CHAPTER 3

Investigation of the probable homodimer model of the Xeroderma pigmentosum complementation group A (XPA) protein to represent the DNA binding core

CHAPTER 3

Investigation of the probable homodimer model of the Xeroderma pigmentosum complementation group A (XPA) protein to represent the DNA binding core

3.1 Abstract

The *Xeroderma pigmentosum* complementation group A (XPA) protein functions as a primary damage verifier and as a scaffold protein in nucleotide excision repair (NER) in all higher organisms. New evidence of XPA's existence as a dimer and the redefinition of its DNA binding domain (DBD) raises new questions regarding the stability and functional position of XPA in NER. Here, we have investigated XPA's dimeric status concerning its previously-defined DBD (XPA_{98-219}) as well as its redefined DBD (XPA_{98-239}). We studied the stability of XPA_{98-210} and XPA_{98-239} homodimer systems using all-atom molecular dynamics simulation, and we have also characterized the protein-protein interactions (PPI) of these two homodimeric forms of XPA. After conducting the root mean square deviation (RMSD) analyses, it was observed that the XPA_{98-239} homodimer has better stability than XPA_{98-210}. It was also found that XPA_{98-239} has a larger number of hydrogen bonds, salt bridges, and hydrophobic interactions than the XPA_{98-210} homodimer. We further found that Lys, Glu, Gln, Asn, and Arg residues shared the major contribution toward the intermolecular interactions in XPA homodimers. The binding free energy (BFE) analysis, which used the molecular mechanics Poisson–Boltzmann method (MM-PBSA) and the generalized Born and surface area continuum solvation model (GBSA) for both XPA homodimers, also substantiated the positive result in favor of the stability of the XPA_{98-239} homodimer.

3.2 Introduction

The *Xeroderma pigmentosum* complementation group A (XPA) protein is an obligate member of the nucleotide excision repair (NER) pathway that identifies and repairs very large DNA-distorting lesions. The most common types of lesions repaired by NER are cyclo-butane pyrimidine dimers (CPD), photoproducts, and cisplatin-DNA intra-strand crosslinks rendered by environmental mutagens, ultra-violet radiations, or antitumor agents [128, 353-355]. Based on the multiple steps of this repair process, NER involves

more than 30 different proteins which perform specific functions at each step, forming multi-protein complexes that are coordinated largely by XPA during the initial steps. The primary role of XPA lies entirely in verifying early damage, and in assembling other NER proteins to the DNA damage site [54, 164, 127].

The smooth functioning of NER and the proper excision of the DNA lesions are often dependent wholly on the sequential assembly and coordination of different proteins to the damage site. Therefore, any defect in these repair proteins, which hinders their ability to perform their tasks, can have repercussions in the form of various skin cancers and genetic disorders, such as *Xeroderma pigmentosum* (XP), [164, 165, 356-359]. Because XPA is common in both global-genome NER (GG-NER), and transcription-coupled NER (TC-NER) [48, 75, 360, 361], any deficits/mutations in XPA can result in a total NER failure, thereby causing classical XP disease phenotype that is characterized by extreme photo-reactivity, neurological disorders and often, skin cancers [54, 162-165, 362-367].

During the NER process, XPA interacts with various types of damaged DNA, as well as fellow repair proteins as part of its scaffolding nature. XPA has been known to exclusively interact with damaged strands by using its globular DBD, which earlier spanned between residues 98-219 [142, 368], but was later redefined between residues 98-239 [58, 126, 176]. XPA's protein-protein interactions (PPIs) with fellow NER proteins, on the other hand, have so far been extensively studied with (i) the 70 kD subunit of replication protein A (RPA70), which protects the undamaged DNA strand [125, 127, 142, 368, 369]; (ii) transcription factor II H (TFIIH) complex, a helicase that forms the NER bubble [48, 165, 370], (iii) another NER damage verifier-*Xeroderma pigmentosum* group C (XPC) protein [123, 355, 371]; and (iv) excision-repair cross-complementing group 1 (*Xeroderma pigmentosum* group F) (ERCC1/XPF), a structure-specific endonuclease [144-146].

XPA, as encoded by the XPA gene, has been described as a 273-residue long protein, consisting of disordered C- and N-terminal regions and a central globular DBD. The central globular DNA binding domain (DBD) consists of three helices, a short stretch of β-strand, and some loop regions [114-116]], as shown in **Figure 3.1**. Though the solution structure of the central globular core of XPA was determined nineteen years

ago, many of the structural insights on XPA's interactions with the DNA, and other NER proteins are still not understood well.

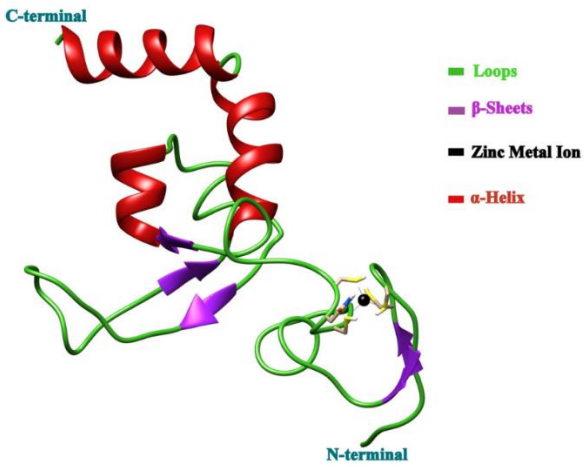


Figure 3.1. Cartoon representation of the central globular DNA-binding domain of XPA.

An earlier experimental study by Yang et al. [130] on XPA-RPA interactions using native gel filtration chromatography, perfluoro-octanoic acid-polyacrylamide gel electrophoresis (PFO-PAGE), and fluorescence spectroscopic analysis uncovered an important finding. They reported for the first time that XPA exists in the form of a dimer, rather than in a monomer, forming the XPA_2-RPA protein complex. Shortly after this, Liu et al. [131] also demonstrated that the binding between XPA and DNA exists as a 2.1 ratio and that the XPA dimer is the dominant form of XPA compared to its monomer form. The next evidence of XPA's dimer status was verified when Gilljam and her team observed XPA in the dimer form to have foci with proliferating cell nuclear antigen (PCNA), another scaffold protein of NER, which is seen only during the ligation phase of NER status. This finding puts forth another possibility for XPA. that it may be involved in other steps of NER apart from damage recognition [83]. Furthermore, Rad14 in yeast, a homolog of XPA, was also observed to be in dimer form, where the Rad dimer complex kinked DNA by 70º [129].

Since XPA's DBD aa98-219 is deficient in positively-charged residues needed to carry out strong DNA binding, a recent study focused on the DBD of XPA has suggested that the DBD should be extended beyond the previously supposed globular

core by at least 20 residues, which is from XPA_{98-219} to XPA_{98-239}, as this region is rich in basic amino acids. This report also suggested that the extended regions of residues aa219-239 are likely to form a helix, hypothesizing the possibility of a fourth helix [58]. Another study was done by Hilton et al. [126], which also emphasized the extension of XPA'S DBD up to position 239. They identified four lysine residues- K221, K222, K224, and K226 that are crucial for efficient DNA binding, and further observed that these lysine residues fall within the proposed DBD of XPA. More insights on the efficacy of the redefined DBD of XPA were provided by Sugitani and her team [176]. They observed higher perturbations upon XPA-DNA interactions, majorly in the globular core of XPA (aa130-210), and in the extended C-terminal residues of XPA (aa215-232). Higher levels of CSP shifts were exhibited mainly by Lys residues- K168, K179, K221, and K222, which lie in redefined DBD region, suggesting that proposed the DBD of XPA (aa98-239) is indeed efficient in conducting a stable XPA-DNA interaction. Earlier C-terminal truncation study was done by Cleaver & States [372], which reported that these truncations cause neurological disorders, wherein XPA_{1-227} truncations had mild symptoms while those with XPA_{1-220} exhibited severe neurological damage.

Thus, with this evidence of the dimer form of XPA and its DNA binding region extending beyond residue 219, there emerges the question as to how the dimer is stabilized, and how the extended region of XPA_{98-239} fits into the usual functioning of XPA in NER. Therefore, in the present study, we modeled the two homodimer structures of the DBD of XPA (i.e., XPA_{98-210} and XPA_{98-239}) by docking their monomeric forms using the SymmDock webserver [344, 345], and analyzed the intermolecular PPIs between them to determine the best model for representing the DBD of XPA. A molecular dynamics (MD) simulation- a complementary approach to experimental methods- was performed on both the monomer and dimer forms of XPA. $XPA_{98-210/239}$ to study their conformational dynamics and stability. Furthermore, using the molecular mechanics Poisson–Boltzmann (MM-PBSA) and generalized Born and surface area continuum solvation (GBSA) methods [322, 373-376,] we also calculated the binding free energy (BFE) of both XPA homodimers.

3.3 Systems and Methodology

3.3.1. Preparation of the initial structures for the DBD of XPA

The 3D structure of the DBD of XPA (PDB ID. 1XPA) determined by Ikegami et al. [114], which has a residue range between 98-210 instead of residues 98-219, was retrieved from Protein Data Bank [377, 378] and was used as the monomer structure for $XPA_{98-210.}$ For the monomer structure of XPA_{98-239}, the residues $XPA_{208-239}$ obtained from UniProt (ID. P23025) [379] were submitted to the RaptorX online server [380-382] for 3D structure prediction. This server works on the principle of protein threading and homology recognition benchmarks. The modeled structure (aa208-239) generated from the RaptorX server was selected based on its (i) P-value, the likelihood score of a predicted model is worse than the best of a set of randomly-generated models for this protein. The relative quality of the model can be assessed by this P-value. It is calculated by threading the target protein against various sets of reference templates and after which the score is provided based on their alignment outcomes. The smaller the P-value, the higher quality of the model. For mainly alpha proteins, a P-value less than 10-3 is a good indicator while for beta proteins, a P-value less than 10-4 is a good indicator [381] (ii) Score, that determines the quality of the predicted model based on alignment score between the target and the reference templates using the template-based modeling (TM) score, a neural network approach. The score ranges between 0 and the sequence length. The score close to the sequence length indicates a good result while the score near the zero value indicates the worst [380] (iii) uSeqId(SeqId), which refers to the total number of identical residues in the alignment, while SeqID is uSeqID normalized by the protein sequence length and multiplied by 100. The modeled structure is said to be of better quality if the uSeqID(SeqID) score is higher (iv) uGDT(GDT), the unnormalized GDT(Global Distance Test) score which measures the absolute quality of the predicted model by providing a value from 0-100, where 0 and 100 indicate worst and the best, respectively. uGDT is defined as $1\times N(1)+0.75\times N(2)+0.5\times N(4)+0.25\times N8$, where N(x) corresponds to the total number of residues with the modeling error (in Å) lesser than x, while GDT is calculated as uGDT divided by the protein length and multiplied by a 100. It is observed that for a protein with greater than 100 residues, uGDT greater than 50 is a

good indicator and for a protein with lesser than 100 residues, GDT greater than 50 is a good indicator. For the peptides/proteins with lesser than 100 residues, uGDT lesser than 50 and GDT greater than 50 is considered to be good for the predicted model [381].

