CHAPTER 1

Introduction and review of literature

CHAPTER 1

1. Introduction and review of literature

1.1 DNA and repair pathways

DNA, being the vital carrier of the genetic material in all living beings has to have its integrity conserved, whether it is eukaryotic or prokaryotic organisms However, the toxins from ultraviolet (UV) light, environment, ionizing radiation, chemotherapeutic drugs, and various other mutagens continuously attack the existence of the DNA. The estimated number ~of 10^{13} cells are subjected to tens of thousands of DNA-damaging events per day. Cellular oxidative metabolic products from mitochondria, replication errors, and cellular inflammation are other variables that further add to the DNA damage 1-5].

Cells, therefore, have many DNA repair systems in place to combat these negative consequences known as DNA damage response (DDR) [2,3]. In some circumstances, if the lesions are not removed—either due to the damage load being too high or to a deficiency in a necessary repair pathway—the cell cycle can be arrested until the damage is repaired, and if this does not happen quickly, the cell may be eliminated by apoptosis or may accumulate mutations and transform into a potentially cancerous cell that may proliferate uncontrollably and result in cancer or tumor development [4]. The frequency of illnesses that put people at risk for developing cancer and neurological conditions, like Ataxia-telangiectasia (AT), Fanconi anemia (FA), Cockayne syndrome (CS) Trichothiodystrophy (TTD), and Xeroderma pigmentosum (XP), which are specially brought on by deficits in DNA repair, highlights the significance of DDR [5-10].

1.1.1 DNA damaging substrates

DNA damages are resultant of either intrinsic/endogenous or extrinsic/exogenous agents. Most of the DNA damage is due to endogenous agents, but sometimes exogenous agents equally create more havoc on the DNA [5-7].

Spontaneous hydrolysis is the most basic type of endogenous DNA damage. The main vulnerable element to acid-catalyzed hydrolysis is the *N*-glycosidic link that connects the DNA base with the deoxyribose, causing abasic or AP sites (apurinic/apyrimidinic sites), at the rate of about 10,000 per cell per day. Chemically vulnerable AP sites experience β-elimination, which causes DNA strand scission. The deamination of DNA bases bearing exocyclic amino groups is a typical hydrolysis reaction [8-10]. The most common of these lesions is the production of

uracil from cytosine, which is thought to occur 100–500 times per cell daily. Adenine and guanine can also spontaneously deaminate to generate hypoxanthine and xanthine, respectively, albeit this happens considerably less frequently [5, 11-14]. Endogenous DNA damages also occur due to deletions, insertions, and mismatched DNA strands. replication-related DNA damage like 8-oxo-dGTP and dUTP, and abortive topoisomerase activities [5, 16-17].

The chemical modifications of reactive molecules also result in DNA damage that usually occurs during normal cellular metabolism. Reactive oxygen species (ROS), which include O2, H2O2, and •OH, are some of the most significant of these molecules). Over 100 different types of oxidative DNA adduct are produced by ROS, such as single- or double-strand (ss/ds) breakage, DNA-protein cross-links base modification, deoxyribose oxidation, and. Similar oxidative adducts can also be produced by endogenous reactive nitrogen species, most notably nitric oxide and its byproducts. The 8-oxoguanine is the most widely researched oxidative DNA lesion and is frequently utilized as a quantitative indicator of oxidative DNA damage in biological systems. Alkylation is another form of DNA deterioration linked to endogenous reactive chemicals. S-adenosylmethionine (SAM), nitrosated amines, methyl radicals produced by lipid peroxidation, and the endogenous methyl donor are some potential possibilities for such agents. Nucleobase O and N atoms are the main locations for alkylation [5, 17-20].

Aside from the many endogenous sources of DNA damage, cellular DNA is always under attack by exogenous or environmental DNA-damaging substances. These include physical strains like exposure to UV from the sun, which predominantly results in two forms of DNA lesions: cyclobutane pyrimidine dimers (CPD) and 6-4 pyrimidine photoproducts (6-4PP). Both of these lesions are caused by an unusual covalent link between neighboring pyrimidine bases. Ionizing radiation, which can come from both organic (like cosmic and gamma radiation) and synthetic (like medical procedures like X-rays and radiotherapy) sources, is another external, physical source of DNA damage. The most dangerous DNA lesions caused by ionizing radiation are double-strand breaks, which are a variety of DNA lesions. Ionizing radiation can also cause indirect damage to DNA by generating ROS [5, 21-22].

1.2 Types of DNA repair pathway

There are four major DNA repair pathways:

I. Base excision repair (BER) pathway
II. Mismatch repair (MMR) pathway
III. Double-strand break (DSB) pathway
IV. Nucleotide excision repair (NER) pathway

The comparison between these repair pathways is shown in **Figure 1.1**.

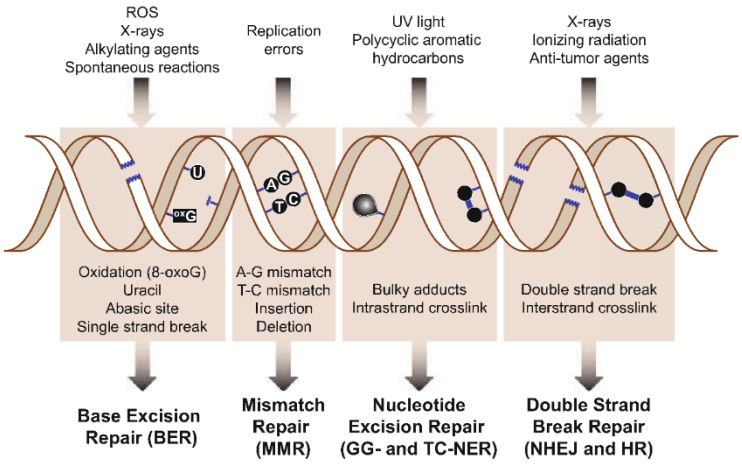

Figure 1.1. Comparative DNA damage and repair mechanisms. Taken from [5].

1.2.1 Base excision repair (BER) pathway

In BER, DNA damage repair usually involves removing the offending base or nucleotide and replacing it with the appropriate base or nucleotide. By cleaving the connection between the broken base and the deoxyribose, a glycosylase enzyme removes the damaged base from the DNA during the repair process. These enzymes produce an apurinic or apyrimidinic site (AP site) by removing a single base. At the AP site, AP endonucleases nick the DNA backbone to repair the damage. The damaged area is then removed by DNA polymerase utilizing 5' to 3' exonuclease activity, and the fresh strand is correctly synthesized using the complementary strand as a template. The DNA ligase enzyme then fills the gap (**Figure 1.2**) [5, 23-27]. Other enzymes that are involved in the BER are shown in **Table 1.1**.

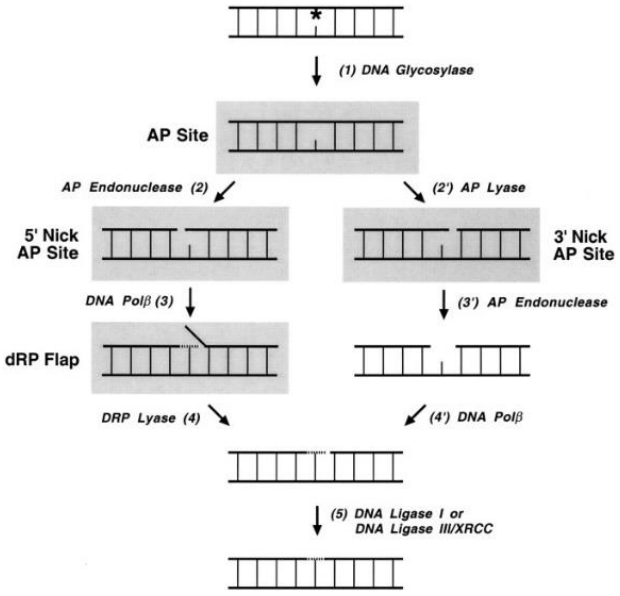

Figure 1.2. Schematic representation base excision (BER) pathway. Taken from [27].

Table 1.1. Enzymes and genes in DNA repair mechanisms. Taken from [27].

Base excision repair (BER)	DNA glycosylase, APE1, XRCC1, PNKP, Tdp1, APTX, DNA polymerase β, FEN1, DNA polymerase δ or ε, PCNA-RFC, PARP
Mismatch repair (MMR)	MutSα (MSH2-MSH6), MutSβ (MSH2-MSH3), MutLα (MLH1-PMS2), MutLβ (MLH1-PMS2), MutLγ (MLH1-MLH3), Exo1, PCNA-RFC
Nucleotide excision repair (NER)	XPC-Rad23B-CEN2, UV-DDB (DDB1-XPE), CSA, CSB, TFIIH, XPB, XPD, XPA, RPA, XPG, ERCC1-XPF, DNA polymerase δ or ε
Homologous recombination (HR)	Mre11-Rad50-Nbs1, CtIP, RPA, Rad51, Rad52, BRCA1, BRCA2, Exo1, BLM-TopIIIα, GEN1-Yen1, Slx1-Slx4, Mus81/Eme1
Non-homologous end-joining (NHEJ)	Ku70-Ku80, DNA-PKc, XRCC4-DNA ligase IV, XLF

1.2.2 Mismatch repair (MMR) pathway

The MMR system generally corrects errors that are not corrected by proofreading. It is crucial in the post-replication repair of incorrectly integrated bases that have evaded the replication polymerases' proofreading function. MMR proteins fix insertion/deletion loops (IDLs), which are caused by polymerase slippage during replication of repetitive DNA sequences, in addition to mismatched bases. The germline mutations in the MMR genes put people at risk for developing several malignancies, including Lynch syndrome-associated hereditary non-polyposis colon cancer. The MMR process can be broken down into three main steps: (i) identification, where mispaired bases are detected, (ii) excision, where the error-containing strand is degraded, and (iii) gap, and repair synthesis, where the gap is

filled by DNA resynthesis. The insertion/deletion loops (IDLs) resulting due to polymerase slippage during the replication of repetitive DNA are also corrected by MMR [5, 28-30].

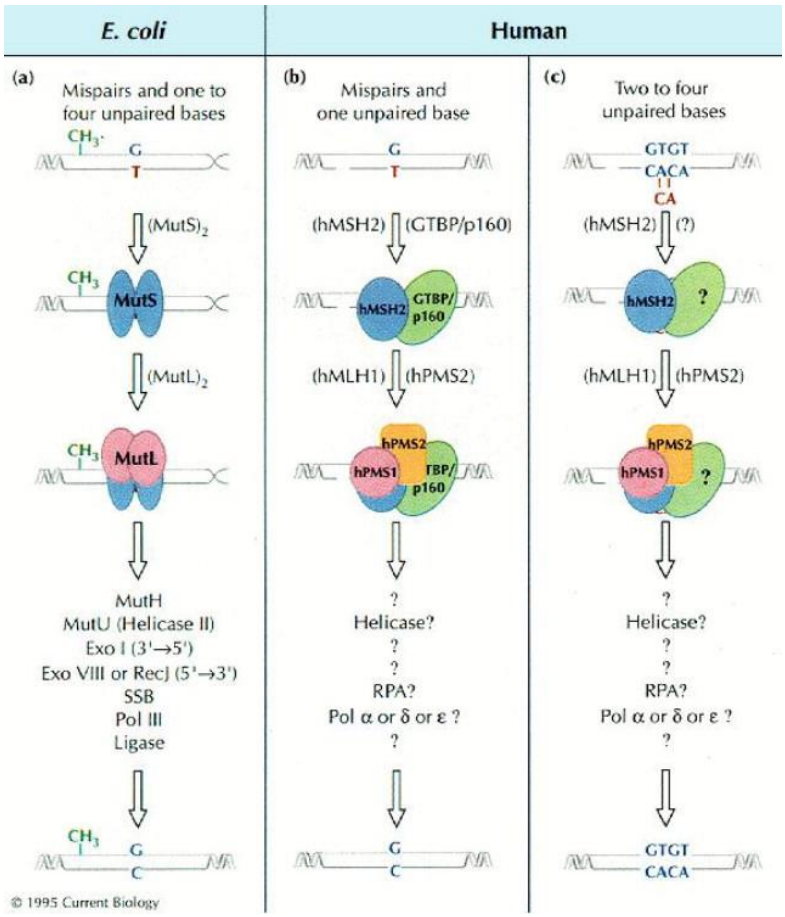

Figure 1.3. Comparison between prokaryotic and eukaryotic MMR pathways. Taken from [33].

