XPA: A key scaffold for human nucleotide excision repair

Norie Sugitani, Robert M. Sivley, Kelly E. Perry, John A. Capra, and Walter J. Chazin
Departments of Biochemistry, Biological Sciences, Biomedical Informatics, Chemistry, and Computer Science, and Vanderbilt Genetics Institute and Center for Structural Biology, Vanderbilt University, Nashville, TN 37232-7917

Abstract

Nucleotide excision repair (NER) is essential for removing many types of DNA lesions from the genome, yet the mechanisms of NER in humans remain poorly understood. This review summarizes our current understanding of the structure, biochemistry, interaction partners, mechanisms, and disease-associated mutations of one of the critical NER proteins, XPA.

Keywords

DNA repair; NER; XPA; Xeroderma pigmentosum

1. Introduction

Nucleotide excision repair (NER) is the primary pathway for the repair of a wide range of bulky DNA adducts, such as those formed by UV irradiation, environmental toxins and certain antitumor agents. The protein XPA is believed to play a key role as a scaffold that organizes the damaged DNA and other proteins to ensure lesions are appropriately excised. Defects in NER can result in the genetic disorder Xeroderma pigmentosum (XP). XP is characterized by extreme sensitivity to sunlight and very high rates of skin cancer, with the most severe cases displaying neurological degeneration with loss of mental and sensory faculties. The association of XPA mutants with the most severe clinical XP symptoms underscores the critical role of this protein in NER.

Substantial progress has been made in elucidating the mechanisms of NER in prokaryotes, but understanding of human NER has lagged behind due to the lack of conservation of proteins and complex regulation of the ~30 proteins involved. NER occurs in coordination with transcription (transcription coupled (TC) NER) and more generally throughout the genome (global genome (GG) NER). Once the presence of damage is recognized, a series of...
protein factors are recruited to verify the presence of damage, cleave the damaged nucleotide 5′ and 3′ of the lesion, fill in the gap using the undamaged strand as template, and seal the resulting gap.

XPA is involved in both TC-NER and GG-NER; the other proteins involved and their roles in these two sub-pathways are described elsewhere. In both pathways, XPA is recruited to the damage site by the transcription factor II H (TFIIH) complex that is responsible for unwinding double-stranded DNA around the damaged nucleotide creating the NER bubble. XPA is generally understood to function in damage-verification and assembly of NER incision complexes. XPA is recruited at the same time, and functions in coordination with, the eukaryotic ssDNA binding protein replication protein A (RPA). Together, they help recruit and properly position the excision nucleases. RPA binds to the undamaged single strand, suggesting that XPA interacts with the damaged strand. However, XPA prefers to bind ss-dsDNA junctions and duplexes with overhangs. Although XPA has been studied for >20 years, several key questions remain about its function, including: 1) What is the structural basis of XPA interaction with protein binding partners and how does this lead to their positioning within the complex? 2) Is XPA involved in pathways other than NER? 3) How do different XPA mutations relate to NER outcomes and disease phenotypes? In the following sections, we highlight current knowledge of the interactions of human XPA with DNA, other NER proteins and proteins outside of NER, and the relationship between XPA mutants and XP disorders. In the last section, we discuss future directions for XPA studies that can enrich our understanding of NER and XP disorders.

2. XPA structure and interactions with DNA

XPA is a modular protein whose primary function is mediated through its interaction with the NER bubble. The DNA binding apparatus of XPA has been mapped to its globular central domain, but the molecular details of how XPA is engaged on the NER bubble have yet to be established. Mutations in the DNA binding region are associated with the most severe symptoms of XP patients, including accelerated aging and neurodegeneration, suggesting the importance of XPA-DNA interaction. However, since some protein interactions also map to this region, understanding the molecular basis of the malfunctions of disease-associated mutations in this region requires a more complete understanding of XPA-DNA interactions in the context of its protein interactions in NER complexes.

2.1 XPA structure

XPA is a relatively small 273 residue protein that does not possess enzymatic activity but interacts with many other NER proteins, consistent with its role as a scaffold. A domain map of XPA is shown in Figure 1. XPA is organized around a central globular domain (XPA98-219). 3D structures of this domain were determined independently by two groups using solution NMR (PDB: 1XPA, 1D4U) (Figure 2). XPA98-219 contains a C4 type zinc-finger motif in the N-terminal region and a shallow basic cleft in the C-terminal region (Figure 2). The N- and C- termini of XPA are disordered and mediate a variety of protein interactions. Interestingly, severe XP symptoms associated with XPA mutations map primarily to the central domain.
2.2 Localization of XPA on the NER bubble

XPA was shown to bind ssDNA-dsDNA junctions more strongly than ssDNA or duplex alone, suggesting that this protein is likely to be located at one end of the NER bubble rather than strictly associated with the damaged (or undamaged) single strand. Whether XPA binds to the junction 5′ or 3′ to the lesion remains unclear as evidence has accumulated supporting both models. XPA is recruited to NER complexes via interactions of its flexible C-terminus with both the p8 and p52 subunits of TFIIH (Figure 1). However, it is difficult to model how XPA is positioned in the NER bubble based on XPA-TFIIH interactions alone due to the lack of knowledge of the orientation of p8 and p52 within the TFIIH complex. The reported interaction of XPA with XPC suggests XPA localization at the 3′ junction as XPC binds to the duplex 3′ to the lesion. This model is also supported by the interaction with RPA. It is well established that RPA binds ssDNA in a 5′-3′ orientation, which matches the direction of the undamaged strand in NER bubble. As noted below, XPA interacts with the tandem high affinity ssDNA binding domains RPA70AB, which are positioned 5′ on the undamaged strand (3′ to the lesion). Support for the opposite model is based on XPA interactions with the 5′ incision nuclease XPF/ERCC1, assuming that in order to recruit and localize XPF/ERCC1 to the 5′ side of the lesion, XPA should also be located 5′ to the lesion. In vitro studies using isolated XPA, RPA and damage containing DNA also support XPA localization 5′ of the lesion in both a duplex and a model bubble. In summary, although most models place XPA 5′ to the lesion, there is conflicting evidence and the controversy over the location of XPA within NER complexes is clearly not settled. One critical issue that has not been considered is that these models are based on viewing the complexes as linear 2-dimensional arrays. In fact, consideration of the 3D topology of the NER bubble and proteins bound to it may allow XPA to be bound to DNA 3′ to the lesion yet still position XPF/ERCC1 to cleave 5′ of the lesion. Clearly, there is a great need for determining the structure of functional NER complexes to truly understand where XPA is bound.

