

# Structure-function analysis of human nucleotide excision DNA repair

Dissertation  
for the award of the degree  
**“Doctor rerum naturalium”**  
of the Georg-August-Universität Göttingen

within the doctoral program IMPRS Molecular Biology  
of the Georg-August University School of Science (GAUSS)

submitted by

**Goran Kokic**

from Zagreb, Croatia  
Göttingen 2019



## **MEMBERS OF THE THESIS COMMITTEE AND THE EXAMINATION BOARD**

Prof. Dr. Patrick Cramer (1<sup>st</sup> Referee)

Department of Molecular Biology

Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

Prof. Dr. Holger Stark (2<sup>nd</sup> Referee)

Department of Structural Dynamics

Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

Prof. Dr. Claudia Höbartner

Institute for Organic Chemistry

University of Würzburg, Germany

Dr. Vladimir Pena

Macromolecular Crystallography Group

Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

## **OTHER MEMBERS OF THE EXAMINATION BOARD**

Prof. Dr. Marina Rodnina

Department of Physical Biochemistry

Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

Prof. Dr. Henning Urlaub

Bioanalytical Mass Spectrometry Group

Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

Date of oral examination: 10.04.2019.

# **AFFIDAVIT**

Herewith I declare, that I prepared the Doctoral Thesis “Structure-function analysis of human nucleotide excision DNA repair” on my own and with no other sources and aids than quoted.

Goran Kokic

February, 2019

Göttingen, Germany

# PUBLICATIONS

Parts of this work have been published (section 2.2. and associated parts of ‘Discussion and Outlook’ and ‘Materials and Methods’).

**G. Kokic**, A. Chernev, D. Tegunov, C. Dienemann, H. Urlaub and P. Cramer (2019) Structural basis of TFIIH activation for nucleotide excision repair. *Nat Commun* **10**, 2885 (2019).

Author contributions: G.K. designed and carried out all experiments except for crosslinking mass spectrometry, which was carried out by A.C. D.T. assisted with image processing and C.D. assisted with cryo-EM data acquisition and model building. H.U. supervised mass spectrometry. P.C. designed and supervised research. G.K. and P.C. interpreted the data and wrote the manuscript, with input from all authors.

M. Boehning, C. Dugast-Darzacq, M. Rankovic, A.S. Hansen, T. Yu, H. Marie-Nelly, D. McSwiggen, **G. Kokic**, G.M. Dailey, P. Cramer, X. Darzacq and M. Zweckstetter (2018) RNA polymerase II clustering through carboxy-terminal domain phase separation, *Nat Struct Mol Biol.*, **25**, 833-840

Author contributions: M.B. designed experiments, generated constructs, and prepared proteins unless otherwise noted. C.D.-D. designed experiments, established and characterized the RPB1 cell lines, and performed and analyzed the in vivo FRAP and SPT experiments. M.R. designed experiments, performed all phase separation assays, DIC and fluorescence microscopy, in vitro FRAP measurements and data analysis. A.S.H. designed, performed, and analyzed SPT experiments and helped with the in vivo FRAP analysis. H.M.-N. designed, performed, and analyzed 3D-PALM experiments. D.T.McS. performed cell-viability experiments and helped in performing 3D-PALM experiments. G.M.D. designed and cloned the different RPB1 expression vectors. G.K. prepared human TFIIH kinase complex. T.Y. performed CD and NMR experiments. C.D.-D., X.D., P.C., and M.Z. designed and supervised research. M.B., M.R., C.D.-D., P.C., X.D., and M.Z. prepared the manuscript with input from all authors

# TABLE OF CONTENTS

<b>MEMBERS OF THE THESIS COMMITTEE AND THE EXAMINATION BOARD</b>	<b>II</b>
<b>AFFIDAVIT</b>	<b>III</b>
<b>PUBLICATIONS</b>	<b>IV</b>
<b>SUMMARY</b>	<b>1</b>
<b>1. INTRODUCTION</b>	<b>2</b>
1.1 Overview of DNA repair	2
1.2 Nucleotide excision DNA repair (NER)	4
1.2.1 Global-genome nucleotide excision repair (GGR)	5
1.2.2 Transcription-coupled nucleotide excision repair (TCR)	9
1.3 General transcription factor IIIH (TFIIH) is a central NER factor	13
1.4 Structural studies of NER	14
1.5 Aims and scope of this work	17
<b>2. RESULTS</b>	<b>18</b>
2.1 Reconstitution and analysis of GGR initiation	18
2.2 Structural basis of TFIIH activation for NER	23
2.2.1 Biochemical characterization of core TFIIH ATPases	23
2.2.2 Core TFIIH-XPA-XPG-DNA complex formation and cryo-EM structure determination	26
2.2.3 Structural rearrangements in TFIIH upon activation for DNA repair	32
2.2.4 XPA interactions with core TFIIH and DNA	33
2.2.5 XPD-DNA interactions	35
2.2.6 XPD inhibition by the kinase module and activation by XPA	36
2.2.7 XPG binding to the core TFIIH-XPA-DNA complex	40
2.3 Reconstitution and analysis of TCR intermediates	42
2.3.1 Biochemical characterization of TCR initiation	42
2.3.2 The fate of lesion-arrested RNAP	47
2.3.3 Active RNAP backtracking exposes DNA lesions to repair	53
<b>3. DISCUSSION AND OUTLOOK</b>	<b>56</b>
3.1 First structure of an NER intermediate	56
3.2 Extended model of the human TFIIH core and comparison to other TFIIH structures	58
3.3 Structural etiology of Xeroderma pigmentosum and Trichothiodystrophy	60