This process, in particular, is highly conserved from *E. coli* to humans. The proteins of MMR in *E. coli* are Mut class proteins: MutS and Mut, where MutS recognizes the mismatch site, and MutL continues to process downstream of MMR events, leading to the erroneous strand removal. the removal of the strand containing the error. Most Eukaryotes have analogs of MutS and MutL, MSH for MutS, and MLH for MutL. MutH is only found in bacteria (**Figure 1.3**). Following the removal of the damaged site, DNA polymerase starts the resynthesis process, and DNA ligase repairs the nick [5, 31-33]. The enzymes that are associated with the MMR pathway are shown in **Table 1.1**.

1.2.3 Double-strand break (DSB) repair

The most dangerous kinds of DNA damage from a biological standpoint are double-strand breaks (DSBs). An inaccurate repair can result in deletions or chromosomal abnormalities, where no strand can be used as a template for the repair mechanisms. As a result, the cell cannot complete mitosis when it divides next and will either die or, in rare instances, result in the emergence of cancer and other diseases of genomic instability and genomic rearrangements.

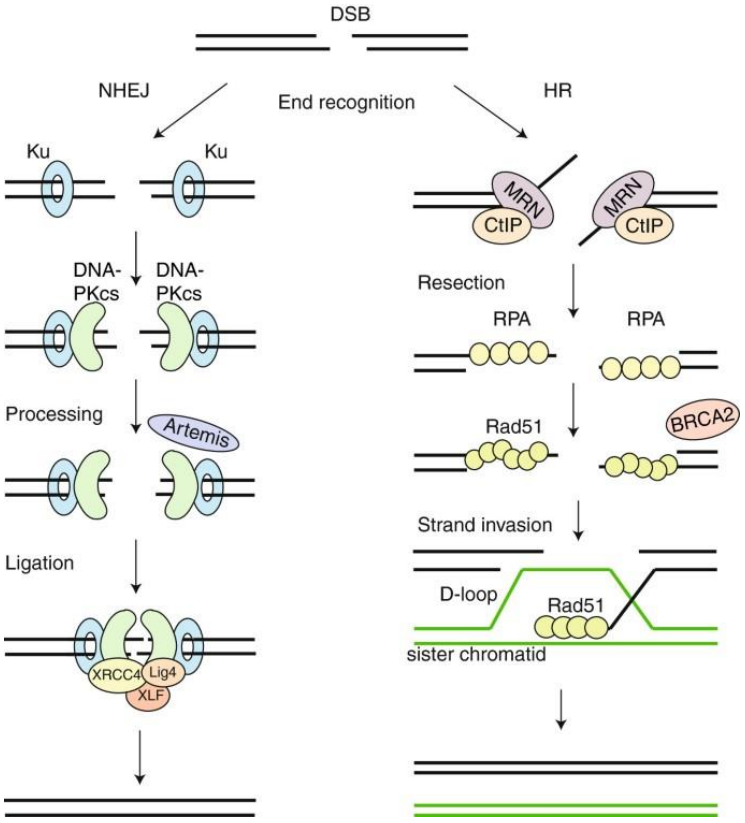

Figure 1.4. Illustration of double-strand break (DSB) pathway. Taken from [41].

Therefore, it is essential for both cell survival and the preservation of genome integrity for DSBs to be repaired [5, 34-36]. Homologous recombination (HR) and non-homologous end-joining (NHEJ) are the two primary processes by which mammalian cells repair DSBs. Because it uses the genetic data found in the sister chromatid that is still intact as a template, the HR pathway is largely an error-free method [37-39], while NHEJ, in contrast, typically entails ligating the damaged ends directly to eliminate DSBs and is error-prone as shown in **Figure 1.4** [41-43]. According to logic, NHEJ predominates over HR in mammalian cells, which are only capable of HR in the late-S and G2 phases of the cell cycle [5, 34-47]. Other remaining enzymes involved in the DSB pathway are shown in **Table 1.1**.

1.2.4 Nucleotide excision repair (NER) pathway in brief

A wide range of large, helix-distorting lesions from DNA can be recognized and repaired by the highly adaptable repair mechanism known as NER. The most significant of these lesions are pyrimidine dimers, which are created by the UV component of sunlight and include 6-4 photoproducts (6-4PP) and cyclobutane pyrimidine dimers (CPD) (**Figure 1.5**). Cisplatin-DNA intrastrand crosslinks are another notable substrate of NER. At the site of the DNA damage, repair proteins are sequentially assembled to mediate NER. Despite sharing a similar function to the BER route, the NER pathway is more complicated and requires about thirty different proteins to complete a multi-step "cut-and-patch" method. Recognition of the DNA damage, the local opening of the DNA helix around the lesion, excision of a brief single-strand DNA segment spanning the lesion, sequential repair synthesis, and strand ligation are the steps in this process [5,46-49]. An in-depth process of NER will be discussed below.

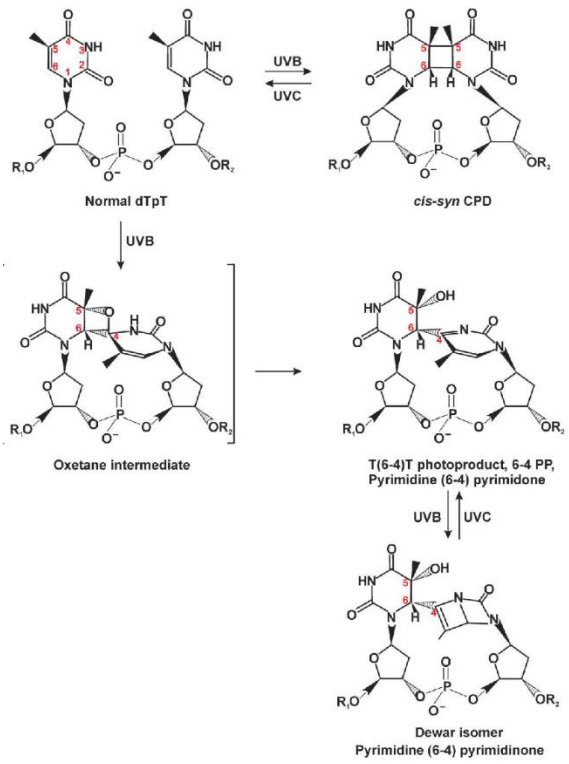

Figure 1.5. The chemical structures of DNA photoproducts are caused by sunlight. Taken from [50].

1.3. Nucleotide excision repair (NER) pathway

NER is a key DNA repair system that can remove DNA damages among all other repair pathways. It protects DNA from various physical and chemical damages. The

massive covalent lesions repaired by NER are (i) cis-syn CPDs and 6-4PPs formed due to UV light, (ii) polycyclic aromatic hydrocarbons induced by cigarette smoking, and (iii) helix distorting intrastrand crosslinks, which are caused by chemotherapeutic drugs like cisplatin, and iv) minor base changes implicated by alkylating and oxidizing agents. Defects in NER protein machinery lead to autosomal recessive conditions, like Xeroderma pigmentosum (XP), Cockayne syndrome (CS), and Trichothiodystrophy (TTD), while enhanced expression of NER proteins can often lead to clinical resistance towards platinum-based chemotherapeutics as it can easily identify cisplatin drugs in the body. [5,9,10, 48-51].

This mechanism, in particular, is carried over by 30-40 different proteins forming various types of protein-protein complexes (PPCs), protein-protein interactions (PPIs), and DNA-protein interactions (DPIs). It is majorly carried by xeroderma pigmentosum proteins- XPA, XPB, XPC/human homolog Rad23 B (HHR23B), XPD, XPE, XPF, and XPG. Other proteins involved are Cockayne syndrome proteins A and B (CSA and CSB), excision repair cross-complementation group 1 (ERCC1), transcription factor II H protein (TFIIH), replication protein A protein (RPA), replication protein C protein (RPC), proliferating cell nuclear antigen (PCNA), DNA polymerase ε and δ proteins, and trichothiodystrophy A protein (TTDA) [51,52]. Among these proteins, XPA, XPC/HHR23B, RPA, TFIIH, XPG, and ERCC1-XPF are NER core members that interact with each other and forms pre-incision complex (PIC). PPIs of PIC are shown in **Table 1.2**.

Table 1.2. Functions, and main interactions of NER core members. Taken from [46 and 57].

NER factors	**Molecular weight (kDa)**	**Interactions with other NER core factors**	**NER function**
XPC–HHR23B	106+43	XPA, TFIIH	Damage recognition and NER factors recruitment
RPA	70+32+14	XPA, XPG	PIC anchor and NER factors recruitment
XPA	40	ERCC1, RPA, TFIIH, XPC-RAD23B	PIC scaffold and NER factors recruitment

TFIIH	460	XPA, XPC, XPG	Damage recognition, helicase, and NER factors recruitment
XPG	113	RPA, TFIIH	3′ endonuclease
ERCC1–XPF	38+112	XPA	5′ endonuclease

NER is generalized in two sub-pathways: global genome NER (GG-NER) and transcription-coupled NER (TC-NER), differing only in the damage recognition step (**Figure 1.6**}. The multi-step cut-and-paste process of NER takes in following steps.

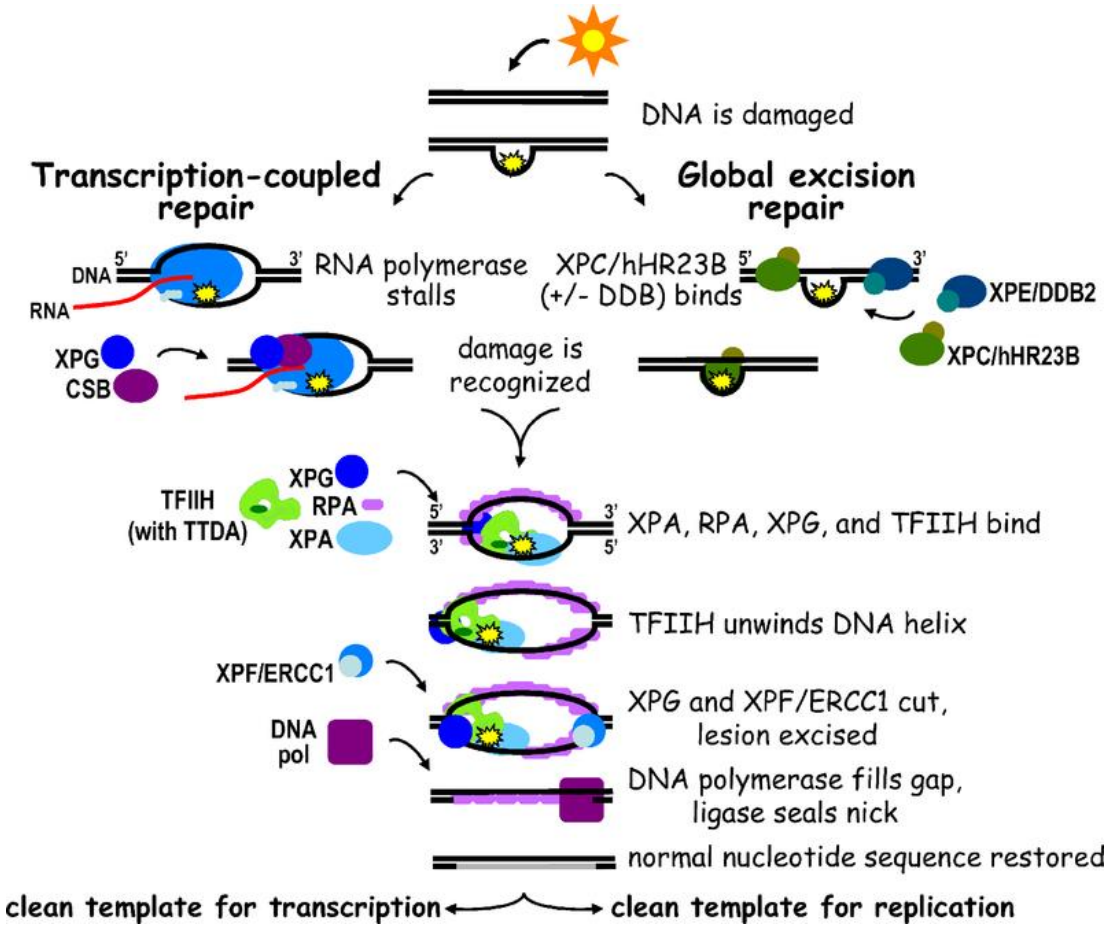

Figure 1.6. Mechanism of eukaryotic nucleotide excision repair. Taken from [53].

i. DNA lesion recognition

The genomic damage is usually repaired by GG-NER, while the TC-NER, is in charge of speeding up the repair of lesions in the template DNA strand of genes that are actively transcribed (**Figure 1.7**). In GG-NER, XPC in complex with proteins

HHRAD23B and centrin 2 (CETN2) is the primary damage recognizer and holds on to the ssDNA injury. The primary subunit of the complex, the XPC, is stimulated to bind to DNA by the small subunits, which stabilize the XPC structure collectively, altering some PPIs, and improving NER effectiveness both in vitro and in vivo conditions. followed by which entire repair machinery is then recruited to the NER pathway. XPE/DNA damaged binding 2 protein (DDB2) usually recognized the UV damaged site (CPDs) [4, 59] XPC/HHR23B does not diffuse along with DNA by "sliding," but rather by "hopping" (diffusion through frequent microscopic dissociation and reassociation with the DNA). One benefit of the hopping mode is that it enables a protein to get over protein barriers on DNA. XPC/HHR23B complex inserts two-hairpin modules from the BHD2/BHD3 domains into the DNA duplex and creates a stable DPC in the context of helical distortion and base-pair disruption. Rad4/XPC flips out damage-containing nucleotide pairs to establish an "open" conformation in this complex by interacting only with the nucleotides on the unharmed strand [4, 52, 60-62]. Because they only result in a slight deformation of DNA, CPDs, the most frequent photolesions are not well detected by XPC but are very well recognized by a unique protein DNA damage-binding protein (DDB), a heterodimeric protein made up of DDB1 and DDB2/XPE.

