UNIVERSITY OF CALIFORNIA
RIVERSIDE

Molecular Mechanisms of DNA Opening by XPB Helicase and its Interaction With Nuclease in Nucleotide Excision Repair

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Biochemistry and Molecular Biology

by

Feng He

June 2021

Dissertation Committee:
Dr. Li Fan, Chairperson
Dr. Jikui Song
Dr. Xuan Liu

The Dissertation of Feng He is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

The text of this dissertation, in part or in full, is based upon work from two publications:

1. He, F., Duprez, K., Hilario, E., Chen, Z. and Fan, L. (2020). Structural basis of the XPB helicase-Bax1 nuclease complex interacting with the repair bubble DNA. Nucleic Acids Research 48(20): 11695–11705.

2. Duprez, K.*, He, F.*, Chen, Z., Hilario, E. and Fan, L. (2020). Structural basis of the XPB-Bax1 complex as a dynamic helicase-nuclease machinery for DNA repair. Nucleic Acids Research 48(11): 6326–6339. (*co-first author)

The corresponding author Dr. Li Fan, listed in these publications, directed and supervised the research which forms the basis for this dissertation. The funding for this work was provided by National Institutes of Health [R01GM108893], and UC Regents Faculty Development Award (to L.F.).

ABSTRACT OF THE DISSERTATION

Molecular Mechanisms of DNA Opening by XPB Helicase and its Interaction With Nuclease in Nucleotide Excision Repair

by

Feng He

Doctor of Philosophy, Graduate Program in Biochemistry and Molecular Biology
University of California, Riverside, June 2021
Dr. Li Fan, Chairperson

The superfamily II (SF2) DNA helicase XPB is an essential protein involved in two biologically important processes: transcription and DNA repair. In eukaryotes, XPB is the largest subunit of the TFIIH complex and is essentially required in both transcription initiation and the nucleotide excision repair (NER) pathway. Previous structural and biochemical data of TFIIH all have showed that eukaryotic XPB binds and translocates on the double-stranded (ds) DNA for DNA unwinding, strikingly different from other conventional DNA helicases that requires a single-stranded (ss) DNA extension for strand separation. XPB was proposed to work as a "molecular wrench" or dsDNA translocase and use its ATPase activity for achieving its DNA opening function in transcription. However, it still remains elusive how this unconventional enzyme achieves DNA opening in the context of NER. XPB also plays a key role in coordinating the DNA incision by nucleases like XPF or XPG, and the underlying mechanisms are unclear. In archaea, there is no TFIIH-like complex. XPB is in complex with a nuclease called Bax1 for DNA

unwinding and incision, but how this helicase-nuclease complex binds, opens and cleaves DNA is unknown. The objective of my research is to obtain structures and biochemical data of XPB with nuclease/DNA substrates for better understanding the DNA opening mechanisms of XPB and the crosstalk between XPB and nuclease for damage incision. A combination of biochemical, structural, and molecular biology techniques has been employed throughout to achieve these goals. Mutational and biochemical analyses based on crystal structures of archaeal XPB-Bax1 complexes from *Sulfurisphaera tokodaii* (St) and *Archaeogloubus fulgidus* (Af) demonstrates that this helicase-nuclease complex is a dynamic machinery and the activities of XPB are regulated by protein-protein interactions. Crystal structures of the StXPB-Bax1$^{\Delta C}$ with and without DNA substrate reveal how the XPB-Bax1 complex interacts with the repair bubble DNA, and adopts different conformations in DNA-free and DNA-bound states. Biochemical and mutational data confirmed that the conserved RED and ThM motifs of XPB play key roles in DNA binding and XPB activities. My investigation uncovered the unconventional helicase activity of archaeal XPB using its ThM motif to unzip DNA while translocating in 3'-5' direction on the duplex. Biochemical data together with structure comparison of the StXPB-Bax1$^{\Delta C}$-DNA ternary complex with the cryo-EM structure of human TFIIH-XPA suggest human XPB cooperates with XPA for the initial DNA opening around the lesion. In sum, my findings provide new insights into the molecular mechanisms of XPB-mediated DNA binding and opening during the repair bubble formation, and how XPB and nucleases may coordinate DNA incision in archaeal and eukaryotic NER.

TABLE OF CONTENTS

LIST OF FIGURES x
LIST OF TABLES xii

CHAPTER 1 1
Introduction 1
XPB is a 3'-5' SF2 DNA helicase 1
XPB is conserved from archaea to human 2
XPB opens DNA as a dsDNA translocase in transcription 5
XPB and XPD helicases drive DNA unwinding in NER 7
Archaeal XPB is in complex with Bax1 nuclease 13

CHAPTER 2 15
XPB-Bax1 is a dynamic helicase–nuclease complex 15
ABSTRACT 15
INTRODUCTION 16
MATERIALS AND METHODS 18
Site-directed mutagenesis of StXPB, StBax1 and AfBax1 constructs 18
Expression and purification of XPB and Bax1 WT and mutants 18
Isothermal titration calorimetry (ITC) analysis 19
Nuclease activity assay 20
RESULTS 21
The XPB-Bax1 complex is a dynamic machine 21
The interaction interface between XPB and Bax1 25
AfBax1 contains two distinguished nuclease active sites 28
DISCUSSION 33

CHAPTER 3 37
Structural insights into DNA binding and opening by the XPB-Bax1 complex and implications for nucleotide excision repair 37
ABSTRACT 37
INTRODUCTION 38
MATERIALS AND METHODS 40
Cloning, expression and purification of StXPB-Bax1 and StXPB-Bax1$^{\Delta C}$ 40
Crystallization and structure determination 41
Cloning, expression and purification of human XPB/p52/p8 trimer 42
Cloning, expression and purification of human XPA 43
Electrophoretic mobility shift assay 43
ATPase activity assay 44
RESULTS 45
Optimizing XPB-Bax1 construct for co-crystallization with DNA 45
Overall structure of the XPB-Bax1-DNA ternary complex 49
Interactions between the XPB-Bax1$^{\Delta C}$ heterodimer and the forked DNA 54
Human XPB has a shortened ThM motif and forms a stable complex with p52/p8 60
The ThM motif of StXPB is most essential for DNA junction binding 64
StXPB has enhanced affinity for dsDNA with a small mismatched bubble 67
Archaeal XPB could play a role in damage recognition 70
StXPB in the DNA-free StXPB-Bax1$^{\Delta C}$ structure is in the ATP-bound state 72
XPB conformational changes induced by DNA binding 75
The RED and ThM motifs are critical for the ATPase activities of StXPB 77
Proposed model of ATP-driven DNA opening by StXPB 79
Comparison to the cryo-EM structure of human TFIIH core complexed with XPA and a forked DNA 81
DISCUSSION 84
REFERENCES 90

APPENDIX A 102
Mapping human XPB-p52 interacting interface and two human disease mutations 102
INTRODUCTION 102
MATERIALS AND METHODS 104
Cloning, expression and purification of human p52/p8 104
Co-expression and purification of human XPB-NTD/p52/p8 104
Cloning, expression and purification of human XPB/p52/p8 WT, F99S and T119P 105
Electrophoretic mobility shift assay 106
RESULTS 107
Human XPB-NTD could be co-expressed with p52/p8 107
Mapping of the minimal p52-interacting domain in human XPB 110
Crystallization trials of human XPB-NTD/p52/p8 113
Biochemical impact of XPB disease –causing mutations F99S and T119P 116
DISCUSSION 118
REFERENCES 121

LIST OF FIGURES

Chapter 1.

Figure 1.1 Helicase motifs from SF1 and SF2 helicases. 2
Figure 1.2 XPB proteins are highly conserved. 4
Figure 1.3 The subunits of TFIIH. 4
Figure 1.4 XPB in TFIIH functions as a dsDNA translocase. 6
Figure 1.5 Some examples of DNA lesions repaired by NER. 10
Figure 1.6 The schematic diagram of the nucleotide excision repair pathway. 11
Figure 1.7 Structures containing XPD in complex with ssDNA. 12
Figure 1.8 Distribution of eukaryotic NER genes in archaea. 14

Chapter 2.

Figure 2.1 The XPB–Bax1 complex is a dynamic complex. 22
Figure 2.2 Sequence and structural alignments of Bax1 orthologues. 24
Figure 2.3 Hydrophobic and charge/polar interactions contribute to the XPB-Bax1 assembly. 26
Figure 2.4 The XPB-Bax1 interaction interfaces are conserved among archaea. 27
Figure 2.5 The two nuclease sites in AfBax1. 31
Figure 2.6 Mutational and biochemical analysis of the two nuclease sites in AfBax1 by ITC and the nuclease activity assay. 32
Figure 2.7 Protein-protein interactions regulate the nuclease activity of Bax1. 36

Chapter 3.

Figure 3.1 StXPB-Bax1 proteins and their interactions with DNA. 47
Figure 3.2 Crystallization of StXPB-Bax1 with DNA substrates. 48
Figure 3.3 Two StXPB-Bax1$^{\Delta C}$-DNA complexes in one asymmetric unit. 50
Figure 3.4 Structure of the StXPB-Bax1$^{\Delta C}$-forked DNA complex. 53
Figure 3.5 Interactions of StXPB-Bax1 with the forked DNA substrate. 56
Figure 3.6 Electron density (Fo-Fc) map of the key residues in Figure 3.5. 57

Figure 3.7 Sequence and structural alignments of XPB orthologues in archaea. 59
Figure 3.8 Human XPB superimposed onto StXPB-forked DNA and purification of XPB/p52/p8 from insect cell. 61
Figure 3.9 Sequence and structural alignment of StXPB and human XPB. 63
Figure 3.10 The RED and ThM motifs of XPB are important for forked DNA binding and comparison to human XPB-XPA. 66
Figure 3.11 The RED and ThM motif are important for XPB binding to dsDNA with mismatched bubbles. 69
Figure 3.12 Model of StXPB binding to the damage site. 71
Figure 3.13 Crystallization and structure of the DNA-free StXPB-Bax1$^{\Delta C}$ dimer. 73
Figure 3.14 DNA interactions induce conformational changes in StXPB. 76
Figure 3.15 A schematic diagram of DNA opening mechanism of StXPB. 80
Figure 3.16 Structure comparison of the StXPB-Bax1$^{\Delta C}$-DNA ternary complex and the cryo-EM structure of the TFIIH-XPA-DNA complex 83
Figure 3.17 Model of DNA incisions by the XPB-nuclease complex at a DNA bubble. 89

Appendix A.

Figure A.1 XPB domain architecture and purification of the XPB/p52/p8 complex. ... 109
Figure A.2 Purification of the XPB-NTD/p52/p8 complexes 112
Figure A.3 Crystallization of the XPB-NTD/p52/p8 complexes. 114
Figure A.4 T119P disrupts interactions of XPB-NTD with p52/p8 while F99S does not. 117
Figure A.5 DNA binding of XPB/p52/p8 complexes analyzed by electrophoretic mobility shift assay (EMSA). 119

LIST OF TABLES

Chapter 3.

Table 1 Statistics for X-ray diffraction data collection and structural refinement of the StXPB-Bax1$^{\Delta C}$-DNA ternary structure 51

Table 2 Statistics for X-ray diffraction data collection and structural refinement of the DNA-free StXPB-Bax1$^{\Delta C}$ structure 74

Table 3 ATPase activities of StXPB variants in the presence and absence of Bax1 and DNA substrates 78

Appendix A.

Table A.1 Summary of human XPB constructs used in co-expression with p52/p8 111

Table A.2 Surface-engineered XPB and p52 constructs for XPB (30-160)/p52/p8 co-expression and crystallization 115

CHAPTER 1

Introduction

XPB is a 3'-5' SF2 DNA helicase

Helicases are ATP-dependent enzymes that translocate along nucleic acids (DNA or RNA) to catalyze strand separation. Based on different conserved motifs, helicases could be divided into six superfamilies SF1-6 [1]. Members of SF and SF2 helicases are monomeric while SF3-6 helicases are hexameric [2]. SF1 and SF2 are the largest two groups that both contain seven sequence motifs (I, Ia, II–VI) [3]. Within these seven motifs, motifs I and II (also called Walker A and B motifs) are the most conserved, which bind and hydrolyze ATP (Figure 1.1). Each monomeric SF1 or SF2 helicase contains two core RecA-like domains [4], and additional domains are present to regulate helicase activities [1, 5]. Therefore, based on the seven conserved motifs in XPB helicase, it belongs to superfamily 2 (SF2).

XPB/Ssl2 (also known as ERCC3 and RAD25) was firstly identified to be responsible for nucleotide excision repair (NER) defects in XP-B patients [6]. Then XPB has been shown to a subunit of the TFIIH complex necessary for transcription and DNA repair and have 3'-5' ATP-dependent DNA helicase activity [7-10].

SF1 helicases	motif	SF2 helicases
(G/A)xxGxGKT	I	GxGKT
(T/S)xxA	Ia	PxxxL
DExx	II	DEx(H/D) (ARL follows in some SF2s)
GDxxQLxx (ARL follows in some SF1s)	III	(S/T)(A/G)T
no consensus	IV	no consensus
(T/S)Y(H/D)	V	(L/V)xTxx(G/A)
VAx(T/S)RA	VI	(Q/H)(R/A)x(G/D)R(A/T)(G/H)R

Figure 1.1 Helicase motifs from SF1 and SF2 helicases.
Conserved helicase motifs I, Ia, and II–VI as are listed using the single letter amino acid code. This figure is adapted from [3].

XPB is conserved from archaea to human

XPB proteins are well conserved from archaea to human [11-15]. The first crystal structure of an A. fulgidus XPB homolog (AfXPB) was determined by Fan *et al.*[15]. The AfXPB structure contains two core RecA-like helicase domains, HD1 and HD2, and these domains bind and hydrolyze ATP to induce conformational changes of the protein for its translocation on DNA substrates. Also this structure uncovered a DNA damage recognition domain (DRD), and unique RED and ThM motifs for XPB (Figure 1.2). All these domains and motifs are highly conserved and they are also found in the corresponding human protein (Figure 1.2A). Human XPB is a 782-residue protein and contains two additional domains flanking the central helicase core: the N terminal domain

(NTD) and the C-terminal domain (CTD). These two domains are primarily involved in protein-protein interactions in the TFIIH complex.

In eukaryotes, XPB is an essential subunit of the basal transcription factor complex TFIIH, which is essentially required for both transcription initiation and nucleotide excision repair (NER) [7, 16-18]. Due to its important roles, mutations in XPB cause xeroderma pigmentosum (XP), trichothiodystrophy (TTD) and Cockayne syndrome (CS) symptoms, sensitivity to UV light including sunlight, developmental disorders or increased frequency of skin cancer [19-21].

TFIIH is a 500 kDa ten-subunit complex containing the core complex (XPB, XPD, p62, p52, p44, p34 and p8) and a kinase module (CDK7, Cyclin H and MAT1) (Figure 1.3) [22]. XPD is also a SF2 DNA helicase in TFIIH with 5'-3' helicase activity [1, 23, 24], which is opposite to XPB helicase. XPB has close interactions with other subunits in TFIIH such as p52 and p8. p52 has been reported to interact with the N terminal of XPB and regulate the activities of XPB [25-28] while p8 subunit makes contacts with the C terminal of human XPB [18]. In addition, p52 was shown to have strong interactions with p8 [18, 26, 29, 30]. Therefore, p52 and p8 should be the most two important regulators of XPB function in TFIIH.

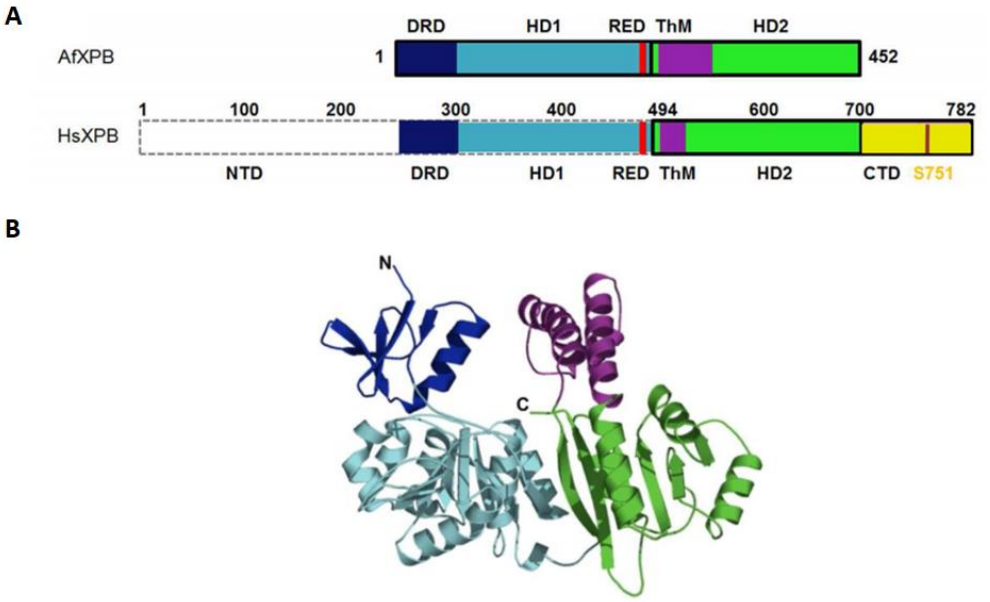

Figure 1.2 XPB proteins are highly conserved.

(A) Domain organizations of AfXPB and human XPB. Two core RecA-like helicase domains (HD1 and HD2) are highly conserved. Human XPB contains additional N-terminal domain and C-terminal domain, which are implicated in protein-protein interactions in TFIIH.

(B) The crystal structure of AfXPB (PDB ID: 2FWR) is shown in cartoon.

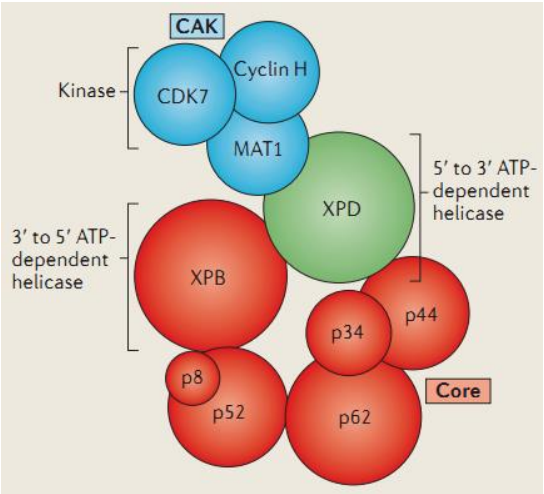

Figure 1.3 The subunits of TFIIH.

XPB and XPD have opposite polarity. This figure is adapted from [22].

XPB opens DNA as a dsDNA translocase in transcription

XPB/Ssl2 is required for promoter opening [31] and escape during transcription [32, 33]. Crosslinking data and previous cryo-EM studies have demonstrated that XPB engages with promoter DNA downstream of the putative transcription start site [34-37], as shown in the cryo-EM structure of yeast pre-initiation complex (PIC) (Figure 1.4A). XPD/Rad3 makes no contact with the DNA duplex (Figure 1.4A), consistent with its dispensable role in transcription [38, 39]. Traditional helicases unwind DNA by loading to the ssDNA overhang of dsDNA and then translocating on this strand with ATP binding and hydrolysis to catalyze strand separation of dsDNA [1-3]. XPB does not directly bind the DNA opening region and translocate on dsDNA instead of ssDNA, so it is believed that XPB is an unconventional DNA helicase [40]. To explain how XPB melts promoter DNA, a "molecular wrench" model was initially proposed [41]. Later more biochemical and structural evidences suggest a translocase model for ATP-dependent DNA opening by XPB [36, 42, 43]. According to this model (Figure 1.4B), XPB/Ssl2 tries to translocate on the downstream DNA away from Pol II with its ATP activity. Because the upstream promoter DNA is fixed by other transcription cofactors like TFIIA and TFIIE, and the location of TFIIH is also physically fixed due to strong protein-protein interactions including TFIIE, the action of XPB would instead result in a reeling of DNA into the active site of RNA polymerase (Pol) II. This model could nicely explain how XPB achieves its function in transcription and further indicates that XPB is not a canonical DNA helicase.

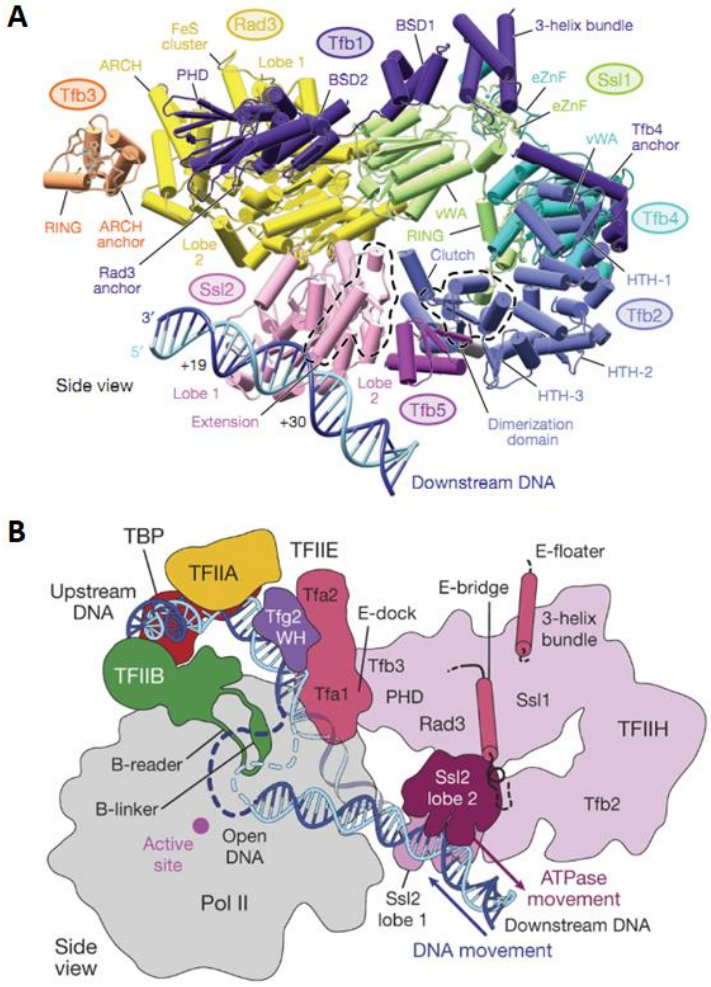

Figure 1.4 XPB in TFIIH functions as a dsDNA translocase.

(A) The side view of TFIIH structure in cylindrical representation from the cryo-EM structure of yeast PIC (PDB ID: 5OQJ). Ssl2 (XPB) is colored in pink and Rad3 (XPD) is colored in yellow.

(B) Side view of the cross-section of the PIC with open and closed DNA. The HD1 of Ssl2 is colored in pink and HD2 is colored in burgundy. The arrows indicate the direction of the Ssl2 ATPase translocation and DNA movement during promoter melting. This figure is adapted from [36].

XPB and XPD helicases drive DNA unwinding in NER

The genome DNA is highly susceptible to different kinds of damaging agents. Among all the DNA repair pathways, the NER pathway is a highly diverse one that removes various bulky and helix-distortion lesions such like cyclobutane pyrimidine dimer (CPD), 6-4 pyrimidine-pyrimidone photoproducts (6-4PP) and cisplatin adducts caused by UV radiation (Figure 1.5) including sunlight [16, 44, 45]. Eukaryotic NER consists of two subpathways (Figure 1.6): global genome NER (GG-NER) that repairs DNA lesions throughout the entire genome, and transcription-coupled NER (TC-NER) that specifically targets DNA lesions blocking the RNA polymerases II during transcription. TC-NER and GG-NER only differ in the first damage recognition step. In TC-NER, damage recognition is triggered by stalled RNA polymerases II [46-48]. The arrested RNAPII firstly recruits CSB (Cockayne syndrome protein B) and Cockayne syndrome proteins CSA (Cockayne syndrome WD repeat protein A) [49], then other TCR factors such as XAB2, UVSSA, USP7, plus some histone-remodeling factors are engaged [50, 51]. In GG-NER, a lesion that causes a significant DNA helix distortion is directly recognized by the XPC/hHR23B complex. But minor distortions like CPD lesions could not be detected by the XPC/hHR23B complex and are recognized by other proteins including DDB2/XPE and DDB1, which would facilitate recruiting the XPC/hHR23B complex (Figure 1.6) [50, 52, 53].

Upon the initial recognition of the DNA lesions, TFIIH is directly recruited by XPC/hHR23B [54-59] and both subpathways share identical downstream mechanisms.

XPB and XPD helicases in TFIIH unwind DNA around the lesion and extend the DNA repair bubble. It is believed that XPB initiates the DNA unwinding around the lesion site to enable the engagement of XPD onto a ssDNA extension, and XPD further translocates along the damaged strand when it gets stalled at the lesion site that impedes its motion [40, 60]. The supporting evidences for this hypothesis include that the helicase activity of yeast XPD homolog Rad3 is inhibited by bulky DNA lesions [61], and the crystal structure of archaeal XPD/ssDNA and a recent cryo-EM structure of TFIIH-XPA-DNA clearly reveal that the domains of XPD form a cleft or DNA binding groove that only allows binding and the passing through of normal ssDNA nucleotides (Figure 1.7) [62, 63]. Therefore, XPD is a conventional DNA helicase that loads and translocates on an overhang of the DNA substrate to unwind the DNA through the "inch worm" mechanism [64], which is in line with the idea that the initial opening of the damaged DNA by TFIIH has to be achieved by the unconventional XPB helicase.

During the damage verification process, TFIIH would generate a ~25-30-nucleotide repair bubble and pinpoint the lesion site to get ready for subsequent DNA incisions. In the meanwhile, other NER factors such as XPA and replication protein A (RPA) are also involved to stabilize the opened state of DNA and facilitate the assembly of a pre-incision complex [56, 65, 66]. XPA is a critical component of the NER pathway as it interacts with XPC-RAD23B, DDB2, TFIIH, RPA, ERCC1-XPF and PCNA [67-74]. These interactions of XPA suggest its key role in ensuring all the NER factors to be in the right place during NER. RPA binds the non-damaged ssDNA strand opposite the lesion, protecting it from degradation and helping to coordinate excision and

repair events [75, 76]. The endonuclease XPG could also stably interact with and stabilize TFIIH [77], while its endonuclease activity is inhibited before the upcoming incision events.

Once the DNA repair bubble is created and the lesion site is precisely verified, two endonucleases including XPG and the ERCC1/XPF complex were engaged. ERCC1-XPF is recruited to the NER machinery through the interaction with XPA [72, 73], which triggers the dual excision reaction. ERCC1-XPF makes the first incision at the 5' side of the lesion, and then XPG makes the second cut at the 3' side [16, 78-81]. Following these dual excision events, the lesion-containing ssDNA fragment is released with TFIIH bound to it [82]. This observation further indicates that XPD is responsible for lesion verification. Finally, the resulting gap in the damaged DNA strand is filled by DNA synthesis and ligation [83-86], and this genome region containing the DNA damage would be restored into the original DNA sequence.

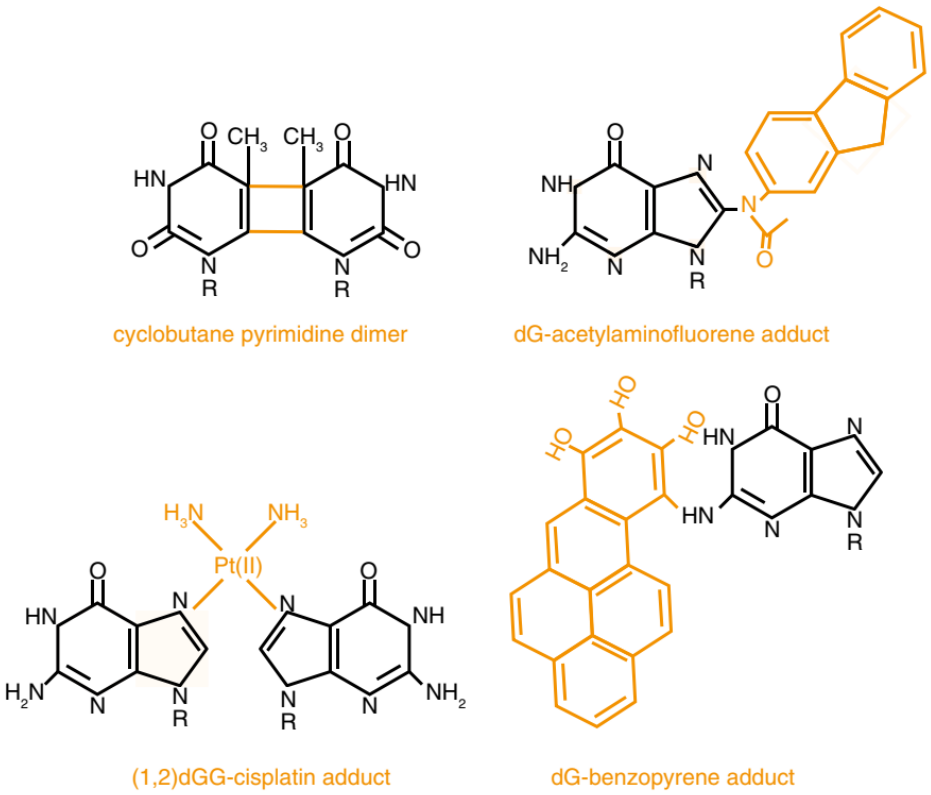

Figure 1.5 Some examples of DNA lesions repaired by NER.

Structures of some NER DNA lesions including Cyclobutane pyrimidine dimer (CPD), dG-acetylaminofluorene, (1,2) dGG-cisplatin and dG-benzopyrene adducts are shown. This figure is adapted from [44].

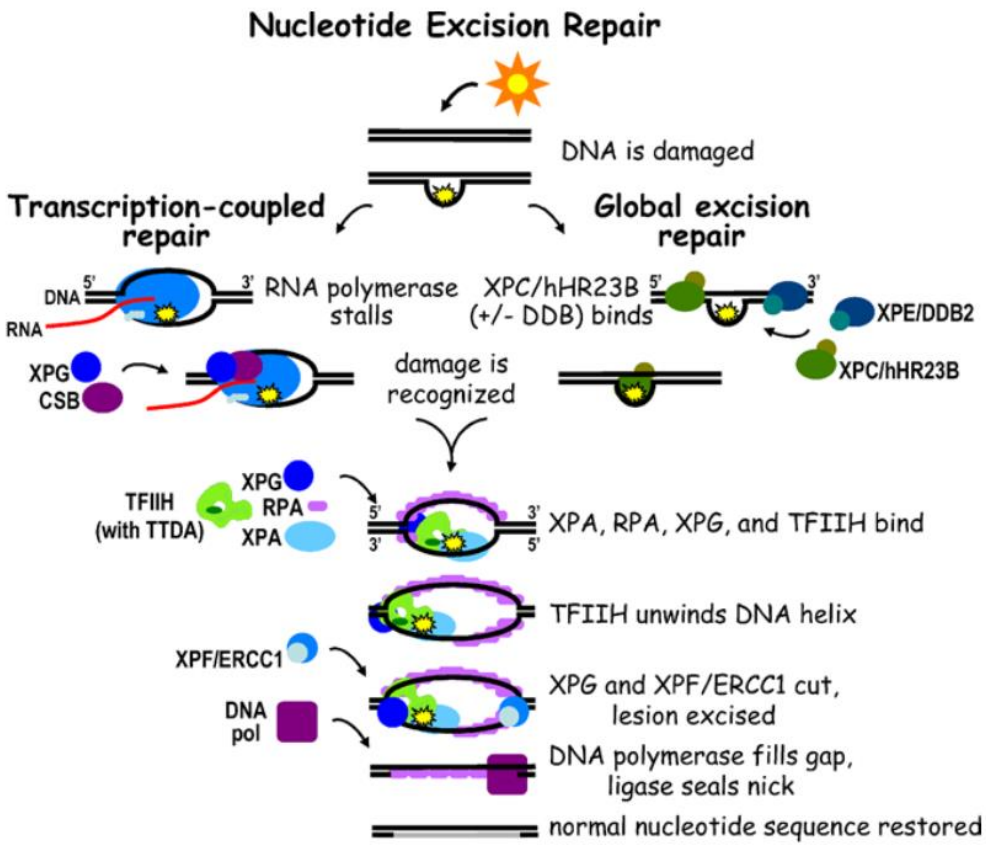

Figure 1.6 The schematic diagram of the nucleotide excision repair pathway.
GG-NER and TCR only differ in the initial step regarding how the damage is recognized. The CSB protein has been involved in initiation of TCR through recognizing stalled RNAP. In GG-NER, some less distorting lesions require initial recognition by the XPE/DDB2 complex. Two subpathways converge when TFIIH is recruited to the damaged DNA. TTDA refers to p8. This figure is adapted from [53].