3.4 Role of XPA in NER and activation of TFIIH for repair .....	63
3.5 Role of XPB in NER and the problem of repair bubble opening.....	66
3.6 Positive and negative regulation of XPD helicase activity.....	69
3.7 Novel insights into the mechanism of TCR and a first case of active RNA- polymerase backtracking in eukaryotes.....	69
3.8 Towards a complete molecular mechanism of human NER .....	73
<b>4. MATERIALS AND METHODS.....</b>	<b>75</b>
4.1 Chemicals and consumables.....	75
4.2 Buffers and growth media .....	77
4.3 Bacterial and insect cell strains .....	78
4.4 Plasmids.....	79
4.5 Gel electrophoresis .....	80
4.5.1 Agarose gel electrophoresis.....	80
4.5.2 Sodium-dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE).....	80
4.5.3 Denaturing urea polyacrylamide gel electrophoresis .....	81
4.6 Cloning .....	81
4.6.1 Polymerase chain reaction (PCR).....	81
4.6.2 Ligation-independent cloning (LIC).....	83
4.6.3 Assembly of multiple genes into a single vector by LIC reactions.....	84
4.6.4 Round-the-horn side-directed mutagenesis .....	84
4.6.5 Transformation of chemically competent cells .....	85
4.6.6 Isolation of vector DNA from bacterial cultures.....	85
4.7 Protein Expression.....	86
4.7.1 Protein expression in <i>E. coli</i> .....	86
4.7.2 Protein expression in insect cells.....	86
4.7.3 Transformation of electrocompetent DH10EMBacY cells .....	86
4.7.4 Isolation of bacmid DNA .....	87
4.7.5 Transfection of Sf9 cells with bacmid DNA and V <sub>0</sub> virus production.....	87
4.7.6 V <sub>1</sub> virus production.....	88
4.7.7 Protein expression (V <sub>2</sub> virus production) .....	88
4.8 Protein purification.....	89
4.8.1 TFIIS and TFIIS:D282AE283A inactive mutant .....	89
4.8.2 Core TFIIH and core TFIIH containing XPD:K48R point mutant .....	90
4.8.3 CsB, UVSSA, Usp7, XPA, XPG, XPF-ERCC1, XPC-Rad23b.....	91

4.8.4 CsA-DDB1 .....	92
4.8.5 Protein quantification .....	92
4.9 DNA and RNA constructs .....	93
4.10 Biochemical assays.....	95
4.10.1 Analytical gel filtration.....	95
4.10.2 Fluorescence anisotropy .....	95
4.10.3 Bulk ATPase assay .....	96
4.10.4 Helicase assay.....	96
4.10.5 Translocase assay .....	97
4.10.6 Transcription assay .....	97
4.10.7 Backtracking assay .....	98
4.10.8 Bead-based RNAP assays.....	99
4.10.9 Kinase activity assay .....	100
4.11 Synthesis of a CPD lesion and preparation of the CPD-containing DNA.....	101
4.12 Mass-spectrometric identification of crosslinking sites .....	102
4.12.1 Sample preparation .....	102
4.12.2 Mass spectrometry .....	102
4.13 Negative stain electron microscopy of the XPC-Rad23b-core TFIIH-XPA-DNA complex .....	103
4.13.1 Complex formation.....	103
4.13.2 Negative staining and electron microscopy.....	103
4.14 TFIIH-XPA-DNA structure determination .....	104
4.14.1 Sample preparation.....	104
4.14.2 Cryo-electron microscopy and image processing.....	105
4.14.3 Model building .....	105
<b>5. SUPPLEMENTARY MATERIALS .....</b>	<b>107</b>
<b>6. REFERENCES .....</b>	<b>122</b>
<b>ACKNOWLEDGMENTS.....</b>	<b>139</b>
<b>List of abbreviations.....</b>	<b>140</b>
<b>List of figures .....</b>	<b>141</b>
<b>List of tables .....</b>	<b>142</b>
<b>Curriculum vitae.....</b>	<b>143</b>

## SUMMARY

An average human cell accumulates tens of thousands of DNA lesions per day. Therefore, DNA repair pathways are set up to continuously proofread the DNA and correct DNA lesions, thus ensuring accurate expression and propagation of genetic material. Nucleotide excision DNA repair (NER) is a major DNA repair pathway which removes UV-induced lesions and bulky DNA modifications. Defects in NER promote cancer development and cause human diseases, such as Xeroderma pigmentosum (XP), Cockayne syndrome and trichothiodystrophy. During NER, lesions are cut out of the DNA as part of a short oligonucleotide and the resulting gap is filled by DNA synthesis using the nondamaged DNA strand as a template. The NER machinery assembles around the heterodecameric transcription factor IIIH (TFIIH), comprised of the core module and the kinase module. The TFIIH core module utilizes the ATP-consuming subunits XPB and XPD to open the DNA repair bubble, scan for the lesion and coordinate the excision of the damaged DNA. However, due to the complete lack of structural information on NER assembly intermediates and difficulties in preparing the NER complexes for in vitro analysis, the molecular mechanism of NER is still not well understood.

