which they are put together to conduct different cellular

functions is known as ‘protein sociology’ [476, 477]. Understanding these cellular activities necessitates an in-depth knowledge of the molecules' 3D structures, as well as the conditions and architectural elements that govern their interactions [476]. Because they are present in their endogenous host in such negligible quantities, many important complexes remain difficult to purify from their natural sources. Over the past few decades, numerous expression systems, particularly for heterologous protein expression in *Escherichia coli*, have been created and improved. Recombinant overproduction can eliminate this bottleneck. Recently, polycistronic mRNA transcripts or multiple coexisting plasmids in the same cell have been used to build *E. coli* expression systems for the coexpression of multiple proteins. *E. coli* cannot, however, efficiently synthesize many eukaryotic protein complexes. They might comprise subunits that are too big for the transcription and translation machinery of *E. coli*, or they might need specialized chaperone systems or protein modifications (such as phosphorylation or acetylation) that *E. coli* is unable to give. Therefore, the availability of strong eukaryotic expression technologies is essential for the successful overproduction of these complexes, which is needed to understand their structure and function [478-484].

One such multiprotein complex difficult to decipher is the protein involved in the nucleotide excision repair (NER) pathway, especially the proteins that form the pre-incision complex (PIC). NER possesses the ability to recognize the DNA distortion, excise it and then repair it [1, 3-5, 10, 46-50 52, 55, 72, 164]. Cyclo-butane pyrimidine dimers (CPD), photoproducts, and cisplatin-DNA intra-strand crosslinks caused by environmental mutagens, ultraviolet light, or anticancer drugs are the most prevalent forms of lesions healed by NER [57, 58, 72, 75-77, 126]. Since the NER process is pivotal to the damage repair process, any defects in the protein involved in this process result in various types of skin cancers, neurological damage, and premature aging as seen in Xeroderma pigmentosum (XP), Cockayne syndrome (CS), and trichothiodystrophy (TTD) [404, 405].

Xeroderma pigmentosum complementation group A protein (XPA) is the principal component of the nucleotide excision repair (NER) pathway. The Scaffold protein XPA, being the primary damage verifier, assembles all the NER protein

members involved in this multi-step process to the damaged site and ensures that the damaged site is excised effectively [57, 58, 72, 73, 75-77, 126, 176, 177]. The majority of XPA mutations have been linked to the Classical XP phenotype [165, 372, 404, 406]. The transcription factor II H (TFIIH) complex attracts XPA to the damaged site by its helicase activity, which relaxes duplex DNA around the mistaken site to create the NER bubble [75, 144, 357, 361, 370, 407-411, 488]. XPA then functions in both global genome NER (GG-NER) and transcription-coupled NER (TC-NER) [86, 361].

XPA is a 273-residue long protein with disordered C- and N-terminal domains and a globular DNA binding domain (DBD). This DBD region was previously described between aa98-219 [114-116], but has now been redefined between aa98-239 and is expressed by the XPA gene. The main PPIs of XPA have been seen with the replication protein A 70AB protein (RPA70AB), replication protein A 32C protein (RPA32C), excision-repair cross-complementing group 1 (ERCC1), p8 and p52 subunits of the TFIIH protein [48, 123, 125, 127, 142, 144-146, 165, 355, 368- 371, 488, 489].

In most of the research studies, XPA was depicted as a monomer, but in the recent past, studies have revealed that XPA exists in the form of a homodimer [73, 83, 129-131]. The studies conducted by Liu et al. [131] showed that XPA and DNA bind in a 2:1 ratio and that the XPA dimer is the more dominant form of XPA than the monomer form. The next piece of evidence for XPA's dimer status was presented by Gilljam and her team [83] where they observed 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. The binding of XPA with RPA in XPA2-RPA [130], and that XPA may be present as a monomer in presence of DNA and dimer in absence of DNA [73].

Even though the solution structure of XPA (aa98-219/239) has been determined by many groups [74, 114-116, 118, 119]. Most of the studies regarding XPA have been done considering only its DBD, as the full-length structure is not available yet due to the disordered nature (aa1-97, and aa240-273) on either side of the DBD region [57, 72]. The unavailability of the 3D structure of full-length XPA

makes it difficult to study the protein-protein interactions (PPIs) and the placement of protein-protein complexes (PPCs) between the XPA and its fellow NER members. Brosey et al. [133] proposed a theoretical model for the PIC consisting of XPA and other fellow NER core members, but experimentally it has not been realized yet.

Here, in this study, we have computationally determined the full-length structure of XPA in its monomer, and dimer form using the Iterative Threading ASSEmbly Refinement (I-TASSER) server [448] and studied its PPIs and PPC between these proteins using the PDBsum server [348, 349, 389].

6.3 Systems and methodology

6.3.1 Structure modeling of full-length XPA

We obtained the FASTA sequence of XPA (aa1-273) from the UniProt database [378] having the ID number P23025 and submitted the amino acid residues to the I-TASSER server for its structure determination, where the server uses the profile-profile alignment (PPA) threading technique to determine the best structure. The five resultant structures are provided, out of which the structure having the highest confidence score or C-score, template modeling (TM) score, and root-mean-square deviation (RMSD), was taken into consideration as the best-modeled structure.