2.3 Structural analysis of XPA bound to damaged DNA

The quest for structurally characterizing how XPA binds to the NER bubble started approximately twenty years ago. Based on the combination of limited proteolysis and filter binding assays, the central globular region of the protein (residues 98-219) was proposed to serve as the DNA binding domain. After determining the NMR solution structure, NMR chemical shift perturbations induced by a 9-nt ssDNA oligomer were used to map the DNA binding site onto the XPA98-219 structure and generate a model of the complex. This study suggested that the C-terminal basic cleft is the site of DNA binding. However, because ssDNA is not a high affinity substrate, questions remain about the accuracy of this model for the interaction of XPA with the NER bubble. In fact, in 2014, we and others showed that in order to bind a junction DNA substrate as does the full-length protein, the globular XPA98-219 core must be extended C-terminally by ~20 residues.

In 2015, Kisker, Carrell and co-workers reported X-ray crystal structures at 1.8-2.8 Å resolution for the S. cerevisiae XPA homolog Rad14 in complex with damage containing DNA (Figure 3). These were the first high-resolution 3D structures of an XPA homolog in complex with DNA: one was with duplex DNA containing cisplatin that forms a 1,2-GG
intra-strand crosslink (PDB: 5A39) and the second was with the same duplex containing a \textit{N}(deoxyguanosin-8-yl)-2-acetylaminofluorene (AAF) (PDB: 5A3D) adduct (Figure 3). Notably, the two structures are nearly identical except for the differences in the lesions, as reflected in the RMSD over all protein atoms of only 0.22 Å. The two key findings from the Rad14 structures are (i) two molecules of Rad14 bind to each side of the lesion-containing duplex, and (ii) the duplexes are kinked by 70° (Figure 3)\textsuperscript{15}. The interaction of Rad14 with the ss-dsDNA junction as observed in these structure is consistent with previous studies indicating that human XPA also preferentially binds to junction DNA\textsuperscript{26}. Also, these structures support the idea that XPA does not make direct contact with the lesion as suggested previously\textsuperscript{16,46}. It is interesting that Rad14 binds to both damaged duplexes as a dimer, consistent with a report that isolated XPA forms a dimer\textsuperscript{47}. Despite these \textit{in vitro} observations, it is difficult to imagine how an XPA dimer can be fit and function within the context of multi-protein NER complexes processing the bubble.

Comparisons of XPA and Rad14 can help assess if the Rad14 crystal structures adequately represent the interactions of human XPA with DNA in NER. Figure 4 shows a structure-guided alignment of XPA homologs from seven diverse species, and Figure 5 maps the evolutionary conservation of each position in XPA onto the 1XPA structure. The human XPA and \textit{S. cerevisiae} Rad14 constructs used for structural studies are also highlighted on the alignment. The Rad14 construct used for crystallization (Rad14\textsubscript{188-302}) has 2 insertions, 1 deletion, is 4 residues shorter at the N-terminus and 4 residues shorter at the C-terminus, and has 29% identity and 57% conservation to XPA\textsubscript{98-219} (Figure 4). As noted above, XPA\textsubscript{98-219} has severely reduced DNA-binding activity; a 20 residue C-terminal extension is required to reproduce the DNA binding activity of FL XPA\textsuperscript{28}. It is therefore surprising that FL-XPA, FL-Rad14, and Rad14t bind duplex DNA containing cisplatin or AAF lesions very tightly\textsuperscript{45}. Moreover, Rad14 does not bind to duplexes containing other commonly studied DNA lesions (e.g. (6-4)photoproduct ((6-4)PP), cyclobutane pyrimidine dimer (CPD)) with appreciable affinity\textsuperscript{45}. In light of these observations, it would be interesting to know if XPA\textsubscript{98-219} binds these substrates with comparable affinity.

Figure 6 compares the Rad14 structure with the solution NMR structure of XPA\textsubscript{98-219}. Although the Rad14 construct is shorter, a larger number of C-terminal residues were observed in the crystal structure. Moreover, Rad14 has more helical character than XPA; this difference may be due to interaction with DNA or from the characteristics of the crystal lattice. A β-hairpin at the N-terminal zinc finger is observed in XPA but not in the yeast structure; this difference is likely due to the truncation of 4 residues at the N-terminus of the Rad14 construct. Overall, the Rad14t and XPA\textsubscript{98-219} structures are very similar (Figure 6C); the only significant differences are minor shifts in the β-hairpin (between β2 and β3 in Rad14, which correspond to β4 and β5 in XPA) and the most C-terminal helix (α7 in Rad14, α3 in XPA). The striking similarities between Rad14t and XPA\textsubscript{98-219} structures imply that XPA will bind DNA in a manner similar to Rad14 overall. However, because these structures were determined for two very unique damaged duplexes, it is unclear if they adequately represent XPA interaction with DNA within NER complexes that process the full range of NER-repaired lesions,

\textit{DNA Repair (Amst).} Author manuscript; available in PMC 2017 August 01.
3. XPA interaction with other proteins

The interaction of XP proteins with their binding partners was reviewed in 2008. This section provides updated information and additional insights. To provide an overview, the binding sites for various XPA binding partners are mapped on the XPA sequence in Figure 1.

