STRUCTURAL AND BIOPHYSICAL CHARACTERIZATION OF THE NUCLEOTIDE EXCISION REPAIR FACTOR XPA

By

Norie Sugitani

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in
Chemistry
May, 2017
Nashville, Tennessee

Approved by:
Walter J. Chazin, Ph. D.
David Cortez, Ph. D.
Brandt Eichman, Ph. D.
Jens Meiler, Ph. D.
Carmelo J. Rizzo, Ph. D.

To my loving hysband,

Takuto Chiba

And my fantastic parents,

Aizo and Hiroko Sugitani

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Walter J. Chazin for the opportunity to pursue my graduate career in his laboratory after working for a year as a research assistant. Time in his groups allowed me to grow as a scientist in many aspects, including laboratory skills, trouble-shooting and scientific writings. I appreciate that he was supportive for my unique career goal of becoming an astronaut and conduct research that contributes to the health of astronauts during long-term space flights.

The members of my thesis committee, Drs. David Cortez, Brandt Eichman, Jens Meiler, and Carmelo J. Rizzo have always provided me with tremendous support and insights to my project. Their expertise in structural and genome maintenance fields was fruitful in revising and carrying the research forward.

I am grateful for the collaboration work with wonderful team of people including Drs Anthony Capra, Larry Marnett, Orlando Schärer, and Zachary Nagel. I am very thankful for the involvement of Dr. Markus Voehler, director for the Biomolecular NMR facility, to this project. Timely advice of Dr. Nicole Eckert from NanoTemper helped me learn and optimize MST experiments quickly. This work would not have been possible without the support and friendship of the members of Center for Structural Biology. Especially, Heather

Darling, Karen Davis, Roy Hoffman, and Joel Harp always made my days brighter.

I would also like to thank current and former members of the Chazin group. I have learned so much from Drs. Steven Damo and Mike Shell at the early days of my research career. While I barely had no experience in NMR coming in, Dr. Sarah Soss patiently provided me with not only NMR lessons but also many hand-made sweets. I enjoyed working with talented undergraduate students Michelle Roh and Caeley Gullett. I am thankful for Dr. Agnieszka Topolska-Woś for continuing the XPA project as well as her friendship. Especially for someone like me whose English is not the first language professional copy editing by Marilyn Holt was a life saver. Lab manager Sophie Jouan was like the mom in the group always being cheerful and supportive.

Vanderbilt also provided me with the opportunity to meet many valuable friends. I enjoyed talking about animals and going to Zumba class with Lauren Salay and Amanda Erwin. I will miss the times sharing lunch an coffee with Victoria Flatt, Karuna Gujar, Xuan Zhang, Lisa Pankewitz, Hanako Kobayashi and Akiko Shimamoto.

Many thanks go to my family. While we are far away, I certainly appreciate my parents as my sister Motomi and brother Eisho for their understanding and support. Finally, my husband Takuto, I can't thank you enough

for everything. There is no ways I could have reached this day without your support even at the toughest days.

TABLE OF CONTENTS

Page

LIST OF TABES

LIST OF FIGURES

LIST OF ABBREVIATIONS

APIM AlkB homolog 2 PCNA interacting motif

ATR ataxa telongiectasia and Rad3-related

CSP Chemical Shift Perturbation

dsDNA Double-stranded DNA

DTT Dithiothreitol

FA Fluorescence Anisotropy

FITC Fluorescein isothiocyanate

HSQC Heteronuclear single-quantum coherence

IPTG Isopropyl thio-β-D-galactopyranoside

MST Microscale Thermophoresis

NER Nucleotide Excision Repair

NLS Nuclear localization signal

NMR Nuclear Magnetic Resonance

nt Nucleotide

PAR Poly(ADP-ribosyl)

PARP-1 PAR polymerase-1

PCNA Proliferating Cell Nuclear Antigen

RF Radio frequency

RPA Replication Protein A

ssDNA Single-stranded DNA

TROSY Transverse relaxation optimizes spectroscopy

XP Xeroderma Pigmentosum

XPA Xeroderma Pigmentosum Complementation Group A

XPB XPA binding

CHAPTER I

INTRODUCTION[1]

Overview

Nucleotide excision repair (NER) is multi-step, multi-protein process responsible for removing bulky lesions from the genome[1–3]. In human NER, xeroderma pigmentosum complementation group A (XPA) and replication protein A (RPA) work together as scaffolds to organize damaged DNA and other repair proteins. The importance of XPA in human NER is underscored by the observation that XPA mutations are frequently associated with severe clinical symptoms of the genetic disorder *xeroderma pigmentosum* (*XP*), including accelerated aging, increased rate of cancer and neurodegeneration[4–7]. The interaction of XPA with DNA is a core function and a number of mutations in the DNA binding domain (DBD) are associated with *XP* disease[8]. Although NMR structures of human XPA and complementary data on DNA binding have been available for many years, the molecular details of human XPA binding to DNA remain unclear. In addition, one of the two interaction sites of XPA with RPA overlaps with XPA DBD and is also poorly characterized. Structural and

[1] Part of this chapter was published in Sugitani, N., Sivley, R. M., Perry, K. P., Capra, J. A., Chazin, W. J., XPA: A Key Scaffold for Human Nucleotide Excision Repair, *DNA Repair*, 2016, 44, 123-135. Contents reused under permission by Elsevier (license #: 4063671121717).

biochemical tools have been employed to develop a model for the molecular basis of XPA-DNA and XPA-DNA-RPA interactions and the effects of disease-associated mutations. Findings from this study enhance the mechanistic understanding of human NER and *XP* disorders.

Mechanisms of Nucleotide Excision Repair

Our genome is constantly exposed to genotoxic agents such as UV-light, endogenous and exogenous reactive oxygen species, and chemical compounds from the environment. These toxins give rise to a variety of DNA lesions ranging from single strand breaks, abasic sites, cross-links and covalent adducts. If these lesions persist in the genome, the cells undergo either apoptosis or replication of the lesion-containing DNA that can result in mutagenesis, which in turn can lead to carcinogenesis. Depending on the nature of the lesion, there are different types of repair pathways to eliminate DNA damage and combat genetic instability. Nucleotide excision repair (NER) is essential for removing bulky DNA lesions from the genome in both prokatyotes and eukatyotes[1,9]. These lesions are typically generated from exposure to sun light and chemical carcinogens in the environment, as well as certain natural metabolites[10].

Defects in NER result in a spectrum of *Xeroderma pigmentosum* (*XP*) disorders[4–7]. These maladies are characterized by extreme sensitivity to sunlight, which leads to sunburn, abnormal pigmentation, and significant increases in skin

cancer[46]. The most severe cases of *XP* result in neurological degeneration, with loss of mental and sensory faculties[411–13]. Currently there is no cure or therapy for *XP* disease; the average life expectancy of *XP* patients is ~30 years, although death in early adolescence is observed for the most severe cases. Given the poor prognosis and dearth of options for *XP* patients, the development of new therapeutics remains a high priority. Detailed mechanistic understanding of the NER process offers a pathway to the discovery of new therapeutic strategies for *XP* and *XP*-associated cancer. *XP* disorders are caused by mutations in eight XP genes (*XPA* - *XPG* and *XPV*, also known as *POLH*) [1,6,14,15]. *XPA* – *XPG* genes encode for the proteins directly involved in NER pathway while *XPV* codes for the translesion polymerase η[6].

NER processing of damaged DNA in all organisms involves four distinct phases: 1) damage recognition, 2) NER bubble formation, 3) dual-incision of the damaged strand, and 4) gap-filling synthesis and ligation[1] (Figure 1.1). NER is effective at removing a wide variety of lesions because it (i) senses the presence of DNA damage without requiring damage-specific sensors[16] and (ii) removes a lesion-containing DNA oligomer instead of removing the damaged base only, which otherwise would require specialized repair mechanisms for each type of lesion. While considerable progress has been made in dissecting the mechanisms of NER in prokaryotes, detailed understanding of NER in eukaryotes has lagged

behind. This lack of progress is due in large part to the inability to transfer the

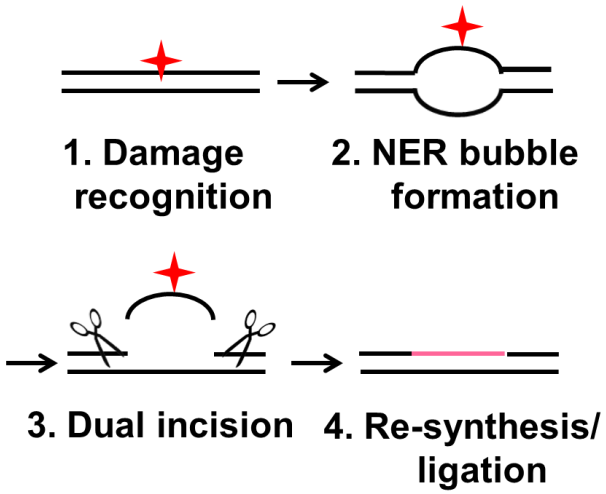


Figure 1.1: DNA processing during NER. This schematic diagram shows DNA processing during NER. Bulky DNA lesion (red star) in DNA strand (black lines) needs to be recognized in step 1. In step 2, helicases unwinds duplex around the lesion. In the third step, nucleases (represented by scissors) excise the nucleotide containing the lesion. At the last step, resulting gap is filled by synthesizing healthy nucleotides complementary to the undamaged strand.

knowledge obtained for lower organisms as a result of the lack of conservation of the repair proteins[17,18]. Human NER is more complicated than the simpler systems of lower organisms, requiring the coordinated action of over 30 proteins[18].

NER is divided into two sub-pathways: a specialized and faster pathway, transcription coupled repair (TCR), which removes lesions from actively transcribed genes; global genome repair (GGR) is responsible for NER of the rest of the genome[1,16,19–22]. Figure 1.2 represents a working model of initiation steps

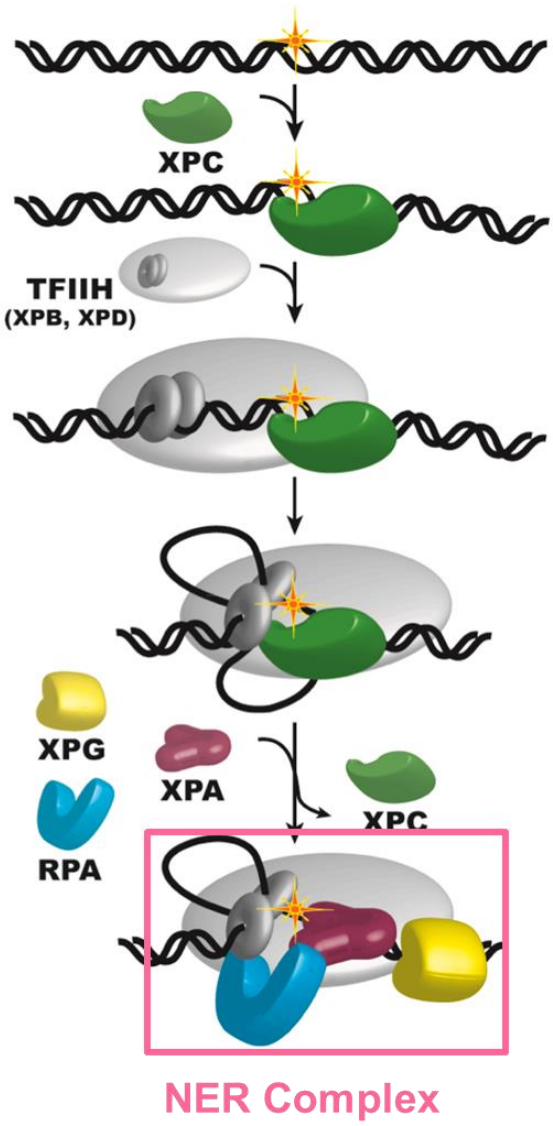


Figure 1.2: A model for initiation steps of human GGR. In human GGR, most helix distorting lesions (red star) are sensed by the protein XPC (green). XPC recruits the TFIIH complex containing two helicases XPB and XPD that are responsible for unwinding the duplex around the lesion to create the NER bubble. Other NER factors including XPA, RPA, and XPG are believed to be recruited to the bubble and XPC is displaced. Together with RPA, XPA works as a scaffold to organize the formation and remodeling of the NER complex. This figure was prepared by a former graduate student Rachel C. Wright based on the information in reference 3.

of

human GGR[1,3,17]. The main difference between TCR and GGR is at the initial step of recognizing the presence of damage. In TCR, the presence of a bulky lesion is recognized by the stalling of RPA polymerase and subsequent recruitment of TCR specific CSA and CSB proteins, whereas in GGR, the damage sensor protein XPC with or without XPE is responsible for recognizing the presence of a bulky lesion[1,19]. After recognition of the presence of damage, the transcription factor IIH (TFIIH) complex containing two helicases XPB and XPD is recruited to the damage site to unwind the duplex around the lesion[3,23,24]. XPB unwinds dsDNA on the 5′ side of the lesion, which enables loading of XPD, which in turn unwinds about 20 nucleotides of duplex towards the lesion[24]. It is generally accepted that XPD stalls when it encounters the lesion, creating the NER bubble structure[24]. Thus, formally speaking, XPD is the damage recognition factor. Once the NER bubble is formed, scaffolding proteins XPA and RPA are recruited to the bubble. RPA is associated with the undamaged strand and XPA with the damaged strand and the adjacent junction[25,26]. The 3′ nuclease XPG is also recruited as this time, which correlates with dissociation of XPC from the complex[3]. This assembly of proteins around the bubble is termed the NER pre-incision complex (Figure 1.2), and represents the formal commitment to NER. Without the correct formation of the NER complex, nucleases will not cut the lesion-containing nucleotide properly. Once the pre-incision complex is formed, the 5′ nuclease XPF/ERCC1 is recruited[27]. XPF cuts the nucleotide 5′ to the

lesion followed by 3′ incision by XPG, removing 24 – 30 nucleotides from the damage containing nucleotide (long)[27–29]. The gap resulting from incision is filled by the replicative polymerase machinery, using the undamaged strand as template[30].

XPA

The *XPA* gene was discovered in the 1970s in cell cultures derived from *XP* patients, along with other disease-associated *XP* genes (XPB – XPG, XPV) [31–35]. Since then, extensive efforts have been made to characterize the biochemical properties of XPA and understand its function in NER as well as the impact of specific disease-associated mutations. XPA is involved in both TCR and GGR; the other proteins involved and their roles in these two sub-pathways are described elsewhere[3,16,20–30]. In both pathways, XPA is recruited to the damage site by the 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[2,36–38]. 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[25,26]. 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? The following sections 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.

XPA Structure and Interactions with DNA

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 1.3. XPA is organized around a central globular domain (XPA_{98-219}). 3D structures of this domain were determined independently by two groups using solution NMR (PDB: 1XPA, 1D4U) (Figure 1.4)[39,40]. XPA_{98-219} contains a C4 type zinc-finger motif[41] in the N-terminal region and a shallow basic cleft in the C-terminal region (Figure 1.4). The N- and C- termini of XPA are disordered and mediate a variety of protein interactions[42–48]. Interestingly, severe *XP* symptoms associated with XPA mutations map primarily to the central domain[8].

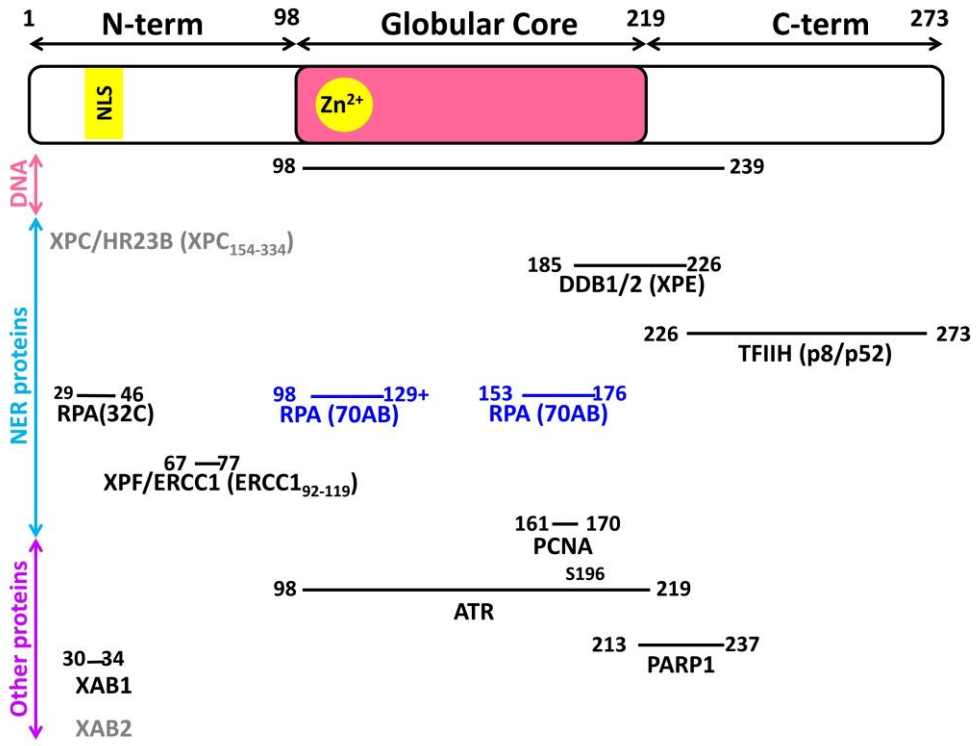


Figure 1.3: Domain map of XPA and interaction partners. Schematic domain structure of human XPA protein (top). The region containing the globular core is colored pink, with the location of the Zn finger indicated as a yellow circle. The nuclear localization signal (NLS) is colored yellow. The N- and 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.

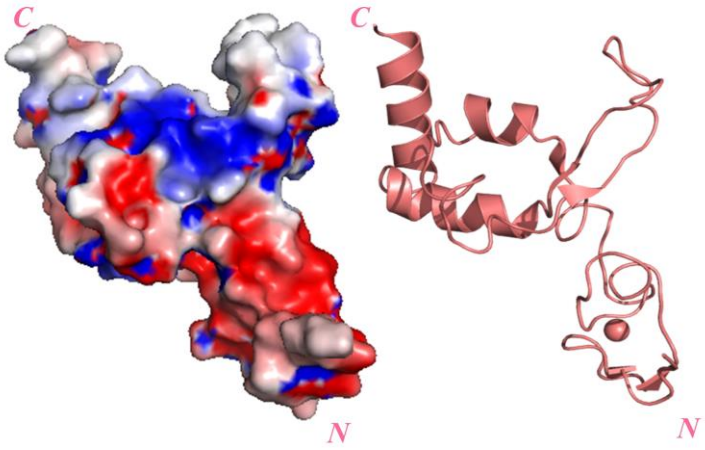


Figure 1.4: A structure of the globular core of XPA. Left - surface representation of the solution NMR structure of the globular core of XPA (PDB: 1XPA) colored by electrostatic field at the surface. Positive charge is in blue tones and negative charge in red tones. Right – Ribbon diagram of 1XPA.

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[37]. Whether XPA binds to the junction 5′ or 3′ to the lesion remains unclear as evidence has accumulated supporting both models[49]. XPA is recruited to NER complexes via interactions of its flexible C-terminus with the p8 subunit of TFIIH (Figure 1.3)[50,51]. 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 p8within 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[2]. 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[26]. 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.

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[41]. 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 XPA_{98-219} structure and generate a model of the complex[52]. 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[37]. 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 XPA_{98-219} core must be extended C-terminally by ~20 residues (see Chapter II)[53,54].

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 1.5)[55]. 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-deoxydiguanosine intrastrand crosslink (PDB: 5A39) and the second was with the same duplex containing a *N*-(deoxyguanosin-8-yl)-2-acetylaminofluorene (AAF) (PDB: 5A3D) adduct

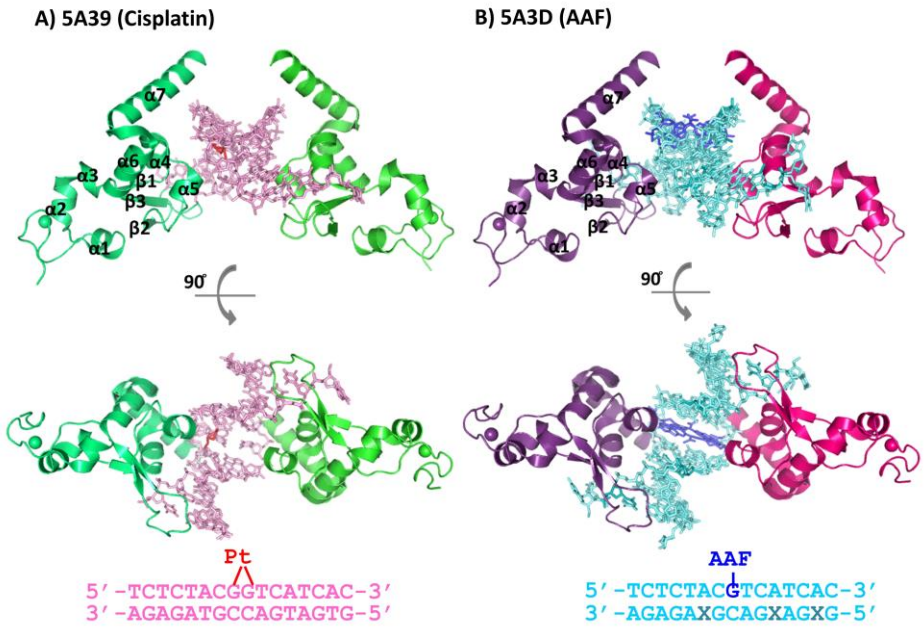


Figure 1.5: Structures of *S. cerevisiae* Rad14 in complex with DNA. A) Upper panel, x-ray crystal structure of Rad14t (dark and light green) bound to a cis-platin-containing duplex (PDB: 5A39). Lower panel, sequence of the DNA duplex. B) Upper panel, x-ray crystal structure of Rad14t molecules (purple and pink) bound to an AAF-containing duplex (PDB: 5A3D). Lower panel, sequence of the DNA duplex. X represents 5-iododeoxyuridine.

(Figure 1.5). 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 70º (Figure 1.5)[55]. 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[37]. Also,

these structures support the idea that XPA does not make direct contact with the lesion as suggested previously[3,56]. It is interesting that Rad14 binds to both damaged duplexes as a dimer, consistent with a report that XPA forms a dimer[57]. Despite these *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 1.6 shows a structure-guided alignment of XPA homologs from seven diverse species, and Figure 1.7 maps the evolutionary conservation of each position in XPA onto the 1XPA structure. The human XPA and *S. cerevisiae* Rad14 constructs used for structural studies are also highlighted on the alignment. The Rad14 construct used for crystallization ($Rad14_{188-302}$) has two insertions, one deletion, is four residues shorter at the N-terminus and four residues shorter at the C-terminus, and has 29% identity and 57% conservation to XPA_{98-219} (Figure 1.6). As noted above, XPA_{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[53]. It is therefore surprising that FL-XPA, FL-Rad14, and Rad14t bind duplex DNA containing cisplatin or AAF lesions very tightly[55]. Moreover, Rad14 does not bind to duplexes containing other commonly studied DNA lesions (e.g. (6-4)photoproduct ((6-4)PP), cyclobutane pyrimidine dimer

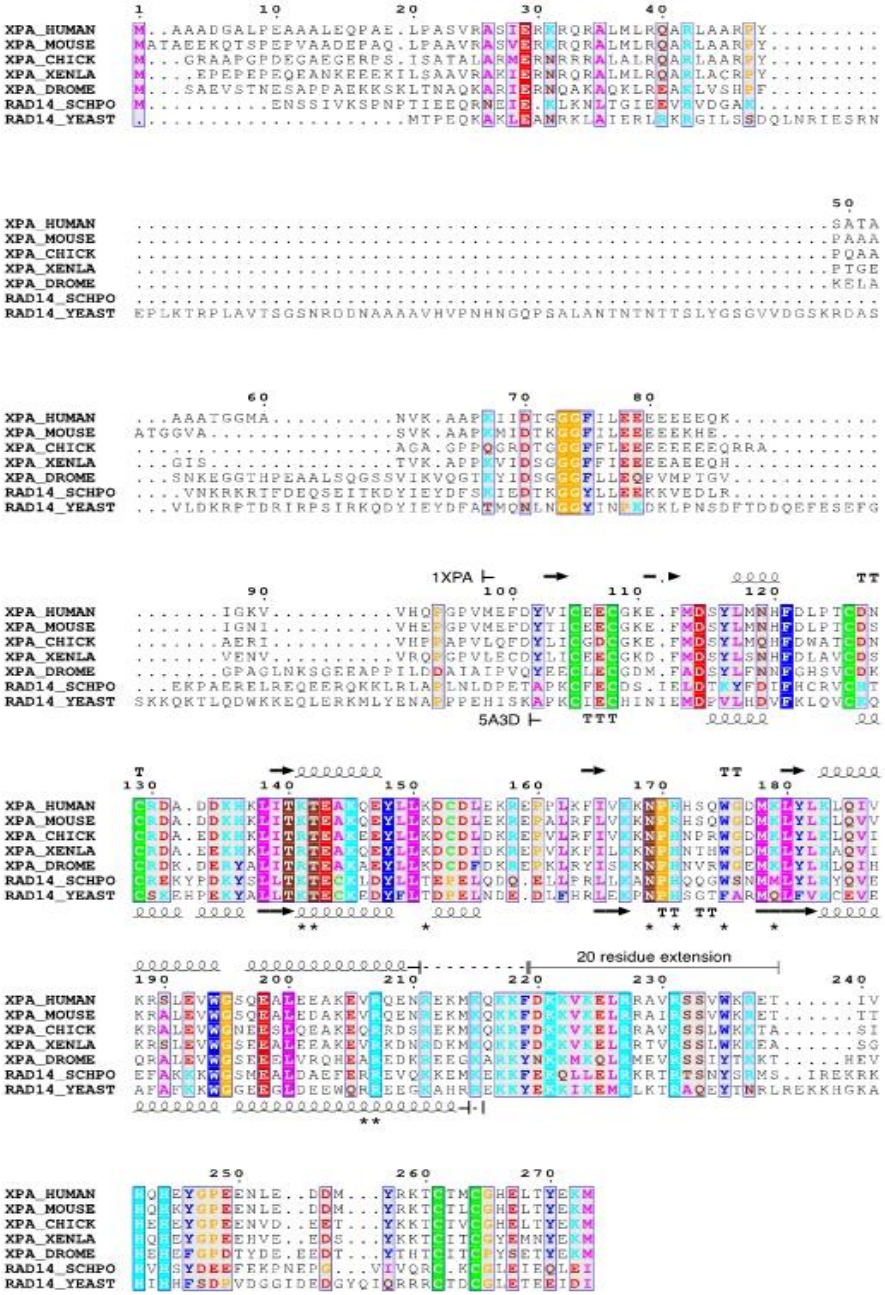

XPA_HUMAN
XPA_MOUSE
XPA_CHICK
XPA_XENLA
XPA_DROME
RAD14_SCHPO
RAD14_YEAST
1XPA
5A3D
20 residue extension

Figure 1.6: 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 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 indicated above the alignment. The residues not visible in the for DNA binding in the Rad14 crystal structures. The alignment was computed by PROMALS3D[133] 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[134].

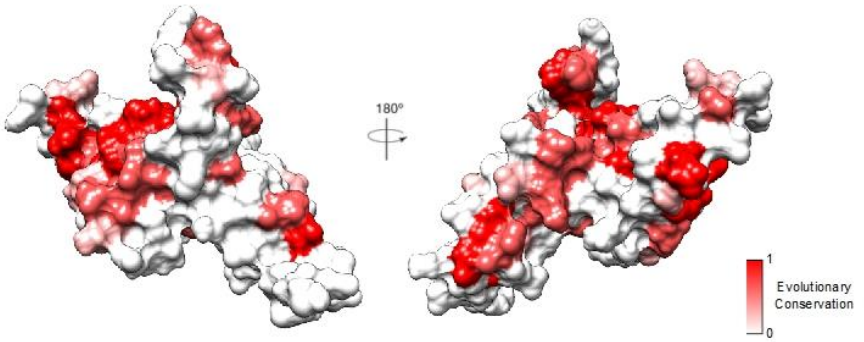


Figure 1.7: Evolutionary conservation of XPA. The surface representation of the globular core of human XPA (PDB: 1XPA) colored by evolutionary conservation computed from the alignment of orthologous XPA sequences from human, mouse, chicken, frog, fruit fly, fission yeast, and baker's yeast (Figure 1.6). The rendering of the structure was created with Chimera[135].

(CPD)) with appreciable affinity[55]. In light of these observations, it is of interest to know if XPA_{98-219} binds these substrates with comparable affinity.

Figure 1.8 compares the Rad14 structure with the solution NMR structure of XPA_{98-219}. Although the Rad14 construct is shorter, a larger number of C-terminal residues were observed in the crystal structure. Moreover, Rad14 has

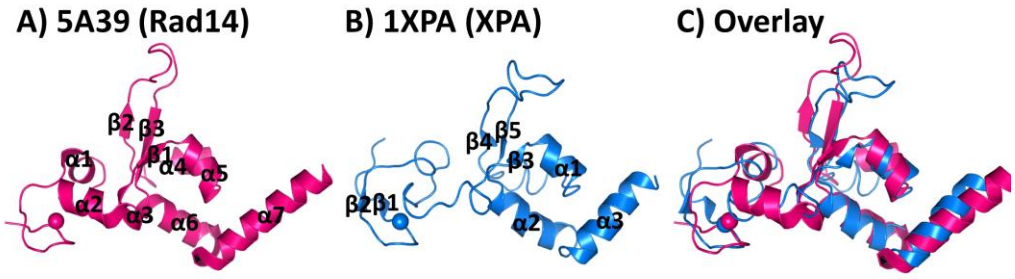


Figure 1.8: Comparing structures of human XPA with *S. cerevisiae* Rad14. A) One molecule from the crystal structure of Rad14t bound to a cis-platin-containing duplex (PDB: 5A3D). B) A representative conformer from the NMR solution structure of the globular core of human XPA (PDB ID:1XPA). C) Overlay of structures in panels A and B.

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_{98-219} structures are very similar (Figure 1.8-C); 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_{98-219} structures imply that XPA will bind DNA in a

manner similar to Rad14 overall. However, because these structures were determined with 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.

XPA-Protein Interactions

The interaction of XP proteins with their binding partners was reviewed in 2008[58]. 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.

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.

XPC

XPC is a 106 kDa protein responsible for detecting the presence of DNA damage in the GG-NER pathway[2]. XPC functions as a heterotrimer with HR23B and centrin-2, which stimulate XPC DNA binding activity and increases cellular stability[59]. Once engaged on the damaged duplex, XPC recruits the TFIIH complex[51]. As discussed in the TFIIH section 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 XPA pulls down XPC_{1-492} and $XPC_{154-334}$, suggesting that N-terminal residues 154-334 are responsible for binding to XPA[60]. 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. In particular, one would like to know if XPA-XPC interaction is 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[51].

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[61], and together they recognize a wide variety of lesions[62]. Mutations in XPE often result in mild *XP* disorders[62]. Although the DDB1/2 complex is dispensable for NER reconstituted *in vitro*, it enhances this activity especially for CPD lesions[62]. 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 bind to DNA containing this lesion[63,64]. 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 [48]. 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 [48]. In other studies, R207 was reported to be involved in DNA binding[65], and the R207Q mutation was discovered in cancer patients (Table 1.1). 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.