6.3.2 Energy minimization and structural validation of XPA

The structural of the best-modeled structure obtained from the previous process was submitted to the KoBaMIN server [485-487] for structural refinement and energy minimization. This server, in particular, uses stereochemistry correction and knowledge-based potential of mean force for its process.

The refined structure was then validated using RAMPAGE [383], MolProbity [450], and ProSA server [382]. We validated our structure before and after the minimization using the KoBaMIN server.

6.3.3 Sequential molecular docking of XPA with the members of PIC

We used ClusPro online server [124, 345-347, 447] for the construction of our full-length XPA homodimer and PPCs involved in the formation of PIC. The working

principle of this server is the same as discussed in the CH6A. The protein-protein docking was done in a sequence as to how XPA is recruited to the damaged site by p8 and p52 subunits of TFIIH protein [48, 123, 125, 127, 142, 144-146, 165, 355, 368- 371, 488, 489], followed by an assemblage of remaining members of PIC to the damaged site by the Scaffold protein, which is XPA. We docked XPA with the fellow members of PIC in monomer as well as in dimer forms. The mechanism of the docking process is depicted in **Figures 6.1** and **6.2.** The PPCs with the highest number of populated clusters, cluster centers, and the lowest energy weighted scores were chosen for the study.

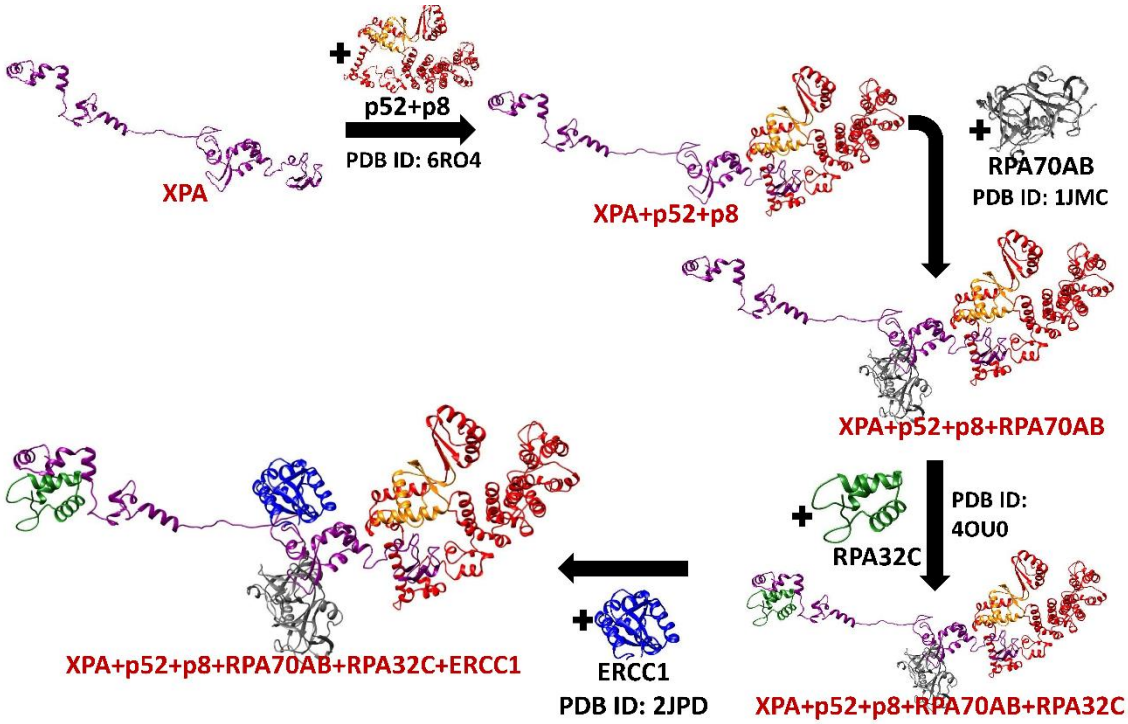

Figure 6.1. Molecular docking of the XPA monomer with other members of PIC.