Here we prepare human TFIIH and other NER factors involved in DNA excision. We reconstitute several steps of the NER pathway and analyze the trapped intermediates with biochemical assays, cross-linking mass-spectrometry and electron microscopy (EM). We systematically dissect the regulation of the TFIIH ATPases XPB and XPD and show that the additional NER factors XPA and XPG stimulate the enzymatic activities of the ATPases. We report the core TFIIH-XPA-DNA structure at 3.6Å resolution, which represents the lesion scanning NER intermediate, and we map the position of XPG within the complex by chemical crosslinking. The structure expands the previous model for the TFIIH core and explains many disease mutations. The structure further elucidates the topology of NER factors around the 5' edge of the repair bubble: XPB binds the DNA duplex, XPD binds the 3' single strand extension and XPA wraps around the duplex single-strand junction and bridges the ATPases. XPA and XPB form a DNA duplex tunnel which anchors the NER machinery to the DNA. Our biochemical analysis and comparison to previous structures reveal how XPA and XPG activate TFIIH for repair. The TFIIH kinase module inhibits NER by repressing the XPD helicase activity. XPA stabilizes a completely novel TFIIH conformation in which the ATPases are dramatically reoriented, which displaces the TFIIH kinase module and removes the “plug” element from the DNA-binding pore in XPD. This allows XPD to move by ~80Å, engage the DNA and rapidly scan for the lesion. XPG facilitates lesion scanning by directly stimulating XPD migration on DNA and by sequestering the kinase module binding site on XPD. The results presented here greatly extend our understanding of human NER and provide the basis for future structure-function analysis of this repair pathway, also in the context of transcription.



# 1. INTRODUCTION

## 1.1 Overview of DNA repair

DNA integrity is constantly threatened by exogenous and endogenous sources of DNA damage resulting in tens of thousands of DNA injuries per human cell per day<sup>1</sup>. DNA damage encompasses a variety of covalent modifications in DNA bases and non-covalent alterations in DNA, such as base-pairing mismatches, bubbles and loops<sup>2</sup>. These unfavorable changes in the DNA interfere with a correct interpretation and propagation of the genetic material, which often leads to disease development<sup>3-5</sup>. To counteract the high load of DNA damage, cells are equipped with different DNA repair systems, each directed towards a different subset of DNA lesions. Thus, DNA repair can be sub-classified into several pathways: direct DNA repair, DNA mismatch repair, base excision repair, nucleotide excision repair and double-strand break repair<sup>2</sup>.

Direct DNA repair is a simple and error-free repair pathway in which a DNA lesion is directly restored to the original DNA sequence<sup>6</sup>. However, the repair pathway focuses on a narrow set of DNA lesions which include the UV-light induced photolesions, O- and N-alkylated DNA<sup>6</sup>. Photolyase and O<sup>6</sup>-alkylguanine DNA alkyltransferase are the best studied enzymes involved in direct DNA repair. Photolyase splits the UV-induced pyrimidine dimers in an electron-transfer reaction driven by blue light<sup>7</sup>. Placental mammals lack the enzyme even though it is wide-spread in all kingdoms of life<sup>2</sup>. O<sup>6</sup>-alkylguanine DNA alkyltransferase recognizes a highly mutagenic O<sup>6</sup>-methyl guanine<sup>8</sup> and transfers the O<sup>6</sup>-methyl group to its active site cysteine<sup>9</sup>. The reaction is suicidal because the modified cysteine cannot be regenerated to participate in further rounds of catalysis<sup>9</sup>.

DNA mismatch repair (MMR) removes single-nucleotide mismatches and DNA loops created by errors in DNA replication<sup>10,11</sup>. Thus, the MMR machinery must be able to differentiate between the parental DNA strand and the newly replicated DNA strand which contains the incorrect DNA sequence. In bacteria this differentiation is enabled by DNA modifications, as the parental DNA strand is methylated and the newly synthesized DNA is unmethylated<sup>12</sup>. During bacterial MMR MutS recognizes the mismatches in DNA<sup>13</sup> with the help of MutL<sup>14</sup>, followed by the recruitment of the MutH endonuclease<sup>15</sup>. MutH binds the hemi-methylated DNA duplex and cleaves the unmethylated DNA strand to initiate the degradation of a newly synthesized DNA strand in the 5' or 3' direction depending on the position of the cleavage site<sup>10,11</sup>. The resulting gap in the DNA is filled by a high-fidelity DNA polymerase and sealed by a DNA ligase. The overall mechanism is highly conserved from bacteria to humans<sup>10,11</sup>. However, it is still unclear how the eukaryotic MMR discriminates between the parental and daughter DNA strands.