Table 1.1: Disease associated mutations in XPA

Mutation in gene	Mutation in protein	Mutation type	Disease phenotype	Possible Effects on XPA Function	Source
171+2T>G	NA	splice site	XP-A; severe	Disrupts 5′ splice donor site of intron 1	Tanioka et al.[136]
268_269insAA	variant1: V9EfsX15, variant2: V9EfsX6 P96-Q185del	insertion/ frameshift	XP-A; severe form		Lehmann et al.[137]
281C>T	P94L	missense	Severe XP; neurological disease or disruption of function		Cleaver and States[13]
323G>T	C108F	missense	XP-A; severe form	zinc finger disruption	Satokata et al./ States et al.[8,138]
331G>T	E111X	nonsense	XP-A; severe form		Amr et al./ Messaoud et al.[139,140]
348T>A	Y116X	nonsense	Severe XP neurological disease or disruption of function		Cleaver and States[13]
349_353 delCTTAT	L117EfsX4	deletion/ frameshift	XP-A; severe form		Ghafouri-Fard et al.[141]
374delC	T125IfsX15	deletion/ frameshift	XP-A; severe form		Amr et al.[139]
377C>T	C126T	missense	XP-A		States et al.[8]
387-1G>A	NA	splice site	XP-A	Disrupts 5′ splice donor site of intron 3	Satokata et al.[138]
388-12A>G	NA	splice site	XP-A	Disrupts 3′ splice acceptor site of intron 3	States et al.[8]
388-2A>G	NA	splice site	XP-A	Disrupts 3′ splice acceptor site of intron 3	Satokata et al. [142]
388-1G>C	NA	splice site	XP-A; severe form	Disrupts 3′ splice acceptor site of intron 3	Tanaka et al. [108]
388-1G>T	NA	splice site	XP-A	Disrupts 3′ splice acceptor site of intron 3	States and Myrand[143]
545_546insTA	L182Ffs	insertion/ frameshift	XP-A		ClinVar
553C>T	Q185X	nonsense	XP-A; severe form		cBio
555G>C,T	Q185H	missense	XP-A		cBio
555-1G>C	NA	splice site	XP-A	Disrupts 5′ splice donor site of intron	Satokata et al. [138]

				4	
555+8A>G	NA	splice site	XP-A	Disrupts 5′ splice donor site of intron 4	Sidwell et al.[144]
619C>T	R207X	nonsense	XP-A, neurological impairment and mild skin abnormality		Santiago et al./Messaoud et al.[140,145]
620G>A	R207Q	missense		inhibition of XPE binding	Wakasugi et al.[48]
622C>T	Q208X	nonsense	XP-A		Maeda et al.[146]
631C>T	R211X	nonsense	Severe XP neurological disease or disruption of function		Cleaver and States[13]
647_648delAG	K217EfsX3	deletion/ frameshift	XP-A; severe form		Sun et al.[147]
672-1G>C	NA	splice site	XP-A	Disrupts 5′ splice donor site of intron 5	Sato et al.[148]
682C>T	R228X	nonsense	Mild XP neurological disease or partial function		Cleaver and States[13]
683G>A	R228G	missense	Improved Adduct Removal		Porter et al.[149]
690insT	R231KfsX15	insertion/ frameshift	XP-A; mild form		Takahashi et al.[111]
700G>T	V234L	missense	Improved Adduct Removal		Porter et al.[149]
731A>G	H244R	missense	XP-A; mild form		Satokata et al.[138]
779_780 insTT, 780_781 insTT	T260IfsX9	insertion/ frameshift	XP-A; mild form		Takahashi et al.[111]

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[66–78]. TFIIH is recruited to the damage site by interacting with XPC through its p62 and XPB domain[42,75,79]. The two NER helicases, XPB and XPD, are responsible for opening of the damaged DNA duplex and creating the NER bubble[24]. A recent study revealed that the helicase activity of TFIIH is inhibited by the presence of bulky lesions and that unwinding is XPC dependent[38].

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[44]. XPA was also shown to enhance the helicase activity of TFIIH, but only in the absence of bulky lesions, apparently to provide further damage verification during NER[38]. XPA interacts with both p8 subunit of TFIIH[50,51]. XPA was reported to mediate the dissociation of CAK domain from TFIIH, which then promotes incision of damage-containing nucleotide[80]. Interestingly, XPA was also reported to interact with another transcription factor TFIIE[81]. However, the physiological role of this interaction has yet to be established.

RPA

RPA is the primary eukaryotic ssDNA binding protein required for virtually all DNA transactions[82–85]. 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[82,83,86]. RPA also plays an important role in the transition between dual incision and the re-synthesis phase of NER[3,51].

Two contact points with XPA have been reported. The primary interaction involves the RPA32C protein recruitment domain and XPA residues 29-46[46]. A secondary weaker interaction occurs between RPA70AB and the XPA_{98-219}, but the specific site has not yet determined[45,87,88] (Figure 1.3). Figure 1.9 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_{98-219} with RPA70 constructs suggested the N-terminus of XPA_{98-219} containing the zinc finger is involved[39,45]. Biochemical pull-down and cell-free NER assays with XPA mutants concluded that C-terminus of XPA_{98-219} is responsible for the interaction[89,90]. In the latter model, XPA residues responsible for RPA70AB and DNA interaction may overlap. A biochemical study to test how each RPA70AB-binding residue within XPA_{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[90]. They also demonstrated that disruption of

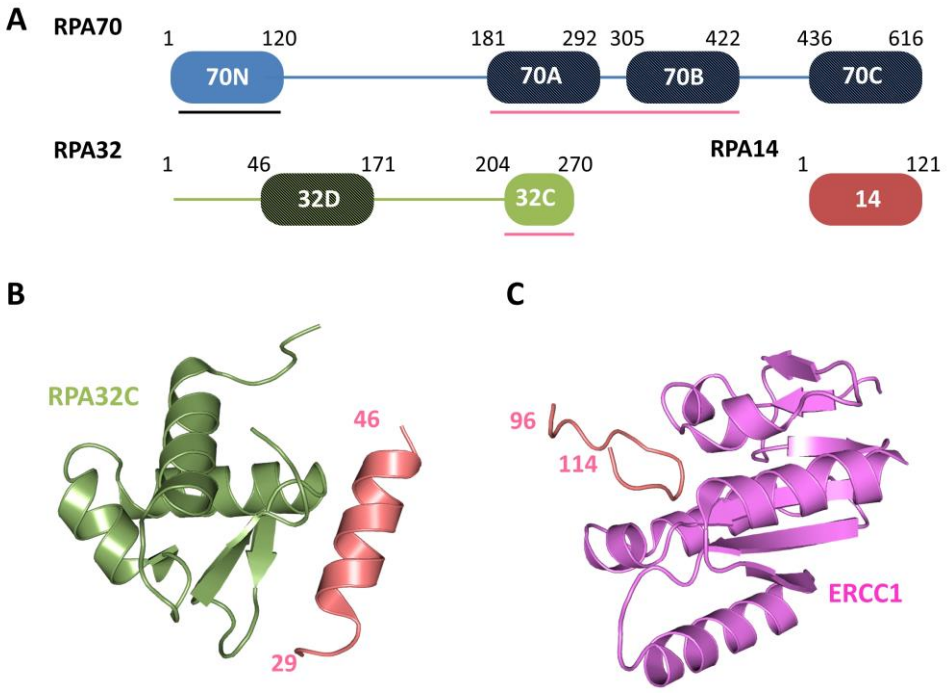


Figure 1.9: Structures of XPA in complex with other NER proteins. A) Schematic domain map of human RPA. DNA binding domains (A, B, C, D) have stipled shading. Domains involved in protein interactions are underlined, with those involved in XPA interactions in pink. B) Ribbon diagram of the solution NMR 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 XPA_{29-46}. 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.

both RPA32C and RPA70AB interactions severely lowered NER activity, supporting the hypothesis that both contacts are critical for NER function[90]. In

contrast, lysine scanning mutagenesis revealed K141 and K179 are involved in DNA binding[65]. 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)[65,90]. A high-resolution structure of an XPA_{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[88]. 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.

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 $ERCC1_{92-119}$ and XPA_{96-114}[43]. An X-ray crystal structure of the ERCC1-XPA_{96-114} complex is available[27] (Figure 1.9-C). 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[49].

PCNA

Proliferating cell nuclear antigen (PCNA) is an essential protein for multiple DNA processing pathways[91]. 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 APIM (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[92]. 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?

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.

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[93]. Proteomic mass spectrometry analysis showed that this interaction is mediated within the globular XPA_{98-219}[94]. 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[95].

PARP-1

Poly(ADP-ribosyl)ation (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[96–100]. XPA was found to be PARylated, with the critical region mapped to C-terminal residues 213–237 that contain a conserved PAR binding motif[101]. Interestingly, while XPA stimulates PARP-1 activity, PARylation of XPA was shown to reduce DNA binding activity of XPA[101]. 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[101]. It is interesting to note that XPC also seems to be PARylated[102].

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[103]. 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[103]. 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[103] and contains a nuclear localization signal (NLS)[104], which suggests that it facilitates the nuclear localization of XPA. However, 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[105]. XAB2 contains 15 TPR (tetratricopeptide repeat) motifs and appears to have a role in TCR and transcription[106]. In addition to XPA, it also interacts with other proteins involved in TCR such as CSA, CSB, and RNA polymerase II[106]. The exact role of XPA-XAB2 interactions in NER remains to be investigated.

XPA Mutations and Disease

Many XPA mutations are associated with *XP*; however, the severity of the symptoms vary dramatically depending on the mutation[8]. 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[4,12,13,107]. 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 1.1).

The mRNA coding for XPA protein is composed of 6 exons (Figure 1.10)[108,109]. 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[8,13,110]. Deletion of exon 1 (coding for N-terminal residues including the RPA32Cand ERCC1 binding regions as well as the NLS, Figure 1.3) 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 1.3) result in marginal NER disruption[109]. This is supported by a clinical report of two C-terminal truncation mutations that result in unusually mild *XP-A* symptoms[111]. Furthermore, there are no characterized mutations in exon 1 associated with severe *XP* (Table 1.1). 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[110]. These results led to the conclusion that the XPA-DNA interaction is critical for NER activity. However, as shown in Figure 1.3, 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 1.1, Figure 1.10).

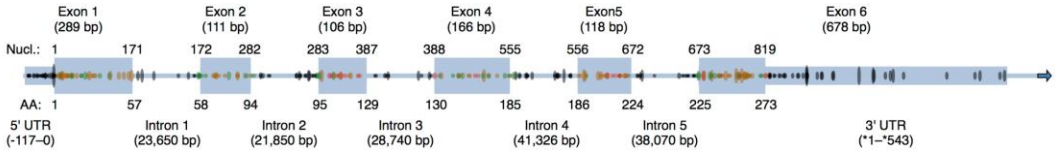


Figure 1.10: XPA gene structure and mutations. The blue boxes give a schematic representation of the human XPA gene structure. Exons are represented by large boxes, introns by medium boxes, and introns by blue lines connecting the exons. Colored ellipses within the gene model show the location and frequency 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%.

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 1.10). 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[112] 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*.

Experimental Overview

Investigation of XPA interactions with other macromolecules such as DNA and RPA would not have been possible without the integration of structural and biophysical tools. This section provides a brief overview of the concepts for the key experimental methods employed in this study.

Nuclear magnetic resonance (NMR) spectroscopy

NMR spectroscopy provides information on the structure, dynamics, and interactions of biomolecules. The NMR phenomenon is based on the induction of resonance in NMR-active atomic nuclei, which can be detected as characteristic frequencies as the system returns to equilibrium[113–117]. The NMR chemical shift parameter (the characteristic frequency) is highly sensitive to the surrounding chemical environment and this provides a valuable tool for determining structures and characterizing interactions. Information on dynamics can be obtained

because the decay of the NMR signal back to equilibrium is highly sensitive to molecular motions. NMR spectroscopy can be performed in the solid state NMR or in solution. In this work, all NMR experiments employed the solution NMR approach.

Although NMR spectroscopy can be applied to a variety of molecules, this section will focus on its application for proteins as this dissertation focuses on the characterization of XPA protein. The most basic 1D NMR spectrum provides valuable insights into protein structure and can be used for example to assay tertiary structure and for monitoring the progress of titrations. However, 1D for proteins is limited because proteins and nucleic acids have so many nuclei that their ^{1}H 1D NMR spectra are extensively overlapped. Higher order hetero-nuclear 2D (e.g. correlation between ^{1}H and ^{15}N or ^{1}H and ^{13}C) or 3D (correlation between ^{1}H, ^{15}N and ^{13}C) NMR experiments resolve the overlap problem by dispersing overlap resonances in the extra dimension(s). Isotopic enrichment is required for nearly all heteronuclear studies. Although the ^{1}H isotope is found in high natural abundance in macromolecules, this is not the case for the key ^{15}N and ^{13}C nuclei. In order to incorporate these isotopes, proteins are usually recombinantly expressed in bacteria grown in minimal media supplemented with ^{15}N-ammonium chloride or ^{13}C-glucose as sole nitrogen and/or carbon sources[118].

Assigning the resonances to specific nuclei in the molecule is an essential step for any in-depth NMR analysis. The assignment process requires acquisition of a series of complementary experiments.

While there are many different types of 2D NMR experiments, this dissertation project employed ^{15}N-^{1}H heteronuclear single quantum coherence (HSQC) experiment, a key tool for structural analysis of proteins[119]. ^{15}N-^{1}H HSQC is often described as 'finger print of a protein'. The NMR spectrum from this experiment gives the chemical shift for each correlated ^{15}N-^{1}H pair as a single peak in the 2D plane; in other words, each chemical shift represents each amino acid in the protein backbone, except for proline. HSQC informs several important properties of proteins. Predictions of folding as well as secondary and tertiary structure of the protein can be obtained by observing the dispersion of the chemical shifts. A series of 3D NMR experiments allows for assignment of the spectra, or identifying which chemical shift corresponds to which amino acid in the protein[120,121]. As mentioned above, positions of each chemical shift in the HSQC spectrum depends on the chemical environment of each nuclear spin. Therefore, as long as assignments are available, HSQC can also reveal which amino acids are involved in interaction with other molecules because only chemical shifts experiencing the chemical environment change (for example, by interaction with other molecules) will change in its intensity or position in the spectrum[122].

While HSQC is a powerful tool for characterizing proteins, it is limited to proteins smaller than ~35 kDa. Larger molecules tumble slowly in solution, resulting in faster (shorter) transverse relaxation times (T_2). This leads to broadening of chemical shift and poses difficulty detecting those peaks. Development of transverse relaxation optimized spectroscopy (TROSY) allowed for the basic heteronuclear correlation experiments to be expanded to proteins or complexes up to ~ 100 kDa[123]. TROSY cancels different physical processes that dictate relaxation rates to attenuate T_2 relaxation, resulting in single, sharp chemical shift in the NMR spectrum[123].

Fluorescence anisotropy (FA) binding assay

Fluorescence anisotropy (FA) is an approach to measure the binding of a ligand to a fluorescently tagged molecule and can be used in a high-throughput mode for quantifying interactions[124]. The application of the FA assay to measure protein-nucleic acids interactions is well established[125–127]. In this project, FA assay was exclusively used to screen binding of XPA constructs to various fluorescently labeled DNA substrates. When the fluorophore attached to DNA is excited, it exhibits fluorescence anisotropy, or the phenomenon where the light emitted by the fluorophore shows different intensities along parallel ($I_{\parallel}$) and

perpendicular ($I_{\perp}$) axes of polarization. Fluorescence anisotropy (r) is defined as the ratio of the polarized component to the total intensity (I_T): $r = (I_{\parallel} - I_{\square})/ I_T$.

As a protein binds to the labeled DNA, fluorescence anisotropy of the fluorophore changes, partially due to the change in size of the molecule to slow tumbling in solution. FA assay monitors this change in fluorescence anisotropy as increasing amount of protein is added to the labeled DNA. The FA assay is typically set up in a 384-well plate where each column of the plate contains one titration and fluorescence anisotropy of each well is read by a plate reader. This allows for high-throughput, collecting eight distinct reactions with each having triplicates in one experimental run. The plate based assay is also economical as it only requires very small amounts of purified components at relatively low concentrations (nm – μm range).

Microscale thermophoresis

Microscale thermophoresis (MST) is a powerful tool to quantify interactions between biomolecules[128]. MST depends on the thermophoresis phenomenon whereby molecules move along a temperature gradient with distinct properties[129]. When a protein is in a solution at a certain temperature, molecules are distributed homogeneously. If the temperature gradient is induced in the

solution, molecules exhibit positive or negative thermophoresis (movement of molecules from warm to cold or cold to warm temperatures, respectively)[129]. In MST experiments, change in thermophoresis is detected by monitoring the movement of fluorescently labeled molecule as the temperature gradient is induced by local exposure to an infrared laser. In the case of the experiments performed for this thesis work, the DNA molecules were fluorescently labeled. The protocol involves labeled molecules at a known concentration being titrated with unlabeled binding partner in glass capillaries. The rate of thermophoresis depends on various factors including the size of the molecule. Therefore, fluorescently labeled DNA mixed with different concentrations of XPA proteins exhibit distinct thermophoresis from those not interacting with XPA. This difference in thermophoresis at each point of titration can be fit to a binding curve to quantify the interaction[128]. While MST requires more reagents and time compared to FA assay, it is still a relatively economical and quick assay: a titration takes about 15 minutes. The main advantage of MST over the FA assay is the ease of troubleshooting. Since it is a specialized method for assaying binding, NanoTemper, Inc. has established a case-by-case troubleshooting protocol, three types of capillaries to accommodate various samples, as well as the support system by the technology specialists for their Monolith instruments.

Circular dichroism

Light can be either linearly or circularly polarized. Circular dichroism (CD) is the phenomenon whereby right and left circularly polarized light are absorbed unequally[130]. CD can be represented in absorption bands of any optically active molecule and it reflects secondary structures of these molecules. Therefore, in biochemistry, CD is most often applied to study the secondary structures as well as folding of macromolecules[131,132]. In this project, CD was applied to compare the thermostabilities of mutant XPA DBD to WT using CD thermodenaturation. Since the CD spectra represent secondary structures of the protein, it can also represent distinct folded and unfolded states of the protein. CD thermodenaturation monitors CD spectra at of a protein at a set wavelength as the temperature increases to capture the loss in secondary structure as the protein unfolds.

Acknowledgements

We thank Dr. Steven M. Shell for critical proofreading of this chapter and Rachel C. Wright for generously providing Figure 1.2.

CHAPTER II

RE-DEFINING THE DNA-BINDING DOMAIN OF HUMAN XPA[2]

Abstract

Xeroderma pigmentosum complementation group A (XPA) protein plays a critical role in the repair of DNA damage via the nucleotide excision repair (NER) pathway. XPA serves as a scaffold for NER, interacting with several other NER proteins as well as the DNA substrate. The critical importance of XPA is underscored by its association with the most severe clinical phenotypes of the genetic disorder *Xeroderma pigmentosum*. Many of these disease-associated mutations map to the XPA_{98-219} DNA-binding domain (DBD) first reported ~20 years ago. Although multiple solution NMR structures of XPA_{98-219} have been determined, the molecular basis for the interaction of this domain with DNA is only poorly characterized. In this report, we demonstrate using a fluorescence anisotropy (FA) DNA-binding assay that the previously reported XPA DBD binds DNA with substantially weaker affinity than the full-length protein. In-depth analysis of the XPA sequence suggested that the original DBD construct lacks critical basic charge and helical elements at its C-terminus. Generation and analysis of a series of C-terminal extensions beyond residue 219 yielded a stable,

[2] This work was published in Sugitani, N., Shell, S. M., Soss, S. E., Chazin, W. J., Re-defining the DNA-Binding Domain of Human XPA, *J. Am. Chem. Soc.* 2014, 136, 10830-10833.

soluble human XPA_{98-239} construct that binds to a Y-shaped ssDNA-dsDNA junction and other substrates with the same affinity as the full-length protein. Two-dimensional ^{15}N-^{1}H NMR suggested XPA_{98-239} contains the same globular core as XPA_{98-219} and likely undergoes a conformational change upon binding DNA. Together, our results demonstrate that the XPA DBD should be redefined and that XPA_{98-239} is a suitable model to examine the DNA binding activity of human XPA.

Introduction, Results and Discussion

Nucleotide excision repair (NER) is a highly versatile DNA damage repair pathway that is able to remove bulky DNA lesions arising from exposure to sunlight, endogenous metabolites, and various environmental toxins[1,9]. Defects in NER result in the genetic disease *Xeroderma pigmentosum* (*XP*), a spectrum of disorders characterized by hypersensitivity to sunlight, dramatically increased incidents of skin cancer, and neurological disorders[4,12]. NER in humans involves the coordinated action of ~30 proteins, including seven that were identified based on their direct association with specific *XP* disorders (XPA-XPG)[3,58,150]. Among these, the essential *XP* complementation group A protein (XPA) is associated with the most severe clinical *XP* phenotypes, leading to neurodegenerative disorders, accelerated aging, and cancer[12].

Despite its key importance to NER, XPA has no known enzymatic function[13]. However, XPA is known to bind to DNA and a number of other NER proteins, suggesting that it serves as a scaffold for the complex multi-protein NER machinery[26,27,58,151]. Genetic and biochemical studies suggest that DNA binding by XPA is crucial for the proper function of NER. Moreover, a number of XPA mutations associated with severe *XP* symptoms map to residues in the DNA-binding domain (DBD)[4,11,13]. Nevertheless, there has yet to be any systematic biophysical and structural characterization of the interactions between XPA and DNA.

The discovery of the human XPA DBD was reported nearly 20 years ago. Biochemical studies revealed a protease resistant domain within residues 98-219 that was associated with binding of DNA[41]. Additional studies suggested that relative to ssDNA or dsDNA, XPA binds preferentially to DNA containing ssDNA-dsDNA junctions[37,125]. This observation was of particular interest because NER requires unwinding of the DNA duplex, which creates ssDNA-dsDNA junctions. Two solution NMR structures of XPA_{98-219} were subsequently determined[39,45]. NMR chemical shift analysis was also used to investigate binding of a 9 nucleotide (nt) ssDNA substrate, which enabled mapping of the interaction to a shallow basic cleft in XPA_{98-219}[52]. However, the affinity for this substrate is extremely weak (Kd estimated to be several mM), which leads to considerable doubt about whether this model accurately represents how XPA interacts with

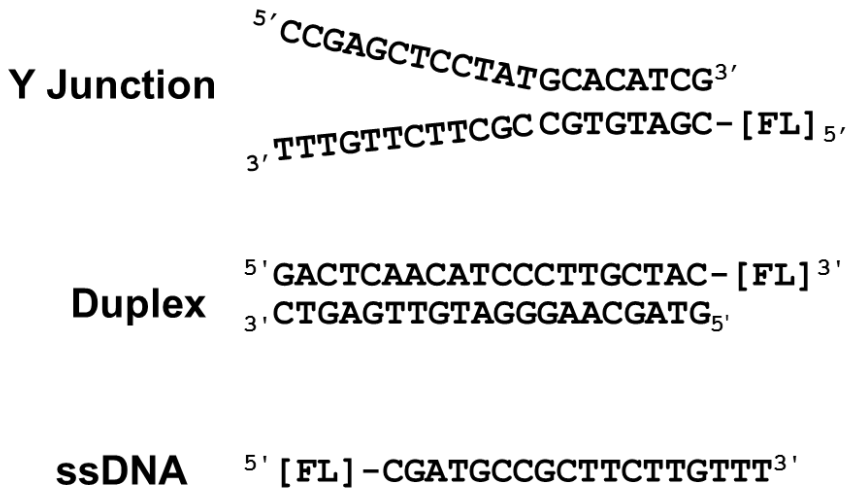


Figure 2.1: DNA substrates used for FA assays. Top- Y-shaped ssDNA-dsDNA junction containing an 8 basepair duplex with 12-nt ssDNA overhangs. Middle- 20 basepair duplex. Bottom- 20-nt ssDNA. FL indicates the FITC fluorophore.

DNA. Nevertheless, this study has remained the prevailing model to explain how XPA binds to the NER bubble[152]. We therefore set out to structurally characterize the interactions between XPA_{98-219} and a high affinity DNA substrate.

We began by setting up crystallization trials for human XPA_{98-219} in complex with a Y-shaped ssDNA-dsDNA junction substrate that contains an 8 basepair duplex extended by two non-complementary 12 nt ssDNA arms on one end of the duplex (Figure 2.1). After standard screening of conditions, crystals were obtained that diffracted to 2.2 Å. A concern arose during the course of refining the data when it was realized that based on the Matthews coefficient the volume of the unit cell was not sufficient to contain the mass of the protein and

the DNA substrate. The molecule in the crystal was assumed to be the protein because the volume of the asymmetric unit at 42 % solvent content could accommodate only one molecule of 15 kDa of XA_{98-219}. Moreover, an absorption peak at 9.67 keV indicated the presence of zinc in the crystal, presumably from the XPA_{98-219} zinc motif. Before progressing with further refinement, we decided to determine the affinity of XA_{98-219} for the DNA substrate.

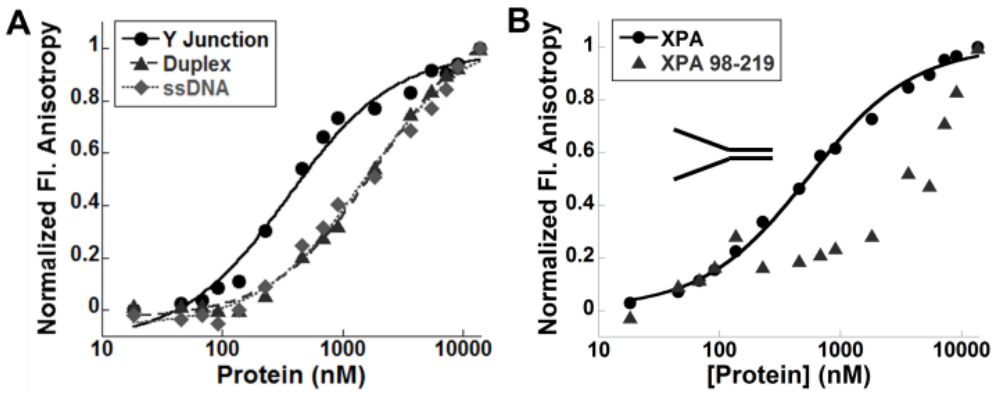


Figure 2.2: XPA_{98-239} does not have full DNA-binding capacity of XPA. A) Fluorescence anisotropy assay of the binding of XPA to Y-shaped ssDNA-dsDNA junction (black circles, solid line), duplex (gray triangles, dashed line) and ssDNA (light gray diamonds, dotted line). B) Comparison of the binding of the Y-shaped ssDNA-dsDNA junction by full-length XPA (black circles, solid line) and XPA_{98-219} (gray triangles). The concentration of the FITC tagged DNA substrate was 20 nM and measurements were performed at room temperature in a buffer containing 20 mM HEPES at pH 7.5, 75 mM KCl, 5 mM $MgCl_2$, 5% glycerol, and 1 mM dithiothreitol.

To this end, a fluorescence anisotropy assay (FA) was employed to directly compare the DNA-binding activity of full-length human XPA and XPA_{98-}

$_{219}$ (Figure 2.2). Fluorescein isothiocyanate (FITC) modified ssDNA, dsDNA, and Y-shaped ssDNA-dsDNA junction substrates (Figure 2.1) were used for these analyses. Substrates of 20 nucleotides were selected for this analysis as this corresponds to the approximate length of DNA predicted to be occluded by one molecule of XPA[37]. The results we obtained for full-length XPA were consistent with previous reports; Figure 2.2-A shows that XPA binds a Y-shaped ssDNA-dsDNA junction (0.29 ± 0.09 μM) with higher affinity than dsDNA (1.7 ± 0.6 μM) or ssDNA (1.5 ± 0.2 μM). In stark contrast, XPA_{98-219} had substantially weaker DNA binding affinity for all three substrates, so weak that it was not possible to extract a Kd value even for the highest affinity Y-shaped ssDNA-dsDNA junction (Figure 2.2-B). In order to verify that XPA_{98-219} was properly folded, a ^{15}N-enriched sample was prepared and a ^{15}N-^{1}H HSQC spectrum was acquired. Comparison to the previously reported spectra for this construct[39,45,52] confirmed that our sample of XPA_{98-219} was properly folded and free of aggregation (Figure 2.3, red spectrum). Taken together, these results demonstrate that XPA_{98-219} lacks critical elements necessary to reproduce the full DNA-binding activity of XPA. Thus, the widely accepted view that XPA_{98-219} is the DBD must be revised.

The observation that XPA_{98-219} does not recapitulate the full activity of XPA led us to perform analyses of the primary sequence to search for indications that other residues might contribute to DNA binding. Having previously showed

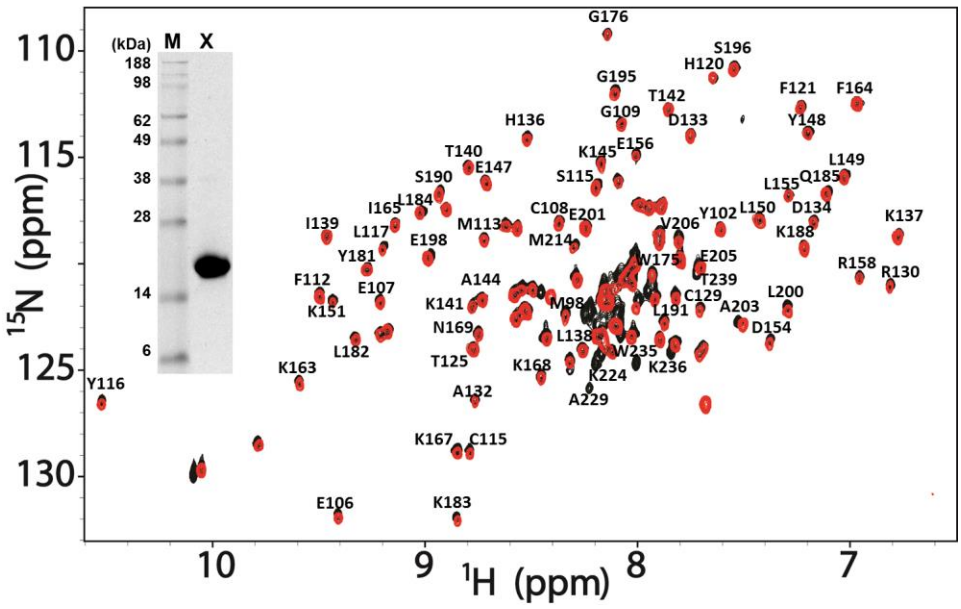


Figure 2.3: HSQC overlay of XPA$_{98-239}$ and XPA$_{98-219}$. Overlay of the 800 MHz ^{15}N-^{1}H TROSY-HSQC spectra of XPA$_{98-239}$ (black) and XPA$_{98-219}$ (red) acquired at 35 °C in a buffer containing 20 mM Tris at pH 7.0, 150 mM KCl, and 1 mM DTT. The inset shows SDS-PAGE of purified XPA$_{98-239}$. Lane M is the molecular weight marker and lane X is purified XPA$_{98-239}$.

that the N-terminal domain of human XPA is disordered[46], we focused on the sequence extending towards the C-terminus. Interestingly, secondary structure predictions indicated that the C-terminus of XPA$_{98-219}$ is located in the midst of a long helical element, with a high probability for helical secondary structure extending well beyond F219 (Figure 2.4). Moreover, there are several lysine and arginine residues in the region C-terminal to F219 that presumably enhance DNA binding affinity through electrostatic interaction with negatively charged DNA. Based on these insights, a series of C-terminally extended constructs were prepared (Figure 2.5). In all, six different human XPA constructs were cloned into

bacterial expression vectors. After screening for soluble expression in *E. coli*, the solubility and stability of each construct was assessed and on this basis, XPA_{98-239} was selected for further analysis.

```
Residue#:         10         20         30         40         50         60
Sequence: MAAADGALPE AAALEQPAEL PASVRASIER KRQRALMLRQ ARLAARPYSA TAAAATGGMA
Jufo9D  : EECCCCCCCC CCCCCCCCCC CHHHHHHHHH HHHHHHHHHH HHHHCCCCCH HHHCCCCCEE
Psipred : CCCCCCCCCC CCCCCCCCCC CHHHHHHHHH HHHHHHHHHH HHHHCCCCCH HHHHCCCCCC

Residue#:         70         80         90        100        110        120
Sequence: NVKAAPKIID TGGGFILEEE EEEEQKIGKV VHQPGPVMEF DYVICEECGK EFMDSYLMNH
Jufo9D  : EEECCCCCCC CCCCEEHHHH HHHHHHHHHH HCCCCCCHHH HHHHHHHCCC HHHHHHHHHH
Psipred : CCCCCCCCCC CCCCCCCCCC CHHHHHCCCC CCCCCCCCCC CCCCCCCCCC CCCHHHHHHH

Residue#:        130        140        150        160        170        180
Sequence: FDLPTCDNCR DADDKHKLIT KTEAKQEYLL KDCDLEKREP PLKFIVKKNP HHSQWGDMKL
Jufo9D  : CCEEEECCCC CCCCCHHHHH HHHHHHHHHH HHHHHHCCCC CCCEEECCCC CCCCCCHHHH
Psipred : CCCCCCCCCC CCCCCCCCCC HHHHHHHHCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCHH

Residue#:        190        200        210        220        230        240
Sequence: YLKLQIVKRS LEVWGSQEAL EEAKEVRQEN REKMKQKKFD KKVKELRRAV RSSVWKRETI
Jufo9D  : HHHHHHHHHH HHHCCCHHHH HHHHHHHHHH HHHHHHHHHH HHHHHHHHHH HHHHHHHCC
Psipred : HHHHHHHHHH HHHHCCHHHH HHHHHHHHHH HHHHHHHHHH HHHHHHHHHH HHHHCCCCCC

Residue#:        250        260        270
Sequence: VHQHEYGPEE NLEDDMYRKT CTMCGHELTY EKM
Jufo9D  : EEEEECCCCC CCCCCCEEEE EEECCCEEHH HHH
Psipred : CCCCCCCCCC CCCCCCEEEE CCCCCCEEEE EEC
```

Figure 2.4: Sequence analysis of XPA. Secondary structure prediction results from BCL::Jufo9D and Psipred with H for helical, E for extended, and C for coil conformation.

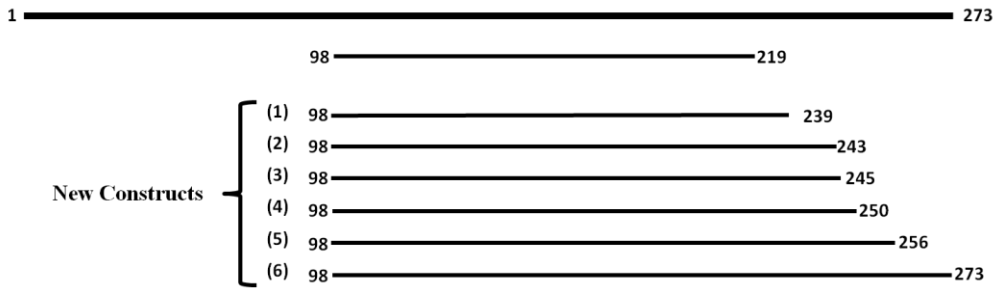


Figure 2.5: C-terminal extension XPA constructs. Top- full-length XPA, middle- XPA_{98-219}, and bottom- the six new C-terminally extended DNA binding domain constructs.