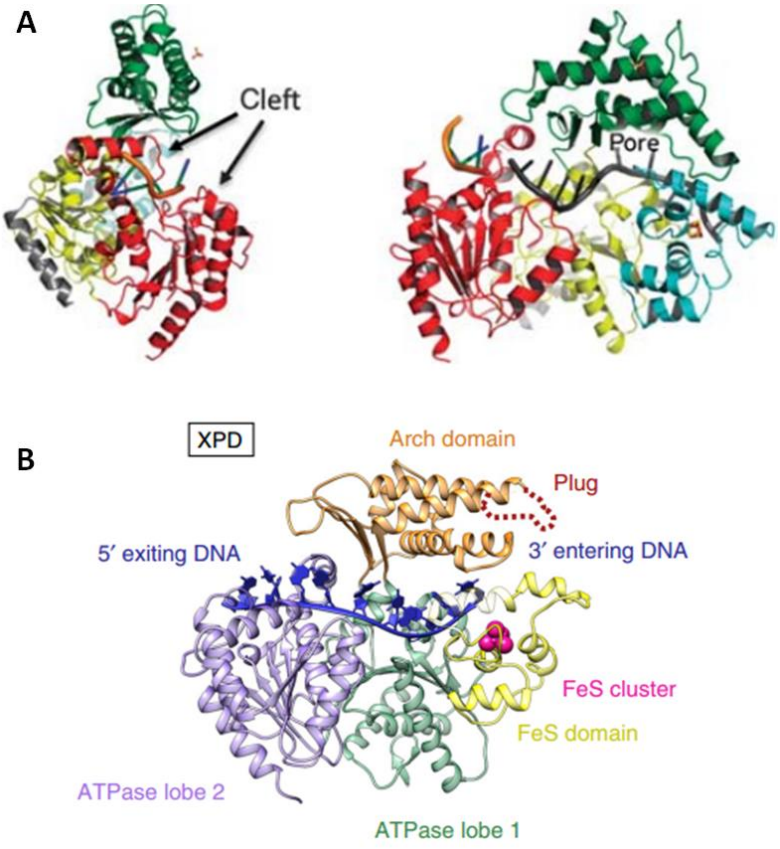

Figure 1.7 Structures containing XPD in complex with ssDNA.

(A) Side view of the *Thermoplasma acidophilum* (Ta) XPD–ssDNA structure. The two RecA-like domains (HD1 and HD2) are in yellow and red, the FeS cluster domain is in cyan, and the arch domain is in green. The DNA identified in the electron density is shown in orange. Left: The cleft where the DNA is bound is indicated with arrows. Right: Combination of experimental DNA (orange) with modeled DNA (grey).

(B) Human XPD is in complex with ssDNA from the cryo-EM structure of TFIIH-XPA-DNA. XPD HD1, FeS, Arch, and HD2 domains are in green, yellow, orange, and medium purple, respectively. DNA is shown in dark blue. This figure is adapted from [62, 63].

Archaeal XPB is in complex with Bax1 nuclease

As an essential subunit of TFIIH, XPB helicase is highly conserved in eukaryotes. Nevertheless, there is no TFIIH-like complex in archaea. Homologs of eukaryotic NER proteins XPB and XPD helicases, RPA (called SSB in archaea), and XPF nucleases (including a short and a long version) are present in most archaea [87] for DNA repair (Figure 1.8). But there are no recognizable homologs of the damage recognition proteins like XPC and XPA in archaea. The archaeal SSB protein (known as RPA in eukaryotes) may play a role in detecting DNA damage, because *Sulfolobus solfataricus* SSB was shown to specifically melt damaged DNA [88].

Even though most archaeal genomes encode homologs of several eukaryotic NER proteins, how archaeal NER works remains very unclear. Expression levels of XPB, XPF homologs in *Sulfolobus solfataricus* were upregulated in response to UV treatment, suggesting the presence of an XP-protein based NER pathway in this archaea [89, 90]. Whereas, deletion of the XPD and XPB genes in *Thermococcus kodakaraensis* did not obviously increase the sensitivity to UV irradiation [91].

Most crenarchaea have two copies of XPB: XPB1 and XPB2 [92]. The XPB1 protein has an N-terminal extension compared to XPB2. The archaeal *XPB* genes (*XPB* gene in euryarchaea or the *XPBII* gene in crenarchaea) are usually found in close proximity to the *Bax1* (Binds archaeal XPB) gene [92]. Bax1 belongs to DUF790 endonuclease-like family based on bioinformatic analyses and sequence conservation [93]. XPB and Bax1 were shown to form a stable 1:1 heterodimer from different archaeal

species in vitro [92, 94-96]. Bax1 is a structure-specific endonuclease and its nuclease activities are regulated by XPB helicase in vitro [94, 95]. On the other hand, Bax1 could stimulate the ATPase activity of XPB while have no effect on its helicase activity [97]. In addition, biochemical data also suggests that XPB-Bax1 complex could function as a helicase-nuclease machine in an ATP-dependent manner for DNA unwinding and incision [96]. Because Bax1 was shown to have a preference for cleavage on 3'-sides of NER substrates [96], Bax1 is likely to be a functional equivalent of eukaryotic XPG in archaea.

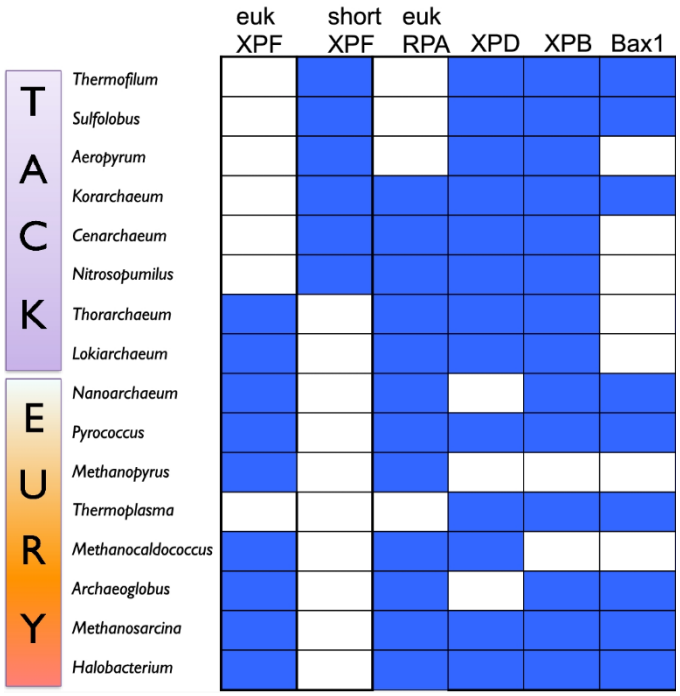

Figure 1.8 Distribution of eukaryotic NER genes in archaea.

Genus names on the left are organised as members of the TACK superphylum and Euryarchaea, which is reviewed in [98]. This figure is adapted from [87].

CHAPTER 2

XPB-Bax1 is a dynamic helicase–nuclease complex

ABSTRACT

In archaea, XPB forms a stable complex with Bax1 nuclease, to open the damaged DNA and mediate damage incision. Bax1 binding has dramatic impact on the activities of XPB and XPB was also reported to regulate the nuclease activities of Bax1, but the underlying mechanisms remain unclear. Based on our crystal structures of XPB in complex with Bax1 from two archaeal species *Archaeoglobus fulgidus* (Af) and *Sulfolobus tokodaii* (St), we further analyzed the interacting interface between XPB and Bax1, and confirmed important residues for XPB-Bax1 interactions by mutagenesis and ITC assays. Nuclease activity assays on nuclease-dead mutants demonstrated that Bax1 is likely to contain two distinguished nuclease active sites to incise DNA damage. Our structural analyses combined with biochemical results for the first time demonstrated how the archaeal XPB–Bax1 complex is built to be a dynamic machine at the molecular level.

INTRODUCTION

XPB helicase is essentially required for DNA opening and coordinating damage incision by nucleases in eukaryotic nucleotide excision repair. Homologs of essential eukaryotic NER proteins including XPB, XPD, RPA and XPF could be found in most archaea, suggesting the presence of the archaeal NER pathway.

Previous investigations on these archaeal proteins have advanced our understanding of key NER steps [87]. Archaeal XPD is a 5'-3' DNA helicase and contains an essential iron-sulfur (FeS) cluster [99]. Crystal structures of apo and ssDNA-bound archaeal XPD revealed how XPD binds DNA to verify DNA damage, and provided explanations for the disease-causing XPD mutations [23, 24, 62, 100]. The crystal structure of AfXPB [15] for the first time revealed the architecture of the two core RecA-like helicase domains, HD1 and HD2, and also uncovered the unique RED and ThM motifs for XPB (Figure 1.2). Since archaeal XPB retains key structural elements of eukaryotic XPB, it could also play a critical role in archaeal NER.

Instead of being involved in a TFIIH-like complex, archaeal XPB was only reported to have close interactions with a nuclease called Bax1. Therefore, this XPB-Bax1 complex is likely to be a helicase-nuclease machine for DNA repair in archaea. The structure-specific nuclease activities of Bax1 were confirmed in different archaeal species including *Thermoplasma acidophilum* (Ta) and *Sulfolobus solfataricus* (Ss). Interestingly, TaBax1 only cleaves DNA substrates on the 3'-overhang while SsBax1 has no such activity by itself. And TaXPB-Bax1 cleaves both 3' and 5'-overhangs

at the junction. In contrast, SsXPB-Bax1 is specific for DNA substrates containing 5'-overhangs [92-97]. Also, the nuclease activity assays with mutations of Bax1 suggest TaBax1 contains an N-terminal nuclease site whereas the active site of SsBax1 is at its C-terminal domain [94-96].

Here we carried out structural analyses according to the recent two crystal structures of AfXPB-Bax1 and StXPB-Bax1 determined by our group, which reveals XPB-Bax1 complex is a structurally dynamic helicase-nuclease machine. Combined with the method of site-directed mutagenesis, we utilized isothermal titration calorimetrythe (ITC) to identify the crucial residues involved in the XPB-Bax1 interactions and characterized the impact of XPB-Bax1 interactions on Bax1 function by nuclease activity assay. We propose that archaeal Bax1 contains two distinguished nuclease sites, which would exhibit different activities on DNA substrates in the absence or presence of XPB. These findings resolved the previously considered confliction of the results on Bax1 nuclease activities.

MATERIALS AND METHODS

Site-directed mutagenesis of StXPB, StBax1 and AfBax1 constructs

XPB-Bax1 interaction interface mutants including StXPB E357A/E360A, StBax1 R86A/R87A, StBax1 L89A/F90A/P94A/V95A, and nuclease-site mutants including AfBax1 D133A/E135A, AfBax1 D305A, AfBax1 D133A/E135A/D305A were all generated by the two step overlap extension PCR [101]. The mutation sites were introduced into the designed primer. The DNA encoding StXPB mutant E357A/E360A and StBax1 mutant L89A/F90A/P94A/V95A were cloned into pET-15b vector with an N-terminal His_6-tag, and the DNA encoding other mutants were cloned into pET-28a vector with an N-terminal His_6-tag. All the DNA sequences are verified by the Sanger Sequencing Services at UCR IIGB Core Facilities.

Expression and purification of XPB and Bax1 WT and mutants

StXPB or Bax1 was expressed in *E. coli* Rosetta (DE3) pLysS cells (Invitrogen). After induction for 18 hours with 0.2 mM IPTG at 28 °C, the cells were harvested by centrifugation and the pellets were resuspended in lysis buffer containing 50 mM Tris-Cl pH 7.5, 500 mM NaCl, 10% glycerol. The cells were then lysed by sonication and the cell debris was removed by centrifugation. The supernatant was purified by affinity chromatography using Ni-NTA resin (Thermo Scientific). For StXPB or StBax1, protein was loaded in low-salt buffer (50 mM MES pH 6.0, 100 mM NaCl, 5% glycerol) and eluted in high-salt buffer (50 mM MES pH 6.0, 1000 mM NaCl, 5% glycerol) by HiTrap

SP FF ion-exchange chromatography (GE). The purification was completed by gel-filtration chromatography (Superdex 200, 16/60, GE) in 25 mM Tris-Cl pH 7.5, 200 mM NaCl, 5% Glycerol. For AfXPB or AfBax1, protein was loaded in low-salt buffer (50 mM Tris-Cl pH 7.0, 100 mM NaCl, 5% glycerol) and eluted in high-salt buffer (50 mM Tris-Cl pH 7.0, 1000 mM NaCl, 5% glycerol) by HiTrap SP FF ion-exchange chromatography. The purification was completed by gel-filtration chromatography in 25 mM Tris-Cl pH 7.5, 500 mM NaCl, 5% Glycerol. All the purified protein samples were concentrated and stored at -80 ℃.

Isothermal titration calorimetry (ITC) analysis

ITC measurements were performed using a MicroCal iTC200 instrument (GE Healthcare). Titrations were performed in Af-protein sample buffer (10 mM Tris-Cl pH 7.5, 500 mM NaCl, 5% (v/v) glycerol) and St-protein sample buffer (10 mM Tris-Cl pH 7.5, 200 mM NaCl, 5% (*v/v*) glycerol). Titrations of XPB in the syringe over Bax1 were carried out at 40 ℃ to inhibit precipitation. The time between each injection was set at 180 s. All data were processed with the MicroCal Origin software and fitted with the single-site binding mode. Two independent experiments were performed for every interaction described here. The values for the stoichiometry of binding (N), enthalpy change (ΔH), and binding constant (K) were determined via least squares analysis performed by the ORIGIN software package provided by the calorimeter manufacturer (GE) following the procedure provided by the manufacturer. Equilibrium dissociation constant K_d was calculated based on K value.

Nuclease activity assay

DNA oligonucleotides were ordered from IDT and purifed from urea denaturing PAGE gel. Then ssDNA was 5' end labeled for 30 min at 37°C in a 20 μL reaction containing 250 nM ssDNA, [γ-^{32}P] ATP, 1 × T4 polynucleotide kinase reaction buffer, and 25 units of T4 polynucleotide kinase (Promega). Labeled ssDNA was subsequently annealed with the complementary ssDNA in a ratio of 1:1.2 in a buffer containing 20 mM Tris-Cl pH 7.5, 50 mM NaCl, 1 mM EDTA. B50 and TAG31 were labeled. 50–16 bubbled DNA were annealed from labeled B50 and B50bub16. DNA oligo sequences are listed below:

B50: 5'-CCT CGA GGG ATC CGT CCT AGC AAG CCG CTG CTA CCG GAA GCT TCT GGA CC-3'

B50bub16: 5'-GGT CCA GAA GCT TCC GGA TAG TTA CCG CAC GAT GGA CGG ATC CCT CGA GG -3'

Reactions were incubated for 1h at 48°C on a heat block in a total volume of 7 μL containing 20 mM Tris-Cl pH 8.0, 40 mM NaCl, 10% glycerol, 0.1 mg/ml BSA, 10 mM $MgCl_2$. Each reaction contained ~100 fmol labeled DNA substrate and ~20000 fmol proteins or as indicated in the text. 3 μL stop buffer containing 90% formamide and 10 mM EDTA was added and reaction mixtures were boiled for 10 min prior to electrophoresis at 1500V in a 18% urea polyacrylamide gel. Gels were exposed to a phosphorimaging screen overnight, visualized by a GE Typhoon 9410 Molecular Imager and edited by the Image Lab software.

RESULTS

The XPB-Bax1 complex is a dynamic complex

The crystal structures of AfXPB-Bax1 and StXPB-Bax1 were previously determined by our group at 3.0 Å resolution (PDB ID: 6P66) and 3.15 Å resolution (PDB ID: 6P4O), respectively. Interestingly, the two Bax1 proteins retain similar conformations while AfXPB and StXPB adopt dintinct conformations (Figure 2.1A-B). The two structures clearly show that Bax1 N-terminal domain (NTD) interacts exclusively with the C-terminal half of XPB.

The two tri-helix bundles of Bax1 NTD (α1–3 and α4–6, Figure 2.2) are mainly responsible for the interactions with XPB: the first tri-helix bundle (α1–3) interacts with the ThM motif of XPB while the second tri-helix bundle (α4–6) interacts with the HD2 of XPB (Figure 2.1A-B). These unique helix-helix interactions provide the flexibility to allow Bax1 to swing back and forth. The flexible hinge connection joining the HD1 and HD2 of XPB allows the N-terminal half XPB to rotate from open conformation to closed conformation while StBax1 rotates away to avoid clashing with the N-terminal half of StXPB compared to the AfXPB–Bax1 complex (Figure 2.1C).

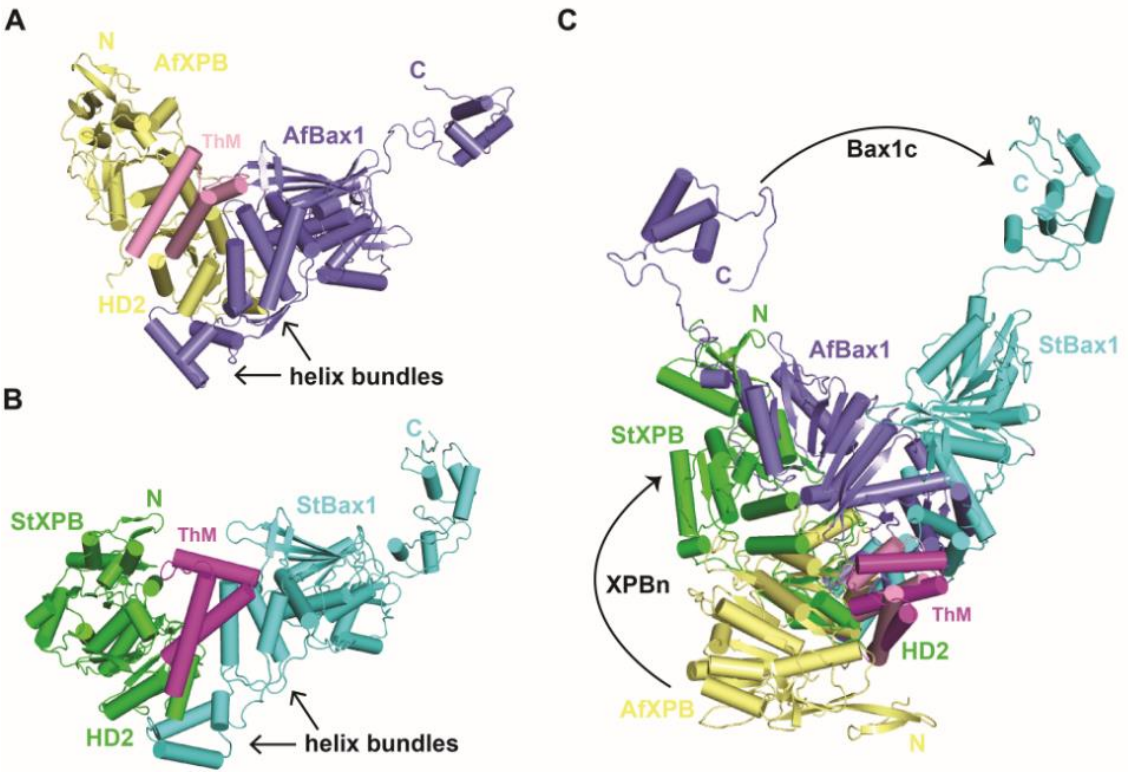

Figure 2.1 The XPB–Bax1 complex is a dynamic machine.

(A) Crystal structure of the AfXPB–Bax1 complex in cartoon. AfXPB is colored in yellow with the ThM motif in pink. AfBax1 is colored in slate blue. The positions of AfBax1 two tri-helix bundles are indicated.

(B) Crystal structure of the StXPB–Bax1 complex in cartoon. StXPB is colored in green with the ThM motif in magenta. StBax1 is colored in cyan. The positions of StBax1 two tri-helix bundles are indicated.

(C) Structural comparison of the StXPB–Bax1 complex with the AfXPB–Bax1 complex. The AfXPB–Bax1 complex is superimposed with the StXPB–Bax1 complex over the HD2 and ThM of AfXPB and StXPB. Different orientations of the N-terminal half XPB (XPBn) and the C-terminal half Bax1 (Bax1c) between the two heterodimers are highlighted by arrows.

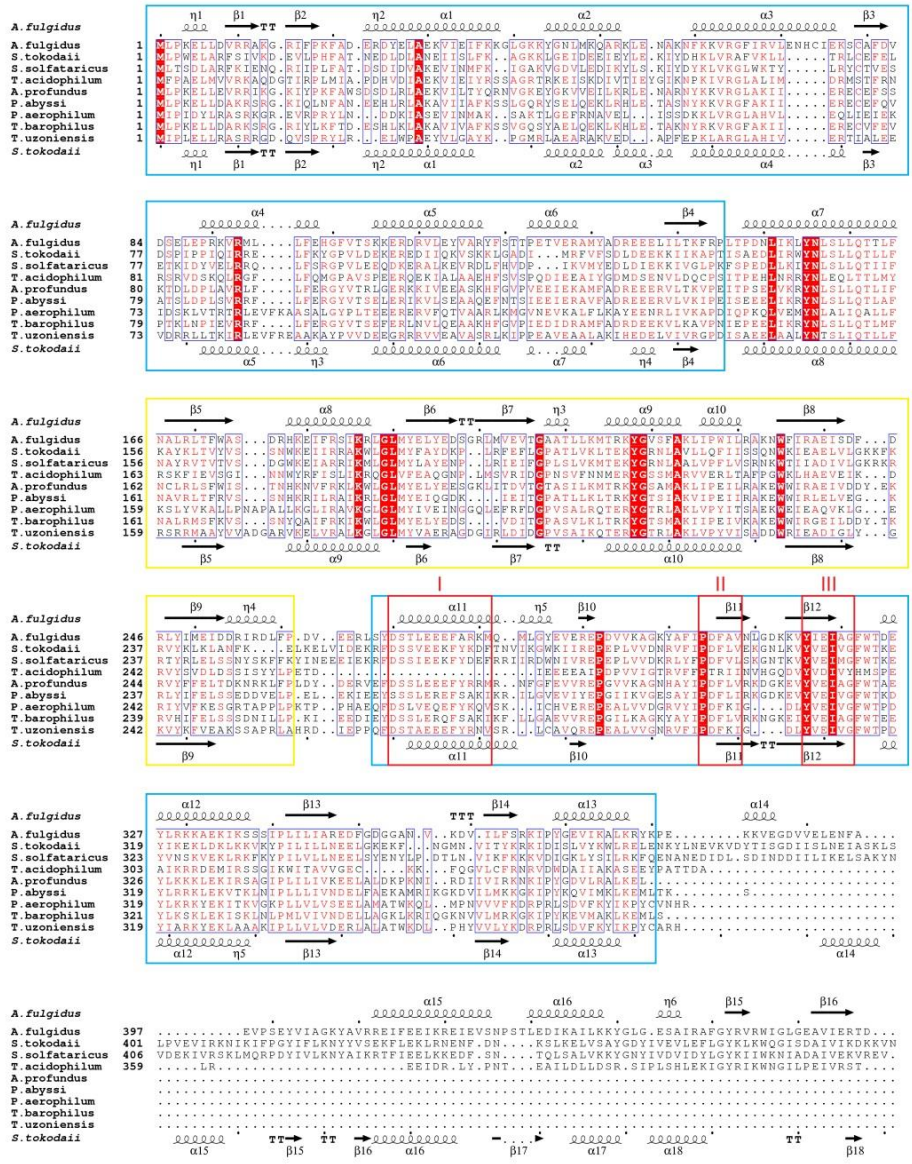

Figure 2.2 Sequence and structural alignments of Bax1 orthologues.

Amino acid sequences of Bax1 orthologues from *Archaeoglobus fulgidus*, *Sulfurisphaera tokodaii*, *Saccharolobus solfataricus*, *Thermoplasma acidophilum*, *Archaeoglobus profundus*, *Pyrococcus abyssi*, *Pyrobaculum aerophilum*, *Thermococcus barophilus*, and *Thermoproteus uzoniensis* are aligned with a 0.5 threshold for similarity. Alignment was performed with Clustal Omega [102] and depicted using ESPript 3.0 server [103]. Secondary structure elements for *A. fulgidus* (top) and *S.tokodaii* (bottom) are numbered and represented according to the PDB files for the AfXPB-Bax1 (PDB entry 6P66) and StXPB-Bax1 (PDB entry 6P4O) structures. Domains of Bax1 are colored in frames: NTD – cyan, CRD – yellow, NUS/nuclease motifs – cyan/red, except the C-terminal domain that is not conserved in all Bax1 orthologues. *S. tokodaii*, *S. solfataricus*, *P. aerophilum*, *T. uzoniensis* belong to the phylum Crenarchaeota while others belong to the phylum Euryarchaeota.

The interaction interface between XPB and Bax1

Further detailed structural analysis reveals that the main interactions contribute to the assembly of the XPB–Bax1 complex consists of a hydrophobic (van der Waals) site and a polar/charge interaction site (Figure 2.3A-B). Thus I tested the impact of some key residues in these two sites by mutagenesis and ITC assays (Figure 2.3C-D). Compared to the interactions between the wild type StXPB and StBax1 (K_d = 2.35 nM), StXPB mutant E357A/E360A interacted with StBax1 mutant R86A/R87A in a much weaker fashion (K_d = 64.9 nM). And the K_d between the StBax1 mutant L89A/F90A/P94S/V95S and the wild type StXPB has the highest value of 3.08 μM under the same condition, indicating a much more weakened interaction. These ITC analyses demonstrate that disruption of the hydrophobic residues has more severe impact on XPB-Bax1 interactions than the charged residues, suggesting a more important role of the hydrophobic site. What's more, the interaction interfaces between XPB and Bax1 are conserved among different archaea species (Figure 2.4).

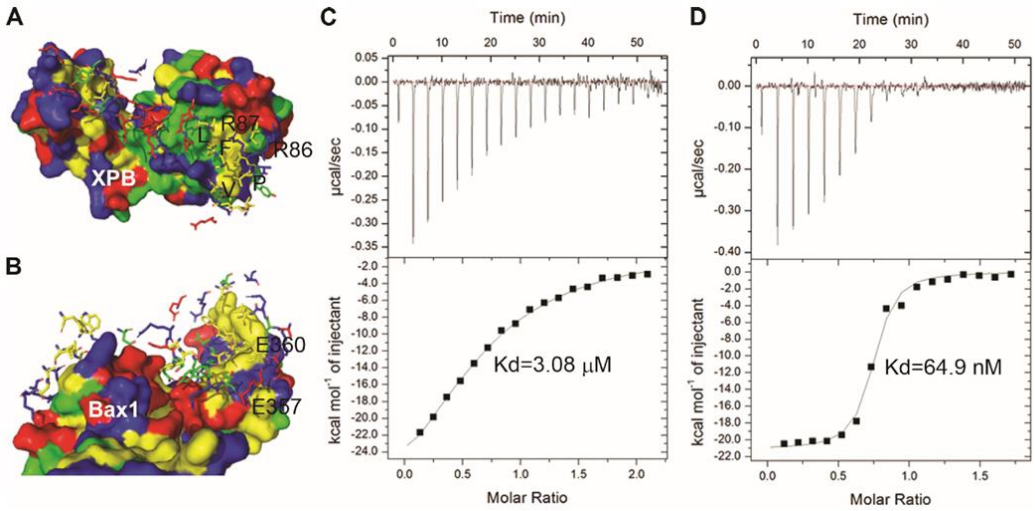

Figure 2.3 Hydrophobic and charge/polar interactions contribute to the XPB-Bax1 assembly.

(A) Biochemical properties of the StXPB surface interacting residues from StBax1: E24, D27, E31, K41, G43, E44, D45, E47, E48, E50YLEKIY56, R62, I83, R86, R87, L89 (L), F90 (F), K91YG93, P94 (P), V95 (V), L96, E98, R101, I104, I105, M117, V120, F121 and D123LDEE127. Residues selected for mutagenesis are indicated.

(B) Biochemical properties of StBax1 surface interacting with residues of StXPB: F278, V282, A285AK287, K289, R292, L295, L296, W298, H299, N303, R316, L319, K323, R332DTQ335, Y338, S341KTFLIPV348, T350YKTD354, E357, E360, I361, K364, E369YRV372, V378 and F379. Residues are represented by sticks and colored according to amino acid properties: yellow for hydrophobic residues, green for polar uncharged residues, red for acidic residues, and blue for basic residues. Biochemical characteristics of residues were determined by the PISA server.

(C) ITC results for WT StXPB and StBax1 L89A/F90A/P94A/V95A mutant.

(D) ITC results for StXPB E357A/E360A mutant and StBax1 R86A/R87A mutant. Each ITC titration represents a typical profile of triplicate experiments with the raw data in the top and data fitting by ORIGIN.

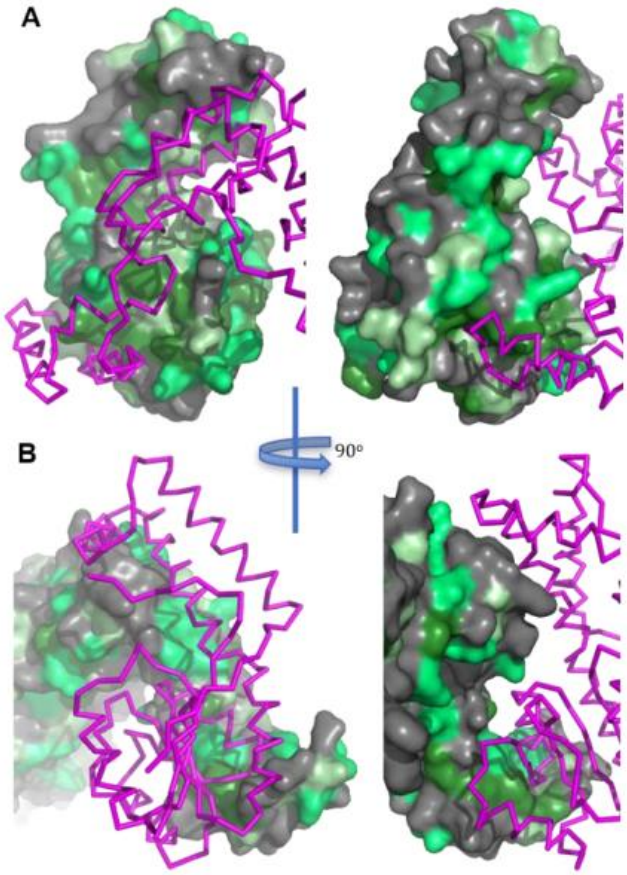

Figure 2.4 The XPB-Bax1 interaction interfaces are conserved among archaea.

A. Sequence conservation on the surface of XPB interacting with Bax1(magenta ribbons). B. Sequence conservation on the surface of Bax1 interacting with XPB (magenta ribbons). Surfaces are colored based on Clustal Omega alignment of XPB sequences (A) or Bax1 sequences (B) from *S. tokodaii*, *A. fulgidus*, *S. solfataricus*, and *T. acidophilum*. Surfaces are color coded as identical residues in dark green, highly similar residues in green, similar residues in pale green, not conserved residues in gray.

AfBax1 contains two distinguished nuclease active sites

Previously research has reported that TaBax1 in the absence of XPB could cleave 3'-flaps at 4-6 nt away from the ds-ss junction [94]. However, SsBax1 works together with SsXPB to cleave 5'-overhang at the ds-ss junction and has no nuclease activities for 3'-overhang DNA substrates [96]. These data show different strand selections of TaBax1 and SsBax1 for cleavage, which lacks a clear explanation. Also, substitution of residues F116, Y128, D130, E132, Y152 and N153 with alanine in TaBax1significantly reduced its nuclease activities [94], leading to the proposal that these residues form the nuclease active site of TaBax1. These residues correspond to F119, Y131, D133, E135, Y155 and N156 of AfBax1 at the interaction interface between XPB and Bax1 (Figure 2.2 and 2.5A). These residues in AfBax1 are away from the conserved Bax1 nuclease domain where substitution of SsBax1 residue Asp-301, a key acidic residue at the conserved nuclease domain (D305 in Figure 2.2 and Figure 2.5B), with alanine eliminated the nuclease activity of the SsXPB–Bax1 complex [96]. These results suggest that there are two nuclease active sites in Bax1 nuclease and protein-protein interactions regulate the polarity of DNA incision by the Bax1 nuclease and the Bax1-XPB complex in order to remove a fragment of damage DNA during DNA repair.

To test this hypothesis, we substituted acidic residues Asp133 and Glu135 with alanine at the N-terminal AfBax1 and acidic residue Asp305 with alanine at the nuclease domain of AfBax1, and tested their influences on protein-protein interactions and nuclease activities (Figure 2.6). AfBax1 with substitutions D133A/E135A still interacted with AfXPB but with much lower affnity (K_d = 157 nM in Figure 2.6A) comparing to the

wild type AfBax1 (K_d = 15.2 nM) based on ITC measurements, confirming the importance of these residues in protein-protein interactions as observed in the crystal structure.

Nuclease activity assays with a 50-bp dsDNA substrate containing a 16-nt bubble (Figure 2.6B and C) reveal that AfBax1 has two nuclease active sites with distinguished nuclease activities. AfBax1 alone shows weak nuclease activity by making an incision (indicated by the black arrow in Figure 2.6C) at the 5' ds region with 5-bp away from the ds-ss junction to produce a 12-nt product (compare lane 3 with lane 2 in Figure 2.6B). The N-terminal nuclease site is likely responsible for this activity since mutation D133A/E135A almost eliminated this activity (compare lane 4 to lane 3 in Figure 2.6B). However, inhibition on the N-terminal nuclease activity by the D133A/E135A mutation enhanced the nuclease activity from the C-terminal nuclease domain, which incises DNA around the ds-ss junction to produce products (indicated by slim light grey arrows in Figure 2.6B and C) with various sizes ranging from 14-nt to 26-nt (lane 4 in Figure 2.6B and C). Mutation D305A in the nuclease domain significantly reduced this new activity (lane 5 in Figure 2.6B) but enhanced the N terminal nuclease activity as revealed by increased level of the 12-nt product (indicated by the black arrow in Figure 2.6B). In addition, the increased levels of the15-nt and 16-nt products for the AfBax1 mutant D133A/E135A suggest that the N-terminal nuclease active site can perform DNA incision at the ds-ss junction (indicated by slim black arrow in Figure 2.6C) as well. These results demonstrated that the N-terminal nuclease active site competes with the nuclease domain for DNA incision, and inhibition on either activity significantly

enhances the other activity. Interestingly, the AfBax1 mutant D133A/E135A/D305A (lane 6 in Figure 2.6B) displayed stronger but disordered activities for both the N-terminal nuclease and the nuclease domain than the wild type AfBax1 (lane 3 in Figure 2.6B), suggesting the regulation of nuclease activities are disrupted by these mutations.