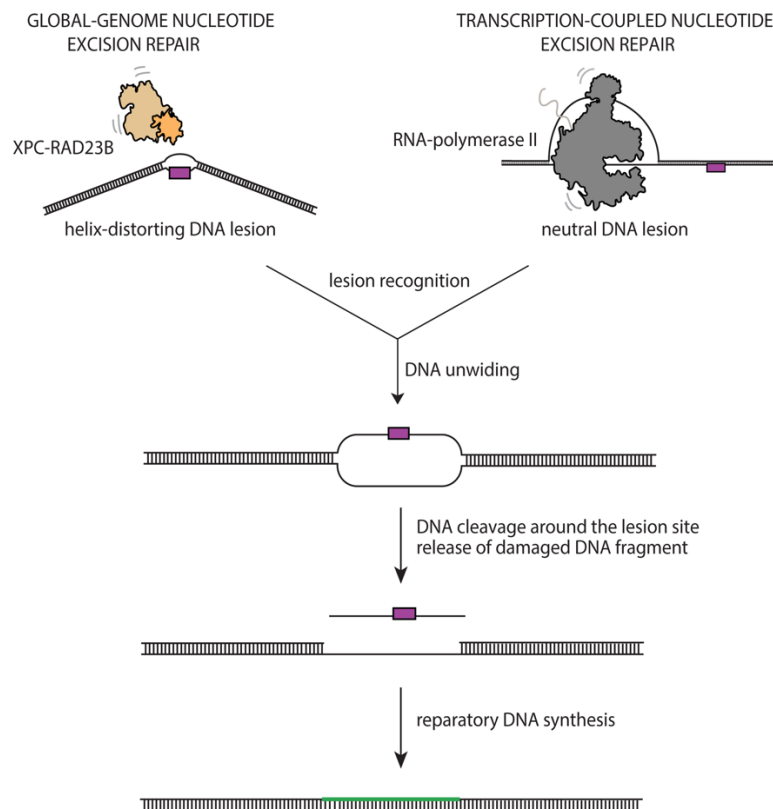
Base excision repair (BER) targets small modifications of DNA bases such as deamination, oxidation and methylation<sup>16</sup>. The repair reaction is mediated by glycosylases which can identify a damaged DNA base within the overwhelming amount of undamaged substrate and cleave off the modified base leaving an abasic site (AP-site)<sup>2</sup>. In mammals, 11 different glycosylases were identified thus far, each recognizing a subset of similar lesions<sup>17</sup>. Uracil-DNA glycosylases, a well-studied group of glycosylases, are used to remove a highly mutagenic uracil generated by cytosine deamination<sup>18</sup>. The AP-site is further processed by an AP-endonuclease which cleaves the sugar-phosphate DNA backbone, a DNA-polymerase which removes the abasic sugar and fills in the single nucleotide gap and a DNA ligase which seals the nick in the DNA<sup>16</sup>.

Nucleotide excision repair (NER) is the main repair pathway for the removal of bulky DNA modifications caused by irradiation, exogenous and endogenous mutagens, and chemotherapeutics<sup>19,20</sup>. NER can remove the biggest variety of chemically and structurally unrelated DNA lesions<sup>21-24</sup>, because the NER machinery recognizes the lesion-induced distortion in the DNA secondary structure, rather than the lesion itself<sup>24-26</sup>. NER is a complex biochemical process which utilizes four different factors in bacteria (UvrA-UvrD)<sup>27</sup> and six factors in humans (XPA-XPG) to perform a controlled sequence of several enzymatic steps<sup>19,20</sup>. After the recognition of a DNA lesion, the DNA is unwound around the lesion site and the damaged DNA strand is incised on both sides of the lesion, resulting in the release of a lesion-containing oligonucleotide from the DNA<sup>19,20</sup>. The resulting gap is filled by a DNA polymerase using the healthy DNA strand as a template for reparatory DNA synthesis<sup>19,20</sup>.

Double-strand break (DSB) repair is required when both DNA strands are broken, mostly due to reactive oxygen species and ionizing radiation<sup>28-30</sup>. DSBs also underlie some biochemical processes in the cell, such as V(D)J recombination in the immune system maturation, replication fork collapses and topoisomerase failures<sup>28-30</sup>. There are two main strategies for repairing DSBs: the non-homologous end joining (NHEJ) and the homologous recombination (HR)<sup>28-30</sup>. NHEJ facilitates a direct ligation of the two DNA ends<sup>31,32</sup>. Since NHEJ is template-independent, this solution is error-prone and it often leads to deletions, insertions or even chromosomal rearrangements. In contrast, HR is a template-dependent and mostly error-free way to repair DSBs<sup>33,34</sup>. A free DNA end is first resected to produce a single-stranded overhang, which invades a homologous DNA and uses it as a template for repair<sup>33,34</sup>. Thus, this repair strategy is only feasible in G and S2 phases of the cell cycle when the sister chromatid is available to serve as a homologous DNA template<sup>33,34</sup>.