To determine if the extra C-terminal residues were important for binding DNA, the affinity of XPA_{98-239} for the 20 nt Y-shaped ssDNA-dsDNA junction, dsDNA and ssDNA substrates (Figure 2.1) was measured using the FA assay (Figures 2.6-A, 2.7). Notably, these data provided Kd values of 0.29 ± 0.08, 1.3 ± 0.2, and 1.5 ± 0.8 μM, respectively, very similar to those for the full-length XPA, including the preference for the Y-shaped ssDNA-dsDNA junction over dsDNA or ssDNA[37,125]. These results suggest XPA_{98-239} may be a more suitable model for XPA DBD than XPA_{98-219}.

To verify that XPA_{98-239} occupies a stable conformation and is not aggregated, ^{15}N-enriched XPA_{98-239} was prepared and a 2D ^{15}N-^{1}H HSQC spectrum was acquired. XPA_{98-239} is seen to have the characteristics of a stably folded 17 kDa protein, with narrow line widths and spectral dispersion evident in the ^{1}H dimension (Figure 2.6-B, black spectrum). An overlay of the spectra of XPA_{98-219} and XPA_{98-239} (Figure 2.3) strongly suggests they adopt a similar topology and that the new construct contains the globular core. Since many of the peaks overlap, a significant number of the previously reported resonance assignments for XPA_{98-219} could be transferred to XPA_{98-239}. There are 16 extra cross peaks in the HSQC spectrum of XPA_{98-239} and a limited number of them

could be tentatively assigned.

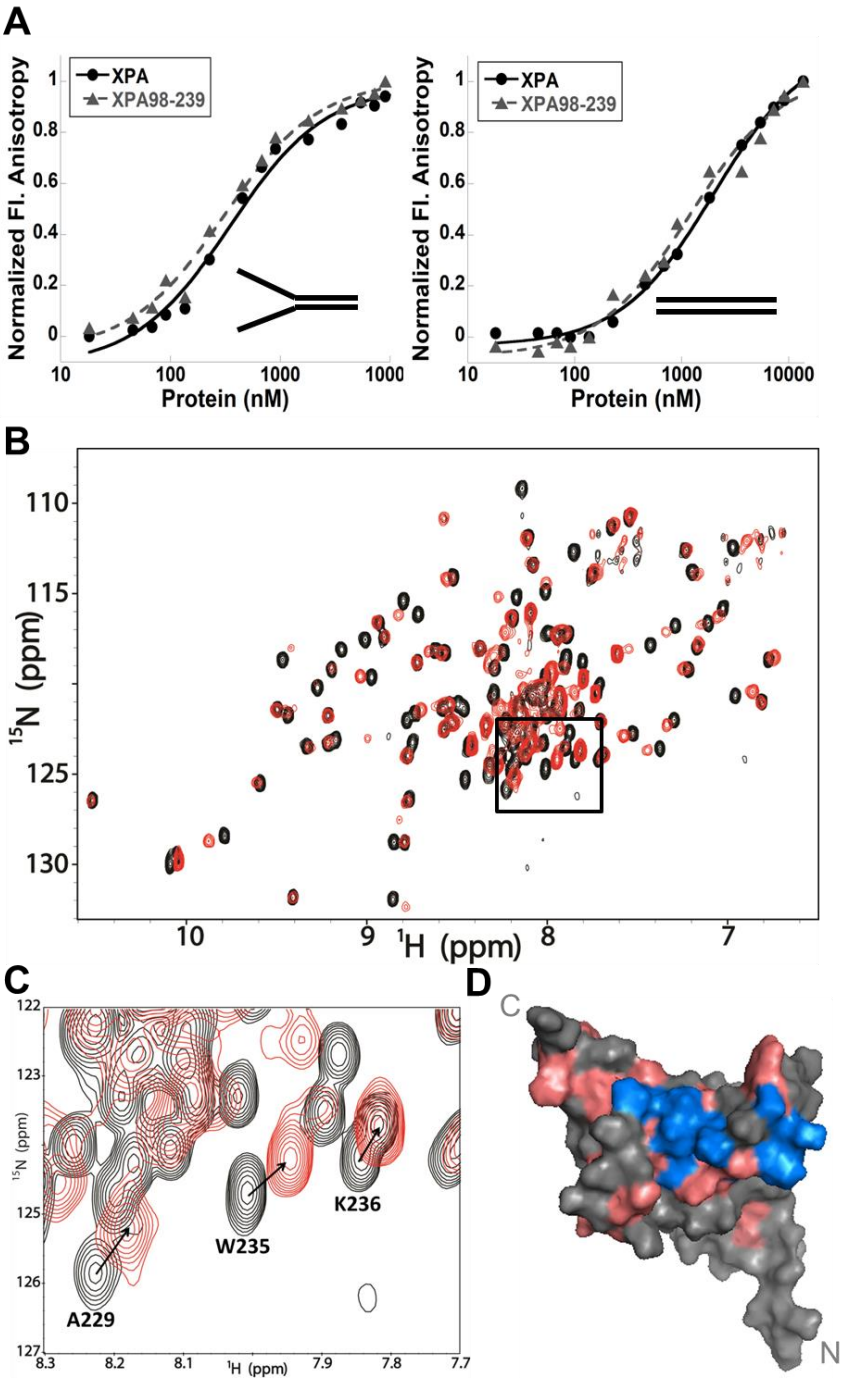


Figure 2.6: XPA_{98-239} exhibits the full-DNA binding of XPA. A) Fluorescence anisotropy assay of the binding of the Y-shaped ssDNA-dsDNA junction (left) and duplex (right) substrates by full-length XPA (black circles, solid line) and XPA_{98-239} (gray triangles, dashed line). The conditions were the same as in Figure 1. B) 900 MHz ^{15}N-^{1}H TROSY HSQC spectra of XPA_{98-239} obtained in the absence (black) and presence (red) of an equimolar amount of Y-shaped ssDNA-dsDNA junction substrate. The data were acquired at 35 °C in a buffer containing 20 mM Tris at pH 7.0, 150 mM KCl, 1 mM DTT. C) Zoomed-in view of the boxed region of B) showing perturbations of cross peaks from A229, W235, and K236 in the C-terminal extension. D) Map of NMR chemical shift perturbations on a surface representation of XPA_{98-219} (PDB ID: 1d4u). Residues identified in the study of XPA_{98-219} binding a 9-nt ssDNA substrate[18] are colored blue. Additional residues with significant perturbations in the study of XPA_{98-239} binding the Y-shaped ssDNA-dsDNA junction substrate are c0lored salmon. See supplementary Methods for a detailed description of how residues with significant perturbations were identified.

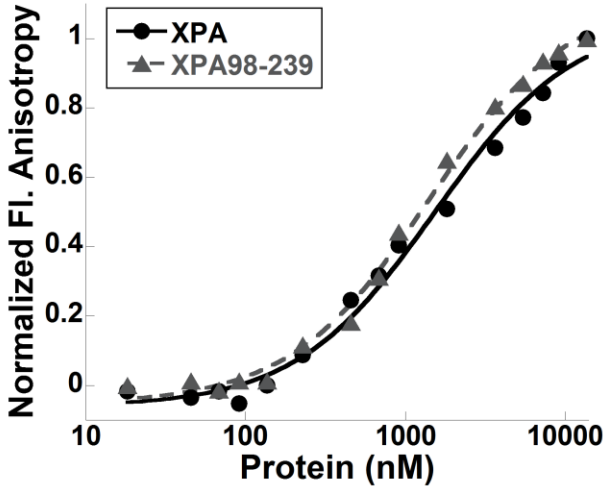


Figure 2.7: Binding of FL-XPA and XPA_{98-239} to ssDNA. Fluorescence anisotropy assay of ssDNA binding by XPA (black circles, solid line) and XPA_{98-239} (gray triangles, dashed line). The concentration of the FITC tagged DNA substrate was 20 nM and measurements were performed at room temperature in a buffer containing 20 mM HEPES at pH 7.5, 75 mM KCl, 5 mM $MgCl_2$, 5% glycerol, and 1 mM DTT.

These cross peaks from the C-terminal extension have the same line shape as the other peaks in the spectrum and are narrowly dispersed. Although consistent with the prediction of helical character, the available data are not sufficient to formally assign the structure of the C-terminal extension.

To further characterize the interaction of XPA_{98-239} with DNA, we monitored a titration of the Y-shaped ssDNA-dsDNA junction substrate using 2D ^{15}N-^{1}H HSQC NMR. This analysis showed perturbation of a select number of cross peaks in the spectrum that saturate at a ratio of ~1:1 (Figure 2.8), consistent with specific binding of this DNA substrate with low μM affinity. Comparison to the corresponding titration of XPA_{98-219} confirms that XPA_{98-239} binds the substrate much more strongly; the shorter construct is far from saturation at the 1:1 ratio and in fact does not saturate even at a substrate ratio of 5:1 (Figure 2.8). These observations support the proposal that the XPA DBD had been incorrectly assigned.

The NMR titration data also enabled us to test the validity of the previous model for the DNA binding site of XPA. In the previous study of a 9-nt ssDNA substrate binding to XPA_{98-219}, chemical shift perturbations in fast exchange between the free and bound states were observed for 13 residues and 3 others were exchange broadened. Cross peaks from a larger number of residues are perturbed in the titration with the much higher affinity 20-nt Y-shaped ssDNA-dsDNA junction substrate. Consistent with the higher affinity for the Y-shaped

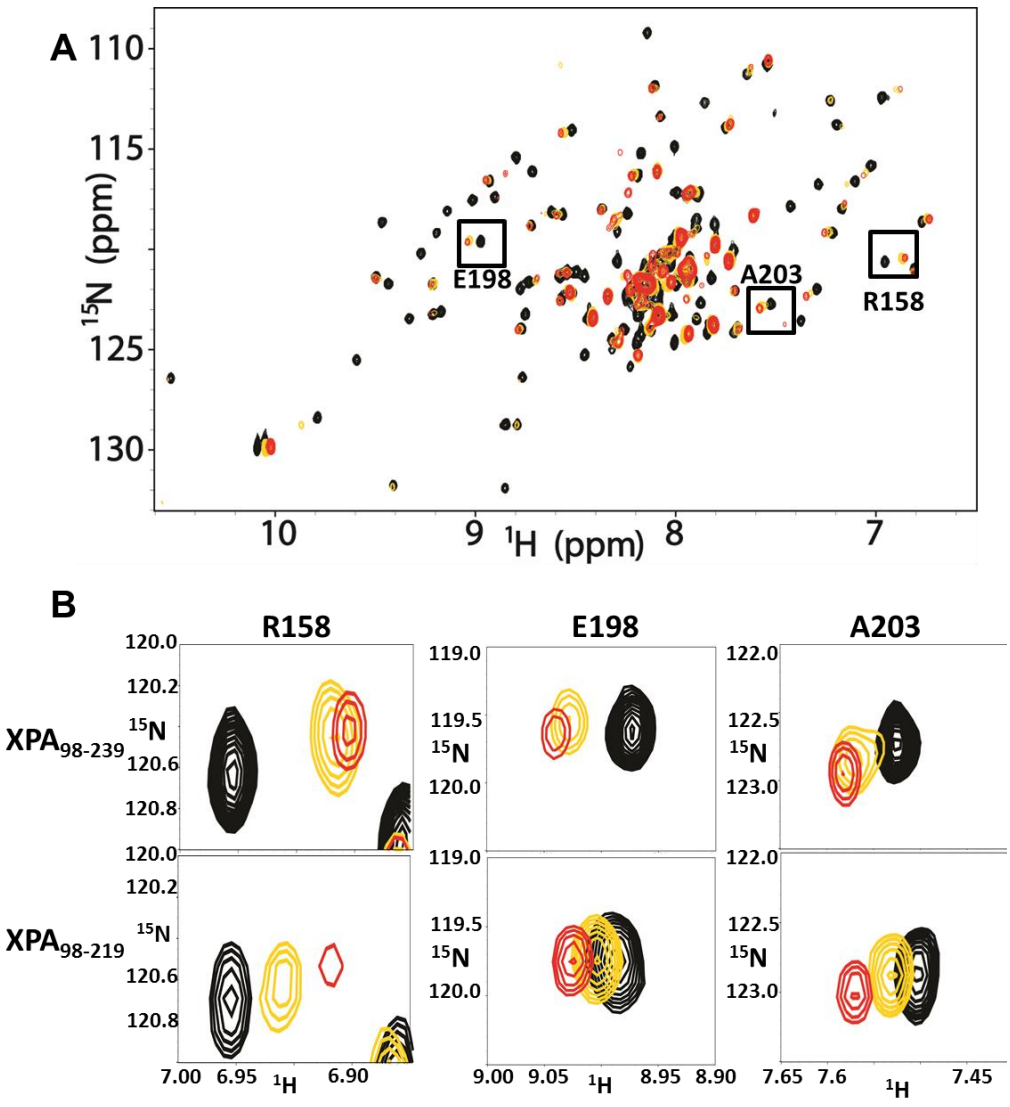


Figure 2.8: DNA-binding comparison of XPA_{98-239} and XPA_{98-219} by NMR titration. NMR analysis of the binding of Y-shaped ssDNA-dsDNA junction substrate. A) 900 MHz ^{15}N-^{1}H TROSY-HSQC titration of XPA_{98-239} with Y-shaped ssDNA-dsDNA substrate acquired at DNA: protein ratios of 0:1 (black), 1:1 (gold), and 4:1 (red) acquired at 35 °C in a buffer containing 20 mM Tris at pH 7.0, 150 mM KCl, and 1 mM DTT. B) Top - Zoomed-in view of the boxed regions in A). Bottom corresponding regions from the 800 MHz ^{15}N-^{1}H TROSY-HSQC titration of XPA_{98-219} acquired under the same conditions as A), but with the final DNA: protein ratio set to 5:1 (red).

substrate, both chemical shift perturbations in fast exchange and line broadening of resonances in intermediate exchange were observed. Figure 2.6-D maps the residues exhibiting significant perturbations on the previously determined NMR structure of the globular core; beyond the residues previously assigned to the DNA binding site in the study of XPA_{98-219} with 9 nt ssDNA (blue), the titration with Y-shaped substrate identified many additional perturbed residues (salmon). The latter include several additional residues in and around the basic cleft (residues L191, K204, and R207). One additional critical observation was the perturbation of cross peaks from three residues in the C-terminal extension (A229, W235, and K236, Figure 2.6-C), which strongly supports our proposal of the need for the C-terminal extension for full DNA binding activity.

Our results show that XPA_{98-239} contains the full DNA- binding apparatus of human XPA, thereby redefining the XPA DBD. Mutations of residues between F219 and T239 are associated with severe *XP* disorders, which implies this region of the protein is critical to the function of XPA[13]. The incorrect assignment of XPA_{98-219} as the DBD may help explain the lack of substantial progress in elucidating the molecular mechanisms of XPA action in NER over the past 20 years[152]. Moreover, our studies of the more physiologically relevant ssDNA-dsDNA junction substrate clearly demonstrate the previous model for the XPA DNA binding site was incomplete. The new XPA_{98-239} DBD provides an excellent target for high resolution structural and biophysical investigations of the XPA-

DNA complex that can better define its role in NER. Additionally, as increased cellular NER activity is often associated with loss of effectiveness of multiple classes of current anticancer treatments, such as radiation therapy and cisplatin[153], XPA has been identified as a possible target for therapeutic intervention due to its critical role in NER[154]. The availability of structural information greatly enhances the pace of drug discovery. Hence, given the high quality NMR data presented here, XPA_{98-239} has potential as a valuable reagent for structural analyses directed to the design and validation of novel small molecule inhibitors and probes.

Materials and Methods

Protein Construction

Table 2.1: PCR primers for XPA constructs.

Oligo	Sequence (5' -> 3')
XPA_FW	aaaaaggatccatggcggcggc
XPA_REV	tttttgcggccgcctcatcacatttttttcatatgtca
XPA98-239_FW	gcgggatccatggaatttgattatgtaatatgcg
XPA98-239_REV	taattgcggccgctcacgtctccctttttccacac
XPA98-219_FW	aaaaaggatccatggaatttgattatgtaata
XPA98-219_REV	taattgcggccgcctcatcaaaatttcttctgtttcattt

XPA constructs were amplified by PCR using indicated oligonucleotides (Table 2.1) to introduce 5′ *BamHI* and 3′ *NotI* cleavage sites. Proteins were cloned into

the pBG100 in-house expression vector (L.S. Mizoue, Center for Structural Biology, Vanderbilt University), which incorporates an N-terminal human rhinovirus 3C (HRV3C) protease cleavable 6xHis tag.

Protein Production

All XPA constructs were overexpressed in BL21(DE3) cells. Cells were grown in either terrific broth or a minimal medium containing 0.5 g/L $^{15}NH_4Cl$ (CIL, Inc.) at 37 °C to an $OD_{600\ nm}$ of ~0.6-0.8, then transferred to 18 °C. Protein expression was induced at $OD_{600\ nm}$ ~1.0 by adding isopropyl thio-β-D-galactopyranoside to 0.2 mM and proceeded for ~16 hours. The purification buffer was adjusted to pH 7.0 for XPA_{98-239} and 8.0 for XPA and XPA_{98-219}. All constructs were purified using Ni-NTA chromatography (Sigma) using standard procedures. The 6xHis fusion tag was removed by H3C protease cleavage followed by re-pass over Ni-NTA resin. Size exclusion chromatography was used as the final purification step using either S200 (XPA) or S75 (XPA_{98-219} and XPA_{98-239}) resin (GE Healthcare).

Preparation of DNA Substrates

Figure 2.1 shows the structures of DNA substrates used in this study. The position of the FITC tag is indicated. The DNA substrate used for the NMR study was not tagged. Desalted oligodeoxynucleotides were purchased from Sigma-Aldrich Co. (St. Louis, MO). Y-shaped ssDNA-dsDNA junctions and duplexes were prepared

by mixing equimolar amount of each strand in TNE buffer (10 mM Tris, 50 mM NaCl, 1 mM EDTA, pH 7.0), then heating in a boiling water bath and allowing the solution to cool to room temperature.

Sequence Analysis

The primary amino acid sequence of XPA was analyzed with a number of secondary structure and disorder prediction algorithms, with the final conclusions drawn from BCL::Jufo9D[155] and Psipred[156].

Fluorescence Anisotropy DNA Binding Assay

Protein and DNA substrate were diluted in binding buffer (20 mM HEPES, pH 7.5, 75 mM KCl, 5 mM $MgCl_2$, 5% glycerol, 1 mM dithiothreitol). Diluted protein was dispensed into Corning #3676 low-volume 384-well microtiter plates. DNA substrate was subsequently added to each well at a concentration of 20 nM. The plate was covered from light and incubated at room temperature for 5 minutes prior to measuring fluorescence anisotropy to ensure samples were homogenized and had reached equilibrium. The fluorescence anisotropy was measured at room temperature using a Synergy H1 plate reader (Bio-Tek) equipped with a GreenFP polarization filter cube set (λ_{Ex} = 485 nm, λ_{Em} = 528 nm). Each binding measurement was performed in triplicate for each DNA substrate. Apparent dissociation constants (K_d) were determined for each

individual titration by plotting fluorescence anisotropy against protein concentration and fitting to a simple two-state binding model using KaleidaGraph (v4.03). The K_d values are reported as the mean and standard deviation of at least two independent measurements made on separate days from separate preparations.

NMR Spectroscopy

The NMR samples were concentrated to 50 – 100 μM in the stated buffer to which 5% ^2H_2O was added prior to performing the experiment. Titrations were performed by preparing two identical solutions of protein, one with no DNA substrate and the other at the highest ratio of DNA to protein, then collecting spectra for these and intermediate ratios created by mixing of the two solutions. ^{15}N-^1H HSQC and TROSY-HSQC spectra were recorded in 3 mm tubes at 35 °C using Bruker *AVANCE* 800 or 900 MHz NMR spectrometers equipped with a cryoprobe. All data were processed and analyzed using SPARKY (Goddard, T.D. & Kneller, D. G. SPARKY 3, University of California, San Francisco.). All residues whose cross peaks disappeared due to intermediate exchange line broadening upon binding DNA were placed in the category of significantly perturbed. In addition, for residues exhibiting chemical shift perturbations in fast exchange, we calculated the change in chemical shift (Δδ) from the spectra acquired with no substrate and a 1:1 ratio using the formula: $\Delta\delta = \sqrt{[(H_a - H_b)^2 +}$

$(0.2*(N_a - N_b))^2]$. The threshold for significant chemical shift perturbation was set to the average $\Delta\delta$ + one standard deviation.

Acknowledgements

We thank Nicholas P. George, Matthew K. Thompson, and Brandt F. Eichman for assistance in analyzing X-ray diffraction data, Markus Voehler for valuable insights for NMR experiments, Joshua Bauer and Sonja Brooks for their help with FA assay, and M. Wade Calcutt and Anindita Basu for assistance with mass spectroscopy. This work was supported by funding from NIH grants R01 ES1065561 and PO1 CA092584. Access to facilities was supported by P30 ES00267 to the Vanderbilt Center in Molecular Toxicology and P30 CA068485 to the Vanderbilt Ingram Cancer Center. SMS is supported by postdoctoral fellowship 119569-PF-11-271-01-DMC from the American Cancer Society. Support for acquisition and upgrade of the NMR instrumentation was obtained from shared instrumentation grants from the NIH (S10 RR025677) and NSF (DBI-0922862).

CHAPTER III

INTERACTIONS OF HUMAN XPA WITH DNA

Abstract

Xeroderma pigmentosum complementation group A (XPA) is an essential scaffolding protein in the multi-protein nucleotide excision repair (NER) machinery. The interaction of XPA with DNA is a core function and a number of mutations in the DNA binding domain are associated with *XP* disease. Although NMR structures of the core globular domain of human XPA and complementary data on DNA binding have been available for many years, the molecular details of how human XPA binds DNA remain unclear. Insights have been obtained from X-ray co-crystal structures of the central globular domain of the yeast XPA (Rad14) in the presence of DNA, but it was unclear if these structural models represent DNA binding of XPA in the context of human NER. In order to better understand the DNA binding activity of human XPA in NER, we used NMR to investigate the interaction with DNA of the human XPA DNA binding domain (DBD). The data show that XPA binds different ss-ds junction DNA substrates similarly. Comparisons to the crystal structures of Rad14-DNA complexes revealed similarities and differences between binding of DNA, including direct evidence of a significant role for the residues extending C-terminally from the

globular core. A key site in Rad14, F262, was previously reported as critical to DNA binding, but mutation of the corresponding W175 in human XPA had only a moderate effect on DNA binding. We also obtained insights into the molecular basis for XPA malfunction in disease-associated mutations in the DBD, suggesting a correlation may exist between the effect of mutations on DNA binding affinity and the severity of symptoms in *XP* patients.

Introduction

Nucleotide excision repair (NER) is a DNA damage repair pathway specialized for removing bulky lesions arising from exposure to various types of endogenous and exogenous toxic agents[1,2,9,157,158]. Human NER is a multi-step process involving coordinated action of over 30 proteins[3,157]. Two NER pathways exist, one repairs lesions in actively transcribed DNA (transcription coupled repair (TCR)) and the other processes lesions more generally throughout the genome (global genome repair (GGR)). These differ only in the mechanism by which the presence of damage is recognized. The subsequent steps, destabilization and unwinding of the DNA by transcription factor II H (TFIIH) to create a DNA structure termed the NER bubble, excision of the damaged nucleotide, and gap-filling synthesis, are understood to be the same[3,16,20–30,38,159–161]. XPA is recruited to the damage site by TFIIH once the duplex is unwound[3,38].

While XPA does not have any enzymatic activity, it acts in concert with replication protein A (RPA) as a critical scaffolding protein through its interactions with DNA and other NER proteins[162].

Defects in NER result in the genetic disorder *Xeroderma pigmentosum (XP)*, which is characterized by hypersensitivity to sunlight and increased incidence of skin cancer[4–7]. In severe cases, neurological defects are also observed[4,11–13]. *XP* arises from mutations in 8 genes, 7 of which (*XPA*-*G*) are directly involved in NER[18,157]. Among these, XPA mutations are usually associated with most severe disease phenotypes, many of which map to the DNA binding domain (DBD)[11–13]. Different XPA mutations lead to different disease phenotypes[4,12,13,107]; complete loss, substantial truncation, and unfolding of the DBD are known to lead to severe *XP* disease phenotypes. However, genotype-phenotype correlations and mechanisms behind missense mutations remain poorly understood[8,13,110,162,163].

Two structures of the globular core of the *S. cerevisiae* homolog of XPA, Rad14, have been determined for complexes with a DNA duplex containing a cisplatin 1,2-deoxydiguanosine intrastrand crosslink or a N-(deoxyguanosin-8-ly)-2-acetylaminofluorene (AAF) lesion[55]. Rad14 has remarkably higher affinity for these two modified duplexes than for unmodified duplexes or duplexes containing other lesions[55]. The structure of the Rad14 globular core is quite similar to the previously reported structures of the globular core of human XPA, obtained in the

absence of DNA[39,40]. However, we and others have shown that the globular core of human XPA (XPA_{98-219}) does not bind DNA with appreciable affinity, and that a significant number of additional C-terminal residues are required for full DNA binding affinity[53,54]. Consequently, the high binding affinity for the two lesion-containing duplexes of the Rad14 globular core that lacks additional C-terminal residues is puzzling and suggests that the Rad14 and XPA DBDs are not functionally equivalent. The uncertainty is particularly confounding in the context of NER because XPA is not recruited until after the presence of damaged DNA is recognized and unwound.

Here we report an investigation of the interaction of human XPA with model NER bubble substrates using the XPA_{98-239} construct (XPA DBD) that contains the full DNA binding affinity. We used NMR and DNA affinity measurements to define the binding site of XPA DBD for model NER substrates, and compared our findings to the structures of Rad14-DNA complexes. To confirm our findings, we determined the effect of selected structure-based and disease-associated mutants on DNA binding. The results revealed differences between the Rad14 structures and how human XPA interacts with the bubble and functions in NER, and insights into the molecular basis for disease association for certain XPA mutations.

Results

Binding affinities of human XPA for ss-ds DNA junction substrates

It had been established previously that XPA binds to a ss-dsDNA junction in the NER bubble[37], although controversy remains over whether it is the 5′ or 3′ junction. Moreover, the footprint of XPA on the junction remains unclear. To address these issues, we measured the affinities of the XPA DBD for different junction structures and lengths. We have previously used a DNA fluorescence anisotropy (FA) assay to characterize affinities, but turned instead here to microscale thermophoresis (MST) as a result of its higher precision and the tendency of XPA-DNA complexes to aggregate in the 384 well plates during the course of FA experiments. Comparisons of the values obtained by the two methods revealed the same trends among substrates but higher dissociation constants (K_d) for MST by ~5-fold. The systematically weaker binding in the MST experiments is attributable to the higher ionic strength of the buffer.

We first confirmed that XPA DBD binds to a previously characterized 8/12 splayed-arm substrate (Figure 3.1, Y-shaped junction with an 8 base pair (bp) duplex and 12 nucleotide (nt) 5′ and 3′ overhangs) with high affinity[53].

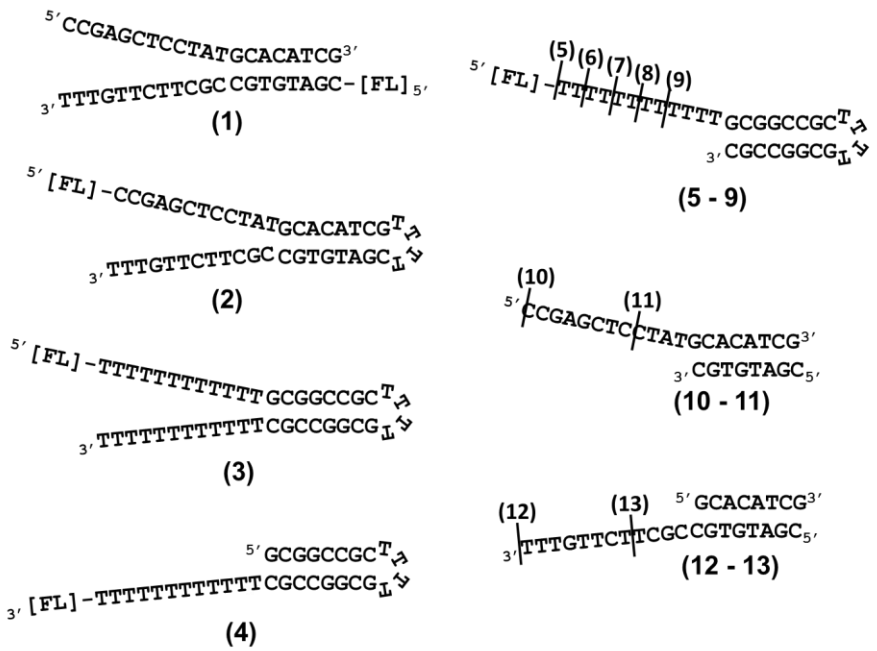
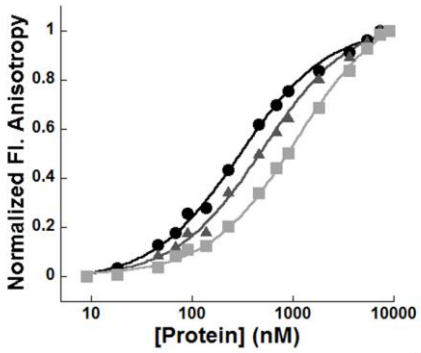


Figure 3.1: DNA Substrates. Structures and sequences of ss-dsDNA junction substrates used for binding assays and NMR analyses. The names are based on the number of basepairs in the duplex region followed by the number of nucleotides in the overhang: (1) 8/12 splayed-arm, (2) 8/12 HP splayed-arm with mixed sequence, (3) 8/12 HP splayed-arm, (4) 8/12 HP 3′ overhang, (5) 8/12 HP 5′ overhang, (6) 8/10 HP 5′ overhang, (7) 8/8 HP 5′ overhang, (8) 8/6 HP 5′ overhang, (9) 8/4 HP 5′ overhang, (10) 8/12 5′ overhang, (11) 8/4 5′ overhang, (12) 8/12 3′ overhang, (13) 8/4 3′ overhang. All hairpins (HP) are composed of four Ts. The positions of fluorescein tags are indicated by [FL]. Substrate (1) without fluorescein tag was employed for the in-depth NMR analysis.

We next sought to find a junction DNA substrate that is optimized in length and shape so that it interacts with XPA in a manner that avoids non-specific secondary binding to the substrate. In order to systematically screen the types of junction and the length of overhangs, a series of DNA substrates were designed containing

a GC-rich duplex (sealed with a 4 nt hairpin (HP) for stability) and oligo-dT overhangs (Figure 3.1). We found a modest but clearly significant difference in affinity for different nucleotide sequences (Figure 3.2).

A)



B)

Substrate	Kd (μM)
1 (●)	0.30 ± 0.02
2 (▲)	0.51 ± 0.04
3 (■)	1.10 ± 0.05

Figure 3.2: Structure and Sequence Dependence of XPA DBD Binding of Substrates. (A) Binding of XPA DBD to 8/12 splayed-arm (circle), 8/12 HP splayed-arm with mixed sequence (triangle), and 8/12 HP splayed-arm (square) (substrates 1, 2, and 3 in Figure 3.1, respectively) as determined by fluorescence anisotropy binding assay. Error bars represent standard deviation from triplicate experiments. (B) Dissociation constants (Kd) determined from each curves in panel A.

The data show that XPA DBD binds both 3′ and 5′ overhang substrates with approximately the same affinity as Y-shaped substrates (Figure 3.3-A, Table 3.1). It is interesting that the length of the overhang could be shortened to 4 nts without any significant effect on the affinity for substrate (Figure 3.3-B, Table

3.1). The shorter overhang was advantageous for NMR studies, so we selected an 8/4 5′ overhang substrate for detailed analysis.

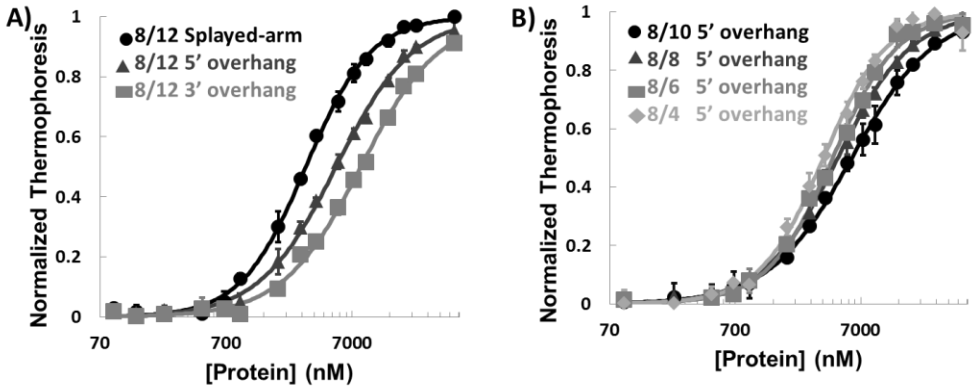


Figure 3.3: XPA Binding to ss-ds Junction DNA Substrates. (A) Plot of MST data for XPA DBD binding 8/12 splayed-arm (circle), 8/12 5′ overhang (triangle) and 8/12 3′ overhang (square) (substrates 3, 5 and 4 in Figure 3.1, respectively). (B) Plot of MST data for XPA DBD binding DNA substrates with 8 nt duplex and different lengths of 5′ overhangs (substrates 6–9 in Figure 3.1). All measurements were made at room temperature in a buffer containing 50 mM Tris-HCl at pH 7.8, 150 mM NaCl, 10 mM $MgCl_2$, 0.05 % Tween-20 and 1 mM DTT.