As being expected, the incision by the N-terminal nuclease active site to produce the 12-nt product was inhibited for the AfXPB–Bax1 complex (compare lane 9 with lane 3 in Figure 2.6B) because the interaction of AfXPB with AfBax1 likely blocks DNA from accessing the N-terminal nuclease active site of AfBax1. This leads to the enhancement on the activity from the nuclease domain to produce products longer than 12-nt (lane 9 in Figure 2.6B). Interestingly, the AfXPB–Bax1 complex displayed the ability to incise DNA at the other ds-ss junction of the 16-nt bubble (indicated by grey arrows in Figure 2.6C) to produce products of 28-nt, 32-nt, and 37-nt DNA oligomers (lane 9 in Figure 2.6B). These products were significantly reduced by the mutation D133A/E135A in AfBax1 (lane 10 in Figure 2.6B) and were almost eliminated by the mutation D305A in AfBax1 (lane 11 in Figure 2.6B). These results indicated that the incision at the other ds-ss junction of the bubble is mediated by the nuclease domain of AfBax1 but is regulated by the interactions between AfXPB and AfBax1 since the AfBax1 mutant D133A/E135A interacts with AfXPB much weaker (K_d = 157 nM) than the wild type AfBax1 (K_d = 15.2 nM).

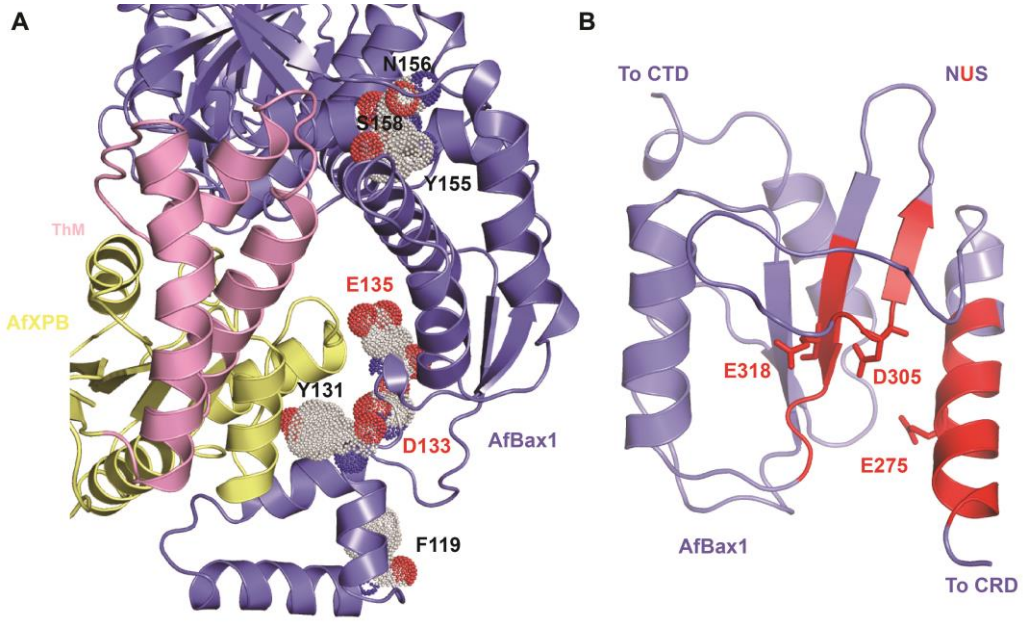

Figure 2.5 The two nuclease sites in AfBax1.

(A) The interface between AfXPB C terminal (including the ThM motif) and Bax1 N terminal domain. Residues equivalent to those mutated in ref. [95] are colored in CPK and shown as dot spheres with labels. Two acidic residues for the potential N-terminal nuclease active site are indicated by red labels.

(B) The conventional NUS domain of AfBax1. The nuclease motifs are colored in red as in Figure 2.2. Active site residues are shown in sticks.

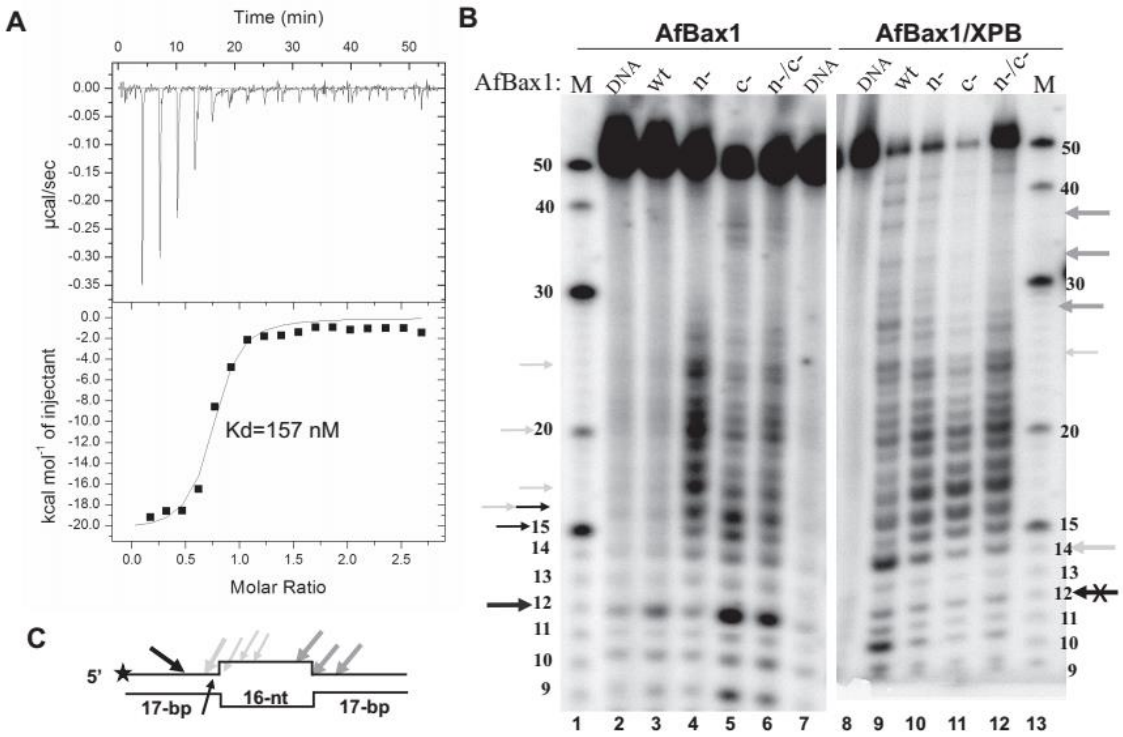

Figure 2.6 Mutational and biochemical analysis of the two nuclease sites in AfBax1 by ITC and the nuclease activity assay.

(A) The N-terminal active site plays a role in Bax1-XPB interactions. Mutation D133A/E135A increases the K_d of the AfXPB–Bax1 complex by 10-fold. ITC titration represents a typical profile of multiple assays with the raw data in the top and data fitting by the ORIGIN software in the bottom.

(B) DNA incisions on a 16-nt bubble DNA substrate by AfBax1 variants and their complexes with AfXPB. M: DNA oligomer markers, DNA: nuclease reaction control without AfBax1 or AfBax1-XPB complex, wt: wild type AfBax1, n-: AfBax1 mutant D133A/E135A, c-: AfBax1 mutant D305A, n-/c-: AfBax1 mutant D133A/E135A/D305A. Black arrows indicate incised products by the N-terminal nuclease active site (X indicates inhibition on the activity); Grey and light gray arrows indicate incised products by the nuclease domain.

(C) Schematic summary of the results from (B). The star indicates P-32 label on the DNA strand.

DISCUSSION

Based on the crystal structures of the XPB-Bax1 complex from two archaeal species, we have identified the essential interaction interface between XPB and Bax1. The two helix bundles in Bax1-NTD exclusively interact with the XPB HD2 and ThM, allowing archaeal XPB-Bax1 to perform its biological functions with substantial conformational flexibilities of XPB HD1 domain and Bax1 CRD/NUS domains. On the other hand, the conformational changes of the HD2 and ThM in XPB could also regulate the domain orientation of XPB HD1 domain to Bax1 protein. Therefore, this interaction feature between XPB and Bax1 fundamentally enables this complex to be a dynamic machine for DNA repair. Our ITC results on the interface mutants confirmed the key residues for the interaction and revealed hydrophobic interactions may play a dominant role in the assembly of the XPB-Bax1 complex.

In the absence of Bax1, AfXPB is in the open conformation while StXPB is in partially closed conformations as observed in the AfXPB (PDB entry: 2FWR) and StXPB (PDB entry: 5TNU) crystal structures, respectively. In the crystal structure of the AfXPB–Bax1 complex, AfXPB remains in the open conformation just like AfXPB alone with loss of the ATP binding groove. In order to form the ATP-binding groove, the N-terminal half (DRD and HD1) of AfXPB has to rotate about 170 °to form the closed conformation. The StXPB–Bax1 complex keeps StXPB in a much more closed conformation, which is favorable for ATP-binding. This explains previous results that StBax1 enhances significantly the ATPase activity of StXPB [97]. These observations combined with our structural analysis in Figure 2.1 demonstrate that Bax1 can adjust its

relative position to XPB allowing XPB to form a closed conformation or open conformation, and this process is likely regulated by ATP binding/hydrolysis by XPB and the interactions of XPB-Bax1 with DNA to coordinate DNA unwinding with incision during DNA repair.

Previous studies reported that TaBax1 has nuclease activities on DNA substrates containing 3'-overhang while SsXPB-Bax1 only cleaves 5'-overhang at the junction. Nuclease activity assays with mutations of Bax1 suggest TaBax1 has a different active site than SsBax1 as well. These results from the two archaeal species remain intriguing and seem to be conflicting with each other. Here we carried out nuclease activity assay on WT and nuclease-site mutants of AfBax1 and AfXPB-Bax1, and demonstrated that AfBax1 contains two distinguished nuclease active sites. One nuclease active site is located in the N-terminal domain of Bax1 at the interaction interface between XPB and Bax1, explaining why TaXPB inhibited the nuclease activity of TaBax1 as the association of XPB will block this active site from access by any DNA substrates. The other site is located in the conserved nuclease domain since Ala-substitution of Asp-301, a key acidic residue at the conserved nuclease domain previously identified by bioinformatic analysis, completely eliminated the nuclease activity of the SsXPB–Bax1 complex. Interestingly, we observed that inhibition of one nuclease site will enhance the activity on the other nuclease site (Figure 2.6). Furthermore, interactions with XPB block the N-terminal nuclease activity and change the properties of DNA incision by the Bax1 nuclease domain as the AfXPB–Bax1 complex shows different DNA incision patterns from DNA incision by AfBax1 alone (Figure 2.6B). Similarly, the TaXPB–Bax1

complex was previously reported to cleave 5'-overhang while TaBax1 alone cleaves 3'-overhang. These results together demonstrate protein-protein interactions regulate DNA incision by the Bax1 nuclease in order to remove a fragment of damage DNA during DNA repair (Figure 2.7A).

We were able to draw a cartoon model to explain this regulation of DNA cleavage by XPB-Bax1. As shown in Figure 2.7A, Bax1 likely contains two active sites: one at the N-terminal domain for 5' cleavage at the DNA bubble and the other at the nuclease domain for 3' cleavage to the bubble. Formation of the XPB–Bax1 complex enhances the nuclease activity on the 3' cleavage (like XPG in eukaryotes) by Bax1. When a bubble is created around DNA lesion during NER, the XPB–Bax1 complex is responsible for the 3' incision to the damage while the 5' incision to the damage is likely achieved by two different mechanisms. In euryarchaea lacking XPF homologs (such as *T. acidophilum*), Bax1 is primarily a monomer and acts like XPF to cleave the damage strand 5' to the lesion. In crenarchaea containing XPF homolog, the 5' incision is likely carried out by XPF nuclease. In this case, Bax1 forms a homodimer to mask both active sites in order to avoid an active Bax1 nuclease for competition with XPF. This can be achieved by the two Bax1 monomers to interact with each other through N-terminal domain to C-terminal domain cross-interactions, resulting in both active sites at the nuclease domain and the N-terminal domain blocked from access by DNA substrates in the Bax1 homodimer. In agreement with this model, we observed that StBax1 forms a homodimer in solution (Figure 2.7B).

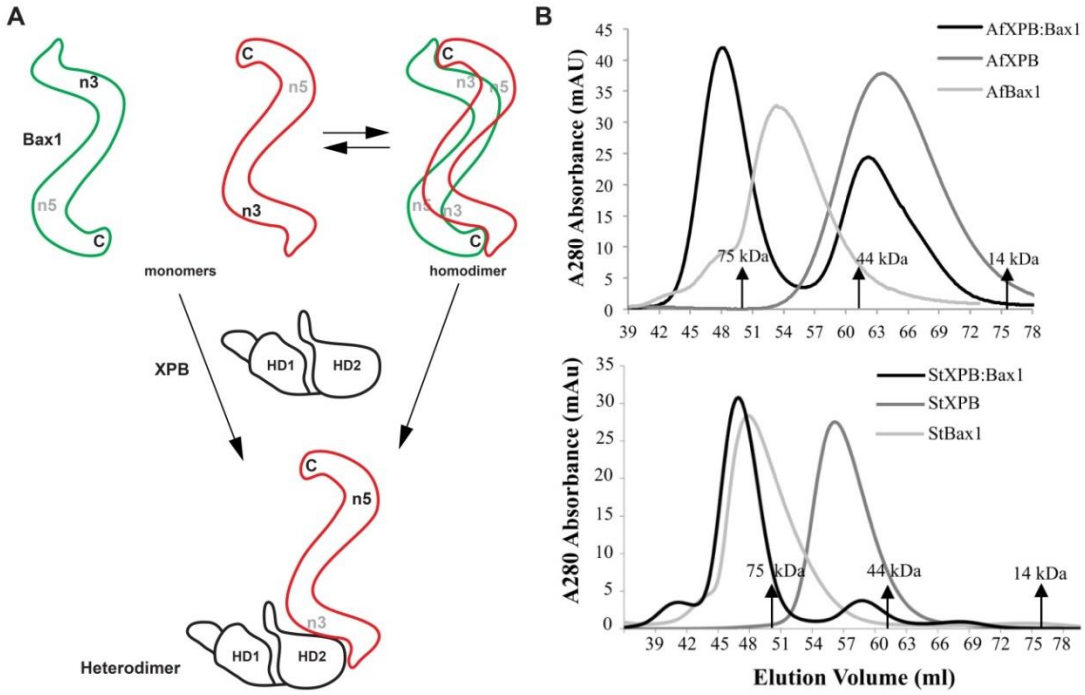

Figure 2.7 Protein-protein interactions regulate the nuclease activity of Bax1.

(A) Bax1 likely contains two nuclease active sites: one (n5) at the N terminal domain and the other (n3) at the nuclease domain. Bax1 from euryacharchaea *A. fulgidus* and *T. acidophilum* is predominantly monomer in solution and is in the conformation preferable for the N-terminal (n5) activity, which presumably incises DNA 5' to the damage. StBax1 and SsBax1 from crenarchaea *Sulfurisphaera* are exclusively homodimers in solution and the two nuclease active sites are mutually masked due to dimerization. Both monomeric and dimeric Bax1 interact with XPB to form the heterodimeric XPB–Bax1 complex which masks the N-terminal active site (n5) but enhances the activity of the nuclease domain (n3). Active nuclease sites are highlighted by black labels (n3 or n5) while inhibited nuclease sites are labeled in gray. Bax1 C-terminus is indicated by letter C.

(B) Profle of S200 size-exclusion chromatography for Af and St proteins. The positions of three protein markers are indicated based on chromatographic calibration profle of Conalbumin (75 kDa), Ovalbumin (44 kDa) and Lactalbumin (14 kDa).

CHAPTER 3

Structural insights into DNA binding and opening by the XPB-Bax1 complex and implications for nucleotide excision repair

ABSTRACT

Nucleotide excision repair (NER) removes various DNA lesions caused by UV light and chemical carcinogens. The DNA helicase XPB plays a key role in DNA opening and coordinating damage incision by nucleases during NER, but the underlying mechanisms remain unclear. Here, we report crystal structures of XPB from *Sulfurisphaera tokodaii* (St) bound to the nuclease Bax1 and their complex with a bubble DNA having one arm unwound in the crystal. XPB and Bax1 together spirally encircle 10 base pairs of duplex DNA at the double-/single-stranded (ds-ss) junction. Furthermore, StXPB has its ThM motif intruding between the two DNA strands and gripping the 3'-overhang while Bax1 interacts with the 5'-overhang. This ternary complex likely reflects the state of repair bubble extension by the XPB and nuclease machine. ATP binding and hydrolysis by StXPB could lead to a spiral translocation along dsDNA and DNA strand separation by the ThM motif, revealing an unconventional DNA unwinding mechanism. Our structural analyses also suggest eukaryotic XPB in TFIIH cooperates with XPA for the initial DNA opening around the lesion. Interestingly, the DNA is kept away from the nuclease domain of Bax1, potentially preventing DNA incision by Bax1 during repair bubble extension.

INTRODUCTION

Transcription and DNA repair are two essential biological processes. As the largest subunit of the transcription factor TFIIH complex, XPB is required for promoter melting in transcription and unwinding damaged DNA in nucleotide excision repair. Mutations in XPB are associated with xeroderma pigmentosum (XP), trichothiodystrophy (TTD) and Cockayne syndrome (CS) symptoms with developmental disorders or increased frequency of skin cancer. In transcription, XPB binds dsDNA downstream from the promoter opening location and has been proposed to function as a "molecular wrench" [41] or dsDNA translocase [36, 42, 43]. In the general genomic NER pathway, DNA lesions are first recognized by the XPC-HR23B complex, which directly recruits the TFIIH complex for DNA unwinding. TFIIH uses its helicase subunits XPB and XPD to generate the repair bubble. XPB likely initiates DNA unwinding at the lesion since XPD is a conventional SF2 helicase and requires a ssDNA extension to start unwinding. Other NER factors including XPA and replication protein A (RPA) are required to facilitate the assembly of the preincision complex. After the DNA lesion is verified by TFIIH, the ERCC1-XPF complex and XPG nucleases incise the damaged strand at 5' and 3' side to the lesion, respectively, to remove a damage-containing fragment of 25-30 nucleotides. The gap is finally filled by the DNA replication machinery [16, 22, 40, 44, 45, 52, 66]. However, it is still unclear how XPB recognizes the DNA substrates and initiates unwinding in NER. Structural analysis on crystal structures of *Archaeoglobus fulgidus* XPB (AfXPB) [15] and StXPB [104] suggested that domain rotation in XPB might generate a supertwist in DNA at the lesion,

leading to the initial unwinding, consistent with the recent cryo-EM structure of XPA and the TFIIH core bound to a forked DNA substrate showing that human XPB acts as a translocase by binding to the dsDNA region ahead of the fork during DNA repair [63]. In archaea, due to the lack of the TFIIH-like complex, XPB is in complex with Bax1, an XPG-like nuclease, to function as a helicase-nuclease machine for DNA unwinding and incision. Our recent crystal structures of the XPB-Bax1 complex from both *Archaeogloubus fulgidus* (Af) and *Sulfurisphaera (*previously named *Sulfolobus) tokodaii* (St) reveal that the XPB-Bax1 complex is a dynamic machinery which can adapt different conformations for protein-protein and protein-substrate interactions.

Here we determined the crystal structures of the StXPB-Bax1$^{\Delta C}$ (a truncated Bax1 without the C-terminal domain, which is absent in many archaeal Bax1 homologs) complex and the StXPB-Bax1$^{\Delta C}$ heterodimer associated with a bubble DNA substrate, which has one dsDNA arm unwound in the crystal to become a forked DNA. StXPB in the DNA-free heterodimeric structure contains a phosphate ion in its ATP-binding site, possibly mimicking the state of StXPB after ATP hydrolysis (ADP+Pi). Structural and mutational analyses reveal that the conserved RED and ThM motifs play key roles in DNA interactions and XPB activities, consistent with previous results on both archaeal and human XPB [63, 95]. These results provide new insights into the molecular mechanisms of XPB-mediated DNA repair bubble formation in archaeal and eukaryotic NER.

MATERIALS AND METHODS

Cloning, expression and purification of StXPB-Bax1 and StXPB-Bax1$^{\Delta C}$

The DNA encoding StXPB (residues 2-439) was cloned into a modified pET28a vector with an N-terminal His_8-tag followed by a PreScission protease cleavage site, while the DNA encoding the full-length StBax1 and a truncated StBax1$^{\Delta C}$ (residues 2-373) were cloned into the pET15b vector by PCR. Purification of StXPB alone was described previously in Chapter 2. StXPB and StBax1 or StBax1$^{\Delta C}$ were co-expressed in *E. coli* Rosetta (DE3) pLysS cells (Invitrogen). After induction for 18 hours with 0.2 mM IPTG at 28 °C, the cells were harvested by centrifugation and the pellets were resuspended in lysis buffer containing 50 mM Tris-Cl pH 7.5, 500 mM NaCl, 10% glycerol. The cells were then lysed by sonication and the cell debris was removed by centrifugation. The supernatant was purified by affinity chromatography using Ni-NTA resin (Thermo Scientific). PreScission protease was then added to remove the His_8-tag. The protein complex was further purified by HiTrap SP FF ion-exchange chromatography (GE). The purification was completed by gel-filtration chromatography (Superdex 200, 16/60, GE) in 25 mM Tris-Cl pH 7.5, 200 mM NaCl or 25 mM HEPES pH 7.5, 200 mM NaCl (for crystallization). The purified protein samples were concentrated and stored at -80 ℃. All the variants of StXPB were expressed and purified following similar procedures.

Crystallization and structure determination

Crystals of the StXPB-Bax1$^{\Delta C}$ complex were prepared from 200 mM NH_4-citrate pH 7.5, 8% PEG3350 by the sitting-drop vapor diffusion at room temperature. Synthesized DNA oligos are used as additives in the drop to promote crystal formation. Crystals grew as plates to maximal size within 1 week. Crystals were gradually transferred into a harvesting solution made of mother liquor supplemented with 26% ethylene glycol, followed by flash-freezing in liquid nitrogen for shipment to synchrotron facilities. X-ray diffraction datasets for StXPB-Bax1$^{\Delta C}$ complex were collected at beamline 5.0.1 at the Advanced Light Source, Lawrence Berkeley National Laboratory, and the diffraction data were indexed, integrated, and scaled using the HKL3000 program [105]. The structure was solved by molecular replacement using Phaser [106], with individual domains of the StXPB-Bax1 structure (PDB entry: 6P4O) as search models. Protein structure refinement was carried out with the REFMAC5 [107].

The StXPB-Bax1$^{\Delta C}$-DNA ternary complex was crystallized by sitting-drop vapor diffusion at room temperature. The StXPB-Bax1$^{\Delta C}$ complex was mixed with the bubble-6 DNA at a protein:DNA ratio of 1:1.2, followed by incubation for 40 min at room temperature. The protein-DNA co-crystals typically grew in a reservoir solution consisting of 50 mM MES pH 5.3, 10 mM $MgCl_2$, 26% 2-methyl-2,4-pentanediol (MPD). The quality of crystals was improved by micro-seeding. Crystals grew as plates to maximal size in 2 weeks. Crystals were transferred into a harvesting solution containing 50 mM MES pH 5.3, 10 mM $MgCl_2$ and 28% MPD, followed by flash-freezing in liquid nitrogen. The dataset for the XPB-Bax1$^{\Delta C}$-DNA complex was collected on the 24-ID-C NE-CAT beamline at

the Advanced Photon Source, Argonne National Laboratory, and the diffraction data were indexed and integrated using iMOSFLM [108], then scaled and merged with SCALA [109]. The structure was solved by molecular replacement with Phaser using individual domains of the StXPB-Bax1$^{\Delta C}$ structure as search models. Positive density appearing in the difference map was identified as DNA, which was manually built into the density and improved in Coot [110], refinement was performed using the PHENIX software package [111]. All the structural figures were prepared with PyMOL (version 1.8.6.2, Schrodinger LLC; http://www.pymol.org/).

Cloning, expression and purification of human XPB/p52/p8 trimer

The DNA encoding full-length human XPB was cloned into a modified Bac-to-Bac vector with an N-terminal His_6-tag followed by a PreScission protease cleavage site. The DNA encoding full-length human p52 and p8 were cloned into MacroBac 438A vector, and then p52 and p8 were combined into a single vector via restriction digestion and ligation-independent cloning [112]. The recombinant baculovirus expressing XPB or p52/p8 was generated using standard protocols. High Five insect cells were co-infected with these two recombinant baculoviruses. The cells were harvested after 65-70 hours by centrifugation. The pellets were resuspended in lysis buffer containing 50 mM Tris-Cl pH 7.0, 500 mM NaCl, 10% glycerol, 5 mM β-mercaptoethanol, 1 mM PMSF. The cells were then lysed by sonication, and the debris was removed by ultracentrifugation. The supernatant was mixed with Ni-NTA resin and rocked for 1 hour at 4 ℃ before elution with 400 mM imidazole. PreScission protease was then added to remove the His_6-tag. The proteins were further purified by ion-exchange chromatography (SP-FF, GE) and

gel-filtration chromatography (Superdex 200, 16/60, GE). The purified protein samples were concentrated in 25 mM Tris-Cl pH 7.5, 200 mM NaCl, 5% glycerol, 2 mM DTT, and stored at -80 ℃.

Cloning, expression and purification of human XPA

The DNA encoding the full-length human XPA was cloned into a modified pET28a vector with a cleavable N-terminal His_6-SUMO tag for expression in *E. coli* Rosetta (DE3) pLysS cells (Invitrogen). After induction for 18 hours with 0.2 mM IPTG at 22 °C, the cells were harvested by centrifugation and the pellets were resuspended in lysis buffer containing 50 mM Tris-Cl pH 7.5, 500 mM NaCl, 10% glycerol, 10 μM $ZnCl_2$, 5 mM β-mercaptoethanol. The cells were then lysed by sonication and the cell debris was removed by centrifugation. The supernatant was purified by Ni-NTA affinity chromatography and SUMO protease was then added to remove the His_6-SUMO tag. XPA was further purified with the Heparin (GE) and Superdex 200 (16/60, GE) columns. The purified XPA protein samples were concentrated in 25 mM Tris-Cl pH 7.5, 200 mM NaCl, 5% glycerol, 2 mM DTT, and stored at -80 ℃.

Electrophoretic mobility shift assay

For the single-stranded DNA (ssDNA) substrates, the protein samples and the ssDNA oligonucleotides were mixed at room temperature (RT) for 1 h with a molar ratio of 1:1 in the binding buffer consisting of 50 mM Tris-Cl pH 7.5, 100 mM NaCl, 5% glycerol. Samples were then loaded to a 1.5% TBE agarose gel, which was run for 90

min (60V) in the TBE buffer at room temperature. The gels were stained by SYBR gold Nucleic Acid Gel Stain (Thermo Fisher Scientific) and visualized under UV light.

For the forked DNA substrates, unless otherwise indicated, 0.8 μM DNA was incubated with 0.4, 0.8 μM StXPB, StXPB-Bax1$^{\Delta C}$, human XPB/p52/p8, human XPA or both XPB/p52/p8 and XPA (each at 0.4 or 0.8 μM) in 10 μL binding buffer at room temperature for 40 min. For the bubble DNA substrates, unless otherwise indicated, 0.3 μM DNA was incubated with 0.3, 0.6 μM StXPB in 10 μL binding buffer at room temperature for 40 min. The binding buffer consists of 25 mM Tris-Cl pH 7.5, 100 mM NaCl, 5% glycerol, 1 mM DTT. All samples were loaded onto 4% native polyacrylamide gel and resolved under 100 V for 30-40 min in 0.5 × TBE buffer. The gels were then stained by ethidium bromide and visualized under UV light.

ATPase activity assay

ATPase reactions were carried out in a 20 μL reaction buffer (50 mM Tris-Cl pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT) with 1 mM ATP. 1 μM StXPB or StXPB-Bax1 WT and mutants were assayed in the absence or presence of 1 μM DNA substrate in a 50 ℃ water bath for 10 min. The concentration of liberated phosphate from hydrolyzed nucleotides was detected as previously described [113]. The absorbance of reactions with nucleotide alone was subtracted from protein reactions to account for ATP auto-hydrolysis.

RESULTS

Optimizing XPB-Bax1 construct for co-crystallization with DNA

Initially full-length AfXPB-Bax1 and StXPB2-Bax1 proteins (Fig. 3.1A) were used for co-crystallization trials with various ssDNA oligonucleotides of different lengths (from 6-nt to 40-nt), since previous publications on XPB-Bax1 from other archaeal organisms demonstrated XPB-Bax1 has enhanced ssDNA binding affinity compared with XPB alone, and also my EMSA results also suggested that full-length St XPB-Bax1 complex has good affinity for ssDNA of different lengths (Fig. 3.1B). Crystals of full-length St XPB2-Bax1 in complex with ssDNA were obtained but only diffracted to ~7.5 Å resolution at best even after extensive optimization. And co-crystallization trials of full length XPB-Bax1 proteins with many other DNA substrates like 3'-overhang DNA, forked DNA did not produce any diffracting quality crystals.

Then I sought to design several new truncated StBax1 constructs by removing its C terminal domain, which seems to be highly flexible in the structure of StXPB-Bax1 and is also absent in many archaeal Bax1 orthologues (Fig. 2.2). The truncated construct StBax1 (1-370 aa) was firstly screened out because of its successful co-expression with StXPB. And this construct was further optimized to StBax1 (1-373 aa, termed as Bax1$^{\Delta C}$, Fig. 3.2A) to further stabilize the last α helix structure of StBax1$^{\Delta C}$ and enhance the crystallization. The new version of StXPB-Bax1$^{\Delta C}$ complex were expressed and purified with higher yield and purity than the full length ones (Fig. 3.1C). The StXPB-Bax1$^{\Delta C}$

complex interacts with different DNA substrates (using the forked DNA as an example) in the same way as the StXPB-Bax1 complex does (Fig. 3.1D).

Different DNA substrates including ssDNA of different sizes, dsDNA with different overhangs, forked DNA with different arms were tested in co-crystallization trials with the purified StXPB-Bax1$^{\Delta C}$ heterodimer. Large rod crystals ranging up to approximately 400 μm in the longest dimension (Fig. 3.2A) were produced with several forked DNA substrates. However, these crystals always had big limitations on X-ray diffraction even after many rounds of refinements such as crystallization buffer optimization, dehydration, annealing, and seeding. Reductive methylation method for the StXPB-Bax1$^{\Delta C}$ proteins was also applied to lower the surface entropy [114, 115] but still could not further improve the diffraction, even though good crystals could still be produced (Fig. 3.2B). The best resolution of those crystals only could reach ~7.5 Å through diffraction tests on hundreds of crystals by an in-house X-ray source or X-ray beam line. Meanwhile, I designed some bubble DNA oligonucleotides (dsDNA containing an internal mismatched bubble region) with different bubble sizes. And a remarkable feature of these DNA is that they have one-base overhangs at both ends of the two DNA duplex arms. New crystals with the plate-shape were screened out with a bubble-6 DNA in a new condition containing 2-methyl-2,4-pentanediol (MPD) as the precipitant, which is a popular precipitant found in crystallization buffers for many protein-DNA structures. Finally, larger and better well-diffracting crystals were obtained after several rounds of refinement and micro-seeding (Fig. 3.2C-D). Finally, one good dataset was successfully collected from one of those crystals at the beam line.

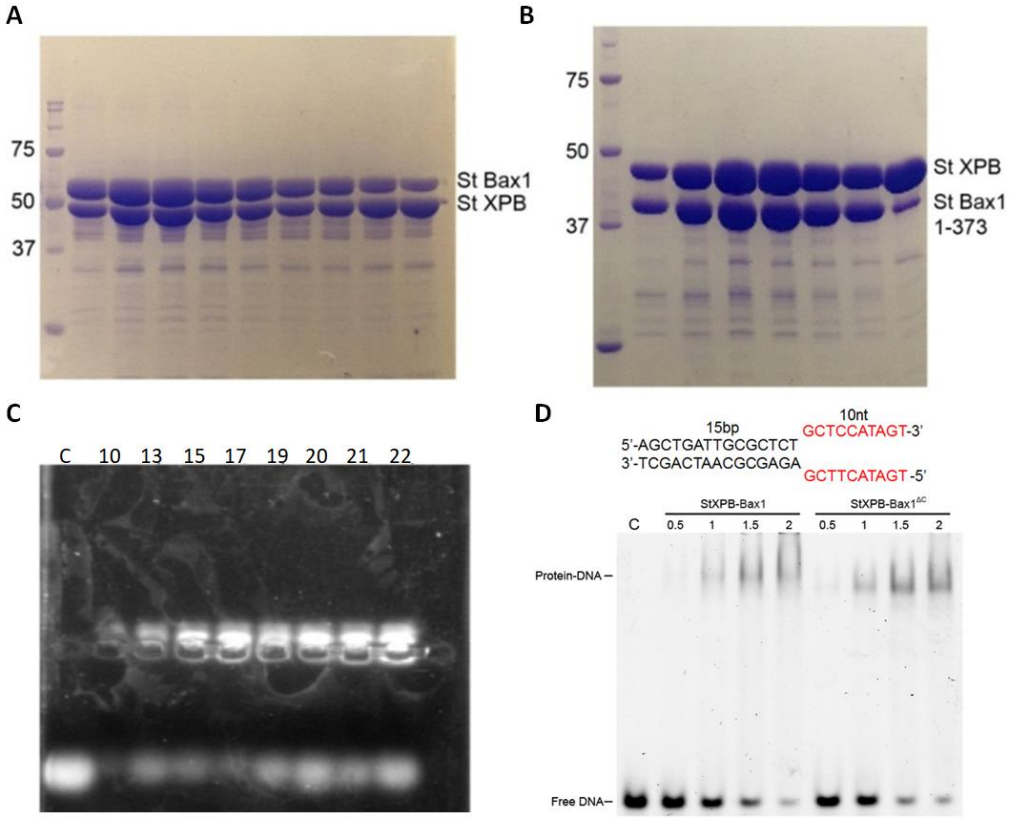

Figure 3.1 StXPB-Bax1 proteins and their interactions with DNA.