## 1.2 Nucleotide excision DNA repair (NER)

NER was discovered in 1964 when several groups observed that DNA lesion-containing oligonucleotides are cut out of the DNA and the resulting gaps in the DNA are filled in by the reparatory DNA synthesis<sup>35-38</sup>. The inability to perform such a DNA correction was linked to the development of the genetic disease Xeroderma pigmentosum (XP)<sup>3,39</sup>, characterized by a high incidence of skin cancer<sup>40</sup>. Genetic complementation studies on patient cell lines revealed that the underlying core DNA repair machinery is comprised of 7 main factors named XPA, XPB, XPC, XPD, XPE, XPF and XPG<sup>41,42</sup>. A major milestone in NER research was the reconstitution of the NER reaction in vitro with purified components<sup>43-45</sup> which defined a minimal set of factors required for a complete NER (listed in Table1), as well as the main sequence of events accompanying the repair process (Fig. 1). Two different strategies are employed for the initial lesion recognition which divides NER into 2 sub-pathways: in global-genome NER (GGR) the XPC-Rad23b complex recognizes the lesion induced distortions in DNA structure and in transcription-coupled NER (TCR) the lesion arrested RNA-polymerase II (RNAP) elicits the repair<sup>19,20</sup> (Fig. 1).



**Figure 1 | The main steps of nucleotide excision DNA repair (NER).**

XPC-Rad23B recognizes the helix-distorting lesions<sup>24,46-48</sup> and the elongating RNA-polymerase II recognizes the transcription blocking lesions<sup>49-51</sup> (lesions are indicated by the purple square). After the lesion recognition, the two sub-pathways converge into a conserved series of reactions which include the DNA unwinding, DNA incisions around the lesion site and the removal of the damaged DNA single-strand<sup>19,20</sup>. The resulting gap in DNA is filled by DNA synthesis templated by the undamaged DNA strand (the newly synthesized DNA is shown in green)<sup>19,20</sup>.

### 1.2.1 Global-genome nucleotide excision repair (GGR)

Efficient substrate for GGR contains the two following structural features: the disruption of base pairing and the presence of a bulky DNA modification<sup>24,52,53</sup> (Fig. 2). Disruption of base pairing is recognized by the main GGR lesion sensor **XPC**<sup>24,46-48</sup> assisted by Rad23B<sup>54</sup> and centrin2<sup>55</sup>. XPC does not interact with DNA lesions directly<sup>26</sup>, which explains how XPC initiates repair on a diverse set of structurally unrelated DNA lesions. Thus, the physical presence of a bulky DNA modification in the distorted DNA recognized by XPC must be additionally verified by a downstream GGR component which is known as bipartite lesion recognition<sup>56</sup>. UV-induced cyclobutane pyrimidine dimers (CPDs) are efficiently repaired by GGR but are poor substrates for XPC because these lesions do not significantly melt the DNA<sup>24</sup>. However, an additional DNA lesion sensor comprised of DDB1 and DDB2 (XPE) binds CPDs with a very high affinity, bends the DNA and facilitates the recruitment of XPC to the lesion site<sup>57-60</sup>.

Following the recognition of damaged DNA, XPC recruits **transcription factor IIIH (TFIIH)** to the lesion site<sup>61-63</sup> (Fig. 2). TFIIH is a heterodecameric complex comprised of two modules, the core and the kinase module<sup>64</sup>. The TFIIH core contains two SF2-type DNA helicases XPB and XPD which are essential for repair<sup>64</sup>. TFIIH recruitment allows XPB to engage the DNA and to open the DNA around the lesion, thereby creating a DNA repair bubble<sup>65-68</sup>. XPB was shown to have a 3'-5' helicase<sup>67</sup> and a 5'-3' translocase activity<sup>69</sup>, however, it is still unclear how these activities are employed for the initial repair bubble opening. It was also reported that only the ATPase, and not the helicase activity of XPB is required for DNA melting<sup>68</sup>. This suggested that the ATP-hydrolysis by XPB is not used for DNA unwinding but rather to fuel the large scale conformational changes in XPB which can lead to twisting of DNA and DNA melting<sup>66</sup>. The yeast XPB homologue, Ssl2, in the context of TFIIH acts as a DNA translocase, rather than a helicase<sup>69</sup>. Thus, XPB may open the repair bubble in analogy to its role in promoter opening during transcription initiation<sup>69,70</sup>. The repair bubble opening creates a binding substrate for the XPD helicase that unwinds DNA in 5'-3' direction<sup>71-73</sup>. XPD most likely verifies the presence of a lesion because the movement of XPD on DNA is arrested by lesions<sup>23,73-75</sup>. In addition, XPC orients TFIIH during the initial stages of NER so that XPD specifically scans the damaged DNA strand<sup>56,73</sup>. It is still uncertain how XPD recognizes DNA lesions. The structure of the archaeal XPD homologue has revealed a narrow pore formed by the accessory domains inserted into the helicase lobe of XPD<sup>76,77</sup>. The dimensions of the pore would allow the passage of single stranded DNA, but would restrict the passage of a DNA duplex. The DNA trajectory was proposed to lead towards the pore, which could be used to sift the DNA single strand for bulky modifications<sup>66,78</sup>. In addition, XPD contains a redox reactive iron-sulfur (FeS) cluster situated next to the narrow pore<sup>66,78</sup>. The FeS cluster is also present in other DNA repair proteins, such as the *E.coli* endonuclease III and the mismatch specific adenine glycosylase MutY, where the cluster is involved in DNA lesion processing<sup>79</sup>. Moreover, the FeS cluster