Using the residue 208-210 as a common factor between the two structures (XPA_{98-210} and $XPA_{208-239}$), the resultant structure was joined with the NMR-determined structure of XPA (1XPA) by using the matchmaker tool, followed by the copy/combine function under the 'model panel' command of the UCSF Chimera software [315]. The residues were then renumbered, followed by the structure validation of this modeled XPA_{98-239} monomer using protein structure assessment (ProSA) server [383], RAMPAGE [384], and VERIFY 3D server [385-387].

The dimers of $XPA_{98-210/239}$ were generated upon the submission of their respective monomers to the SymmDock webserver [344, 345]. The complex structures were predicted based on the geometric C_n symmetry of the monomers with the rotational symmetry angle of alpha 360/n degrees, where 'n' represents the number of unit molecules the user wants to obtain from the submitted protein unit as an output. A total of ten complexes were generated for $XPA_{98-210/239}$ homodimers, and they were ranked upon their atomic contact energy, geometric score, and surface area. Both homodimers were selected by taking their maximum surface area, highest geometric score, and minimum free energy value into consideration.

Using the leap module of the AMBER 12 software package [388], we obtained topology and coordinate files for monomer and dimer forms of XPA as an input file for subsequent steps. All XPA systems were solvated using the TIP3P [297] water box with a buffer distance of at least 10 Å between systems and the periodic box wall; charge-balancing containers were then added to neutralize the systems (six Na^+ ions for the XPA_{98-210} monomer, three Cl^- ions for the XPA_{98-239} monomer, twelve Na^+ ions for the XPA_{98-210} homodimer, and six Cl^- ions for the XPA_{98-239} homodimer).

3.3.2. MD simulation of $XPA_{98-210/239}$ monomer and dimer systems

All MD runs were performed using the AMBER 12 software package, using the Amberff99 force field [283], while all the long-range-electrostatic interactions of the systems were addressed by the Particle Mesh Ewald algorithm (PME) [386, 389]. After

undergoing xleap preparations, the structures were forwarded for two-step minimization cycles with a cut-off of 8 Å for non-bonded interactions. The first minimization imposed the constraints over the solute for the first 500 steps using the steepest descent algorithm, and another 500 steps with the conjugate gradient method. The second minimization was devoid of any such restraints. All four systems were then placed in a heat bath by gradually increasing the temperature from 0-300 K under constant volume (NVT) conditions, after which the whole arrangement was equilibrated at constant pressure for 100 ps at NPT condition. $XPA_{98\text{-}210/239}$, in particular, was checked for their energy and temperature plots to ascertain sufficient equilibration (**Figures 3.2** and 3.**3**). To assess the equilibration for the big XPA homodimer, All XPA systems were then subjected to 60 ns of full NPT MD production run.

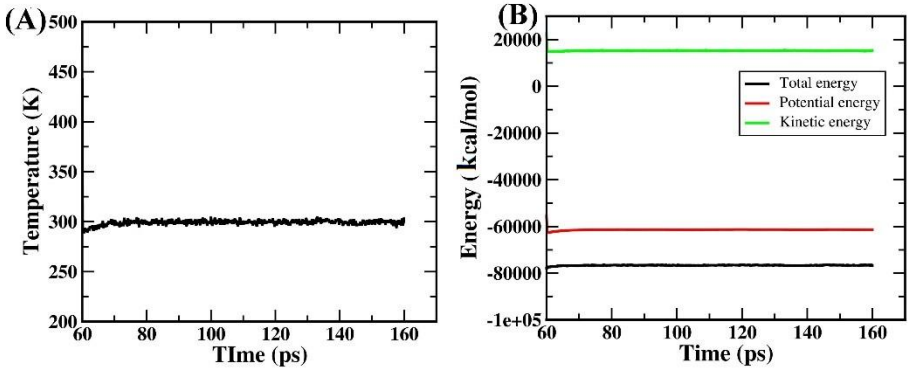


Figure 3.2. (A) Temperature plot and (B) Energy plot for $XPA_{98\text{-}210}$ homodimer.

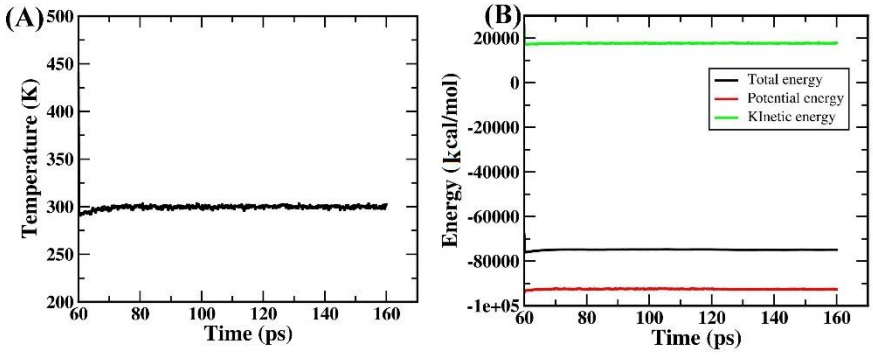


Figure 3.3. (A) Temperature plot and (B) Energy plot for $XPA_{98\text{-}239}$ homodimer.

The whole system was optimized under isobaric and isothermal conditions (T = 300 K; P = 1 atm), along with standard periodic boundary conditions with the time step of 2 fs using the SHAKE algorithm [293]. The Berendsen weak coupling algorithm [295] (0.5 ps of heat bath and 0.2 ps of pressure relaxation) was used for controlling the temperature throughout the simulation process.

3.3.3. MD analyses

All the clustered MD trajectories of $XPA_{98\text{-}210/239}$ systems were analyzed by the PTRAJ algorithm [317] and were visualized in the UCSF Chimera package alpha v.1.12 [315] and VMD v.1.9.3 [316]. The comparative RMSD, the radius of gyration (Rg), and solvent accessible surface area (SASA) analyses were performed to draw conclusive results. We also employed the hydrogen bond analysis for both homodimers based on the potential donors (HD) and acceptors (HA) of the hydrogen atom. The results were stipulated according to the occupancy, bond length, and the bond angle formed (HA–H–HD) between the HD and HA atoms.

3.3.4 Determination of the interface residues

For the determination of the PPI of XPA-homodimers, we extracted the lowest energy conformers for these homodimers from the trajectory files. The generated structure was uploaded to the PDBsum server [349, 389, 390] to visualize the intermolecular/inter-monomeric interface residues of the homodimers.

3.3.5 BFE calculations of the homodimers

To determine a better homodimer model to represent the XPAs DBD, the BFE of $XPA_{98\text{-}210/239}$ monomers to form $XPA_{98\text{-}210/239}$ homodimers was calculated using the molecular mechanics Poisson–Boltzmann method (MM-PBSA), and the molecular mechanics generalized Born and surface area continuum solvation (MM-GBSA) method ([322, 373-376]. We extracted 100 snapshots from the MD trajectories of the last 2 ns and conducted independent MM-GBSA/PBSA analyses for (i) $XPA_{98\text{-}210/239}$ monomer 1, (ii) $XPA_{98\text{-}210/239}$ monomer 2, and (iii) the homodimer $XPA_{98\text{-}210/239}$.

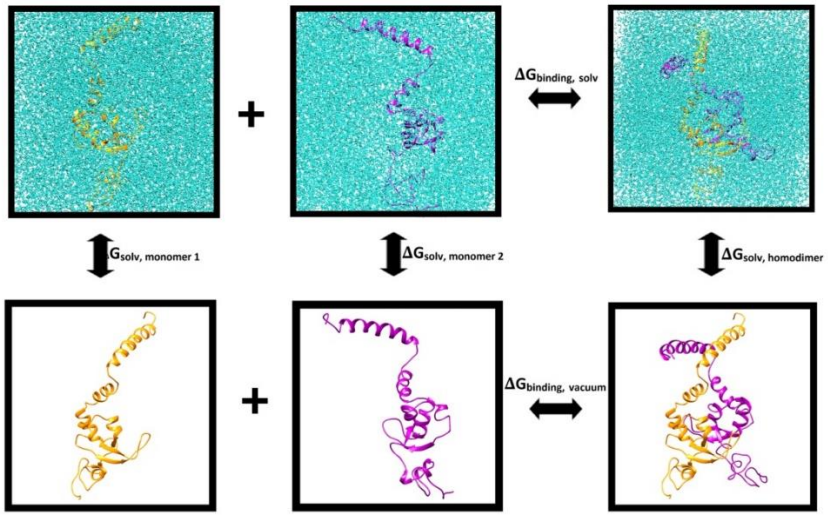


Figure 3.4. Schematic representation of the thermodynamic cycle used for the calculation of binding free energies (BFE).

BFE calculations between individual monomers were computed using the second law of thermodynamic integration, as shown in **Figure 3.4.** The binding free energy of $XPA_{monomer\ 1} + XPA_{monomer\ 2} = XPA_{homo-dimer}$ was calculated using the following thermodynamic equations, wherein all the participants involved were subjected to both gas (vacuum) and aqueous environments.

$$\Delta G_{binding} = \Delta G_{homo-dimer} - [\Delta G_{monomer\ 1} + \Delta G_{monomer\ 2}] \quad (1)$$

Here, $G_{binding}$ is the total binding free energy; $G_{homodimer}$, $G_{monomer\ 1}$ and $G_{monomer\ 2}$ are the relative free energies of the $XPA_{98\text{-}210/239}$ homodimer, $XPA_{98\text{-}210/239}$ monomer 1 and $XPA_{98\text{-}210/239}$ monomer 2, respectively. As per the second law of thermodynamics, free energy (G) for each component (homodimer complex, monomer 1, and monomer 2) is also enumerated by subtracting the total solute entropy (-TΔS) at absolute temperature (T) [373, 374] from the enthalpy (ΔH) of the system. Entropy is usually enumerated using a normal-mode analysis [391], using the AMBER 12 software package. However, we did not consider entropy in our calculations since our main objective here was to determine the binding free energy of our respective systems and not Gibbs free energy. The enthalpic state of each component was computed from molecular mechanical energy (E_{MM}) and solvation free energy (G_{solv}), respectively, as shown in Eqn. [3].

$$\Delta G_{binding} = \Delta H - T\Delta S \quad (2)$$

$$\Delta H = \Delta E_{MM} + \Delta G_{solv} \quad (3)$$

$$\Delta E_{MM} = \Delta E_{internal} + \Delta E_{vdW} + \Delta E_{ele} \quad (4)$$

$$\Delta G_{solv} = \Delta G_{PB/GB} + \Delta G_{surf} \quad (5)$$

$$\Delta G = \Delta E_{internal} + \Delta E_{vdW} + \Delta E_{ele} + \Delta G_{PB/GB} + \Delta G_{surf} - T\Delta S \quad (6)$$

Here, E_{MM} can be split into internal energy ($E_{internal}$), van der Waals forces (E_{vdW}), and electrostatic energy (E_{ele}) while G_{solv} is the total of polar solvation free energy of the PB/GB model ($G_{PB/GB}$), and the non-polar /surface solvation free energy (G_{surf}). Thus, from Eqn. [6], the free energy of any system can be explained by the changes in molecular mechanical energy, solvation free energy, and conformational entropies.