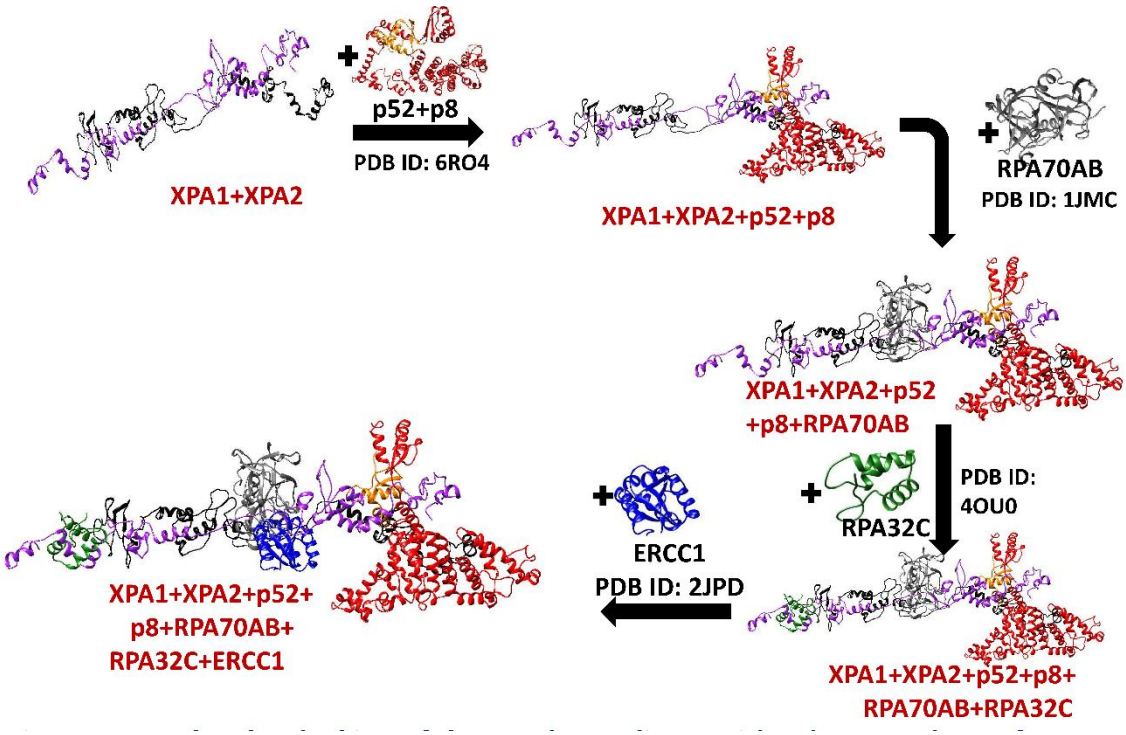

Figure 6.2. Molecular docking of the XPA homodimer with other members of PIC.

6.3.4 PPI study using PDBsum server

A region where two or more sets of proteins come into contact with each other is typically referred to as an interface area. Surface residues with larger surface regions accessible to the solvent available are common characteristics observed in an interface region [393]. Final structures of PIC with XPA in monomer and dimer forms were uploaded to the PDBsum server to determine the intermolecular interactions between them.

6.4 Results and discussion

6.4.1 Structural modeling and the validation of XPA protein, and the selection of an ideal XPA

We obtained five solution structures (**Figure 6.3**) for full-length XPA upon the submission of its FASTA sequence to the I-TASSER server. We chose model 1 as our best-fit structure as it had a good C-score of -2.73, TM-score of 0.40±0.13, and RMSD of 12.4±4.30Å.

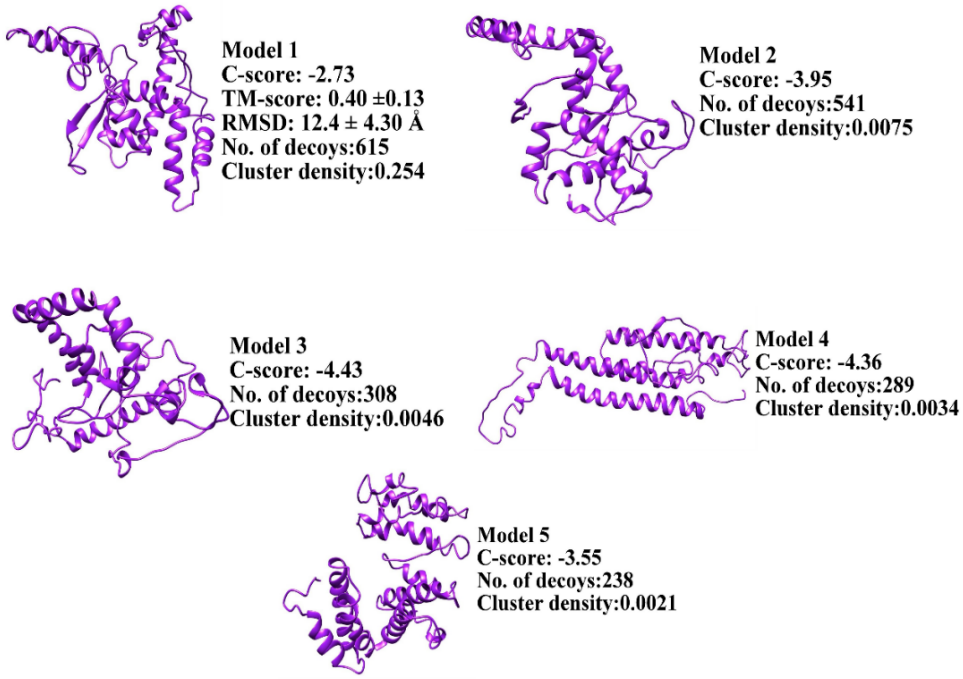

Figure 6.3. Top five models generated for full-length XPA by I-TASSER server.

Initially, upon submission of our model to RAMPAGE, ProSA server, it showed poor results. A RAMPAGE analysis of our model showed 86.8% of the results in the favored region, 9.2% residues in the allowed region3% in the outlier region. the energy plot of the same had some of the residues in the higher level, with the Z-score -4.58, while the MolProbity score of our model was 51%. Hence, we energy minimized our model using the KoBaMIN server. After minimization, our model showed better results, where the RAMPAGE server showed 94.5% of residues in the allowed region, 4.1% residues in the favored region, and 1.5% residues in the outlier region. ProSA server showed a Z-score of -5.76, and the energy of all the residues was lower than before meaning that our minimized model is near to the native structure conformation. Further, the MolProbity score for the minimized model was 100%. The comparative analysis of full-length XPA with respect to before and minimization is depicted in **Figure 6.4**, while the structure validation using RAMPAGE and ProSA server, and MolProbity has been depicted in **Figure. 6.5**, and **Table 6.1**.