3.1 XPA binding partners in human NER

XPA binds proteins involved in every step of NER, from damage recognition to gap-filling synthesis. These proteins are introduced in the order of their recruitment to the site of damage.

3.1.1 XPC—XPC is a 106 kDa protein responsible for detecting the presence of DNA damage in the GG-NER pathway. XPC functions as a heterotrimer with HR23B and centrin-2, which stimulate XPC DNA binding activity and increases cellular stability. Once engaged on the damaged duplex, XPC recruits the TFIIH complex. As discussed in section 3.1.3 below, XPA is recruited to the damaged site after formation of the NER bubble through an interaction with TFIIH. However, XPC also binds XPA; using a pull-down assay, Bunick et al. showed that XPC N-terminal residues 154-334 are responsible for binding to XPA. There is currently no structure of the complex of XPA and XPC or more detailed mapping of XPC interaction sites on XPA sequence. So the functional role of this interaction has yet to be established, i.e., is XPA-XPC interaction responsible for the recruitment of XPA to the damaged site or guiding XPA to a certain site on the NER bubble? XPA (and RPA) was originally thought to contribute to damage recognition and verification, in part due to its interaction with XPC. However, more recent experiments showed that XPA (in concert with RPA) is recruited to the damaged site after the formation of the NER bubble.

3.1.2 XPE—Damaged DNA-binding protein 2 (DDB2), also named XPE, is another protein involved in damage recognition in GG-NER. XPE exists as heterodimer with DDB1, and together they recognize a wide variety of lesions. Mutations in XPE often result in mild XP disorders. Although the DDB1/2 complex is dispensable for NER reconstituted in vitro, it enhances this activity especially for CPD lesions. The DDB1/2 complex binds to CPD-containing duplexes and creates a kink in the DNA that is recognized by XPC; XPC alone does not directly recognize this lesion. Wakasugi et al. showed that XPA interacts with the DDB2 (XPE) subunit of the XPE/DDB1 dimer and that this interaction is mediated by XPA residues 185–226. They also showed that XPA R207G mutation diminishes XPA-XPE binding, prevents XPA recruitment to the NER bubble, and fails to stimulate CPD removal by NER. In other studies, R207 was reported to be involved in DNA binding, and the R207Q mutation was discovered in cancer patients (Table 1). It is generally accepted that XPA is primarily recruited to the repair site by the TFIIH (see below), so XPA interaction with damage recognition proteins such as XPC and XPE presumably functions to position XPA to specific positions within NER complexes, although the details are yet to be elucidated.

3.1.3 TFIIH—The TFIIH complex is composed of 10 subunits that are independently folded proteins capable of forming a range of sub-assemblies and other complexes. These
subunits are divided in three groups: the cyclin-activated kinase (CAK) domain, the core domain, and XPD. The CAK domain is composed of CDK7, cyclin H, and MAT1. The core domain comprises p44, p34, p62, p52, trichothiodystrophy A (TTDA, also termed p8) and XPB. XPD plays a key role in linking the CAK and core domains. High resolution structures of domains and subdomains, as well as an EM structure of human TFIIH have been reported\textsuperscript{56-68}. TFIIH is recruited to the damage site by interacting with XPC through its p62 and XPB domain\textsuperscript{34,65,69}. The two NER helicases, XPB and XPD, are responsible for opening of the damaged DNA duplex and creating the NER bubble\textsuperscript{18}. A recent study revealed that the helicase activity of TFIIH is inhibited by the presence of bulky lesions and that the unwinding is XPC dependent\textsuperscript{27}.

Although XPA preferentially binds to ss-ds junction DNA, it is generally accepted that it is first recruited to the NER bubble through an interaction with TFIIH\textsuperscript{36}. XPA was also shown to enhance helicase activity of TFIIH, but only in the absence of bulky lesions, apparently to provide further damage verification during NER\textsuperscript{27}. XPA interacts with both p8 and p52 subunits of TFIIH\textsuperscript{42,43}. XPA was reported to mediate the dissociation of CAK domain from TFIIH, which then promotes incision of damage-containing nucleotide\textsuperscript{70}. Interestingly, XPA was also reported to interact with another transcription factor TFIIE\textsuperscript{71}. However, the physiological role of this interaction has yet to be established.

### 3.1.4 RPA—

RPA is the primary eukaryotic ssDNA binding protein required for virtually all DNA transactions\textsuperscript{72-75}. In NER, RPA functions together with XPA to scaffold the assembly and stabilize NER complexes. The primary function of RPA is to bind and protect the undamaged strand in the NER bubble\textsuperscript{72,73,76}. RPA also plays an important role in the transition between dual incision and the re-synthesis phase of NER\textsuperscript{16,43}.

Two contact points with XPA have been reported. The primary interaction involves the RPA32C protein recruitment domain and XPA residues 29-46. A secondary weaker interaction occurs between RPA70AB and the XPA\textsubscript{98-219}, but the specific site has not yet determined\textsuperscript{37,38,77,78} (Figure 1). Figure 7 shows the XPA-binding domains within RPA, as well as a model for XPA-RPA32C complex. There are two hypotheses for the RPA70AB binding site in XPA. NMR titration of XPA\textsubscript{98-219} with RPA70 constructs suggested the N-terminus of XPA\textsubscript{98-219} containing the zinc finger is involved\textsuperscript{31,37}. Biochemical pull-down and cell-free NER assays with XPA mutants concluded that C-terminus of XPA\textsubscript{98-219} is responsible for the interaction\textsuperscript{79,80}. In the latter model, XPA residues responsible for RPA70AB and DNA interaction may overlap. A systematic biochemical study to test how each RPA70AB-binding residue within XPA\textsubscript{98-219} affects DNA binding and NER activity concluded that K141 and K179 are involved in RPA70 interaction but not binding DNA; mutation of these residues decreases damage incision efficiency\textsuperscript{81}. They also demonstrated that disruption of both RPA32C and RPA70AB interactions severely lowered NER activity, supporting the hypothesis that both contacts are critical for NER function\textsuperscript{81}. In contrast, lysine scanning mutagenesis revealed K141 and K179 are involved in DNA binding\textsuperscript{55}. The inconsistency in DNA binding results from these studies are likely due to differences in the approaches to characterize the interaction (filter binding assay versus EMSA, different DNA substrates)\textsuperscript{55,81}. A high-resolution structure of an XPA\textsubscript{98-219}-DNA-RPA70AB complex would be extremely useful to clarify how XPA simultaneously engages DNA and protein.
binding partners. Interestingly, an NMR study revealed that the ssDNA and XPA binding sites on RPA70AB overlap. This competition may play a role in how substrates are handled and processed. Further investigation is required to map RPA70AB and DNA binding sites on XPA with greater specificity.