(A) Protein fractions of full length StXPB-Bax1 proteins from S200 gel filtration analyzed by SDS-PAGE and Coomassie staining.

(B) Protein fractions of StXPB-Bax1$^{\Delta C}$ proteins from S200 gel filtration analyzed by SDS-PAGE and Coomassie staining.

(C) EMSA of full length StXPB-Bax1 with ssDNA substrates from 10-nt to 22-nt by the agarose gel.

(D) Comparison of DNA binding affinities between full length StXPB-Bax1 and StXPB-Bax1$^{\Delta C}$ by EMSA. Top: Sequence of the forked DNA substrate used for EMSA. The molar ratio of protein:DNA (0.5, 1.0, 1.5, and 2.0) for each reaction is indicated on the top of the gel. Experiments were repeated twice with consistent results.

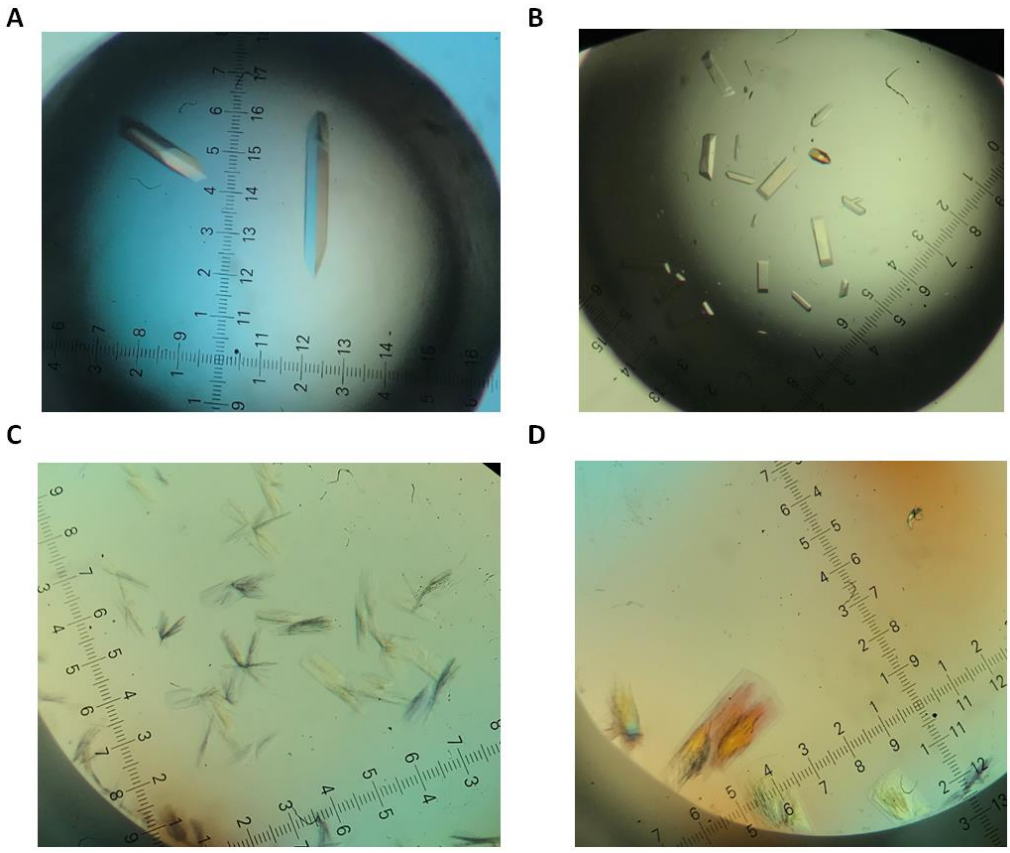

Figure 3.2 Crystallization of StXPB-Bax1 with DNA substrates.

(A) Example crystal of StXPB-Bax1 with forked DNA substrates.

(B) Example crystal of reductively methylated StXPB-Bax1 with forked DNA substrates.

(C) Initial crystal of StXPB-Bax1 with the bubble-6 DNA substrate from the 96-well screening plate.

(D) Crystal of StXPB-Bax1 with the bubble-6 DNA substrate after optimization.

Overall structure of the XPB-Bax1-DNA ternary complex

The crystal structure of the StXPB-Bax1$^{\Delta C}$-DNA ternary complex was determined at 3.55 Å resolution (Table 1 for statistics of data collection and structure refinement). There are two StXPB-Bax1$^{\Delta C}$-DNA complexes in one asymmetric unit. The two ends of the longer dsDNA arms are stacked against each other, suggesting the one-base overhang strategy contributed to the crystal packing (Fig. 3.3). The bubble-6 DNA substrate used for the co-crystallization is a 24 base-pair (bp) DNA duplex containing a 6-nucleotide unpaired region (Fig. 3.4B). Surprisingly, the 6-bp short arm of the bubble-6 DNA was unwound in the crystal (Fig. 3.3 and 3.4C), which is consistent with our previous observation that binding of XPB to DNA induces changes in DNA electrochemical properties even in the absence of ATP [104]. The dsDNA region retains the B form while the two ssDNA tails are bent and split apart by XPB and Bax1$^{\Delta C}$, respectively (Fig. 3.4D). Bax1$^{\Delta C}$ contains three domains (Fig. 3.4A and 3.4D): the N-terminal domain consisting of two helix-bunddles (NTD), the central Cas2-like domain (CRD) and the nuclease domain (NUS). The DNA-bound StXPB-Bax1$^{\Delta C}$ heterodimer spirally encircles the DNA substrate by the HD1/HD2/ThM of XPB and the NTD/CRD of Bax1 (Fig. 3.4D), forming a tunnel for 10-bp DNA duplex binding with XPB closer to the fork (Fig. 3.4A). Furthermore, the ThM motif of XPB intrudes between the two ssDNA tails like a wedge with the 3'-overhang extending through the channel formed by the HD2/ThM of XPB (Fig. 3.4D and 3.4A and 3.4B) and the 5'-overhang extending into the space between two N-terminal β-hairpins of Bax1$^{\Delta C}$ (Fig. 3.4D and 3.4A). These observations are consistent with the 3'-5' helicase polarity of archaeal XPB [15] (moving along the 3'-overhang strand toward the

fork junction) and the nuclease activity of SsBax1 on the DNA substrate containing a 5'-overhang in vitro [96]. Neither XPB nor Bax1$^{\Delta C}$ interacts with the remaining nucleotides of the two ssDNA tails further away from the fork, leading to poor electron density for this portion of the DNA.

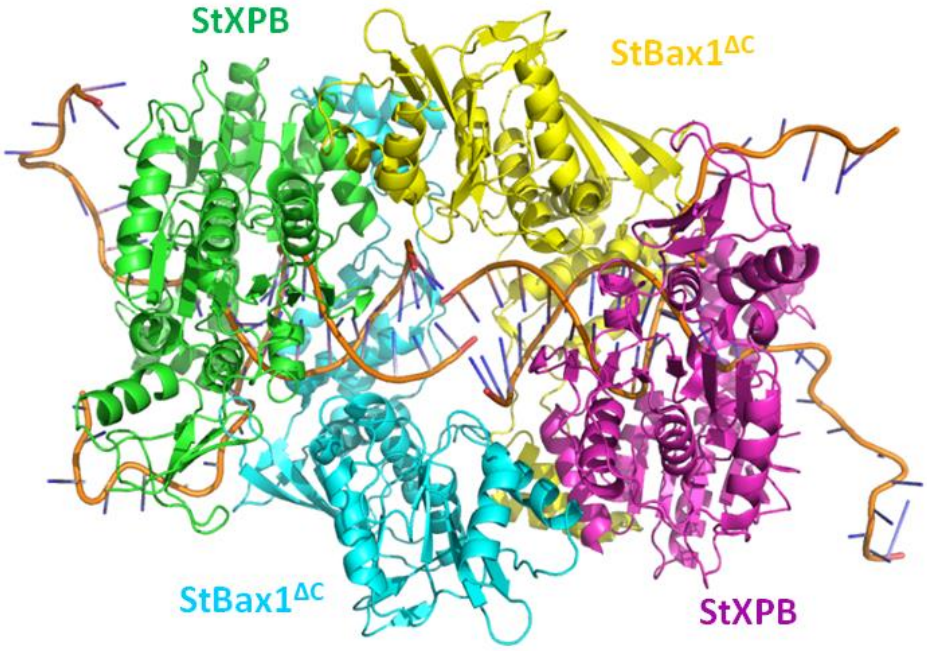

Figure 3.3 Two StXPB-Bax1$^{\Delta C}$-DNA complexes in one asymmetric unit.
In the StXPB-Bax1$^{\Delta C}$-DNA ternary structure, two StXPB molecules in the asymmetric unit were colored in green and magenta, respectively, and two StBax1$^{\Delta C}$ molecules in the asymmetric unit were colored in cyan and yellow, respectively. The bubble-6 DNA became a forked DNA in the crystal and was represented in cartoon.

Table 1 Statistics for X-ray diffraction data collection and structural refinement of the StXPB-Bax1$^{\Delta C}$-DNA ternary structure

Structure	StXPB:Bax1$^{\Delta C}$-DNA
PDB ID	6P4F
Data collection	
Space group	C1 2 1
Cell dimensions: a, b, c (Å)	214.69, 92.15, 172.13
α, β, γ (°)	90.00, 132.24, 90.00
Resolution (Å)	39.86 – 3.55 (3.74 – 3.55)
R_{pim}	0.148 (0.973)
I/σI	4.9 (1.3)
Completeness (%)	94.1 (94.3)
Multiplicity	3.0 (3.0)
CC1/2 (%)	97.7 (30.5)
Refinement	
Resolution (Å)	39.74 – 3.55 (3.68-3.55)
No. reflections	28335
R_{work} / R_{free}*	25.02 / 27.16
Number of atoms	14209
Protein	12310
Ligands	1899
Water	0
Ramachandran favored	91.80%
Ramachandran allowed	8.20%
Ramachandran outliers	0.00%
R.m.s.d Bond length	0.004 Å
R.m.s.d Bond angles	0.73°
Fo, Fc correlation	0.91
Anisotropy	0.128
Averaged B factor	108.8 Å^2
MolProbity score	2.05

Values in parenthesis are for the highest resolution shell. *5% data was used for R_{free}.

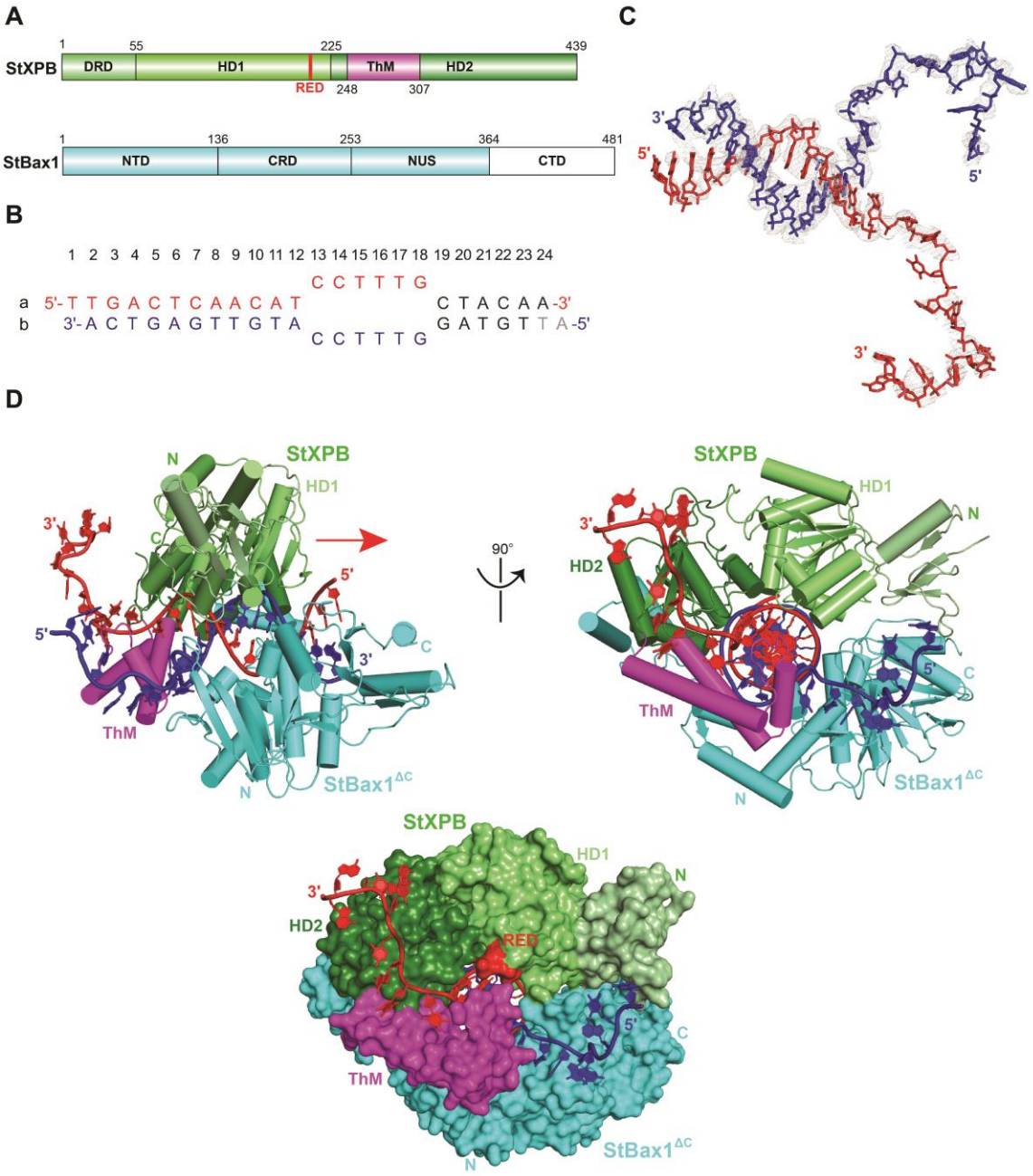
A
StXPB
1
55
225
439
DRD
HD1
ThM
HD2
RED
248
307
StBax1
1
136
253
364
481
NTD
CRD
NUS
CTD
B
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
C C T T T G
a 5'-T T G A C T C A A C A T C T A C A A-3'
b 3'-A C T G A G T T G T A G A T G T T A-5'
C C T T T G
C
3'
5'
5'
3'
D
StXPB
N
HD1
3'
C
5'
5'
3'
C
ThM
N
StBax1ΔC
90°
StXPB
HD1
N
3'
HD2
C
5'
ThM
N
StBax1ΔC
StXPB
HD1
N
3'
HD2
RED
5'
C
ThM
N
StBax1ΔC

Figure 3.4 Structure of the StXPB-Bax1$^{\Delta C}$-forked DNA complex.

(A**)** Diagrams of domain arrangements in StXPB and StBax1. Domains are presented as boxes in different colors with labels: DRD (damage recognition domain), HD1 (helicase domain 1), HD2 (helicase domain 2) and ThM (thumb-like) domains of StXPB are colored in palegreen, lime, forest green, and magenta; StBax1 are colored in cyan with the truncated C-terminal domain in white.

(B) Sequence of the bubble-6 DNA substrate used for crystallization. The two DNA strands (strand a and b) are colored in red and blue, respectively. Unwound bases in the crystal are in black and missing bases are in gray.

(C) The electron density (Fo-Fc) map for the forked DNA is contoured at 2σ level in gray.

(D) Left: Orthogonal views of the StXPB-Bax1$^{\Delta C}$-forked DNA complex structure in cylindrical representation. The red arrow indicates the direction for StXPB to move along the red DNA strand for dsDNA unwinding. Right: the StXPB-Bax1$^{\Delta C}$-forked DNA complex structure with both proteins presented in surfaces and DNA in ribbons. Protein domains/motifs are colored as in (A).

Interactions between the XPB-Bax1$^{\Delta C}$ heterodimer and the forked DNA

Close examination of the interface between the StXPB-Bax1$^{\Delta C}$ heterodimer and the forked DNA reveals how XPB and Bax1 interact with DNA at the ds-ss DNA junction. XPB makes extensive contacts to dsDNA (base pairs 6–12) immediately adjacent to the junction, the first mismatching base pair $C13_a$-$C13_b$, and the next two unpaired nucleotides $C14_a$ and T15a on the 3'-overhang (Fig. 3.5). The interactions of XPB with the ds-ss DNA junction region are mainly mediated by residues from the RED and ThM motifs (Fig. 3.5A), two unique and important motifs among XPB homologues. The ThM motif grips the 3'-overhang like a claw (Fig. 3.5B). Residues N274, L275, F278, H279, V282, L295 intrude between the two ssDNA tails and interact with $C13_a$-$C13_b$, $C14_a$ and $T15_a$, and the aromatic side chain of residue F278 approaches and stacks with the mismatched $C13_a$-$C13_b$ (Fig. 3.5A and 3.6A), very similar to the F633 [116] or Y621 [117] at the separation pin of UvrD. This Phenylalanine residue of the ThM motif is also highly conserved among different archaeal species (Fig. 3.7). Residue R205 (of the RED motif) forms hydrogen bonds with the phosphate of nucleotide $A11_a$ and residue D206 (of the RED motif) stabilizes the unpaired base of $C14_a$ while residue D207 (of the RED motif) interacts with the mismatched base of $C13_a$ (Fig. 3.5C and 3.6C). The side chains of R258 and W298 (of the ThM motif) interact with the phosphate backbone of $T15_a$ and W298 also interacts with the phosphate backbone of nucleotide $C14_a$ (Fig. 3.5B and 3.6B).

The DNA duplex immediately adjacent to the fork sits in the upper section of the groove formed between the two RecA-like motor (HD1, HD2) domains (Fig. 3.4D). The bottom of the same groove is the site for ATP binding and hydrolysis. Therefore,

conformational changes induced by ATP binding and hydrolysis likely push StXPB to move along the dsDNA. When XPB translocates along the dsDNA ahead of the fork, the ThM motif grips the 3' overhang tail and the tip of the ThM motif, particularly residue F278, functions as a wedge to break the base pairs along the way. Collectively, these interactions allow StXPB to function as a dsDNA translocase with 3'-5' helicase activity. In the ternary complex, $Bax1^{\Delta C}$ interacts with the unpaired 5'-overhang nucleotides $C13_b$, $C15_b$, $T16_b$ and stablizes the strand separation, likely enhancing the DNA unwinding by XPB. In addition, $Bax1^{\Delta C}$ has some contacts with the dsDNA (base pairs 3–8, 10, 12) next to XPB (Fig. 3.5A) and extends the protein-dsDNA interactions, possibly increasing the processivity of DNA unwinding by XPB. However, the nuclease domain of $Bax1^{\Delta C}$ does not interact with DNA at all, suggesting that the nuclease activity is inhibited when the repair bubble is being created and extended by XPB helicase during DNA repair. This is in good agreement with the previous study showing that XPB inhibits the endonuclease activity of Bax1 [94].

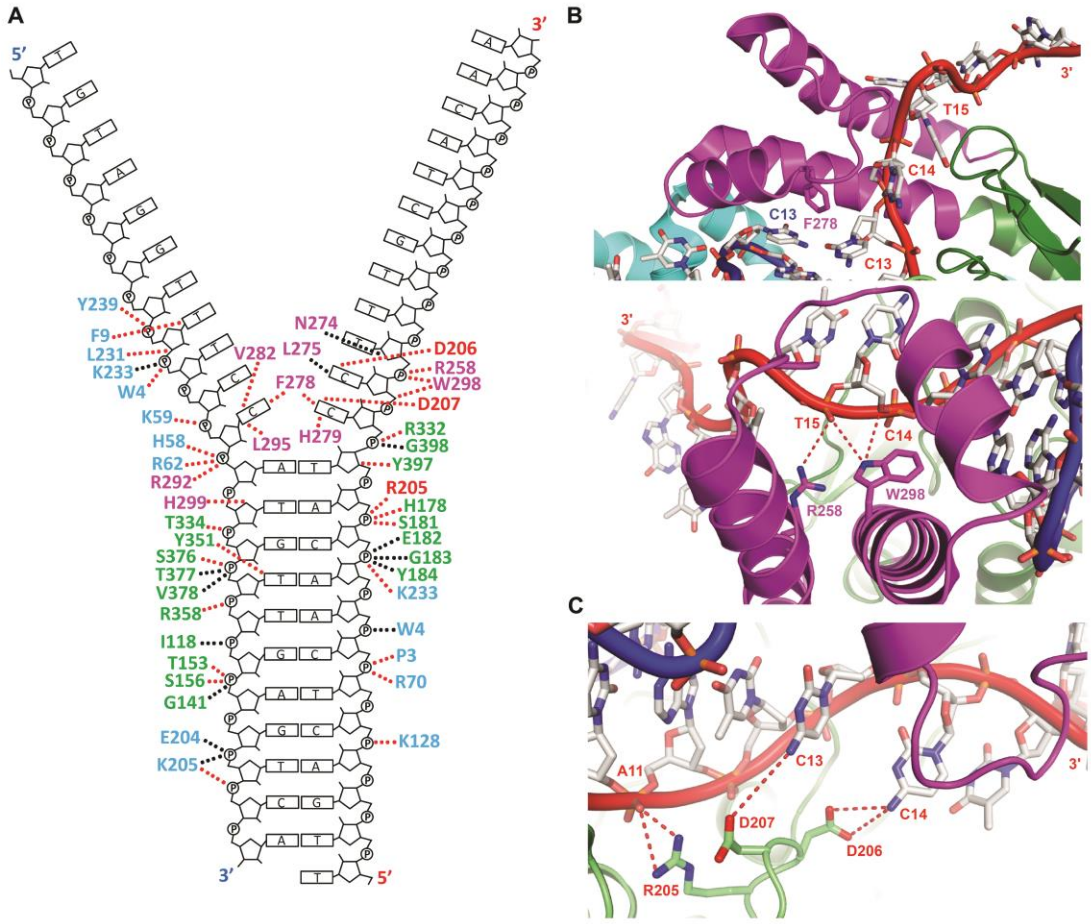

Figure 3.5 Interactions of StXPB-Bax1 with the forked DNA substrate.

(A) Diagram of DNA-protein interactions. Residues directly contacting DNA are shown with interactions to DNA highlighted by dashed lines: red lines for interactions from residue side-chain and black lines for interactions from peptide backbone. Residues from Bax1 are in cyan, residues from the HD1/HD2 and the ThM of XPB are in green and magenta, respectively. The RED motif residues (R205, D206 and D207) are highlighted by red labels.

(B) The ThM motif intrudes between the two ssDNA arms (DNA backbones are in red and blue ribbons, respectively) and grips the 3'-overhang (red ribbon) with residue F278 stacking with $C13_a$-$C13_b$ (Top) and residue R258 interacting with $T15_a$ while W298 interacting with both $C14_a$ and $T15_a$ (bottom).

(C) The RED motif interacts with the junction with residue R205 forming hydrogen bonds with $A11_a$ and residues D206 and D207 stabilizing the unpaired bases of $C14_a$ and $C13_a$, respectively. DNA and protein backbones are displayed as ribbons with the same colors as in Fig. 3.4D. Nucleotides and key amino acid residues are shown in sticks with oxygen atoms in red and nitrogen atoms in blue.

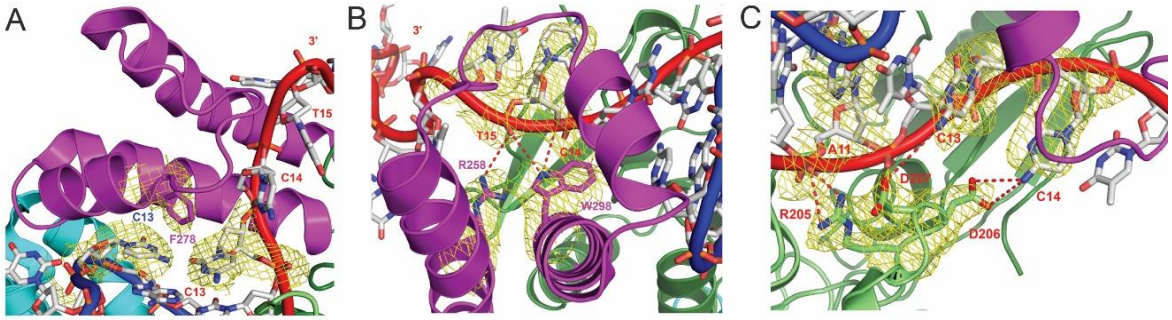

Figure 3.6 Electron density (Fo-Fc) map of the key residues in Figure 3.5.

(A**)** The same view as in Figure 3.5B (upper). The electron density (Fo-Fc) map for residue F278 of StXPB and C13a, C13b of the forked DNA is contoured at 2σ level in yellow.

(B**)** The same view as in Figure 3.5B (lower). The electron density (Fo-Fc) map for residues R258 and W298 of StXPB as well as bases C14a and T15a of the forked DNA is contoured at 2σ level in yellow.

(C**)** The same view as in Figure 3.5C. The electron density (Fo-Fc) map for the RED motif residues R205, D206, D207 and bases A11a, C14a, C13a of the forked DNA is contoured at 2σ level in yellow.

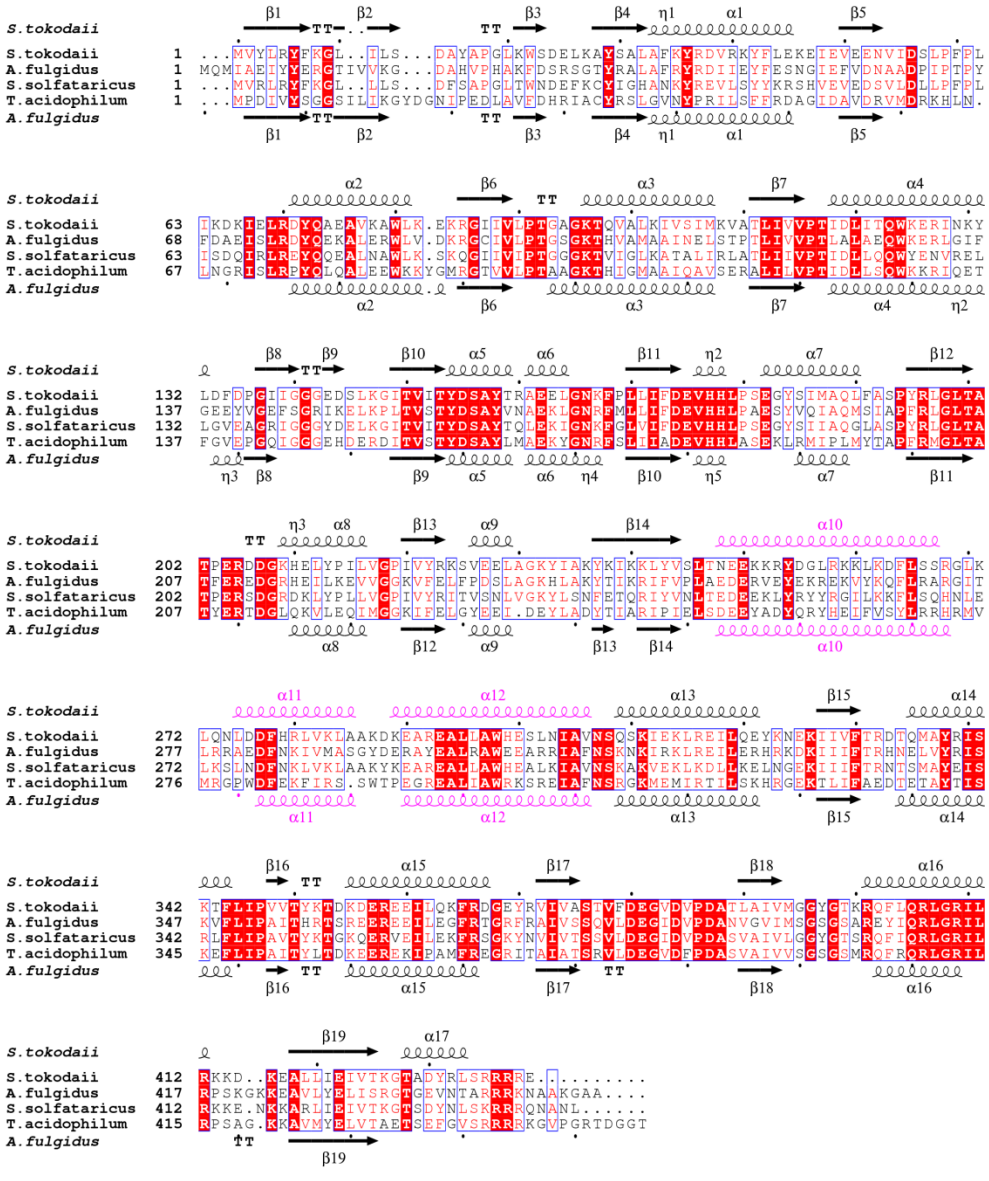
S.tokodaii
S.tokodaii 1 ...MVYLRYFKGL..ILS...DAYAPGLKWSDELKAYSALAFKYRDVRKYFLEKEIEVEENVIDSLPFPL
A.fulgidus 1 MQMIAEIYYERGTIVVKG...DAHVPHAKFDSRSGTYRALAFRYRDIIEYFESNGIEFVDNAADPIPTPY
S.solfataricus 1 ...MVRLRYFKGL..LLS...DFSAPGLTWNDEFKGYIGHANKYREVLSYYKRSHVEVEDSVLDLLPFPL
T.acidophilum 1 ...MPDIVYSGGSILIKGYDGNIPEDLAVFDHRIACYRSLGVNYPRILSFFRDAGIDAVDRVMDRKHLN.
A.fulgidus
S.tokodaii 63 IKDKEELRDYQAEAVKAWLK.EKRGIIVLPTGAGKTQVALKIVSIMKVATLIVVPTIDLITQWKERINKY
A.fulgidus 68 FDAEISLRDYQEKALERWLV.DKRGCIVLPTGSGKTHVAMAAINELSTPTLIVVPTLALAEQWKERLGIF
S.solfataricus 63 ISDQIRLREYQQEALNAWLK.SKQGIIVIPTGGGKTVIGLKATALIRLATLIIVPTIDLLQQWYENVREL
T.acidophilum 67 LNGRISLRPYQLQALEEWKKYGMRGTVVLPTAAGKTHIGMAAIQAVSERALILVPTIDLLSQWKKRIQET
A.fulgidus
S.tokodaii 132 LDFDPGIIGGGEDSLKGITVITYDSAYTRAEELGNKFPLLIFDEVHHLPSEGYSIMAQLFASPYRLGLTA
A.fulgidus 137 GEEYVGEFSGRIKELKPLTVSTYDSAYVNAEKLGNRFMLLIFDEVHHLPAESYVQIAQMSIAPFRLGLTA
S.solfataricus 132 LGVEAGRIGGGYDELKGITVITYDSAYTQLEKIGNKFGLVIFDEVHHLPSEGYSIIAQGLASPYRLGLTA
T.acidophilum 137 FGVEPGQIGGGEHDERDITVSTYDSAYLMAEKYGNRFSLHIADEVHHLASEKLRMIPLMYTAPERMGLTA
A.fulgidus
S.tokodaii 202 TPERDDGKHELYPILVGPIVYRKSVEELAGKYIAKVKIKKLYVSLTNEEKKRYDGLRKKLKDFLSSRGLK
A.fulgidus 207 TFERDDGRHEILKEVVGGKVFELFPDSLAGKHLAKYTIKRIFVPLAEDERVEYEKREKVYKQFLRARGIT
S.solfataricus 202 TPERSDGRDKLYPLLVGPIVYRITVSNLVGKYLSNFETQRIYVNLTEDEEKLYRYYRGILKKFLSQHNLE
T.acidophilum 207 TYERTDGLQKVLEQIMGGKIFELGYEEI.DEYLADYTIARIPIELSDEEYADYQRYHEIFVSYLRRHRMV
A.fulgidus
S.tokodaii 272 LQNLDDFHRLVKLAAKDKEAREALLAWHESLNIAVNSQSKIEKLREILQEYKNEKIIVFTRDTQMAYRIS
A.fulgidus 277 LRRAEDFNKIVMASGYDERAYEALRAWEEARRIAFNSKNKIRKLREILERHRKDKIIFTRHNELVYRIS
S.solfataricus 272 LKSLNDFNKLVKLAAKYKEAREALLAWHEALKIAVNSKAKVEKLKDLLKELNGEKIIFTRNTSMAYEIS
T.acidophilum 276 MRGPWDFEKFIRS.SWTPEGREALLAWRKSREIAFNSRGKMEMIRTILSKHRGEKTLIFAEDTETAYTIS
A.fulgidus
S.tokodaii 342 KTFLIPVVTYKTDKDEREEILQKFRDGEYRVIVASTVFDEGVDVPDATLAIVMGGYGTKRQFLQRLGRIL
A.fulgidus 347 KVFLIPAITHRTSREEREEILEGFRTGRFRAIVSSQVLDEGIDVPDANVGVIMSGSGSAREYIQRLGRIL
S.solfataricus 342 RLFLIPAVTYKTGKQERVEILEKFRSGKYNVIVTSSVLDEGIDVPDASVAIVLGGYGTSRQFIQRLGRIL
T.acidophilum 345 KEFLIPAITYLTDKEERERKIPAMFREGRITAIATSRVLDEGVDFPDASVAIVVSGSGSMREQFRQRLGRIL
A.fulgidus
S.tokodaii 412 RKKD..KEALLIEIVTKGTADYRLSRRRRE.........
A.fulgidus 417 RPSKGKKEAVLYELISRGTGEVNTARRKNAAKGAA....
S.solfataricus 412 RKKE.NKKARLIEIVTKGTSDYNLSKRRRQNANL......
T.acidophilum 415 RPSAG.KKAVMYELVTAETSEFGVSRRRRKGVPGRTDGGT

Figure 3.7 Sequence and structural alignments of XPB orthologues in archaea.
Amino acid sequences of XPB orthologues from *Sulfurisphaera tokodaii*, *Archaeoglobus fulgidus*, *Saccharolobus solfataricus*, *Thermoplasma acidophilum*, are aligned with a 0.5 threshold for similarity. Alignment was performed with Clustal Omega [102] and depicted using ESPript 3.0 server [103]. Secondary structure elements for *S.tokodaii* (top) and *A. fulgidus* (bottom) XPB proteins are numbered and represented according to StXPB-Bax1 (PDB entry 6P4O) and AfXPB-Bax1 (PDB entry 6P66) structures. Secondary structure elements for the ThM motifs of XPB are colored in magenta. *S. tokodaii* and *S. solfataricus* belong to the phylum Crenarchaeota while *Archaeoglobus fulgidus* and *Thermoplasma acidophilum* belong to the phylum Euryarchaeota.