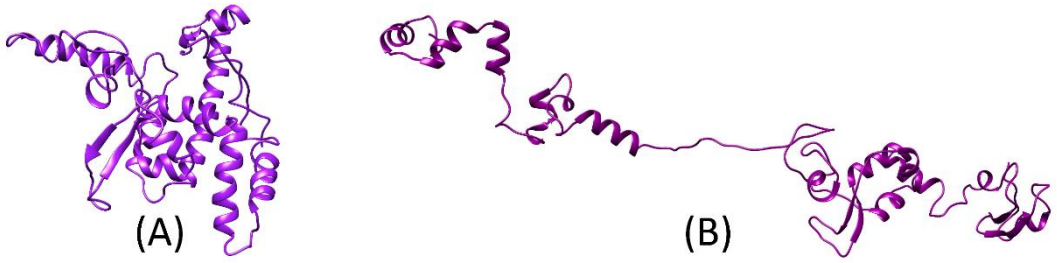

Figure 6.4 Full-length XPA (A) before minimization, (B) after minimization.

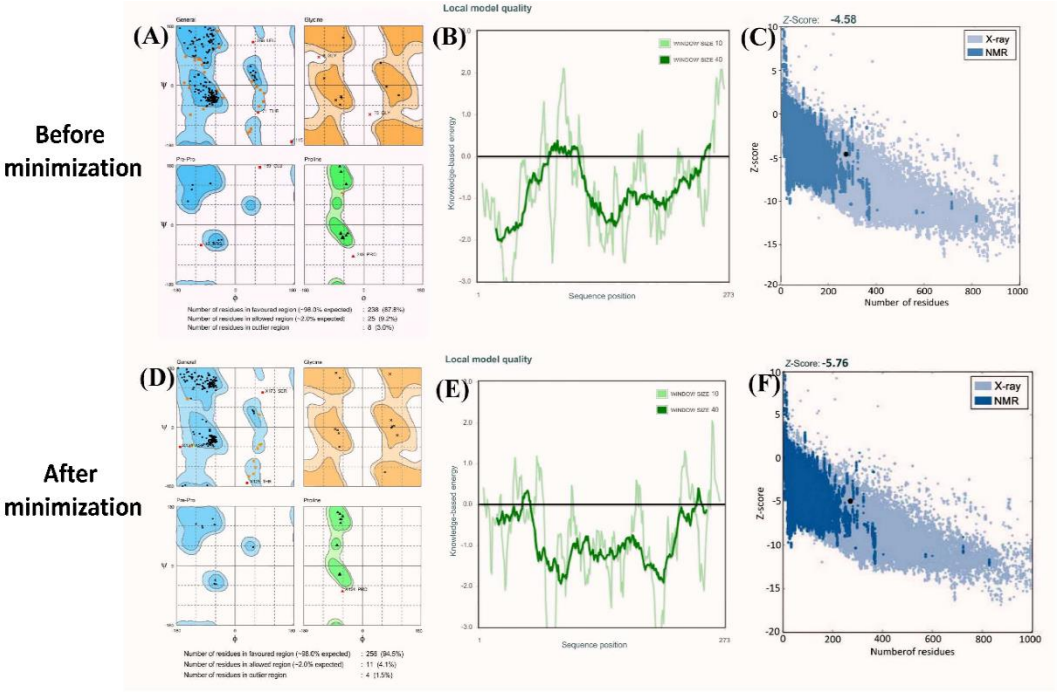

Figure 6.5. Structure validation of the modeled XPA: (A) Ramachandran plot, (B) energy plot, (C) Z-score of XPA before minimization; and (D) Ramachandran plot, and (B) energy plot, (F) Z-score of XPA after minimization.

Table 6.1. MolProbity summary for the modeled structure of full-length XPA before and after minimization.

MolProbity properties		XPA before minimization	XPA after minimization	Goal
All-Atom Contacts	Clash score, all atoms:	20.74 (30th percentile*, N=1784, all resolutions)	0 (100th percentile*, N=1784, all	100th percentile

				resolutions)		
Protein Geometry	Poor rotamers	1	0.43 %	4	1.72 %	<0.3%
	Favored rotamers	226	96.00%	226	96.00%	>98%
	Ramachandran outliers	10	3.69 %	4	1.48 %	<0.05%
	Ramachandran favored	235	86.72%	253	93.36%	>98%
	MolProbity score^	2.44 (51st percentile)		1.11 (100th percentile)		100th percentile* (N=27675, 0Å - 99Å)
	Cβ deviations >0.25Å	35	13.46%	8	3.08 %	0
	Bad bonds	0 / 2233	0.00 %	0 /232 3	0.00 %	0%
	Bad angles	93/2996	3.10 %	30 / 2996	1.00 %	<0.1%
Peptide Omegas	Cis Prolines	0 / 12	0.00 %	0 / 12	0.00 %	≤1 per chain, or ≤5%
	Twisted Peptides	1 / 272	0.37 %	1 / 272	0.37 %	0