3.1.5 XPF/ERCC1—XPF is the structure-specific endonuclease responsible for incision 5’ to the lesion. XPF functions as a heterodimer with ERCC1. XPF/ERCC1 is recruited to the NER bubble by an interaction between ERCC192-119 and XPA96-114. An X-ray crystal structure of the ERCC1-XPA96-114 complex is available23 (Figure 7C). The ability of XPF/ERCC1 to bind DNA and XPA simultaneously has been investigated, but there remains some debate as to how XPA is positioned in the NER bubble relative to the 5’ XPF/ERCC1 cleavage site.

3.1.6 PCNA—Proliferating cell nuclear antigen (PCNA) is an essential protein for multiple DNA processing pathways. In NER, PCNA appears at the gap-filling synthesis phase to facilitate replication of the incised nucleotide using the undamaged strand as the template. It is widely accepted that all proteins in the NER incision complex, except for RPA, are displaced between the incision and gap-filling synthesis phases. However, XPA contains a PCNA binding API (AlkB homolog 2 PCNA interacting motif) sequence, and it has been shown that XPA and PCNA co-localize to damaged DNA foci in cell culture. This finding opens up a new set of mysteries: 1) Is XPA needed for gap-filling synthesis? 2) Is the XPA-PCNA interaction essential for the NER function? 3) If not, is this interaction required for DNA processing pathways other than NER?

3.2 XPA binding partners not directly involved in NER

Besides the proteins directly involved in NER, XPA is also known to interact with proteins involved in the regulation of NER, including ATR and PARP-1. Moreover, while XPA is most well recognized for its function in NER, there are also additional proteins interacting with XPA that are neither established as a part of NER nor known to be involved in the regulation of NER.

3.2.1 ATR—The serine/threonine protein kinase ATR (ataxia telangiectasia and Rad3-related, also known as FRP1 (FPAP-related protein 1)) is a central protein in the DNA damage response. ATR is known to be capable of regulating NER. In particular, ATR phosphorylation of Ser196 in XPA enhances nuclear import of XPA so that it can be localized to the sites of damage. Proteomic mass spectrometry analysis showed that this interaction is mediated within the globular XPA96-219. A recent study also showed that XPA phosphorylation by ATR enhances XPA stability by inhibiting ubiquitination by the E3 ubiquitin ligase HERC2 and subsequent degradation.

3.2.2 PARP-1—Poly(ADP-ribosylation) (PARylation) is an increasingly recognized post-transcriptional protein modification. PARylation by PARP-1 (PAR polymesase-1) is reported to be involved in the repair of DNA single and double strand breaks, as well as in NER. XPA was found to be PARylated, with the critical region mapped to C-terminal residues 213–237 that contain a conserved PAR binding motif. Interestingly, while XPA...
stimulates PARP-1 activity, PARylation of XPA was shown to reduce DNA binding activity of XPA. Cell based imaging experiments showed that PARP inhibition results in the impairment of XPA localization to sites of DNA damage, suggesting that PARylation of XPA may play a role in formation of the PIC. It is interesting to note that XPC also seems to be PARylated.

### 3.2.3 Additional XPA binding proteins

An XPA yeast two-hybrid screen identified an additional set of five XPA binding (XAB) proteins not previously known as binding partners. The validity of the approach was supported by the detection of several previously identified XPA binding partners such as RPA and ERCC1. Among the XAB proteins, XAB3, XAB4 and XAB5 were known proteins or closely related to known proteins: XAB3 is the metallopeptidase PRSM1, XAB5 is the Golgi reassembly stacking protein GRASP65, and XAB4 contained a region homologous to XAB5. The role of these XPA interactions in NER, or if these interactions suggest involvement of XPA in other pathways, is currently unclear.

XAB1 and 2 were novel proteins. XAB1 is a GTPase that interacts with residues 30–34 of XPA and contains a nuclear localization signal (NLS), which suggests that it facilitates the nuclear localization of XPA. However, as mentioned above, ATR has been shown to play an important role in XPA nuclear localization, so further investigation is needed to determine if both are required and to clarify the biological significance of the XPA-XAB1 interaction.

XAB2 is an essential protein in mice as the disruption of the XAB2 gene resulted in embryonic lethality. XAB2 contains 15 TPR (tetratricopeptide repeat) motifs and appears to have a role in TC-NER and transcription. In addition to XPA, it also interacts with other proteins involved in TC-NER such as CSA, CSB, and RNA polymerase II. The exact role of XPA-XAB2 interactions in NER remains to be investigated.

### 3.3 A structural model for XPA protein-protein interactions in NER

Figure 8 presents an initial model using available structural data for a NER incision complex containing XPA, RPA, and XPF/ERCC1, using a combination of mapped interactions between NER proteins and currently available structures. An homology model of human XPA in complex with DNA constructed using the Rad14 structure was used for placing XPA at the ssDNA-dsDNA junction 3′ to the lesion. SAXS data for the RPA DNA binding core bound to 30-nts of ssDNA was used to generate the model for RPA bound to the undamaged strand in the NER bubble. While not incorporated in Figure 8 for clarity, further modeling can incorporate the structure of XPF-ERCC1 in complex with the XPA ERCC1-binding region and the structurally characterized portions of TFIIH and XPG. XPA interactions with XPC and DDB1-XPE complexes are also relevant to modeling the early stages of assembling the NER incision complex. XPE interaction is especially interesting because it maps to the C-terminal side of the XPA DBD (residues 185–226, Figure 1), and most likely overlaps with the DNA binding site.