Human XPB has a shortened ThM motif and forms a stable complex with p52/p8

With the solved StXPB-Bax1$^{\Delta C}$-DNA ternary structure, we then superimposed human XPB from the cryo-EM structure of TFIIH-XPA-DNA [63] on StXPB to see if human XPB could bind to the forked DNA in a similar fashion (Fig. 3.8A). Strikingly, StXPB and human XPB are in almost identical conformations, indicating the core motor domains of XPB are highly conserved in dsDNA binding and translocation (Fig. 3.8A). However, the human XPB has an apparently shorter ThM motif (Fig. 3.9), which may prevent it from clamping the ds-ss DNA junction like archaeal XPB. To further confirm whether human XPB could strongly bind forked DNA, human XPB protein is needed to be isolated from TFIIH for biochemical studies. Our extensive efforts to express human XPB in *E.coli* or insect cell did not succeed, presumably because human XPB alone has folding problems during expression or extremely low solubility. I speculated that protein-protein interactions in TFIIH may contribute to its expression and stabilizing XPB in solution. Since p52 has been reported to interact with XPB, thus I started co-expressing p52 with XPB. When full length XPB was co-expressed with different p52 constructs in *E.coli* or insect cell, no soluble complex was detected. Then we realized p52 itself may also not be soluble and added p8 in this co-expression, which interacts with the C terminal of p52 [18, 30]. This strategy finally led to the successful expression and purification of XPB/p52/p8 from insect cells (Fig. 3.8B), allowing us to study the activities of XPB in vitro. Therefore, we generated several StXPB mutants that focus on the RED and ThM motifs to examine our structural analysis and provide more insights into how StXPB-Bax1 recognizes the ds-ss junction around the lesion site.

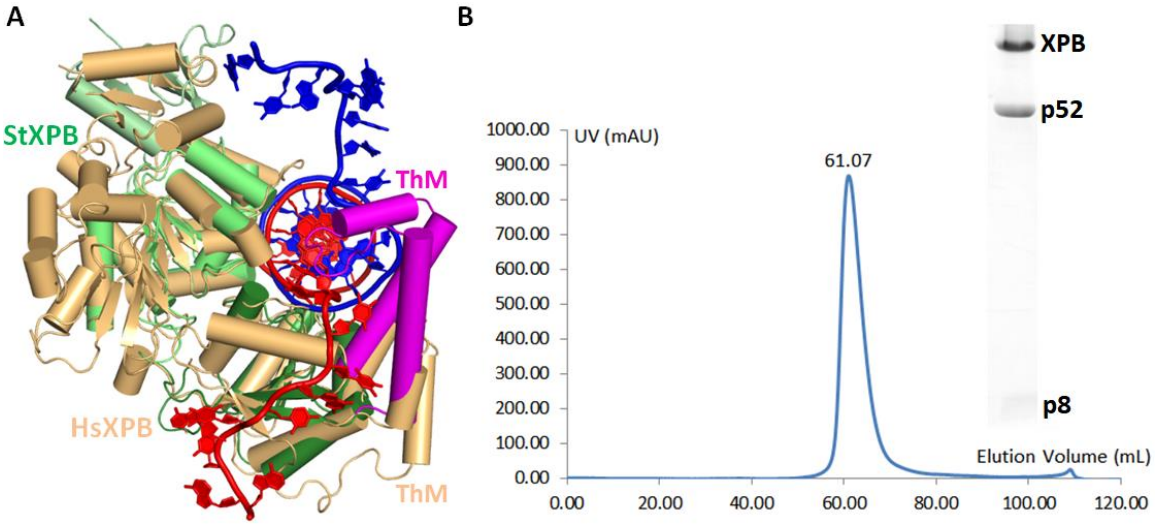

Figure 3.8 Human XPB superimposed onto StXPB-forked DNA and purification of XPB/p52/p8 from insect cell.

(A**)** Superimposition of StXPB and DNA in the StXPB-Bax1$^{\Delta C}$-DNA structure with human XPB in the core TFIIH–XPA–DNA cryo-EM structure (PDB entry: 6RO4). Human XPB is shown in light orange. StXPB and the forked DNA are colored as in Figure 3.4D.

(B**)** Left: Gel-filtration chromatography profiles of the full length XPB/p52/p8 complex. The elution position is indicated. Right: SDS PAGE gel shows the purity of XPB/p52/p8. Samples are from the fractions eluted from the gel-filtration chromatography.

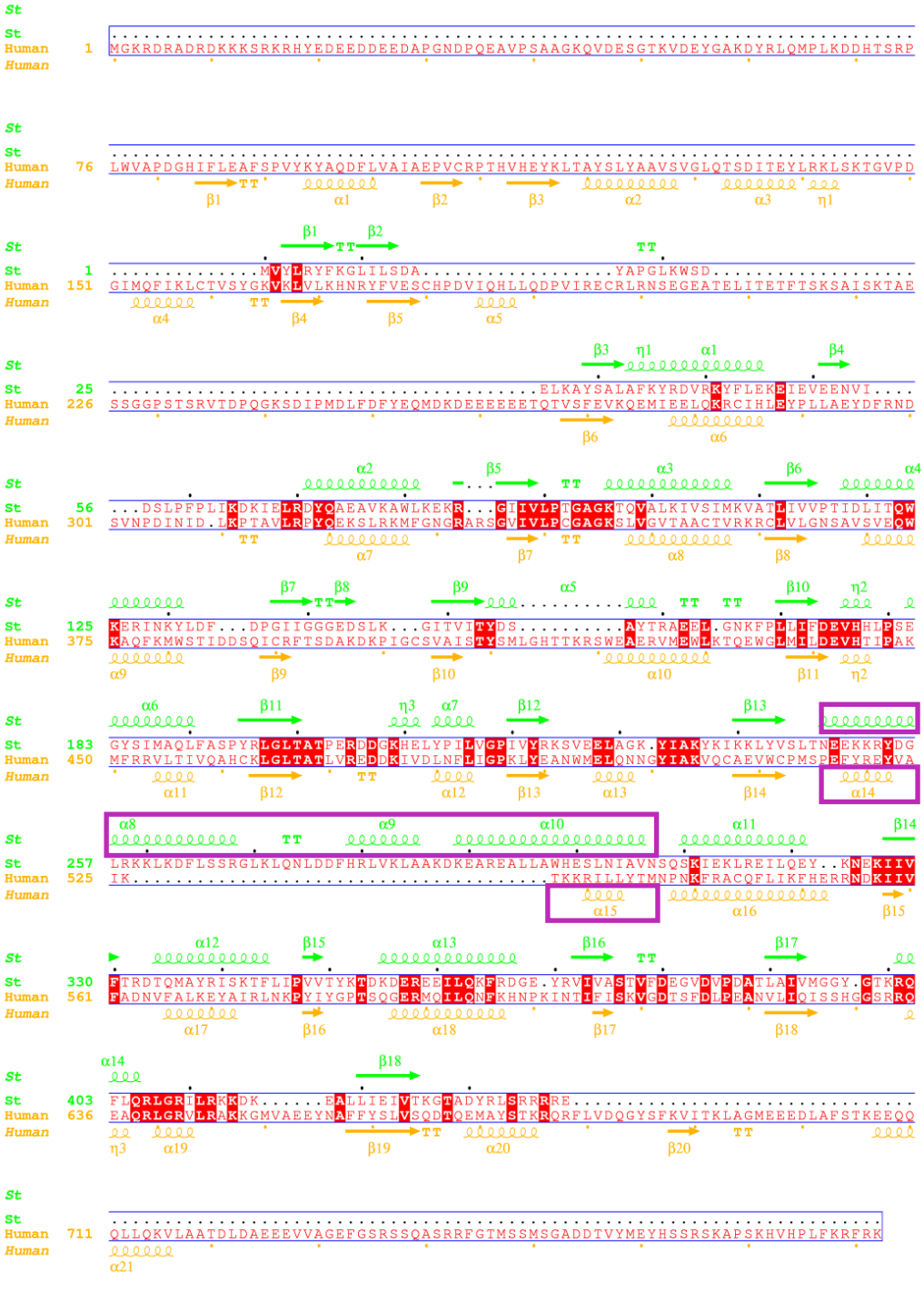

Figure 3.9 Sequence and structural alignment of StXPB and human XPB.

Alignment was performed with PROMALS3D [118] and depicted using ESPript 3.0 server. Secondary structure elements for StXPB (top) and human XPB (bottom) are numbered and represented according to the PDB files of the StXPB-Bax1$^{\Delta C}$-DNA (PDB entry: 6P4F) and core TFIIH-XPA-DNA (PDB entry: 6RO4) structures. The secondary elements of the ThM motifs for both human XPB and StXPB are highlighted by magenta boxes.

The ThM motif of StXPB is most essential for DNA junction binding

To better understand how the two conserved motifs contribute to the activities of StXPB, variants of StXPB containing substitutional mutations R205A/D206A/D207A in the RED motif or deletion of residues 270 to 280 (ΔThM1) and residues 258 to 299 (ΔThM2) in the ThM motif were designed and prepared alone or in complex with Bax1 (Fig. 3.10C). The StXPB ΔThM2 mutant was expected to mimic the length of human XPB ThM motif (Fig. 3.9). The effects of these mutations on DNA binding of the StXPB-Bax1 complex or StXPB alone were analyzed (Fig. 3.10). Results from the electrophoretic mobility shift assay (EMSA) assay revealed that Bax1 enhances the affinity of StXPB binding to the forked DNA (comparing lane 2-3 in Fig. 3.10A to lane 2-3 in Fig. 3.10B). As indicated by the ternary complex structure, mutations of the RED or ThM motif could disrupt the interactions of the heterodimer or StXPB with the forked DNA. Substitutions of three charged residues in the RED motif with alanine reduced the affinity of StXPB or the StXPB-Bax1 complex with forked DNA substrate (compare lane 4 with lane 2 in Fig. 3.10A-B) while deletion (ΔThM1) of the tip of the ThM motif reduced the affinity even further (lane 6 in Fig. 3.10A-B). Furthermore, the heterodimer containing the deletion mutant ΔThM2 has the lowest DNA binding affinity and forms unstable protein-DNA complexes (lane 8-9 in Fig. 3.10A). This ΔThM2 mutant for StXPB alone could not be obtained because the protein was very unstable and suffered from severe degradation during purification. Therefore, it could only be purified with its partner Bax1 that somehow stabilizes this XPB mutant by protein-protein interactions.

XPA helps human XPB hook onto the forked DNA

For comparison with human XPB, we also analyzed the DNA-binding affinity of human XPB expressed in insect cell culture by baculovirus expression system together with p52 and p8 (as mentioned above, human XPB could not be obtained when being expressed alone). Remarkably, human XPB (p52/p8) formed even weaker and unstable complexes with the forked DNA substrate (the smear bands in lane 10-11 in Fig. 3.10A and lane 8-10 in Fig. 3.10B). This observation suggests that human XPB does not have the capability like StXPB to directly recognize the forked DNA or the DNA junction. However, XPA helps human XPB (p52/p8) to form a stable complex with the forked DNA substrate as shown by the distinct supershift (lane 14-15 in Fig. 3.10A and lane 11 in Fig. 3.10B), consistent with the recent cryo-EM structure showing XPA-XPB forms a DNA duplex tunnel and XPA helps clamp TFIIH to DNA [63]. Interestingly, XPA itself also forms unstable complexes with the forked DNA substrate as shown by the smear bands (lane 12-13 in Fig. 3.10A). These results together suggest that XPA hooks human XPB at the fork of the DNA repair bubble and may play a critical role in lesion verification following the damage recognition by XPC in human NER. Consequently, XPB and XPA are likely to work together for the DNA opening around the lesion during the initial stages of repair bubble formation, generating the 3'-overhang for XPD binding and induce conformational changes of TFIIH [63]. The longer ThM motif of archaeal XPB (compared to human XPB) may replace the need for XPA in archaeal NER since no XPA homologs have been identified so far in archaea.

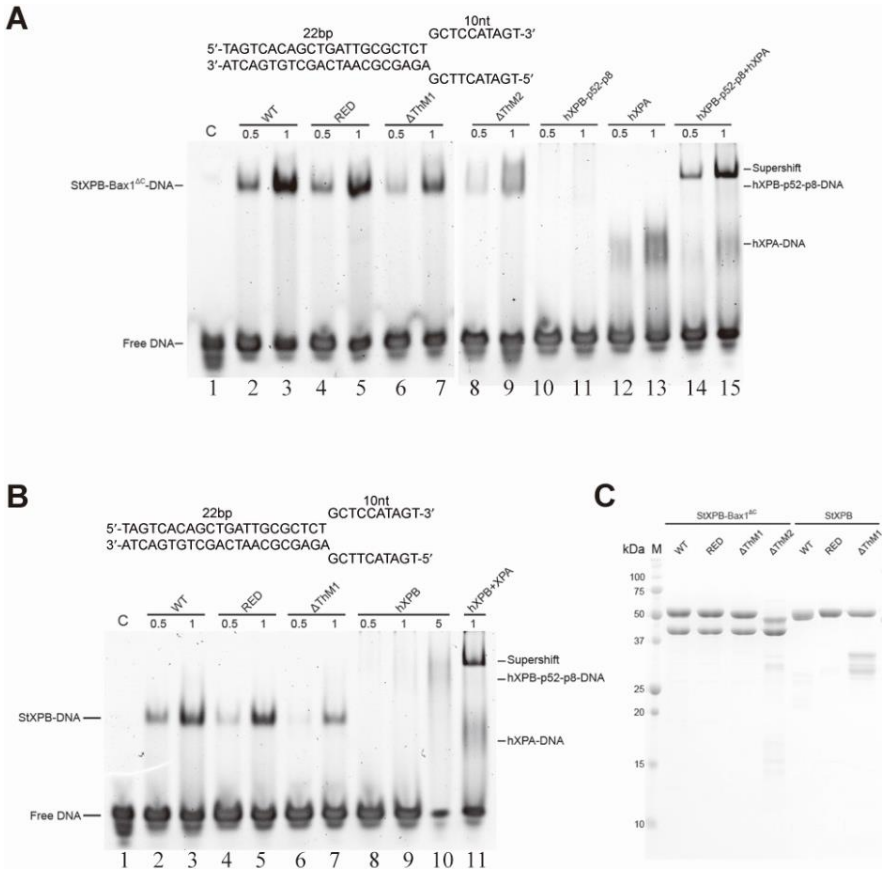

Figure 3.10 The RED and ThM motifs of XPB are important for forked DNA binding and comparison to human XPB-XPA.

EMSA analysis on the interactions of a forked DNA substrate (see Methods) with increasing concentrations of the StXPB-Bax1$^{\Delta C}$ complex containing StXPB variants (A), and StXPB variants compared with human XPB/p52/p8 and XPA (B). The molar ratios of protein to DNA are indicated one top of each gel. Lane numbers are marked at the bottom of the gels. Sequences of DNA substrates used are shown above the gels. Each EMSA gel is a representative of the same EMSA experiment repeated at least twice. (C) SDS-PAGE results of protein samples used for EMSA. WT: the StXPB-Bax1$^{\Delta C}$ heterodimer or StXPB, RED: StXPB mutant with R205A/D206A/D207A substitutions, ΔThM1: StXPB mutant with deletion of residues 270 to 280, ΔThM2: StXPB mutant with deletion of residues 258 to 299. hXPB: human XPB/p52/p8 complex, hXPA: human XPA.

StXPB has enhanced affinity for dsDNA with a small mismatched bubble

A typical NER DNA damage usually induces local melting of DNA. The structure of the StXPB-Bax1$^{\Delta C}$-DNA complex shows the XPB ThM motif clamps into the DNA fork via interactions with the first 3 mismatched DNA nucleotides next to the ds-ss junction. To further test if the ThM motif could enhance the interactions of StXPB with the DNA distortions caused by NER lesions, we applied EMSA again to compare the affinities of StXPB binding to normal dsDNA and dsDNA substrates with mismatched bubbles ranging from 2-nt to 6-nt (Fig. 3.11A). StXPB formed weak and unstable complexes with a 32-bp dsDNA substrate (smear in lane 2) and the bubble-2 (dsDNA with a 2-nt mismatched bubble) substrate (smear in lane 4). The interactions of StXPB with bubble DNA substrates increased when the size of the mismatched bubble increases from 2-nt to 5-nt (Fig. 3.11A) and a distinguished band of the StXPB-DNA complex indicates that StXPB formed a stable complex with the bubble-4, bubble-5 and bubble-6 substrate (lane 8, 10, and 12 in Fig. 3.11A). This observation suggests that StXPB is capable of binding to helix-distortions like small mismatch bubbles, possibly reflecting a key step of repair bubble formation in archaeal NER.

At high StXPB:DNA ratio, StXPB even formed a stable complex with the bubble-3 substrate (lane 4 in Fig. 3.11B). Substitution of the RED motif with alanine residues (AAA) significantly reduced the interactions of StXPB with the bubble-3 (lane 5-7 in Fig. 3.11B) and bubble-5 substrates (lane 4-5 in Fig. 3.11C). Deletion of the tip of the ThM motif (ΔThM1) almost eliminated the interactions of StXPB with the bubble-3 (lane 8-10 in Fig. 3.11B) and bubble-5 substrates (lane 8-9 in Fig. 3.11C). These results

indicate that both the RED and ThM motifs are important for StXPB binding to distorted DNA, possibly playing a role in archaeal DNA damage recognition. To our surprise, substitution of F278 with Ala (F278A) did not show noticeable effects on the interaction of StXPB with the bubble-5 substrate (lane 6-7 in Fig. 3.11C). This observation indicates that F278 is not critically involved in binding to the DNA fork, while it is more likely to play a key role in DNA unwinding by XPB since it interacts with the first mismatched base pairs in the structure and is highly conserved across different archaeal species (Fig. 3.7).

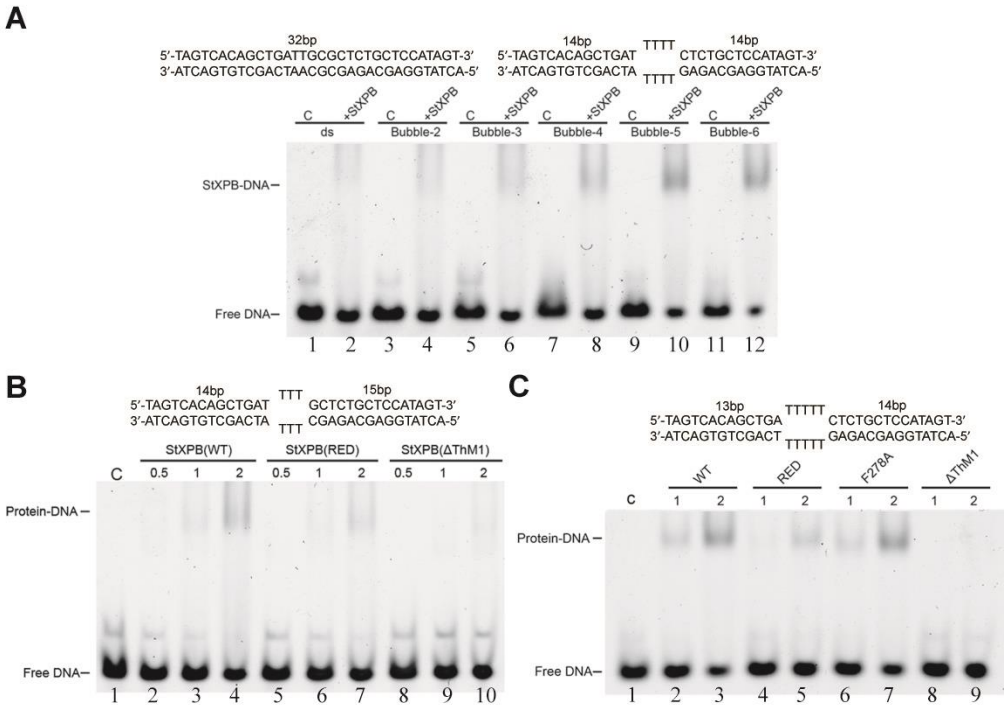

Figure 3.11 The RED and ThM motif are important for XPB binding to dsDNA with mismatched bubbles.

(A) EMSA analysis on the interactions of StXPB with dsDNA substrate (sequence shown above the gel) and dsDNA with 2-nt mismatch (bubble-2), 3-nt mismatch (bubble-3), 4-nt mismatch (bubble-4, sequence shown above the gel), 5-nt mismatch (bubble-5), and 6-nt mismatch (bubble-6).

(B) EMSA analysis on the interactions of StXPB variants with bubble-3 DNA.

(C) EMSA analysis on the interactions of StXPB variants with bubble-5 DNA. In all panels, C: control reaction of the DNA substrate alone. The molar ratios of protein to DNA are indicated on the top of the gels. Lane numbers are marked at the bottom of the gels. Sequences of DNA substrates used are shown above the gels. Each EMSA gel is a representative of the same EMSA experiment repeated at least twice.

Archaeal XPB could play a role in damage recognition

When we replaced the forked DNA in our structure with the CPD lesion-containing DNA from the structure of yeast XPC-HR23B-DNA complex [119], we found the dsDNA regions of the two DNA molecules were aligned very well (Fig. 3.12A). Thus the CPD DNA could fit nicely in the structure of the StXPB-Bax1 heterodimer, likely reflecting the initial binding of XPB to the damage site (Fig. 3.12B). In eukaryotic NER, the XPC-HR23B complex firstly recognizes the lesion site. The β-hairpin of XPC that inserts into the double helix and flips out two base pairs (on the opposite strand of the damage) is very similar to the ThM tip of archaeal XPB which also intrudes between two strands of the forked DNA. Since there are no XPC homologs existing in archaea, it is possible that archaeal XPB may also play a role in damage recognition. Upon initial damaged DNA binding, XPB holds the dsDNA between the two RecA-like domains (HD1, HD2) with the ThM motif clamping at the lesion site. When the ThM motif clamps down, the tip of the ThM motif fits well into the void space created by the CPD and flipping out of the two bases on the other DNA strand (Fig. 3.12C), leading to the enhanced affinity of XPB binding to UV-damaged DNA over normal DNA, which would prevent the ThM motif from clamping down without melting the dsDNA. This speculation is consistent with our EMSA results showing that StXPB forms a weak and unstable complex with dsDNA substrate but forms a stable complex with substrates containing a small bubble (Fig. 3.11). It also explains why disruption of the ThM tip (ΔThM1 mutant) almost abolished the interactions between StXPB and the 3-nt or 5-nt bubble DNA.

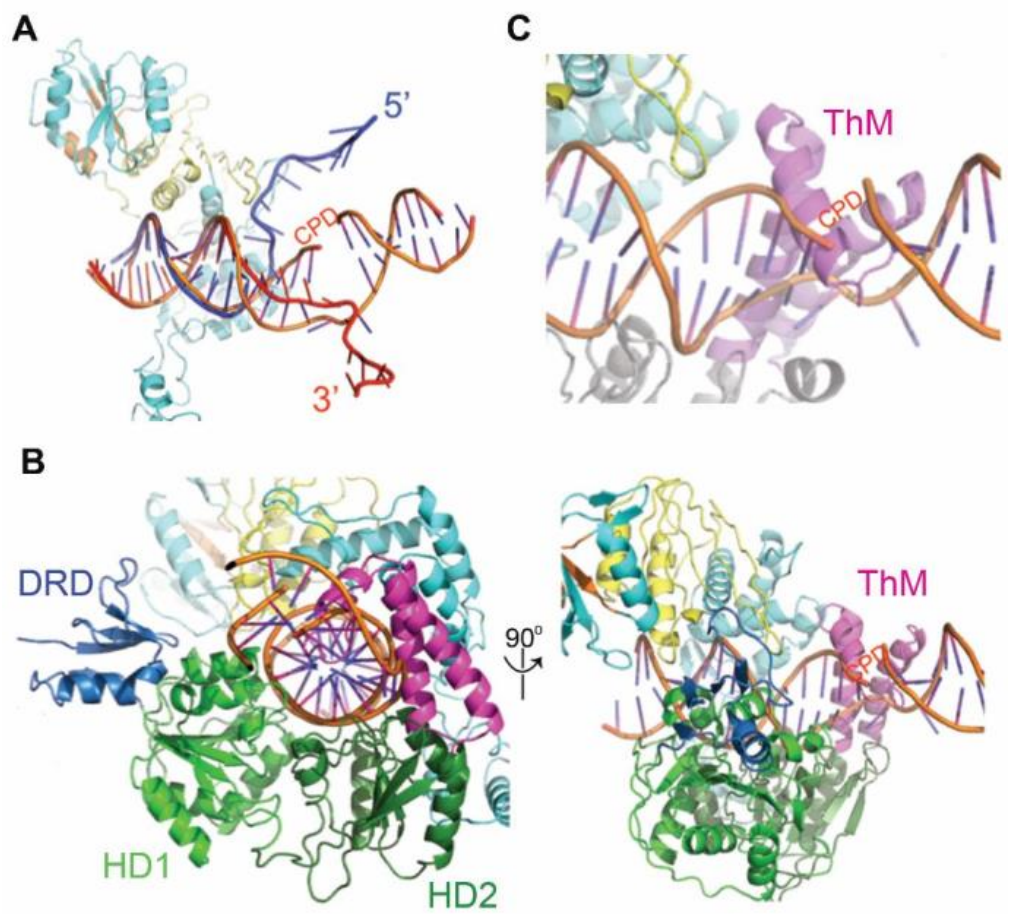

Figure 3.12 Model of StXPB binding to the damage site.

(A) CPD-containing DNA (orange) from the Rad4-Rad23-DNA complex (PDB entry: 2QSG) [119] is superimposed with the forked DNA in the ternary complex over the dsDNA region. Bax1 is shown in cyan ribbons with the Cas2-like (CRD) domain highlighted in yellow and the nuclease motifs in red.

(B) Two views of the ternary complex with the forked DNA replaced by the CPD-containing DNA from (A). Domains of StXPB are colored differently.

(C) Zoom-in view of the ThM tip clamping into the void space created by the lesion CPD.

StXPB in the DNA-free StXPB-Bax1$^{\Delta C}$ structure is in the ATP-bound state

In order to identify protein conformational changes induced by DNA binding, we also needed to solve the crystal structure of the StXPB-Bax1$^{\Delta C}$ heterodimer in the DNA-free state. With extensive crystallization screening and refinement efforts, crystals of StXPB-Bax1$^{\Delta C}$ were obtained but these crystals only diffracted to 4.5 Å resolution at maximum. I accidentally found introducing some synthesized DNA oligos (ssDNA or Y-shaped DNA) in the drop could further improve the crystal quality (Fig. 3.13A). I think the reason is that these DNA oligos are not ideal substrates for StXPB-Bax1$^{\Delta C}$ binding but could somehow induce StXPB-Bax1$^{\Delta C}$ to maintain in certain conformation ready for DNA binding. Finally the crystal structure of the StXPB-Bax1$^{\Delta C}$ heterodimer was determined the at 2.96 Å resolution (see Table 2 for statistics of data collection and structure refinement), in which the StXPB-Bax1$^{\Delta C}$ heterodimer overall has similar conformation as in the full length StXPB-Bax1 structure (PDB ID: 6P4O, Fig. 3.13B). Strikingly, the ATP binding site of XPB contains a bound phosphate ion (Fig. 3.13C) and the position of this phosphate ion is similar to that of the β-phosphate group of the ADP in the ADP-bound UvrB [120] (PDB entry: 2D7D, Fig. 3.13D), an SF2 DNA helicase involved in bacterial NER. Therefore, this heterodimeric structure likely reflects StXPB in the ATP or (ADP+Pi)-bound conformation while the ternary structure presents StXPB in the ATP-free conformation. It is worth mentioning that we also tried to determine structures of StXPB, StXPB-Bax1$^{\Delta C}$ or StXPB-Bax1in complex with ADP or ATP analogs, but we did not ever succeed.

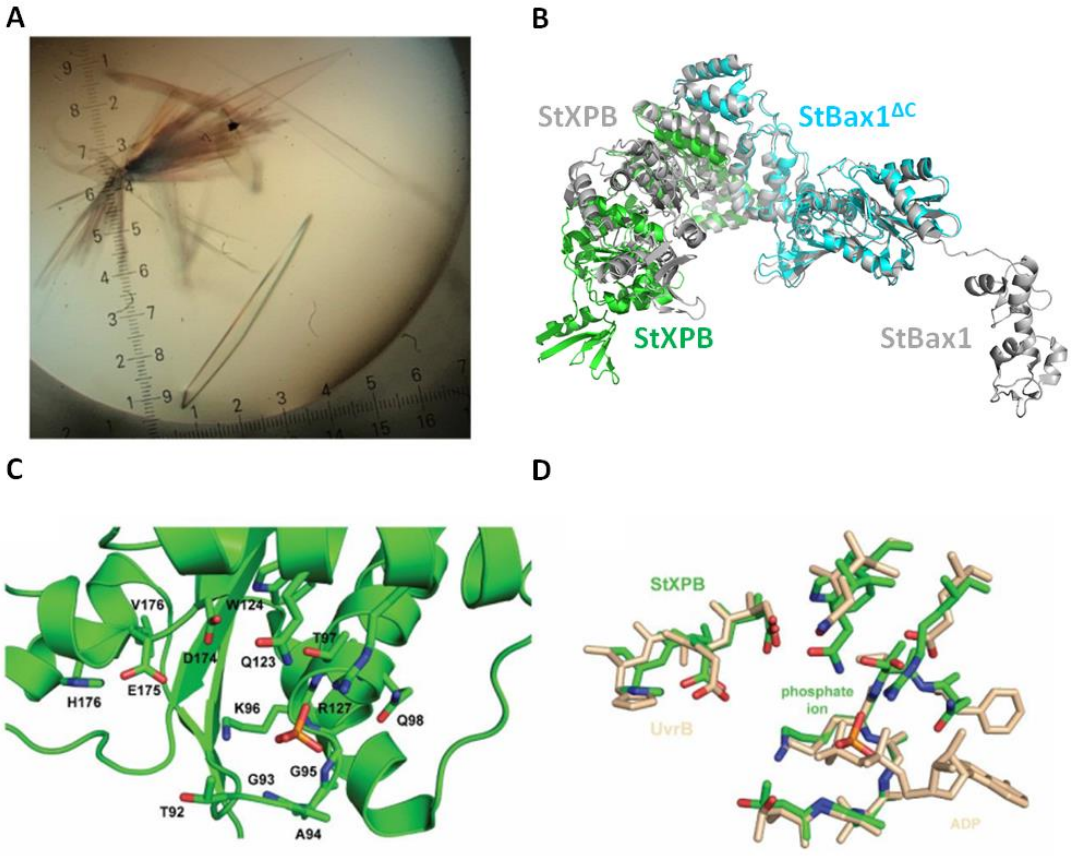

Figure 3.13 Crystallization and structure of the DNA-free StXPB-Bax1$^{\Delta C}$ dimer.

(A) Example crystal of StXPB-Bax1$^{\Delta C}$ used for data collection.

(B) Structural alignment of full length StXPB-Bax1 (PDB ID: 6P4O) with StXPB-Bax1$^{\Delta C}$. StXPB is colored in green with Bax1$^{\Delta C}$ colored in cyan while full length StXPB-Bax1 proteins are colored in gray.

(C) The bound phosphate ion colored in red and orange. The residues possibly involved in forming the ATP-binding pocket are shown in sticks.

(D) Overlay of key residues at the ATP-binding sites in StXPB (green) and UvrB (PDB entry: 2D7D, wheat). The bound phosphate ion in StXPB-Bax1$^{\Delta C}$ structure is colored as in (A) and the bound ADP in the UvrB structure is colored in wheat.