6.4.2 Understanding the docked nature of PIC and determination of PPIs between PIC

XPA'S interaction with the members of PIC in the monomer formed intricate PPCs with each other as summarized in **Table 6.2**, and **Figure 6.6**. The detailed PPIs of this PPC have been shown in **Figure 6.7**, where we observed that p8 and p52 subunits of TFIIH were attached to the C-terminal region of XPA (aa240-273) as mentioned in earlier studies [72, 122, 123]. P52 interacted with aa157-273 of XPA. P8 subunit of TFIIH formed PPI with aa198-270 of XPA. XPA. The same findings were observed by Li et al. [356], wherein TFIIH had an affinity toward the acidic residues belonging to the N-terminal region of XPA, meaning that XPA aids in the helicase activity of TFIIH during the early steps of NER. XPA has also been noted to have helped TFIIH to prevent further damage to DNA [27]. XPA's association with TFIIH further helps disassociation of the CAK domain of TFIIH, which initiates the incision process of the lesioned DNA by ERCC1 protein [488]. RPA70AB, being the protein to protect the undamaged strand of DNA has been known to interact well with the DBD of XPA, where the sites of interaction of XPA and RPA7070AB overlap with ssDNA, but the specific location of the PPI was not deciphered. Here, we observed that RPA70AB was indeed bound to the DBD region of XPA as mentioned before [116, 368, 489], and found out that RPA70AB interacted with the aa102-178 of XPA. RPA32C on the other hand was bound to the N-terminal region XPA (aa2-46). Mer et al. [134] reported that RPA32C has a much stronger interaction with aa29-46 of XPA. ERCC1 being the last member to join the PIC was bound to the DBD region of XPA (aa80-175), especially on the opposite side of RPA70AB. Li et al. [135], and Li et al. [145] reported that aa92-119 of ERCC1 and aa96-114 had a stronger affinity for forming PPC.

Table 6.2. Interface statistics for XPA monomer with PIC.

Chains	No. of interface residues	Interface area (Å^2)	No. of salt bridges	No. of hydrogen bonds	No. of non-bonded contacts
P52	30	1474	5	8	158
P8	29	1367			
P52	32	1487	1	4	361
XPA	27	1857			
P8	3	220	3	4	24
XPA	3	223			
RPA32C	22	1002	1	11	667
XPA	22	1021			
ERCC1	39	1443	2	17	2679
XPA	29	1810			
RPA70AB	29	1365	5	13	233
X	23	1402			
ERCC1	26	896	1	13	1726
RPA70AB	25	1268			