can participate in an electron exchange between the protein and the bound DNA which could be used for lesion scanning in a process called a DNA-mediated charge transfer<sup>79</sup>. A similar mechanism was also shown for the archaeal XPD in vitro<sup>74</sup>, suggesting a role for the FeS cluster in lesion detection and verification during NER.

**Table 1 | List of NER factors and their role in DNA repair.\***

	<b>Mammals</b>	<b>Yeast</b>	<b>Catalytic activity, function in NER</b>
<b>CORE NER FACTORS</b>	XPA	Rad14	recruits NER factors to the lesion site, stimulates XPD helicase activity
	XPB**	Rad25	3'-5' helicase, 5'-3' translocase, opens DNA repair bubble, facilitates XPD loading on the DNA
	XPD	Rad3	5'-3' helicase, verifies lesions
	p62	Tfb1	binds XPC during TFIIH recruitment to the lesion site
	p52	Tfb2	stimulates XPB
	p44	SSL1	stimulates XPD, ubiquitin ligase
	p34	Tfb4	
	TTDA	Tfb5	stimulates XPB
	MAT1	Tfb3	
	CDK7	Kin28	kinase activity, required for transcription initiation but not for NER
	Cyclin H	Ccl1	stimulates Cdk7
	XPG	Rad2	endonuclease, cleaves 3' of the lesion
	XPF-ERCC1	Rad10-Rad1	endonuclease, cleaves 5' of the lesion
RPA	Rpa	binds single-stranded DNA	
<b>GGR- SPECIFIC FACTORS</b>	XPC	Rad4	recognizes lesions in GGR
	DDB1-DDB2	not found	recognizes lesions in GGR
<b>TCR- SPECIFIC FACTORS</b>	CsB	Rad26	translocase, recognizes lesion-arrested RNAP, stimulates RNAP elongation
	CsA	not found	part of E3-ubiquitin ligase, ubiquitinates CsB
	UVSSA	not found	stabilizes CsB during TCR
	Usp7	not found	ubiquitin protease, deubiquitinates CsB
<b>GAP FILLING FACTORS</b>	PCNA	Pcna	stimulates processivity of DNA polymerase
	RFC	Rfc	loads PCNA on the DNA
	DNA Pol $\delta, \epsilon$	Pol2, Pol3	resynthesizes excised DNA
	FEN1	Rad27	exonuclease, degrades damaged DNA
	DNA ligase 1	Cdc9	seals the gap after the reparatory DNA synthesis