Unlike XPC, DDB2 interacts directly with UV light-induced photolesions in DNA, generates a kink in the duplex, and makes a substrate for XPC that is more appropriate. The two CPD nucleotides are flipped out by DDB2 into a shallow binding pocket, which can accommodate lesions like CPDs or 6-4PPs through form complementarity, according to structural analyses. DDB2 is believed to assist XPC by promoting the rearranging of the chromatin DNA. DDB1 is additionally a connector protein for the ubiquitin ligase, CUL4-RBX1. Upon DDB2 interaction, CUL4-RBX1 is activated and ubiquitinates XPC and DDB2, after which the proteasomal destruction occurs, following NER complex extraction. Damage transfer from DDB2 to XPC occurs in tandem with the TFIIH complex arrives, further promoting DDB2 dissociation [4, 52, 63-66].

In TC-NER, the damage site is recognized by stalling of elongating RNA polymerase II (RNAPII). CSB protein, a member of the SNF2 family of DNA-dependent ATPases binds loosely with the elongating RNAPII to induce transcription but becomes more tightly bound following transcription arrest. As

RNAPII stalls at a lesion, the RNAPII-bound CSB recruits the CSA protein. CSA/CSB work together to polyubiquitinate the K1268 residue of RPB1, a subunit of RNAPII. A master switch for switching between transcription, RNAPII breakdown, and the start of DNA repair is the RPB1 ubiquitination. The UVSSA (UV-sensitive syndrome protein A) and the stalled RNAPII are then brought together by CSA. The main element that attracts the TFIIH complex is UVSSA, which in turn recruits XPA to the injury site, followed by which the unwinding of the DNA starts [4, 52, 67-71].

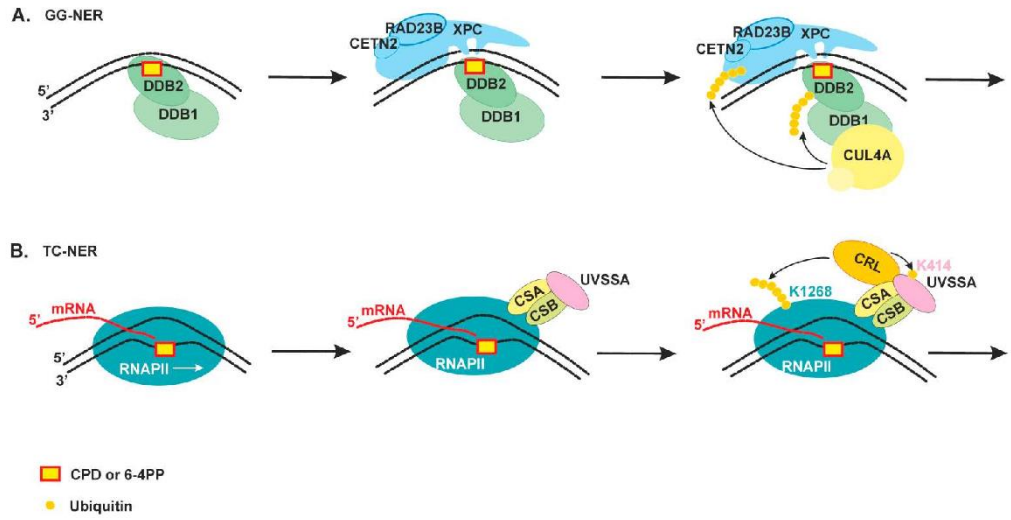

Figure 1.7. An overview of the damage recognition step of NER. Taken from [4].

ii. DNA damage verification, the unwinding of the DNA helix around the lesion, and Pre-incision complex (PIC) formation.

In this step, the unwinding of the DNA often called the NER bubble and DNA lesion demarcation occurs (**Figure 1.8**). XPA protein, known as a scaffold protein, acts as a primary damage verifier in NER and orchestrates the assemblage of the NER core members- RPA, TFIIH, ERCC1/XPF to the damage site. Since this step occurs before the dual incision of the damage site, it is called a pre-incision complex (PIC) [4, 72].

XPA verifies the damage site with or without the help of TFIIH. Recent research has demonstrated that XPA uses episodic one-dimensional diffusion to look for DNA damage. This was demonstrated by atomic force microscopy, scanning force microscopy, and mathematical modeling [73]. TFIIH, a helicase comprises ten subunits, seven of which make up the core complex (ERCC2/XPD, ERCC3/XPB,

GTF2H1/p62, GTF2H4/p52, GTF2H2/p44, GTF2H3/p34, and GTF2H5/TTDA/p8). XPD protein connects the cyclin activating kinase (CAK) -subcomplex, which includes CDK7, MAT1, and cyclin H, to the core complex. The ERCC2/XPD and ERCC3/XPB subunits, which have helicase and ATPase functions, aid in the formation of the transcription/NER bubble. TFIIH is enlisted in the instance of GG-NER through XPB engagement to the DNA duplex, interaction with the C terminus of XPC, and an additional interaction of the p62 subunit with XPC's N terminus. The XPD helicase is then loaded onto the DNA due to its position at the opposite end of the TFIIH arch. The fact that XPC and UVSSA have an interaction surface on the p62 subunit of TFIIH, which suggests that the two routes at least partially share a mechanism for engaging TFIIH with the lesion site, is a startling resemblance between GG-NER and TC-NER. Human TFIIH interacts with the PIC downstream of RNAPII, which moves in the 3′->5′′direction. The adaptable TFIIH structure enables XPD to unwind DNA while tracking in the 5′–3′ direction. The association of XPB and XPG triggers the release of CAK from TFIIH [5, 72, 73].

Biochemical evidence suggests that XPA can increase TFIIH's overall helicase activity while, in the presence of lesions, inhibiting it. As a result, XPA also aids in damage verification. Additionally, XPA can recognize bulky damage as well and prefers to bind DNA structures that are kinked and branched [72]. The effect of XPA and XPG on TFIIH has been further evaluated by Kokic and team [74] from a cryo-electron microscopy study: (1) an alternative conformation of TFIIH is stabilized by XPA and XPG where the XPD helicase is opened for functioning; (2) XPA and XPG stimulate XPB and XPD, and this event facilitates DNA opening; consequently, they are present in a ternary complex in the lesion-scanning mode; (3) XPA interacts with an XPB subunit in the TFIIH-DNA complex and marks the DNA repair at 5′ site of the lesion; (4) XPA helps XPD location on the single-stranded 3′ extension by acting as a bridge between XPB and XPD.

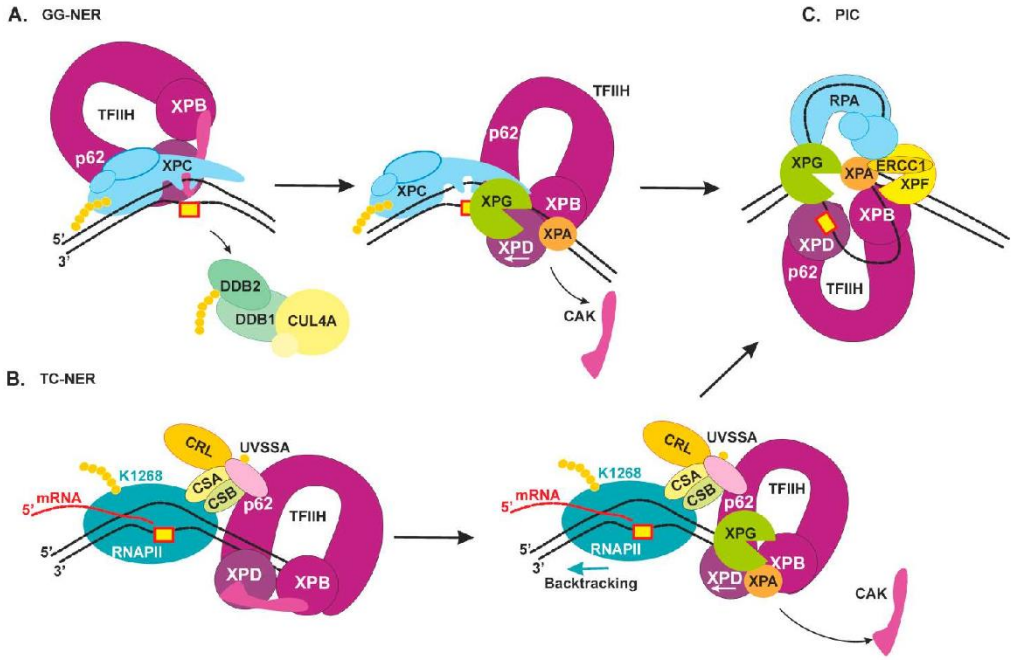

Figure 1.8. Schematic view of the damage verification step of NER and pre-incision complex formation. Taken from [4].

The unharmed strand interacts with RPA, which shields it from a nuclease attack [5,75]. The length of the NER-excised fragment is approximately 30 nt, which is the maximum length of the ssDNA platform for RPA binding. RPA binds to this platform with a specified 5′–3′ polarity [76]. Within the repair bubble, RPA and XPA have close interactions that allow them to jointly control the proper orientation and activation of NER nucleases [77]. Additionally, it has been observed that RPA and XPA form a complex without DNA as well as a ternary complex when DNA is present, and XPA interaction with RPA is necessary for NER [75]. In a study on the crystal structure of *Ustilago maydis*, RPA was seen to be complex with ssDNA in a U-shaped model [78].

Overall, TFIIH stops at the lesion, RPA covers the unharmed opposite strand, XPA and XPG indicate the 5′ and 3′ edges of the NER bubble, respectively, and XPF-ERCC1 binds in the back of XPA. The entry of XPF-ERCC1, which is recruited by XPA, completes the development of the NER PIC.

iii. Dual incision

The lesion which consists of 24-32 nucleotides are removed after the local unwinding (NER bubble) and demarcation of the damage site (**Figure 1.9**). Structure-specific endonucleases, ERCC1/XPF, and XPG are needed for this step.

XPF-ERCC1 initially cleaves the DNA by moving from the 5′ site to the damage site while forming a free 3′ -OH group, while XPG cuts the bubble site at the 3′ site [52, 79-81]. After that, replication equipment can be loaded to begin repair synthesis. As it interacts with PCNA, RPA encourages the arrival and placement of replication factor C (RFC) protein and improves repair synthesis with potential assistance from XPA [5, 52, 82,83]. In the absence of an XPG-made incision, repair synthesis can continue halfway across the gap [5,52]. Possible PCNA-XPG interaction is what causes the XPG-made 3′ incision. The interaction between PCNA and XPG results in the XPG-made 3′ incision [84]. The repair bubble releases the lesion-containing oligonucleotide in combination with TFIIH. Following ATP attachment, TFIIH gradually separates from the excised oligonucleotide, which is then bound by RPA or broken down by cellular nucleases [5].