While there are a number of structures and models for NER proteins, the current body of information is insufficient to build complex NER incision complexes. Since the NER incision complexes progressively incorporate the key factors, it is important to investigate...
the trajectory of structure of complexes over time. Such an endeavor is within reach of current biophysical/structural techniques, in particular with the recent developments in the application of cryo-electron microscopy (cryo-EM) to structural analysis of multi-protein complexes.

4. XPA Mutations and disease

Many XPA mutations are associated with XP, however, the severity of the symptoms vary dramatically depending on the mutation. Some XPA mutations do not produce noticeable defects or only result in mild skin abnormalities, while others give rise to more severe symptoms, including progressive neurological degeneration and skin cancer. The differences in clinical outcomes are presumed to arise from partial versus complete inactivation of XPA, although the precise mechanisms remain unclear. However, it is well established that complete deletion of XPA results in very severe disease. To characterize the current understanding of how XPA mutations affect disease phenotype, we catalogued all known disease-causing XPA mutations and their biochemical effects, as well as patterns of non-disease-associated germline and somatic variation in XPA (Table 1).

The mRNA coding for XPA protein is composed of 6 exons. Severe XP symptoms are correlated with mutations resulting in little to no production of functional XPA protein, e.g., severe truncations and disruptions of the zinc finger. Deletion of exon 1 (coding for N-terminal residues including the RPA32C and ERCC1 binding regions as well as the NLS, Figure 1) was previously reported to be dispensable for NER activity and deletion of exon 6 (coding for C-terminal residues including the TFIIH binding region, Figure 1) result in marginal NER disruption. This is supported by a clinical report of two C-terminal truncation mutations that result in unusually mild XP-A symptoms. Furthermore, there are no characterized mutations in exon 1 associated with severe XP (Table 1). Deletion of any of the remaining exons (2-5), which code for the DNA binding domain, resulted in complete loss of NER activity. Biochemical studies have also shown that mutation of any of the four cysteines coordinating the zinc finger results in unfolded protein. These results led to the conclusion that the XPA-DNA interaction is critical for NER activity. However, as shown in Figures 1 and 8, these exons also code for regions important for interactions with target proteins including the DDB1/2 complex, RPA, and PCNA, as well as sites for post-translational modification. In addition, many variants that influence splice donor and acceptor sites, particularly in intron 3, have been associated with XP-A (Table 1, Figure 9).

Analyzing the frequency and patterns of germline genetic variation in XPA within relatively healthy individuals unaffected by severe XP illustrated the strength of selection on XPA and highlighted regions tolerant of mutation. We identified all missense, loss-of-function (LOF), and intronic variants observed in whole exome sequences from 60,706 unrelated individuals of diverse genetic ancestries from the Exome Aggregation Consortium (ExAC) (Figure 9). The ExAC is a multiple-cohort dataset that combines whole-exome sequencing data from several projects to provide a dense catalog of variant locations and frequencies across global populations. XPA is devoid of common protein-coding variation; the most common missense or LOF variant is at a frequency of 0.3%. This indicates considerable negative
selection on the coding sequence. Considering all rare variation in the analysis, XPA contains fewer missense and LOF variants than expected based on mutation patterns across all genes (95 sites versus 109). Exon 6 exhibits the highest density of variation with missense or LOF variants at ~21% of its translated nucleotides. This is consistent with the marginal functional disruption observed with its deletion described earlier.

XP patients have dramatically increased risk for early development of skin cancers, including basal cell carcinomas and malignant melanomas, presumably due to defects in their ability to repair UV induced DNA damage. To assess whether somatic mutations in XPA are also associated with cancer development, we identified 56 somatic mutations in XPA in 121 cancer studies from the cBio\textsuperscript{104} Portal for Cancer Genomics. No mutation was observed in more than three samples; this low recurrence rate suggests that somatic mutations in XPA are not themselves significant drivers of cancer in general; however, additional studies focused on skin cancers are needed.

Taken together, these observations suggest considerable constraint on the protein sequence of XPA; however, many rare mutations are observed in XPA in individuals without XP. Mutations that result in misfolding or severe truncation of XPA often lead to severe XP. Disruption of XPA-DNA interactions may not be sufficient to completely disturb NER and produce severe XP symptoms. It remains to be determined how disruptions of XPA’s protein interactions relate to XP severity. Understanding the mechanisms by which each mutation affects the protein, which aspects of NER are affected, and the relationship to disease symptoms will require additional genetic and structural analysis of families and individuals with XP.

**Discussion and Future Directions**

Interactions of XPA with the NER bubble and several other NER proteins make a strong case for XPA functioning as a scaffold protein. Current evidence suggests XPA facilitates the assembly and structural organization of human NER incision complexes. Because XPA interacts with NER regulatory proteins, other DNA processing proteins, as well as other proteins not related to DNA processing, XPA may also be involved in additional roles in NER or in other cellular processes. Mutations in XPA give rise to defective NER and the most disabled XP-A patients present with very severe symptoms, underscoring the importance of the XPA protein.

Mechanistic understanding of XPA function has the potential to inform drug development. On one hand, understanding the mechanism of action can be used directly to find strategies to compensate or even elevate DNA repair activities of patients with XP disease. On the other, the suppression of NER has been increasingly recognized as a potential adjuvant therapy during treatments with DNA damaging agents such as radiation and cis-platin drugs\textsuperscript{105}. These treatments result in covalent adducts and DNA cross-links, lesions that are most commonly repaired by NER. It has been increasingly recognized that resistance to treatment with DNA damaging agents arises over time due to up-regulation of the DNA damage response and repair pathways. Hence, inhibitors targeting XPA interfaces could potentially enhance the efficiency of treatment with DNA damaging agents by suppressing...
NER. Knowledge of the structure of XPA bound to the NER bubble substrate and/or other NER proteins is of interest because it reveals critical sites to target for the development of inhibitors of NER. 3D structures are of particular interest because small molecule inhibitors that target interaction interfaces are efficiently identified by structure-based approaches.