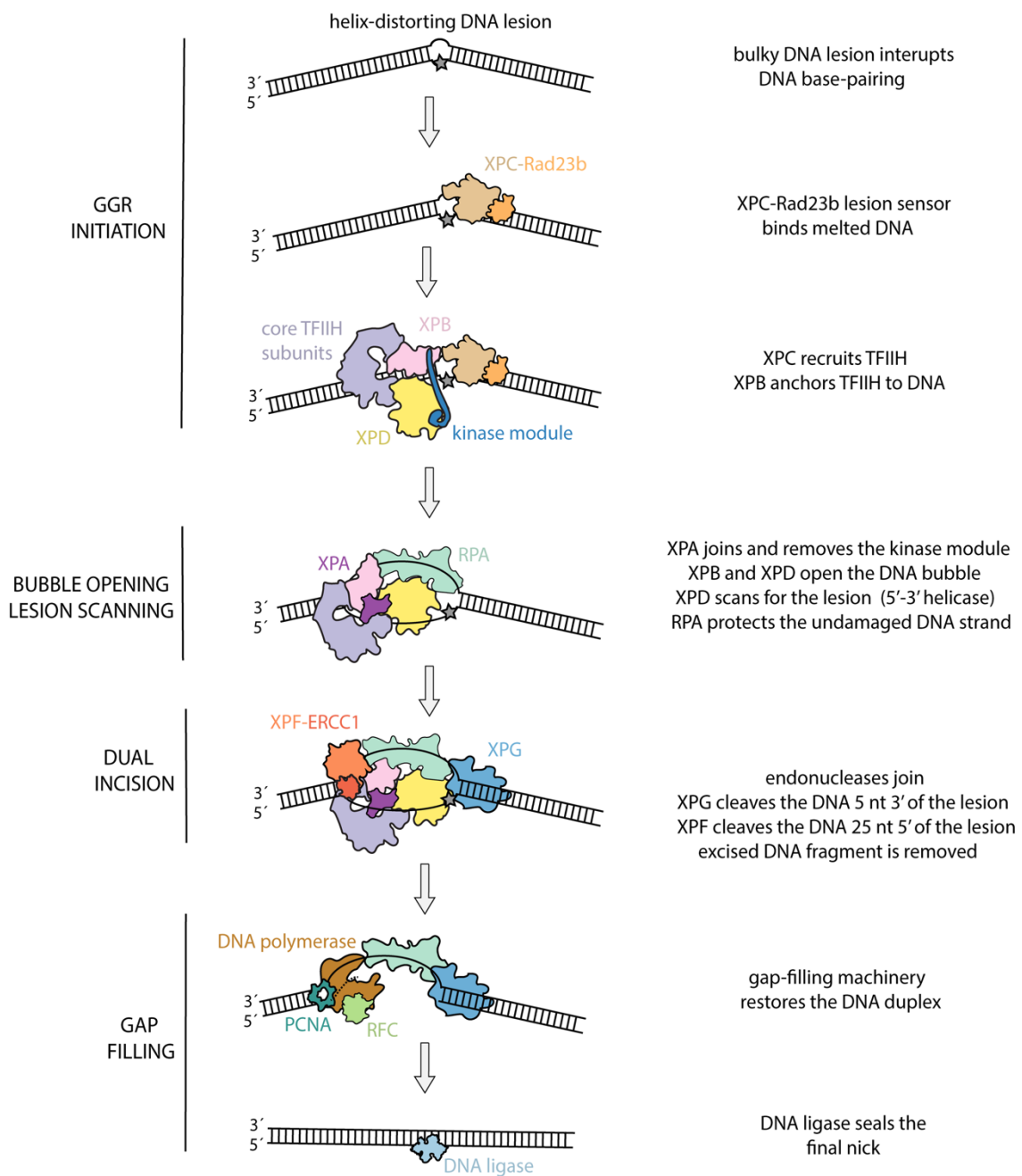
\* adapted from<sup>80</sup>

\*\* XPB, XPD, p62, p52, p44, p34 and TTDA form the TFIIH core module and MAT1, CDK7 and cyclin H form the TFIIH kinase module<sup>64</sup>.

**XPA** is recruited in the early stages of the repair reaction, most likely following the recruitment of TFIIH<sup>61,63</sup> (Fig. 2). XPA is a DNA binding protein with a preference for kinked DNA duplex structures<sup>81</sup> and duplex-single strand DNA junctions<sup>82</sup>. Even though XPA is essential for repair and mutations in XPA cause the most severe cases of Xeroderma pigmentosum<sup>83</sup>, its role in repair is not yet clear. The arrival of XPA to the NER assembly was linked to the dissociation of the TFIIH kinase module *in vivo*<sup>84</sup>. The kinase module inhibits XPD helicase activity<sup>72,73</sup> and needs to be removed before the lesion scanning by XPD can begin. Moreover, XPA stimulates XPD helicase activity *in vitro* but the mechanism of this stimulation is not known<sup>73</sup>. Despite its small size, XPA is considered the main scaffolding protein during repair<sup>83</sup> because it interacts with almost all NER factors, including XPC-Rad23b, TFIIH, RPA, XPF-ERCC1, DDB2 and PCNA<sup>85-98</sup>. The location of XPA within the NER assembly is unknown but some model studies suggest it may sit on the 5' edge of the repair bubble<sup>99</sup> where it would be ideally positioned to recruit the XPF-ERCC1 endonuclease<sup>95,96</sup>. XPA was also proposed to play a role during the initial recognition of DNA lesions<sup>100</sup>.

The replication protein A (**RPA**) binds the DNA cooperatively with XPA<sup>101,102</sup>. RPA is a eukaryotic single-stranded DNA binding protein comprised of three subunits, RPA1-3<sup>103</sup>. The largest RPA subunit contains three DNA-binding domains, while the other two subunits contain only one such domain<sup>103</sup>. This allows RPA to bind DNA fragments of different lengths by successively engaging the DNA-binding domains<sup>103,104</sup>. The cooperative binding of RPA subunits to DNA results in a very high overall affinity for the single-stranded DNA (sub-nanomolar range)<sup>100</sup>. During NER, RPA most likely engages the nondamaged DNA strand because the damaged DNA strand needs to be scanned by XPD<sup>56</sup>. By coating the nondamaged DNA strand, RPA might stabilize the open DNA bubble and also protect the solvent-exposed DNA strand from the DNA damage<sup>19</sup>. Interestingly, when all RPA subunits are bound to DNA, RPA covers 30 nucleotides of DNA which corresponds to the size of the excised DNA fragment during NER<sup>103</sup>. RPA binds the DNA in a defined polarity and it was suggested that RPA helps to position the endonucleases on the opposite sites of the DNA repair bubble for the dual-incision reaction<sup>105</sup> (Fig. 2).