We used the modified GB model 1 (GB^{OBC1}) [392], and the PB solver of the AMBER 12 PBSA module developed by Lu and Luo [393] for the calculation of polar solvation energies of GB and PB models. We set a probe radius of 1.4 Å and used a grid space of 0.5 Å for our estimations. The internal dielectric constant of the solvent (water) and the external dielectric constant of the solute were set to 80 and 1. The non-polar solvation energy was addressed using a solvent-accessible surface area, calculated by Eqn. [7].

$$\Delta G_{surf} = \gamma \times SASA + \beta \quad (7)$$

The non-polar solvation energy for the GB method [323] was achieved by taking γ = 0.0072 kcal ($mol^{-1}Å^{-2}$) and β = 0 kcal mol^{-1}. For the PB method, γ was set to 0.00542 kcal (mol^{-1} $Å^{-2}$) and β was set to 0.92 kcal mol^{-1}.

3.4.Results and Discussion

3.4.1. Prediction of the extended DNA binding region of XPA and the selection of good fit dimer models for $XPA_{98-210/239}$ systems

Upon the submission of the protein FASTA sequence of $XPA_{208-239}$ to the RaptorX server, we obtained a total 4 solution structures, out of which the best-ranked structure had the P-value of 1.12e-01, an overall score of 28, uSeqId(SeqId)=6(18) and

uGDT(GDT)= 26(75) designated our model a good fit structure. Taking aa208-210 as the common factor between the two structures, the final structure of XPA_{98-239} (**Figure 3.6**) was obtained by joining the aa208-239 peptide (211-REKMKQKKFDKKVKELRRAVRSSVWKRET-239) to the NMR structure of XPA_{98-210} using the UCSF Chimera copy/combine tool. We next validated the fitness of this modeled structure using online computational tools. The ProSA-web server assigns a Z-score as well as provides an energy plot to judge the fitness of the modeled structures by analyzing the atomic coordinates of the candidate model. Our modeled structure for the XPA_{98-239} monomer as shown in **Figure 3.5B** was given the Z-score of -3.35, which was within the range of native protein conformation. According to the energy plot, our model was observed to have energy in the negative value (**Figure 3.5C**), indicating the robustness and quality of our modeled structure. RAMPAGE analysis for XPA_{98-239} showed 85.0% residues to be in the favored region, 15.0% residues in the allowed region, and 0% in the disallowed region (**Figure 3.5C**). The Verify-3D server validates the accuracy of the predicted structure concerning its residues (1D) by scoring based on their location and environment. As per Figure S1D, 78.6% of the residues had an average 3D-1D score >= 0.2, which confirms that the predicted model is near to the correct structure of the protein. According to **Figure 3.6**, the appended peptide takes an α-helical form (fourth helix), as hypothesized [58] and as demonstrated earlier [126]. This helix, highly rich with positively-charged lysine and arginine residues, will be able to enhance the DNA binding affinity of XPA in a much better way.

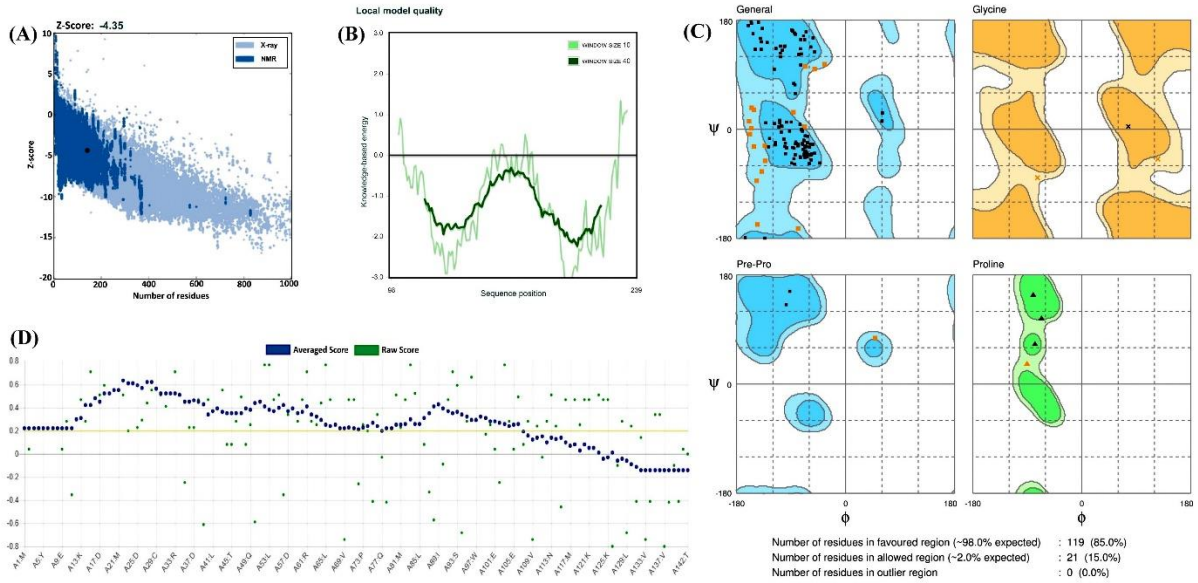


Figure 3.5. Structure validation of the modeled XPA$_{98-239}$ monomer. (A) z-Score plot and (B) energy plot obtained using ProsA server, (C) Ramachandran plot, and (D) Verify 3D plot for the modeled XPA$_{98-239}$ monomer.

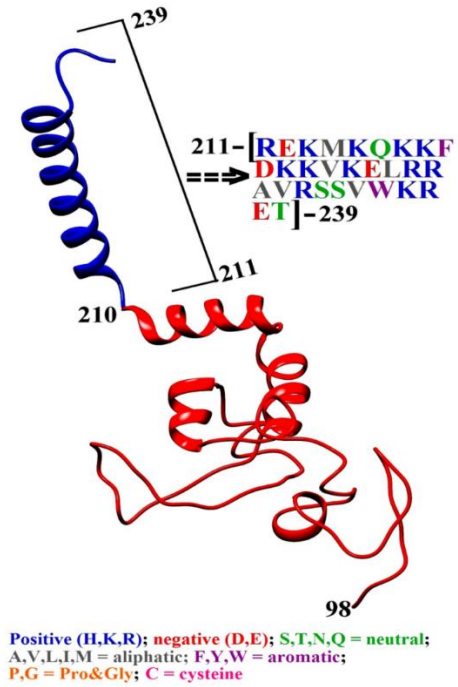


Figure 3.6. Redefined DNA binding domain of XPA (aa98-239) along with the fourth helix.

Next, we submitted the monomer forms of XPA$_{98-210/239}$ to the SymmDock web server for the prediction of probable structures for XPA dimers. The SymmDock web

server provided us with ten solution structures for each of the XPA dimers (**Figures 3.7A** and **3.7B**), ranked as per their maximum surface area, highest geometric score, and lowest atomic contact energy. We chose our model structures by factoring in the above-mentioned criteria among the rest. Our chosen structure to represent the XPA_{98-210} dimer area of 1576.40 $Å^2$, and an atomic contact energy (ACE) value of 410.82 kcal mol^{-1} (Solution Structure 1 from **Figure 3.7A)**. For the XPA_{98-239} homodimer, we chose the structure with a geometric score of 15,140, an interface surface area of 2086.40 $Å^2$, and an ACE value of 1230.78 kcal mol^{-1} (Solution Structure 1 from **Figure 3.7B**).

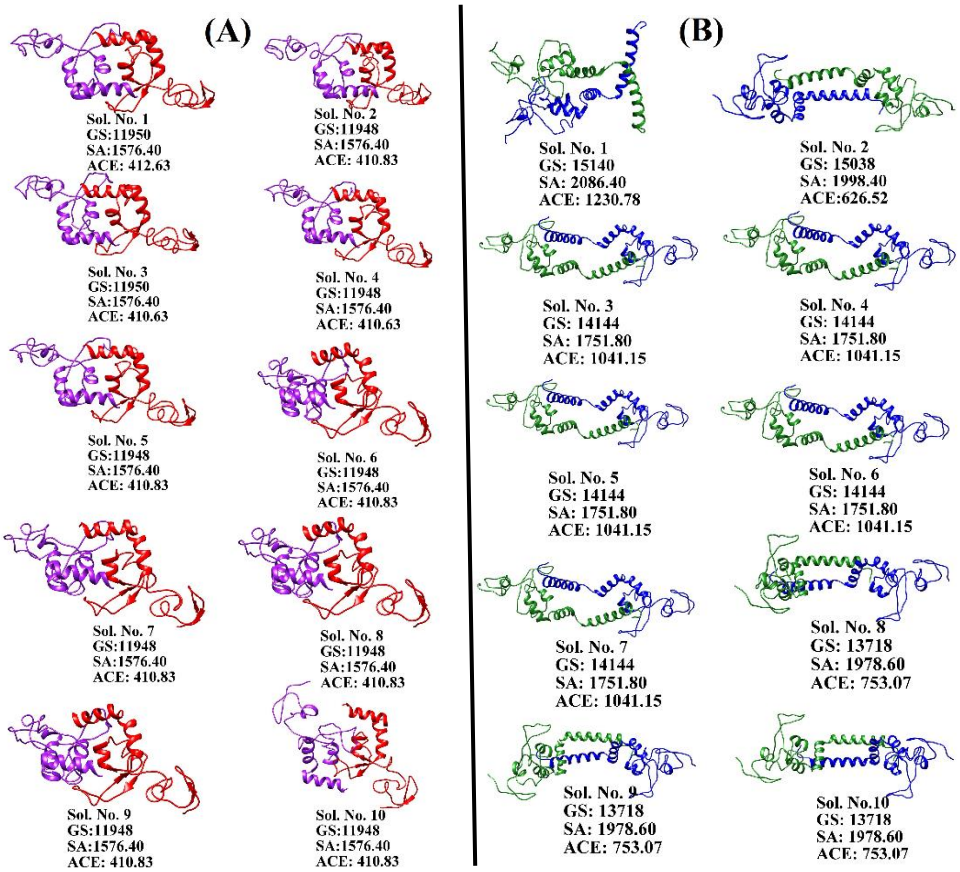


Figure 3.7. (A) Docked structures of XPA_{98-210} homodimer. (B) Docked structures of XPA_{98-239} homodimer.

3.4.2. Interpretations of MD trajectories for XPA$_{98-210/239}$ monomers and homodimers

After the equilibration step, we plotted energy and temperature to study the conformational stability of the XPA monomer and dimers, we used MD trajectories from our 60 ns simulation to generate RMSD, SASA, and Rg plots, as well as snapshots at different time intervals concerning their initial built structures.

The RMSD values of all Cα-atoms, referenced to their starting structures for all systems, were checked to see whether they had attained stability or not. The RMSD plots for all the systems are given in **Figure 3.8**, wherein the RMSD values of XPA$_{98-210/239}$ monomers and dimers are presented in **Figure 3.8A**. The XPA$_{98-210}$ monomer was observed to be well-settled throughout the simulation with a mean value of 5 Å, while the XPA$_{98-239}$ monomer showed initial stability of 6 Å, after which it deviated and then settled at 8 Å from 52 ns onwards. Their dimer counterparts fluctuated initially, stabilizing later on as the simulation proceeded. The XPA$_{98-210}$ dimer settled down from 37 ns onwards at 9 Å while the XPA$_{98-239}$ dimer settled down after 30 ns with an RMSD value of 13 Å. We also investigated the stability of the monomeric units present in the respective XPA homodimers. **Figure 3.8B** shows RMSD values for the individual monomeric units of the XPA$_{98-210}$ homodimer, where both monomers 1 and 2 were stable with the mean values of 4 Å and 12 Å, respectively. **Figure 3.8C** displays the individual RMSD values for each monomer of the XPA$_{98-239}$ homodimer; both monomers showed a coherent pattern of deviation in their movements before settling down from 45 ns onwards. Analyzing the RMSD results of all XPA systems revealed that the XPA$_{98-210}$ monomer is relatively more stable than the XPA$_{98-239}$ monomer. In the case of their dimeric states, the XPA$_{98-239}$ homodimer is much more stable than the XPA$_{98-210}$ homodimer.