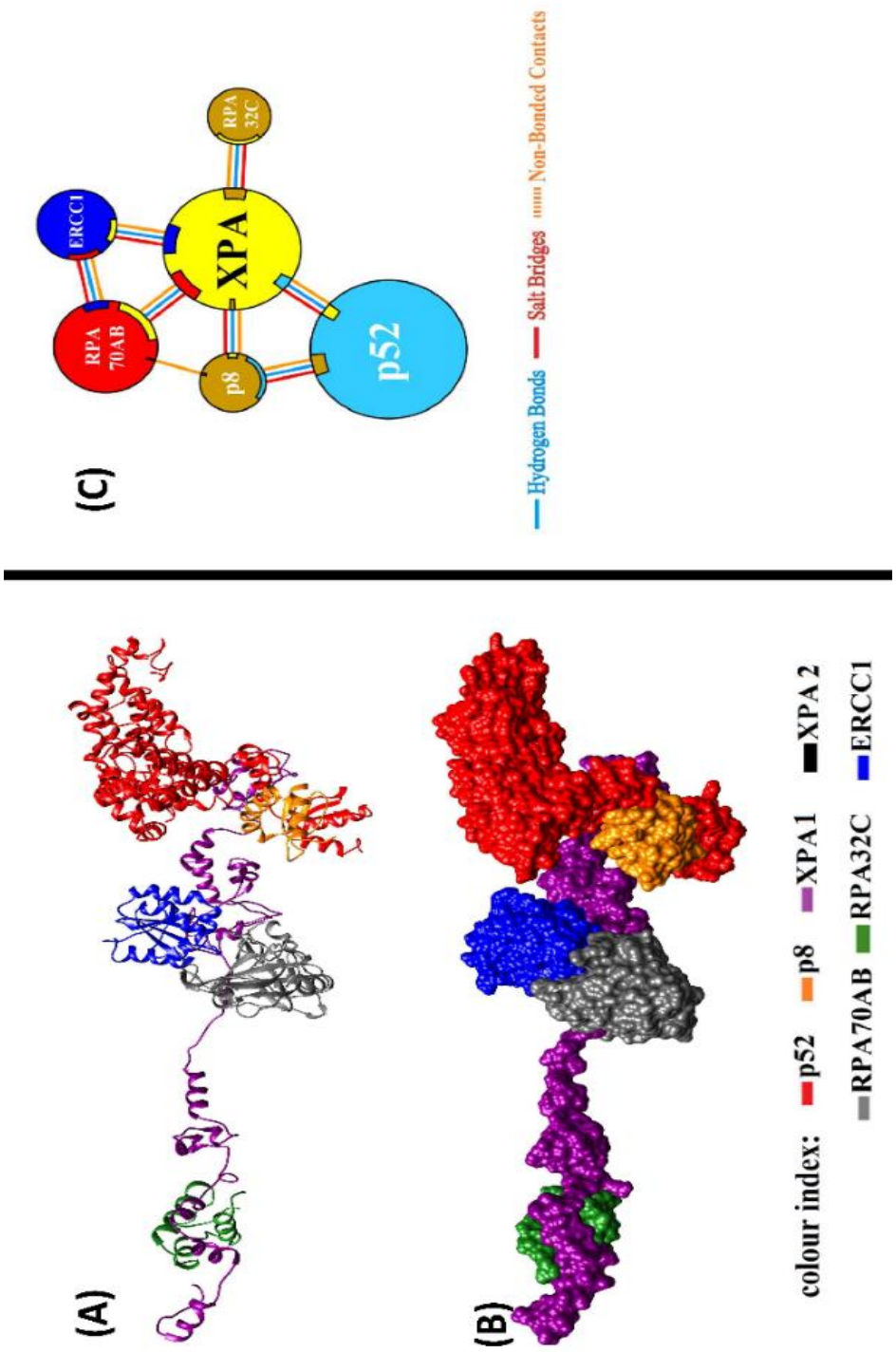

Figure 6.6 (A) cartoon representation and (B) surface diagram of the docked structure. of XPA monomer with the members of PIC. (C) Schematic diagram of PPI between the members of PIC as obtained from PDBsum server.

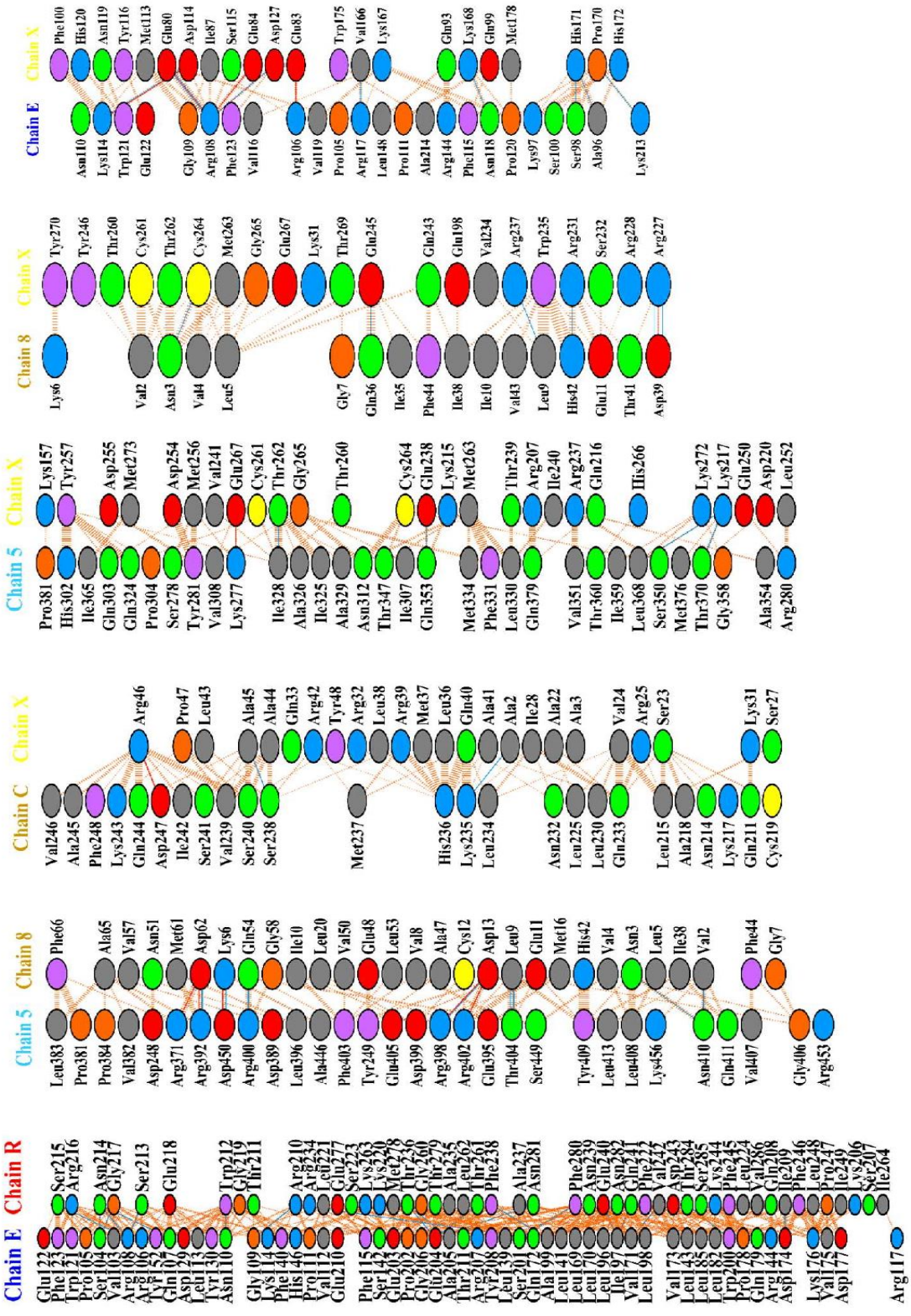

Figure 6.7. Schematic diagram of PPI between the XPA monomer and the members of PIC as obtained from PDBsum server.