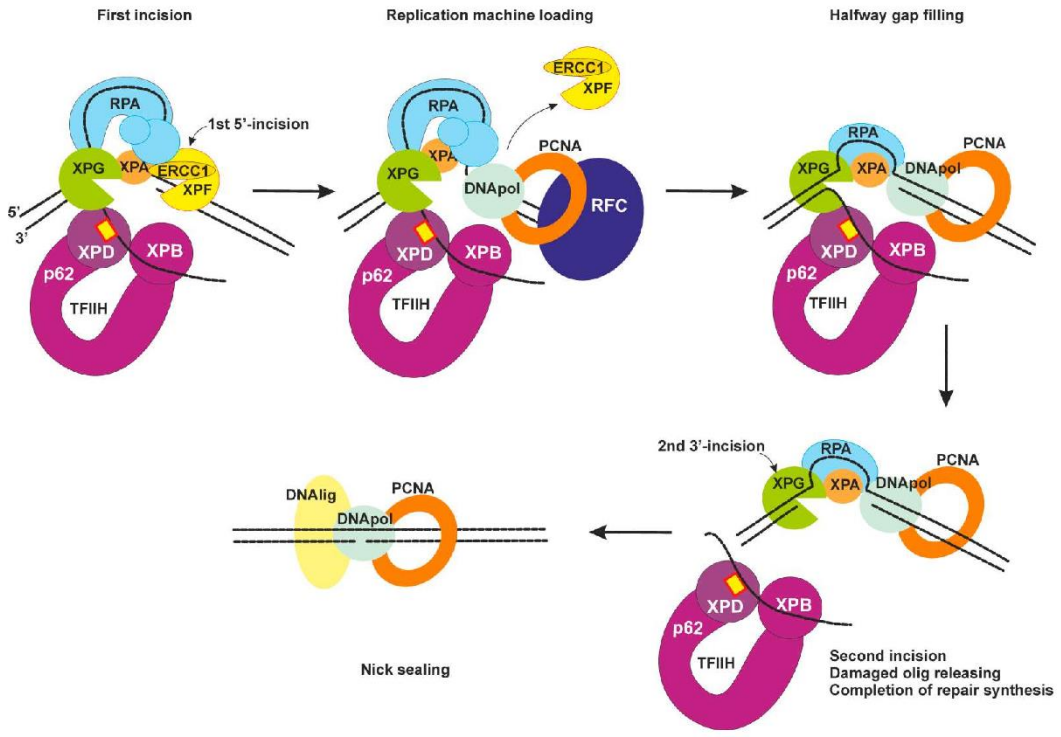

Figure 1.9. Later stages of the NER pathway: dual incision, gap-filling, and ligation. Taken from [4].

iv. Gap-filling and replication of new strand

As per the *in vitro* studies, different groups of replication machinery carry out this step of repair synthesis. These include DNA polymerases δ and ε, sliding clamp PCNA, clamp loader RFC, and RPA (**Figure 1.8**) [85].

v. Ligation of newly synthesized strand

Ligation or the nick-sealing of the newly synthesized strand to the original strand is done by either DNA ligase I or DNA ligase III (**Figure 1.9**). DNA ligase I or DNA ligase IIIα along with XRCC1 uses the incision site left at 5′-phosphate by XPG for nick-sealing [4,5, 57, 72, 86-88].

1.4. NER deficiency outcomes

The three major resultants of NER defects can be seen in recessive conditions line Xeroderma pigmentosum (XP), Cockayne syndrome (CS), and Trichothiodystrophy (TTD) diseases. There are seven complementation groups of proteins resulting in XP (XPA_{A-G}), two in CS (CSA and CSB), three in combined XP/CS patients (XP-B, XP-D, XP-G), and three in TTD (XP-D, XP-B, and TTDA). Each of these diseases has gene defects in a particular gene. The predisposition of skin cancer and sun sensitivity is seen in people affected with XP, except for CS and TTD. Out of these three NER defected disease outcomes, XP disease is the most detrimental one as it causes carcinomas, photosensitivity, neurological damage, sensorineural deafness, pre-eclampsia, colorectal cancers, etc. [52, 89, 90]. The comparative analysis between XP, CS, and TTD has been shown in **Table 1.3** and **Table 1.4**.

Table 1.3. Comparison between XP, CS, XP/CS complex, and TTD. Taken from [89].

Feature	XP	XP with neurolgical disease	TTD	CS	XP/CS complex
Skin					
Skin sun sensitivity	Yes	Severe	Yes/No	Yes	Yes
Increased skin pigmentation	Yes	Yes	No	No	Yes
Sunlight-induced skin cancer	Yes	Yes	No	No	Yes
Eyes					
Photophobia	Yes	Yes	Yes/No	Yes	Yes
Conjunctival growths	Yes	Yes	No	No	Yes
Cancer (anterior eye/lids)	Yes	Yes	No	No	Not reported
Congenital cataracts	No	No	Yes	Yes	No
Pigmentary retinal degeneration	No	No	No	Yes	Yes
Somatic					
Short stature	No	No/Yes	Yes	Yes	Yes
Immature sexual development	No	No	No/Yes	Yes	Yes
Nervous system					
Sensorineural deafness	No	Yes	No	Yes	Yes
Developmental delay	No	Yes	Yes	Yes	Yes
Progressive neurological degeneration	No	Yes	Unknown	Yes	Yes
Primary neuronal degeneration	No	Yes	No	No	No
Dysmyelination	No	No	Yes	Yes	Yes
Cerebral atrophy	No	Yes	No/Yes	Yes	Yes
Cerebellar atrophy	No	Yes	No	Yes	Yes
Calcification (basal ganglia)	No	No	No/Yes	Yes	Yes
Disease mechanism					
NER defect	Yes	Yes	Yes	Yes	Yes
Reaction to exogenous or endogenous damaging agents	Yes, severe	Yes, severe	No	Yes	Yes
Developmental defect	No	Yes	Yes, severe	Yes, severe	Yes, severe

Table 1.4. Comparison of clinical features of patients with XP, XP neurological disease, XP/CS, and CS. Taken from [4].

Clinical Features	XP	XP Neurological Disease	XP/CS	CS
Molecular defects	XP-A, -B, -C, -D, -E, -F and XP-V	XP-A and XP-D followed by XP-B, XP-G, and XP-F	XP-B, XP-D, XP-G	CS-A, CS-B
Skin sun sensitivity	+	+	+	+
Increased freckling	+	+	+	-
Sunlight—induced skin cancer	+	+	+	-
Photophobia	+	+	+	+
Anterior eye cancer	+	+	+	-
Retinal degeneration	-	-	+	+
Sensorineural deafness	-	+	+	+
Developmental delay	-	+	+	+
Progressive neurological degeneration	-	+	+	+
Primary neuronal degeneration	-	+	-	-
Demyelination of brain	-	-	+	+
Cerebral atrophy	-	+	+	+
Cerebellar atrophy	-	+	+	+
Calcification (basal ganglia)	-	-	+	+

1.4.1 Xeroderma pigmentosum (XP)

Xeroderma (skin parchment) and pigmentosum (freckles) is the characteristic feature of this disease. XP patients have defects in their XP genes, which comprises XPA, XPB, XPC, XPD, XPE, XPF, XPG, and XPV genes (**Table 1.5**). All XP diseases from the defective genes exhibit similar symptoms of photophobia, and skin cancers (**Table 1.6)**, except for the XP phenotype due to defective XPA, which results in neurological degeneration (50%), making it the most lethal one [4, 91]. XPB, XPD, and XPG also exhibited severe NER deficiencies. >50% defect in the repair process, and inactivation or deletion of XPD genes causes the disease. XP patients with defective XPF are due to poor coordination with ERCC1. Since XPC is only required in GG-NER, its defect causes a wide range of sunburns in patients [90]. XP with XPA deficit displays high sensitivity to sun exposure and is more prone to cancers of the eyes and skin, and neurological impairments. In 60% of cases, the early onset of this disease is within the first week of life to 2 years. It is observed widely in Japan (frequency of 1 per 22,000), the Middle East, North Africa, and India, and rarely seen in Europe and North America (frequency of ~2.3 per million live births) [4, 57, 72, 91, 92]. According to recent studies, estimates of basal and squamous cell carcinomas and malignant melanoma are 10,000 and 2000 times more likely, respectively, to develop in XP individuals before the age of 20 [63, 65]. A significant rise in the prevalence of oral cancer, notably squamous cell carcinoma of the tongue

tip, is also seen in XP patients [4, 54, 92, 93]. Blepharospasm along with serious keratitis and malignancies (epithelioma, squamous cell carcinoma, and melanoma) have also been observed [92, 94]. Many deaths observed in XP patients have been accounted to neoplasia, reducing the lifespan by thirty years [95, 96].

Table 1.5. Loci and genes associated with XP complementation groups and XP variants. Taken from [91].

Groups	Gene involved	locus	frequency (%)
XPA	XPA	9q22.33	30
XPB	ERCC3/XPB	2q14.3	0.5
XPC	XPC	3p25.1	27
XPD	ERCC2/XPD	19q13.32	15
XPE	DDB2/XPE	11p11.2	1
XPF	ERCC4/XPF	16p13.12	2
XPG	XPG	13q33.1	1
XPV	DNA pol η	6p21.1	23.5

Table 1.6. Clinical signs of XP based on complementation groups. Taken from [91].

Group	Gravity	photosensitivity	xerosis	pigmentation	skin cancer	neurological defects	Eye-disorders
XPA	M/S	+/-	+	+	+	+	+
XPB	M/S	+	+	+	+	+	+
XPC	M/S	+	+	+	+	-	+
XPD	M	+	+	+	+	+	+
XPR	M	+	+	+	+/-	+-	+
XPF	V	+	+	+	+	-	+
XPG	M/S	+	+	+	+	+	+
XPV	V	+/-	+	+/-	-	-	+

M: mild; S: severe; V: varies

Another aspect of XP symptoms is a neurological deformity, wherein the patients suffer from reduced tendon reflexes and high-frequency sensorineural hearing loss, ataxia, a loss of the ability to swallow, areflexia, microcephaly, and progressive cognitive impairments, premature aging, neurodegeneration (**Figure 1.10**) [4, 52, 90]. Other symptoms of XP patients have been discussed in **Table 1.3** and **Table 1.4**.

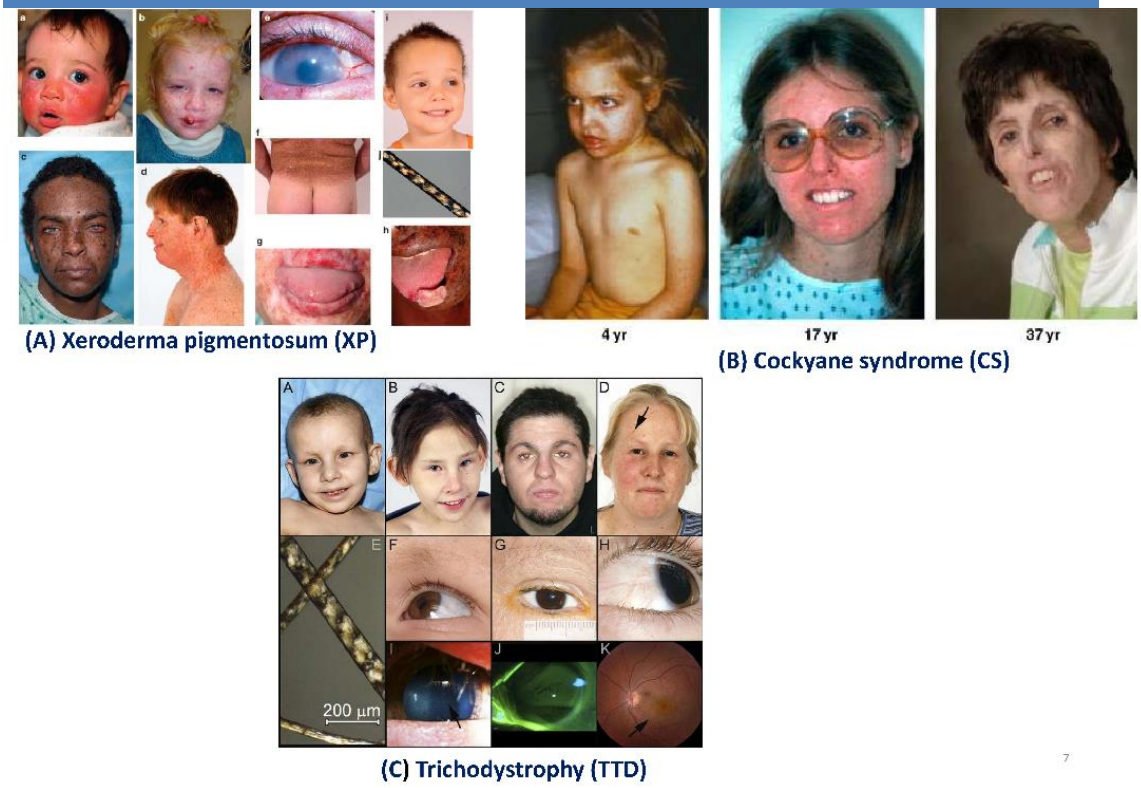

Figure 1.10. Symptoms of NER defects: (A) XP, (B) CS, and (C) TTD. Taken from [89].