Table 3.1: Kd of XPA Binding to ss-ds Junction DNA Substrates.
Dissociation constants determined from the plots in panels A and B in Figure 3.3. All measurements were made at room temperature in a buffer containing 50 mM Tris-HCl at pH 7.8, 150 mM NaCl, 10 mM $MgCl_2$, 0.05 % Tween-20 and 1 mM DTT.

Substrate	Kd (μM)
8/12 HP splayed-arm	3.0 ± 0.1
8/12 HP 5′ overhang	5.2 ± 0.2
8/12 HP 3′ overhang	8.2 ± 0.5
8/10 HP 5′ overhang	5.9 ± 0.2
8/8 HP 5′ overhang	4.6 ± 0.1
8/6 HP 5′ overhang	4.2 ± 0.2
8/4 HP 5′ overhang	3.5 ± 0.2

Structural analysis of the interaction of human XPA with DNA

The binding of DNA substrates by XPA DBD (XPA_{98-239}) was investigated using NMR spectroscopy. The first step in any detailed analysis is the assignment of the NMR signals (resonances) to specific atoms within the molecule. The NMR backbone resonance assignments for XPA DBD were obtained using a standard series of double and triple resonance 2D and 3D experiments[164]. This analysis produced assignments for 96%, 96%, 92%, 96%, and 94% of the ^{15}N, ^{1}H, ^{13}CO, $^{13}C\alpha$ and $^{13}C\beta$ resonances, respectively (Table 3.2). Figure 3.4 shows the 2D ^{15}N-^{1}H HSQC spectrum labeled with the corresponding backbone peak assignments used for the titration analyses of the binding of DNA substrates. Due to the extensive overlap in the spectrum, assignment of the

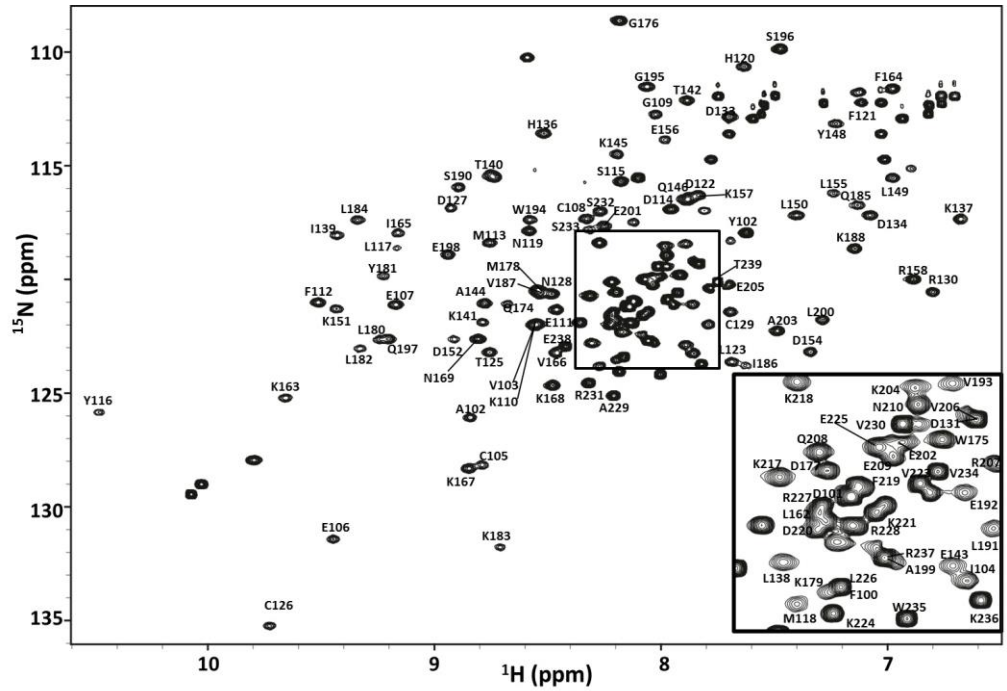


Figure 3.4: NMR Backbone Resonance Assignment of XPA DBD. Region from the 600 MHz ^{15}N-^{1}H HSQC spectrum of XPA DBD acquired at 25 °C in a buffer containing 20 mM Tris at pH 7.0, 500 mM KCl, 1 mM TCEP and 5 % $^{2}H_2O$. To avoid cluttering of the figure, some assignment labels were removed. The inset is an expansion of the central region within the rectangle.

stretch of residues between Q208 to R228 was challenging. In this region, there are groups of residues that connect well; however, there are some breaks in between these stretches that creates more than one possibility for the assignments that makes sense. In this chapter, we employed the assignment that is most self-consistent for the analysis of NMR titration data described below.

Table 3.2: List of Backbone Resonance Assignment of XPA DBD.

Residue #	Assignment #	Atom	^{15}N (ppm)	^{1}H (ppm)	CO (ppm)	Cα (ppm)	Cβ (ppm)
His-1	1	G					
His-2	2	P			177.527	63.533	32.136
His-3	3	G	110.299	8.653	174.258	45.321	
His-4	4	S	115.578	8.161	174.631	58.417	63.71
98	5	M	121.95	8.419	177.6	55.632	32.707
99	6	E	121.505	8.274	175.65	56.742	30.152
100	7	F	119.381	7.895	174.709		40.123
101	8	D	121.248	8.207	175.022	53.952	41.206
102	9	Y	118.01	7.683	174.985	57.509	39.808
103	10	V	121.979	8.61	173.652	60.545	33.118
104	11	I	123.328	7.919	175.919	58.969	37.713
105	12	C	128.223	8.854	178.312	59.895	30.999
106	13	E	131.505	9.509	176.128	58.346	29.959
107	14	E	121.17	9.236	177.55	58.009	30.733
108	15	C	117.376	8.388	177.421	59.078	32.97
109	16	G	112.778	8.095	173.548	46.552	
110	17	K	122.075	8.62	176.321	56.257	33.501
111	18	E	121.368	8.525	177.049	55.97	30.383
112	19	F	121.052	9.576	172.18	56.354	41.38
113	20	M	118.485	8.819	175.026	56.555	34.271
114	21	D	116.936	8.018	174.776	53.346	44.182
115	22	S	115.75	8.248	172.442	56.162	66.315
116	23	Y	125.94	10.55	178.576	62.626	39.476
117	24	L	118.708	9.236	179.621	57.859	40.756
118	25	M	123.924	8.34	178.733	57.934	32.627
119	26	N	117.906	8.645	177.425	56.094	38.447
120	27	H	110.576	7.698	175.319	57.938	29.591
121	28	F	111.855	7.206	174.34	57.061	40.975
122	29	D	116.43	7.902	174.54	55.885	39.459
123	30	L	123.651	7.744		51.589	44.619
124	31	P			173.514	62.586	28.028
125	32	T	123.279	8.819	174.631	61.362	71.847
126	33	C	135.292	9.797	174.917	58.544	30.548
127	34	D	116.912	8.989	178.12	57.351	40.263
128	35	N	120.704	8.547	176.661	56.266	38.935
129	36	C	121.521	7.759	174.848	61.368	30.678
130	37	R	120.61	6.863	176.714	56.853	30.106
131	38	D	123.466	8.226	176.486	55.584	42.184
132	39	A	126.113	8.906	177.172	53.971	19.013
133	40	D	112.934	7.763	176.313	55.368	41.965
134	41	D	117.235	7.144		55.499	
135	42	K			178.229	60.756	32.955
136	43	H	113.643	8.581	173.435	55.565	29.997
137	44	K	117.42	6.739	176.176	56.434	33.598
138	45	L	122.926	8.366	176.94	53.291	
139	46	I	118.114	9.492	174.075	59.001	42.316

140	47	T	115.451	8.81	175.879	61.911	71.052
141	48	K	121.956	8.851	177.87	60.487	33.533
142	49	T	112.155	7.95	176.324	66.809	68.932
143	50	E	122.934	7.957	178.529	59.721	29.702
144	51	A	121.125	8.843	179.393	55.275	18.44
145	52	K	114.533	8.257	176.836	60.669	32.397
146	53	Q	116.514	7.943	177.964	58.748	29.498
147	54	E	115.577	8.801	177.323	57.526	29.866
148	55	Y	113.212	7.29	172.449	58.121	37.234
149	56	L	115.574	7.036	176.154	55.633	39.186
150	57	L	117.262	7.464	176.313	52.931	44.952
151	58	K	121.369	9.495	178.551	54.488	35.634
152	59	D	122.697	8.979	179.52	59.122	39.878
153	60	C	115.154	8.621	176.364	60.197	
154	61	D	123.277	7.406	176.144	57.186	42.887
155	62	L	116.249	7.306	177.633	56.8	42.637
156	63	E	113.895	8.048	179.545	57.337	32.442
157	64	K	116.531	7.946	176.684	56.147	33.333
158	65	R	120.047	6.948	174.543	56.511	30.355
159	66	E	122.302	8.255		54.391	32.641
160	67	P					
161	68	P			177.481	62.985	31.932
162	69	L	122.38	8.214	178.059	55.237	42.446
163	70	K	125.273	9.724	173.94	56.163	34.427
164	71	F	111.647	7.042	175.617	54.834	41.384
165	72	I	118.005	9.232	174.925	59.514	
166	73	V	123.288	8.522	176.014	61.439	33.151
167	74	K	128.394	8.914	175.156	54.865	35.313
168	75	K	124.7	8.539	176.332	56.079	32.698
169	76	N	122.668	8.88		51.253	39.472
170	77	P					
171	78	H			177.621	58.09	32.65
172	79	H	120.334	8.129		56.441	29.188
173	80	S			175.256	59.509	63.626
174	81	Q	121.085	8.736	176.088	56.632	28.732
175	82	W	119.848	7.979	176.799	56.976	29.864
176	83	G	108.651	8.254	173.663	45.556	
177	84	D	120.646	8.259	176.076	54.625	41.515
178	85	M	120.52	8.604	174.424		35.007
179	86	K	123.609	8.252	174.968	56.343	
180	87	L	122.722	9.296	175.046	54.005	44.476
181	88	Y	119.88	9.291	174.999	56.557	42.786
182	89	L	123.095	9.386	178.243	54.746	41.5
183	90	K	131.862	8.778	177.686	61.742	32.163
184	91	L	117.438	9.397	179.859	58.723	41.862
185	92	Q	116.779	7.191	178.552	58.394	27.874
186	93	I	123.882	7.689	177.298	60.927	34.341
187	94	V	120.707	8.6	179.147	66.815	31.617
188	95	K	118.67	7.212	178.723	60.06	32.368

189	96	R	119.42	8.038	177.954	56.669	27.915
190	97	S	116.016	8.956	176.179	62.631	
191	98	L	121.937	7.855	180.738	57.775	41.212
192	99	E	121.163	7.926	178.823	58.977	29.665
193	100	V	118.519	7.954	178.033	65.827	32.44
194	101	W	117.443	8.639	177.696	58.809	30.115
195	102	G	111.573	8.129	174.109	46.689	
196	103	S	109.931	7.54	173.65	57.132	65.676
197	104	Q	122.665	9.262	177.975	58.459	28.133
198	105	E	118.993	9.001	178.869	60.71	28.635
199	106	A	122.793	8.109	180.71	54.969	19.065
200	107	L	121.845	7.349	177.668	58.095	41.045
201	108	E	117.687	8.313	179.887	59.422	28.356
202	109	E	120.033	8.098	178.352	59.264	29.307
203	110	A	122.349	7.552	180.727	54.749	18.645
204	111	K	118.703	8.044	178.883	59.722	32.427
205	112	E	120.268	7.769	178.711	58.865	28.95
206	113	V	119.241	7.91	177.882	64.754	31.933
207	114	R	120.449	7.843		58.057	30.11
208	115	Q			175.267	56.34	30.481
209	116	E	118.305	7.762	175.387	55.883	29.779
210	117	N	117.843	8.403	176.714		32.917
211	118	R	121.939	8.188	175.92	56.846	32.69
212	119	E	121.219	8.197	178.032	58.173	30.387
213	120	K	120.202	8.28	178.041	57.956	29.602
214	121	M	120.264	8.097	177.952	57.768	32.432
215	122	K	119.024	8.039	177.178	56.502	32.36
216	123	Q	120.999	8.008	177.234	57.342	32.615
217	124	K	120.081	8.134	177.296	57.561	28.53
218	125	K	120.773	8.377	177.809	58.271	29.676
219	126	F	118.425	8.335	176.633	54.652	38.433
220	127	D	121.018	8.184	175.555	58.028	39.477
221	128	K	122.049	8.281	176.351	56.94	32.85
222	129	K	121.547	8.134	176.91	56.924	32.802
223	130	V	121.73	8.275	177.178	56.949	32.448
224	131	K	120.94	8.032	176.585	63.208	32.747
225	132	E	124.097	8.245		56.898	32.747
226	133	L					
227	134	R			177.008	63.832	32.148
228	135	R	117.486	8.184	176.122	56.224	30.81
229	136	A	125.164	8.271	177.776	52.531	19.239
230	137	V	119.463	8.074	176.391	62.497	32.64
231	138	R	124.598	8.38	176.335	56.213	30.792
232	139	S	117.062	8.33	174.703	58.383	63.908
233	140	S	117.894	8.373	174.617	58.544	63.686
234	141	V	120.635	7.989	175.852	62.494	32.44
235	142	W	124.227	8.064	175.744	57.324	29.597
236	143	K	123.776	7.885	175.47	55.852	33.426
237	144	R	122.722	8.12	176.145	56.135	30.633

238	145	E	123.009	8.483	175.845	56.697	30.44
239	146	T	120.154	7.813		63.196	70.675

The results from NMR titrations of ^{15}N-enriched XPA DBD with different DNA substrates were used to map the interaction surface of XPA DBD (Figure 3.5). Overall, the trends in chemical shift perturbations (CSPs) were the same for the different substrates. Major perturbations were found in the globular core between residues 130 – 210, and in residues extending C-terminally from the core between residues 215 – 230 (Figure 3.5). The N-terminal region of the globular core containing a zinc finger was unaffected except for a single residue D152, whose side chain is presumably engaged in a salt bridge that spans to the DNA binding site.

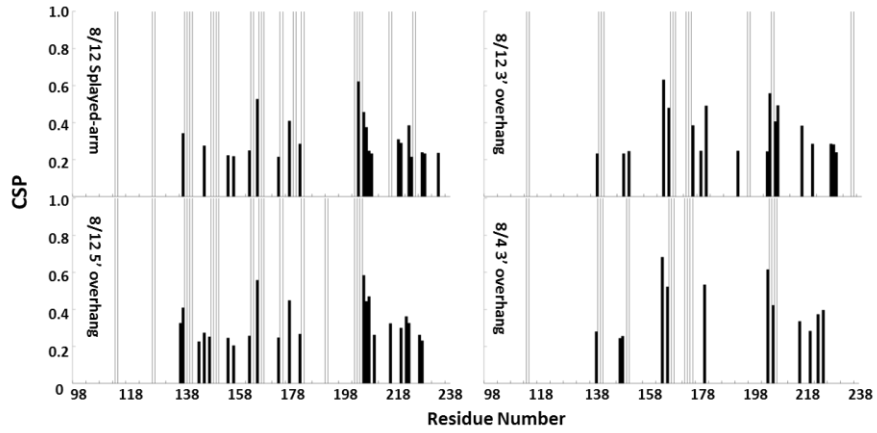


Figure 3.5: Chemical Shift Perturbations (CSPs) Induced by DNA Substrates. CSPs for each residues observed from the NMR titration of XPA DBD with 8/12 splayed-arm, 8/12 5′ overhang, 8/12 3′ overhang, and 8/4 3′ overhang DNA (substrates 1 without fluorescein tag, 10, 12, and 13 in Figure 3.1, respectively).

Peaks exchange broadened upon DNA binding are shown as open bars. Only the chemical shifts that were perturbed above the threshold are shown.

While the overall patterns are similar, each substrate had some unique features. For example, only the junctions with 5′ overhangs induced CSPs of the resonances of L138, E156, and V223. These observations reflect the XPA DBD adapting to the substrate. Notably, XPA has appreciably lower affinity for blunt end duplex substrates[37] and the CSPs were small relative to junction substrates. However, no direct correlation was observed between the magnitude of the CSPs and the binding affinities among the junction substrates.

An in-depth analysis was performed for the titration of ^{15}N labeled XPA DBD with the 8/4 5′ overhang substrate (Figure 3.1). Significant CSPs were observed primarily in the C-terminal portion of XPA DBD including β3, α1, the hairpin between β4 and β5, the C-terminal end of α3, and a number of residues in the C-terminal extension (Figure 3.6). Of the 33 residues perturbed by addition of the substrate, 17 exhibited fast exchange on the NMR time scale, whereas the 16 others exhibited intermediate exchange and were broadened beyond detection. Making the logical assumption that all CSP arise from the same DNA binding phenomenon, the residues with broadened signals have the largest chemical shift differences between the free and bound state and are presumed to be centered in

the binding site. This group of residues is highlighted in Figure 3.6-B, which shows the CSPs mapped on the NMR structure of the XPA globular core.

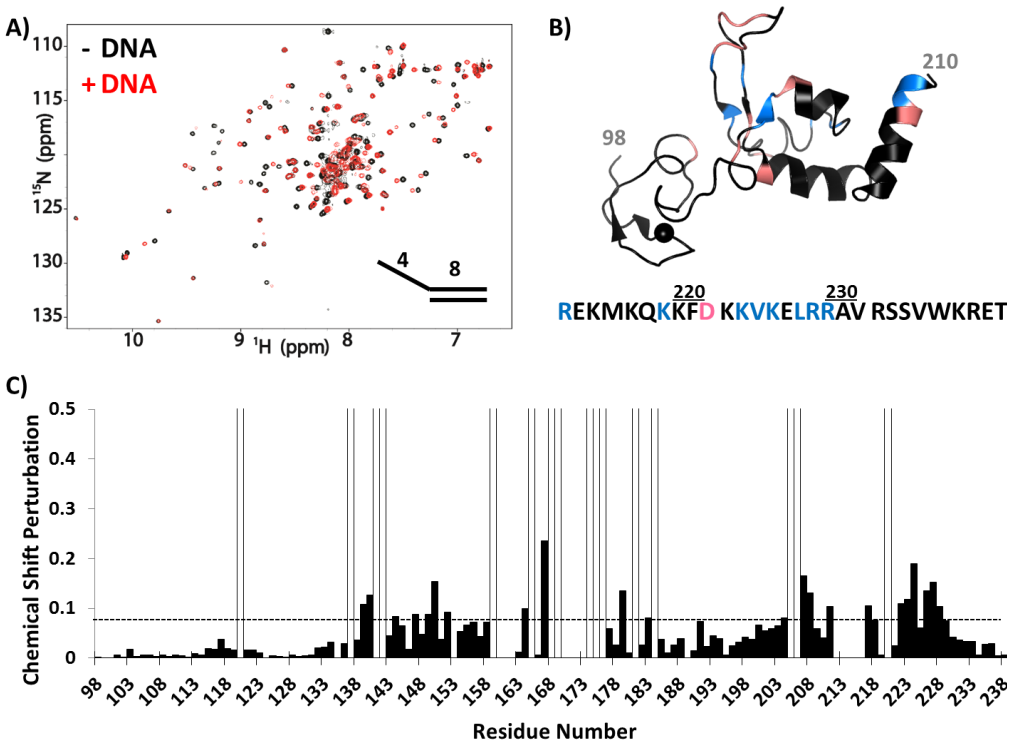
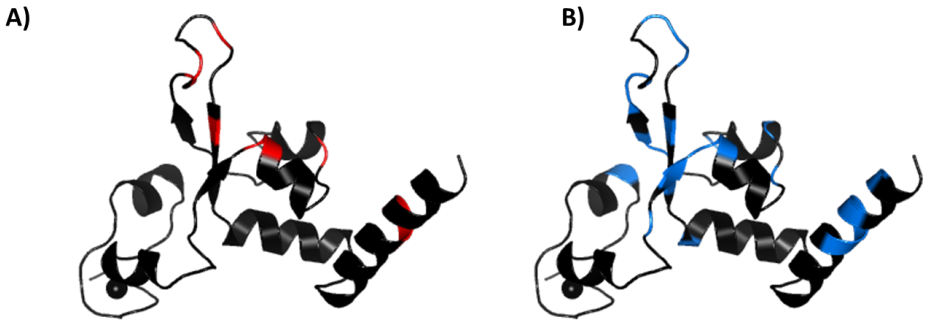


Figure 3.6: NMR Titration of XPA DBD with 8/4 5′ Overhang DNA. A) Overlay of the 900 MHz ^{15}N-^{1}H HSQC spectrum of XPA DBD in the presence (red) and absence (black) of 8/4 5′ overhang DNA (Figure 3.1, **11**). Spectra were acquired in a buffer containing 20 mM Tris, pH 7.0, 150 mM KCl, 1 mM Tris(2-carboxymethil) phosphine (TCEP), and 5 % $^{2}H_2O$. B) CSPs from spectra shown in A mapped on the 1XPA structure. Significant CSPs of C-terminal residues are mapped on the amino acid sequence below the structure. C) Plot of CSPs versus residue number from the spectra shown in A. Peaks exhibiting exchange broadening are shown as open bars. The threshold for significant CSP is indicated by the dashed line.

Comparison of DNA binding by human XPA DBD versus the yeast Rad14 globular core

The CSPs were also used to map the residues involved in binding DNA by XPA onto the structures of the Rad14 X-ray crystal structures and onto a homology model of the XPA DBD. Figure 3.7 shows ribbon diagrams of the yeast crystal structures and the XPA homology model based on one of the Rad14 crystal structures (PDB ID: 5A3D), with residues highlighted that are involved in DNA binding identified by the NMR.

Most of the residues involved in contacts with DNA in the Rad14 crystal structures exhibit CSPs in the NMR titrations of the human XPA DBD. However, the NMR study revealed many additional residues with significant CSPs induced by the binding of DNA. The most important difference were the CSPs observed in the extra 20 C-terminal residues extending beyond the globular core, which are not present in the Rad14 construct (Figure 3.7). These CSPs are fully consistent with our previous analysis showing that the C-terminal extension beyond the globular core is essential to recapitulating the DNA binding affinity of full-length XPA[53]. Two residues, T239 and H258, contact the DNA in the structure of Rad14, but the corresponding residues in XPA DBD (K151 and H171) did not exhibit a CSP in the DNA titrations. The absence of an effect on K151 is likely due to a structural adjustment in the flexible loop in which it is found; both




```
HUMAN         MEFDYVICEECGKE-FMDSYLMNHFDLPTCDNCRDA-DDKHKLITKTEAKQEY 14
YEAST         ----APKCIECHINIEMDPVLHDVFKLQVCKQCSKEHPEKYALLTKTECKEDY 23

HUMAN LLKDCDLEKREPPLKFIVKKNPHHSQWGDMKLYLKLQIVKRSLEVWGSQEALEEAKEVRQ 20
YEAST FLTDPELNDED-LFHRLEKPNPHSGTFARMQLFVRCEVEAFAFKKWGGEEGLDEEWQRRE 29

HUMAN ENREKMKQKKFDKKVKELRRAVRSSVWKRET 239
YEAST EGKAHRR 302
```

Figure 3.7: Comparing DNA-binding Residues Identified in Rad14 Crystal Structures and NMR Analyses of Human XPA. Mapping of DNA-binding residues identified in the Rad14 crystal structures (A) and NMR titrations (B) on the homology model of human $XPA_{102-214}$. (C) Sequence alignment of human XPA (top) and *S. cerevisiae* Rad14 (bottom). DNA binding residues are colored red. The residues reported to be mutated in cancer patients are indicated by underlines (missense mutations only).

neighboring residues around K151, L150 and D152, showed significant CSPs (Figure 3.7) suggesting one of these residues replaces the T238 contact. As for H171, the NMR analysis indicates K168, K179, K221, K222, and K224 are involved in DNA binding, but the Rad14 crystal structure suggests that among

these only K179 contacts the DNA. We note that our findings are consistent with the results of a previous mass spectrometry foot printing study that reported 6 XPA lysine residues (K168, K179, K221, K222, K224, K236) are involved in the binding of junction DNA[54].

The F262A mutation in Rad14 was reported to cause complete loss of binding to damage-containing duplexes, suggesting the critical importance of this residue in DNA binding by yeast Rad14[55]. The equivalent residue in XPA, W175, showed a significant CSP (Figure 3.6), consistent with its involvement in binding DNA and the positioning of the side chain in the homology model. However, when the DNA binding of a W175A mutant was tested, only a slight reduction in DNA binding affinity to the NER model substrates was observed (Figure 3.8-A, Table 3.3).

Mutation of residues in the C-terminal extension of XPA DBD inhibit binding of DNA

In order to investigate the contribution of residues 220–239 to DNA binding more closely, we prepared a series of mutations of basic residues in this region. Three charge reversal, single-site mutations (K221E, K222E, R228E) resulted in mild reduction in DNA binding affinity, whereas the double mutation

K221E/R228E had a much more dramatic effect (Figure 3.8-B, Table 3.3).

Truncations of the DBD also caused significant reductions in DNA binding

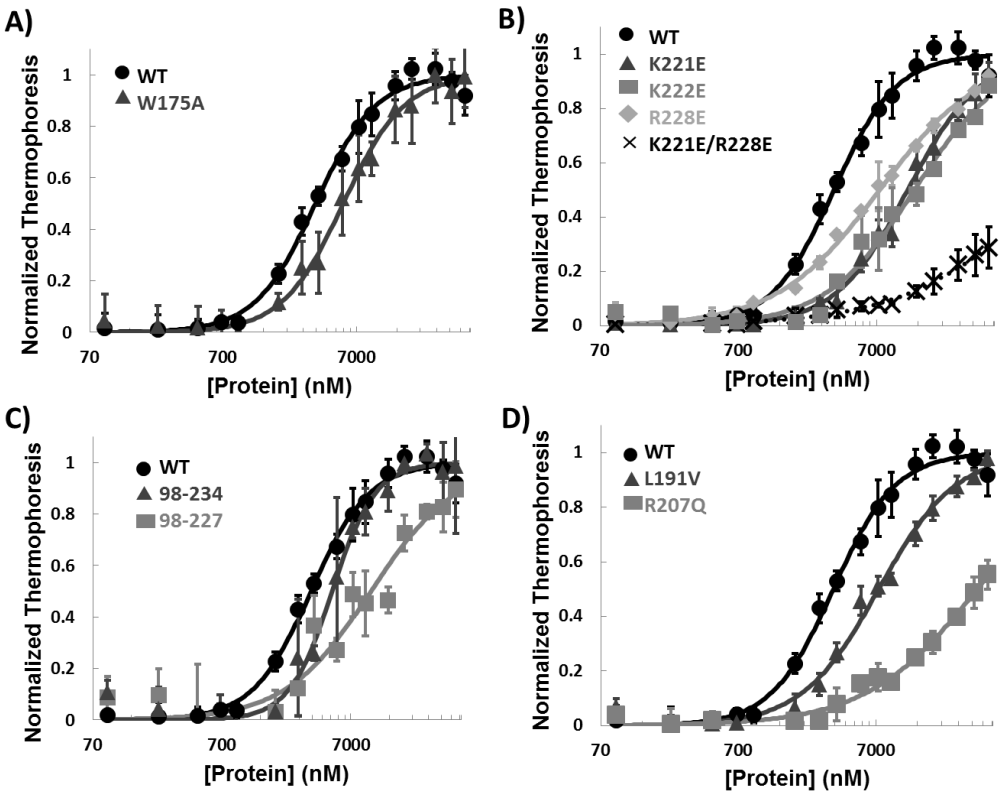


Figure 3.8: DNA Binding of Mutant XPA. MST analyses of DNA binding constants of WT XPA DBD and (A) W175A mutant, (B) mutations of residues in the C-terminal extension from the globular core, (C) truncation mutants, and (D) disease-associated missense mutants. All experiments used the 8/4 HP 5′ overhang DNA (Figure 3.1, **9**).

Table 3.3: DNA Binding of Mutant XPA. Table of dissociation constants (Kd) extracted from the MST data plotted in panels A–D in Figure 3.8. All measurements were made at room temperature in a buffer containing 50 mM Tris-HCl at pH 7.8, 150 mM NaCl, 10 mM $MgCl_2$, 0.05 % Tween-20 and 1 mM DTT. Asterisk shows that the binding was too weak to accurately determine Kd values.

XPA Construct	Kd (μM)
XPA98-239	3.4 ± 0.2
XPA98-239 W175A	5.5 ± 0.4
XPA98-239 K221E	12.0 ± 1.0
XPA98-239 K222E	13.0 ± 3.1
XPA98-239 R228E	7.4 ± 0.6
XPA98-239 K221E/R228E	*
XPA98-234	5.0 ± 0.3
XPA98-227	10.0 ± 3.6
XPA98-239 L191V	7.4 ± 0.7
XPA98-239 R207Q	*

affinity (Figure 3.8-C, Table 3.3). Although both truncation mutants retained some DNA binding activity, the shorter construct XPA_{98-227} had weaker DNA binding activity than the longer XPA_{98-234}. These results confirm that the cluster of basic residues in the C-terminal of XPA DBD contribute significantly to DNA binding.

Insights into the molecular basis of XPA disease-associated mutations

Our biophysical and structural studies of DNA binding by the XPA DBD provide information to enhance the general understanding of genotype-phenotype correlations for XPA mutations. The available data is limited, especially for many missense mutations discovered in cancer patients[162], simply because there have been few investigations of phenotypes or biochemical malfunctions of

specific XPA mutants. The results reported here enable some genotype-phenotype correlations to be made. For example, *XP-A* patients expressing truncation mutant XPA_{1-227} are known to exhibit mild neurological disorders while those with XPA_{1-220} have severe neurological symptoms[13]. Our current and previous data show that XPA DBD truncated at R227 retains weak DNA binding (Figure 3.8-C, Table 3.3), whereas truncation at F219 barely retains any DNA binding activity[53]. These observations suggest a correlation exists between the degree of inhibition of XPA DNA binding activity and *XP* disease phenotype. The observation of charge reversal mutations of K221 and R228 causing only very small reductions in DNA binding (Figure 3.8-B, Table 3.3) therefore suggests that the K221Q and R228Q mutations will result in mild disease symptoms.

Insights were obtained for three additional disease-associated mutants, V166A, L191V and R207Q. These mutations were prepared in the XPA DBD construct, expressed and purified, then their structural integrity and DNA binding affinity were characterized. In WT XPA DBD, V166 did not have a significant CSP upon binding of DNA but is between residues that did (I165 and K167), L191 is distant from residues affected by DNA binding, and R207 has a significant CSP (Figure 3.7). As anticipated, L191V did not cause any significant effect on DNA binding, whereas R207Q caused a dramatic decrease in DNA binding affinity. The data for V166A were more surprising.

Interestingly, we were unable to concentrate V166A sufficiently to conduct the DNA binding assay, suggesting reduced stability of this mutant. We were however able to establish that the structural integrity of this mutant was maintained, as both the NMR and CD spectra were similar to the WT protein (Figure 3.9). To determine if the mutation altered the stability of the domain, thermal denaturation experiments were performed by CD for the mutant and the WT protein (Figure 3.9-B), revealing that the V166A mutation has a 4 °C lower apparent thermal denaturation mid-point (Tm) than the WT. Hence, although the V166A mutation does not grossly alter folding, it does significantly reduce the stability of the globular core. Interestingly, in the Rad14 crystal structures and homology models of XPA, the V166 residue is a part of the β-hairpin that intercalates into the ss-dsDNA junction. A likely explanation for the decreased stability could be the loss of hydrophobic interactions that support the cross-strand interaction in the β-hairpin. Given its important role in XPA binding of DNA, destabilization of the β-hairpin may well prove to be the biochemical malfunction at the origin of disease-association.