Similar characteristics were observed within the members of PIC and XPA homodimer, but with little or negligible changes. XPA in homodimer form had more PPI with the TFIIH domain as compared to the other PIC proteins. An in-depth analysis of the aforementioned studies has been shown in **Table 6.3** and **Figures 6.8** and **6.9**.

Table 6.3. Interface statistics for XPA homodimer with PIC.

Chains	No. of interface residues	Interface area (Å^2)	No. of salt bridges	No. of hydrogen bonds	No. of non-bonded contacts
P52	26	1224	1	8	145
P8	24	1182			
P52	13	660	5	4	87
XPA1	9	678			
P8	3	220	3	4	24
XPA1	3	223			
RPA32C	22	1002	1	11	124
XPA1	22	1022			
ERCC1	39	848	1	17	2679
XPA1	29	1268			
RPA70AB	29	1302	5	13	233
XPA1	23	1401			
ERCC1	57	1443	2	17	2679
RPA70AB	52	1810			
p52	21	1118	3	8	145
XPA2	19	1128			

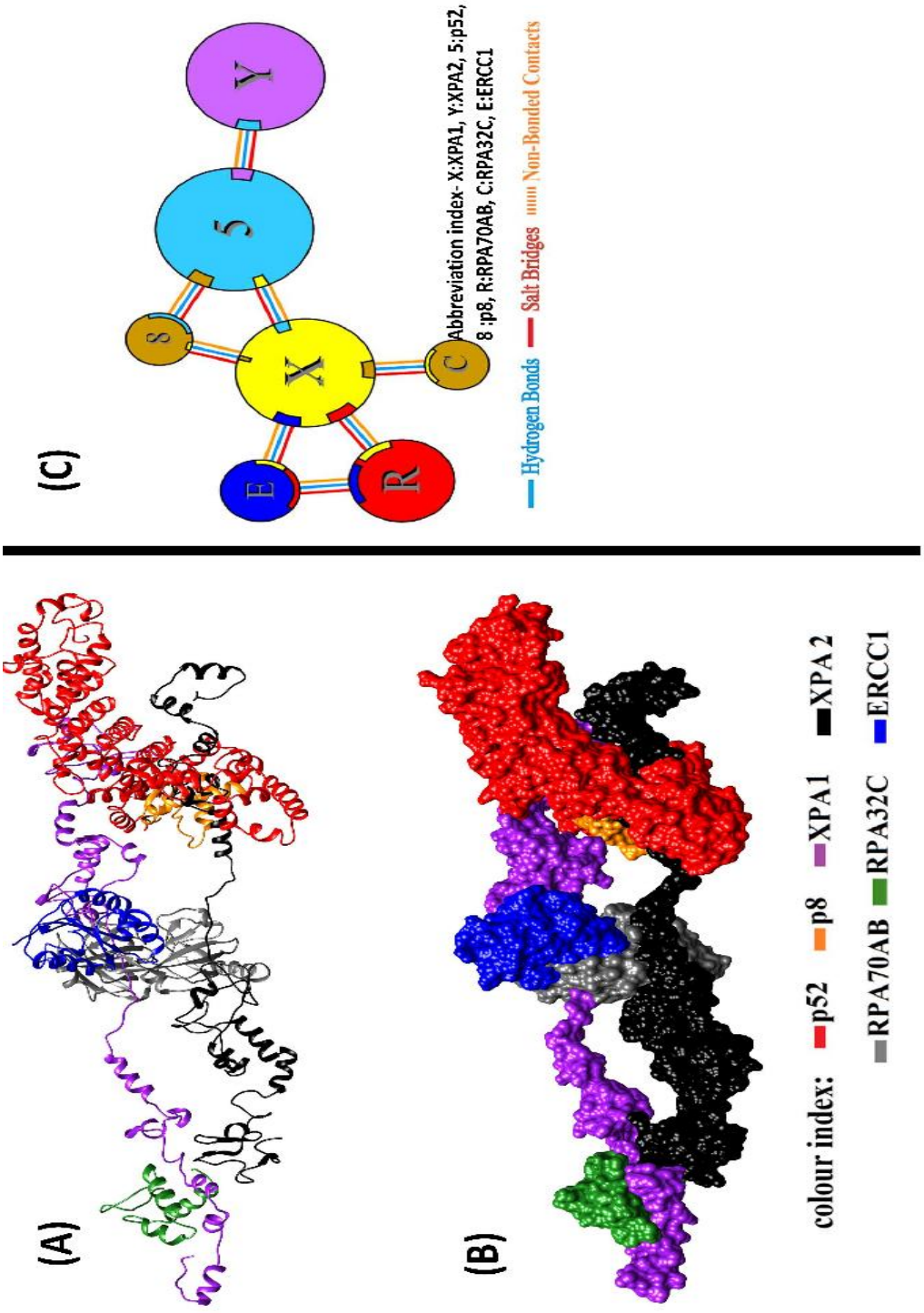

Figure 6.8 (A) cartoon representation and (B) surface diagram of the docked structure. of XPA homodimer with the members of PIC. (C) Schematic diagram of PPI between the members of PIC as obtained from PDBsum server.