Table 2 Statistics for X-ray diffraction data collection and structural refinement of the DNA-free StXPB-Bax1$^{\Delta C}$ structure

Structure	StXPB:Bax1$^{\Delta C}$
PDB ID	6P4W
Data collection	
Space group	P 1
Cell dimensions: a, b, c (Å)	56.42, 101.37, 114.48
α, β, γ (°)	83.09, 81.15, 90.17
Resolution (Å)	29.28 – 2.96 (3.00 – 2.96)
R_{pim}	0.061 (0.529)
I/σI	8.9 (1.0)
Completeness (%)	99.1 (98.3)
Multiplicity	3.5 (3.3)
CC1/2 (%)	98.8 (63.2)
Refinement	
Resolution (Å)	29.28 – 2.96 (3.06-2.96)
No. reflections	48672
R_{work} / R_{free} *	18.30 / 23.34
Number of atoms	13068
Protein	12930
Ligands	42
Water	96
Ramachandran favored	95.71%
Ramachandran allowed	4.29%
Ramachandran outliers	0.00%
R.m.s.d Bond length	0.011 Å
R.m.s.d Bond angles	1.43°
Fo, Fc correlation	0.95
Anisotropy	0.040
Averaged B factor	80.22 Å^2
MolProbity score	1.86

Values in parenthesis are for the highest resolution shell. *5% data was used for R_{free}.

XPB conformational changes induced by DNA binding

When the StXPB-Bax1$^{\Delta C}$ heterodimer and the ternary complex are aligned on Bax1$^{\Delta C}$, StXPB has substantial changes in domain orientation while Bax1$^{\Delta C}$ shows only local changes in the NTD caused by the movement of HD2 of StXPB: the ThM motif clamps down to intrude between the two arms at the junctions and the HD1(and the N-terminal StXPB) rotates toward the dsDNA at the junction (Fig. 3.14A). These domain re-arrangements in StXPB could be simply explained as a sequential two-step action induced by the forked DNA substrate and ATP binding/hydrolysis for StXPB to unwind DNA at the fork (Fig. 3.14B). First, the initial DNA binding puts the forked DNA in the groove between the HD1 and HD2 of StXPB, ATP binding/hydrolysis allows the ThM motif to clamp down onto the ds-ss junction by intruding between the two ssDNA arms and gripping the 3'-overhang; this ThM movement changes the position of the HD2 since ThM is rigidly connected with HD2 (Fig. 3.14B), which pushes HD1 and DRD to rotate toward the DNA duplex in order to maintain the forked DNA in the groove between HD1 and HD2. This second rotation shifts the HD1 of StXPB about 11.5 Å from the 3'-ss tail into the duplex, equivalent to 2 bps (about 10.8 Å apart along the phosphate backbone) 3' to 5' forward movement along the 3'-overhang strand (Fig. 3.14B, insertion), suggesting XPB could unwind two base pairs of dsDNA upon ATP binding and hydrolysis.

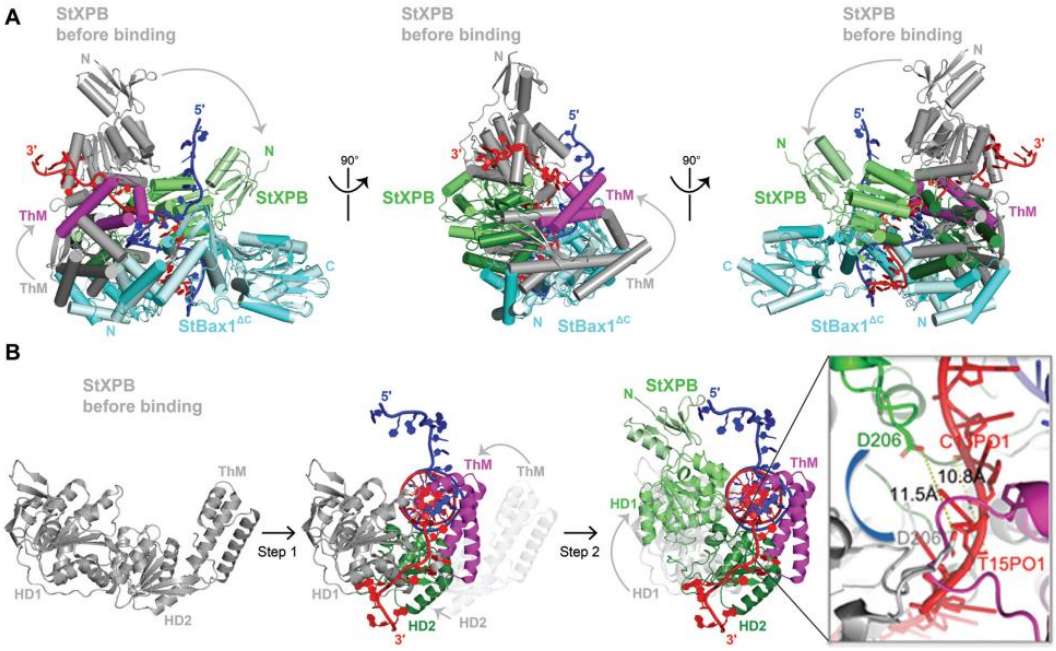

Figure 3.14 DNA interactions induce conformational changes in StXPB.

(A**)** Comparison of the StXPB-Bax1$^{\Delta C}$ heterodimer with the DNA bound ternary complex by superimposing the two StBax1$^{\Delta C}$ molecules. The StXPB-Bax1$^{\Delta C}$-DNA structure is colored as in Fig. 3.4D. In the DNA-free StXPB-Bax1$^{\Delta C}$ structure, StXPB is colored in gray and StBax1 is colored in palecyan. The arrows indicate the domain movements of StXPB from DNA-free state (gray) to the DNA-bound state (green).

(B) Structure-based mechanism for StXPB to unwind DNA in two steps. Step one: DNA-free StXPB (and Bax1, omitted for simplicity) binds to a forked DNA. DNA sits at the groove between HD1 and HD2 of StXPB to allow the ThM motif to clamp down at the fork; this ThM movement changes the position of HD2 and brings out the second step: the rotation of the HD1 of StXPB to shift HD1 two bases along the 3'-overhang strand toward the duplex. Insertion: zoom-in view on the RED motif shifting along the 3'-overhang strand. The curved blue line indicates the rotation. The shift of the RED motif from the DNA-free state (gray) to the DNA-bound state (green) is measured as 11.5 Å between the two positions of the RED motif residue D206. For comparison, the distance (10.8 Å, dash line) between two nucleotides (C13PO1 and T15PO1) is also shown.

The RED and ThM motifs are critical for the ATPase activities of StXPB

The conventional helicase assay uses gel electrophoresis to observe the displaced labeled ssDNA fragment due to the unwinding of dsDNA by a helicase. We attempted to detect DNA opening by XPB using the conventional helicase assay but it did not work out. I think the reason is that StXPB always needs to grasp 10-bp dsDNA region when it translocates on DNA duplex to catalyze the strand separation with the ThM motif. So, the DNA bubble is only locally extended by StXPB and there would never have any ssDNA fragment fully displaced from the DNA duplex. This working mode of XPB may offer a nice mechanism to ensure local and controlled DNA opening around the lesion.

Since the DNA opening is achieved through ATP binding and hydrolysis by XPB, we analyzed the roles of the two key motifs including the F278 residue in StXPB by the ATPase activity assay (Table 3). Mutations in the RED and ThM motifs including F278A significantly reduced the ATPase activity in the presence and absence of the bubble-5 DNA substrate and Bax1. These results indicate the importance of these motifs to the ATPase activity of StXPB in the order from the most important to the least important: the ThM motif (ΔThM2) > the RED motif > the ThM tip (ΔThM1) > residue F278. However, in the presence of the forked DNA substrate (Table 1), substitution of F278 with Ala has a much more severe effect (ATPase activity reduced to 47%) than the substitution of the RED motif with AAA does (ATPase activity reduced to 67%). These results together indicate that the RED and ThM motifs are critical for DNA unwinding at the fork. Combined with the EMSA results (Fig. 3.11), the impaired ATPase activities of the RED andΔThM1 mutants could be also partially caused by the reduced binding of these StXPB

mutants to the bubble-5 DNA. Also we could conclude that residue F278 is important only for DNA unwinding at the fork that is consistent with its role as the wedge to break the base pairing at the fork (Fig. 3.5).

Table 3 ATPase activities of StXPB variants in the presence and absence of Bax1 and DNA substrates

StXPB Variant	alone	+StBax1	+StBax1 +Bubble-5 DNA	+StBax1 +forked DNA
WT	4.22 ± 0.22 (100%)[a]	12.05 ± 0.45 (100%) 2.9x[b]	53.99 ± 1.81 (100%) 12.8x[b]	86.74 ± 3.81 (100%) 20.6x[b]
RED/AAA	1.27 ± 0.18 (30%)	3.87 ± 0.29 (32%)	10.95 ± 0.92 (20%)	58.08 ± 1.14 (67%)
F278A	3.01 ± 0.04 (71%)	9.94 ± 0.26 (82%)	13.78 ± 0.19 (26%)	40.81 ± 0.94 (47%)
ΔThM1	2.07 ± 0.14 (49%)	7.02 ± 0.23 (58%)	12.05 ± 0.39 (22%)	31.52 ± 1.98 (36%)
ΔThM2	N.A.	2.48 ± 0.32 (21%)	3.71 ± 0.23 (6%)	9.84 ± 0.30 (11%)

[a]: numbers in the parenthesis represent the relative ATPase activities in the same column;
[b]: ATPase activity enhancement over StXPB (WT).
The ATPase activities were obtained from at least three replicated experiments as described in the Method.
ATPase activity unit: uM ATP hydrolyzed per uM protein per minute.
RED/AAA: StXPB mutant with R205A/D206A/D207A substitutions,
ΔThM1: StXPB mutant with deletion of residues 270 to 280,
ΔThM2: StXPB mutant with deletion of residues 258 to 299.
Bubble-5 DNA:

```
5′-TAGTCACAGCTGATTTTTCTCTGCTCCATAGT-3′
3′-ATCAGTGTCGACTTTTTTGAGACGAGGTATCA-5′
```

Forked DNA:

```
5′-TAGTCACAGCTGATTGCGCTCTGCTCCATAGT-3′
3′-ATCAGTGTCGACTAACGCGAGAGCTTCATAGT-5′
```

Proposed model of ATP-driven DNA opening by StXPB

Taken together, our structural analyses and biochemical results point out the essential roles of the RED and ThM motifs in DNA binding and unwinding by StXPB and how the conformational changes of StXPB could lead to strand separation of the DNA duplex. This unconventional DNA opening mechanism allows XPB to unwind DNA while working as a dsDNA translocase. These findings inspired us to further propose an ATP-driven DNA opening model to explain how StXPB unwinds 2-bp per ATP binding and hydrolysis cycle (Fig. 3.15). StXPB in our structures of the StXPB-Bax1$^{\Delta C}$ heterodimer and StXPB-Bax1$^{\Delta C}$-DNA complex could represent the ATP-bound and ATP-free states, respectively. In the absence of DNA, StXPB only binds and hydrolyzes ATP in a relatively low rate (Table 3). If DNA substrates containing distortions or ds-ss DNA junctions are around, StXPB would bind to both DNA and ATP simultaneously, catalyzing ATP hydrolysis with much higher ATPase activities (Table 3). With the RED and ThM motifs recognizing the fork, ATP hydrolysis powers the HD2 domain and ThM motif to clamp onto the DNA junction and leads to 2 bps unwinding. Next the HD1 domain of StXPB rotates toward the DNA duplex to maintain DNA in its binding groove for translocation. Meanwhile StXPB clears its ATP-binding site and gets ready for binding new ATP. This intermediate is captured by our StXPB-Bax1$^{\Delta C}$-DNA structure. When another ATP binds, StXPB goes back to the ATP-bound conformation as shown in the DNA-free StXPB-Bax1$^{\Delta C}$ structure before another round of ATP hydrolysis and binding (Fig. 3.15).

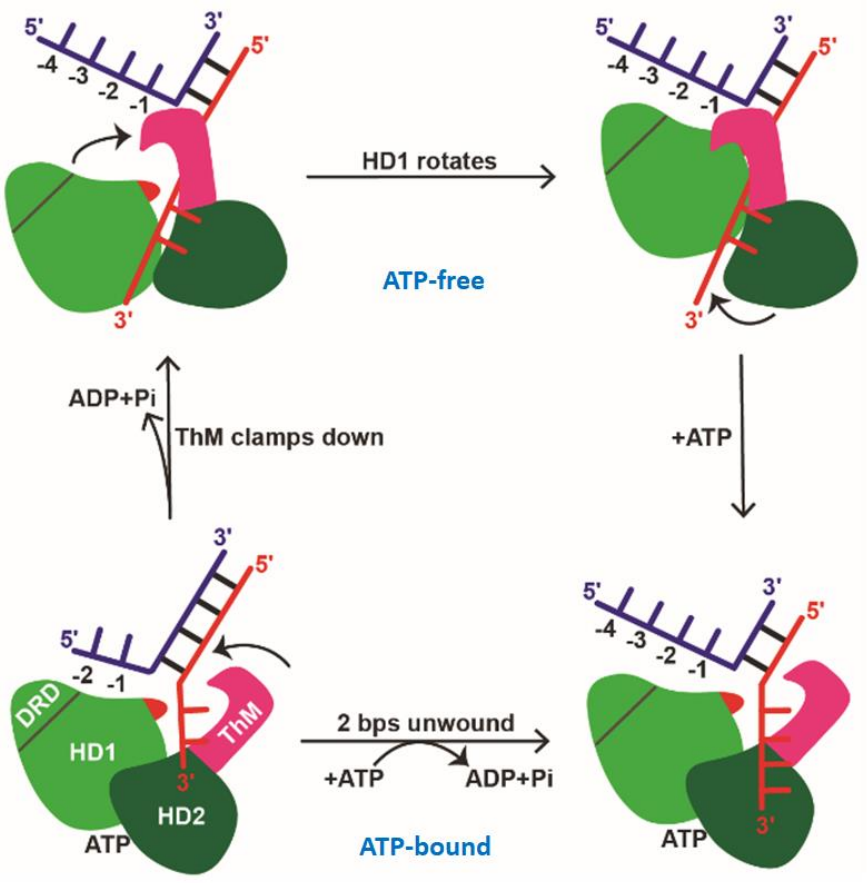

Figure 3.15 A schematic diagram of DNA opening mechanism of StXPB.

StBax1 is omitted for clarity. In the starting (lower left) and ending steps (lower right), StXPB is the ATP-bound conformation, which could be thought as the phosphate-bound DNA-free StXPB in the StXPB-Bax1$^{\Delta C}$ structure. Comparison of these two steps indicates 2 bps are unwound per ATP hydrolysis and binding. The DRD and HD1 domains are colored in green, the HD2 domain is colored in forest green, and the ThM motif is highlighted in magenta. Two strands of the DNA are colored in blue and red, respectively. The conformational changes are indicated by arrows.

Comparison to the cryo-EM structure of human TFIIH core complexed with XPA and a forked DNA

XPB is conserved from archaea to human even though there is no TFIIH-like transcription/DNA repair factor in archaea. When our ternary complex is superimposed with the TFIIH-XPA-DNA cryo-EM structure (PDB entry: 6RO4) [63], a repair intermediate in human NER, over the HD2 domains of StXPB and human XPB (Fig. 3.16A), not only are StXPB and human XPB aligned very well with both in the same closed conformation, but also the duplex regions of both DNA substrates in these two structures are surprisingly well matched (Fig. 3.16B), sitting in the upper section of the groove formed between the two RecA-like (HD1, HD2) domains, indicating StXPB and human XPB interact with dsDNA in the same way as a dsDNA translocase. However, the two forked DNA substrates in our ternary complex and the cryo-EM structure point to the opposite directions (Fig. 3.16A-B). In addition, human XPB is positioned about 5 bps away from the ds-ss junction while StXPB is right at the junction (Fig. 3.16C). These observations further confirm that human XPB is more of translocase in the context of TFIIH, which is due to the fact that human XPB has a much shorter ThM motif (Fig. 3.9) and thus cannot clamp on the forked DNA like StXPB as shown by our EMSA data that human XPB forms unstable complex with the forked DNA substrate (Fig. 3.10).

Remarkably, XPA seems to clamp on the forked DNA with a hook like the long ThM motif of StXPB (Fig. 3.16C). The hook at the fork by XPA and the interactions of XPB at the duplex DNA complement each other and therefore enhance the overall protein-DNA interactions to form a stable ternary complex of human XPB-XPA with the

forked DNA, strongly supporting our EMSA results (Fig. 3.10) and the previous observation that XPA can activate DNA unwinding by the TFIIH core [121]. Interestingly, XPA grips the 5'-ss arm instead of the 3'-ss arm, which is bound by XPD in the cryo-EM structure (Fig. 3.16D). In the StXPB-Bax1$^{\Delta C}$-DNA structure, StXPB grips the 3'-ss arm while Bax1 stabilizes the 5'-ss arm (Fig. 3.4D). Taken together, it is very likely that the ATPase and translocase activities of human XPB could power the spiral movement of XPB-XPA on the DNA fork, leading to strong mechanical torque and DNA unwinding. This hypothesis perfectly explains several remaining questions in NER including how XPB could initiate DNA opening around the lesion only with its ATP-dependent translocase activities [36, 42, 43] and why TFIIH does not have the 3'-5' helicase activities [27, 43, 63], why TFIIH and XPA could be both directly recruited to the damaged DNA by XPC-RAD23B or DDB[55-59, 66-68] and TFIIH strongly interacts with and recruits XPA [63, 69, 122, 123], and why XPA plays a key role in facilitating DNA unwinding and the assembly of pre-incision complex during NER [56, 65, 66, 121].

In addition, the nuclease StBax1 fits nicely with XPA together at the forked DNA in the cryo-EM structure (Fig. 3.16D), suggesting that nuclease XPF or XPG could bind similarly like Bax1 to the DNA junction with XPA and XPB (TFIIH core) for damage incision during eukaryotic NER. This hypothesis is consistent with previous data that XPA directly interacts with XPF-ERCC1 [72, 73] and XPG crosslinks to XPB and p52 [63].

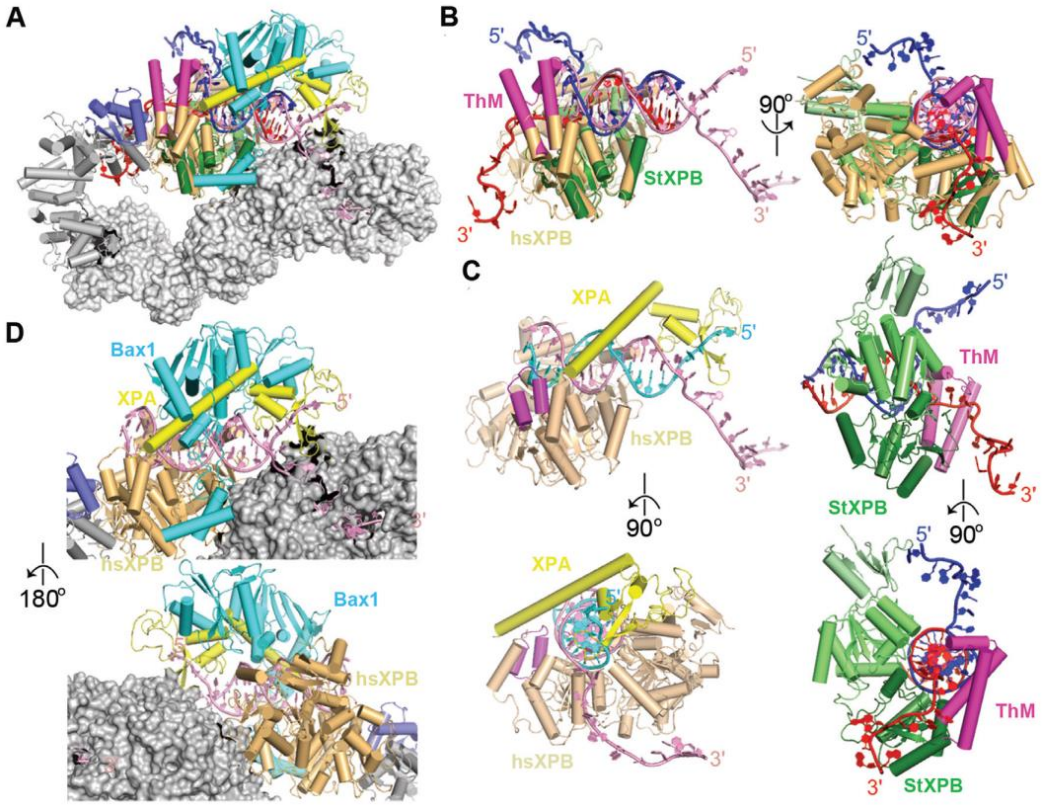

Figure 3.16 Structure comparison of the StXPB-Bax1$^{\Delta C}$-DNA ternary complex and the cryo-EM structure of the TFIIH-XPA-DNA complex.

(A) Superimposition of StXPB in the StXPB-Bax1$^{\Delta C}$-forked DNA ternary complex with human XPB in TFIIH-XPA-DNA cryo-EM structure. Human XPB, p52, p8 and XPA are shown in light orange, gray, slate, and yellow ribbons, respectively. The rest of the TFIIH is shown in gray surfaces. Forked DNA in cryo-EM structure is shown in pink ribbons.

(B) Orthogonal views (left and right) of DNA-protein interactions for StXPB and human XPB as in (A) with other proteins omitted.

(C) Orthogonal views (top and bottom) of human XPB-XPA-DNA subcomplex (left) and StXPB-DNA subcomplex (right). The two strands of DNA in the cryo-EM structure are colored in pink and cyan, respectively. The ThM of human XPB is colored in magenta.

(D) Bax1 fits nicely with XPA at the DNA junction. Zoom-in of the front (top) and back (bottom) views as in (A) with both StXPB and the forked DNA omitted for simplicity.

DISCUSSION

Conventional DNA helicases unwind DNA by loading to the ssDNA overhang of dsDNA and then translocating on this strand with cycles of ATP binding and hydrolysis to "unzip" the dsDNA. Even though XPB proteins possess the conserved helicase motifs like most other SF2 DNA helicases, most previous structural and biochemical evidences indicate that eukaryotic XPB only binds to dsDNA with no appreciable 3'-5' helicase activity [27, 36, 43, 63]. Also mutations in helicase motifs of XPB did not affect its function in NER [27, 124] while its ATPase activity is essentially required for DNA opening in both transcription and NER [27, 31, 121, 124]. Therefore, XPB is believed to be an unconventional DNA helicase to translocate along dsDNA instead of ssDNA with ATPase activity for DNA opening. And this feature makes the traditional helicase assay not applicable to detect DNA unwinding by XPB. However, it remains unclear how XPB unwinds duplex DNA as a translocase in NER.

Our structure of the StXPB-Bax1$^{\Delta C}$-DNA complex has uncovered that archaeal XPB homologs recognize the ds-ss DNA junction by interacting simultaneously with a short 3'-overhang and the DNA duplex immediately adjacent to the junction. This is the first structural evidence that XPB could directly recognize the forked DNA substrates besides the dsDNA substrate. Our structural and biochemical analyses provide new insights to the unconventional DNA unwinding mechanism of archaeal and eukaryotic XPB. Disruption of either the key RED or ThM motif impaired StXPB's ability to interact with DNA, supporting that our DNA-bound ternary structure captures the state of repair bubble extension by the XPB-Bax1 machinery. The DNA-free StXPB-Bax1$^{\Delta C}$ structure

allows us to identify the conformational changes induced by DNA binding and further propose the 2-bp DNA opening model for archaeal XPB. Because the ThM of XPB and Bax1 hold different strands of the melted DNA, this would further split apart the two DNA strands to create the initial repair bubble, which is then extended by the XPB-Bax1 machinery through ATP binding and hydrolysis. Due to its shortened ThM motif, human XPB is more a translocase than a helicase, but XPA may complement this shortage and enhance its helicase activity for the DNA unwinding in consistence with the recent cryo-EM structure revealing that XPA interacts with XPB and helps XPB to hook on the DNA fork [63]. Thus eukaryotic XPB-XPA may work together via the ATPase activity of XPB to initiate the bubble extension around the lesion following damage recognition by XPC, which allows XPD binding to DNA and TFIIH activation for subsequent lesion verification.

A typical NER DNA damage usually induces local melting of DNA. In eukaryotic GG-NER, the DNA lesions are mainly recognized by the XPC-RAD23B complex. XPC-RAD23B detects the DNA distortion through indirect recognition mechanism [125-127] that XPC binds to the flipped-out 2 bases from the undamaged strand rather than the lesion itself [119]. Also yeast homologs of XPA could bind to the kinked lesion-containing DNA without directly contacting the lesion [128]. Because there are no archaeal homologs of both XPC and XPA, we speculate that archaeal XPB may also play a role in damage recognition (Fig. 3.12). Consistent with this idea, StXPB could form distinct protein-DNA bands starting from bubble-3 DNA and these StXPB-bubble DNA interactions increase as the bubble becomes larger (Fig. 3.11). Admittedly, StXPB only

interacts with 3-bp bubble DNA weakly especially at low protein-DNA ratios, suggesting minor helix-distorting lesions like CPD lesion may not be efficiently detected by archaeal NER. Archaeal NER pathway may just be a simplified system with a limited number of proteins involved compared to eukaryotical NER, which has the evolutional advantage and develops a more sophisticated system to better confirm the existence of the DNA damage and accurately pinpoint the lesion site. This muti-protein NER system in eukaryotes also could provide precise regulation mechanisms at different stages during the progression of damage recognition and verification, preventing wasting energy on irrelevant or false-positive DNA lesions in the genome.

Interestingly, in the crystal structure of the StXPB-Bax1$^{\Delta C}$-forked DNA complex, the DNA is kept away from the nuclease domain of Bax1, therefore potentially preventing DNA incision by Bax1 until the DNA bubble is big enough for DNA repair. It is not yet known how the bubble size is determined during NER for any species. One possible mechanism is proposed in Figure 3.17. We recently reported the crystal structures of the XPB-Bax1 complex from both *A. fulgidus* and *S. tokodaii*. In the AfXPB-Bax1 complex, AfXPB remains the open conformation as observed in the AfXPB crystal structure [15], which has the ATP binding groove open to the solution. When the AfXPB-Bax1 heterodimer structure is superimposed with the ternary complex structure over the N-terminal half XPB (DRD and HD1), the completely open conformation of XPB turns the Bax1 away from the DNA fork (Figure 3.17A). This could allow the Bax1 to bind to the other fork of a bubble DNA while keeping the HD1, particularly the RED motif, at the ds-ssDNA junction of the first fork (Figure 3.17B). The distance between the residue

F278 at the ThM tip of StXPB and the residue D305 at the nuclease active site of AfBax1 is 73.9 Å (Figure 3.17A), equivalent to about 14 bps along the phosphate backbone of the DNA arm. The completely open XPB conformation will break the ATP-binding groove of XPB and inhibit the ATPase activity of XPB while the swing of Bax1 will align the nuclease active site of Bax1 right at the ds-ssDNA junction of the second fork of a 14-bp mismatched bubble DNA allowing Bax1 to incise the DNA (Figure 3.17B), in agreement with the nuclease activity of AfXPB-Bax1 complex (Figure 2.6B-C) and *S. solfataricus* XPB-Bax1 complex [96]. If the bubble is shorter than 14 bps, the swing of Bax1 will load Bax1 on a dsDNA region, which cannot be incised by the structure specific DNA nuclease Bax1. In this case, the interactions of Bax1 with dsDNA will be weaker allowing the XPB-Bax1 complex to switch back to the closed XPB conformation for further DNA unwinding until a 14-bp bubble is created. Therefore, the XPB-Bax1 complex could use this conformational switch from closed XPB to completely open XPB as the means to measure the size of the bubble and coordinate DNA unwinding (active XPB and isolated Bax1) with DNA incision (inactive XPB but active Bax1). After the first DNA incision (5' to the lesion in the blue strand, Figure 3.17B) is carried out, the XPB-Bax1 heterodimer could switch to the other fork for the second incision or a Bax1 nuclease will bind directly to the other fork for the second DNA incision (3' to the lesion) to remove the DNA damage (Figure 3.17B).

Notably, the first DNA incision by the XPB-Bax1 heterodimer may correspond to the 5' incision by the ERCC1-XPF complex in human NER, where the interactions of XPB with XPF regulate the incision as phosphorylation at the C-terminus of XPB was

reported to inhibit the incision by the ERCC1-XPF complex [129]. The second incision by Bax1 corresponds to the 3' incision by XPG, which is associated with TFIIH but does not cut the DNA until XPF incises DNA at the other fork of the DNA bubble [16]. Our results suggest that XPG, like Bax1 associated with XPB, is kept away from the unwinding fork by XPB at the damage so that XPG does not incise the forked DNA prematurely.

When our ternary complex is docked onto the TFIIH-XPA-DNA cryo-EM structure (PDB entry: 6RO4) with StXPB superimposed with the human XPB, both the forked DNA and Bax1 fit nicely on the surface of the TFIIH core complex (Fig. 3.16A). When the TFIIH complex extends the bubble large enough, the ERCC1-XPF complex can bind at the other fork for incision. Therefore, the interactions between XPB and XPF could regulate the size of the bubble. Once XPF performs the 5' incision to the lesion, the interactions of XPF with XPB likely change and alter the interactions of XPB with XPG at the unwinding fork to allow XPG to access the fork for the 3' incision to the lesion.

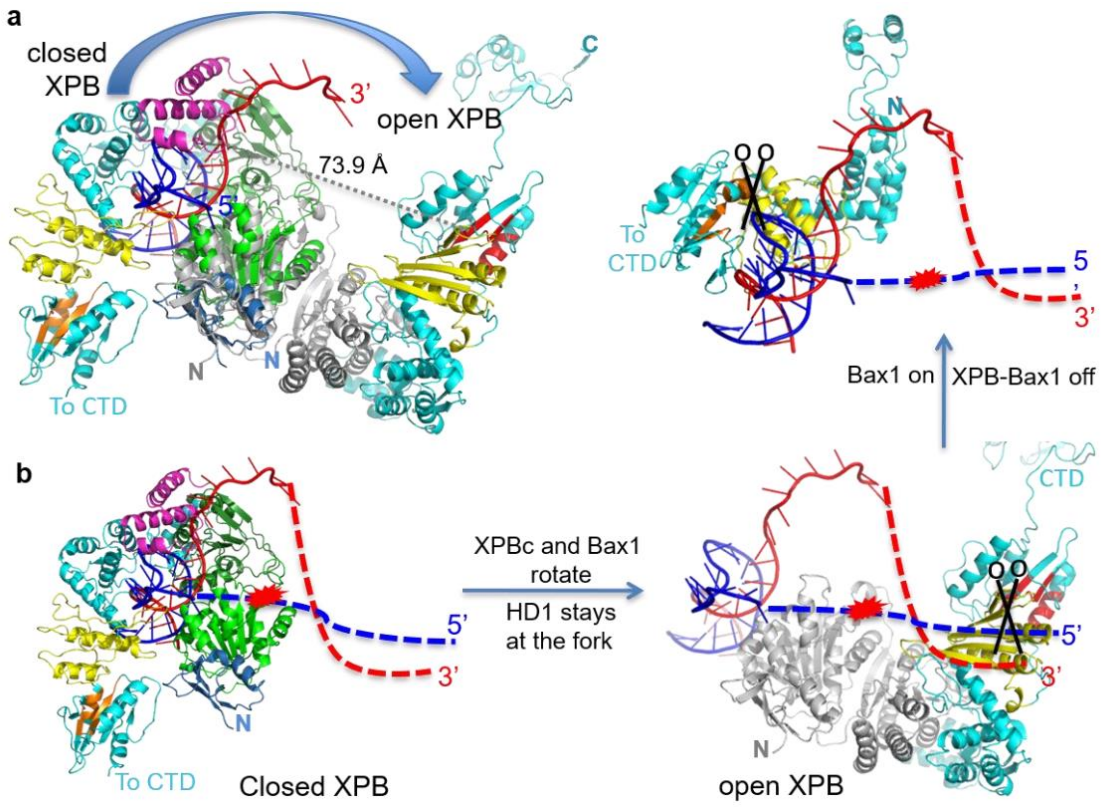

Figure 3.17 Model of DNA incisions by the XPB-nuclease complex at a DNA bubble.

(A) Crystal structure of the StXPB-Bax1$^{\Delta C}$-DNA complex and the AfXPB-Bax1 heterodimer superimposed over the DRD/HD1 of XPB. AfBax1 and Bax1$^{\Delta C}$ are colored in cyan with the CRD domain highlighted in yellow and the nuclease motifs in red. AfXPB is colored in gray while StXPB is colored with DRD in blue, HD1/HD2 in green and ThM in magenta. The distance (gray dash line) between the residue F278 at the ThM tip of StXPB and the residue D305 at the nuclease active motif of AfBax1 is 73.9 Å.

(B) Proposed model of DNA incisions by the XPB-nuclease machinery at the bubble with a DNA lesion. Repair bubble created by the StXPB-Bax1 complex (Left). The two arms of the forked DNA in the ternary complex structure are extended by dashed lines to form the bubble with ~14-bp mismatches. A red star in the bubbled blue line represents a DNA lesion. DNA incisions by the AfXPB-Bax1 complex (right, bottom) and Bax1 (right, top) are shown. Black scissors represent DNA incisions by the nuclease Bax1.

REFERENCES

1. Singleton MR, Dillingham MS, Wigley DB: **Structure and mechanism of helicases and nucleic acid translocases**. *Annu Rev Biochem* 2007, **76**:23-50.

2. Gao Y, Yang W: **Different mechanisms for translocation by monomeric and hexameric helicases**. *Current opinion in structural biology* 2020, **61**:25-32.