1.4.1.1 Treatments for XP patients

XP can be treated with a variety of medical/procedural, surgical, and combination treatments. Antioxidant medications, retinoic acid derivatives, isotretinoin, imiquimod 5% acitretin, 5-fluorouracil immunomodulators, chemical peeling (Trichloroacetic acid) topical liposome lotion containing T4N57 bacteriophage endonuclease, photodynamic therapy (PDT), aminolevulinic acid, cemiplimab, and cryotherapy have all been used in clinical practice (**Table 1.7**). Excision of skin lesions with primary closure, skin grafts, and local flaps, simple or composite, or distant flaps are the surgical techniques that are used the most frequently [91, 94, 96].

Table 1.7. Drugs used in the treatment of XP. Taken from [96].

Age	Sex	Group[b]	Treated tumor	Localization	Treatment	Outcome	Non-target lesions	AEs
51 y	M	XPE	Metastatic melanoma (pT1b, N2a, M1c)	Left cheek with lung, lymph node and right infraorbital metastases	Pembrolizumab 2 mg/kg every 3 weeks	A 90% decrease of the largest lung metastasis and a complete disappearance of the others after 3 months administration	Disappearance of almost all NMSCs and AK in the first 3 months of therapy	Reddish swelling of the skin affected by tumor lesions, mild itching
6 y	F	XPC	Sarcomatoid carcinoma	Scalp with bone lysis, meningeal contact and superior sagittal sinus compression	At first Nivolumab 3 mg/kg every 2 weeks and the monthly (with Cetuximab for 4 cycles)	A 65% tumor volume reduction after 6 infusions with 2-weekly administration	Appearance of a SCC and two melanomas of the scalp, and others skin tumors on the skin, lip and tongue	-
48 y	F	-	Metastatic SCC	Left tight with lymph node metastases	Pembrolizumab 2 mg/kg every 3 weeks	Partial response with regression of all metastases after 3 cycles	-	-
17 y	M	XPC	Metastatic melanoma (pT4b, N0, M1c)	Scalp with hepatic and lung metastases	Pembrolizumab 2 mg/kg every 3 weeks	Partial response with regression of all metastases after 4 cycles	Regression of many pre-existing NMSCs and AK after 4 cycles	Vitiligo-like depigmentation
18 y	F	XPC	Unresectable SCC	Limbus of right eye	Pembrolizumab 2 mg/kg every 3 weeks	Complete regression after 8 months of therapy	Not response of BCCs on the face surgically removed	-
19 y	M	XPE	Unresectable SCC	Right nasal cavity and orbit	Pembrolizumab 2 mg/kg every 3 weeks	Partial radiographic regression	-	-
20 y	F	XPV	(i) Metastatic melanoma; (ii) unresectable SCC	(i) Unknown primary origin; (ii) maxillary sinus	(i) Ipilimumab 10 mg/kg every 3 weeks; (ii) Pembrolizumab 140 mg once a month	(i) Remarkable response; (ii) well response for 31 months until radiographic progression	Development of one BCC on right eyebrow treated with Mohs surgery	-
7 y	F	XPC	Metastatic SCC	Right side of the face with, at first, the involvement of the right sphenoid bone, the cavernous sinus and the right carotid artery, and then the extension to surrounding tissues with lymph node metastases and leptomeningeal spread	Pembrolizumab 2 mg/kg every 3 weeks	A considerable decrease in tumor bulk and the resolution of leptomeningeal disease after five cycles; a long-term sustained stable disease	-	-
32 y	M	XPC	Cutaneous angiosarcoma	Left supraorbital area with submandibular, lung, pleural, mediastinal, pericardial, liver, and bone metastases	Pembrolizumab 200 mg every 3 weeks	Resolution of the lung and bone disease, almost complete resolution of the cardiac and pericardial involvement, and significant reduction in the liver metastases after 4 cycles	-	-

1.4.2. Cockayne syndrome (CS)

CS is named after an English physician Edward Alfred Cockayne (1880–1956), who described it in 1936 and was later described by scientists, Mary M. Dingwall and Catherine A. Neill in two patients, hence CS is also known as Neill-Dingwall syndrome [97], is characterized by cutaneous photosensitivity, growth failure, mild neurological and physical abnormalities, and ocular disorders [90, 98, 99]. It is caused due to the transcriptional defects of CSA and CSB genes involved in the TC-NER pathway (damage recognition step). Oxidative DNA damage initiates the early onset of the symptoms [98-100]. CS patients exhibit severe cachectic dwarfism,

"bird-like" facies, dental caries, kyphosis of the spinal cord, and significant postnatal development failure of the soma and brain in conjunction with early senescence and progressive multiorgan degeneration, segmental demyelinating peripheral neuropathy and very severe patchy myelin loss (tigroid leukodystrophy) are visible in the brains of CS patients (**Figure 1.10**). XP and other neurological conditions linked to poor DNA repair do not exhibit this particular sort of pathology. The basal ganglia and other areas of the brain are also seen to calcify in CS but not in XP. The average lifespan of CS patients is 12.5 years, resulting due to respiratory failures and pneumonia though some may last longer [52, 98-104].

1.4.2.1 Treatments for CS patients

CS patients are treated with physical therapies, minor surgeries like cataract removal, application of sunscreen, and protective clothing [101].

1.4.3 Trichothiodystrophy (TTD)

TTD is often characterized in an individual by their sulfur-deficient brittle hair and ichthyosis in a combination of slight mental and physical retardation (**Figure 1.10**). It is often referred to as Amish brittle hair syndrome, Pollitt syndrome, Tay's syndrome, Sabinas syndrome, and Marinesco-Sjogren syndrome [57]. Stefanini et al. observed that photophobia is seen only in major cases of TTD but not all the time, owing to the defective XPD [105]. A study of twenty families with UV-sensitive TTD cases showed NER defects in the XPD gene [106, 107]. Apart from the symptoms mentioned earlier, other rarely discussed symptoms are decreased fertility, and short stature [108-112].

1.4.3.1 Treatments for TTD patients

Currently, there is no concrete treatment for individuals suffering from TTD, except for the topical usage of sun blockers. However, in 2021, Guber et al. were able to successfully treat TTD in an eight-year-old boy using dupilumab, a monoclonal antibody that disrupted the effects of interleukin (IL)-4 and IL-13 [113].

1.5 Xeroderma pigmentosum complementation group A (XPA) protein

XPA protein is an important part of the NER process, which repairs all the bulky DNA adducts and lesions from the genome. As a scaffold protein, XPA accommodates all the NER core members to the damage site and ensures that proper excision of the lesion occurs from the DNA helix. It acts as a primary agent in damage recognition and damage verification in both GG-NER and TC-NER. Due to its involvement in both sub pathways of NER, the role of XPA in NER becomes even more undeniably strong. It can easily recognize a wide range of DNA damaging substrates such as CPDs, 6-4PPs, PAHs, and alkylating agents like platinum-based chemotherapeutics in the DNA strand. The genetic defect in the XPA gene has been known to cause XP disease with severe occurrences of carcinomas, and neurological damage in a patient in such a way that the symptomatic outcomes of such instances have been known to be devastating, especially since it is the main protein to orchestrate the events before and after the dual incision and excision of the DNA lesion from the strand. However, at the same time, an enhanced expression level of XPA also causes clinical resistance to anti-cancer treatments.

1.5.2 XPA and its structure

XPA is a highly conserved gene right from lower beings to higher mammals (**Figure 1.11**) and is encoded by the XPA gene in human beings, located at 9q22.33 on chromosome 9. It consists of six exons coding for 273 residues and is 40kDa in molecular weight as shown in **Figure 1.12** and **Table 1.2** [57, 58, 72]. It does not possess any enzymatic activity but functions as a scaffold protein by tethering all the requisite NER proteins to the lesioned site [72]. It is composed of a disordered N-terminal region (aa1-98), globular DNA binding domain (DBD) site (aa98-239), and disordered C-terminal region as shown in **Figure 1.13**.

The DBD region of XPA consists of a C4 type zinc-finger motif required for DNA interactions towards the N-terminal region and a shallow basic cleft towards its C-terminal region. Previous studies had mapped the DBD of XPA from aa98-219 [72, 114-116] (**Fig. 1.13 and Fig. 1.14**), but have now been redefined between aa98-239 as the previous region lacked basic residues to bind properly to DNA (**Figure 1.15**) [58, 124]. This was observed using chemical shift perturbation (CSP) assays where Lys residues- K168, K179, K221, and K222, lying in the redefined DBD

region has a stronger affinity towards DNA than XPA aa98-219 [58]. It has been noted that most of the XP phenotypes due to XPA mutation are mainly accorded to the DBD region of XPA [57, 58, 72, 114-117]. 3D structure of XPA has so far been determined by various groups (PDB ID: 1XPA [115], 1D4U [116]), 6J44 [118], 7AD8 [120]), PDB ID: 6R04 [117], and 6LAE [119]).

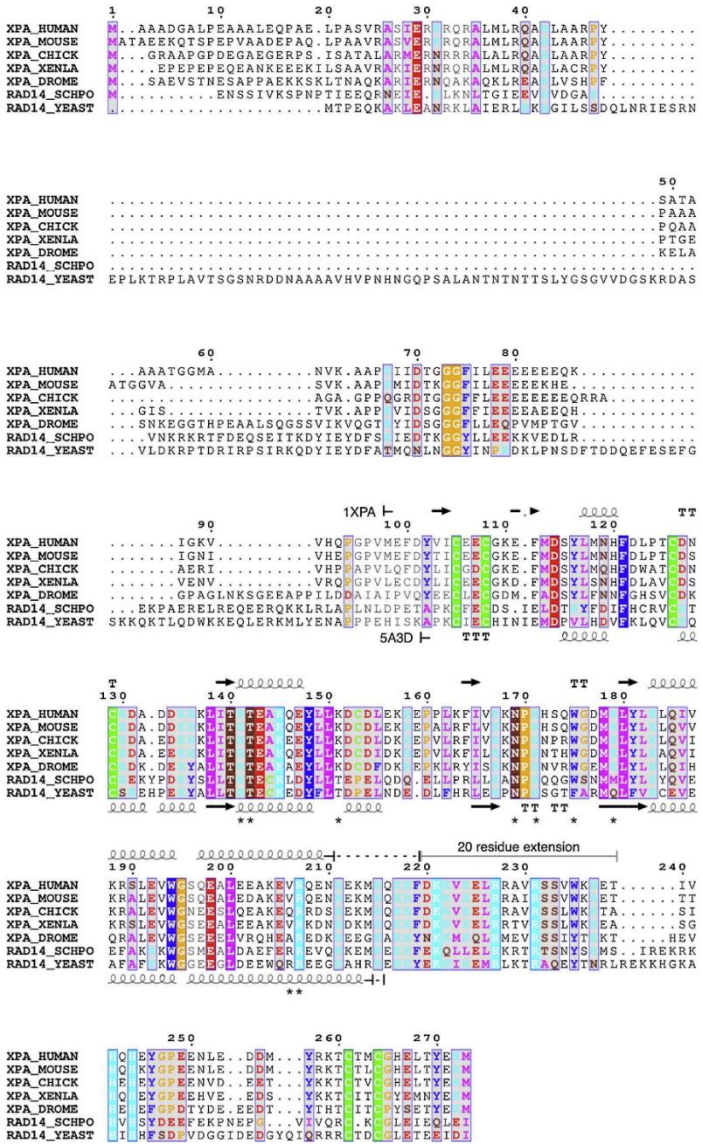

Figure 1.11. The protein sequences of XPA across seven diverse species. Taken from [72].

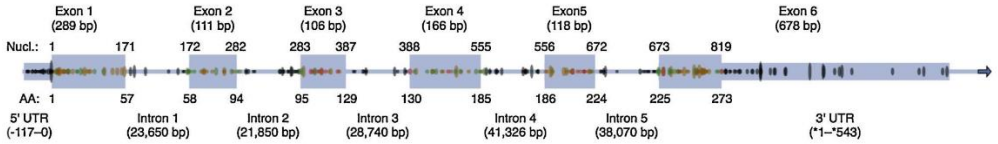

Figure 1.12. Structure and mutations of the XPA gene. Taken from [72].