Even though a significant amount of biochemical, genetic, and functional data has been accumulated on XPA and other NER proteins, a dearth of structural information has limited progress towards understanding how XPA, and eukaryotic NER in general, actually works. Because XPA has a central role in NER through its network of protein and DNA interactions, to fully understand the function of XPA it is necessary to study it in the context of NER incision complexes. Determining structures of full-length XPA and of complexes with DNA and fragments of its partner proteins will be useful steps, but ultimately complete understanding of function requires structures of full complexes. The most significant challenge in these pursuits is the preparation of the complexes. Although in vitro NER has been achieved via reconstitution of purified components, much higher quantities are required for structural analyses and so production techniques must be optimized. One promising direction is the development of new types of expression systems for the production of protein complexes. These include new highly modularized polycistronic and polypromoter approaches, and high yield insect and mammalian cell culture technologies.

Advances in the past ~10 years in techniques for structure determination have set the stage for comprehensive studies of complex multi-domain proteins like XPA, and of multi-protein complexes like the NER incision complexes. X-ray crystallography in particular has realized a number of key developments including the shift to robotic systems for crystal screening, increased automation at synchrotron beamlines, and the availability of microfocus beamlines and FELs. In addition, exciting recent advances in cryo-EM through the development of direct electron detectors, fast data acquisition, and protocols for tracking particle movement during data acquisition, are poised to revolutionize structural biology of NER incision complexes. Equally important developments have been made in recognizing that structural snapshots are insufficient to understand the function of multi-protein complexes; the complexes are not static but rather dynamic assemblies, and even the constituent multi-domain proteins are intrinsically dynamic and constantly remodeling their architecture. This critical advance in understanding dynamic proteins and complexes has been driven by applications of small angle scattering (in particular with X-rays, SAXS) and NMR spectroscopy in combination with computational modeling.

Looking forward, we stand at the precipice of tremendous advances in understanding the mechanistic basis for the function of XPA. We anticipate that within the next 1-2 years, the structure of full-length XPA will be revealed. However, this advance will be but a stepping stone along the path to characterizing NER incision complexes and ultimately the full complex macromolecular machinery responsible for NER.

Acknowledgement

We thank Dr. Rémy Le Meur for critical proofreading of the manuscript. XPA research in our laboratory is supported by the National Institutes of Health grants R01 ES1065561 and P01 CA092584.
Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>NER</td>
<td>Nucleotide excision repair</td>
</tr>
<tr>
<td>XP</td>
<td>Xeroderma pigmentosum</td>
</tr>
<tr>
<td>TFIIH</td>
<td>transcription factor II H</td>
</tr>
<tr>
<td>RPA</td>
<td>replication protein A</td>
</tr>
<tr>
<td>WT</td>
<td>wild type</td>
</tr>
<tr>
<td>FL</td>
<td>full-length</td>
</tr>
<tr>
<td>AAF</td>
<td>N-(deoxyguanosin-8-yl)-2-acetylaminofluorene</td>
</tr>
<tr>
<td>(6-4)PP</td>
<td>(6–4)photoprotein</td>
</tr>
<tr>
<td>CPD</td>
<td>cyclobutane pyrimidine dimer</td>
</tr>
<tr>
<td>DDB</td>
<td>damaged DNA-binding protein</td>
</tr>
<tr>
<td>CAK</td>
<td>cyclin-activated kinase</td>
</tr>
<tr>
<td>TTD</td>
<td>trichothiodystrophy</td>
</tr>
<tr>
<td>PCNA</td>
<td>proliferating cell nuclear antigen</td>
</tr>
<tr>
<td>APIM</td>
<td>AlkB homolog 2 PCNA interacting motif</td>
</tr>
<tr>
<td>NLS</td>
<td>nuclear localization signal</td>
</tr>
</tbody>
</table>

References


65. Okuda M, Kinoshita M. Structural Insight into the Mechanism of TFIIH Recognition by the Acidic String of the Nucleotide Excision Repair Factor XPC. Article Structural Insight into the Mechanism of TFIIH Recognition by the Acidic String of the Nucleotide Excision Repair Factor X. Struct. Des. 2015; 23:1–11.


DNA Repair (Amst). Author manuscript; available in PMC 2017 August 01.
References


114. Tanaka, Kiyoji; Miura, Naoyuki; Satokata, Ichiro; Miyamoto, Iwai; Yoshida, Michihiro C.; Satoh, Yoshiaki; Kondo, Seiji; Yasui, Akira; Okayama, Hiroto; O, Y. Analysis of a human DNA excision repair gene involved in group A xeroderma pigmentosum and containing a zinc-finger domain. Nature. 1990; 348:73–76. [PubMed: 2234061]


Figure 1. Domain map of XPA and interaction partners
Schematic domain structure of human XPA protein (top). The region containing the globular core is colored pink, with the location of the Zn finger indicated as a yellow circle. The nuclear localization signal (NLS) is colored yellow. The N- and C-termini are dynamically disordered. Known interaction partners are shown below the domain map, aligned with the XPA residues involved in each interaction. Gray proteins are those known to interact with XPA but for which the sites of interaction have not been determined. Blue indicates a binding partner for which the binding sites on XPA remain controversial. If known, the domain or residues involved in XPA binding are given in parenthesis.
Figure 2. A structure of the globular core of XPA
Left - surface representation of the solution NMR structure of the globular core of XPA (PDB: 1XPA) colored by electrostatic field at the surface. Positive charge is in blue tones and negative charge in red tones. Right – Ribbon diagram of 1XPA.
Figure 3. Structures of *S. cerevisiae* Rad14 in complex with DNA
A) Upper panel, x-ray crystal structure of Rad14t (dark and light green) bound to a cis-platin-containing duplex (PDB: 5A39). Lower panel, sequence of the DNA duplex. B) Upper panel, x-ray crystal structure of Rad14t molecules (purple and pink) bound to an AAF-containing duplex (PDB: 5A3D). Lower panel, sequence of the DNA duplex.
Figure 4. Alignment of the XPA protein sequence across seven diverse species