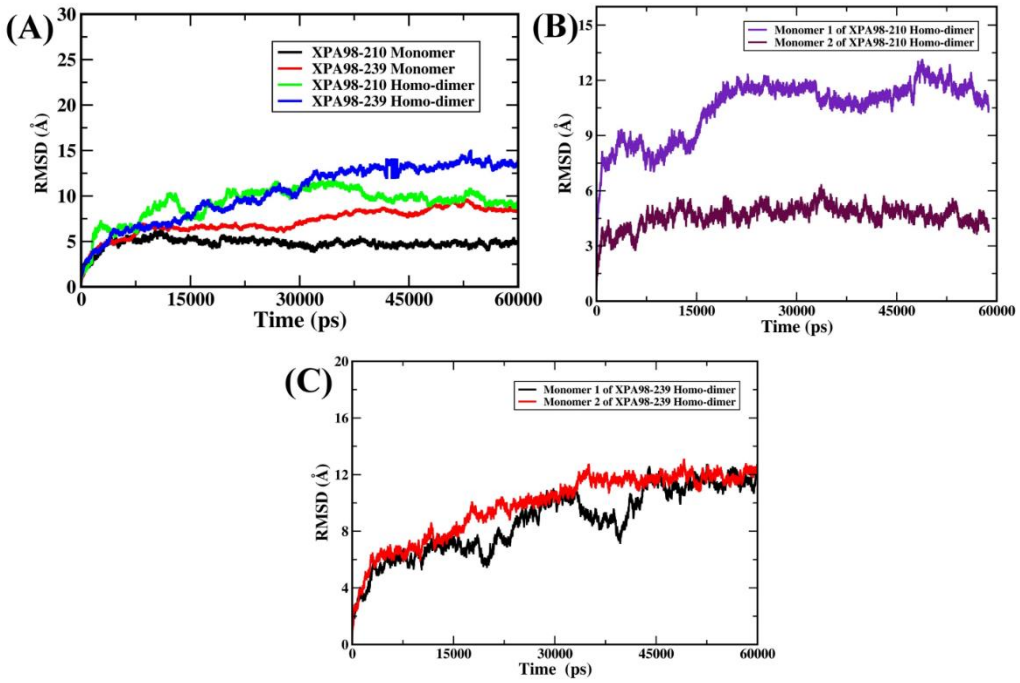


Figure 3.8. Root-mean-square deviation (RMSD) values of all Cα-atoms concerning their starting structure. (A) Comparative RMSD analyses for monomers and dimers of XPA$_{98-210/239}$. (B) RMSDs for the monomers of XPA$_{98-210}$ homodimer. (C) RMSDs for the monomers of XPA$_{98-239}$ homodimer.

Rg is usually calculated to estimate the overall dispersion of atoms of a particular biomolecule from their common center of gravity/axis. The Rg values of all XPA monomers and dimers are given in **Figure 3.9**. **Figure 3.9A** represents the Rg graphs for the XPA$_{98-210/239}$ monomer as well as for the dimers. Here, we observed the Rg values for the XPA$_{98-210}$ monomer and dimer to oscillate within the mean value of 16 Å and 27 Å, respectively. XPA$_{98-239}$ systems, in contrast, exhibited wide movement initially, as its monomer oscillated within 23 Å and its dimer within 25 Å, after which it settled at 27 Å. **Figures 3.9B** and **3.9C** show the Rg values for the individual monomer of XPA dimers. The XPA$_{98-210}$ dimer had a greater dispersion of its atoms in the case of monomers, while the atoms of monomer 2 were stable most of the time. Both the monomers of XPA$_{98-239}$ had their atoms widely dispersed, showing a wide range of motion, though monomer 1 fluctuated more than the other. The changes we see in the Rg values are the reflections endured by each structure as a result of their intermolecular interactions during the simulation.

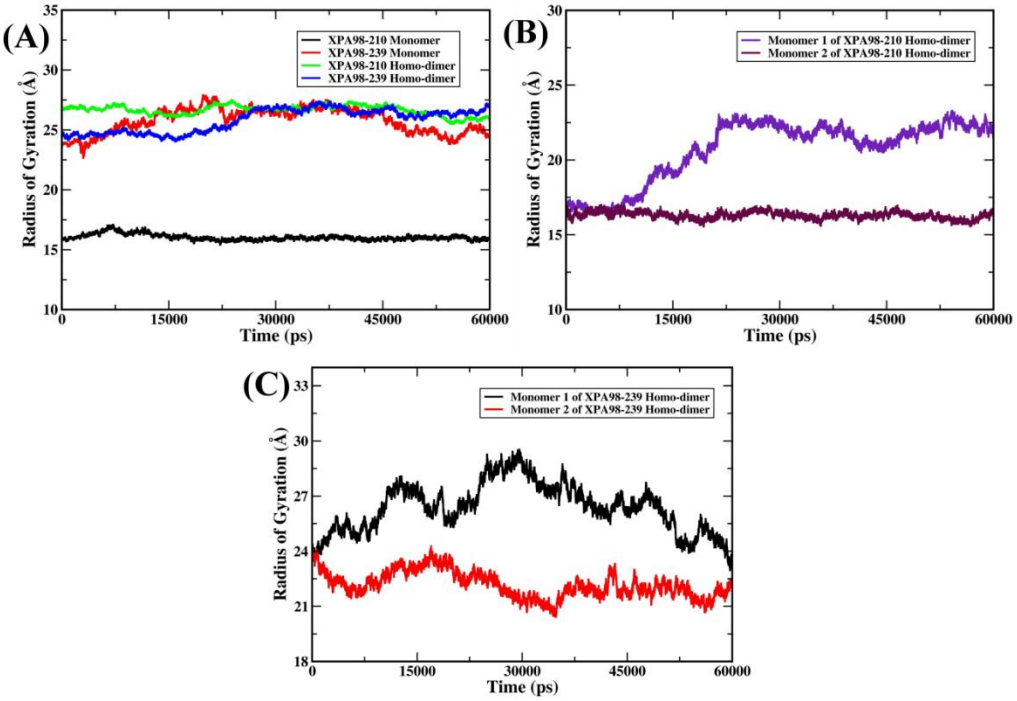


Figure 3.9. Radius of gyration (Rg) values of all Cα-atoms with respect to their starting structure. (A) Comparative Rg analyses for monomers and dimers of XPA$_{98\text{-}210/239}$. (B) Rg for the monomers of XPA$_{98\text{-}210}$ homodimer. (C) Rg for the monomers of XPA$_{98\text{-}239}$ homodimer.

The overall variations in the total SASA of all XPA systems are shown in **Figure 3.10A**. The SASA values (Zhang, Wang, Ling, Liu, & Liu, 2010) are analogous and directly reflective of all the unsuitable (hydrophobic) contacts between the water molecules and biomolecules. To map out the surface area accessible by the water solvent for our XPA systems, we used a probe with a radius of 1.4 Å. The SASAs of the XPA$_{98\text{-}210}$ monomer and dimer, and the XPA$_{98\text{-}239}$ monomer, remained constant at 8,000 Å^2, 11,000 Å^2, and 17,500 Å^2. SASA values for the XPA$_{98\text{-}239}$ homodimer were initially mapped at 20,000 Å^2 and later shifted to 22,500 Å^2. This increase can be accorded to the increase in the hydrophobic contacts within the dimer. Individual studies of the monomers of the XPA dimers (**Figures 3.10B** and **3.10C**) show the SASA values of monomers 1 and 2 of XPA$_{98\text{-}210}$ to be 8,000 and 9,000 ±500 Å^2; the SASA values of monomers 1 and 2 of XPA$_{98\text{-}239}$ alternated between 10,000-11,000 Å^2 and 10,000-10,500 Å^2, respectively.

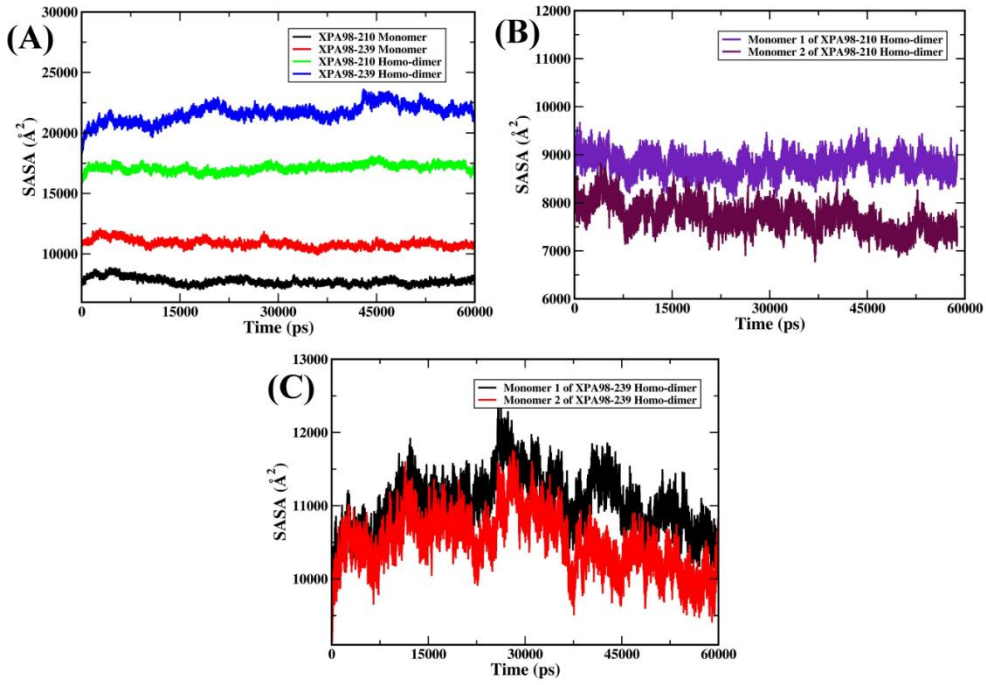
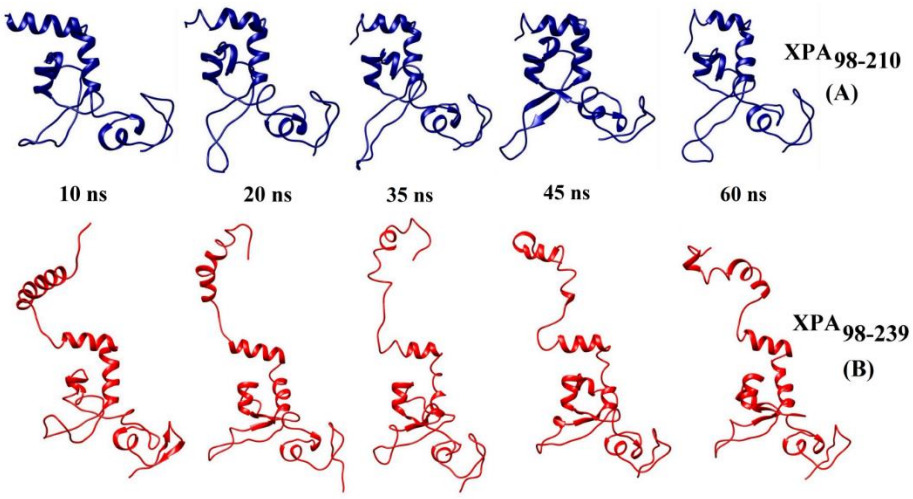


Figure 3.10. Solvent accessible surface area (SASA) values of all Cα-atoms with respect to their starting structure. (A) Comparative SASA analyses for monomers and dimers of XPA$_{98-210/239}$. (B) SASA analysis for the monomers of XPA$_{98-210}$ homodimer. (C) SASA analysis for the monomers of XPA$_{98-239}$ homodimer.