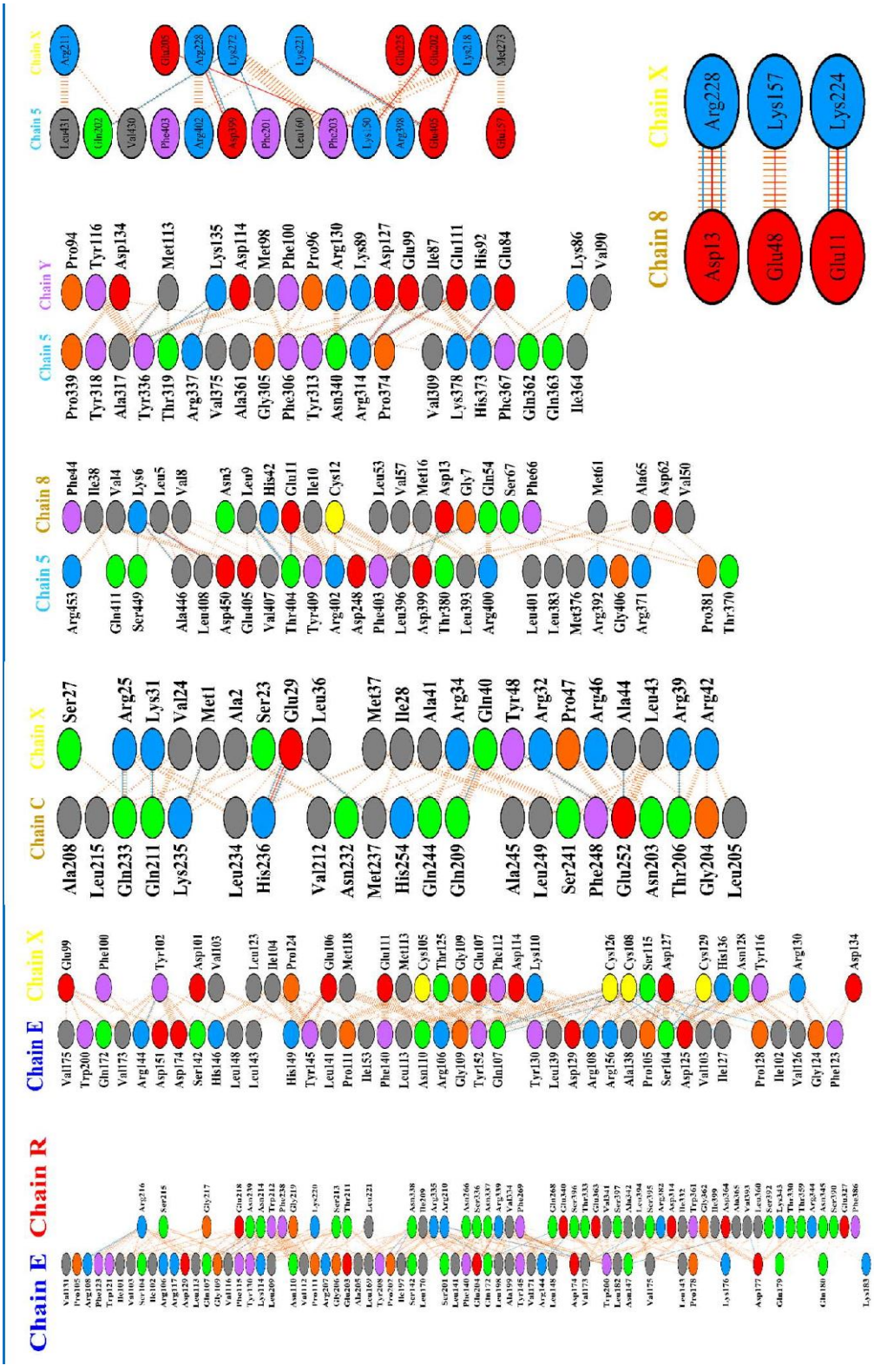

Figure 6.9. Schematic diagram of PPI between the XPA homodimer and the members of PIC as obtained from PDBsum server.

6.5 Conclusion

The modeled 3D structure of full-length XPA was validated using various online tools. The results after the minimization using the KoBaMIN server yielded better results compared to the structure before minimization. The values of the RAMPAGE server, ProSA, and MolProbity servers showed that our structure was a good-fit model to represent the full-length structure of XPA in both monomer and homodimer forms. Using the ClusPro server, we created XPA homodimer. The docking process done following the sequential steps gave us a clear picture of how these members of PIC were placed during the early process of the NER pathway. Next, we characterized the PPIs between the PPCs of PIC. We observed that RPA70AB and ERCC1 were bound to either side of the DBD region of XPA. P52 and p8 subunits of the TFIIH complex were bound to the C-terminal region of XPA, which is mostly composed of acidic regions of XPA. RPA32C on the other hand was bound to the N-terminal region of XPA. We also observed PPIs between ERCC1 and RPA70AB, and p8 of TFIIH complex and RPA70AB stabilized main by large numbers of hydrogen bonds, salt bridges, and hydrophobic interactions. XPA homodimer exhibited similar characteristics to the members of PIC as shown by the XPA monomer. XPA2 in particular was bound to the p5 protein of TFIIH, indicating that it aids TFIIH in helicase activity to initiate the NER process.