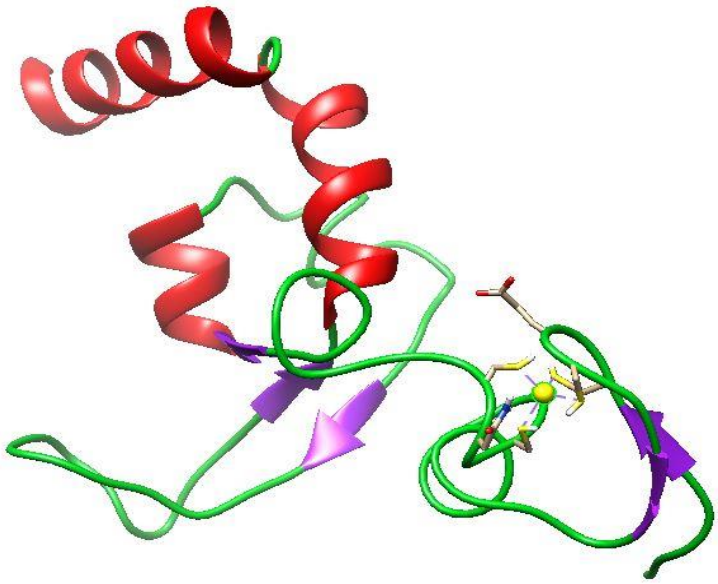

Figure 1.13. NMR structure of XPA (aa98-219). Taken from [115].

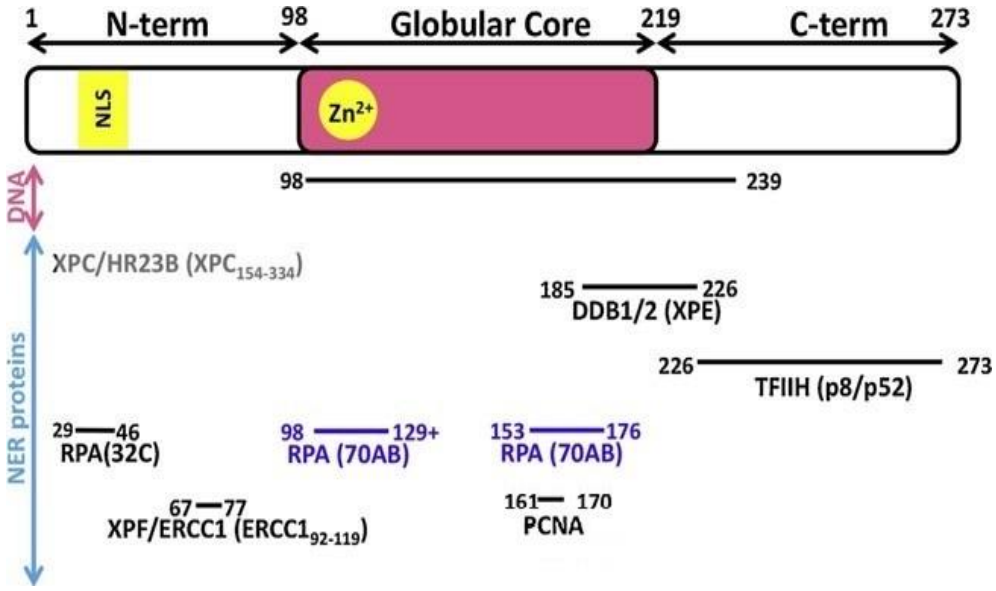

Figure 1.14. XPA domain and its NER interacting members. Taken from [72].

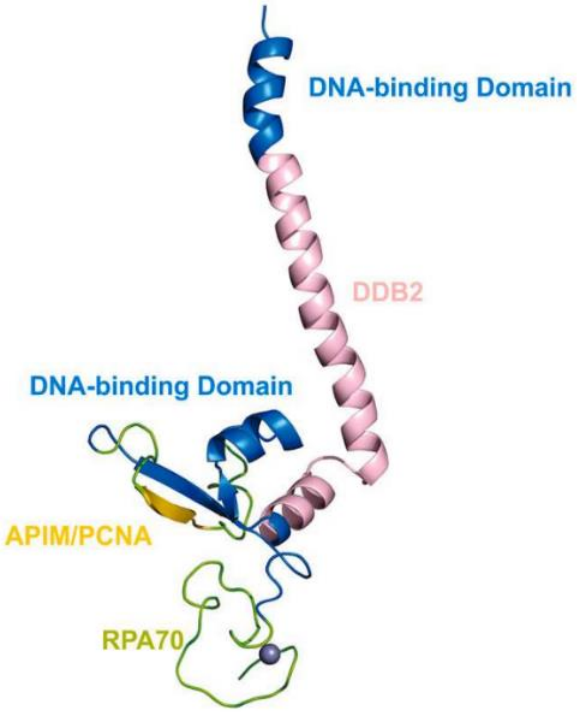

Figure 1.15. Redefined DBD of XPA. Taken from [117].

1.5.3 Positioning of XPA with DNA in the NER bubble

Zn-finger motif is responsible for binding to ssDNA/dsDNA or Y junction equally with ease and with RPA70AB protein. XPA binds to DNA via recruitment of the p8 and p52 subunit of TFIIH. Whether it binds in 3′->5′ or 5′->3′, is not clearly understood [118-120]. One of the pieces of evidence suggests that XPA binds to a 3′ junction like XPC to the DNA duplex. This is supported by the study where RPA binds to ssDNA in binds to 5′->3′ [121]. XPA is also known to interact with ssDNA with RPA70AB at 5′ while XPA at 5′ direction, and had a 2.5-fold faster reaction to the cisplatin-lesioned DNA in the XPA-RPA complex (**Figure 1.16**) [122, 123]. PPI study conducted using the ClusPro online server [124] showed the same results, which agreed with the previous studies [125-127]. Another *in vitro* study showed that XPA interacts with RPA, DNA duplex, and XPA/ERCC1 endonuclease at the 5′ site, suggesting that XPA may bind to the 5′ site of the NER bubble and damaged site [75]. NMR chemical shift perturbation (CSP) shift showed that XPA was bound to 9nt ssDNA using aa98-219 [128], but the 2014 study showed that XPA binds to DNA better using aa98-239 [58, 120], which was shown by the prediction of the fourth helix.

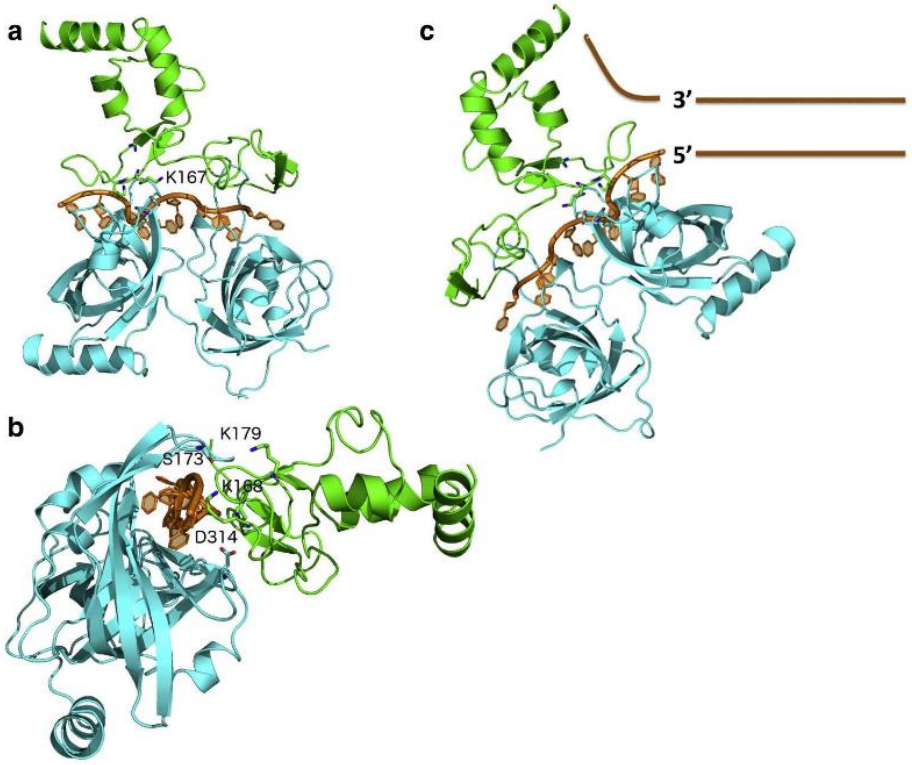

Figure 1.6. (a) The highest-scoring pose for the XPA98–219–ssDNA-RPA70 complex was predicted by the ClusPro online tool [124]. The XPA (PDB ID: 1XPA) is shown in green, the RPA70 in complex with the ssDNA strand (PDB ID:1JMC) are shown in cyan and orange, respectively; (b)) Side view into the three-body XPA98–219–ssDNA-RPA70 complex showing RPA70 Ser 173 and Asp 314 as a possible contact for XPA Lys 179 and Lys 167, respectively and (c) Y junction of the NER pre-incision complex is a three-body XPA98-219-ssDNA-RPA70 complex. Taken from [57].

Most of the studies showed that XPA binds to DNA in monomer [128, 129] form however the XPA homolog in yeast Rad14 (PDB ID: 5A3D, 5A39) revealed that Rad14 existed as a dimer and had bound to lesion site at either side (**Figure 1.17**) while kinking the DNA at 70° [129]. This theory that XPA indeed exists as a dimer was later proved by Yang et al., using fluorescence spectroscopic analysis, perfluoro-octanoic acid-polyacrylamide gel electrophoresis (PFO-PAGE), and native gel filtration chromatography [130]. It was later revealed that XPA binds to DNA in a 2:1 ratio [131], and binds to DNA in both monomeric dimeric forms 132] and was observed to have formed PPI with PCNA in dimer form [83].

Even though many studies have been done concerning the DBD of XPA, still there arises a question as to how XPA interacts with the DNA in monomeric and dimeric forms?

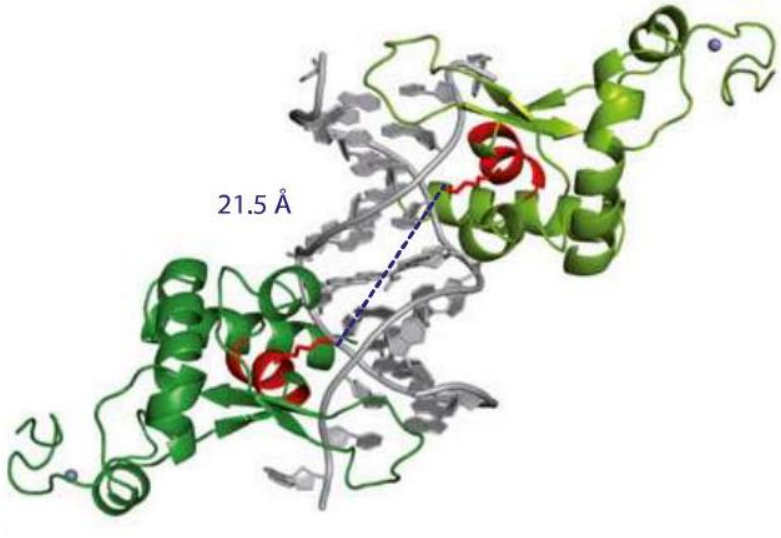

Figure 1.17. Two Rad14 proteins bind to the DNA strand containing the lesion. Taken from [129].

1.5.4 XPA and its interactome with other NER member proteins

Brosey et al. [133] proposed a model showing XPA interacting with fellow NER members of PIC (**Figure 1.18**). This model was created using Rad14 as a homology model for XPA-DNA interaction with other fellow members of NER, where XPA was positioned at the 3′ Y junction of DNA [72, 134].