To further strengthen our study, we extracted the conformational snapshots of the dimers from the MD trajectories at 10, 20, 35, 45, and 60 ns. **Figures 3.11** and **3.12** show the snapshots of the conformational and structural changes undergone by XPA$_{98-210/239}$ monomers and homodimers. In **Figure 3.11A**, it can be seen that XPA$_{98-210}$ has not undergone much change while XPA$_{98-239}$ monomers (**Figure 3.11B**) presented changes mainly in their fourth helix region. From **Figure 3.11A**, we noticed that monomer 1 showed no significant changes in its structure until 35 ns, after which the helices slowly started converting into a coil. Monomer 2, on the other hand, exhibited changes more towards its N-terminal, while its remaining secondary structural elements remained intact. These changes are well reflected in their RMSD and Rg plots. From the snapshots, we noticed the changes in the distances between each monomer of XPA homodimers. The monomers of the XPA$_{98-239}$ homodimer were seen to have drifted apart slightly with respect to its initial structure from 20 ns onwards (**Figure 3.12B**). Similarly, the monomers of the XPA$_{98-210}$ homodimer seemed to have moved closer to

one another. We measured the initial distance between the monomers of XPA homodimers, which was seen to have settled between 20-21 ±1.5 Å, as shown in **Figure 3.13**.



.

Figure 3.11. Comparative snapshots for (A) XPA98-210 monomer and (B) XPA98-239 monomer as a function of time.

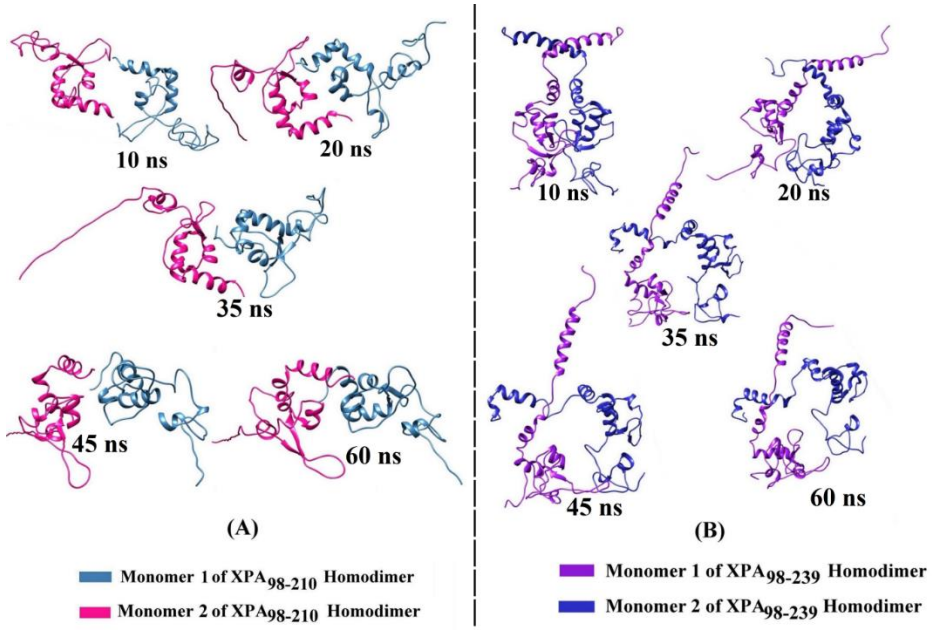


Figure 3.12. Comparative snapshots for (A) XPA98-210 homodimer and (B) XPA98-239 homodimer as a function of time.

We measured the initial distance between the monomers of XPA98-239 homo-dimers to be 11.94 Å, which later increased to 20-21 ±1.5 Å, as shown in **Figure 3.13A**.

Similarly, the initial distance between the monomeric units of the XPA_{98-210} homo-dimer was 28.78 Å, which later decreased to 20-22 ±1.5 Å, as shown in **Figure 3.13B**. The XPA homolog, Rad14 in yeast, had dimerized, binding the DNA from either side, wherein their inter-monomeric distance was calculated to be 21.5 Å [129]. Since our structure, too, became distanced or moved closer during the simulation, they may also use this gap between them to bind and hold the DNA as done by Rad14 molecules.

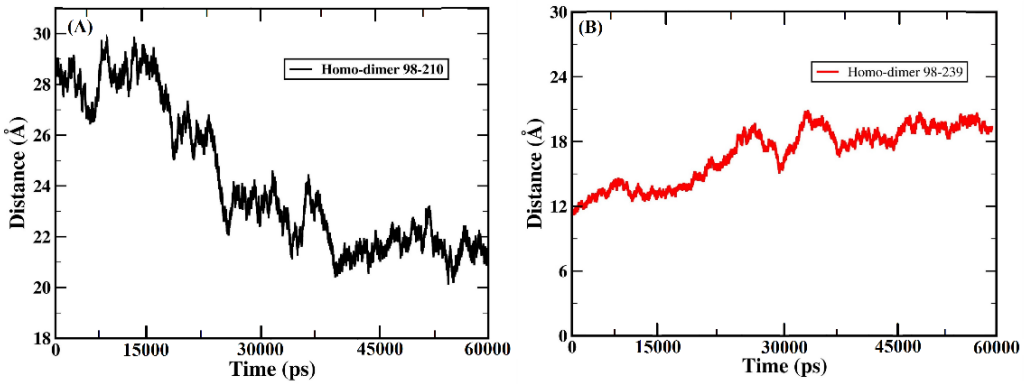


Figure 3.13. Distance between the monomeric units of (A) XPA_{98-210} homodimer and (B) XPA_{98-239} homodimer as a function of time.

Additionally, we also analyzed the hydrogen bond formation between two $XPA_{98-210/239}$ monomers, as these hydrogen bonds play a crucial role in conferring the stability of the protein complexes. The hydrogen bonds we obtained are shown in **Figure 3.14** and were within the ideal range as proposed earlier [444]. The XPA_{98-210} homo-dimer gained its maximum of eight hydrogen bonds (**Figure 3.14A**) with an average of 5-6 hydrogen bonds formed. The highest number of hydrogen bonds observed in the XPA_{98-239} homo-dimer (**Figure 3.14B**) was thirteen, with an average of 8-10 hydrogen bonds. The occupancies of the hydrogen bond formation of these dimers, along with their respective bond distances and bond angles, are provided in **Table 3.1**.

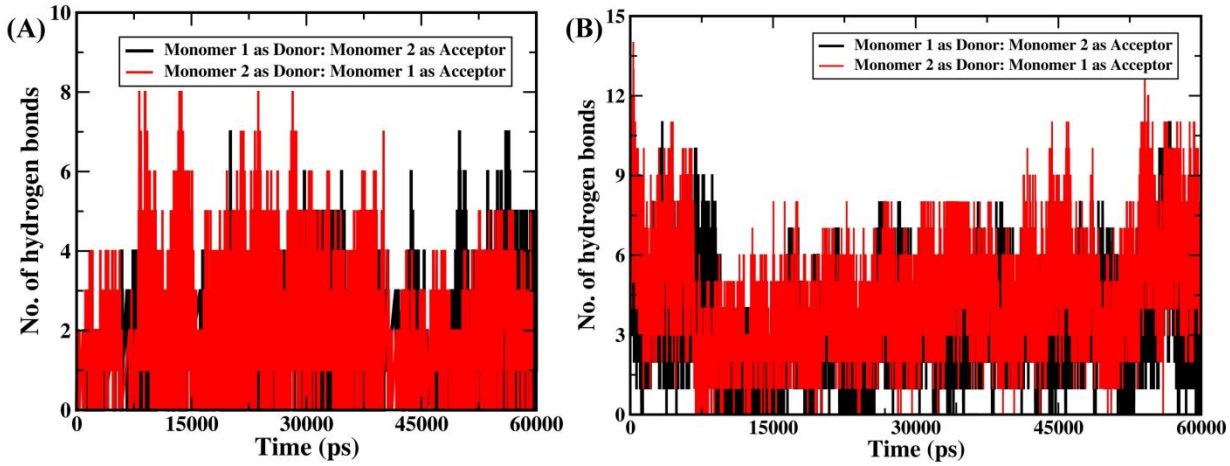


Figure 3.14. Inter-molecular hydrogen bond analyses of (A) XPA_{98-210} homodimer, and (B) XPA_{98-239} homodimer as a function of time.

Table 3.1. Intermolecular Hydrogen bond occupancy of $XPA_{98-210/239}$ homodimer.

Table 3.1A. Intermolecular hydrogen bond occupancy of XPA_{98-210} homodimer.

	Donor atom	Acceptor atom	Fractions	Bond distance (Å)	Bond angle
XPA_{98-210} monomer 1 as Donor. XPA_{98-210} monomer 2 as Acceptor	ASN_210@ND2	ASP_177@OD1	0.8212	2.8243	159.7947
	LYS_198@NZ	ASN_210@O	0.8145	2.7952	157.6399
	LYS_198@NZ	ASN_210@OXT	0.6459	2.7913	156.2205
	LYS_198@NZ	ASN_210@O	0.6343	2.7928	155.2498
	LYS_198@NZ	ASN_210@O	0.6341	2.8073	153.9492
	LYS_198@NZ	GLU_209@O	0.0326	2.8275	156.4583
	LYS_280@NZ	ASN_210@OD1	0.5314	2.8066	155.6701
	LYS_280@NZ	ASN_210@OXT	0.5308	2.7979	157.2485
	LYS_280@NZ	ASN_210@OD1	0.4228	2.8143	153.9806
	LYS_198@NZ	ASN_210@OXT	0.3272	2.7872	155.3173
	LYS_198@NZ	GLU_209@O	0.3269	2.8198	156.7902
	LYS_280@NZ	ASN_210@OXT	0.2261	2.7912	153.6551
	ASN_210@ND2	ASP_177@OD1	0.2216	2.8387	161.3973
	LYS_198@NZ	ASN_210@OD1	0.2189	2.8194	152.8281
	LYS_280@NZ	ASN_210@OD1	0.2176	2.814	152.9007
	LYS_198@NZ	GLU_209@O	0.2175	2.8198	156.0743
	ASN_210@ND2	HIE_172@O	0.2172	2.8707	159.367
	LYS_280@NZ	ASN_210@O	0.2104	2.7917	155.2914
XPA_{98-210} monomer	ASN_210@ND2	TYR_148@OH	0.739	2.8922	141.5984
	THR_142@OG1	GLU_209@OE1	0.645	2.7021	163.7317
	HIE_171@NE2	ASN_210@O	0.627	2.8304	160.2596
	LYS_141@NZ	ASN_210@O	0.58	2.7879	155.6725
	LYS_145@NZ	GLU_209@OE1	0.556	2.8005	153.9723
	THR_142@OG1	GLU_209@OE2	0.552	2.7029	146.8115
	SER_157@OG	GLU_209@OE2	0.547	2.6609	163.4769

2 as Donor. XPA_{98-210} monomer 1 as Acceptor	LYS_145@NZ	GLU_209@OE2	0.148	2.7993	155.0495
	LYS_145@NZ	GLU_209@OE2	0.1476	2.8026	153.8315
	LYS_141@NZ	ASN_210@OXT	0.44	2.7882	156.1439
	LYS_141@NZ	ASN_210@O	0.433	2.7792	153.1475
	LYS_151@NZ	GLU_147@OE1	0.24	2.788	153.5639
	LYS_145@NZ	GLU_209@OE2	0.24	2.803	153.8673
	LYS_145@NZ	GLU_209@OE1	0.1388	2.8066	153.0318
	HIE_172@NE2	ASN_210@O	0.134	2.83143	142.6013

Table 3.1B. Intermolecular Hydrogen bond occupancy of XPA_{98-239} homodimer.