3. Bhattacharyya B, Keck JL: **Grip it and rip it: structural mechanisms of DNA helicase substrate binding and unwinding**. *Protein Science* 2014, **23**(11):1498-1507.

4. Ye J, Osborne AR, Groll M, Rapoport TA: **RecA-like motor ATPases—lessons from structures**. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 2004, **1659**(1):1-18.

5. Fairman-Williams ME, Guenther U-P, Jankowsky E: **SF1 and SF2 helicases: family matters**. *Current opinion in structural biology* 2010, **20**(3):313-324.

6. Weeda G, van Ham RC, Vermeulen W, Bootsma D, van der Eb AJ, Hoeijmakers JH: **A presumed DNA helicase encoded by ERCC-3 is involved in the human repair disorders xeroderma pigmentosum and Cockayne's syndrome**. *Cell* 1990, **62**(4):777-791.

7. Schaeffer L, Roy R, Humbert S, Moncollin V, Vermeulen W, Hoeijmakers J, Chambon P, Egly J-M: **DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor**. *Science* 1993, **260**(5104):58-63.

8. Hwang JR, Moncollin V, Vermeulen W, Seroz T, van Vuuren H, Hoeijmakers JH, Egly JM: **A 3′ → 5′ XPB helicase defect in repair/transcription factor TFIIH of xeroderma pigmentosum group B affects both DNA repair and transcription**. *Journal of Biological Chemistry* 1996, **271**(27):15898-15904.

9. Schaeffer L, Moncollin V, Roy R, Staub A, Mezzina M, Sarasin A, Weeda G, Hoeijmakers J, Egly J-M: **The ERCC2/DNA repair protein is associated with the class II BTF2/TFIIH transcription factor**. *The EMBO journal* 1994, **13**(10):2388-2392.

10. Lin YC, Choi WS, Gralla JD: **TFIIH XPB mutants suggest a unified bacterial-like mechanism for promoter opening but not escape**. *Nature structural & molecular biology* 2005, **12**(7):603-607.

11. Gulyas KD, Donahue TF: **SSL2, a suppressor of a stem-loop mutation in the HIS4 leader encodes the yeast homolog of human ERCC-3**. *Cell* 1992, **69**(6):1031-1042.

12. Park E, Guzder SN, Koken M, Jaspers-Dekker I, Weeda G, Hoeijmakers J, Prakash S, Prakash L: **RAD25 (SSL2), the yeast homolog of the human xeroderma pigmentosum group B DNA repair gene, is essential for viability**. *Proceedings of the National Academy of Sciences* 1992, **89**(23):11416-11420.

13. Koken MH, Vreeken C, Bol SA, Cheng NC, Jaspers-Dekker I, Hoeijmakers JH, Eeken JC, Weeda G, Pastink A: **Cloning and characterization of the Drosophila homolog of the xeroderma pigmentosum complementation-group B correcting gene, ERCC3**. *Nucleic acids research* 1992, **20**(21):5541-5548.

14. Ribeiro DT, Machado CR, Costa RM, Praekelt UM, Van Sluys M-A, Menck CF: **Cloning of a cDNA from Arabidopsis thaliana homologous to the human XPB gene**. *Gene* 1998, **208**(2):207-213.

15. Fan L, Arvai AS, Cooper PK, Iwai S, Hanaoka F, Tainer JA: **Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair**. *Molecular cell* 2006, **22**(1):27-37.

16. Schärer OD: **Nucleotide excision repair in eukaryotes**. *Cold Spring Harbor perspectives in biology* 2013, **5**(10):a012609.

17. Luo J, Cimermancic P, Viswanath S, Ebmeier CC, Kim B, Dehecq M, Raman V, Greenberg CH, Pellarin R, Sali A: **Architecture of the human and yeast general transcription and DNA repair factor TFIIH**. *Molecular cell* 2015, **59**(5):794-806.

18. Greber BJ, Nguyen THD, Fang J, Afonine PV, Adams PD, Nogales E: **The cryo-electron microscopy structure of human transcription factor IIH**. *Nature* 2017, **549**(7672):414-417.

19. Cleaver JE, Thompson LH, Richardson AS, States JC: **A summary of mutations in the UV‐sensitive disorders: xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy**. *Human mutation* 1999, **14**(1):9-22.

20. Lehmann AR: **DNA repair-deficient diseases, xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy**. *Biochimie* 2003, **85**(11):1101-1111.

21. Oh KS, Khan SG, Jaspers N, Raams A, Ueda T, Lehmann A, Friedmann PS, Emmert S, Gratchev A, Lachlan K: **Phenotypic heterogeneity in the XPB DNA**

helicase gene (ERCC3): xeroderma pigmentosum without and with Cockayne syndrome. *Human mutation* 2006, **27**(11):1092-1103.

22. Compe E, Egly J-M: **TFIIH: when transcription met DNA repair**. *Nature reviews Molecular cell biology* 2012, **13**(6):343-354.

23. Fan L, Fuss JO, Cheng QJ, Arvai AS, Hammel M, Roberts VA, Cooper PK, Tainer JA: **XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations**. *Cell* 2008, **133**(5):789-800.

24. Liu H, Rudolf J, Johnson KA, McMahon SA, Oke M, Carter L, McRobbie A-M, Brown SE, Naismith JH, White MF: **Structure of the DNA repair helicase XPD**. *Cell* 2008, **133**(5):801-812.

25. Jawhari A, Lain éJ-P, Dubaele S, Lamour V, Poterszman A, Coin F, Moras D, Egly J-M: **p52 Mediates XPB function within the transcription/repair factor TFIIH**. *Journal of Biological Chemistry* 2002, **277**(35):31761-31767.

26. Coin F, De Santis LP, Nardo T, Zlobinskaya O, Stefanini M, Egly J-M: **p8/TTD-A as a repair-specific TFIIH subunit**. *Molecular cell* 2006, **21**(2):215-226.

27. Coin F, Oksenych V, Egly J-M: **Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair**. *Molecular cell* 2007, **26**(2):245-256.

28. Fregoso M, Lain éJ-P, Aguilar-Fuentes J, Mocquet V, Reynaud E, Coin F, Egly J-M, Zurita M: **DNA repair and transcriptional deficiencies caused by mutations in the Drosophila p52 subunit of TFIIH generate developmental defects and chromosome fragility**. *Molecular and cellular biology* 2007, **27**(10):3640-3650.

29. Zhou Y, Kou H, Wang Z: **Tfb5 interacts with Tfb2 and facilitates nucleotide excision repair in yeast**. *Nucleic acids research* 2007, **35**(3):861-871.

30. Kainov DE, Vitorino M, Cavarelli J, Poterszman A, Egly J-M: **Structural basis for group A trichothiodystrophy**. *Nature structural & molecular biology* 2008, **15**(9):980-984.

31. Guzder SN, Sung P, Bailly V, Prakash L, Prakash S: **RAD25 is a DMA helicase required for DNA repair and RNA polymerase II transcription**. *Nature* 1994, **369**(6481):578-581.

32. Goodrich JA, Tjian R: **Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II**. *Cell* 1994, **77**(1):145-156.

33. Moreland RJ, Tirode F, Yan Q, Conaway JW, Egly J-M, Conaway RC: **A role for the TFIIH XPB DNA helicase in promoter escape by RNA polymerase II**. *Journal of Biological Chemistry* 1999, **274**(32):22127-22130.

34. He Y, Yan C, Fang J, Inouye C, Tjian R, Ivanov I, Nogales E: **Near-atomic resolution visualization of human transcription promoter opening**. *Nature* 2016, **533**(7603):359-365.

35. Murakami K, Tsai K-L, Kalisman N, Bushnell DA, Asturias FJ, Kornberg RD: **Structure of an RNA polymerase II preinitiation complex**. *Proceedings of the National Academy of Sciences* 2015, **112**(44):13543-13548.

36. Schilbach S, Hantsche M, Tegunov D, Dienemann C, Wigge C, Urlaub H, Cramer P: **Structures of transcription pre-initiation complex with TFIIH and Mediator**. *Nature* 2017, **551**(7679):204-209.

37. Guzmán E, Lis JT: **Transcription factor TFIIH is required for promoter melting in vivo**. *Molecular and cellular biology* 1999, **19**(8):5652-5658.

38. Winkler GS, Araújo SJ, Fiedler U, Vermeulen W, Coin F, Egly J-M, Hoeijmakers JH, Wood RD, Timmers HTM, Weeda G: **TFIIH with inactive XPD helicase functions in transcription initiation but is defective in DNA repair**. *Journal of Biological Chemistry* 2000, **275**(6):4258-4266.

39. Kuper J, Braun C, Elias A, Michels G, Sauer F, Schmitt DR, Poterszman A, Egly J-M, Kisker C: **In TFIIH, XPD helicase is exclusively devoted to DNA repair**. *PLoS Biol* 2014, **12**(9):e1001954.

40. Fan L, DuPrez KT: **XPB: an unconventional SF2 DNA helicase**. *Progress in biophysics and molecular biology* 2015, **117**(2-3):174-181.

41. Kim T-K, Ebright RH, Reinberg D: **Mechanism of ATP-dependent promoter melting by transcription factor IIH**. *Science* 2000, **288**(5470):1418-1421.

42. Grünberg S, Warfield L, Hahn S: **Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening**. *Nature structural & molecular biology* 2012, **19**(8):788.

43. Fishburn J, Tomko E, Galburt E, Hahn S: **Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA**

polymerase II open complex formation. *Proceedings of the National Academy of Sciences* 2015, **112**(13):3961-3966.

44. Alekseev S, Coin F: **Orchestral maneuvers at the damaged sites in nucleotide excision repair**. *Cellular and Molecular Life Sciences* 2015, **72**(11):2177-2186.

45. Gillet LC, Schärer OD: **Molecular mechanisms of mammalian global genome nucleotide excision repair**. *Chemical reviews* 2006, **106**(2):253-276.

46. Selby CP, Sancar A: **Mechanisms of transcription-repair coupling and mutation frequency decline**. *Microbiological reviews* 1994, **58**(3):317-329.

47. Iyer N, Reagan MS, Wu K-J, Canagarajah B, Friedberg EC: **Interactions involving the human RNA polymerase II transcription/nucleotide excision repair complex TFIIH, the nucleotide excision repair protein XPG, and Cockayne syndrome group B (CSB) protein**. *Biochemistry* 1996, **35**(7):2157-2167.

48. Lainé JP, Egly JM: **Initiation of DNA repair mediated by a stalled RNA polymerase IIO**. *The EMBO journal* 2006, **25**(2):387-397.

49. Beerens N, Hoeijmakers JH, Kanaar R, Vermeulen W, Wyman C: **The CSB protein actively wraps DNA**. *Journal of Biological Chemistry* 2005, **280**(6):4722-4729.

50. Spivak G: **Transcription-coupled repair: an update**. *Archives of toxicology* 2016, **90**(11):2583-2594.

51. Pani B, Nudler E: **Mechanistic insights into transcription coupled DNA repair**. *DNA repair* 2017, **56**:42-50.

52. Spivak G: **Nucleotide excision repair in humans**. *DNA repair* 2015, **36**:13-18.

53. Fuss JO, Cooper PK: **DNA repair: dynamic defenders against cancer and aging**. *PLoS Biol* 2006, **4**(6):e203.

54. Van Der Spek PJ, Eker A, Rademakers S, Visser C, Sugasawa K, Masutani C, Hanaoka F, Bootsma D, Hoeijmakers JH: **XPC and human homologs of RAD23: intracellular localization and relationship to other nucleotide excision repair complexes**. *Nucleic acids research* 1996, **24**(13):2551-2559.

55. Yokoi M, Masutani C, Maekawa T, Sugasawa K, Ohkuma Y, Hanaoka F: **The xeroderma pigmentosum group C protein complex XPC-HR23B plays an**

important role in the recruitment of transcription factor IIH to damaged DNA. *Journal of Biological Chemistry* 2000, **275**(13):9870-9875.

56. Volker M, Mon é MJ, Karmakar P, van Hoffen A, Schul W, Vermeulen W, Hoeijmakers JH, van Driel R, van Zeeland AA, Mullenders LH: **Sequential assembly of the nucleotide excision repair factors in vivo**. *Molecular cell* 2001, **8**(1):213-224.

57. Ara újo SJ, Nigg EA, Wood RD: **Strong functional interactions of TFIIH with XPC and XPG in human DNA nucleotide excision repair, without a preassembled repairosome**. *Molecular and cellular biology* 2001, **21**(7):2281-2291.

58. Uchida A, Sugasawa K, Masutani C, Dohmae N, Araki M, Yokoi M, Ohkuma Y, Hanaoka F: **The carboxy-terminal domain of the XPC protein plays a crucial role in nucleotide excision repair through interactions with transcription factor IIH**. *DNA repair* 2002, **1**(6):449-461.

59. Okuda M, Kinoshita M, Kakumu E, Sugasawa K, Nishimura Y: **Structural insight into the mechanism of TFIIH recognition by the acidic string of the nucleotide excision repair factor XPC**. *Structure* 2015, **23**(10):1827-1837.

60. Fuss JO, Tainer JA: **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* 2011, **10**(7):697-713.

61. Naegeli H, Modrich P, Friedberg EC: **The DNA helicase activities of Rad3 protein of Saccharomyces cerevisiae and helicase II of Escherichia coli are differentially inhibited by covalent and noncovalent DNA modifications**. *Journal of Biological Chemistry* 1993, **268**(14):10386-10392.

62. Kuper J, Wolski SC, Michels G, Kisker C: **Functional and structural studies of the nucleotide excision repair helicase XPD suggest a polarity for DNA translocation**. *The EMBO journal* 2012, **31**(2):494-502.

63. Kokic G, Chernev A, Tegunov D, Dienemann C, Urlaub H, Cramer P: **Structural basis of TFIIH activation for nucleotide excision repair**. *Nature communications* 2019, **10**(1):1-9.

64. Soultanas P, Wigley DB: **DNA helicases: ‘inching forward’**. *Current opinion in structural biology* 2000, **10**(1):124-128.

65. Rademakers S, Volker M, Hoogstraten D, Nigg AL, Mon é MJ, Van Zeeland AA, Hoeijmakers JH, Houtsmuller AB, Vermeulen W: **Xeroderma pigmentosum**

group A protein loads as a separate factor onto DNA lesions. *Molecular and Cellular Biology* 2003, **23**(16):5755-5767.

66. Riedl T, Hanaoka F, Egly JM: **The comings and goings of nucleotide excision repair factors on damaged DNA**. *The EMBO journal* 2003, **22**(19):5293-5303.

67. Bunick CG, Miller MR, Fuller BE, Fanning E, Chazin WJ: **Biochemical and structural domain analysis of xeroderma pigmentosum complementation group C protein**. *Biochemistry* 2006, **45**(50):14965-14979.

68. Wakasugi M, Kasashima H, Fukase Y, Imura M, Imai R, Yamada S, Cleaver JE, Matsunaga T: **Physical and functional interaction between DDB and XPA in nucleotide excision repair**. *Nucleic acids research* 2009, **37**(2):516-525.

69. Park C-H, Mu D, Reardon JT, Sancar A: **The general transcription-repair factor TFIIH is recruited to the excision repair complex by the XPA protein independent of the TFIIE transcription factor**. *Journal of Biological Chemistry* 1995, **270**(9):4896-4902.

70. Matsuda T, Saijo M, Kuraoka I, Kobayashi T, Nakatsu Y, Nagai A, Enjoji T, Masutani C, Sugasawa K, Hanaoka F: **DNA repair protein XPA binds replication protein A (RPA)**. *Journal of Biological Chemistry* 1995, **270**(8):4152-4157.

71. Li L, Lu X, Peterson CA, Legerski RJ: **An interaction between the DNA repair factor XPA and replication protein A appears essential for nucleotide excision repair**. *Molecular and cellular biology* 1995, **15**(10):5396-5402.

72. Li L, Elledge SJ, Peterson CA, Bales ES, Legerski RJ: **Specific association between the human DNA repair proteins XPA and ERCC1**. *Proceedings of the National Academy of Sciences* 1994, **91**(11):5012-5016.

73. Li L, Peterson CA, Lu X, Legerski RJ: **Mutations in XPA that prevent association with ERCC1 are defective in nucleotide excision repair**. *Molecular and cellular biology* 1995, **15**(4):1993-1998.

74. Gilljam KM, Müller R, Liabakk NB, Otterlei M: **Nucleotide excision repair is associated with the replisome and its efficiency depends on a direct interaction between XPA and PCNA**. *PloS one* 2012, **7**(11):e49199.

75. Fan J, Pavletich NP: **Structure and conformational change of a replication protein A heterotrimer bound to ssDNA**. *Genes & development* 2012, **26**(20):2337-2347.

76. Overmeer RM, Moser J, Volker M, Kool H, Tomkinson AE, van Zeeland AA, Mullenders LH, Fousteri M: **Replication protein A safeguards genome integrity by controlling NER incision events**. *Journal of Cell Biology* 2011, **192**(3):401-415.

77. Ito S, Kuraoka I, Chymkowitch P, Compe E, Takedachi A, Ishigami C, Coin F, Egly J-M, Tanaka K: **XPG stabilizes TFIIH, allowing transactivation of nuclear receptors: implications for Cockayne syndrome in XP-G/CS patients**. *Molecular cell* 2007, **26**(2):231-243.

78. O'Donovan A, Davies AA, Moggs JG, West SC, Wood RD: **XPG endonuclease makes the 3′ incision in human DNA nucleotide excision repair**. *Nature* 1994, **371**(6496):432-435.

79. Matsunaga T, Mu D, Park C-H, Reardon JT, Sancar A: **Human DNA repair excision nuclease: analysis of the roles of the subunits involved in dual incisions by using anti-XPG and anti-ERCC1 antibodies**. *Journal of Biological Chemistry* 1995, **270**(35):20862-20869.

80. Sijbers AM, De Laat WL, Ariza RR, Biggerstaff M, Wei Y-F, Moggs JG, Carter KC, Shell BK, Evans E, De Jong MC: **Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease**. *Cell* 1996, **86**(5):811-822.

81. Fagbemi AF, Orelli B, Schärer OD: **Regulation of endonuclease activity in human nucleotide excision repair**. *DNA repair* 2011, **10**(7):722-729.

82. Kemp MG, Reardon JT, Lindsey-Boltz LA, Sancar A: **Mechanism of release and fate of excised oligonucleotides during nucleotide excision repair**. *Journal of Biological Chemistry* 2012, **287**(27):22889-22899.

83. Mocquet V, Laine JP, Riedl T, Yajin Z, Lee MY, Egly JM: **Sequential recruitment of the repair factors during NER: the role of XPG in initiating the resynthesis step**. *The EMBO journal* 2008, **27**(1):155-167.

84. Moser J, Kool H, Giakzidis I, Caldecott K, Mullenders LH, Fousteri MI: **Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase III α in a cell-cycle-specific manner**. *Molecular cell* 2007, **27**(2):311-323.

85. Ogi T, Limsirichaikul S, Overmeer RM, Volker M, Takenaka K, Cloney R, Nakazawa Y, Niimi A, Miki Y, Jaspers NG: **Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells**. *Molecular cell* 2010, **37**(5):714-727.

86. Lehmann AR: **DNA polymerases and repair synthesis in NER in human cells**. *DNA repair* 2011, **10**(7):730-733.

87. White MF, Allers T: **DNA repair in the archaea—an emerging picture**. *FEMS microbiology reviews* 2018, **42**(4):514-526.

88. Cubeddu L, White MF: **DNA damage detection by an archaeal single-stranded DNA-binding protein**. *Journal of molecular biology* 2005, **353**(3):507-516.

89. Fröls S, Gordon PM, Panlilio MA, Duggin IG, Bell SD, Sensen CW, Schleper C: **Response of the hyperthermophilic archaeon Sulfolobus solfataricus to UV damage**. *Journal of bacteriology* 2007, **189**(23):8708-8718.

90. Salerno V, Napoli A, White MF, Rossi M, Ciaramella M: **Transcriptional response to DNA damage in the archaeon Sulfolobus solfataricus**. *Nucleic acids research* 2003, **31**(21):6127-6138.

91. Fujikane R, Ishino S, Ishino Y, Forterre P: **Genetic analysis of DNA repair in the hyperthermophilic archaeon, Thermococcus kodakaraensis**. *Genes & genetic systems* 2010, **85**(4):243-257.

92. Richards JD, Cubeddu L, Roberts J, Liu H, White MF: **The archaeal XPB protein is a ssDNA-dependent ATPase with a novel partner**. *Journal of molecular biology* 2008, **376**(3):634-644.

93. Kinch LN, Ginalski K, Rychlewski L, Grishin NV: **Identification of novel restriction endonuclease-like fold families among hypothetical proteins**. *Nucleic acids research* 2005, **33**(11):3598-3605.

94. Roth HM, Tessmer I, Van Houten B, Kisker C: **Bax1 Is a Novel Endonuclease**. *Journal of Biological Chemistry* 2009, **284**(47):32272-32278.

95. Roth HM, Römer J, Grundler V, Van Houten B, Kisker C, Tessmer I: **XPB helicase regulates DNA incision by the Thermoplasma acidophilum endonuclease Bax1**. *DNA repair* 2012, **11**(3):286-293.

96. Rouillon C, White MF: **The XBP-Bax1 helicase-nuclease complex unwinds and cleaves DNA: implications for eukaryal and archaeal nucleotide excision repair**. *Journal of Biological Chemistry* 2010, **285**(14):11013-11022.

97. Ma X, Hong Y, Han W, Sheng D, Ni J, Hou G, Shen Y: **Single-stranded DNA binding activity of XPBI, but not XPBII, from Sulfolobus tokodaii causes double-stranded DNA melting**. *Extremophiles* 2011, **15**(1):67-76.

98. Eme L, Spang A, Lombard J, Stairs CW, Ettema TJ: **Archaea and the origin of eukaryotes**. *Nature Reviews Microbiology* 2017, **15**(12):711.

99. Rudolf J, Makrantoni V, Ingledew WJ, Stark MJ, White MF: **The DNA repair helicases XPD and FancJ have essential iron-sulfur domains**. *Molecular cell* 2006, **23**(6):801-808.

100. Wolski SC, Kuper J, Hänzelmann P, Truglio JJ, Croteau DL, Van Houten B, Kisker C: **Crystal structure of the FeS cluster–containing nucleotide excision repair helicase XPD**. *PLoS Biol* 2008, **6**(6):e149.

101. Ling MM, Robinson BH: **Approaches to DNA mutagenesis: an overview**. *Analytical biochemistry* 1997, **254**(2):157-178.

102. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J: **Fast, scalable generation of high‐quality protein multiple sequence alignments using Clustal Omega**. *Molecular systems biology* 2011, **7**(1):539.

103. Robert X, Gouet P: **Deciphering key features in protein structures with the new ENDscript server**. *Nucleic acids research* 2014, **42**(W1):W320-W324.

104. Kahanda D, DuPrez KT, Hilario E, McWilliams MA, Wohlgamuth CH, Fan L, Slinker JD: **Application of electrochemical devices to characterize the dynamic actions of helicases on DNA**. *Analytical chemistry* 2018, **90**(3):2178-2185.

105. Otwinowski Z, Minor W: **[20] Processing of X-ray diffraction data collected in oscillation mode**. *Methods in enzymology* 1997, **276**:307-326.

106. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ: **Phaser crystallographic software**. *Journal of applied crystallography* 2007, **40**(4):658-674.

107. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA: **REFMAC5 for the refinement of macromolecular crystal structures**. *Acta Crystallographica Section D: Biological Crystallography* 2011, **67**(4):355-367.

108. Battye TGG, Kontogiannis L, Johnson O, Powell HR, Leslie AG: **iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM**. *Acta Crystallographica Section D: Biological Crystallography* 2011, **67**(4):271-281.

109. Evans P: **Scaling and assessment of data quality**. *Acta Crystallographica Section D: Biological Crystallography* 2006, **62**(1):72-82.

110. Emsley P, Lohkamp B, Scott WG, Cowtan K: **Features and development of Coot**. *Acta Crystallographica Section D: Biological Crystallography* 2010, **66**(4):486-501.

111. Adams PD, Grosse-Kunstleve RW, Hung L-W, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC: **PHENIX: building new software for automated crystallographic structure determination**. *Acta Crystallographica Section D: Biological Crystallography* 2002, **58**(11):1948-1954.

112. Gradia SD, Ishida JP, Tsai M-S, Jeans C, Tainer JA, Fuss JO: **MacroBac: new technologies for robust and efficient large-scale production of recombinant multiprotein complexes**. *Methods in enzymology* 2017, **592**:1-26.

113. Lanzetta PA, Alvarez LJ, Reinach PS, Candia OA: **An improved assay for nanomole amounts of inorganic phosphate**. *Analytical biochemistry* 1979, **100**(1):95-97.

114. Walter TS, Meier C, Assenberg R, Au K-F, Ren J, Verma A, Nettleship JE, Owens RJ, Stuart DI, Grimes JM: **Lysine methylation as a routine rescue strategy for protein crystallization**. *Structure* 2006, **14**(11):1617-1622.

115. Kim Y, Quartey P, Li H, Volkart L, Hatzos C, Chang C, Nocek B, Cuff M, Osipiuk J, Tan K: **Large-scale evaluation of protein reductive methylation for improving protein crystallization**. *Nature methods* 2008, **5**(10):853-854.

116. Stelter M, Acajjaoui S, McSweeney S, Timmins J: **Structural and mechanistic insight into DNA unwinding by Deinococcus radiodurans UvrD**. *PLoS One* 2013, **8**(10):e77364.

117. Lee JY, Yang W: **UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke**. *Cell* 2006, **127**(7):1349-1360.

118. Pei J, Kim B-H, Grishin NV: **PROMALS3D: a tool for multiple protein sequence and structure alignments**. *Nucleic acids research* 2008, **36**(7):2295-2300.

119. Min J-H, Pavletich NP: **Recognition of DNA damage by the Rad4 nucleotide excision repair protein**. *Nature* 2007, **449**(7162):570-575.

120. Eryilmaz J, Ceschini S, Ryan J, Geddes S, Waters TR, Barrett TE: **Structural insights into the cryptic DNA-dependent ATPase activity of UvrB**. *Journal of molecular biology* 2006, **357**(1):62-72.

121. Li C-L, Golebiowski FM, Onishi Y, Samara NL, Sugasawa K, Yang W: **Tripartite DNA lesion recognition and verification by XPC, TFIIH, and XPA in nucleotide excision repair**. *Molecular cell* 2015, **59**(6):1025-1034.

122. Li R-Y, Calsou P, Jones CJ, Salles B: **Interactions of the transcription/DNA repair factor TFIIH and XP repair proteins with DNA lesions in a cell-free repair assay**. *Journal of molecular biology* 1998, **281**(2):211-218.

123. Ziani S, Nagy Z, Alekseev S, Soutoglou E, Egly J-M, Coin F: **Sequential and ordered assembly of a large DNA repair complex on undamaged chromatin**. *Journal of Cell Biology* 2014, **206**(5):589-598.

124. Oksenych V, De Jesus BB, Zhovmer A, Egly JM, Coin F: **Molecular insights into the recruitment of TFIIH to sites of DNA damage**. *The EMBO journal* 2009, **28**(19):2971-2980.

125. Buterin T, Meyer C, Giese B, Naegeli H: **DNA quality control by conformational readout on the undamaged strand of the double helix**. *Chemistry & biology* 2005, **12**(8):913-922.

126. Chen X, Velmurugu Y, Zheng G, Park B, Shim Y, Kim Y, Liu L, Van Houten B, He C, Ansari A: **Kinetic gating mechanism of DNA damage recognition by Rad4/XPC**. *Nature communications* 2015, **6**(1):1-10.

127. Puumalainen M-R, Rüthemann P, Min J-H, Naegeli H: **Xeroderma pigmentosum group C sensor: unprecedented recognition strategy and tight spatiotemporal regulation**. *Cellular and Molecular Life Sciences* 2016, **73**(3):547-566.

128. Koch SC, Kuper J, Gasteiger KL, Simon N, Strasser R, Eisen D, Geiger S, Schneider S, Kisker C, Carell T: **Structural insights into the recognition of cisplatin and AAF-dG lesion by Rad14 (XPA)**. *Proceedings of the National Academy of Sciences* 2015, **112**(27):8272-8277.

129. Coin F, Auriol J, Tapias A, Clivio P, Vermeulen W, Egly JM: **Phosphorylation of XPB helicase regulates TFIIH nucleotide excision repair activity**. *The EMBO journal* 2004, **23**(24):4835-4846.

APPENDIX A

Mapping human XPB-p52 interacting interface and two human disease mutations

INTRODUCTION

In eukaryotes, XPB is an essential subunit of TFIIH, which is crucial for both transcription and DNA repair. Human XPB is a 782-residue protein composed of 4 domains: N terminal domain (NTD+DRD), two conserved core RecA-like helicase domains (HD1 and HD2) and a C-terminal extension domain (CTD) (Figure A.1A). Since human TFIIH was discovered in 1991 [1], extensive biochemical and structural studies have been focusing on XPB. The crystal structure of XPB HD2 domain (502-730) was firstly determined at 1.8 Angstrom in 2013 [2] and then the cryo-EM structures of yeast and human TFIIH in 2017 provided the overall conformation and position of XPB in TFIIH and also revealed that XPB interacts with p52 and the C-terminal extension domain of XPB binds to p8 [3, 4].

However, these structures did not reveal enough detailed information for the N terminal domain of human XPB because of limited resolution. It remains largely unclear about which region in XPB-NTD specifically interacts with p52. And there are two mutations F99S and T119P in human XPB-NTD associating with human diseases XP and TTD, respectively [5]. The F99S substitution was reported to possibly weaken the interactions between the p52 and XPB while the T119P mutation is with unknown

mechanisms [6-8]. Therefore, the molecular mechanisms underlying these human diseases were still not well understood.

In this chapter, I mapped the minimal domain in human XPB that interacts with p52 (in complex with p8) and the biochemical impact of disease-causing mutations in XPB-NTD. I also attempted to solve the structure of human XPB in complex with p52/p8 but did not succeed due to limited X-ray diffraction of obtained crystals. These results could help provide a better understanding of the XPB-p52 interacting interface, the role of the XPB/p52/p8 sub-complex in NER and how these two mutations in XPB-NTD lead to human diseases.

MATERIALS AND METHODS

Cloning, expression and purification of human p52/p8

DNA encoding full-length p52 was cloned into the cloning site I of a modified pET-Duet vector with an N-terminal His_6-tag, p8 was cloned into the cloning site II. His_6-p52 and p8 were co-expressed in *E. coli* Rosetta (DE3) pLys-S cells and the transformed cells were grown at 37 °C in LB medium until OD600 reached 0.8 and then induced with 0.1 mM IPTG for 20 h at 18 ℃. The cells were harvested by centrifugation, resuspended in a lysis buffer (50 mM Tris-HCl, pH 7.0, 500 mM NaCl and 10% glycerol) and then lysed by sonication. The cell debris was removed by ultracentrifugation. The supernatant containing the protein complex was purified by affinity chromatography using a Ni-NTA column and was eluted using the lysis buffer containing 300 mM imidazole. The p52/p8 complex was further purified by ion-exchange (SPFF) and gel filtration chromatography using SPFF and Hiload Superdex 200 columns (25 mM Tris-HCl, pH 7.5, 250 mM NaCl, 5% glycerol, 1mM DTT). The purified proteins were concentrated and stored at -80 ℃.

Co-expression and purification of human XPB-NTD/p52/p8

The DNA sequences encoding different WT human XPB-NTD (Table A.1) or human XPB-NTD F99S and T119P mutants were amplified by PCR and cloned into a modified pET28a vector with a SUMO protein fused at the N terminus after the His_6-tag. DNA encoding full-length p52 was cloned into the cloning site I of a tag-free pET-Duet vector, p8 was cloned into the cloning site II. XPB-NTD, p52, and p8 were co-expressed in *E.*

coli Rosetta (DE3) pLys-S cells and the transformed cells were grown at 37 °C in LB medium until OD600 reached 0.8 and then induced with 0.2 mM IPTG for 22 h at 18 ℃. The cells were harvested by centrifugation, resuspended in a lysis buffer (50 mM Tris-HCl, pH 7.0, 400 mM NaCl and 10% glycerol, 5 mM β-mercaptoethanol) and then lysed by sonication. The cell debris was removed by ultracentrifugation. The supernatant containing the protein complex was purified by affinity chromatography using a Ni-NTA column and was eluted using 500 mM imidazole. The sumo protease was added to remove the His-SUMO tag of XPB-NTD. The XPB-NTD, p52 and p8 complex was further purified by ion-exchange (SPFF) and gel filtration chromatography using SPFF and Hiload Superdex 200 columns (25 mM Tris-HCl, pH 7.5, 200 mM NaCl, 5% glycerol, 1mM DTT). The purified proteins were concentrated and stored at -80 ℃.

Cloning, expression and purification of human XPB/p52/p8 WT, F99S and T119P

The DNA encoding WT, F99S or T119P mutant of human XPB was cloned into a modified pFastBac vector with an N-terminal His_6-tag. The DNA encoding full-length human p52 and p8 were cloned into MacroBac 438A vector, and then p52 and p8 were combined into a single vector via restriction digestion and ligation-independent cloning [9]. The recombinant baculovirus expressing XPB or p52/p8 was generated using standard protocols. High Five insect cells were co-infected with these two recombinant baculoviruses. The cells were harvested after 65-70 hours by centrifugation. The pellets were resuspended in lysis buffer containing 50 mM Tris-Cl pH 7.0, 500 mM NaCl, 10% glycerol, 1 mM PMSF. The cells were then lysed by sonication, and the debris was removed by ultracentrifugation. The supernatant was mixed with Ni-NTA resin and rocked

for 1 hour at 4 ℃ before elution with 400 mM imidazole. PreScission protease was then added to remove the His_6-tag. The proteins were further purified by ion-exchange chromatography and gel-filtration chromatography (Superdex 200, 16/60). The purified protein samples were concentrated in 25 mM Tris-Cl pH 7.5, 200 mM NaCl, 5% glycerol, 2 mM DTT, and stored at -80 ℃.