A structure-guided sequence alignment of XPA proteins from seven species. The extent and secondary structure in human XPA_{98-219} construct as determined in the NMR structure (PDB: 1XPA) is indicated above the alignment. The residues not visible in the structure are indicated with the dotted line above the sequence. The secondary structure in the S. cerevisiae Rad14_{188-302} construct as determined in the crystal structures (PDB: 5A39, 5A3D) is given below the alignment. The 20-residue extension of XPA required for full DNA binding is also highlighted. Asterisks mark residues identified as critical for DNA binding.
binding in the Rad14 crystal structures. The alignment was computed by PROMALS3D\textsuperscript{1}
using 1XPA\_A and 5A3D\_A as guides. Residues are colored and conserved alignment
columns are boxed according to the default similarity scores in ESPript\textsuperscript{2}.

\textit{DNA Repair (Amst).} Author manuscript; available in PMC 2017 August 01.
Figure 5. Evolutionary conservation of XPA
The surface representation of the globular core of human XPA (PDB: 1XPA) colored by evolutionary conservation computed from the alignment of orthologous XPA sequences from human, mouse, chicken, frog, fruit fly, fission yeast, and baker’s yeast (Figure 4). The rendering of the structure was created with Chimera³.

DNA Repair (Amst). Author manuscript; available in PMC 2017 August 01.
Figure 6. Comparing structures of human XPA with *S. cerevisiae* Rad14
A) One molecule from the crystal structure of Rad14t bound to a cis-platin-containing
duplex (PDB: 5A3D). B) A representative conformer from the NMR solution structure of
the globular core of human XPA (PDB ID: 1XPA). C) Overlay of structures in panels A and
B.
Figure 7. Structures of XPA in complex with other NER proteins

A) Schematic domain map of human RPA. DNA binding domains (A, B, C, D) have stippled shading. Domains involved in protein interactions are underlined, with those involved in XPA interactions in pink.

B) Ribbon diagram of the solution NMR structure (PDB: 1DPU) of RPA32C (light green) in complex with a peptide fragment of UNG2 (salmon), which binds to RPA32C in the same manner as XPA29-46.

C) X-ray crystal structure of a peptide fragment of XPA (salmon) in complex with ERCC1 (violet) (PDB: 2JNW). XPA residue numbers are indicated in panels B and C.
Figure 8. Model of some XPA interactions in NER incision complexes