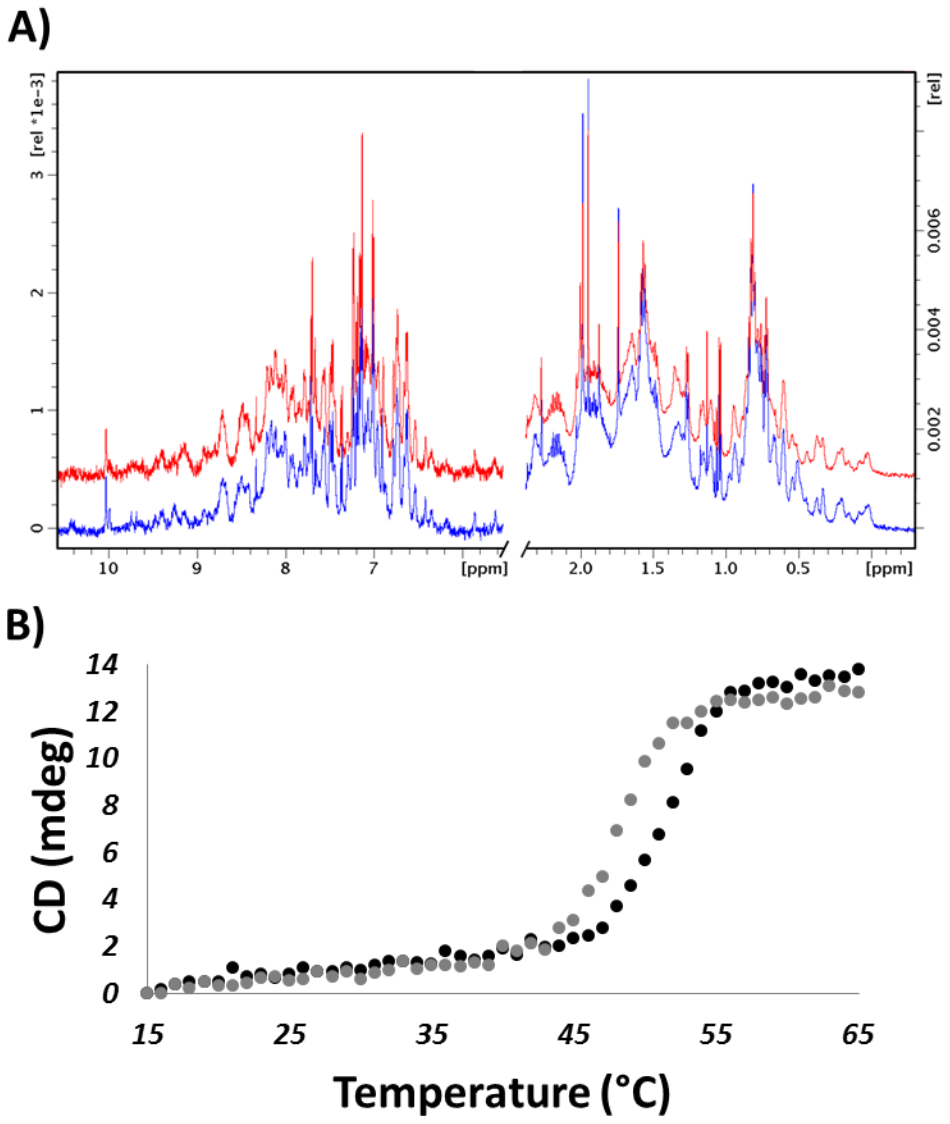


Figure 3.9: V166A Mutant is Less Stable than WT XPA ((A) Overlay of ^{1}H 1D NMR spectra of XPA DBD WT (blue) and V166A (red). (B) CD thermal denaturation curves of WT (black) and V166A (gray) XPA DBD. Melting temperatures (Tm) were measured from the maxima of the derivatives of each curve, resulting in values of 45.5 °C for WT and 41.5 °C for V166A.

Discussion

XPA is central scaffold for human NER machinery, and although loss of DNA binding activity of XPA is anticipated to lead to severe *XP* symptoms[8], XPA binding to the NER bubble and its effect on NER efficiency or correlation to *XP* symptoms have not been investigated. One contributing factor was that the DNA binding domain had been incorrectly assigned to the XPA globular core, which has only weak affinity for DNA[41,53]. Important new insights were obtained recently when the first high-resolution structures in the presence of DNA were reported for the yeast homolog of XPA, Rad14[55]. However, whether or not these structures serve as accurate models for human XPA in the context of NER remains unclear because the DNA substrates in the crystal structures were 15 or 16-bp DNA duplexes modified at the center with specific lesions. These substrates are not representative of canonical NER as XPA is not recruited to the site of damage until TFIIH unwinds the damaged DNA[3,38]. The specific lesion-containing duplexes appear to be very unique substrates because in general, XPA binds ss-ds junctions much more tightly than duplexes. Indeed, Rad14 binds unmodified duplexes and other lesion-containing duplexes with substantially lower affinity than duplexes modified with cisplatin and AAF[55]. XPA has been shown to interact with proteins not involved in NER (e.g. XPA binding proteins[103–106] and PCNA[92,162]) and it is conceivable that the Rad14 structures reveal the molecular basis of DNA interactions in the other pathways for cellular

processing or bypass of damaged DNA. Regardless, structural information on human XPA in complex with NER substrates is required to better understand the molecular basis of XPA function in NER.

We have investigated the interaction of human XPA with DNA using a DBD construct that exhibits the same DNA binding affinity as the full-length protein and ss-dsDNA junction substrates that model the NER bubble created after the damaged duplex is unwound by TFIIH. NMR backbone resonance assignments allowed identification of the residues affected by binding to DNA substrates. Analysis of different DNA substrates resulted in similar CSPs, suggesting that the same set of residues is involved in DNA binding for variety DNA structures with different binding affinities. For the most part, there was a good correlation between NMR CSPs of residues in the globular core and the corresponding DNA binding site in the yeast Rad14 crystal structures. The NMR analysis revealed significant DNA contacts with residues in the C-terminal residues that extend beyond the globular core and are missing from the yeast construct. As a result, uncertainties remain as to whether the Rad14 structures provide a sufficient model for DNA binding by XPA.

The relevance of the Rad14 model is further limited by the fact that two Rad14 molecules are engaged, one on either end of the lesion-containing duplex. Interaction in this manner is inconsistent with current NER models, which include

only a single XPA molecule in the pre-incision and incision complexes. That said, it is conceivable that the end-on mode of interaction of Rad14 is in some way mimicked when XPA is engaged at the ss-dsDNA junction within the NER bubble. However, the yeast and human proteins do not completely correspond; for example, the W175A mutant of XPA DBD binds DNA with similar affinity as WT, whereas the corresponding F262A mutant in the Rad14 globular core completely abolished binding to DNA. These results provide further evidence of the need for high-resolution structures of complexes of XPA with NER substrates.

Residues C-terminal to the globular core contribute significantly to the binding of NER model substrates. Our mutational analysis of basic residues at the C-terminus of XPA DBD shows that multiple residues collectively contribute to the high affinity to DNA. These studies also shed light on genotype-phenotype relationships for certain XPA mutants; in particular, we believe a correlation exists between the effect of mutation on DNA binding affinity and severity of symptoms in *XP* patients. However, DNA binding affinity is but one of several factors that can contribute to biochemical malfunction of disease-associated mutations. For example, we found that the V166A mutation is destabilized compared to the WT protein, which could affect not only DNA binding, but also protein interactions and cellular turnover. R207 presents another interesting example. We found that the disease-associated R207Q mutation has significantly weakened DNA binding affinity, whereas a host reactivation assay showed that

the R207E mutant is capable of the repair of UV lesions[65]. The assay is not sensitive to effects on the rate of repair, which may be compromised when DNA binding affinity is diminished. Moreover, this residue is also important for interactions with the NER factor XPE[48]. The remarkably rapid progress in recent years in structural analysis of large multi-protein complexes, such as those assembled along the trajectory of NER, holds promise for dissecting such complex relationships. Systematic functional analyses, for example of UV lesion repair efficiency, will also be required to attain the ultimate objectives of a more complete understanding of the relationship of XPA DNA binding affinity to NER activity and the genotype-phenotype correlation of disease-associated mutations.

Materials and Methods

XPA DBD Mutant Construction

We previously reported the construction and purification of XPA DBD (XPA_{98-239})[53]. Using this pBG100 XPA DBD plasmid as the template, site single mutations were introduced using the Q5 site-directed mutagenesis kit (New England BioLabs, Inc.) following the manufacturer's protocols. The K221E/R228E double mutant was created using K221E primers and pBG100 XPA DBD R228E plasmid as the template. Primers for mutagenesis are listed in

Table 3.4. For the truncation mutants R228X (XPA_{98-227}) and W235X (XPA_{98-234}), XPA constructs were amplified by PCR using the oligonucleotides indicated in Table 3.4 to introduce 5′ *BamHI* and 3′ *NotI* cleavage sites. All mutants were cloned into the pBG100 in-house expression vector (L. S. Mizoue, Center for Structural Biology, Vanderbilt University), which incorporates an N-terminal human rhinovirus 3C (HRV3C) protease cleavable 6xHis tag.

Table 3.4: Primer Sequences for XPA DBD Mutant Construction

Mutation	Primer	Sequence (5' → 3')
V166A	Fwd	AAATTTATTGcGAAGAAGAATCCACATC
	Rev	AAGAGGTGGCTCTCTTTTTTC
W175A	Fwd	TCATTCACAAgcGGGTGATATGAAACTC
	Rev	TGTGGATTCTTCTTCACAATAAATTTAAG
L191V	Fwd	GAAGAGGTCTgTTGAAGTTTGG
	Rev	ACAATCTGTAACTTTAAGTAGAG
R207Q	Fwd	AAGGAAGTCCaACAGGAAAACC
	Rev	TGCTTCTTCTAATGCTTCTTG
K221E	Fwd	GAAATTTGATgAAAAAGTAAAAGAATTGC
	Rev	TTCTGTTTCATTTTTTCTCGG
K222E	Fwd	ATTTGATAAAgAAGTAAAAGAATTGCGG
	Rev	TTCTTCTGTTTCATTTTTTCTCG
R228E	Fwd	AGAATTGCGGgaAGCAGTAAGAAG
	Rev	TTTACTTTTTTATCAAATTTCTTCTGTTTC
R228X	Fwd	GCGGGATCCATGGAATTTGATTATGTAATATGCG
	Rev	TAATTGCGGCCGCTCACCGCAATTCTTTTACTTTTTTATC
W235X	Fwd	GCGGGATCCATGGAATTTGATTATGTAATATGCG
	Rev	TAATTGCGGCCGCTCACACGCTGCTTCTTACTGC

DNA Substrate Preparation

Figure 3.1 shows the structures of DNA substrates used in this study. The position of the FITC tag is indicated if applicable. Desalted oligodeoxynucleotides were purchased from Sigma-Aldrich Co. (St. Louis, MO). Y-shaped ssDNA-dsDNA junctions, and duplexes were prepared by mixing an equimolar amount at 0.5 – 1 mM of each strand in the buffers listed below for NMR or DNA binding experiments. Then the mixture was heated in a boiling water bath and allowed to cool to room temperature for annealing. Hairpin-containing DNA substrates were dissolved in TNE buffer (10 mM Tris, 50 mM NaCl, 1 mM EDTA, pH 7.0) at 2 μM concentration and annealed by heating in a boiling water bath followed by immediate cooling on ice.

XPA DBD Production

WT and mutant XPA DBD proteins were expressed and purified as described previously[53]. The XPA DBD used for NMR titrations was expressed in minimal media containing 0.5 g/ L of $^{15}NH_4Cl$ (CIL, Inc.). The preparations of samples for backbone resonance assignments also contained 2 g/ L $^{13}C_6$-glucose (CIL, Inc.).

Analysis of DNA Binding by Fluorescence Anisotropy

The basic protocol for the fluorescence anisotropy DNA binding assay was described previously[53].

Analysis of DNA Binding by Microscale Thermophoresis[128,129,165,166]

Proteins were dialyzed into the MST buffer (50 mM Tris-HCl at pH 7.8, 150 mM NaCl, 10 mM $MgCl_2$, 0.05 % Tween-20 and 1 mM DTT). Fluorescein-labeled DNA stocks were also diluted in MST buffer. All samples and buffer were passed through a 0.2 μm filter. For each experiment, sixteen dilutions of the protein were prepared to varying concentrations. DNA was added to each of the sample to the final concentration of 40 nM in the tubes provided in the Monolith NT.115 Series Standard Treated Capillaries kit (NanoTemper, Inc.). All experiments were carried out at room temperature using the capillaries in a Monolith NT.115 Blue/Red instrument (NanoTemper, Inc.) at 20 % LED power and 40 % MST power. Data were analyzed using MO.Affinity software (NanoTemper, Inc.).

Circular Dichroism

Samples of XPA DBD WT and V166A mutant were dialyzed into a buffer containing 20 mM Tris at pH 7.0, 150 mM KCl, and 1 mM DTT. All samples and buffer were passed through a 0.2 µm filter before data collection. The concentration of protein was adjusted to 11 µM prior and far-UV CD data were collected at 220 nm over the range 15–65 °C using a Jasco J-810 CD spectropolarimeter (Easton, MD).

Generation of the XPA DBD Homology Model

A homology model of human $XPA_{102-214}$ was generated using the X-ray crystal structure of *S. cerevisiae* Rad14 (PDB ID: 5A3D) as template for calculations using Modeller 9.14[167].

^{1}H 1D NMR Analysis of XPA DBD V166A

XPA DBD WT and V166A mutant were concentrated to 30 µM in a buffer containing 20 mM Tris, pH 7.0, 75 mM KCl, 1 mM TCEP to which 5% $^{2}H_2O$ was added prior to performing the experiment. ^{1}H 1D spectrum of each

sample was acquired in a 3 mm tube at 25 °C using a Bruker *AVANCE* 600 MHz spectrometer equipped with a cryoprobe.

NMR Backbone Resonance Assignments

^{13}C,^{15}N-enriched XPA DBD was concentrated to 460 μM in a buffer containing 20 mM Tris, pH 7.0, 500 mM KCl, 1 mM TCEP and 5 % $^{2}H_2O$. ^{15}N-^{1}H HSQC and a series of heteronuclear triple resonance 3D experiments (HNCO, HNCA, HNCACB, and CBCACONH)[120] were acquired in a shaped tube at 25 °C using a Bruker *AVANCE* 600 MHz spectrometer equipped with a cryoprobe. All 3D data were collected using non-uniform sampling (NUS)[168–173] and were reconstructed using Topspin (Bruker). Further data processing and figure preparations were carried out using SPARKY (Goddard, T.D. & Kneller, D. G. SPARKY 3, University of California, San Francisco). Resonance assignments were obtained using NMRview software (One Moon Scientific).

NMR Titration of XPA DBD with DNA

All samples for DNA titrations were concentrated to 50 μM in the NMR buffer containing 20 mM Tris, pH 7.0, 150 mM KCl, 1 mM Tris(2-carboxymethil) phosphine (TCEP), and 5 % $^{2}H_2O$. Titrations were performed at

25 °C using a Bruker *AVANCE* 800 or 900 MHz spectrometer equipped with a cryoprobe following the procedure described previously[53]. A second titration with 8/4 5′ overhang substrate was performed in the NMR buffer with 250 mM instead of 150 mM KCl to enable transfer of the assignments from the conditions used for obtaining resonance assignments.

Acknowledgements

This work would not have been possible without the contributions of Dr. Markus Voehler in backbone resonance assignment as well as giving advice for 3D NMR data collections. We appreciate Michelle Roh for her hard work and patience in producing and performing preliminary DNA binding analysis of disease associated XPA DBD mutants. We thank Dr. Agnieszka Topolska-Woś for producing ^{15}N XPA DBD and critical proofreading of this chapter in preparation for publication. We also thank Dr. Nicole Eckart from NanoTemper Technologies for the valuable input to optimize and interpret MST data and Dr. Sarah Soss for her advice with backbone resonance assignment. This work was supported by funding from P01 CA092584, R35 GM 118089, P30 CA068485, P30 ES000267, grants for NMR instrumentation from the NSF (0922862), NIH (S10 RR025677), Vanderbilt University matching funds, as well as MST

instrumentation grant from NIH (S10 OD021483). AMTW is supported in part by a Kosciuszko Foundation scholarship.

CHAPTER IV

DISCUSSION AND FUTURE DIRECTIONS[3]

Summary of This Work

Among numerous factors involved in human NER, XPA seems to play a critical role as indicated by the fact that XPA mutations are associated with severe *XP* disorder symptoms. This is especially true for those mutations found within XPA DBD, containing interaction motifs for DNA and RPA70AB. This dissertation research set out to elucidate molecular basis of XPA-DNA and XPA-RPA interactions at XPA DBD. Key findings from this project are summarized below.

Revised DNA binding domain of XPA

For the past ~20 years, XPA_{98-219} had been termed the XPA DNA binding domain[41]. Numerous biochemical and structural studies have been performed based on this study. However, my DNA binding assays showed that the DNA binding of XPA_{98-219} is significantly weaker than that of FL-XPA[53]. Based on sequence analysis of XPA suggesting that the C-terminus of XPA_{98-219} is in the

[3] Part of this chapter was published in Sugitani, N., Sivley, R. M., Perry, K. P., Capra, J. A., Chazin, W. J., XPA: A Key Scaffold for Human Nucleotide Excision Repair, *DNA Repair*, 2016, 44, 123-135. Contents reused under permission by Elsevier (license #: 4063671121717).

middle of a long C-terminal helix, new XPA constructs extending towards the C-terminus were constructed. Subsequent DNA binding assays and NMR analysis confirmed that XPA_{98-239} construct is a well folded construct that contains full DNA binding activity comparable to FL-XPA[53]. Consistent with this discovery, another group also reported that there are critical lysine residues beyond residue F219 that are involved in DNA binding[54].

Comparison of DNA binding by human XPA and S. cerevisiae Rad14

Although NMR structures of the core globular domain of human XPA[39,174] and complementary data on DNA binding[26,37] have been available for many years, the molecular details of how human XPA binds DNA remain unclear. Insights have been obtained from X-ray co-crystal structures of the central globular domain of the yeast XPA (Rad14) in the presence of DNA[55], but the construct lacked the corresponding C-terminal residues that were shown to be critical to the DNA binding activity of XPA (Figures 1.6, 3.7). Moreover, the duplex DNA substrates do not correspond to DNA structures encountered by XPA as it functions in NER machinery[3,37,38]. NMR studies of human XPA DBD using ss-dsDNA junction substrates revealed which residues are affected upon DNA binding, including those from C-terminal extension. Mutations of these C-terminal residues confirmed that the cluster of basic residues in this region is important for DNA binding (Figure 3.8-B and C). Moreover, a key site in Rad14,

F262, was previously reported as critical to DNA binding[55], but mutation of the corresponding W175 in human XPA had only a moderate effect on DNA binding (Figure 3.8-A).

DNA binding activity of disease-associated XPA mutants

Effects of select disease-associated XPA mutants on DNA binding were investigated. Consistent with the NMR titration of XPA DBD with ss-dsDNA junction substrates (Figure 3.7-C), missense mutant R207Q showed significant reduction in DNA binding while L191V did not (Figure 3.8-D). Comparing the DNA binding of truncation mutants XPA_{98-219}, XPA_{98-234}, and XPA_{98-227} to XPA DBD confirmed that the shorter the C-terminal extension is, the weaker is the DNA binding activity of XPA (Figures 2.2, 2.6-A, and 3.8-C). Patients expressing truncation mutant XPA_{1-227} are known to exhibit mild neurodegeneration[13]. Interestingly, it was also observed that while it doesn't result in complete unfolding of the protein, the V166A mutant is less stable compared to WT XPA (Figure 3.9). This reduction in stability could influence XPA activity within NER pathway.

Defining the site of interaction of the XPA globular core with RPA70AB

There are two interaction sites between XPA and RPA. The major interaction occurs between N-terminus of XPA and RPA32C domain[46] while a

weaker, secondary interaction occurs between the globular core of XPA and the RPA70AB domain (Figure 1.3) [45,87,88]. The molecular basis of the interaction between the XPA globular core and RPA70AB was unclear because there were two conflicting proposals in the field. Whereas one model suggested that the N-terminus of the XPA globular core containing the zinc finger is responsible for RPA70AB interaction[39,45], a second suggested that the C-terminal region of the globular core is involved and that RPA70AB binding site overlaps with DNA binding residues[89,90]. My NMR titration of XPA globular core with RPA70AB showed that the N-terminal region of XPA DBD is responsible for the interaction with RPA70AB and this binding site is independent of the DNA binding residues within XPA DBD. A reverse NMR titration indicated that the binding sites for ssDNA and XPA globular core are spatially proximate on RPA70AB. When the XPA globular core was added to RPA70AB-ssDNA complex, it did not displace the DNA but induced CSPs to the residues near the ssDNA binding site of RPA70AB. These results indicate that the XPA globular core, ssDNA and RPA70AB can form a ternary complex.

Implication of the Results

As summarized above, this study has expanded understanding of XPA interactions with DNA and RPA as well as mutations at the XPA DBD. This

section discusses implication of these results in understanding human NER and *XP* disorders.

Differences in DNA binding of human XPA versus the structural model of S. cerevisiae Rad14-DNA complex

Recently determined crystal structures of yeast homolog of XPA, Rad14, in the presence of DNA provided important new insights in XPA-DNA interaction[55]. However, whether or not these structures serve as accurate models for human XPA in the context of NER remains unclear. The DNA substrates used for the study were duplex DNA with lesions at the center. Moreover, high affinity binding of the Rad14 DNA binding construct was observed only for duplex DNA containing either the cisplatin or AAF adduct lesions; much lower affinity was observed for unmodified duplexes or those containing different lesions[55]. Importantly, these substrates are not representative of canonical NER as XPA is not recruited to the site of damage until TFIIH unwinds the damaged DNA[3,38].

XPA binds ss-ds junctions much more tightly than duplexes[37]. XPA has been shown to interact with proteins not involved in NER (e.g. XPA binding proteins[103–106] and PCNA[92,162]) and it is conceivable that the Rad14 structures reveal the molecular basis of DNA interactions in unforeseen ways or in other pathways for cellular processing or bypass of damaged DNA. It is also possible

that these crystal structures suggest specialized sub-pathways of NER that are not yet discovered. Regardless, structural information on human XPA in complex with proper NER substrates is required to better understand the molecular basis of XPA function in canonical NER.

In this dissertation work, we employed the re-defined XPA DBD construct that exhibits the same DNA binding affinity as the full-length protein and ss-dsDNA junction substrates that model the NER bubble created after the damaged duplex is unwound by TFIIH. NMR backbone resonance assignments as well as NMR titrations revealed that the same set of residues is involved in DNA binding for variety DNA structures with different binding affinities. There was some agreement between the DNA binding residues determined from our NMR studies and from the Rad14 crystal structures in the globular core. However, NMR analysis showed significant DNA contacts with residues in the C-terminal residues that extend beyond the globular core and are absent from the yeast construct. Our mutational analysis supports that the proposal that multiple basic residues in the C-terminal extension collectively contribute to the high affinity to DNA. Therefore, uncertainties remain as to whether the Rad14 structures provide a sufficient model for DNA binding by XPA. It is conceivable that C-terminal extension of Rad14 to better correspond to human XPA DBD would result in DNA binding properties similar to human XPA DBD.

It is possible that the end-on mode of interaction of Rad14 may in some way be mimicked when XPA is engaged at the ss-dsDNA junction within the NER bubble. However, it is difficult to imagine that two molecules are engaged in the manner of Rad14, with one at each end of the lesion containing duplex as shown in the crystal structures. This type of intermediate DNA structure with XPA would not be present at any point of NER. It was also striking to find that W175A mutant of XPA DBD binds DNA with similar affinity as WT, whereas the corresponding F262A mutant in the Rad14 globular core completely abolished binding to DNA. These results further support the importance of high-resolution structures of complexes of XPA with appropriate NER substrates to fully characterize XPA-DNA interactions in terms of human NER.

XPA is not likely dimerized at the NER bubble

Binding of XPA to a ss-dsDNA junction of the NER bubble is well established[37,55]. As discussed in Chapter I, whether XPA binds to the junction 5′ or 3′ to the lesion remains unclear. Interestingly, characterization of XPA *in vitro* has revealed that XPA can form a dimer[92]. It may be possible for two molecules of XPA to interact with each ss-dsDNA junction of the NER bubble, particularly *in vitro* with an isolated NER bubble and purified XPA. However, considering the extensive network of protein-protein interactions of XPA, it is difficult to imagine two molecules of XPA could interact simultaneously with the NER

bubble. For example, as described in Appendix A, the XPA globular core interacts with DNA and RPA70AB at one of the ss-dsDNA junction of the NER bubble. Moreover, the undamaged ssDNA strand of the NER bubble is just long enough to accommodate one molecule of RPA. So, if a second XPA molecule was bound to the bubble, it will have to compete with the first molecule for the binding site of RPA as well as other NER proteins.

Correlation between XPA mutations and XP disease symptoms

This work has provided insights into the correlation between XPA mutations and *XP* disease symptoms. Different XPA mutations lead to different disease phenotypes[4,12,13,107]; complete loss, substantial truncation, and unfolding of the DBD are known to lead to severe *XP* disease phenotypes. However, genotype-phenotype correlations and mechanisms behind missense mutations remain poorly understood[8,13,110,162,163]. *XP* patients with the gene truncated to XPA_{1-210} exhibit severe neurodegeneration, whereas those with the gene truncated to XPA_{1-227} show only mild neurodegeneration[13]. My study showed that XPA_{98-227} has mildly reduced and XPA_{98-219} has severely reduced DNA binding. Since XPA_{1-210} is even further truncated at the C-terminus than XPA_{98-219}, it is expected that this construct retains very little or no DNA binding activity. These observations suggest that the loss of DNA binding activity of XPA is related to severe *XP* symptoms.

It is worth noting that DNA binding affinity is one of several factors that can contribute to biochemical malfunction of disease-associated mutations. As in the case of the V166A mutation, which was found to be destabilized compared to the WT protein, an effect on function may arise not only DNA binding, but also protein interactions and cellular turnover. For example, although the R207Q mutation has significantly weakened DNA binding affinity, the corresponding R207E mutant is capable of the repair of UV lesions in a host reactivation assay[65]. The assay is not sensitive to effects on the rate of repair, which may be compromised when DNA binding affinity is diminished. Moreover, this residue is also important for interactions with the NER factor XPE[48]. Clearly, multiple factors have the potential to lead to *XP* disease in the case of R207Q.

A structural model for the NER complex

As the initial steps in characterizing how NER complex is organized, we began to investigate the overall architecture of the full length XPA-DNA-RPA complex by SAXS. For multi-domain proteins and protein complexes, interpretation of SAXS data is greatly facilitated by high resolution structures or models of the components. High resolution models and NMR titration data are available to model XPA-DNA interaction in atomic detail. While structures were available for individual proteins, information on the interaction between the XPA

globular core and RPA70AB was limited. In order to fill this gap in knowledge, this work has identified the residues critical in this interaction with and without

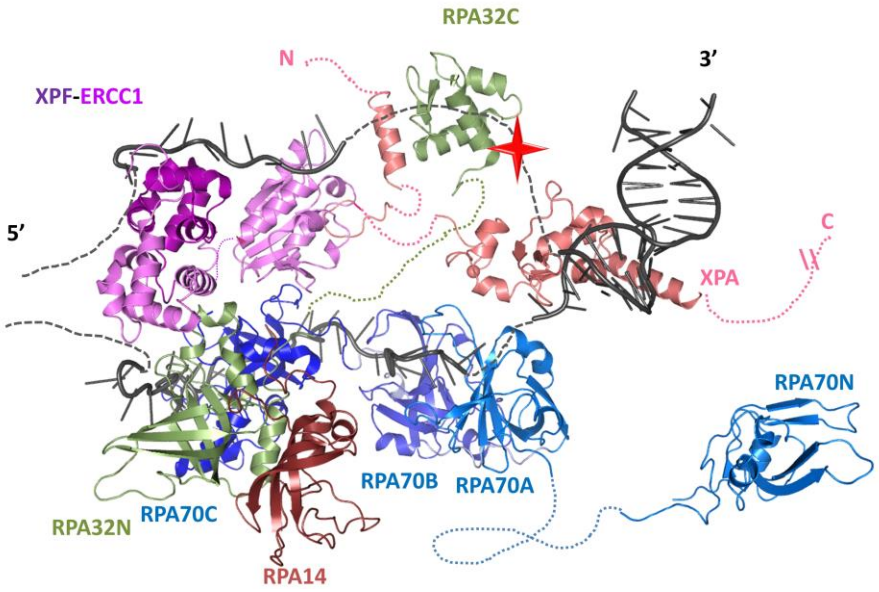


Figure 4.1: Model of some XPA interactions in the NER complexes. An homology model of $XPA_{102-214}$ in complex with an AAF-containing duplex was built based on the Rad14t structure (PDB: 5A3D). A SAXS model was used for the RPA DNA binding core in complex with ssDNA. The structure of RPA70N is taken from an X-ray crystal structure (PDB: 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.

the presence of ssDNA. Figure 4.1 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. A 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[175]. While not included in Figure 4.1 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[66–77,176]. 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 1.3), 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. Biophysical studies described in chapters 3 for characterizing the XPA-DNA interaction in human NER will further refine the model in combination with high resolution structures or computational models of XPA DBD in complex with appropriate DNA substrates. Maturation of the studies

described in Appendix A to elucidate interactions of the XPA globular core with RPA70AB, as well as the characterization of the full-length XPA-DNA-RPA complex, would also aid in the refinement of models of the NER complex. Since the NER incision complexes progressively incorporate key proteins, it is important to investigate the trajectory of structures 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.

Future Directions

XPA as a drug target

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 cisplatin drugs[154]. 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. Unlike other XP proteins, XPA is understood to be a factor specific to NER and mutations of XPA are detrimental to NER function. Therefore, targeting XPA could selectively inhibit NER function without affecting other pathways. However, as will be discussed below, there is increasing realization that XPA may be involved in non-NER pathways. Interestingly, circadian oscillation of XPA was observed in mice liver but not in testis[177]. Further investigation of tissue-specific distribution and behavior of XPA could contribute to potential interest in tissue-specific NER inhibitors.

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. It is important to note that increased efficiency of chemotherapy by combining with a NER inhibitor would result in increased toxicity. Therefore, knowledge of differences in XPA expression levels and function in cancer and normal cells may be useful for selectively targeting cancer cells.

Potential involvement of XPA in other pathways

Currently, there is no clear evidence that XPA could play a role in pathways other than NER. However, there are several reports showing the interaction of XPA with non-NER proteins[92,103,104,177]. As mentioned in Chapter 3, recently reported structures of yeast XPA in complex with DNA may represent a DNA binding mode of XPA that is operative outside of NER. While dimerization of XPA was observed *in vitro*[57], our data together with current structural results suggest XPA does not form a dimer within NER complexes. These observations suggest that XPA may be involved in DNA processing pathways other than NER. One approach to test this possibility is the flow-cytometric host cell reactivation assay (FM-HCR)[178]. FM-HCR allows for high-throughput investigation of the involvement of certain protein in different DNA repair pathways. To this end, our group is planning collaboration with Dr. Zachary Nagel at Harvard, who has developed the FM-HCR approach.

Organization and remodeling of the NER complex

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[3], 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[179,180].

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[181–183]. 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.

Concluding Remarks

The studies presented here offer insights into the molecular basis of XPA interactions in human NER. We updated the DNA binding construct of XPA, which was misidentified about 20 years ago. Key residues involved in DNA and RPA70 interactions, as well as initial insights into structural organization of XPA-RPA-DNA complex in NER were also demonstrated. We also cataloged disease-associated XPA mutations and showed initial steps in dissecting the molecular basis for *XP* symptoms correlated to each mutation. These findings serve as stepping stones in elucidating the molecular details of human NER pathway,

which could then contribute to development of treatments for *XP* disorders as well as cancer therapeutics.

APPENDIX

INTERACTIONS OF XPA AND RPA

A. INTERACTIONS OF THE GLOBULAR CORE OF XPA AND RPA70AB

Introduction

XPA and RPA work together as the central scaffold in human NER. While new data on XPA-DNA interaction have been reported, the current body of information and structures are not sufficient to understand how human XPA interacts with the NER bubble. Since XPA DBD contains DNA and RPA70 interaction residues (Figure 1.3), it is possible that these two interactions are coupled. However, the interaction of XPA with RPA70AB is poorly characterized, in part because there are conflicting reports for where RPA70AB binds within XPA from indirect methods. An early NMR titration of XPA_{98-219} with RPA70 truncation constructs suggested the zinc finger in the N-terminus of XPA_{98-219} is involved[39,45]. On the other hand, biochemical pull-down and cell-free NER assays with XPA mutants concluded that C-terminal residues of XPA_{98-219}, overlapping with the DNA binding residues described in chapters 2 and 3, are responsible for the interaction[89,90].

In my studies using NMR titrations, we confirmed that the N-terminal region of the XPA globular core is responsible for RPA70AB binding. We found that XPA and ssDNA elicit chemical shift perturbations of some common residues of RPA70AB, but they do not compete for the same binding site. In combination with available high resolution structures, these data will be used to build structural models for XPA globular core-DNA-RPA70AB complex.

Results and Discussion

A series of NMR titration experiments were performed in order to fill in the gap in knowledge regarding how XPA interacts with RPA70AB. An initial titration of ^{15}N-labeled XPA DBD with RPA70AB showed significant exchange broadening of resonances of residues in the N-terminal zinc finger region of XPA DBD (Figure A.1). This data suggests that there are separate binding sites for DNA and RPA70AB on the XPA globular core (Figure 3.6). Importantly, this initial experiment revealed that XPA DBD is less soluble at lower salt concentration buffers, and is unable to be concentrated to an appropriate range to fully characterize its weak binding to RPA70AB. Since the C-terminal extension required for DNA binding was not involved in RPA70AB interaction, further NMR studies were performed using more soluble XPA_{98-219} construct. Titration of ^{15}N-RPA70AB and unlabeled XPA_{98-219} induced broadening of residues that overlap with the ssDNA binding site of RPA70AB (Figure A.2). In order to

determine if there is competition between ssDNA and XPA globular core on RPA70AB binding site, a follow up titration was performed in the presence of ssDNA. These data indicate that the XPA globular core, RPA70AB, and ssDNA form a ternary complex with independent binding sites (Figure A.3). Importantly, together the experiments identified residues involved in the XPA-RPA70AB interaction, which will allow the creation of a model of the complex.