Electrophoretic mobility shift assay

The protein samples of XPB(54-312)/p52/p8 were mixed with different oligonucleotides on ice for 1 h with a molar ratio of 1:1 in the binding buffer consisting of 50 mM Tris-Cl pH 7.5, 100 mM NaCl, 5% glycerol, 1 mM DTT. Samples were then loaded to a 1.2% TBE agarose gel, which was run for 90 min (60V) in the TBE buffer at cold room. The gels were stained by ethidium bromide and visualized under UV light.

RESULTS

Human XPB-NTD could be co-expressed with p52/p8

Before I started this project, other lab members had already tested expressing numerous XPB full-length and the N terminal domain constructs in *E.coli*. All those XPB constructs suffered from low expression or insolubility, so apparently this approach did not work out. Therefore, I established the insect cell system for protein expression in lab and tried to express XPB alone from insect cells. I cloned several constructs of XPB with different protein tags, but unfortunately, all these recombinant XPB fragments were all aggregates and thus not active. Then I attempted to co-express XPB-NTD with p52, considering that p52 was reported to interact with and regulate the activity of XPB [10, 11]. However, it was not working either. At that time, the available cryo-EM structures of TFIIH were at low resolution and it was unclear how XPB interacts with other subunits in TFIIH. Based on the fact that p52 C terminal domain also interacts with p8 [10, 12, 13], we decided to co-express these three proteins XPB, p52 and p8 together. And we speculated that p52/p8 together would help XPB fold correctly during expression and also improve its solubility.

First, co-expression of the full length p52/p8 was tested in *E.coli*, and p52 could form a stable complex with p8. This protein complex was soluble and active without any aggregations (Figure. A.1B). Noteworthy, I also tested co-expression of some truncated p52 with p8 but all those modified p52 were not expressed well with p8 or might result in insoluble protein complexes (data not shown). Based on the secondary structure

predictions, new XPB-N constructs (No. 1-7) containing the intact DRD domain or HD1 were designed to co-express with p52/p8 in *E.coli* (Table A.1). Three of those XPB constructs including XPB (residues 54-312), XPB (residues 65-312), and XPB (residues 73-312) resulted in soluble and stable ternary protein complex of XPB/p52/p8, which were readily purified to homogeneity by affinity chromatography followed by ion-exchange and size-exclusion chromatograph (Figure. A.1C). Next these XPB-N/p52/p8 complexes were used for extensive crystallization screening but no crystals could be produced.

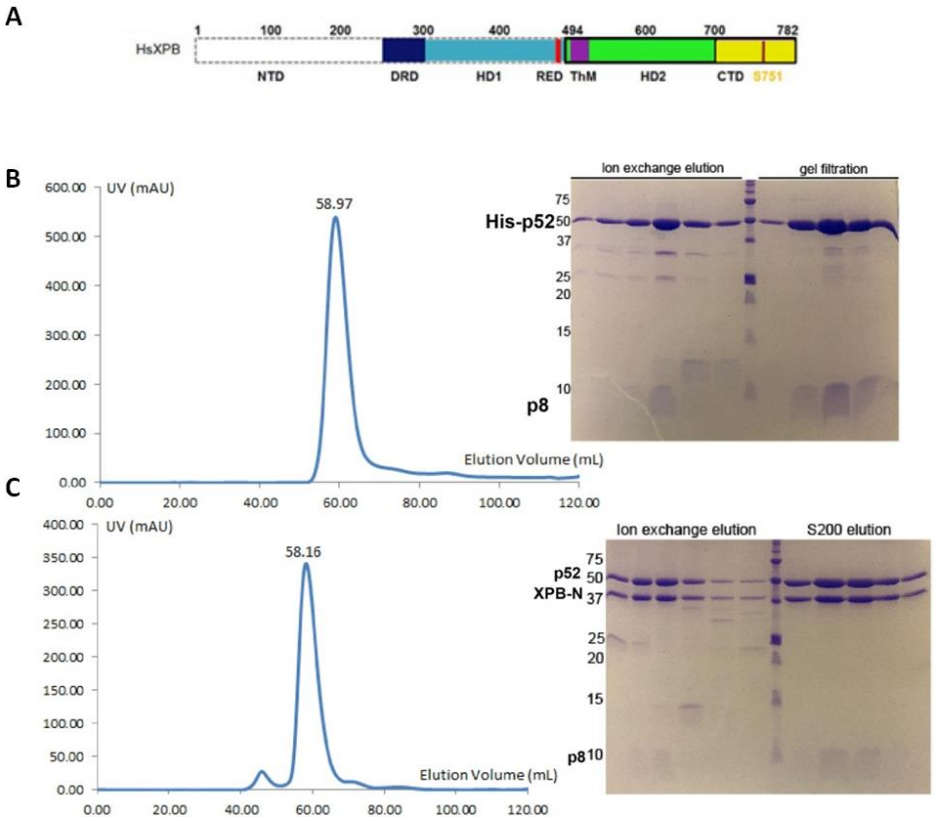

Figure A.1 XPB domain architecture and purification of the XPB/p52/p8 complex.

(A) Domains of XPB include the N-terminal domain (NTD; white), the damage-recognition domain (DRD; blue), helicase domains 1 (HD1; cyan) and 2 (HD2; green), the thumb-like insert (ThM; magenta) and the C-terminal extension (CTE; pink). The known phosphate-site residue Ser751 is indicated by an orange line.

(B-C) Gel-filtration chromatography profiles of p52/p8 and XPB-N/p52/p8 (aa 54-312). SDS gels show samples in the fractions eluted from ion-exchange and gel-filtration chromatography.

Mapping of the minimal p52-interacting domain in human XPB

To further map the minimal p52-interacting domain in XPB and improve the crystallization results, I designed several truncated XPB constructs (No. 8-14 in Table A.1) based on the secondary structure of XPB and further tested their co-expressions with p52/p8. To my surprise, only two XPB constructs (No. 11 and 13) worked nicely when being co-expressed with p52/p8 (Fig. A.2A). Notably, XPB (residues 30-179)/p52/p8 (No. 13) suffers from lower solubility and degradation during purification, leading to extremely low yield. These results suggest that the C terminal boundary of XPB-NTD is residue 160 and the minimal domain in XPB that could interact with p52/p8 is from residue 30 to 160. Interestingly, XPB-N (residues 54-312)/p52/p8 (No. 2) could be smoothly expressed and purified while soluble XPB (residues 54-160)/p52/p8 proteins (No. 12) could not be obtained, suggesting the truncated NTD construct (residues 54-160) of XPB is not as stable as the minimal NTD domain (residues 30-160) but adding an intact DRD domain (residues 54-312) could somehow stabilize the NTD to maintain its solubility. Furthermore, in comparison to the S200 profile of the XPB-N/p52/p8 (No. 2, residues 54-312) complex (Figure. A.1C), the peak of XPB-NTD/p52/p8 (No. 11, residues 30-160) shifts to a higher elution volume (65 mL compared to 58 mL), which means this complex has a substantially smaller molecular weight (MW) and thus could probably be a better sample to be applied for crystallization trials (Fig. A.2A). We thus focused on determining the structure of this XPN-NTD/p52/p8 complex, since there was no available structure that reveals how XPB N terminal interacts with p52 at that time.

Table A.1 Summary of human XPB constructs used in co-expression with p52/p8

No.	XPB construct	Result
1	aa 54-301	N
2	aa 54-312	Y
3	aa 65-312	Y
4	aa 73-312	Y
5	aa 90-312	N
6	aa 60-494	N
7	aa 40-500	N
8	aa 54-242	N
9	aa 54-212	N
10	aa 54-160	Y
11	aa 30-160	Y
12	aa 30-146	N
13	aa 30-179	Low
14	aa 30-201	N

aa represents amino acids. WT p52/p8 construct was used in each co-expression test. Y means the co-expression worked well and the expressed proteins were soluble while N means the co-expression was not working (extremely low expression level or proteins were insoluble).

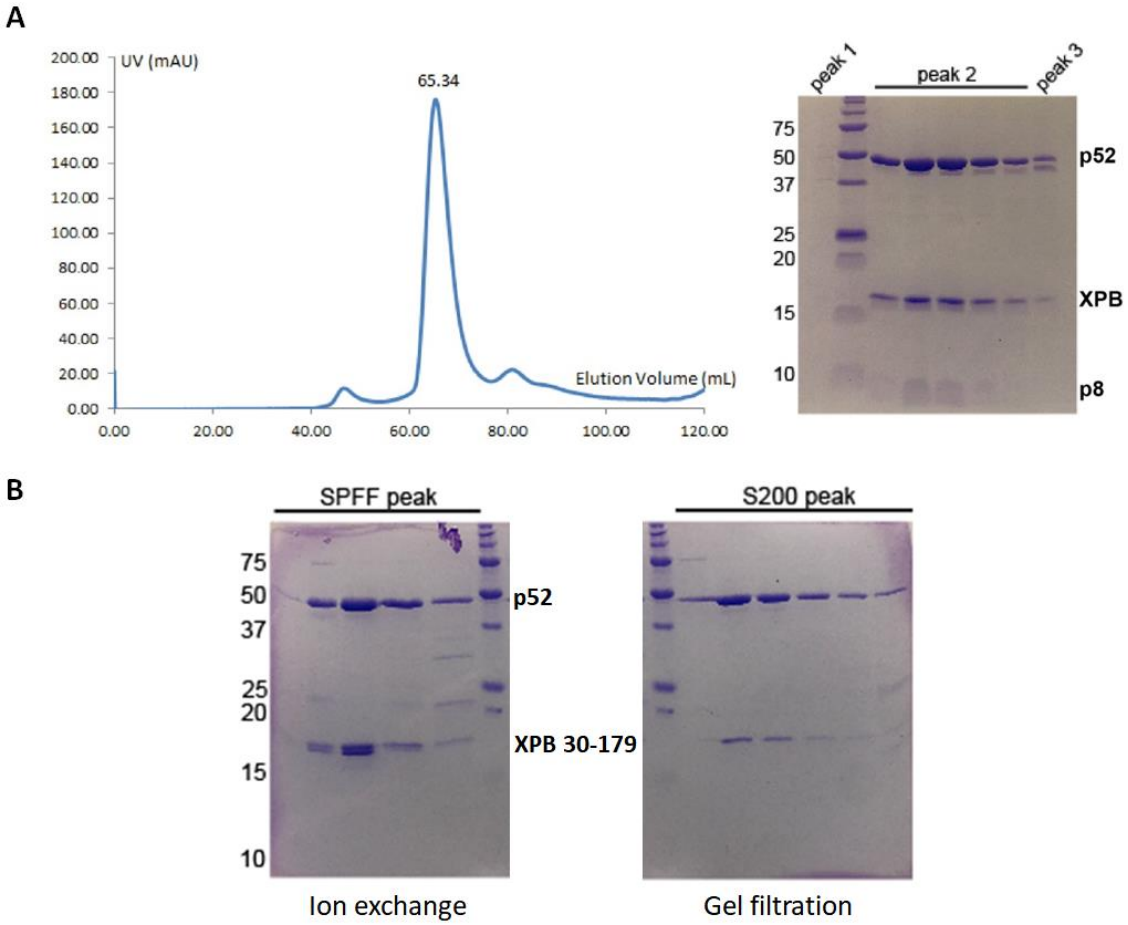

Figure A.2 Purification of the XPB-NTD/p52/p8 complexes.

(A) Gel-filtration chromatography profile of XPB (residues 30-160)/p52/p8 (No. 11). The elution position of peak 2 is indicated. The SDS gel shows samples in the fractions eluted from the S200 gel-filtration chromatography.

(B) SDS gels show the protein samples of XPB (residues 30-179)/p52/p8 (No. 13) after ion-exchange purification (SPFF) and S200 gel filtration. The protein amount of XPB (residues 30-179) reduced during the purification process (as shown by the decreased band intensities), indicating an unstable complex suffering from degradation.

Crystallization trials of human XPB-NTD/p52/p8

Initial attempts to crystallize the XPB-N/p52/p8 proteins (No. 2-4 in Table A.1) were not successful, and also expression of XPB-NTD/p52/p8 using truncated p52 or without p8 did not work well. Therefore, I focused on crystallizing the XPB (residues 30-160)/p52/p8 protein (No. 11 in Table A.1), which has the minimal MW among all XPB/p52/p8 proteins I could obtain. Tiny crystals of WT XPB (residues 30-160)/p52/p8 appeared in several crystallization conditions (Figure. A.3A-B) containing similar precipitating reagents such as PEG200, PEG300 and PEG400. Extensive refinements including crystallization condition optimization, additive screening and micro-seeding (Figure. A.3C) did not further improve the size and quality of crystals, which only diffracted to ~20 Å by an in-house X-ray source. To further enhance the crystallization results, I applied the surface engineering strategy on XPB and p52 proteins (Table A.2). I selectively mutated the residues with flexible side chains like K/R/E/Q/D to alanines according to SER server [14]. Among them, only XPB (residues 30-160)/p52 (K140A+E189A)/p8 could substantially increase the sizes of crystals (Figure. A.3D). Nonetheless, the diffraction resolutions still had huge limitations which are between 12 and 15 Å. In addition, a new cryo-EM structure of TFIIH [15] was reported in 2019 while my refinements of these crystals were still ongoing. The new cryo-EM structure of TFIIH was determined at 3.7 Å resolution and revealed how XPB-NTD interacts with p52. The author assigns XPB residues 44–160 and residues 165–300 to the NTD and DRD domain, respectively, consistent with my co-expression results in vitro.

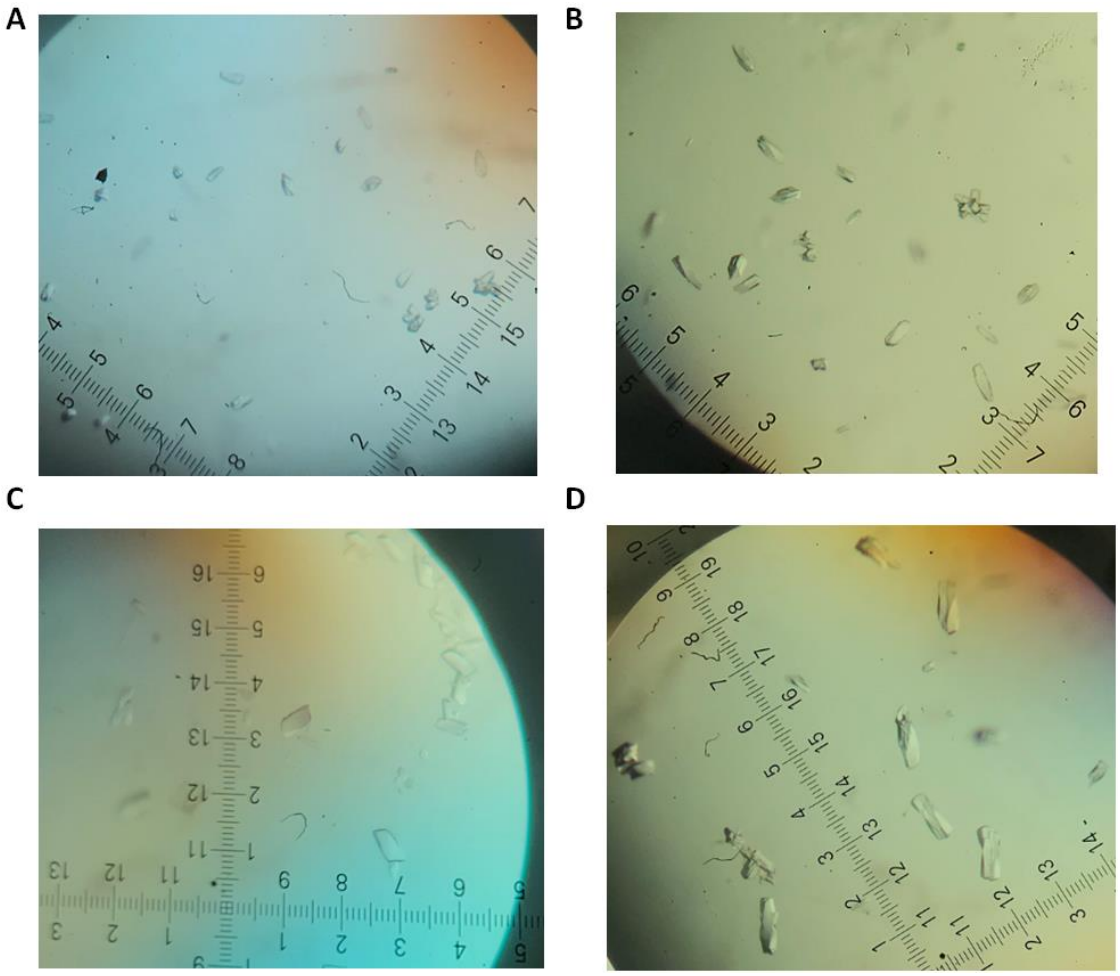

Figure A.3 Crystallization of the XPB-NTD/p52/p8 complexes.

(A) Crystals of WT XPB (residues 30-160)/p52/p8 grew in 100 mM Tris-Cl, pH 7.0, 0.2 M Sodium acetate, 8% PEG 200.

(B) Crystals of WT XPB (residues 30-160)/p52/p8 grew in 100 mM Bis-tris propane, pH 6.35, 0.2 M Ammonium acetate, 7% PEG 400.

(C) Crystals of WT XPB (residues 30-160)/p52/p8 grew in the same condition as in **(**B) with addition of 100 mM Guanidine hydrochloride.

(D) Crystals of surface-engineered XPB (residues 30-160)/p52 (K140A+E189A)/p8 100 mM Bis-tris propane, pH 7.1, 0.2 M Ammonium acetate, 100 mM NaF, 7% PEG 400.

Table A.2 Surface-engineered XPB and p52 constructs for XPB (30-160)/p52/p8 co-expression and crystallization

Combination	XPB (30-160) construct	p52 construct	Result
1	WT	K140A	Y
2	WT	E189A	Y
3	WT	K140A/E189A	Y
4	WT	K378A/Q379A	N
5	WT	R266S/K267S	N
6	WT	E63D/Q64H	N
7	K44S	K140A/E189A	Y
8	E48A	K140A/E189A	N
9	K68R/D69P	K140A/E189A	Y
10	Q132T	K140A/E189A	N

Mutation sites in XPB and p52 in each co-expression trial are indicated. WT-p8 construct was used in each co-expression test. Y means the co-expression worked well and the expressed proteins were soluble while N means the co-expression was not working (extremely low expression level or proteins were insoluble).

Biochemical impact of XPB disease –causing mutations F99S and T119P

Next I tested how the two human disease mutations F99S and T119P in XPB would affect the co-expression of XPB with p52/p8. Remarkably, the XPB-NTD F99S mutant could be stably co-expressed and form soluble complex with p52/p8 only with a slightly decreased expression level (~80% compared to WT) whereas the XPB T119P mutant was almost insoluble (Figure. A.4A-B). Since XPB-NTD does not directly interact with p8 [3], these observations indicate F99S mutation slightly weakened the interaction between XPB-NTD (30-160) and p52 while T119P mutation almost abolished these interactions that are critical for the correct folding or the solubility of XPB-NTD. However, I could successfully co-express and purify full length human XPB F99S and T119P mutants with p52/p8 from insect cells (data not shown), suggesting these two mutations have limited impact on interactions between full length XPB and p52.

Combined with previous data [7], my biochemical results could further explain how these two mutations lead to human diseases. We have shown F99S mutation did not significantly weaken the interaction between human XPB and p52 but this mutation impaired the ATPase activity and damaged DNA opening by XPB [7], suggesting this mutation may induce a local conformational change in XPB to lower its DNA binding ability and thus lead to human XP/CS disease. On the other hand, T119P mutation severely affects the stability of the XPB N-terminal or its interaction with p52/p8. This impairment caused by T119P in XPD-NTD should be considerably diminished by the interactions between XPB-CTD and p52/p8 since this mutation only leads to mild disease symptoms for TTD patients.

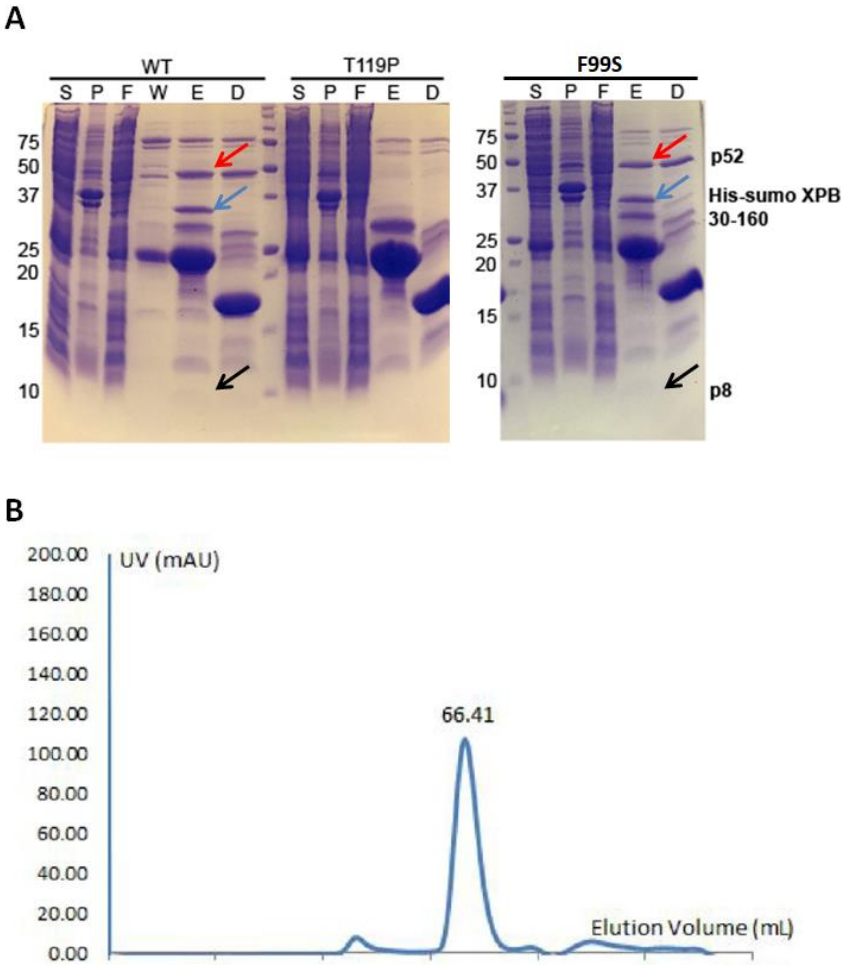

Figure A.4 T119P disrupts interactions of XPB-NTD with p52/p8 while F99S does not.

(A) SDS PAGE gel shows comparison of XPB-NTD with/without mutation F99S or T119P interacting with p52/p8. Human XPB-NTD (WT) and mutant F99S or T119P was fused with SUMO tag and co-expressed with p52/p8 in bacteria. Cell extract (S) was purified by Ni-NTA column. P: cell pellet, F: flow through, W:wash, E: elution, D: protease digestion to remove SUMO tag. Red arrow: p52, blue arrow: XPB, black arrow: p8.

(B) S200 gel-filtration chromatography profile of the XPB (30-160, F99S)/p52/p8 complex.

DISCUSSION

Since the discovery of XPB many years ago, it remains challenging to obtain active human XPB proteins in vitro. Members in our lab successfully purified the CTD of human XPB and solved the crystal structure [2]. I have established the methods of expressing and purifying XPB N terminal domain with p52/p8 in *E.coli* and full length XPB with p52/p8 in insect cells. According to my knowledge, I for the first time mapped out the minimal region (residues 30-160) in human XPB that interacts with p52/p8 by biochemical experiments. There had been some progress in the crystallization trials of human XPB-NTD/p52/p8 but the crystals diffracted poorly. Also a new cryo-EM structure of TFIIH [15] was determined at 3.7 Å resolution in 2019 and reveals the molecular basis of XPB-NTD interacting with p52. This cryo-EM structure shows that F99 is located in a conserved hydrophobic pocket and the side chain of T119 points towards the solvent [15], suggesting T119 may be more crucial for the solubility of XPB-NTD. This observation is highly in agreement with my co-expression data that XPB-NTD containing T119P could barely be co-expressed with p52/p8 due to its lower solubility or less efficient domain-folding. This could also explain the lower expression levels of TFIIH in TTD patients caused by T119P [16] but this mutation will not significantly affect the function of XPB in NER [7]. Since F99S mutation seems to have less impact on the solubility of XPB-NTD as shown by our co-expression data, this F99 residue is more likely to cause local changes in the hydrophobic pocket of XPB-NTD and its neighboring region in TFIIH, which could impair the ATPase activity of XPB [7], and induce defect in nucleotide excision repair [17].

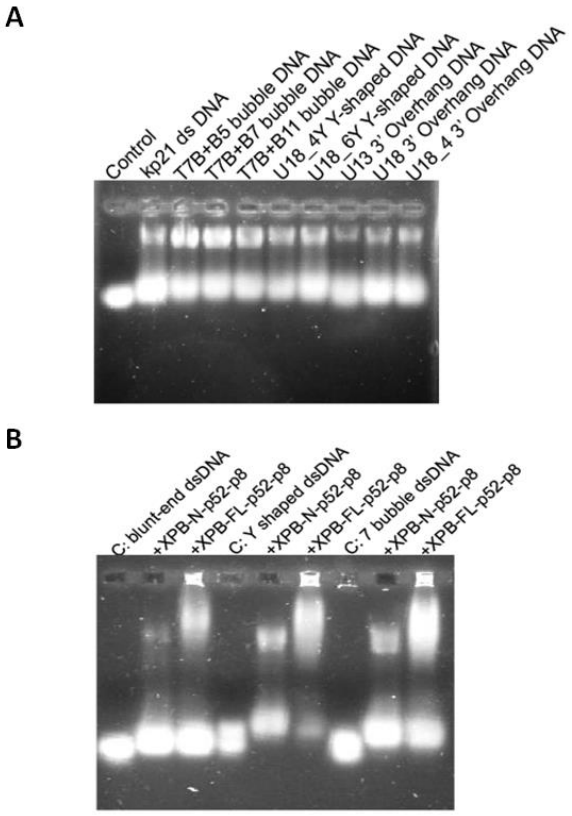

Figure A.5 DNA binding of XPB/p52/p8 complexes analyzed by electrophoretic mobility shift assay (EMSA).

(A) DNA binding of XPB (54-312 aa)/p52/p8 using different DNA substrates including dsDNA, bubble DNA, Y-shaped DNA (forked), and 3'-overhang DNA.

(B) Comparison of full length XPB/p52/p8 and XPB (54-312 aa) /p52/p8 on binding DNA substrates including dsDNA, bubble DNA, Y-shaped DNA. Protein and DNA with equal molar ratio were used for EMSA. Experiments were repeated at least twice with consistent results.

The DNA damage recognition domain (DRD) of XPB was firstly discovered in AfXPB by Fan *et al.* [18], which shares structural similarity with the mismatch recognition domain of MutS involved in the mismatch repair pathway [19]. Biochemical assays suggested that the DRD domain plays a role in binding damaged DNA [18, 20]. Sequence alignment shows that this DRD domain in archaeal XPB might correspond to residues 243-301 in human XPB, but it remains elusive about the function of the DRD domain of human XPB. Research on human XPB or the DRD domain was largely impeded due to the difficulties in generating soluble proteins of human XPB or XPB N terminal domain in vitro. Here I was able to obtain several XPB (NTD+DRD) proteins including XPB (residues 54-312), XPB (residues 65-312), and XPB (residues 73-312) in complex with p52/p8. Interestingly, XPB (residues 54-312) in complex with p52/p8 was capable of binding different types of DNA substrates as shown in the EMSA assay (Figure. A.5A). This result suggests the DRD domain of human XPB may also play a role in DNA binding activities of human XPB during NER even though the DNA affinities of XPB (NTD+DRD)/p52/p8 are weaker in comparison to the full-length XPB/p52/p8 complex (Figure. A.5B). The importance of the DRD domain was supported by biochemical and structural evidence that deletion of the DRD was lethal in yeast [21] and the DRD of human XPB was shown to directly contact XPD [15]. However, the cryo-EM structures of yeast and human TFIIH in complex with DNA [4, 22] have shown that only the two RecA-like domains (HD1 and HD2) of XPB are involved in dsDNA binding. Thus it is still unclear how the DRD could contribute to the function of XPB and such inquiry could be further aided by future studies of human XPB.

REFERENCES

1. Gerard M, Fischer L, Moncollin V, Chipoulet JM, Chambon P, Egly JM: **Purification and interaction properties of the human RNA polymerase B (II) general transcription factor BTF2**. *Journal of Biological Chemistry* 1991, **266**(31):20940-20945.

2. 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 Crystallographica Section D: Biological Crystallography* 2013, **69**(2):237-246.

3. Greber BJ, Nguyen THD, Fang J, Afonine PV, Adams PD, Nogales E: **The cryo-electron microscopy structure of human transcription factor IIH**. *Nature* 2017, **549**(7672):414-417.

4. Schilbach S, Hantsche M, Tegunov D, Dienemann C, Wigge C, Urlaub H, Cramer P: **Structures of transcription pre-initiation complex with TFIIH and Mediator**. *Nature* 2017, **551**(7679):204-209.

5. Cleaver JE, Thompson LH, Richardson AS, States JC: **A summary of mutations in the UV‐sensitive disorders: xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy**. *Human mutation* 1999, **14**(1):9-22.

6. Oh KS, Khan SG, Jaspers N, Raams A, Ueda T, Lehmann A, Friedmann PS, Emmert S, Gratchev A, Lachlan K: **Phenotypic heterogeneity in the XPB DNA helicase gene (ERCC3): xeroderma pigmentosum without and with Cockayne syndrome**. *Human mutation* 2006, **27**(11):1092-1103.

7. Coin F, Oksenych V, Egly J-M: **Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair**. *Molecular cell* 2007, **26**(2):245-256.

8. Weeda G, van Ham RC, Vermeulen W, Bootsma D, van der Eb AJ, Hoeijmakers JH: **A presumed DNA helicase encoded by ERCC-3 is involved in the human repair disorders xeroderma pigmentosum and Cockayne's syndrome**. *Cell* 1990, **62**(4):777-791.

9. Gradia SD, Ishida JP, Tsai M-S, Jeans C, Tainer JA, Fuss JO: **MacroBac: new technologies for robust and efficient large-scale production of recombinant multiprotein complexes**. *Methods in enzymology* 2017, **592**:1-26.

10. Coin F, De Santis LP, Nardo T, Zlobinskaya O, Stefanini M, Egly J-M: **p8/TTD-A as a repair-specific TFIIH subunit**. *Molecular cell* 2006, **21**(2):215-226.

11. Jawhari A, Lain éJ-P, Dubaele S, Lamour V, Poterszman A, Coin F, Moras D, Egly J-M: **p52 Mediates XPB function within the transcription/repair factor TFIIH**. *Journal of Biological Chemistry* 2002, **277**(35):31761-31767.

12. Zhou Y, Kou H, Wang Z: **Tfb5 interacts with Tfb2 and facilitates nucleotide excision repair in yeast**. *Nucleic acids research* 2007, **35**(3):861-871.

13. Kainov DE, Vitorino M, Cavarelli J, Poterszman A, Egly J-M: **Structural basis for group A trichothiodystrophy**. *Nature structural & molecular biology* 2008, **15**(9):980-984.

14. Goldschmidt L, Cooper DR, Derewenda ZS, Eisenberg D: **Toward rational protein crystallization: A Web server for the design of crystallizable protein variants**. *Protein Science* 2007, **16**(8):1569-1576.

15. Greber BJ, Toso DB, Fang J, Nogales E: **The complete structure of the human TFIIH core complex**. *Elife* 2019, **8**:e44771.

16. Botta E, Nardo T, Lehmann AR, Egly J-M, Pedrini AM, Stefanini M: **Reduced level of the repair/transcription factor TFIIH in trichothiodystrophy**. *Human molecular genetics* 2002, **11**(23):2919-2928.

17. Riou L, Zeng L, Chevallier-Lagente O, Stary A, Nikaido O, Ta ëb A, Weeda G, Mezzina M, Sarasin A: **The relative expression of mutated XPB genes results in xeroderma pigmentosum/Cockayne's syndrome or trichothiodystrophy cellular phenotypes**. *Human molecular genetics* 1999, **8**(6):1125-1133.

18. Fan L, Arvai AS, Cooper PK, Iwai S, Hanaoka F, Tainer JA: **Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair**. *Molecular cell* 2006, **22**(1):27-37.

19. Obmolova G, Ban C, Hsieh P, Yang W: **Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA**. *nature* 2000, **407**(6805):703-710.

20. Rouillon C, White MF: **The XBP-Bax1 helicase-nuclease complex unwinds and cleaves DNA: implications for eukaryal and archaeal nucleotide excision repair**. *Journal of Biological Chemistry* 2010, **285**(14):11013-11022.

21. Warfield L, Luo J, Ranish J, Hahn S: **Function of conserved topological regions within the Saccharomyces cerevisiae basal transcription factor TFIIH**. *Molecular and cellular biology* 2016, **36**(19):2464-2475.

22. Kokic G, Chernev A, Tegunov D, Dienemann C, Urlaub H, Cramer P: **Structural basis of TFIIH activation for nucleotide excision repair**. *Nature communications* 2019, **10**(1):1-9.