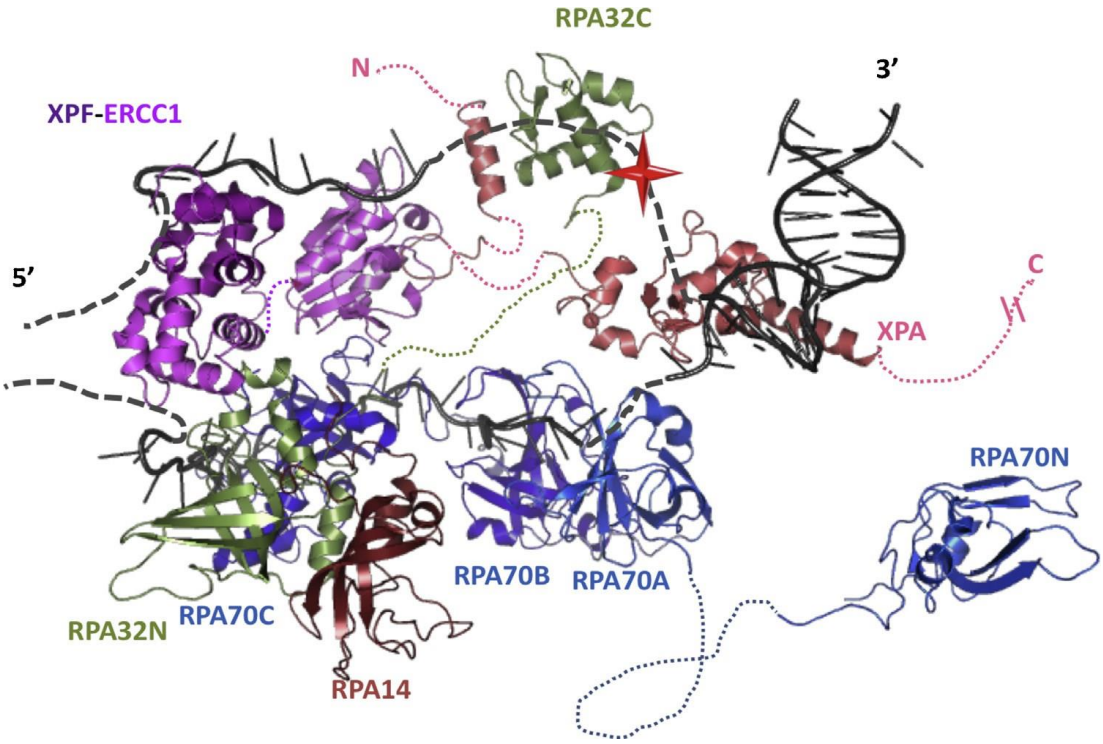

Figure 1.18. Proposed Model of some XPA interactions with NER incision complexes. Taken from [133].

1.5.4.1 PPIs of XPA with NER members

XPA interacts with different sets of protein during each step of NER, which can be seen in **Figures 1.13** and **1.18** [121]. These PPIs take place right from the DNA lesion recognition step till the gap-filling stage of NER. These interactions have been discussed in detail below.

1.5.4.1.1 XPC

XPC functions as a DNA damage detector in GG-NER and is present in a form of a heterotrimer that comprises XPA/HHR23B/CEN2 protein. XPC's interaction with TFIIH (subunits p62 and XPB) initiates the entry point of XPA to the DNA damage site [72, 123]. Using pull-down assay, Bunick et al. and team [134] demonstrated the interaction of XPA with aa154-334 residues of XPC (**Figures 1.14** and **1.19**).

1.5.4.1.2 TFIIH

TFIIH comprises 10 subunits as discussed above in step (ii) of the NER pathway. XPA is called onto the NER damage site by p8 and p52 subunits of TFIIH (**Figures 1.14** and **1.19)** and binds to ssDNA/dsDNA or Y junction of DNA [122, 123, 135]. It has also been seen that XPA helps TFIIH in unwinding the DNA helix so that it can accommodate fellow NER core proteins. XPA further helps CAK dissociate from TFIIH, triggering the dual incision step [136]. However new pieces of evidence show that (i) Pro160, Arg227, and Thr239 form PPI with Gly402, Ala572, and Lys529 of XPB; (II) Pro66, and His92 forms PPI with Arg373, and Pro170 of XPA forms PPI with Leu637 of XPD; (iii) Thr239 of XPA forms PPI with Ala1007 of XPG [121]. Here, the interacting residues of XPA and TFIIH are not known still.

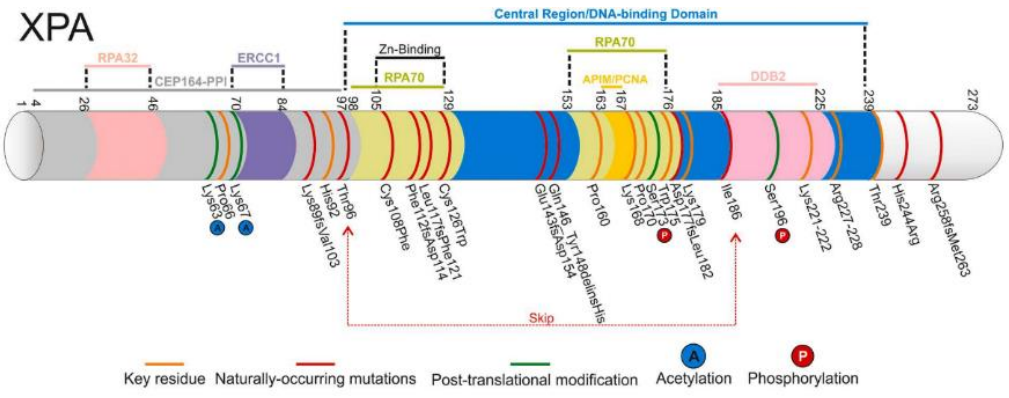

Figure 1.19. Model of XPA depicting known interacting domains, key residues, mutations, and post-translational modifications. Taken from [4].

1.5.4.1.3 RPA

RPA is an ssDNA binding protein for all eukaryotic organisms. The primary function of RPA is to protect the undamaged DNA strand and acts as a transition between dual incision and synthesis of the new strand in NER. RPA interacts with XPA using its 2 subunits- RPA32C and RPA70AB. RPA32C binds to aa29-46 residues of XPA [134], while RPA70AB interacts somewhere aa98-239 residues of XPA at the Zn-finger motif (**Figures 1.14** and **1.19**) [115, 116, 136], but the exact location is not known so far. According to biochemical pull-down assay Lys-residues, K141 and K179 of XPA have been reported to be involved in PPI with RPA70AB but not in binding to the DNA [127, 137]. A study reported that XPA and RPA participate in recognizing triple-helix DNA distortions [138]. Many studies showed that XPA with RPA recognizes CPD, and then aids in the dual incision of distorted DNA lesions [139-143]. Disorientation of RPA32C and RPA70AB from the NER process is implicated in the poor functioning of NER, thereby inhibiting the excision of damaged strands [127].

1.5.4.1.4 XPF/ERCC1

These are structure-specific nucleases where they bind to either for dual incision of the lesion DNA. The PPI between ERCC1 and XPA (**Figures 1.14** and **1.19**) is modulated via aa96-114 of XPA and aa92-119 of ERCC1 [144, 146].

1.5.4.1.5 DDB2/XPE

This protein exists as a heterodimer unit DDB1 and is exclusively involved in the damage recognition step of GG-NER. Mild XP is observed in the case of XPE mutations [72, 147, 148]. DDB1/XPE creates a kink in the lesion which is then recognized by XPC [149]. XPA residues aa185-226 have been known to interact with XPE during the NER process (**Figures 1.4** and **1.19**). It was further found that R207 of XPA is vital for damage verification, and if mutated to R207G, it can cause disassociation of PPC between XPA and XPE, thereby inhibiting the recognition and healing of the damaged strand [150]. R207Q mutation in XPA is often implicated in the occurrence of cancer. Even though Wakasugi et al. studied the PPI between XPA and XPE [150], the exact location by which XPE interacts with XPA is not elucidated yet.

1.5.4.1.6 PCNA

The involvement of PCNA is usually seen gap-filling stage of NER, which is after the disassociation of the PIC members for the NER bubble [150]. However, Gilljam et al saw the PCNA binding sequence in XPA, APIM (AlkB homolog 2 PCNA interacting motif) (**Figures 1.14** and **1.19**), and together they co-localize to form foci in the damaged DNA [83]. However, this creates another debate about whether XPA is involved ed in the gap-filling stage of NER or not?

1.5.4.2 Sidekicks of XPA not involved in NER directly

1.5.4.2.1 Ataxia telangiectasia and Rad-related (ATR) protein

This protein is a serine/threonine-protein kinase, also known as the FPAP-related protein 1 (FRP1), acting as one of the associative proteins in DNA damage response in NER [152, 153]. Phosphorylation of serine196 in XPA that lies in between the DBD region of XPA is regulated by ATR, which increases the probability of the presence of XPA at the NER site. This phosphorylation stabilizes XPA by inhibiting the disruption and ubiquitination by E3 ubiquitin ligase HER2 protein in NER, therefore ensuring the smooth functioning of NER [72, 154].

1.5.4.2.2 Newly identified XPA binding protein

In the year 2000, five XPA binding proteins- XAB1, XAB2, XAB3, XAB4, and XAB5 were identified using the yeast hybrid-screening method by Nitta et al. [155], out of which XAB1, and XAB2 were novel proteins, and further validated by comparing them with previously XPA binding proteins like ERCC1, and RPA. XAB1 functions as a GTPase interacting aa30-34 residues of XPA and houses the nuclear localization signal (NLS) motif [156], which suggests that it may aid in the localization of XPA in NER. Although ATR is a known NLS protein of XPA, so whether they both equally function in NER is debatable. XAB2, on the other hand, comprises 15 tetracopeptide repeats (15 TPR), generally seen during TC-NER, and when unregulated caused embryonic damage in mice. Further, it acts and forms PPIs with TC-NER proteins [157,158].

XAB3 consists of PRSM1, a metallopeptidase protein while XAB4 and XAB5 are homologs of each other containing GRASP65, which reassemble and stack the Golgi complex [155]. The relation of XAB3, XAB4, and XAB5 proteins

with XPA in NER is currently not understood well enough and may require further study.

1.5.4.2.3 PAR polymerase 1 protein (PARP-1)

Post-translational modification protein, Poly(ADP-ribosyl)ation (PARylation), when PARylated by PARP-1 functions as a repair machine in NER as well in ss/dsDSB. C-terminal residues aa213-217 of XPA have a conserved motif for PARylation, which when PARylated, induces the decrease of XPA's role in NER, also inhibiting PIC formation as per the cell-based imaging study conducted by Fischer et al. in 2014 [159].

1.5.5 The complacency of XPA and disease outcomes

It has been noted that most severe cases of XP due to XPA mutations have been mapped to the DBD of XPA. The severity of XP phenotype due to XPA mutation gives rise to detrimental conditions like neurological damages and carcinomas, while in rare cases, the symptoms are mild [160]. It has been observed that even partial loss of the XPA can result in NER complications [54, 161-165] Most of the phenotypes of XPA mutations have been shown in **Tables 1.8** and **1.9**. It is seen that the mutation Zn-finger motif gives rise to an irregular folding of the protein leading to the non-functional form of XPA, especially the mutation of four cysteine residues [166]. Omission of exons 1 disturbed the PPIs between RPA32C and ERCC1, while the C-terminal truncation resulted in mild interference of NER (**Figures 1.14**) [167, 168]. DBD mutations of XPA (exons 2-5) have been known to cause the loss of NER functioning as a whole. These exons code for PPIs with RPA70AB [115, 116, 127 136], PCNA [83, 151], XPF/ERCC1 [144, 146], XPE [72, 147, 148, 150] and damaged DNA (**Figures 1.13** and **1.18**).

Sugitani et al. [72] listed all the disease phenotypes of XPA mutants as seen in Table 2.7, in which they determined all the missense, intron variants, and loss of functions (LOF) of XPA using exome aggregation consortium (ExAC) datasets from 60,706 independent individuals of various ethnicities and ancestries. They obtained ~ 21% density of variants in exon 6 [72]. Gao et al. further identified 56 somatic mutations in 121 cancer patients having mutated XPA [169]. Sehgal and Singh [170] identified 191 deleterious nsSNPs of XPA using a computational approach.

XPA sees novel mutations every year. A frameshift mutation Asp177fsSer182 was observed in a Japanese patient with SCC, but presented no neurological damage [171], while another frameshift deletion of exon 3 (c.349_353delCTTAT) Leu117fsPhe121 saw a patient with severe microcephaly and mental retardation but no presence skin tumors [172]. BCC and neurological impairments were seen in two novel mutations, c.772_785delCGTAAGACTTGTAC (Arg258fsMet263), and c.438_443delAGAATA (Gln146_Tyr148delinsHis), with one rare case Arg258fsMet263, which showed no such damages. The PPI of XPA with DDB2, PCNA, and RPA70 was reported to have been disturbed by Gln146_Tyr148delinsHis and Leu117fsPhe121, while Arg258fsMet263, caused misfolding of XPA [173]. Truncation of XPA was observed to have caused SCC and cognitive impairments, which had resulted from skipping of exons 3 and 4 [174] Three frameshift and one missense mutation were reported from thirteen Indian people causing neurological damages [175].