Initial sets of XPA mutations to disrupt each of the two RPA binding residues were designed and will be created in both FL and DBD constructs. We plan to collaborate on functional analysis of these mutations with Dr. Orlando Schärer at Cold Spring Harbor and Dr. Zachary Nagel at Harvard. Mutant FL protein cDNA will be sent to their laboratories so that these mutations can be tested for their effect on NER efficiency or recruitment of downstream NER factors in cellular studies at the Schärer lab, and possible involvement in other DNA repair pathways in the Nagel lab.

Materials and Methods

Protein Production

Recombinant expression plasmids for pBG100 XPA_{98-219}, XPA_{98-239} were described previously[53]. pSV281 RPA70AB ($RPA70_{181-422}$) was constructed and made available by L. S. Mizoue, Center for Structural Biology, Vanderbilt University. All constructs contain N-terminal 6xHis tag cleavable by H3C (for

pBG100 plasmids) or TEV (for pSV281 plasmid) proteases. All proteins were purified as described previously[53]. Purification of RPA70AB followed the same protocol as XPA constructs with the exception of the use of different proteases for His-tag cleavage.

DNA Substrate

Desalted 8-mer ssDNA (dC8) substrates were purchased from Sigma-Agldrich Co. (St. Louis, MO). dC8 were dissolved in the NER buffer (see below) to the stock concentration of 1 mM.

NMR Titrations

All samples for DNA titrations were concentrated to 50 - 300 μM in the NMR buffer containing 20 mM Tris, pH 7.0, 75 mM KCl, 1 mM Tris(2-carboxymethil) phosphine (TCEP), and 5 % $^{2}H_2O$. Titrations were performed at 25 °C using a Bruker *AVANCE* 900 MHz spectrometer equipped with a cryoprobe following the procedure described previously[53]. Titration with the presence of dC8 were performed by first collecting ^{15}N-^{1}H TROSY-HSQC of labeled protein, add 1 equimolar amount of DNA and take another spectrum, followed by the addition of unlabeled protein and the third data collection. Titration data were analyzed using available backbone resonance assignment of XPA_{98-219} (BMRB 4249), XPA_{98-239} (chapter III), or RPA70AB[122].

Acknowledgements

We thank Dr. Markus Voehler for his assistance with NMR spectroscopy. Figure 4.3 was created by Dr. Agnieszka Topolska-Woś and she will be continuing the project described in this chapter with a talented undergraduate student Caeley Gullett. We are also excited and thankful for the ongoing collaborations with the Schärer and the Nagel laboratories.

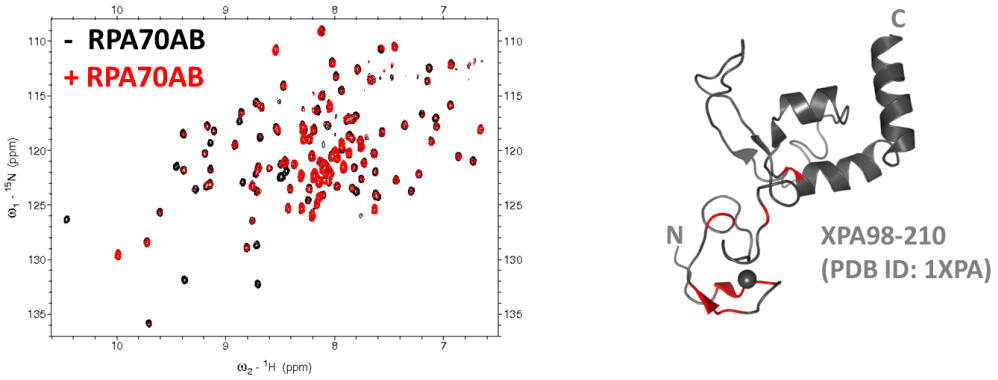


Figure A.1: NMR titration of ^{15}N XPA DBD with RPA70AB. Left - overlay of the 900 MHz ^{15}N-^{1}H TROSY-HSQC spectrum of XPA DBD in the presence (red) and absence (black) of unlabeled RPA70AB. Spectra were acquired in a buffer containing 20 mM Tris, pH 7.0, 75 mM KCl, 1 mM Tris(2-carboxymethil) phosphine (TCEP), and 5 % $^{2}H_2O$ at 25 °C. Right - CSPs from spectra shown in the left panel mapped on the 1XPA structure.

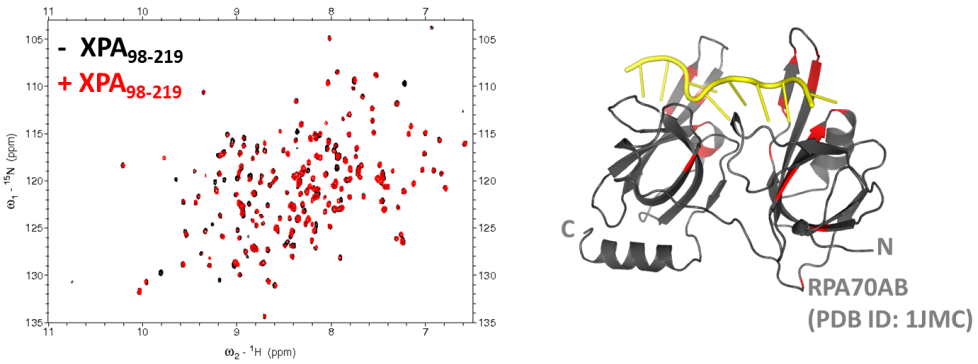


Figure A.2: NMR titration of ^{15}N RPA70AB with XPA$_{98\text{-}219}$. Left - overlay of the 900 MHz ^{15}N-^{1}H TROSY-HSQC spectrum of RPA70AB in the presence (red) and absence (black) of unlabeled XPA$_{98\text{-}219}$. Spectra were acquired in a buffer containing 20 mM Tris, pH 7.0, 75 mM KCl, 1 mM Tris(2-carboxymethil) phosphine (TCEP), and 5 % $^{2}H_2O$ at 25 °C. Right - CSPs from spectra shown in the left panel mapped on the 1JMC structure.

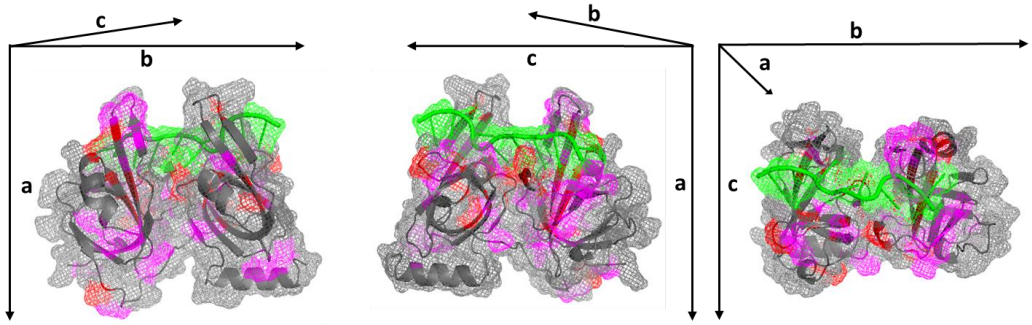


Figure A.3: NMR titration of ^{15}N RPA70AB with XPA$_{98\text{-}219}$ in the presence of ssDNA. Three orientations of 1JMC structures indicating CSPs induced by the addition of ssDNA (purple), and the addition of XPA$_{98\text{-}219}$ to the pre-formed RPA70AB-DNA complex (red). DNA molecule in the 1JMC structure is highlighted in green.

B. TOWARDS STRUCTURAL ANALYSIS OF THE XPA-RPA COMPLEX

Introduction

Previous work from the lab defined the molecular basis of the first contact point between XPA and RPA (N-terminus of XPA and RPA32C domain). Together with the new data reported here, we now have the complete map of XPA-RPA interaction residues. As a first step towards elucidating structures of the NER complex, structural investigation of full-length XPA-DNA-RPA complexes is critical. XPA and RPA are the two scaffolding proteins in human NER where each protein interacts with the NER bubble to organize the substrate. Initial SAXS analysis of XPA-DNA-RPA was conducted. However, interpretation of SAXS data for multi-domain proteins and protein complexes relies on high resolution structures or models of the components. Construction of high resolution structural model of the ternary complex of XPA globular core, ssDNA and RPA70AB is ongoing using the information obtained from this section and the resulting structural model will be a critical piece in interpreting the full-length ternary complexes.

Sample Design and Preparation

This section summarizes the preliminary results from the two sets of data collected on December 2015 at SSRL beamline and March 2016 at the SIBYLS

beamline as well as insights for the future experiments and interpretation of these results.

From December/ SSRL

1. XPA_{98-239}
2. XPA_{98-239} + DNA
3. XPA_{98-239} + DNA + RPA70AB
4. XPA
5. XPA + DNA + RPA

From March/ SIBYLS

1. XPA_{98-239}
2. XPA_{98-239} + DNA
3. XPA_{98-239} + DNA + RPA70AB
4. XPA + DNA
5. XPA + DNA + RPA

All samples were prepared in identical buffer (50 mM HEPES pH 8.5, 75 mM KCl, 10 % glycerol, 1 mM DTT). All samples were run through size-exclusion column prior to data collection. Samples were further concentrated when needed (SIBYLS recommends sample concentrations over 1 mg/ml). For each sample, 3 different concentrations (ideally, the concentration right off the sizing column and 2 serial dilutions of that concentration) and blank(s) were prepared. One blank was provided for SSRL samples and two for SIBYLS. To ensure the identity of blank to the sample buffer, first fractions from size-exclusion chromatography were used as blank for each sample. Samples containing FL proteins were provided with quite low concentrations and required concentration prior to data collection. Low concentrations of these samples could have been due to availability of the components. However, to avoid aggregations, it should be avoided in the future by running concentrated samples over the column.

While 8/12 5′ overhang (Figure A.4-B) was used for the first SAXS samples 2 and 3 sent on mid-December due to availability, I used a DNA construct similar to those used for FL complex in March (Figure A.4-C). The asymmetric Y-shaped junction for the FL complex has 7 nt duplex, 25 nt top overhang and 8 nt bottom overhang (Figure A.4-A). For samples 2 and 3 prepared on March, I used the same sequence but trimmed the top strand to 10 nt for RPA70AB binding (based on crystal structure of RPA70AB + ssDNA

complex, PDB: 1JMC, which contains 8 nt ssDNA) and bottom strand to 4 nt (Figure A.4-C). This DNA bound well to XPA_{98-239} as expected, however, when XPA_{98-239} + DNA complex was mixed with RPA70AB and run over s75 column, they didn't elute together (although there was overlap and I sent out overlapped parts to SAXS, Figure A.4-E). This separation was not observed for SSRL sample (Figure A.4-D) when similar but different DNA substrate was used.

Data Analysis

This section lists brief description of how the data was analyzed, SAXS parameters for each sample, and implications of these data. In general, data collected at SIBYLS had higher quality than those from SSRL. Each buffer and 3 concentrations of samples were scanned 10 times at SSRL and 32 times at SIBYLS. After data collection, they provided buffer subtracted intensity data. For SIBYLS data, 3 types of buffer subtracted data were provided – subtraction of first buffer, second buffer, average of the two. SIBYLS also provided a spreadsheet describing which type of buffer subtraction is most appropriate for further analysis along with other notes from data collection. SSRL also provided averaged data from 10 scans for each sample at each of the 3 concentrations.

All SAXS data were analyzed using SCATTER.

SSRL Data from December 2015

1. XPA_{98-239}

Due to low data quality, I was not able to obtain reasonable P(r) curve and analysis was aborted. Concentration dependence was observed for Rg values which could be due to aggregation. However, signal was quite weak for lowest concentration of 0.5 mg/ml. Figure A.5 summarizes parameters obtained for this sample. As indicated by the first plateau in Kratky-Debye plot in flexibility plots and low Porod exponent, the sample is flexible as expected.

2. XPA_{98-239} + DNA

Figure A.6 summarizes parameters obtained for this sample. Concentration dependence was also observed for Rg values but signal was better at higher concentrations. While still flexible, Porod exponent was slightly higher than sample 1. This indicates that XPA DBD becomes less flexible upon binding to DNA. However, as shown in Figure A.4, DNA substrate was designed to accommodate RPA70AB binding and without this molecule, the ssDNA chain that's not bound to XPA DBD can provide flexibility, although the contribution would be minor due to the small size. C-terminus of XPA is also expected to be flexible.

As indicated in Figure A.6, P(r) analysis didn't go well due to poor fit to the data. This gave Dmax value of 72 Å which could be possible considering estimated Dmax value of ~ 60 Å from the structure of the globular core of XPA (PDB: 1XPA). *Ab initio* shape calculation was performed using DAMMIF (Figure A.7). While we can't trust these shapes too much due to the low Porod exponent, 7 calculated shapes showed common elongated features. Comparing shapes 2, 3, and 5 with 1XPA structure, the structure could fit into the bottom part of these shapes and the extra tail on the top could be attributed to C-terminal extension of the protein and/or the ssDNA overhang of the substrate.

3. XPA_{98-239} + DNA + RPA70AB

Figure A.8 summarizes parameters obtained for this sample. Highest concentration had aggregation as shown from Gunier region of the intensity plot. Rg values were again concentration dependent. Porod exponent was higher than samples 2, indicating RPA70AB binding to ssDNA (and XPA) makes the complex more globular. As noted in pink in Figure A.8, P(r) analysis was tricky. Dmax value of 66 is reported here but when higher value was applied, there was better agreement between real and reciprocal Rg and I(0) values. However, at higher Rg, P(r) curve gave strange features. Figure A.9 compares calculated *ab initio* shapes and a model of this complex composed of available structures at

PDB and my NMR titrations. While Porod exponent is still not high enough to fully rely on these shapes, most calculated shapes has common diamond shapes and are similar in size with the model structure.

4. XPA

Figure A.10 summarizes parameters obtained for this sample. Concentration dependence was observed for Rg values. Smiling was observed for all intensity plots during Gunier analysis. Due to low data quality, I was not able to obtain reasonable P(r) curve and analysis was aborted. As indicated by the first plateau in Kratky-Debye plot in flexibility plots and low Porod exponent, the sample is flexible as expected.

5. XPA + DNA + RPA

Figure A.11 summarizes parameters obtained for this sample. As evident from the intensity plots, signals were quite weak for all samples most likely due to low concentrations. They also seem to be aggregated as smiling was observed from Gunier analysis.

SIBYLS Data from March 2016

The data collected at SIBYLS had higher quality than those from SSRL. As mentioned above, each sample was scanned for 32 times and averaging was done manually using AdjustTwoRegions.py.

1. $XPA_{98\text{-}239}$

Figure A.12 summarizes parameters obtained for this sample. Only highest concentration data was used for analysis because lowest concentration had bubble and medium concentration was aggregated. Flexibility plots and Porod exponent show the sample is flexible as expected. P(r) curve was asymmetric and had a feature at long distance which is most likely due to flexibility of the C-terminal residues. Main peak from P(r) lands around ~70 Å which is close to expected Dmax of the globular core of XPA (determined from 1XPA structure containing residues 98 – 210). All 13 *ab initio* shapes look elongated (Figure A.13).

2. XPA_{98-239} + DNA

Figure A.14 summarizes parameters obtained for this sample. Like SSRL samples, slight increase in Porod exponent was observed compared to sample 1 indicating that DNA binding reduces flexibility of XPA DBD. This is also observed from the change in the shape of P(r) curve between Figures A.12 and A.14. The extra feature at high Dmax shrunk upon DNA addition. There is still small tail in this region but as discussed in SSRL sample 2 section, this is most likely due to extra ssDNA in the substrate to accommodate RPA70AB which is not present in this sample.

Calculated *ab initio* shapes for this sample (Figure A.14) look less flexible compared to XPA DBD by itself (Figure A.12). However, in both cases, the envelopes look much bigger than 1XPA structure.

3. XPA_{98-239} + DNA + RPA70AB

As noted in Samples section, separation of XPA DBD and RPA70AB was observed during preparation (Figure A.4). Therefore, this data set may not represent that of homogeneous ternary complex. Figure A.15 summarizes parameters obtained for this sample. Porod exponent of 3.5 was obtained and this is highest of all samples analyzed here. As noted in Figure A.15 in pink, P(r)

curve shows feature at high Dmax, which could be due to flexible DNA extensions. *Ab initio* shapes for this sample (Figure A.16) look more rigid compared to Figures A.13 and A.14 and the envelopes are similar in size with the model structure. Interestingly, Dmax obtained for this sample was almost 2-fold bigger than the comparable sample sent to SSRL. This could be due to the difference in DNA substrate and the separation problem encountered when this sample was prepared. I'm suspecting this sample may be just RPA70AB bound to DNA.

4. XPA + DNA

Figure A.18 summarizes parameters obtained for this sample. As noted in Samples section, this sample was provided with very low concentration and required significant concentration. Due to low sample quality, I was not able to complete P(r) analysis for this sample.

5. XPA + DNA + RPA

Figure A.19 summarizes parameters obtained for this sample. Although sample concentrations were quite low, medium and high concentration samples had reasonable signal due to the size of the complex. This sample appeared to be

highly flexible with the Porod exponent of 2.1. This is expected from the flexibility arising from N- and C-terminus of XPA as well as the linkers between RPA. During first analysis, I obtained Dmax above 200 Å. However, this value gave strange feature in P(r) curve as seen in Figure A.16. Therefore, I repeated P(r) analysis with lower Dmax of 190 Å and more reasonable curve was obtained. Main peak in this curve lands around 120 Å and other features are observed beyond this point.

Ab initio shapes were also calculated for this sample (Figure A.20). As expected, these shapes are much bigger compared to those from XPA DBD complexes.

Conclusion

Although data introduced here are preliminary, these initial data collection and analysis revealed hints to improve future SAXS analysis of XPA complexes. For the initial analysis, the goal was to see how XPA-DNA-RPA complex looks like. For this reason, samples with truncated proteins were designed to resemble the condition from FL complex. However, it sometimes caused problems interpreting the data by creating extra flexibility (as in the case for sample 2) or

prevented complex formation due to different DNA substrate design (sample 3). Below is the list of things to consider for future SAXS analysis.

1. Data qualities were higher for those collected at SIBYLS than SSRL.
2. It would be interesting to compare XPA_{98-219} with XPA_{98-239}.
3. For sample 2 (XPA_{98-239} + DNA), DNA substrate with shorter ssDNA overhang should be used to avoid flapping of nucleotides not engaged in XPA binding. From my previous DNA substrate screening, 8/4 5′ overhang substrate (8 nt duplex and 4 nt ssDNA overhang) had sufficient affinity to XPA DBD.
4. For sample 3 (XPA_{98-239} + DNA + RPA70AB), DNA substrate with single overhang and not double overhang should be used to ensure complex formation.
5. In order to avoid aggregation problems, running sizing column just before SAXS analysis should be considered.
6. Sample 4 and 5 are very flexible systems. Since we know that RPA70N doesn't interact with XPA nor DNA, a complex with FL-XPA + DNA + and RPA construct without 70N should also be analyzed for comparison. C-terminal of XPA beyond residue 239 is not engaged in DNA nor RPA interaction so XPA1-239 would be a better construct for SAXS analysis.
7. For the analysis of sample 5 (XPA-DNA-RPA) comparisons with Dr. Chris Brosey's SAXS data on RPA DBC with ssDNA is required.

8. For both samples 3 and 5, different junction DNA substrates should be tested.

Acknowledgements

We thank: SSRL and SIBYLS beamlines and Kathryn Burnett for the assistance for data collection; Dr. Greg Hura and Dr. Matthew Thompson for their help with the data analysis; Dr. Steven M. Shell, Reid Putney, and Kateryna Ogorodnik for optimization and preparation of full length XPA-RPA-DNA complex. This work will be continued by Dr. Agnieszka Topolzka-Woś

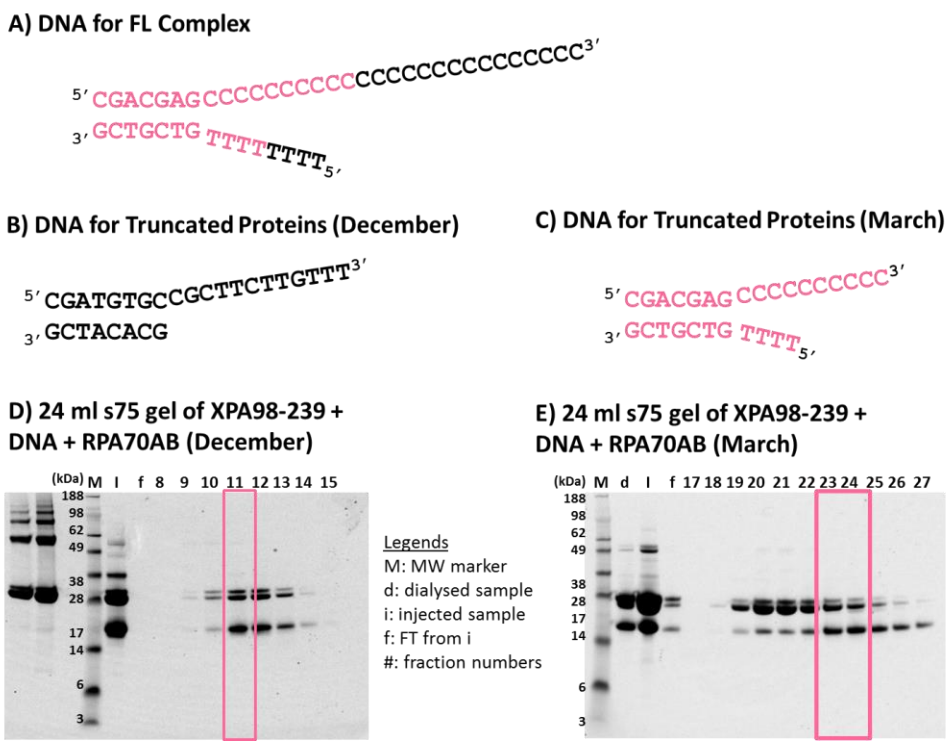
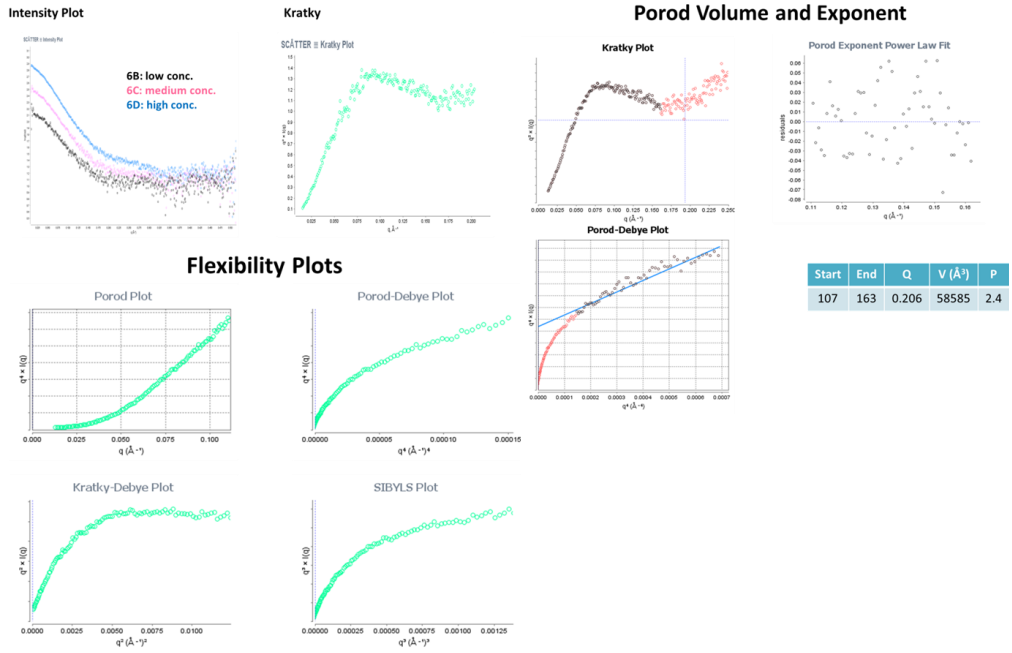


Figure A.4: DNA Substrates and SDS-Page gel for Sample 3. DNA substrates used for samples 5 and 4 (A, no DNA for sample 4 for SSRL), samples 2 and 3 for SSRL (B), and for SIBYLS (C). Pink indicates identical sequence. SDS-Page gel from size-exclusion purification of sample 3 for SSRL (D) and SIBYLS (E). Pink box indicate fractions pooled for SAXS analysis.

		SCATTER						
Sample #	Conc. (mg/ml)		strat	end	I(0) real	I(0) reci	Rg real	Rg reci	Dmax
6B	0.5	not used	9	553		1.35E+02		21.4	
6C	1	merged	8553 (92)			3.06E+02		24	
6D	2	merged	9 (63)	554		7.13E+02		26.5	
		ScaleMerge							



Start	End	Q	V (Å³)	P
107	163	0.206	58585	2.4

Figure A.5: Summary of SAXS analysis of XPA_{98-239} from SSRL. Table on the top lists sample ID, concentrations, data used for further analysis are indicated. 'Start' and 'End' columns indicate the range of data points used for analysis, if merged data was created for further analysis, the range of merged curves are indicated in parenthesis. Real and reciprocal I(0) and Rg as well as Dmax values are indicated if available. Below the table, intensity plots, Kracky plot, plots for flexibility and Porod analysis are also shown. P(r) curve as well as fit of calculated data to the low data are also shown if available.

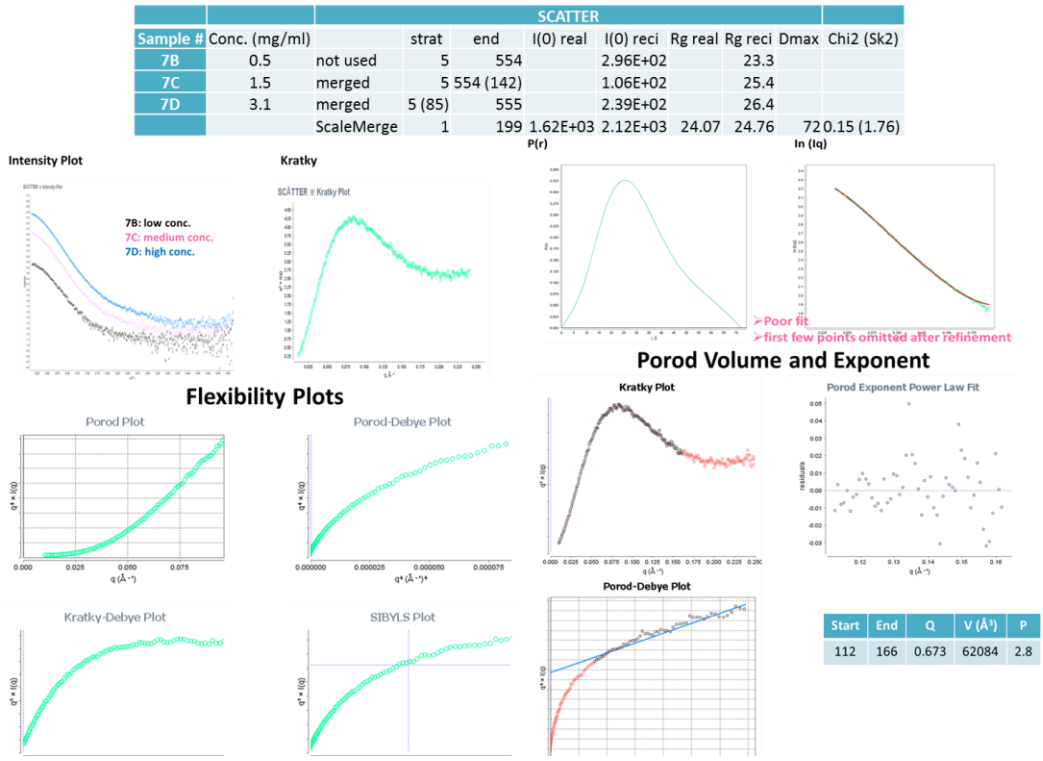


Figure A.6: Summary of SAXS analysis of XPA_{98-239} + DNA from SSRL. Table on the top lists sample ID, concentrations, data used for further analysis are indicated. 'Start' and 'End' columns indicate the range of data points used for analysis, if merged data was created for further analysis, the range of merged curves are indicated in parenthesis. Real and reciprocal I(0) and Rg as well as Dmax values are indicated if available. Below the table, intensity plots, Kracky plot, plots for flexibility and Porod analysis are also shown. P(r) curve as well as fit of calculated data to the low data are also shown if available.

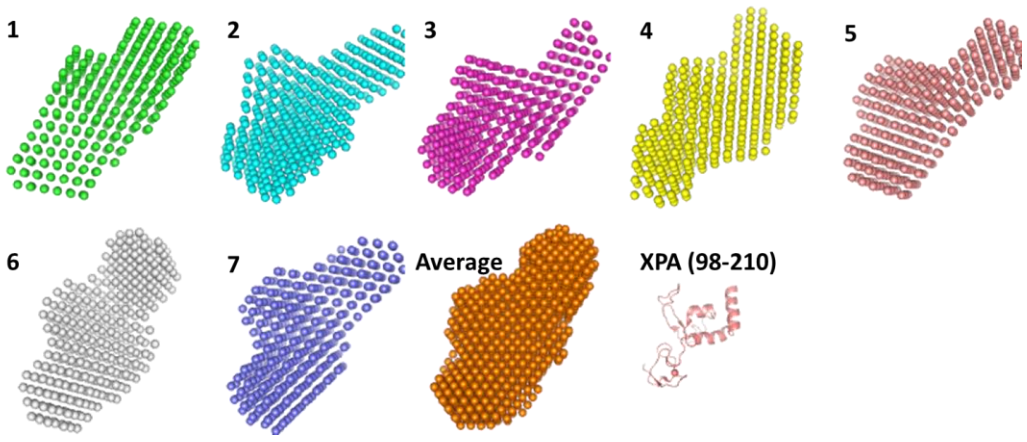
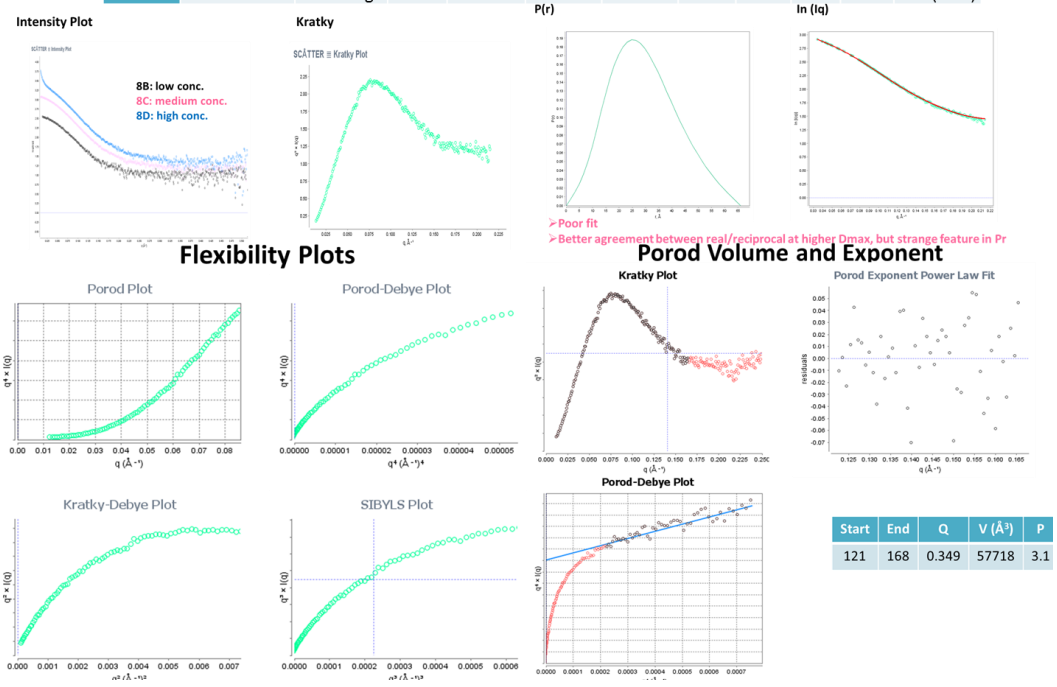


Figure A.7: *Ab initio* shape calculations for XPA_{98-239} + DNA from SSRL.