	Donor atom	Acceptor atom	Fractions	Bond distance (Å)	Bond angle
XPA_{98-239} monomer 1 as Donor. XPA_{98-239} monomer 2 as Acceptor	LEU_138@N	ASP_133@OD1	0.913	2.8631	163.0901
	GLN_208@NE2	GLN_208@OE1	0.903	2.8567	159.877
	ARG_227@NH2	ASP_220@OD1	0.8181	2.8025	159.8594
	LYS_218@NZ	ASP_131@OD1	0.7811	2.8012	156.4888
	LYS_218@NZ	GLU_202@OE2	0.6424	2.7985	157.118
	LYS_222@NZ	GLU_205@OE2	0.6415	2.7873	156.6412
	LYS_222@NZ	GLU_202@OE1	0.6409	2.7955	158.3446
	LYS_222@NZ	GLU_205@OE2	0.5421	2.7919	157.5471
	LYS_218@NZ	GLU_202@OE2	0.5403	2.806	157.2897
	LYS_215@NZ	GLU_198@OE2	0.5389	2.7927	156.4132
	LYS_218@NZ	GLU_202@OE2	0.5381	2.8053	155.4826
	LYS_215@NZ	GLU_198@OE1	0.5365	2.7937	155.6763
	LYS_215@NZ	GLU_198@OE2	0.5351	2.792	156.3356
	LYS_215@NZ	GLU_198@OE2	0.3512	2.7805	157.1491
	ARG_227@NH2	ASP_220@OD1	0.3381	2.7845	155.581
	LYS_222@NZ	GLU_205@OE1	0.3293	2.796	157.4168
	LYS_215@NZ	GLU_198@OE1	0.3245	2.7973	156.5131
	LYS_218@NZ	GLU_202@OE1	0.3236	2.8036	153.1143
XPA_{98-239} monomer 2 as Donor. XPA_{98-239} monomer 2 as Acceptor	LYS_222@NZ	GLU_202@OE1	0.9038	2.7896	158.2899
	LYS_222@NZ	GLU_205@OE1	0.9015	2.7922	156.3216
	ARG_227@NH2	ASP_220@OD2	0.9013	2.8099	158.3146
	LYS_222@NZ	GLU_205@OE1	0.892	2.7857	157.7273
	LYS_222@NZ	GLU_205@OE1	0.8651	2.7854	155.8971
	ARG_227@NH1	ASP_220@OD1	0.8315	2.8019	158.104
	LYS_222@NZ	GLU_205@OE2	0.8232	2.7898	156.9231
	LYS_222@NZ	GLU_202@OE1	0.756	2.7912	158.6096
	LYS_222@NZ	GLU_205@OE2	0.7278	2.8098	155.1849
	ARG_211@NH2	GLU_201@OE2	0.6649	2.8046	158.4043
	LYS_222@NZ	GLU_205@OE2	0.6568	2.7887	157.168
	ARG_227@NH2	ASP_220@OD1	0.6011	2.8247	156.3326
	ARG_227@NH1	ASP_220@OD2	0.5772	2.8075	155.8179
	GLN_208@NE2	GLN_208@OE1	0.5413	2.8557	160.5261
	LYS_222@NZ	GLU_202@OE1	0.5342	2.7876	157.8085
	ARG_211@NH1	GLU_201@OE2	0.5019	2.8077	157.4414
	LYS_167@NZ	ASP_133@OD2	0.4280	2.7749	157.3798
	LYS_222@NZ	GLU_202@OE2	0.4257	2.7892	157.54

3.4.3 Determination of the interface interactions of the XPA homodimers

An interface area is usually defined as a region where two sets of proteins come into contact with each other. They are usually characterized by surface residues with large surface regions accessible to the solvent available [394]. The interface statistics for both XPA homodimers, which were obtained upon the submission of the lowest energy structures of XPA homodimers (**Figures 3.15** and **3.12C**) to the PDBsum server are shown in **Table 3.2**; **Figures 3.15B** and **3.15D** show the interface site between the two monomers for the $XPA_{98\text{-}210/239}$ homodimers. The summarized intermolecular interactions between each monomeric unit of the XPA dimers' residue levels are shown in **Figure 3.16**; the detailed contributions of each interface residue stabilizing the homodimer complexes are accordingly given in **Tables 3.3** and **3.4**

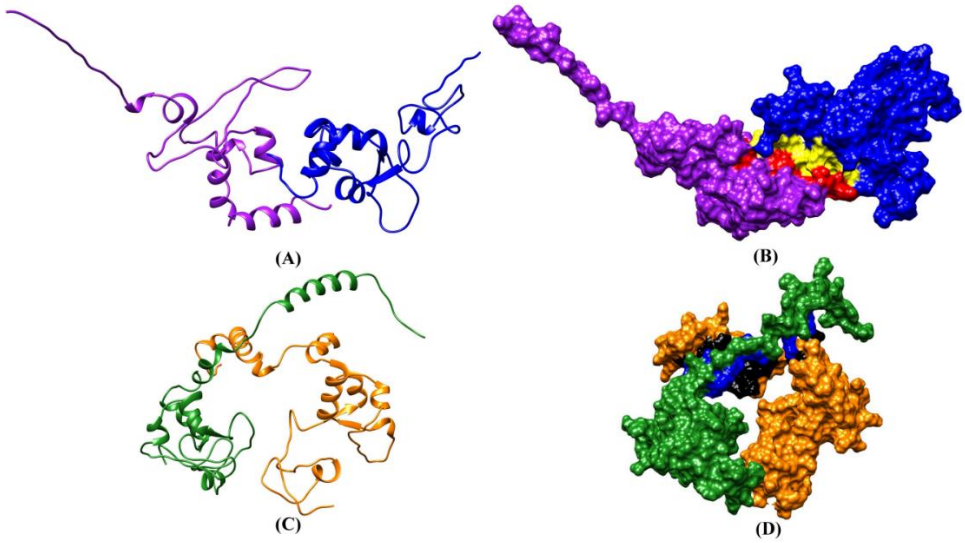


Figure 3.15. The lowest energy conformers of $XPA_{98\text{-}210/239}$ homodimers. (A) Cartoon structure $XPA_{98\text{-}210}$ homodimer (B) Surface diagram of $XPA_{98\text{-}210}$ homodimer showing the interface site of both monomers. (C) Cartoon structure $XPA_{98\text{-}239}$ homodimer (B) Surface diagram of $XPA_{98\text{-}239}$ homodimer showing the interface site of both monomers.

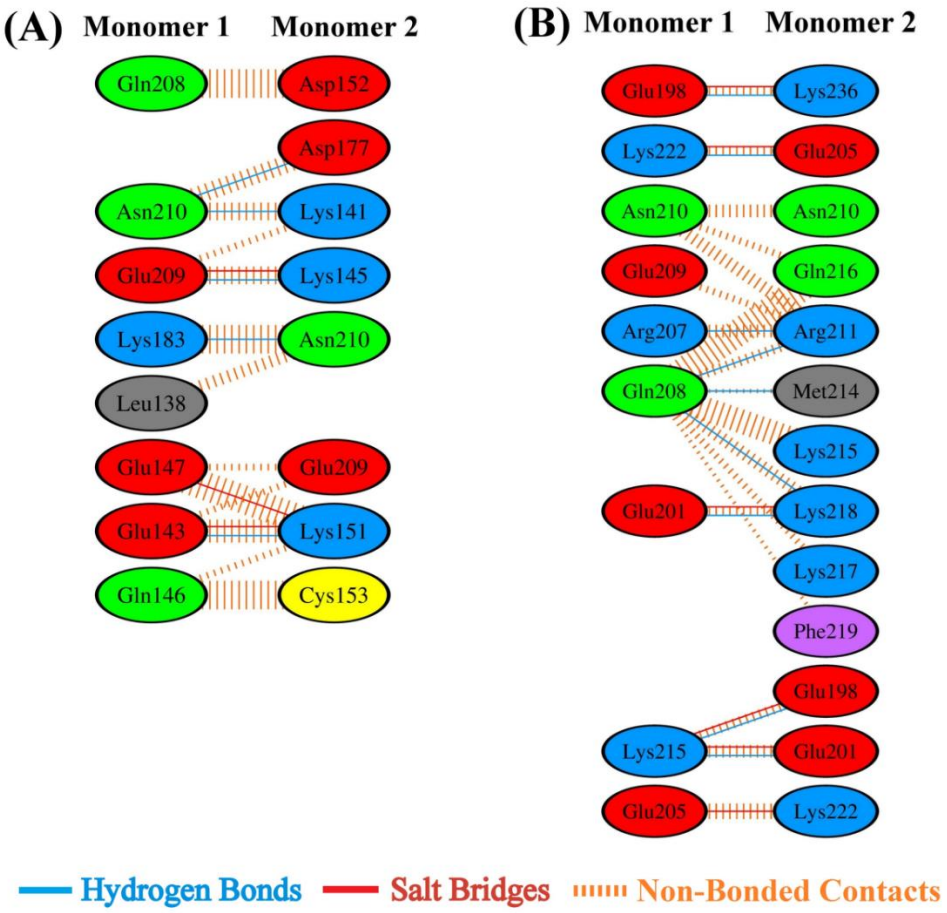


Figure 3.16. Intermolecular interactions between the monomers of. (A) XPA_{98-210} homodimer, and (B) XPA_{98-239} homodimer.

Table 3.2. Interface statistics for XPA homodimers.

Homodimer Systems	Monomers	No. of Interface Residues	Interface Area (Å^2)	No. of Salt Bridges	No. of Disulphide Bonds	No. of Hydrogen Bonds	No. of Non-Bonded Contacts
XPA_{98-210}	1	8	418	3	-	5	39
	2	8	418				
XPA_{98-239}	1	9	536	6	-	9	65
	2	13	461				

Table 3.3. Intermolecular interactions across the monomer-monomer interface of XPA_{98-210} homodimer.

Table 3.3A. Intermolecular hydrogen bond formation.