An homology model of XPA_{102-214} in complex with an AAF-containing duplex was built based on the Rad14t structure (PDB: 5A3D). A SAXS model was used for the RPA DNA binding core in complex with ssDNA. The structure of RPA70N is taken from an X-ray crystal structure (PDB: 1EWI structure). The structure of RPA32C in complex with a peptide fragment of UNG2 (PDB: 1DPU) was used to represent RPA32C bound to XPA_{29-46}. The XPF-ERCC1 model combined ERCC_{196-214} in complex with XPA_{67-80}, XPF_{842-916} in complex with ssDNA (PDB: 2KN7), and ERCC1_{220-297} (PDB: 1Z00). Dashed lines indicate potential path of linkers or DNA. The DNA lesion is represented by a red star. Colors: XPA – pink, RPA70 – blue, RPA32 – green, RPA14 – dark red, XPF – purple, ERCC1 – violet, DNA – dark grey.
Figure 9. XPA gene structure and mutations
The blue boxes give a schematic representation of the human XPA gene structure. Exons are represented by large boxes, introns by medium boxes, and introns by blue lines connecting the exons. Colored ellipses within the gene model show the location and frequency of XPA mutations observed in the ExAC database of 60,706 human exome sequences. Missense mutations and inframe indels are colored yellow; frameshifts, gained stop codons, and mutations to splice acceptor/donor sites are colored red; synonymous mutations are in green; and non-coding variants are colored black. The eccentricity of each ellipse indicates the mutation’s frequency in the ExAC population. Coding variation is rare in XPA; the most common coding variant has a frequency of 0.3%.
<table>
<thead>
<tr>
<th>Mutation in gene</th>
<th>Mutation in protein</th>
<th>Mutation type</th>
<th>Disease phenotype</th>
<th>Possible Effects on XPA Function</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>171+2T&gt;G</td>
<td>NA</td>
<td>splice site</td>
<td>XP-A; severe</td>
<td>Disrupts 5' splice donor site of intron 1</td>
<td>Tanioka (2005) J Invest Dermatol 125, 244</td>
</tr>
<tr>
<td>268_269insAA</td>
<td>variant1: V9EfsX15, variant2: V9EfsX6 P96-Q185del</td>
<td>insertion/frameshift</td>
<td>XP-A; severe form</td>
<td></td>
<td>Lehmann et al. (2014) European Academy of Dermatology and Venereology (Short Report pp 1-4)</td>
</tr>
<tr>
<td>281C&gt;T</td>
<td>P94L</td>
<td>missense</td>
<td>Severe XP; neurological disease or disruption of function</td>
<td></td>
<td>Cleaver and States (1997) Biochem. J. 328, 1-12</td>
</tr>
<tr>
<td>323G&gt;T</td>
<td>C108F</td>
<td>missense</td>
<td>XP-A; severe form</td>
<td>zinc finger disruption</td>
<td>Satokata et al. (1992) Human Genetics 88, 603-607</td>
</tr>
<tr>
<td>331G&gt;T</td>
<td>E111X</td>
<td>nonsense</td>
<td>XP-A; severe form</td>
<td></td>
<td>Amr et al. (2013) Gene 553, 52-56</td>
</tr>
<tr>
<td>348T&gt;A</td>
<td>Y116X</td>
<td>nonsense</td>
<td>Severe XP neurological disease or disruption of function</td>
<td></td>
<td>Cleaver and States (1997) Biochem. J. 328, 1-12</td>
</tr>
<tr>
<td>349_353 delCTTAT</td>
<td>L117EfsX4</td>
<td>deletion/frameshift</td>
<td>XP-A; severe form</td>
<td></td>
<td>Ghafouri-Fard et al. (2015) Gene</td>
</tr>
<tr>
<td>374delC</td>
<td>T125IfsX15</td>
<td>deletion/frameshift</td>
<td>XP-A; severe form</td>
<td></td>
<td>Amr et al. (2013) Gene 553, 52-56</td>
</tr>
<tr>
<td>377C&gt;T</td>
<td>C126T</td>
<td>missense</td>
<td>XP-A</td>
<td></td>
<td>States et al. 1998</td>
</tr>
<tr>
<td>387-1G&gt;A</td>
<td>NA</td>
<td>splice site</td>
<td>XP-A</td>
<td>Disrupts 5' splice donor site of intron 3</td>
<td>Satokata (1992) Mutat Res 273, 203</td>
</tr>
<tr>
<td>388-12A&gt;G</td>
<td>NA</td>
<td>splice site</td>
<td>XP-A</td>
<td>Disrupts 3' splice acceptor site of intron 3</td>
<td>States (1998) Hum Mutat 12, 103</td>
</tr>
<tr>
<td>388-1G&gt;C</td>
<td>NA</td>
<td>splice site</td>
<td>XP-A; severe form</td>
<td>Disrupts 3' splice acceptor site of intron 3</td>
<td>Tanaka (1990) Nature 348, 73</td>
</tr>
<tr>
<td>388-1G&gt;T</td>
<td>NA</td>
<td>splice site</td>
<td>XP-A; severe form</td>
<td>Disrupts 3' splice acceptor site of intron 3</td>
<td>States (1996) Mutat Res 363, 171</td>
</tr>
<tr>
<td>545_546insTA</td>
<td>L182Ffs</td>
<td>insertion/frameshift</td>
<td>XP-A</td>
<td></td>
<td>ClinVar</td>
</tr>
<tr>
<td>553C&gt;T</td>
<td>Q185X</td>
<td>nonsense</td>
<td>XP-A; severe form</td>
<td></td>
<td>cBio</td>
</tr>
<tr>
<td>555G&gt;C,T</td>
<td>Q185H</td>
<td>missense</td>
<td>XP-A</td>
<td></td>
<td>cBio</td>
</tr>
<tr>
<td>Mutation in gene</td>
<td>Mutation in protein</td>
<td>Mutation type</td>
<td>Disease phenotype</td>
<td>Possible Effects on XPA Function</td>
<td>Source</td>
</tr>
<tr>
<td>------------------</td>
<td>---------------------</td>
<td>---------------</td>
<td>------------------</td>
<td>----------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>555-1G&gt;C</td>
<td>NA</td>
<td>splice site</td>
<td>XP-A</td>
<td>Disrupts 5' splice donor site of intron 4</td>
<td>Satokata (1992) Mutat Res 273, 203</td>
</tr>
<tr>
<td>631C&gt;T</td>
<td>R211X</td>
<td>nonsense</td>
<td>Severe XP neurological disease or disruption of function</td>
<td></td>
<td>Cleaver and States (1997) Biochem. J. 328, 1-12</td>
</tr>
<tr>
<td>647_648delAG</td>
<td>K217EfsX3</td>
<td>deletion/frameshift</td>
<td>XP-A; severe form</td>
<td></td>
<td>Sun et al. (2015) British Journal of Dermatology 172, 1096-1100</td>
</tr>
<tr>
<td>672-1G&gt;C</td>
<td>NA</td>
<td>splice site</td>
<td>XP-A</td>
<td>Disrupts 5' splice donor site of intron 5</td>
<td>Sato (1996) Mutat Res 362, 199</td>
</tr>
<tr>
<td>682C&gt;T</td>
<td>R228X</td>
<td>nonsense</td>
<td>Mild XP neurological disease or partial function</td>
<td></td>
<td>Cleaver and States (1997) Biochem. J. 328, 1-12</td>
</tr>
<tr>
<td>683G&gt;A</td>
<td>R228G</td>
<td>missense</td>
<td>Improved Adduct Removal</td>
<td></td>
<td>Porter et al. (2005) DNA Repair 4, 341-349</td>
</tr>
<tr>
<td>690insT</td>
<td>R231KfsX15</td>
<td>insertion/frameshift</td>
<td>XP-A; mild form</td>
<td></td>
<td>Takahashi et al. (2010) Journal of Investigative Dermatology 130, 2481-2488</td>
</tr>
<tr>
<td>700G&gt;T</td>
<td>V234L</td>
<td>missense</td>
<td>Improved Adduct Removal</td>
<td></td>
<td>Porter et al. (2005) DNA Repair 4, 341-349</td>
</tr>
<tr>
<td>731A&gt;G</td>
<td>H244R</td>
<td>missense</td>
<td>XP-A; mild form</td>
<td></td>
<td>Satokata et al. (1992) Mutation Research 273, 203-212</td>
</tr>
<tr>
<td>779_780 insTT, 780_781 insTT</td>
<td>T260IfsX9</td>
<td>insertion/frameshift</td>
<td>XP-A; mild form</td>
<td></td>
<td>Takahashi et al. (2010) Journal of Investigative Dermatology 130, 2481-2488</td>
</tr>
</tbody>
</table>

The full citation of the source for each entry is listed below.