		SCATTER									
Sample #	Conc. (mg/ml)		strat	end	I(0) real	I(0) reci	Rg real	Rg reci	Dmax	R(ave)	Chi2 (Sk2)
8B	0.5	merged	7	552 (115)		3.60E+02		24.4			
8C	1.5	merged	13 (64)	552		1.17E+02		25.9			
8D	3.1	not used	1	553		8.15E+02		124.9			
		Scalemerge	1	221	8.09E+02	1.02E+03	22.54	24.03	66	29.3	0.59 (0.93)



Start	End	Q	V (Å³)	P
121	168	0.349	57718	3.1

Figure A.8: Summary of SAXS analysis of XPA_{98-239} + DNA + RPA70AB from SSRL. Table on the top lists sample ID, concentrations, data used for further analysis are indicated. 'Start' and 'End' columns indicate the range of data points used for analysis, if merged data was created for further analysis, the range of merged curves are indicated in parenthesis. Real and reciprocal I(0) and Rg as well as Dmax values are indicated if available. Below the table, intensity plots, Kracky plot, plots for flexibility and Porod analysis are also shown. P(r) curve as well as fit of calculated data to the low data are also shown if available.

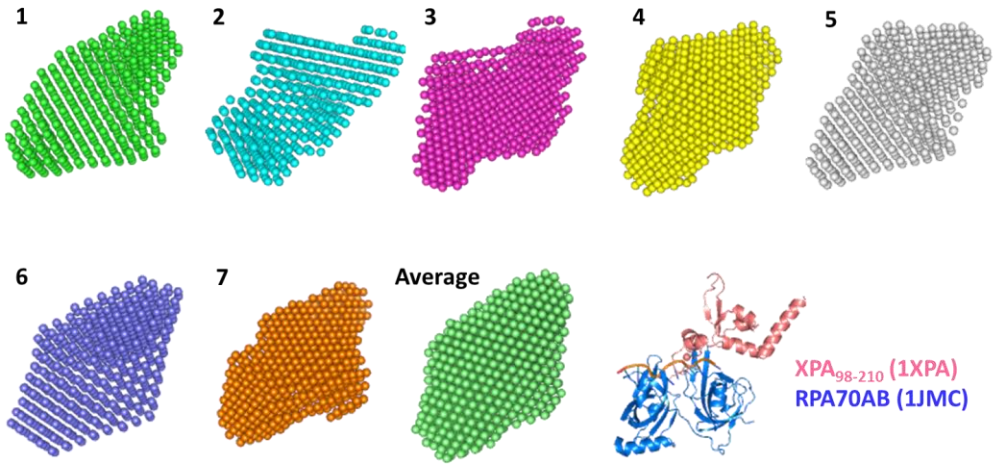
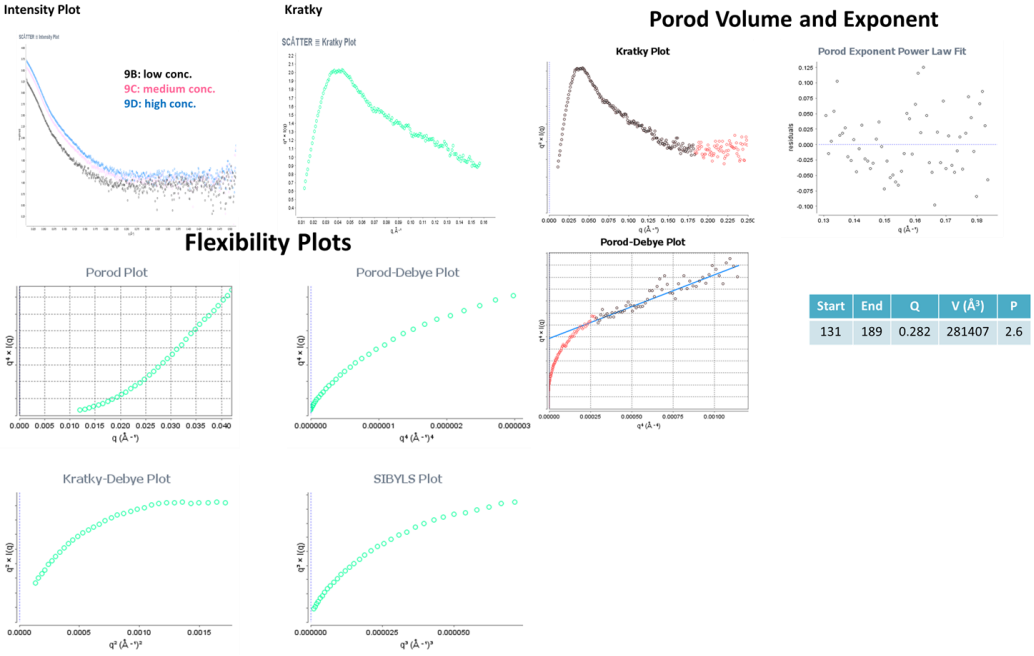


Figure A.9: *Ab initio* shape calculations for XPA_{98-239} + DNA + RPA70AB from SSRL.

		SCATTER							
Sample #	Conc. (mg/ml)		strat	end	I(0) real	I(0) reci	Rg real	Rg reci	Dmax
9B	0.5	merged	6	556 (69)		1.80E+03		55.9	
9C	1	not used	4	553		3.81E+03		57.6	
9D	1.3	merged	6 (40)	554		5.17E+03		58.3	
		ScaleMerge	1	160					N/A



Start	End	Q	V (Å³)	P
131	189	0.282	281407	2.6

Figure A.10: Summary of SAXS analysis of XPA from SSRL. Table on the top lists sample ID, concentrations, data used for further analysis are indicated. 'Start' and 'End' columns indicate the range of data points used for analysis, if merged data was created for further analysis, the range of merged curves are indicated in parenthesis. Real and reciprocal I(0) and Rg as well as Dmax values are indicated if available. Below the table, intensity plots, Kracky plot, plots for flexibility and Porod analysis are also shown. P(r) curve as well as fit of calculated data to the low data are also shown if available.

		SCATTER							
Sample #	Conc. (mg/ml)		strat	end	I(0) real	I(0) reci	Rg real	Rg reci	Dmax
10B	0.1	not used	1	527		6.14E+01		56.5	
10C	0.3	not used	1	527		5.94E+02		84.9	
10D	0.52	not used	1	527		2.16E+03		112.1	

Intensity Plot

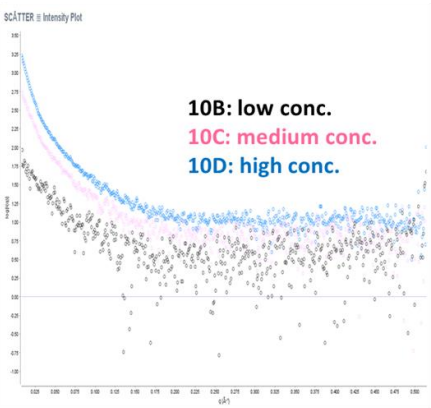


Figure A.11: Summary of SAXS analysis of XPA + DNA + RPA from SSRL.

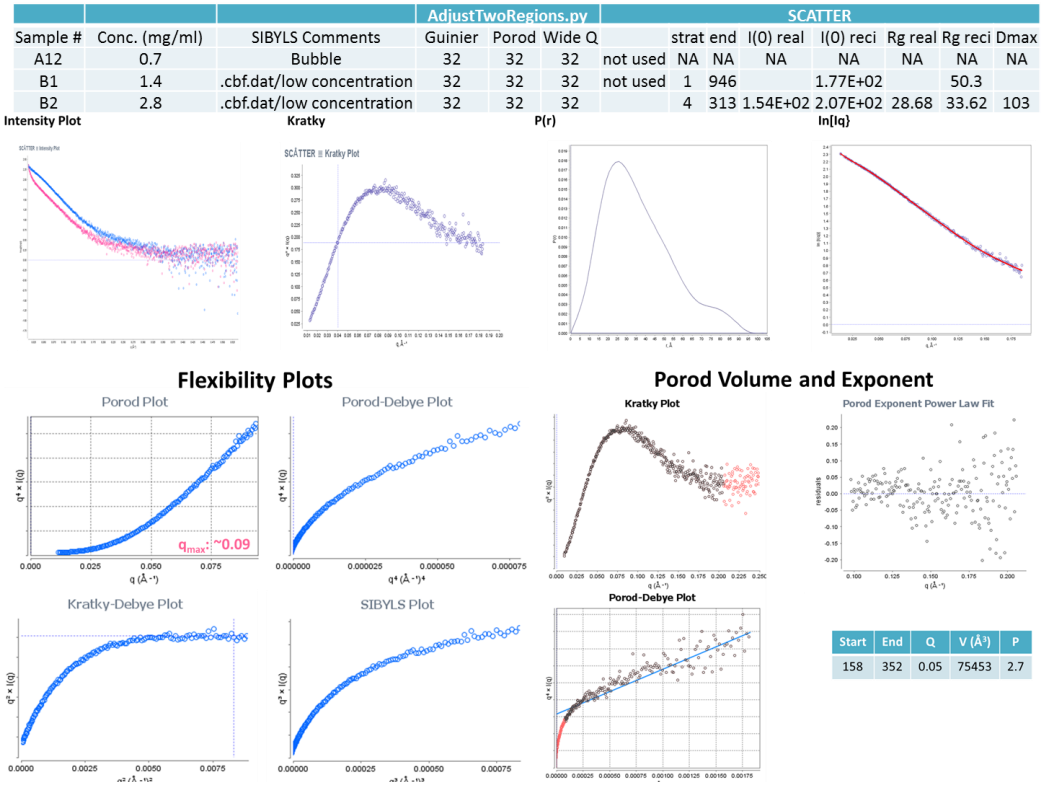


Figure A.12: Summary of SAXS analysis of XPA$_{98-239}$ from SIBYLS. Table on the top lists sample ID, concentrations, data used for further analysis are indicated. 'Start' and 'End' columns indicate the range of data points used for analysis, if merged data was created for further analysis, the range of merged curves are indicated in parenthesis. Real and reciprocal I(0) and Rg as well as Dmax values are indicated if available. Below the table, intensity plots, Kracky plot, plots for flexibility and Porod analysis are also shown. P(r) curve as well as fit of calculated data to the low data are also shown if available.

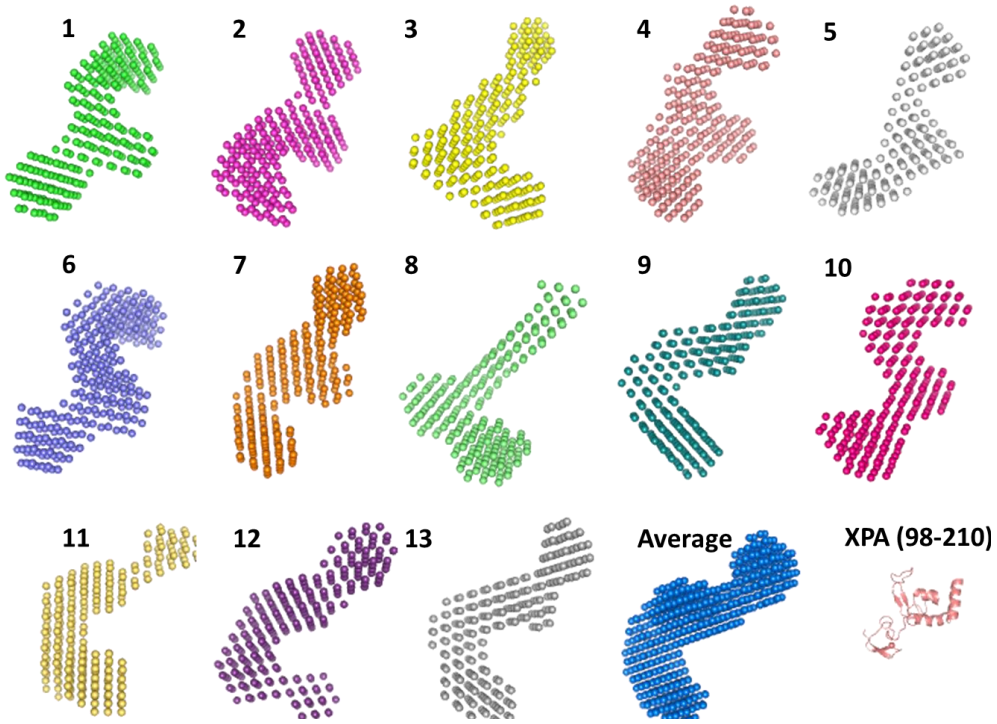


Figure A.13: *Ab initio* shape calculations for XPA_{98-239} from SIBYLS.

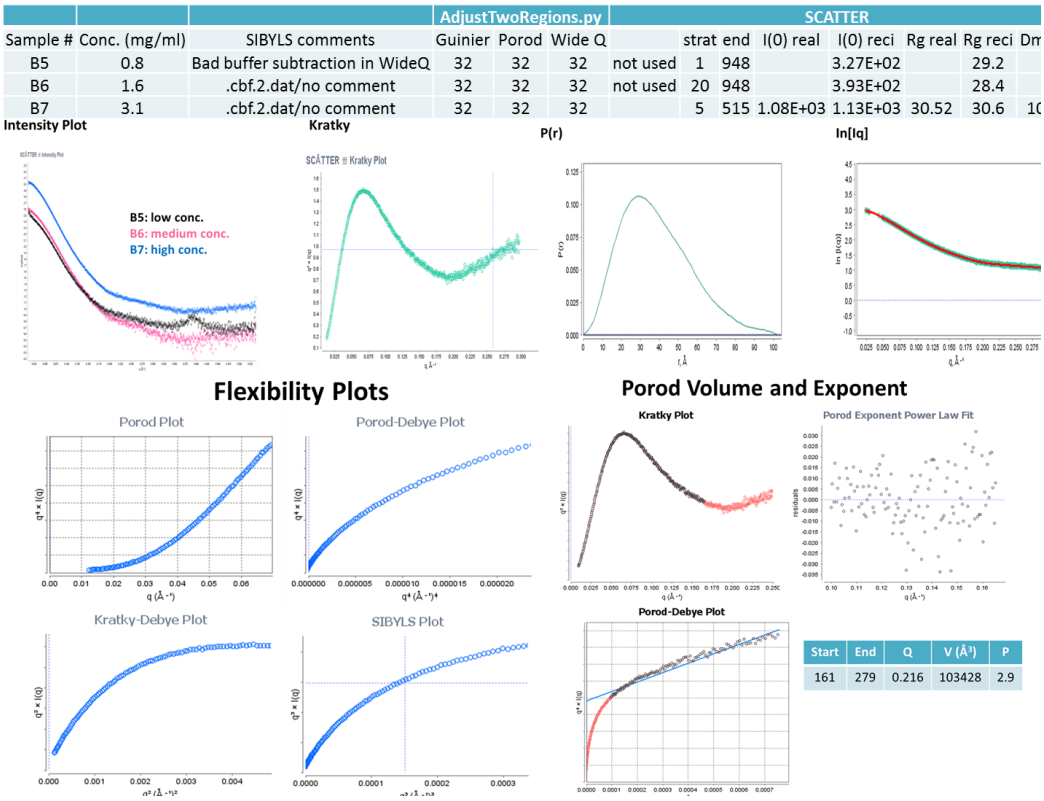


Figure A.14: Summary of SAXS analysis of XPA_{98-239} + DNA from SIBYLS. Table on the top lists sample ID, concentrations, data used for further analysis are indicated. 'Start' and 'End' columns indicate the range of data points used for analysis, if merged data was created for further analysis, the range of merged curves are indicated in parenthesis. Real and reciprocal I(0) and Rg as well as Dmax values are indicated if available. Below the table, intensity plots, Kracky plot, plots for flexibility and Porod analysis are also shown. P(r) curve as well as fit of calculated data to the low data are also shown if available.

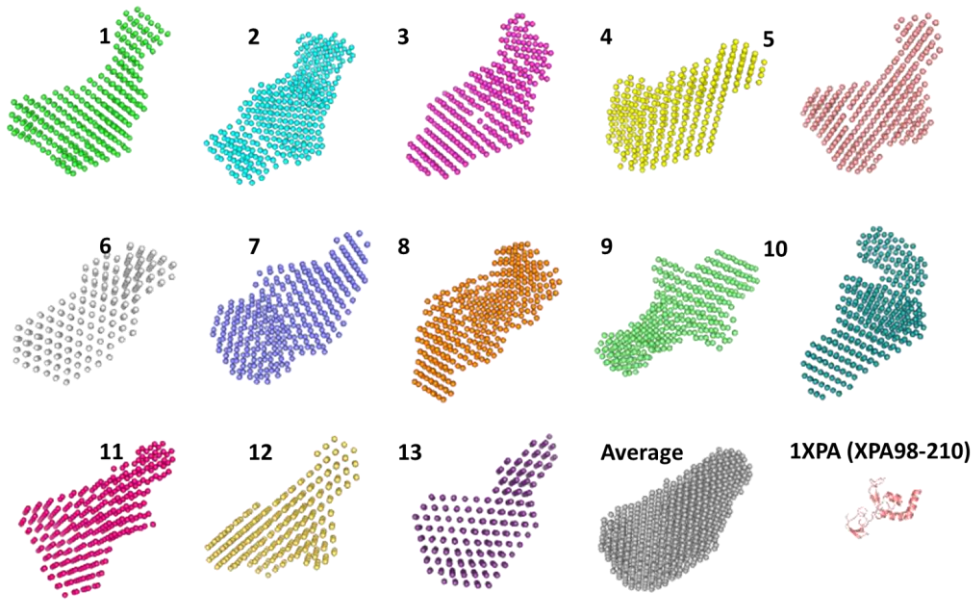


Figure A.15: *Ab initio* shape calculations for XPA_{98-239} + DNA from SIBYLS.

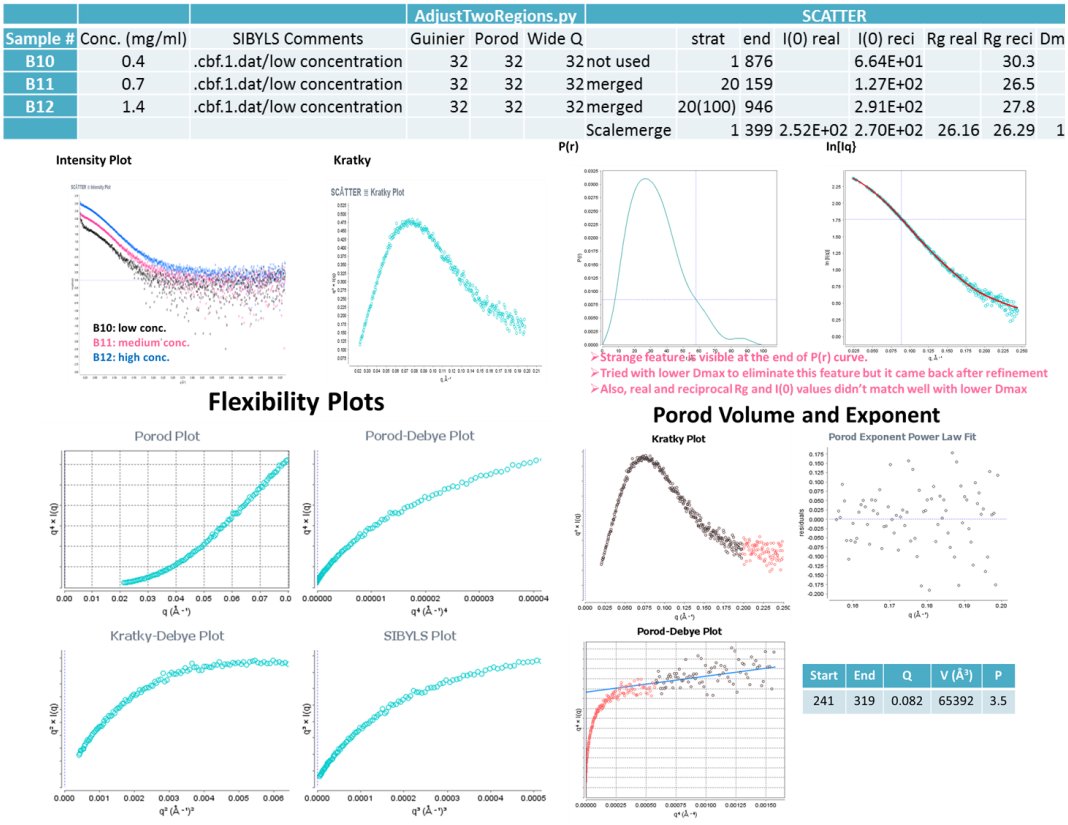


Figure A.16: Summary of SAXS analysis of XPA_{98-239} + DNA + RPA70AB from SIBYLS. Table on the top lists sample ID, concentrations, data used for further analysis are indicated. 'Start' and 'End' columns indicate the range of data points used for analysis, if merged data was created for further analysis, the range of merged curves are indicated in parenthesis. Real and reciprocal I(0) and Rg as well as Dmax values are indicated if available. Below the table, intensity plots, Kracky plot, plots for flexibility and Porod analysis are also shown. P(r) curve as well as fit of calculated data to the low data are also shown if available.

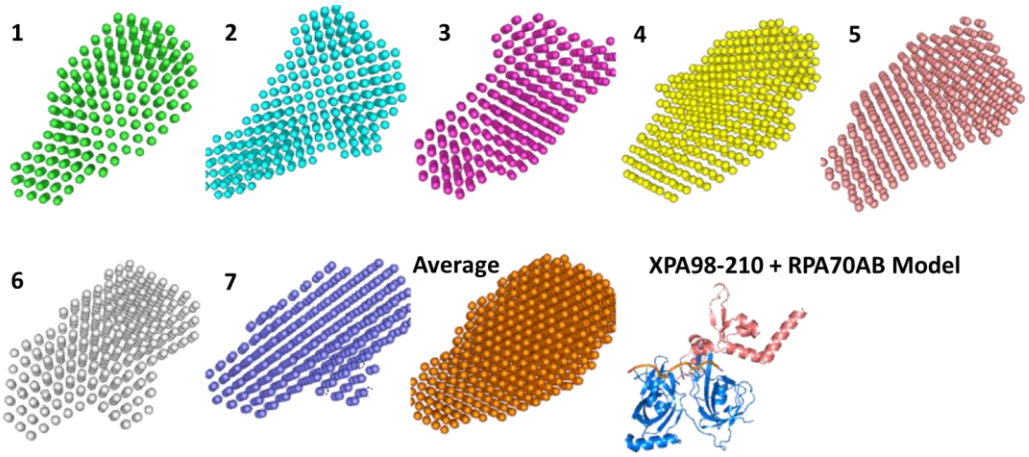


Figure A.17: *Ab initio* shape calculations for XPA_{98-239} + DNA + RPA70AB from SIBYLS.

			AdjustTwoRegions.py			SCATTER						
Sample #	Conc. (mg/ml)	SIBYLS Comments	Guinier	Porod	Wide Q		strat	end	I(0) real	I(0) reci	Rg real	Rg reci	Dmax
C3	0.4	.cbf.dat/low concentration	32	32	32	merged	4	184		2.01E+02		48.2	
C4	0.7	.cbf.dat/low concentration	32	32	32	not used	1	909		3.05E+02		159	
C5	1.3	.cbf.dat/low concentration	32	32	32	merged	4 (113)	925		1.24E+03		66.4	
						ScaleMerge	1	299					N/A

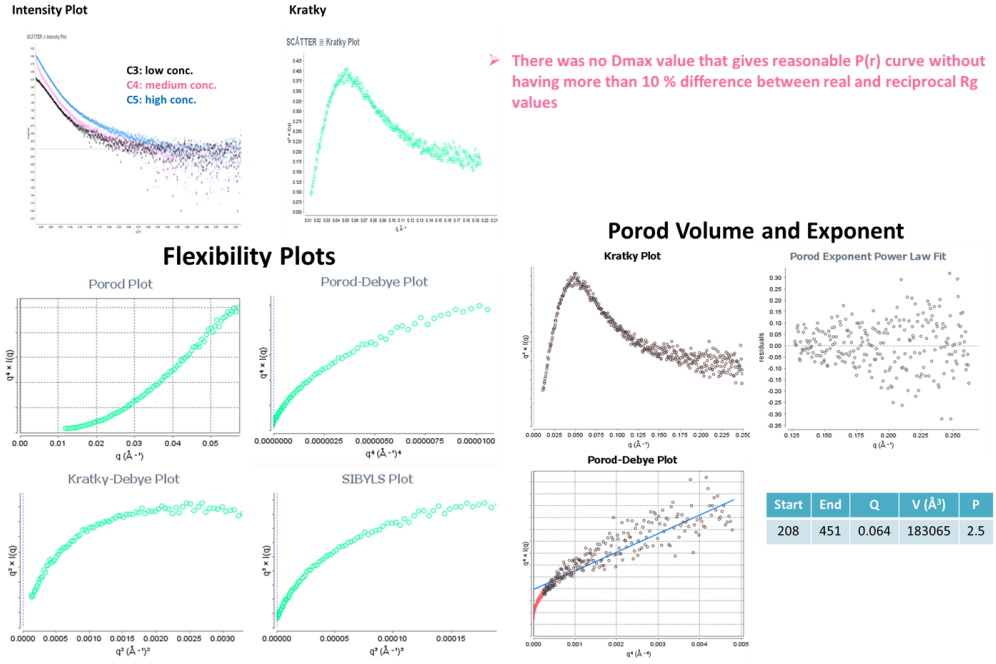
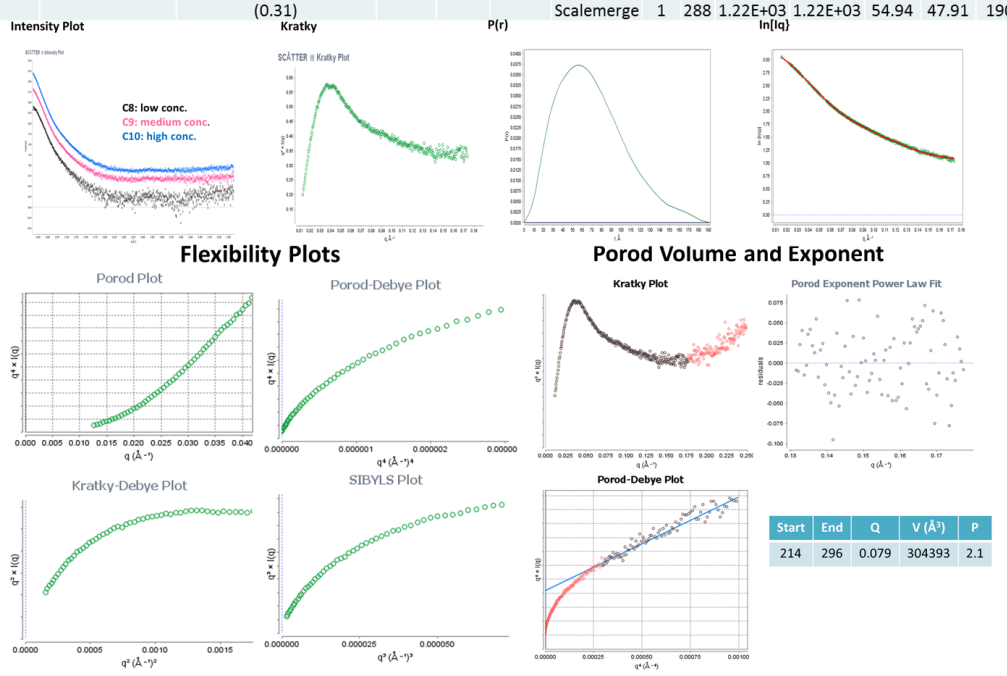


Figure A.18: Summary of SAXS analysis of XPA + DNA from SIBYLS. Table on the top lists sample ID, concentrations, data used for further analysis are indicated. 'Start' and 'End' columns indicate the range of data points used for analysis, if merged data was created for further analysis, the range of merged curves are indicated in parenthesis. Real and reciprocal I(0) and Rg as well as Dmax values are indicated if available. Below the table, intensity plots, Kracky plot, plots for flexibility and Porod analysis are also shown. P(r) curve as well as fit of calculated data to the low data are also shown if available.

			AdjustTwoRegions.py			SCATTER							
Sample #	Conc. (mg/ml)	SIBYLS comments	Guinier	Porod	Wide Q		strat	end	I(0) real	I(0) reci	Rg real	Rg reci	Dmax
C8	0.3	.cbf.dat/low concentration	32	32	32	not used	4	948		2.79E+02		50.2	
C9	0.6	.cbf.dat/no comment	32	32	32	merged	5	130		6.70E+02		54.9	
C10	1.2	.cbf.dat/no comment	32	32	32	merged	56	948		1.51E+03		63.3	
		Rave=69.5, chi2(Sk2)=0.12 (0.31)				Scalemerge	1	288	1.22E+03	1.22E+03	54.94	47.91	190



Start	End	Q	V (Å³)	P
214	296	0.079	304393	2.1

Figure A.19: Summary of SAXS analysis of XPA + DNA + RPA from SIBYLS. Table on the top lists sample ID, concentrations, data used for further analysis are indicated. 'Start' and 'End' columns indicate the range of data points used for analysis, if merged data was created for further analysis, the range of merged curves are indicated. Real and reciprocal I(0) and Rg as well as Dmax values are indicated if available. Below the table, intensity plots, Kracky plot, plots for flexibility and Porod analysis are also shown. P(r) curve as well as fit of calculated data to the low data are also shown if available.

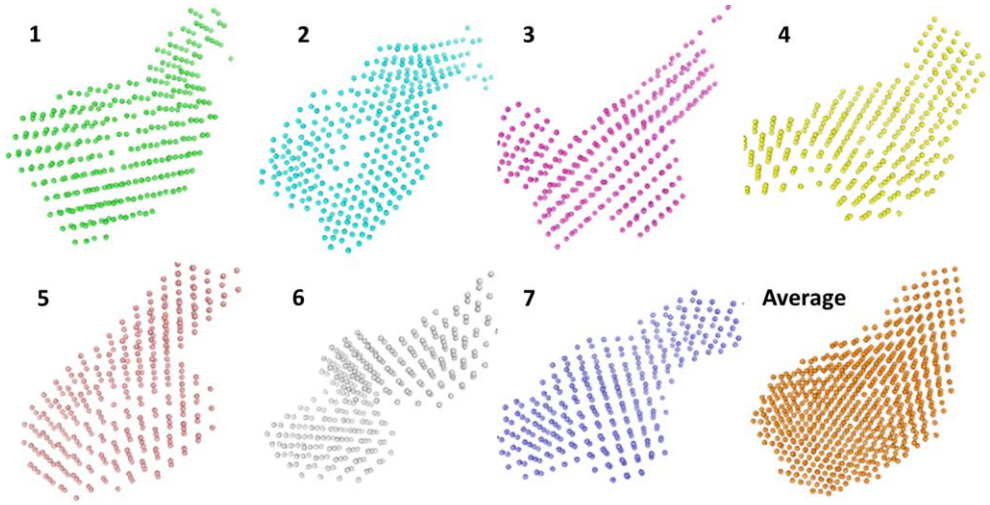


Figure A.20: *Ab initio* shape calculations for XPA + DNA + RPA from SIBYLS.

BIBLIOGRAPHY

1. Gillet, L. C. J. & Scharer, O. D. Molecular Mechanisms of Mammalian Global Genome Nucleotide Excision Repair. *Chem. Rev.* **106,** 253–276 (2006).
2. Scharer, O. D. Nucleotide Excision Repair in Eukaryotes. *Cold Spring Harb. Perspect. Biol.* **5,** 1–19 (2013).
3. Riedl, T., Hanaoka, F. & Egly, J. The comings and goings of nucleotide excision repair factors on damaged DNA. *EMBO J.* **22,** 5293–5303 (2003).
4. DiGiovanna, J. J. & Kraemer, K. H. Shining a light on xeroderma pigmentosum. *J. Invest. Dermatol.* **132,** 785–96 (2012).
5. Hoeijmakers, J. H. J. DNA damage, aging, and cancer. *N. Engl. J. Med.* **361,** 1475–85 (2009).
6. Lehmann, A. R., McGibbon, D. & Stefanini, M. Xeroderma pigmentosum. *Orphanet J. Rare Dis.* **6,** 70 (2011).
7. Lehmann, A. R. DNA repair, DNA replication and human disorders: A personal journey. *DNA Repair (Amst).* **11,** 328–334 (2012).
8. States, J. C., McDuffie, E. R., Myrand, S. P., McDowell, M. & Cleaver, J. E. Distribution of mutations in the human xeroderma pigmentosum group A gene and their relationships to the functional regions of the DNA damage recognition protein. *Hum. Mutat.* **12,** 103–113 (1998).

9. Truglio, J. J., Croteau, D. L., van Houten, B. & Kisker, C. Prokaryotic nucleotide excision repair: The UvrABC system. *Chem. Rev.* **106,** 233–252 (2006).

10. Mahadevan, B. *et al.* Genetic toxicology in the 21st century: reflections and future directions. *Environ. Mol. Mutagen* **52,** 339–354 (2011).

11. Bradford, P. T. *et al.* Cancer and neurologic degeneration in xeroderma pigmentosum: long term follow-up characterises the role of DNA repair. *J. Med. Genet.* **48,** 168–76 (2011).

12. Cleaver, J. E. Cancer in xeroderma pigmentosum and related disorders of DNA repair. *Nat. Rev. Cancer* **5,** 564–73 (2005).

13. Cleaver, J. E. & States, J. C. The DNA damage-recognition problem in human and other eukaryotic cells : the XPA damage binding protein. *Biochem J* **12,** 1–12 (1997).