Sl.No.	Monomer 1	Monomer 2	Hydrogen

	Atom Name	Residues		Atom Name	Residues	Bond Distance (Å)
1.	OE1	GLU143	<-->	NZ	LYS151	2.76
2.	NZ	LYS183	<-->	O	ASN210	2.63
3.	OE2	GLU209	<-->	NZ	LYS145	2.69
3.	ND2	ASN210	<-->	OD2	ASP177	3.30
5.	OXT	ASN210	<-->	NZ	LYS141	2.90

Table 3.3B. Intermolecular salt bridge formation.

Sl. No.	Monomer 1			Monomer 2		Salt Bridge Distance (Å)
	Atom Name	Residues		Atom Name	Residues	
1.	OE1	GLU143	<-->	NZ	LYS151	2.76
2.	OE2	GLU147	<-->	NZ	LYS151	2.85
3.	OE2	GLU209	<-->	NZ	LYS145	2.69

Table 3.3C. Intermolecular non-bonded contacts.

Sl. No.	Monomer 1			Monomer 2		Non-bonded Contacts Distance (Å)
	Atom Name	Residues		Atom Name	Residues	
1.	O	LEU138	<-->	OD1	ASN210	3.25
2.	O	LEU138	<-->	ND2	ASN210	3.45
3.	CD	GLU143	<-->	NZ	LYS151	3.26
3.	OE1	GLU143	<-->	CE	LYS151	3.60
5.	OE1	GLU143	<-->	NZ	LYS151	2.76
6.	OE2	GLU143	<-->	NZ	LYS151	3.10
7.	OE2	GLU143	<-->	OE1	GLU209	3.72
8.	CB	GLN146	<-->	CB	CYS153	3.73
9.	CD	GLN146	<-->	CB	CYS153	3.55
10.	OE1	GLN146	<-->	CE	LYS151	3.51
11.	OE1	GLN146	<-->	C	CYS153	3.84
12.	OE1	GLN146	<-->	O	CYS153	3.87
13.	OE1	GLN146	<-->	CB	CYS153	3.32
13.	CD	GLU147	<-->	NZ	LYS151	3.13
15.	OE1	GLU147	<-->	CD	LYS151	3.57
16.	OE1	GLU147	<-->	CE	LYS151	3.48
17.	OE1	GLU147	<-->	NZ	LYS151	3.00
18.	OE2	GLU147	<-->	NZ	LYS151	2.85
19.	OE2	GLU147	<-->	OE1	GLU209	3.83
20.	C	LYS183	<-->	O	ASN210	3.71
21.	CE	LYS183	<-->	OD1	ASN210	3.23
22.	NZ	LYS183	<-->	C	ASN210	3.80
23.	NZ	LYS183	<-->	O	ASN210	2.63
23.	NZ	LYS183	<-->	OD1	ASN210	3.41
25.	CA	GLN208	<-->	OD2	ASP152	3.60
26.	O	GLN208	<-->	CB	ASP152	3.30
27.	O	GLN208	<-->	CG	ASP152	3.85
28.	CB	GLN208	<-->	OD2	ASP152	3.83
29.	CG	GLN208	<-->	OD2	ASP152	3.86
30.	O	GLU209	<-->	CE	LYS141	3.24
31.	CD	GLU209	<-->	NZ	LYS145	3.83
32.	OE2	GLU209	<-->	CE	LYS145	3.65
33.	OE2	GLU209	<-->	NZ	LYS145	2.69

33.	C	ASN210	<-->	NZ	LYS141	3.87
35.	CG	ASN210	<-->	OD2	ASP177	3.87
36.	OD1	ASN210	<-->	OD2	ASP177	3.75
37.	ND2	ASN210	<-->	OD2	ASP177	3.30
38.	OXT	ASN210	<-->	CE	LYS141	3.77
39.	OXT	ASN210	<-->	NZ	LYS141	2.90

Table 3.4. Intermolecular interactions across the monomer-monomer interface of XPA_{98-239} homodimer

Table 3.4A. Intermolecular hydrogen bond formation.

Sl. No.	Chain A			Chain B		Hydrogen Bond Distance (Å)
	Atom Name	Residues		Atom Name	Residues	
1.	OE1	GLU198	<-->	NZ	LYS236	2.79
2.	OE1	GLU201	<-->	NZ	LYS218	2.96
3.	O	ARG207	<-->	NH2	ARG211	2.87
3.	O	GLN208	<-->	NH2	ARG211	3.01
5.	OE1	GLN208	<-->	N	LYS218	2.95
6.	NE2	GLN208	<-->	O	MET214	3.15
7.	NZ	LYS215	<-->	OE1	GLU198	2.73
8.	NZ	LYS215	<-->	OE2	GLU201	2.77
9.	NZ	LYS222	<-->	OE1	GLU205	2.9

Table 3.4B. Intermolecular salt bridge formation.

Sl. No.	Chain A			Chain B		Salt Bridge Distance (Å)
	Atom Name	Residues		Atom Name	Residues	
1.	OE2	GLU198	<-->	NZ	LYS236	2.79
2.	OE2	GLU201	<-->	NZ	LYS218	2.96
3.	OE2	GLU205	<-->	NZ	LYS222	2.69
3.	NZ	LYS215	<-->	OE1	GLU198	2.73
5.	NZ	LYS215	<-->	OE1	GLU201	2.77
6.	NZ	LYS222	<-->	OE1	GLU205	2.96

Table 3.4C. Intermolecular non-bonded contacts.

Sl. No.	Chain A			Chain B		Non-bonded Contacts Distance (Å)
	Atom Name	Residues		Atom Name	Residues	
1.	CD	GLU198	<-->	NZ	LYS236	3.79
2.	OE1	GLU198	<-->	CE	LYS236	3.59
3.	OE1	GLU198	<-->	NZ	LYS236	2.79
3.	OE1	GLU201	<-->	CD	LYS218	3.36
5.	OE1	GLU201	<-->	CE	LYS218	3.53
6.	OE1	GLU201	<-->	NZ	LYS218	2.96
7.	CD	GLU205	<-->	NZ	LYS222	3.30
8.	OE1	GLU205	<-->	NZ	LYS222	3.13
9.	OE2	GLU205	<-->	CD	LYS222	3.30
10.	OE2	GLU205	<-->	CE	LYS222	3.58
11.	OE2	GLU205	<-->	NZ	LYS222	2.69
12.	C	ARG207	<-->	NH2	ARG211	3.89

13.	O	ARG207	<-->	CZ	ARG211	3.60
13.	O	ARG207	<-->	NH1	ARG211	3.59
15.	O	ARG207	<-->	NH2	ARG211	2.87
16.	CA	GLN208	<-->	NH2	ARG211	3.88
17.	CA	GLN208	<-->	O	LYS215	3.51
18.	C	GLN208	<-->	NH2	ARG211	3.42
19.	C	GLN208	<-->	O	LYS215	3.84
20.	O	GLN208	<-->	NH2	ARG211	3.01
21.	O	GLN208	<-->	O	LYS215	3.31
22.	O	GLN208	<-->	CA	GLN216	3.56
23.	O	GLN208	<-->	CD	GLN216	3.51
23.	O	GLN208	<-->	OE1	GLN216	2.93
25.	CD	GLN208	<-->	C	LYS215	3.73
26.	CD	GLN208	<-->	O	LYS215	3.79
27.	CD	GLN208	<-->	N	GLN216	3.77
28.	OE1	GLN208	<-->	N	GLN216	3.67
29.	OE1	GLN208	<-->	CA	GLN216	3.71
30.	OE1	GLN208	<-->	C	GLN216	3.13
31.	OE1	GLN208	<-->	O	GLN216	3.04
32.	OE1	GLN208	<-->	N	LYS217	3.52
33.	OE1	GLN208	<-->	C	LYS217	3.88
33.	OE1	GLN208	<-->	N	LYS218	2.95
35.	OE1	GLN208	<-->	CA	LYS218	3.48
36.	OE1	GLN208	<-->	CB	LYS218	3.27
37.	OE1	GLN208	<-->	CG	LYS218	3.59
38.	OE1	GLN208	<-->	N	PHE219	3.37
39.	NE2	GLN208	<-->	O	MET214	3.15
40.	NE2	GLN208	<-->	CA	LYS215	3.41
41.	NE2	GLN208	<-->	C	LYS215	3.16
42.	NE2	GLN208	<-->	O	LYS215	3.38
43.	NE2	GLN208	<-->	N	GLN216	3.52
43.	C	GLU209	<-->	NH2	ARG211	3.48
45.	O	GLU209	<-->	NH2	ARG211	3.43
46.	N	ASN210	<-->	NH2	ARG211	3.64
47.	N	ASN210	<-->	OE1	GLN216	3.89
48.	CA	ASN210	<-->	NH2	ARG211	3.78
49.	CB	ASN210	<-->	ND2	ASN210	3.82
50.	CB	ASN210	<-->	NE	ARG211	3.84
51.	CB	ASN210	<-->	OE1	GLN216	3.53
52.	CG	ASN210	<-->	ND2	ASN210	3.83
53.	OD1	ASN210	<-->	CD	ARG211	3.78
53.	ND2	ASN210	<-->	OD2	ASN210	3.83
55.	ND2	ASN210	<-->	ND2	ASN210	3.47
56.	CD	LYS215	<-->	OE2	GLU201	3.71
57.	CE	LYS215	<--5	OE1	GLU198	3.73
58.	CE	LYS215	<-->	OE2	GLU201	3.59
59.	NZ	LYS215	<-->	CD	GLU198	3.78
60.	NZ	LYS215	<-->	OE1	GLU198	2.73
61.	NZ	LYS215	<-->	CD	GLU201	3.65
62.	NZ	LYS215	<-->	OE2	GLU201	2.77
63.	CD	LYS222	<-->	OE1	GLU205	3.81
63.	CE	LYS222	<-->	OE1	GLU205	3.84
65.	NZ	LYS222	<-->	OE1	GLU205	2.96

According to **Figure 3.16A**, out of thirty-nine non-bonded interactions, along with three salt bridges and five hydrogen bonds, Asn210 of the XPA_{98-210} monomer 1 formed two hydrogen bonds with the residues of monomer 2, while K151 of monomer 2 contributed two salt bridge formations from its end. From **Figure 3.16B**, we observed that out of nine hydrogen bonds and six salt bridges, residue Q208 of XPA_{98-239} monomer 1 contributed three hydrogen bonds; K215 of monomer 1 lent two hydrogen bonds and two salt bridges. R211 of monomer 2 also formed two hydrogen bonds with the interface residues of monomer 1 in the XPA_{98-239} homodimer complex. The homodimer complex of the XPA_{98-239} system conferred further stability by the achievement of sixty-five non-bonded interactions. In both cases, XPA homodimers had massive surface (interface) areas had more hydrophobic contacts, which is a crucial property of a stable homodimer as reported in earlier findings [395, 396].