Most of the XPA disease phenotypes have been reported to have emerged from the mutations of DBD of XPA [72, 121, 171-173, 175-178]. rs1800975 A/G polymorphism of XPA in 71 control studies was observed to have susceptibility towards developing colorectal, skin and lung cancer [179], digestive system cancer [180, 181], breast cancer [182], oral squamous cell carcinoma (OSCC) [183], hepatocellular cancer (HCC) [184], ovarian cancer [185], prostate cancer [186-188], pre-eclampsia [190-192], premature aging [193-195, neuroblastoma [196] accelerated aging and Alzheimer's disease [197], and various kinds of SNPs' effects (**Table 1.10**) [188].

Table 1.8. XPA mutations and their effects Taken from [72].

Mutation in gene	Mutation in protein	Mutation type	Disease phenotype	Possible Effects on XPA Function
171 + 2T > G	NA	splice site	XP-A; severe	Disrupts 5′ splice donor site of intron 1
268_269insAA	variant1: V9EfsX15, variant2: V9EfsX6 P96-Q185del		XP-A; severe form	
281C > T	P94L	missense	Severe XP; neurological disease or disruption of function	
323G > T	C108F	missense	XP-A; severe form	zinc finger disruption
331G > T	E111X	nonsense	XP-A; severe form	
348T > A	Y116X	nonsense	Severe XP neurological disease or disruption of function	
349_353 delCTTAT	L117EfsX4	deletion/frameshift	XP-A; severe form	
374delC	T125IfsX15	deletion/frameshift	XP-A; severe form	
377C > T	C126T	missense	XP-A	
387-1G > A	NA	splice site	XP-A	Disrupts 5′ splice donor site of intron 3
388-12A > G	NA	splice site	XP-A	Disrupts 3′ splice acceptor site of intron 3
388-2A > G	NA	splice site	XP-A	Disrupts 3′ splice acceptor site of intron 3
388-1G > C	NA	splice site	XP-A; severe form	Disrupts 3′ splice acceptor site of intron 3
388-1G > T	NA	splice site	XP-A	Disrupts 3′ splice acceptor site of intron 3
545_546insTA	L182Ffs		XP-A	
553C > T	Q185X	nonsense	XP-A; severe form	
555G > C,T	Q185H	missense	XP-A	
555-1G > C	NA	splice site	XP-A	Disrupts 5′ splice donor site of intron 4
555 + 8A > G	NA	splice site	XP-A	Disrupts 5′ splice donor site of intron 4
619C > T	R207X	nonsense	XP-A, neurological impairment and mild skin abnormality	
620G > A	R207Q	missense		inhibition of XPE binding
622C > T	Q208X	nonsense	XP-A	
631C > T	R211X	nonsense	Severe XP neurological disease or disruption of function	
647_648delAG	K217EfsX3	deletion/frameshift	XP-A; severe form	
672-1G > C	NA	splice site	XP-A	Disrupts 5′ splice donor site of intron 5
682C > T	R228X	nonsense	Mild XP neurological disease or partial function	
683G > A	R228G	missense	Improved Adduct Removal	
690insT	R231KfsX15		XP-A; mild form	
700G > T	V234L	missense	Improved Adduct Removal	
731A > G	H244R	missense	XP-A; mild form	
779_780 insTT, 780_781 insTT	T260IfsX9		XP-A; mild form	

Table 1.9. Mutations of XPA and their effects. Taken from [121].

Mutation/Res/ Domain	Identity	Located region	Disease-Country
Lys89fsVal103	Frameshift (Insertion)	CEP164 PPI.	XP-A, Germany
96–185del	In-Frame (Skipping)	DNA-binding domain and Zn-binding domain.	XP-A, Germany
Phe112fsAsp114	Frameshift (Del_Ins)	DNA-binding domain and Zn-binding domain.	XP-A, India
Leu117fsPhe121	Frameshift (Deletion)	DNA-binding domain, RPA70, and Zn-BD	XP-A, Iran
Cys126Trp	Frameshift (Deletion)	DNA-binding domain.	XP-A, India
Glu143fsAsp154	Frameshift (Deletion)	DNA-binding domain.	XP-A, India
Gln146_Tyr148delinsHis	In-Frame (Deletion)	DNA-binding domain.	XP-A, Japan
Asp177fsLeu182	Frameshift (Splicing)	DNA-binding domain.	XP-A, Hungary
Arg258fsMet263	Frameshift (Deletion)	None in particular.	XP-A, Hungary
XPA_{26-46}	Domain	RPA32 and CEP64 PPIs	NA
XPA_{98-126} and $XPA_{153-176}$	Domains	RPA70 (DNA-binding/ Zn-Binding domains) and RPA70 (DNA-binding/ APIM/ PCNA PPIs).	NA
Lys168, Trp175, Lys179, Lys221, Lys222 and Arg228	Interacts with DNA	DNA-binding domain and DDB2 PPI.	NA
Pro160, Arg227, and Thr239	Protein Interaction	XPB PPI.	NA
Pro66 and His92, and Pro170	Protein Interaction	XPD PPI.	NA
Thr239	Protein Interaction	XPG PPI.	NA
Cys108Phe * *	Missense	Zn-binding, RPA70.	XP-A, Japan
His244Arg * *	Missense	None in particular.	XP-A, Japan

Table 1.10. SNPs of XPA with respect to its cancer outcomes. Taken from [188].

SNP ID	Location	Allelic Variant	Effect	Association with Cancer Risk	Response to Therapy
rs2808668	5′-UTR	T/C	Binding of transcription factors	No association with cancer risk within overall analysis;	NA
				Decreased cancer risk with the exception of digestive system cancer in subgroup analysis;	
				No association with OSCC risk and/or prognosis	
rs10817938	5′-UTR	T/C	Binding of transcription factors; Decreased transcription of the XPA gene	Homozygous CC genotype, C allele, and CC/CT genotype in dominant setting associates with an increased cancer risk within overall analysis;	CT and TT genotypes have longer OS in CRC patients receiving oxaliplatin-based chemotherapy
				TC and CC genotypes display higher risk of developing OSCC compared to the TT genotype;	
				It associates with HCC risk in stage 1, where the CC genotype displays an increased risk of HCC compared with the TT wild-type and TT plus TC genotype;	
				It contributes to an increased CRC risk in its variant homozygote and recessive model both in overall and stratification analyses	
rs1800975	5′-UTR	A/G	Binding of 40S ribosomal subunit	No association with BC risk in the pooled analysis for all genetic settings;	No association with chemotherapy efficacy and prognosis in EC;
				In subgroup analysis, it decreases BC risk in some ethnic groups;	Homozygous GG genotype shows a higher response rate than the GA or AA genotype in LC;
				GG genotype shows an increased LC risk in some ethnic groups;	The GA and AA genotype has an increased risk of death in inoperable LC treated with radiotherapy with or without platinum-based chemotherapy;
				When combined with rs3176752, it increases neuroblastoma risk;	It plays an important role in response to radiotherapy in HNSCC;
				It contributes to a risk from basal and SCC, oral SCC, and OC;	The AG genotype imposes with a higher risk of mortality after cancer treatment compared with the GG genotype;
				AG and GG genotypes significantly decrease the ESCC risk compared to AA genotype;	No association with OS or disease progression regarding clinical outcome to 5-fluorouracil/oxaliplatin combination therapy in refractory CRC
				No association with risk of testicular, prostate, and gastric cancers, CRC, SCC of the oropharynx, and melanoma	
rs3176658	Intron	C/T	-	Modest association with LC risk	Significantly associates with PFS in LC;
					Significantly associates with the response to neoadjuvant radiochemotherapy treatment of locally advanced rectal cancer
rs3176721	Intron	C/A	-	NA	Significantly associates with toxicity and efficiency of platinum-based chemotherapy in LC
rs2808667	Intron	T/C	-	Association with risk of EC	NA
-	Intron	G709A	-	A significant protective effect in AG heterozygotes in LC	
rs3176752	3′-UTR	G/T	Binding of microRNA	When combined with rs1800975, it increases neuroblastoma risk	NA

1.5.6 Therapeutic aspects of XPA in cisplatin-resistant cancer patients

It has been noticed that higher expression of NER protein, especially harms the chemotherapy treatment in cancer patients, by removing the cisplatin adducts. Higher levels of XPA during NER lead to chemoresistance in lung cancer, ovarian cancer, head and neck cancer, gastric cancer, prostate cancer, glioblastoma, and neuroblastoma [198-212]. Hence, the adjuvant therapy for platinum-based chemotherapeutic resistance is now being researched so that cancer patients are relieved from their relapses [72]. So far, many studies have been conducted to delay or disrupt NER in cancer treatments [213-215].

UCN-01, an anticancer drug used during stage-I cancer treatment was reported to have inhibited NER by preventing PPI between XPA and ERCC1. inhibits DNA repair: association with attenuation of the interaction of XPA and ERCC1 [215]. NER inhibitor 01; NERI01 (AB-00026258) was reported to have caused the failure of forming PPC between XPA and ERCC1 [216]. This inhibitor formed a hydrophobic cleft with the interacting residues of ERRC1, which is responsible for PPI with XPA [217]. Another potent inhibitor identified for the same was AB-00027849 [218].

Researchers have also attempted to design small molecule inhibitors (SMIs) against XPA to increase the efficacy of anticancer drugs in patients [219-221]. Neher et al. identified novel small molecule inhibitors (SMI)- X57, X60, and X80 using virtual screening [220], where they observed the DPI between XPA and DNA being disrupted. They observed that out of three SMIs, X80 showed a 95% efficacy rate in inhibiting XPA-DNA interaction, by forming a salt bridge with K137 of XPA, which is essential for DNA binding. They also designed SMIs for disrupting PPIs between XPC/HHR23B and cisplatin cross-linked DNA by positioning exo-N-{2-[N-(4-azido-2,5-difluoro-3-chloropyridin-6-yl)-3-aminopropionyl]aminoethyl}-2′-deoxycytidine 5′-monophosphate)FAP-dCMP at 3′ side of the adduct instead of the 3′ side [221]. Gavande et al. designed 24 amide-based inhibitors using structure-guided drug design and docked them with XPA (PDB ID: 1XPA) using the Glide SP protocol. They found that analogs of compound X80 (3′-COOH, 4′-C) had the highest specificity with an IC_{50} value of 0.82 ± 0.18 μM $_{for}$ disruption of PPI between XPA and RPA, and DPI between XPA and DNA [222]. The mutant variant of XPA-A23G showed promising results for decreasing PT-based resistance in non-small cell

lung cancer (NSCLC) [223]. Synthesis of anti-XPA IgG monoclonal antibodies was shown to inhibit the scaffold property of XPA in NER [224]. The gene silencing XPA by enhancer of zeste homolog 2 (EZH2) was shown to have decreased resistance toward anticancer drugs in patients with nasopharyngeal carcinomas [225].

1.6 Scope of our present work

Although many studies have been conducted so far, some gaps are still persistent in the knowledge of XPA.

I. What is the structural basis of XPA interaction with DNA?
II. How does this lead to their positioning within the complex (NER DPC/PPC complexes)?
III. How does XPA fit and function within the context of multi-protein NER complexes processing the NER bubble?

By understanding these problems, we have attempted to answer a few unresolved queries: -

a) Determined a better homodimer model for DBD of XPA in **chapter 3**.
b) In **chapter 4A,** we studied the structural dynamics and interactions of XPA with the damaged DNA.
c) We studied the binding mechanism of XPA homodimer with the DNA in **chapter 4B.**
d) We characterized the protein-protein interactions between the DNA binding proteins, XPA and XPE in **chapter 5A**.
e) In **chapter 5B**, we studied the effect of XPA's R207G mutation on its binding affinity with XPE in a dynamic system.
f) We determined the full-length structure of XPA, and studied the PPI/PPC of XPA with fellow members of PIC involved in NER in **chapter 6**. Both PPIs/PPCs studies were conducted taking XPA in monomer and dimer states.

1.7 Main objectives of this thesis:

Objective 1: To study the salient structural features of the DNA binding domain (DBD) of XPA and the determination of a better homodimer model for the DNA binding domain of XPA

Objective 2: To study structural dynamics and molecular interactions of XPA with the DNA

Objective 3: Investigating the role of XPE in modulating the functioning of XPA

Objective 4: To study the protein-protein interaction between the members of the pre-incision complex (PIC)