14. Sancar, A. & Hearst, J. E. Molecular matchmakers. *Science* **259,** 1415–1420 (1993).

15. Lehmann, A. R. DNA polymerases and repair synthesis in NER in human cells. *DNA Repair (Amst).* **10,** 730–3 (2011).

16. Shell, S. M. *et al.* Xeroderma pigmentosum complementation group C protein (XPC) serves as a general sensor of damaged DNA. *DNA Repair (Amst).* **12,** 947–53 (2013).

17. Sancar, B. A. & Reardon, J. T. NUCLEOTIDE EXCISION REPAIR IN E .

COLI AND MAN. *Adv. Protein Chem.* **69,** 43–71 (2004).

18. Rouillon, C. & White, M. F. The evolution and mechanisms of nucleotide excision repair proteins. *Res. Microbiol.* **162,** 19–26 (2011).

19. Spivak, G. Transcription-coupled repair: an update. *Arch. Toxicol.* 1–12 (2016). doi:10.1007/s00204-016-1820-x

20. Batty, D., Rapic'-Otrin, V., Levine, a S. & Wood, R. D. Stable binding of human XPC complex to irradiated DNA confers strong discrimination for damaged sites. *J. Mol. Biol.* **300,** 275–290 (2000).

21. Uchida, A. *et al.* The carboxy-terminal domain of the XPC protein plays a crucial role in nucleotide excision repair through interactions with transcription factor IIH. *DNA Repair (Amst).* **1,** 449–461 (2002).

22. Min, J. & Pavletich, N. Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* **449,** 570–575 (2007).

23. Yokoi, M. *et al.* The xeroderma pigmentosum group C protein complex XPC-HR23B plays an important role in the recruitment of transcription factor IIH to damaged DNA. *J. Biol. Chem.* **275,** 9870–9875 (2000).

24. Fuss, J. O. & Tainer, J. a. XPB and XPD helicases in TFIIH orchestrate DNA duplex opening and damage verification to coordinate repair with transcription and cell cycle via CAK kinase. *DNA Repair (Amst).* **10,** 697–713 (2011).

25. Hermanson-Miller, I. L. & Turchi, J. J. Strand-specific binding of RPA and

XPA to damaged duplex DNA. *Biochemistry* **41,** 2402–2408 (2002).

26. Krasikova, Y. S., Rechkunova, N. I., Maltseva, E. a, Petruseva, I. O. & Lavrik, O. I. Localization of xeroderma pigmentosum group A protein and replication protein A on damaged DNA in nucleotide excision repair. *Nucleic Acids Res.* **38,** 8083–94 (2010).

27. Tsodikov, O. V *et al.* Structural basis for the recruitment of ERCC1-XPF to nucleotide excision repair complexes by XPA. *EMBO J.* **26,** 4768–76 (2007).

28. O'Donovan, A., Davies, A. & Moggs, J. XPG endonuclease makes the 3′ incision in human DNA nucleotide excision repair. *Nature* **371,** 432–435 (1994).

29. Staresincic, L. *et al.* Coordination of dual incision and repair synthesis in human nucleotide excision repair. *EMBO J.* **28,** 1111–20 (2009).

30. Shivji, M., Moggs, J., Kuraoka, I. & Wood, R. Dual-incision assays for nucleotide excision repair using DNA with a lesion at a specific site. *DNA Repair Protoc.* **113,** 373–392 (1999).

31. Kraemer, K. H. *et al.* Genetic Heterogeneity in Xeroderma Pigmentosum : Complementation Groups and Their Relationship to DNA Repair Rates Complen. *Proc. Natl. Acad. Sci. U. S. A.* **72,** 59–63 (1975).

32. Kraemer, K. Five complementation groups in xeroderma pigmentosum. *Mutat. Res. …* **33,** 327–339 (1975).

33. Arase, S., Kozuka, T. & Tanaka, K. A sixth complementation group in xeroderma pigmentosum. *Mutat. Res.* **59,** 143–146 (1979).

34. Keijzer, W. & Jaspers, N. A seventh complementation group in excision-deficient xeroderma pigmentosum. *Mutat. Res. …* **62,** 183–190 (1979).

35. Cleaver, J. Defective repair replication of DNA in xeroderma pigmentosum. *Nature* **218,** 652–656 (1968).

36. Feltes, B. C. & Bonatto, D. Overview of xeroderma pigmentosum proteins architecture, mutations and post-translational modifications. *Mutat. Res. Rev. Mutat. Res.* **763,** 306–20 (2015).

37. Yang, Z. *et al.* Specific and efficient binding of xeroderma pigmentosum complementation group A to double-strand/single-strand DNA junctions with 3′- and/or 5′-ssDNA branches. *Biochemistry* **45,** 15921–15930 (2006).

38. Li, C.-L. *et al.* Tripartite DNA Lesion Recognition and Verification by XPC, TFIIH, and XPA in Nucleotide Excision Repair. *Mol. Cell* **59,** 1025–34 (2015).

39. Ikegami, T. *et al.* Solution structure of the DNA- and RPA-binding domain of the human repair factor XPA. *Nat. Struct. Biol.* **5,** 701–706 (1998).

40. Buchko, G. W., Ni, S., Thrall, B. D. & Kennedy, M. A. Structural features of the minimal DNA binding domain (M98 – F219) of human nucleotide excision repair protein XPA. *Nucleic Acids Res.* **26,** 2779–2788 (1998).

41. Kuraoka, I. *et al.* Identification of a damaged-DNA binding domain of the

XPA protein. *Mutat. Res.* **362,** 87–95 (1996).

42. Bernardes de Jesus, B. M., Bjørås, M., Coin, F. & Egly, J. M. Dissection of the molecular defects caused by pathogenic mutations in the DNA repair factor XPC. *Mol. Cell. Biol.* **28,** 7225–7235 (2008).

43. Li, L., Elledge, S. J., Peterson, C. a, Bales, E. S. & Legerski, R. J. Specific association between the human DNA repair proteins XPA and ERCC1. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 5012–6 (1994).

44. Nocentini, S., Coin, F., Saijo, M., Tanaka, K. & Egly, J.-M. DNA Damage Recognition by XPA Protein Promotes Efficient Recruitment of Transcription Factor II H. *J. Biol. Chem.* **272,** 22991–22994 (1997).

45. Buchko, G. W. *et al.* Interactions of human nucleotide excision repair protein XPA with DNA and RPA70 Delta C327: chemical shift mapping and 15N NMR relaxation studies. *Biochemistry* **38,** 15116–15128 (1999).

46. Mer, G. *et al.* Structural basis for the recognition of DNA repair proteins UNG2, XPA, and RAD52 by replication factor RPA. *Cell* **103,** 449–456 (2000).

47. Neher, T. M., Shuck, S. C., Liu, J.-Y., Zhang, J.-T. & Turchi, J. J. Biochemical analysis of the damage recognition process in nucleotide excision repair. *J. Biol. Chem.* **278,** 7476–85 (2003).

48. Wakasugi, M. *et al.* Physical and functional interaction between DDB and XPA in nucleotide excision repair. *Nucleic Acids Res.* **37,** 516–25 (2009).

49. McNeil, E. M. & Melton, D. W. DNA repair endonuclease ERCC1-XPF as a novel therapeutic target to overcome chemoresistance in cancer therapy. *Nucleic Acids Res.* **40,** 9990–10004 (2012).

50. Giglia-Mari, G. *et al.* Dynamic interaction of TTDA with TFIIH is stabilized by nucleotide excision repair in living cells. *PLoS Biol.* **4,** 0952–0963 (2006).

51. Ziani, S. *et al.* Sequential and ordered assembly of a large DNA repair complex on undamaged chromatin. *J. Cell Biol.* **206,** 589–98 (2014).

52. Buchko, G. W. *et al.* DNA-XPA interactions: a (31)P NMR and molecular modeling study of dCCAATAACC association with the minimal DNA-binding domain (M98-F219) of the nucleotide excision repair protein XPA. *Nucleic Acids Res.* **29,** 2635–2643 (2001).

53. Sugitani, N., Shell, S. M., Soss, S. E. & Chazin, W. J. Redefining the DNA-Binding Domain of Human XPA. *J. Am. Chem. Soc.* **136,** 10830–10833 (2014).

54. Hilton, B. *et al.* A new structural insight into XPA-DNA interactions. *Biosci. Rep.* **34,** 831–840 (2014).

55. Koch, S. C. *et al.* Structural insights into the recognition of cisplatin and AAF-dG lesion by Rad14 (XPA). *Proc. Natl. Acad. Sci. U. S. A.* **112,** 8272–7 (2015).

56. Rademakers, S. *et al.* Xeroderma Pigmentosum Group A Protein Loads as

a Separate Factor onto DNA Lesions. *Mol. Cell. Biol.* **23,** 5755–5767 (2003).

57. Yang, Z., Liu, Y., Mao, L., Zhang, J. & Zou, Y. Dimerization of human XPA and formation of XPA2-RPA protein complex. *Biochemistry* 13012–13020 (2002). at <http://pubs.acs.org/doi/abs/10.1021/bi026064z>

58. Shell, S. & Zou, Y. Other proteins interacting with XP proteins. *Adv. Exp. Med. Biol.* **637,** 103–112 (2008).

59. Krasikova, Y. S. *et al.* Influence of centrin 2 on the interaction of nucleotide excision repair factors with damaged DNA. *Biochem. Biokhimiia* **77,** 346–53 (2012).

60. Bunick, C., Miller, M. & Fuller, B. Biochemical and structural domain analysis of xeroderma pigmentosum complementation group C protein. *Biochemistry* **45,** 14965–14979 (2006).

61. Wittschieben, B. Ø. & Wood, R. D. DDB complexities. *DNA Repair (Amst).* **2,** 1065–1069 (2003).

62. Tang, J. & Chu, G. Xeroderma pigmentosum complementation group E and UV-damaged DNA-binding protein. *DNA Repair (Amst).* **1,** 601–16 (2002).

63. Fujiwara, Y. *et al.* Characterization of DNA recognition by the human UV-damaged DNA-binding protein. *J. Biol. Chem.* **274,** 20027–20033 (1999).

64. Fei, J. *et al.* Regulation of nucleotide excision repair by UV-DDB:

prioritization of damage recognition to internucleosomal DNA. *PLoS Biol.* **9,** e1001183 (2011).

65. Camenisch, U., Dip, R., Vitanescu, M. & Naegeli, H. Xeroderma pigmentosum complementation group A protein is driven to nucleotide excision repair sites by the electrostatic potential of distorted DNA. *DNA Repair (Amst).* **6,** 1819–28 (2007).

66. Di Lello, P. *et al.* Structure of the Tfb1/p53 Complex: Insights into the Interaction between the p62/Tfb1 Subunit of TFIIH and the Activation Domain of p53. *Mol. Cell* **22,** 731–740 (2006).

67. Vitorino, M. *et al.* Solution structure and self-association properties of the p8 TFIIH subunit responsible for trichothiodystrophy. *J. Mol. Biol.* **368,** 473–80 (2007).

68. Khoo, K. H., Joerger, A. C., Freund, S. M. V & Fersht, A. R. Stabilising the DNA-binding domain of p53 by rational design of its hydrophobic core. *Protein Eng. Des. Sel.* **22,** 421–30 (2009).

69. Hilario, E., Li, Y., Nobumori, Y., Liu, X. & Fan, L. Structure of the C-terminal half of human XPB helicase and the impact of the disease-causing mutation XP11BE. *Acta Crystallogr. D. Biol. Crystallogr.* **69,** 237–46 (2013).

70. Natan, E. *et al.* Interaction of the p53 DNA-binding domain with its N-terminal extension modulates the stability of the p53 tetramer. *J. Mol. Biol.*

409, 358–368 (2011).

71. Okuda, M. & Nishimura, Y. Extended String Binding Mode of the Phosphorylated Transactivation Domain of Tumor Suppressor p53. *J. Am. Chem. Soc.* **136,** 14143–14152 (2014).

72. Andersen, G., Poterszman, A., Egly, J. M., Moras, D. & Thierry, J.-C. The Crystal Structure of Human Cyclin H. *FEBS Lett.* **397,** 65–69 (1996).

73. Gervais, V. *et al.* Solution structure of the N-terminal domain of the human TFIIH MAT1 subunit: new insights into the RING finger family. *J. Biol. Chem.* **276,** 7457–64 (2001).

74. Lolli, G., Lowe, E. D., Brown, N. R. & Johnson, L. N. The crystal structure of human CDK7 and its protein recognition properties. *Structure* **12,** 2067–2079 (2004).

75. 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.* **23,** 1–11 (2015).

76. Gervais, V. *et al.* TFIIH contains a PH domain involved in DNA nucleotide excision repair. *Nat. Struct. Mol. Biol.* **11,** 616–22 (2004).

77. Kellenberger, E. *et al.* Solution structure of the C-terminal domain of TFIIH P44 subunit reveals a novel type of C4C4 ring domain involved in

protein-protein interactions. *J. Biol. Chem.* **280,** 20785–92 (2005).

78. Schultz, P., Fribourg, S. & Poterszman, A. Molecular structure of human TFIIH. *Cell* **102,** 599–607 (2000).

79. Lafrance-Vanasse, J., Arseneault, G. V., Cappadocia, L., Legault, P. & Omichinski, J. G. Structural and functional evidence that Rad4 competes with Rad2 for binding to the Tfb1 subunit of TFIIH in NER. *Nucleic Acids Res.* **41,** 2736–2745 (2013).

80. Coin, F. *et al.* Nucleotide Excision Repair Driven by the Dissociation of CAK from TFIIH. *Mol. Cell* **31,** 9–20 (2008).

81. Park, C.-H., Mu, D., Reardon, J. T. & Sancar, A. The general transcription-repair factor TFIIH is recruited to hte excision repair complex by the XPA protein independent of the TFIIE transcription factor. *J. Biol. Chem.* **270,** 4896–4902 (1995).

82. Wold, M. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu. Rev. Biochem.* **66,** 61–92 (1997).

83. Fanning, E., Klimovich, V. & Nager, A. R. A dynamic model for replication protein A (RPA) function in DNA processing pathways. *Nucleic Acids Res.* **34,** 4126–37 (2006).

84. Stauffer, M. E. & Chazin, W. J. Structural mechanisms of DNA replication, repair, and recombination. *J. Biol. Chem.* **279,** 30915–8 (2004).

85. Sugitani, N. & Chazin, W. J. Characteristics and concepts of dynamic hub proteins in DNA processing machinery from studies of RPA. *Prog. Biophys. Mol. Biol.* **117,** 206–11 (2015).

86. De Biasio, A. & Blanco, F. J. *Proliferating cell nuclear antigen structure and interactions: too many partners for one dancer? Advances in protein chemistry and structural biology* **91,** (Elsevier Inc., 2013).

87. Lee, S.-H., Kim, D.-K. & Drissi, R. Human Xeroderma Pigmentosum Group A Protein Interacts with Human Replication Protein A and Inhibits DNA Replication. *J. Biol. Chem.* **270,** 21800–21805 (1995).

88. Daughdrill, G. W. *et al.* Chemical shift changes provide evidence for overlapping single-stranded DNA- and XPA-binding sites on the 70 kDa subunit of human replication protein A. *Nucleic Acids Res.* **31,** 4176–4183 (2003).

89. Li, L., Lu, X., Peterson, C. a & Legerski, R. J. An interaction between the DNA repair factor XPA and replication protein A appears essential for nucleotide excision repair. *Mol. Cell. Biol.* **15,** 5396–5402 (1995).

90. Saijo, M., Takedachi, A. & Tanaka, K. Nucleotide excision repair by mutant xeroderma pigmentosum group A (XPA) proteins with deficiency in interaction with RPA. *J. Biol. Chem.* **286,** 5476–83 (2011).

91. Mailand, N., Gibbs-Seymour, I. & Bekker-Jensen, S. Regulation of PCNA-protein interactions for genome stability. *Nat. Rev. Mol. Cell Biol.* **14,** 269–

82 (2013).

92. Gilljam, K. M., Müller, R., Liabakk, N. B. & Otterlei, M. Nucleotide excision repair is associated with the replisome and its efficiency depends on a direct interaction between XPA and PCNA. *PLoS One* **7,** e49199 (2012).

93. Wu, X., Shell, S. M., Yang, Z. & Zou, Y. Phosphorylation of nucleotide excision repair factor xeroderma pigmentosum group A by ataxia telangiectasia mutated and Rad3-related-dependent checkpoint pathway promotes cell survival in response to UV irradiation. *Cancer Res.* **66,** 2997–3005 (2006).

94. Shell, S. M. *et al.* Checkpoint kinase ATR promotes nucleotide excision repair of UV-induced DNA damage via physical interaction with xeroderma pigmentosum group A. *J. Biol. Chem.* **284,** 24213–24222 (2009).

95. Farley, S. M. *et al.* Menin localizes to chromatin through an ATR-CHK1 mediated pathway after UV-induced DNA damage. *J. Surg. Res.* **133,** 29–37 (2006).

96. Vodenicharov, M. D., Ghodgaonkar, M. M., Halappanavar, S. S., Shah, R. G. & Shah, G. M. Mechanism of early biphasic activation of poly(ADP-ribose) polymerase-1 in response to ultraviolet B radiation. *J. Cell Sci.* **118,** 589–99 (2005).

97. Robu, M. *et al.* Role of poly (ADP-ribose) polymerase-1 in the removal of UV-induced DNA lesions by nucleotide excision repair. 1–6 (2013). doi:10.1073/pnas.1209507110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1209507110

98. Pines, A. *et al.* PARP1 promotes nucleotide excision repair through DDB2 stabilization and recruitment of ALC1. *J. Cell Biol.* **199,** 235–49 (2012).

99. Luijsterburg, M. S. *et al.* DDB2 promotes chromatin decondensation at UV-induced DNA damage. *J. Cell Biol.* **197,** 267–81 (2012).

100. King, B. S., Cooper, K. L., Liu, K. J. & Hudson, L. G. Poly(ADP-ribose) contributes to an association between poly(ADP-ribose) polymerase-1 and xeroderma pigmentosum complementation group A in nucleotide excision repair. *J. Biol. Chem.* **287,** 39824–33 (2012).

101. Fischer, J. M. F. *et al.* Poly(ADP-ribose)-mediated interplay of XPA and PARP1 leads to reciprocal regulation of protein function. *FEBS J.* **281,** 3625–41 (2014).

102. Jungmichel, S. *et al.* Proteome-wide identification of poly(ADP-Ribosyl)ation targets in different genotoxic stress responses. *Mol. Cell* **52,** 272–85 (2013).

103. Nitta, M. *et al.* A novel cytoplasmic GTPase XAB1 interacts with DNA repair protein XPA. *Nucleic Acids Res.* **28,** 1–3 (2000).

104. Lembo, F. *et al.* MBD in , a Novel MBD2-Interacting Protein , Relieves

MBD2 Repression Potential and Reactivates Transcription from Methylated Promoters. *Mol. Cell. Biol.* **23,** 1656–1665 (2003).

105. Yonemasu, R. *et al.* Disruption of mouse XAB2 gene involved in pre-mRNA splicing, transcription and transcription-coupled DNA repair results in preimplantation lethality. *DNA Repair (Amst).* **4,** 479–91 (2005).

106. Nakatsu, Y. *et al.* XAB2, a novel tetratricopeptide repeat protein involved in transcription-coupled DNA repair and transcription. *J. Biol. Chem.* **275,** 34931–7 (2000).

107. Anttinen, A. *et al.* Neurological symptoms and natural course of xeroderma pigmentosum. *Brain* **131,** 1979–89 (2008).

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

109. Miyamoto, I., Miura, N., Niwa, H., Miyazaki, J. & Tanaka, K. Mutational analysis of the structure and function of the xeroderma pigmentosum group A complementing protein. Identification of essential domains for nuclear localization and DNA excision repair. *J. Biol. Chem.* **267,** 12182–7 (1992).

110. Morita, E. H. *et al.* Implications of the zinc-finger motif found in the DNA-binding domain of the human XPA protein. *Genes to Cells* **1,** 437–442

(1996).

111. Takahashi, Y. *et al.* XPA gene mutations resulting in subtle truncation of protein in xeroderma pigmentosum group A patients with mild skin symptoms. *J. Invest. Dermatol.* **130,** 2481–8 (2010).

112. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6,** (2013).

113. Rabi, I. I., Zacharias, J. R., Millman, S. & Kusch, P. A New Method of Measuring Nuclear Magnetic Moment. *Phys Rev* **53,** 318 (1938).

114. Taylor, D. G., Inamdar, R. & Bushel, M. NMR Imaging in Theory and in Practice. *Phys. Med. Biol.* **33,** 635–670 (1988).

115. Bloch, F., Hansen, W. W. & Packard, M. Nuclear Induction. *Phys Rev* **69,** 127 (1946).

116. Purcell, E., Torrey, H. & Pound, R. Resonance absorption by nuclear magnetic moments in a solid. *Phys Rev* **69,** 37–38 (1946).

117. Bloch, F., Hansen, W. W. & Packard, M. The Nuclear Induction Experiment. *Phys Rev* **70,** 474–485 (1946).

118. Li, M. X., Corson, D. C. & Sykes, B. D. Structure Determination by NMR: Isotope Labeling. *Methods Mol. Biol.* **173,** 255–265 (2002).

119. Bodenhausen, G. & Ruben, D. J. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **69,** 185–189 (1980).

120. Whitehead, B., Craven, C. J. & Waltho, J. P. Double and Triple Resonance NMR Methods for Protein Assignment. *Protein NMR Tech.* **60,** 29–52 (1997).

121. Leopold, M. F., Urbauer, J. & Wand, A. J. Resonance assignment strategies for the analysis of nmr spectra of proteins. *Mol. Biotechnol.* **2,** 61–93 (1994).

122. Arunkumar, A. I., Stauffer, M. E., Bochkareva, E., Bochkarev, A. & Chazin, W. J. Independent and coordinated functions of replication protein A tandem high affinity single-stranded DNA binding domains. *J. Biol. Chem.* **278,** 41077–82 (2003).

123. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci.* **94,** 12366–12371 (1997).

124. Chris Le, X., Wan, Q. H. & Lam, M. T. Fluorescence polarization detection for affinity capillary electrophoresis. *Electrophoresis* **23,** 903–908 (2002).

125. Hey, T., Lipps, G. & Krauss, G. Binding of XPA and RPA to damaged DNA investigated by fluorescence anisotropy. *Biochemistry* **40,** 2901–10 (2001).

126. Jantz, D. & Berg, J. M. Probing the DNA-binding affinity and specificity of designed zinc finger proteins. *Biophys. J.* **98,** 852–860 (2010).

127. Liu, Y. *et al.* Interactions of human replication protein A with single-stranded DNA adducts. *Biochem J* **385,** 519–526 (2005).

128. Jerabek-Willemsen, M. *et al.* MicroScale Thermophoresis: Interaction analysis and beyond. *J. Mol. Struct.* **1077,** 101–113 (2014).

129. Duhr, S. & Braun, D. Why molecules move along a temperature gradient. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 19678–82 (2006).

130. Beychok, S. Rotatory Dispersion and Circular Dichroism. *Annu. Rev. Biochem.* **37,** 437–462 (1968).

131. Beychok, S. Circular Dichroism of Biological Macromolecules. *Science (80-.).* **154,** 1288–1299 (1966).

132. Dodero, V. I., Quirolo, Z. B. & Sequeira, M. A. Biomolecular Studies by Circular Dichroism. *Front. Biosci.* **16,** 61–73 (2011).

133. Pei, J., Kim, B.-H. & Grishin, N. V. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36,** 2295–2300 (2008).

134. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42,** W320–W324 (2014).

135. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

136. Miki Tanioka, Arief Budiyant, Takahiro Ueda, Tohru Nagano, Masamitsu Ichihashi, Yoshiki Miyachi, C. N. A novel XPA gene mutation and its functional analysis in a Japanese patient with xeroderma pigmentosum group A. *J. Invest. Dermatol.* **125,** 244–246 (2005).

137. Lehmann, J. *et al.* A novel mutation in the XPA gene results in two truncated protein variants and leads to a severe XP/neurological symptoms phenotype. *J. Eur. Acad. Dermatol. Venereol.* 1–4 (2014). doi:10.1111/jdv.12841

138. Satokata, I., Tanaka, K. & Okada, Y. Molecular basis of group A xeroderma pigmentosum : a missense mutation and two deletions. *Hum. Genet.* **88,** 603–607 (1992).

139. Amr, K., Messaoud, O., El Darouti, M., Abdelhak, S. & El-Kamah, G. Mutational spectrum of Xeroderma pigmentosum group A in Egyptian patients. *Gene* **533,** 52–6 (2014).

140. Messaoud, O. *et al.* Severe phenotypes in two Tunisian families with novel XPA mutations: evidence for a correlation between mutation location and disease severity. *Arch. Dermatol. Res.* **304,** 171–6 (2012).

141. Ghafouri-Fard, S., Fardaei, M. & Miryounesi, M. A novel 5 nucleotide deletion in XPA gene is associated with severe neurological abnormalities. *Gene* 1–2 (2015). doi:10.1016/j.gene.2015.08.039

142. Satokata, I., Uchiyama, M. & Tanaka, K. Two novel splicing mutations in

the XPA gene in patients with group A xeroderma pigmentosum . *Hum. Mol. Genet.* **4,** 1993–1994 (1995).

143. States, J. C. & Myrand, S. P. Splice site mutations in a xeroderma pigmentosum group A patient with delayed onset of neurological disease. *Mutat. Res.* **363,** 171–177 (1996).

144. Sidwell, R. U. *et al.* A novel mutation in the XPA gene associated with unusually mild clinical features in a patient who developed a spindle cell melanoma. *Br. J. Dermatol.* **155,** 81–88 (2006).

145. Santiago, K. M., França de Nóbrega, A., Rocha, R. M., Rogatto, S. R. & Achatz, M. I. Xeroderma pigmentosum: low prevalence of germline XPA mutations in a Brazilian XP population. *Int. J. Mol. Sci.* **16,** 8988–96 (2015).

146. Maeda, T. *et al.* Compound heterozygous group A xeroderma pigmentosum patient with a novel mutation and an inherited reciprocal translocation. *Br. J. Dermatol.* **143,** 174–9 (2000).

147. Sun, Z. *et al.* Genotype-phenotype correlation of xeroderma pigmentosum in a Chinese Han population. *Br. J. Dermatol.* **172,** 1096–102 (2015).

148. Sato, M., Nishigori, C., Yagi, T. & Takebe, H. Aberrant splicing and truncated-protein expression due to a newly identified XPA gene mutation. *Mutat. Res.* **362,** 199–208 (1996).

149. Porter, P. C., Mellon, I. & States, J. C. XP-A cells complemented with

Arg228Gln and Val234Leu polymorphic XPA alleles repair BPDE-induced DNA damage better than cells complemented with the wild type allele. *DNA Repair (Amst).* **4,** 341–349 (2005).

150. Park, C.-J. & Choi, B.-S. The protein shuffle. Sequential interactions among components of the human nucleotide excision repair pathway. *FEBS J.* **273,** 1600–1608 (2006).

151. Missura, M. *et al.* Double-check probing of DNA bending and unwinding by XPA-RPA : an architectural function in DNA repair. *EMBO J.* **20,** 3554–3564 (2001).

152. Petruseva, I. O., Evdokimov, a N. & Lavrik, O. I. Molecular mechanism of global genome nucleotide excision repair. *Acta Naturae* **6,** 23–34 (2014).

153. Woźniak, K. & Błasiak, J. Recognition and repair of DNA-cisplatin adducts. *Acta Biochim. Pol.* **49,** 583–596 (2002).

154. Tracy M. Neher, Sarah C. Shuck, Jing-Yuan Liu, Jian-Ting Zhang, J. J. T. Identification of novel small molecule inhibitors of the XPA protein using in silico based screening. *ACS Chem. Biol.* **5,** 953–965 (2010).

155. Leman, J. K., Mueller, R., Karakas, M., Woetzel, N. & Meiler, J. Simultaneous prediction of protein secondary structure and transmembrane spans. *Proteins Struct. Funct. Bioinforma.* **81,** 1127–1140 (2013).

156. Buchan, D. W. A., Minneci, F., Nugent, T. C. O., Bryson, K. & Jones, D. T. Scalable web services for the PSIPRED Protein Analysis Workbench.

Nucleic Acids Res. **41,** 349–357 (2013).

157. Alekseev, S. & Coin, F. Orchestral maneuvers at the damaged sites in nucleotide excision repair. *Cell. Mol. Life Sci.* **72,** 2177–86 (2015).

158. Marteijn, J. a, Lans, H., Vermeulen, W. & Hoeijmakers, J. H. J. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat. Rev. Mol. Cell Biol.* **15,** 465–81 (2014).

159. Spievak, G. Nucleotide excision repair in humans. *DNA Repair (Amst).* **36,** 13–18 (2015).

160. Spivak, G. & Ganesan, A. K. The complex choreography of transcription-coupled repair. *DNA Repair (Amst).* **19,** 64–70 (2014).

161. Vermeulen, W. & Fousteri, M. Mammalian transcription-coupled excision repair. *Cold Spring Harb Perspect Biol* **5,** a012625 (2013).

162. Sugitani, N., Sivley, R. M., Perry, K. E., Capra, J. A. & Chazin, W. J. XPA: A key scaffold for human nucleotide excision repair. *DNA Repair (Amst).* **44,** 123–135 (2016).

163. Asahina, H. *et al.* The XPA protein is a zinc metalloprotein with an ability to recognize various kinds of DNA damage. *Mutat. Res.* **315,** 229–237 (1994).

164. Campagne, S., Gervais, V. & Milon, A. Nuclear magnetic resonance analysis of protein – DNA interactions. *J. R. Soc. Interface* **8,** 1065–78 (2011).

165. Seidel, S. A. I. *et al.* Microscale Thermophoresis Quantifies Biomolecular Interactions under Previously Challenging Conditions. *Methods* **59,** 301–315 (2013).

166. Breitsprecher, D. *et al.* Aptamer Binding Studies Using MicroScale Thermophoresis. *Methods Mol. Biol.* **1380,** 99–111 (2016).

167. Marti-Renom, M. A. *et al.* Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol Struct* **29,** 291–325 (2000).

168. Hyberts, S. G., Arthanari, H. & Wagner, G. in *Topics in Current Chemistry* **316,** 125–148 (Springer, 2012).

169. Hyberts, S. G., Robson, S. A. & Wagner, G. Exploring signal-to-noise ratio and sensitivity in non-uniformly sampled multi-dimensional NMR spectra. *J. Biomol. NMR* **55,** 167–178 (2013).

170. Hyberts, S. G., Arthanari, H., Robson, S. A. & Wagner, G. Perspectives in magnetic resonance: NMR in the post-FFT era. *J. Magn. Reson.* **241,** 60–73 (2014).

171. Hyberts, S. G., Frueh, D. P., Arthanari, H. & Wagner, G. FM reconstruction of non-uniformly sampled protein NMR data at higher dimensions and optimization by distillation. *J. Biomol. NMR* **45,** 283–294 (2009).

172. Hyberts, S. G., Takeuchi, K. & Wagner, G. Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and

sensitivity of protein NMR data. *J. Am. Chem. Soc.* **132,** 2145–2147 (2010).

173. Hyberts, S. G., Milbradt, A. G., Wagner, A. B., Arthanari, H. & Wagner, G. Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling. *J. Biomol. NMR* **52,** 315–327 (2012).

174. Buchko, G. W., Ni, S., Thrall, B. D. & Kennedy, M. a. Structural features of the minimal DNA binding domain (M98-F219) of human nucleotide excision repair protein XPA. *Nucleic Acids Res.* **26,** 2779–88 (1998).

175. Brosey, C. A. *et al.* A new structural framework for integrating replication protein A into DNA processing machinery. *Nucleic Acids Res.* **41,** 2313–27 (2013).

176. Miętus, M. *et al.* Crystal structure of the catalytic core of Rad2: insights into the mechanism of substrate binding. *Nucleic Acids Res.* **42,** 10762–75 (2014).

177. Kang, T.-H., Lindsey-Boltz, L. A., Reardon, J. T. & Sancar, A. Circadian control of XPA and excision repair of cisplatin-DNA damage by cryptochrome and HERC2 ubiquitin ligase. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 4890–5 (2010).

178. Nagel, Z. D. *et al.* Multiplexed DNA repair assays for multiple lesions and multiple doses via transcription inhibition and transcriptional mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E1823-32 (2014).

179. van Oers, M. M., Pijlman, G. P. & Vlak, J. M. Thirty years of baculovirus-insect cell protein expression: From dark horse to mainstream technology. *J. Gen. Virol.* 6–23 (2014). doi:10.1099/vir.0.067108-0

180. Sakuma, T., Barry, M. a & Ikeda, Y. Lentiviral vectors: basic to translational. *Biochem. J.* **443,** 603–18 (2012).

181. Giegé, R. A historical perspective on protein crystallization from 1840 to the present day. *FEBS J.* **280,** 6456–97 (2013).

182. Manjasetty, B. a, Turnbull, A. P., Panjikar, S., Büssow, K. & Chance, M. R. Automated technologies and novel techniques to accelerate protein crystallography for structural genomics. *Proteomics* **8,** 612–25 (2008).

183. Juranić, P. N. *et al.* High-precision x-ray FEL pulse arrival time measurements at SACLA by a THz streak camera with Xe clusters. *Opt. Express* **22,** 30004 (2014).