As another important observation, we noticed that $XPA_{98-210/239}$ homodimers have more of their C-terminal residues taking an active part in the interface interactions, with the residues Lys, Glu, Gln, and Asn contributing in a major way. The residues located towards C-terminals play a vital role in the scaffolding nature of XPA, as these contain binding sites for other NER proteins, RPA70AB [125, 127, 369], *Xeroderma pigmentosum* complementation group E proteins, damaged DNA-binding proteins 1 and 2 (XPE/DDB1/2) [151], and TFIIH sites [48, 165, 397]. Furthermore, the importance of Lys residues in DNA interactions has been well documented in previous work [126]. K217E mutation, in particular, has been found to cause severe XPA with neurological damage, growth stoppage, and tumorigenesis [398]. In the current study, we, too, have observed Lys residues involved in the PPI, where K141, K145, K151, and K183 were seen in the interface communication of the XPA_{98-210} homodimer, while K215, K217, K218, K222, and K236 were involved in the inter-monomeric interactions of the XPA_{98-239} homodimer. One peculiar feature of these $XPA_{98-210/239}$ homodimers is that they dimerized without any disulfide linkage despite the presence of several cysteine residues. Instead, XPA dimerized with the help of other means as described above. The first group who reported the existence of XPA as homodimers, where XPA bound RPA in XPA_2-RPA mode also noticed that dimerization had taken place with no disulfide bonds, despite maintaining the absence of reducing agents in their experiments [130].

3.4.4. BFE estimation of XPA$_{98-210/239}$ homodimers

The BFE calculations of XPA$_{98-210/239}$ monomers to form the XPA$_{98-210/239}$ homodimers were done using MM-PBSA/GBSA methods. As mentioned earlier, we did not consider entropy in our study; therefore, these values represent only the binding free energy rather than absolute or total binding energy. The binding free energies determined for both dimer complexes, along with the energy terms, are given in **Table 3.5A**, and **3.5B.**

Table 3.5A displays the individual contribution of both monomers of XPA$_{98-210}$ towards forming the XPA$_{98-210}$ homodimer, while the contributions of the XPA$_{98-239}$ monomers 1 and 2 for the formation of the XPA$_{98-239}$ homodimer are given in **Table 3.5B**. Analyzing both **Table 3.5A** and **3.5B**, we observed that all the derived components for the BFE analysis contributed greatly to the binding of each of the XPA monomers to form XPA dimers. However, in comparison, the BFE for the XPA$_{98-239}$ homodimer (PBSA= -76.52 kcal mol^{-1}; GBSA= -63.24 kcal mol^{-1}) is much greater than that for the XPA$_{98-210}$ homodimer (PBSA= -57.38 kcal mol^{-1}; GBSA= -46.26 kcal mol^{-1}). This may be due to the energies accorded by vdW forces (-81.62 kcal mol^{-1}) and the electrostatic contributions (-313.45 kcal mol^{-1}) to the XPA$_{98-239}$ homodimer. Additionally, we observed a huge difference between the binding energy values of both XPA homodimers, whether it was with the PBSA or GBSA method. For the PBSA method, the difference was -76.52 - (- 57.38) = -19.14 kcal mol^{-1}; for the GBTOT analysis, the difference was -63.24 - (-46.26) = -16.98 kcal mol^{-1}. Thus, considering the different energy contributions of various components as well the BFE for both XPA homodimers, we found that the XPA$_{98-239}$ homodimer can indeed be an ideal model to represent the DBD of XPA.

Table 3.5. BFE (kcal mol^{-1}) analysis for XPA homodimers.

Table 3.5A. BFE (kcal mol^{-1}) analysis for XPA$_{98-210}$ homodimer.

Method	Components	Homodimer complex (kcal mol^{-1})	Monomer 1 (kcal mol^{-1})	Monomer 2 (kcal mol^{-1})	ΔΔGcalc (kcal mol^{-1})
MM	ΔE_{ele}	-4853.58	-2348.87	-2286.56	-259.158
	ΔE_{vdW}	-848.94	-399.27	-390.04	-59.63
	ΔE_{int}	5103.40	2556.85	2547.55	-0.00
	ΔE_{MM}	**-599.12**	**-191.29**	**-129.05**	**-318.78**
PBSA	ΔG_{surf}	86.64	43.54	43.35	-1.25
	ΔG_{cal}	-6086.85	-2860.12	-3563.14	262.65
	ΔG_{solv}	-6000.21	-2816.58	-3519.79	261.40
	ΔG_{ele}	-10941.43	-5208.99	-5250.64	3.50
	PB$_{TOT}$	**-6599.33**	**-3007.87**	**-3648.84**	**-57.38**
GBSA	ΔG_{surf}	86.64	43.54	43.35	-1.25
	ΔG_{cal}	-5883.98	-2977.26	-3233.01	273.77
	ΔG_{solv}	-5797.35	-2933.72	-3188.66	272.52
	ΔG_{ele}	-10738.56	-5126.13	-5122.01	13.62
	GB$_{TOT}$	**--6396.46**	**-3125.01**	**-3317.71**	**-46.26**

Table 3.5B. BFE (kcal mol^{-1}) analysis for XPA$_{98-239}$ homodimer

Method	Components	Homodimer complex (kcal mol^{-1})	Monomer 1 (kcal mol^{-1})	Monomer 2 (kcal mol^{-1})	ΔΔGcalc (kcal mol^{-1})
MM	ΔE_{ele}	-6888.32	-3713.54	-3487.23	-313.45
	ΔE_{vdW}	-1065.54	-548.71	-598.45	-81.62
	ΔE_{int}	-7215.87	3713.54	3502.33	-0.00
	ΔE_{MM}	**-737.99**	**-549.71**	**-583.35**	**-395.07**
PBSA	ΔG_{surf}	303.02	103.35	192.54	-7.13
	ΔG_{cal}	-8728.73	-4736.21	-4315.20	322.68
	ΔG_{solv}	-8423.71	-4631.86	-4122.66	315.55
	ΔG_{ele}	-15617.05	-8450.75	-7802.43	9.23
	PB$_{TOT}$	**-9162.70**	**-5181.57**	**-4706.01**	**-76.52**
GBSA	ΔG_{surf}	303.02	103.35	192.54	-7.13
	ΔG_{cal}	-9860.78	-4865.61	04656.21	338.96
	ΔG_{solv}	-9556.76	-4761.26	-4463.67	331.83

	ΔG_{ele}	-16749.10	-8580.15	-8143.44	25.51
	GB_{TOT}	**-10293.75**	**-5310.97**	**-5047.02**	**-63.24**

ΔE_{ele} = electrostatic energy as calculated by the MM force field; ΔE_{vdW} = van der Waals contribution from MM; ΔE_{int} = internal energy arising from the bond, angle, and dihedral terms in the MM force field; ΔE_{MM} = total gas-phase energy (sum of ELE, VDW, and INT); ΔG_{surf} = non-polar contribution to the solvation free energy calculated by an empirical model; ΔG_{cal} = the electrostatic contribution to the polar solvation free energy calculated by PB or GB, respectively; ΔG_{sol} = sum of non-polar and polar contributions to solvation; ΔG_{ele} = sum of the electrostatic solvation free energy and MM electrostatic energy; PB_{TOT}/GB_{TOT} = final estimated binding free energy in kcal mol^{-1}calculated from the terms above.

Over the years, we have learned that the globular DBD of XPA (aa98-219) is the one that interacts with the DNA and that it readily recognizes kinked DNA distortions better than other types of DNA anomalies [142, 368]. However, this region (aa98-219,) lacking a considerable amount of positively-charged residues, was not sufficient enough to make a strong interaction with the negatively-charged DNA. The redefined DBD, on the other hand, is much more effective for DNA binding, with a greater number of positively-charged residues, especially in the predicted fourth helix (which contains 13 positively-charged residues). These residues alone were able to increase the XPA's affinity to Y-shaped ssDNA-dsDNA junctions by five times [58]. The high BFE values and the large number of intermolecular contacts exhibited by the XPA_{98-239} homodimer in this study as compared to the XPA_{98-210} homodimer are indicative of the fact that the former is indeed a very stable DBD of XPA.

Using an MS footprinting method, Hilton and his group [126] identified Lys residues K221, K222, K224, and K226 present in the redefined DBD model of XPA, which rendered the XPA flexible enough to accommodate ssDNA/dsDNA as a clamp. As shown earlier in **Figure 3.6**, we now know that the fourth helix of XPA's refined DBD (aa210-239) has more positively-charged and neutral residues to counter the weak binding shown earlier in XPA_{98-219} DBD. According to the study conducted on 129 DNA-protein interactions, the residues Arg, Lys, Ser, Thr, Asn, and Gln have been reported to have a higher affinity toward DNA, forming a large number of hydrogen bonds with both the bases as well with backbone phosphates of the DNA [399]. The same study also showed that Arg and Lys, in particular, have a greater bonding inclination with guanine while Asn and Gln favor adenine. They further observed that Ser and Thr preferred binding the DNA via backbone interaction rather than with base

interactions, while Lys, His, Thr, Arg, and Phe residues interacted with the sugar and phosphate regions of DNA using vdW forces. Our previous study on the behavior of XPA in the presence and absence of DNA has also shown that the stability of XPA-DNA is dependent on these types of DNA-protein interactions [400].

The biochemical data provided by Sugitani et al. [176] was able to shed more light on the significance of the $XPA_{219-239}$ region. To test the importance of the residues belonging to the extended region of XPA in ensuring stronger XPA-DNA interaction, they conducted a mutational analysis of C-terminal residues. K221E, K222E, and R228E. These mutations led to the collapse of the NER process. Their XPA-DNA interaction uses two XPA constructs. XPA_{98-227} and XPA_{98-234} further confirmed that truncating the C-terminal residues led to the reduction of its DNA-binding affinity as a whole.

Most of the DNA repair proteins have been known to exist in the form of homodimers, heterodimers, or oligomers. Even *E. coli*'s UvrA, a prokaryotic NER repair protein, is known to exist as a dimer [401]. So, it was no exception when XPA was also reported to exist as a homodimer by various research teams. As observed in the previous works of literature, we also observed the precedence of XPA_{98-239} homodimer over XPA_{98-210} after analyzing the complex stability, intermolecular PPIs between the respective monomers, and the BFE analysis for both the homodimers. Though we have tried to address the stability and the viability of XPA's dimer status, and its refined DBD region through this study, many questions about XPA still lie ahead that needs to be answered. We can surely hope that with more scientific advancements in the study of XPA, the time may come when we may be able to address the shortcomings of the adjuvant therapy site for the reduction of cisplatin drug resistance and potential drug therapy for the treatment of various disease phenotypes based on the NER outcomes.

3.5. Conclusion

Our computational study on the investigation of XPA's two homodimer models (XPA_{98-210} and XPA_{98-239}) has provided us with considerable insights regarding their structural and conformational stability to determine the best representative of the DBD of XPA. From the comparative analyses of the MD trajectories (RMSD, Rg, SASA, and

hydrogen bond analysis) and the PPI profile studies for both the homodimer systems, it was concluded that the XPA_{98-239} homodimer is much more stable than the other XPA dimer variant. Both homodimers were seen to be largely stabilized by hydrogen bonding, salt bridges, and numerous hydrophobic interactions, whereas the XPA_{98-239} homodimer had a larger count of these interactions than the XPA_{98-210} homodimer. In both cases, the residues Lys, Glu, Gln, Asn, and Arg were observed to play an active part in the stability and dimerization status of the proteins, as deduced from the PPI analysis. The BFE calculations also indicated the XPA_{98-239} homodimer to be a better model than the XPA_{98-210} homodimer, as their PBSA and GBSA differences were -19.14 kcal mol^{-1} and -16.98 kcal mol^{-1}, respectively.