**XPG and XPF-ERCC1** are structure-specific endonucleases which cleave the DNA around the lesion site<sup>106</sup> (Fig. 2). XPG belongs to the FEN1 family of endonucleases and cleaves the duplex-single strand junctions and other DNA substrates containing a 5' DNA overhang<sup>107,108</sup>. The XPF-ERCC1 endonuclease is comprised of two polypeptides which together form a unique active site<sup>109</sup>. XPF-ERCC1, like XPG, recognizes and cleaves the duplex single-strand DNA junctions<sup>110</sup>. However, XPF-ERCC1 cleaves the DNA substrates containing a 3' overhang<sup>110</sup>. Thus, both endonucleases bind the duplex-single strand junctions which are formed during the repair bubble opening but cleave the DNA with different polarities. *In vitro* repair assays have shown that XPG cleaves the damaged DNA 5 nucleotides 3' of the lesion site and the XPF-ERCC1 cleaves the DNA ~25 nucleotides 5' of the lesion site<sup>65,110,111</sup> (Fig. 2).



**Figure 2 | Schematic representation of human global-genome nucleotide excision repair (GGR).**

A bulky DNA modification (indicated by the gray star) induces local DNA melting which is recognized by the main lesion sensor XPC-Rad23b<sup>24,46-48</sup>. XPC recruits TFIIH to the lesion site<sup>61-63</sup>. This allows the XPB subunit of the TFIIH complex to bind the DNA duplex and to open the DNA repair bubble<sup>65-68</sup>. XPA and RPA cooperatively join the NER assembly<sup>101,102</sup>. XPA facilitates the removal of the TFIIH kinase module because the kinase module inhibits the repair<sup>84</sup>. RPA may coat the undamaged DNA strand and stabilizes the open DNA conformation. XPD engages the damaged DNA single-strand and starts to unwind the DNA in 5'-3' direction (towards the lesion)<sup>71-73</sup>. XPG and XPF-ERCC1 endonucleases bind the 3' and the 5' edge of the bubble, respectively, where they cleave the DNA after the presence of the lesion was verified<sup>106</sup>. Lesions are most likely verified by XPD because the helicase activity of XPD is inhibited by lesions<sup>23,73-75</sup>. The cleaved lesion-containing oligonucleotide is released from the NER complex and the components of the reparatory DNA synthesis are recruited<sup>19,20</sup>. DNA polymerase synthesizes a new DNA strand by using the nondamaged DNA strand as a template<sup>19,20</sup>. The final nick in DNA is sealed by a DNA ligase<sup>19,20</sup>.

Despite having the opposite cleavage polarities, the two endonucleases have to be properly positioned around the repair bubble by interactions with other NER factors<sup>105</sup>. XPG is constitutively associated with TFIIH *in vivo*<sup>112</sup>. XPG is thus recruited early during NER and also has a non-catalytic role in facilitating the DNA repair bubble opening<sup>61,113</sup>. In contrast, XPF-ERCC1 is the last core NER factor recruited to the lesion site, which is mediated by the interaction between XPA and ERCC1<sup>95,96</sup>. XPF-ERCC1 makes the first incision in the DNA, creating a free 3' DNA end which can be used by a DNA-polymerase to initiate the repair DNA synthesis<sup>114</sup>. XPG likely cleaves the DNA when the reparatory synthesis is in progress to ensure that all components required for restoring the DNA duplex are present before the DNA fragment is excised<sup>115</sup>. The DNA gap is filled by a machinery comprised of the clamp (PCNA), the clamp loading complex (RFC), DNA polymerase  $\delta$  and DNA ligase 1<sup>44</sup>. Depending on the proliferative status of the cell, DNA polymerases  $\epsilon$  and  $\kappa$ , and XRCC1-DNA ligase 3 are also employed<sup>116,117</sup>.

### 1.2.2 Transcription-coupled nucleotide excision repair (TCR)

Transcription is an initial stage of gene expression in which the information stored in a DNA molecule is copied into a strand of RNA<sup>118</sup>. This process is mediated by RNA polymerases, multiprotein complexes that elongate the RNA molecule by successive addition of nucleotides to its 3' end<sup>118</sup>. Human cells host several types of RNA polymerases<sup>119</sup>, however, here we focus only on the RNA-polymerase II (RNAP) which can participate in DNA repair<sup>120</sup>. The RNAP transcribes the protein coding genes in three main stages: initiation, elongation and termination<sup>121</sup>. During the initiation, the RNAP is correctly positioned at the beginning of a gene and the DNA duplex is opened to expose the DNA strand that needs to be transcribed<sup>122</sup>. The transcribed DNA strand, i.e. the strand copied into RNA, is called the template strand (TS), and the complementary DNA strand is called the non-template strand (NTS)<sup>123</sup> (Fig. 3a). The elongation phase includes the extension of the RNA, while the RNAP translocates on the TS in 3'-5' direction<sup>118</sup> (Fig. 3a). Termination is the last stage of transcription in which the RNAP dissociates from the DNA template and the RNA product is released<sup>121</sup>.

Transcription-coupled nucleotide excision repair (TCR) was first discovered in human cells where it was observed that UV-induced DNA lesions are preferentially repaired in the transcribed strands of active genes<sup>124,125</sup> and that the repair requires actively transcribing RNAP<sup>126</sup>. Thus, TCR is initiated by RNAP arrested by transcription-blocking lesions<sup>49-51</sup>. RNAP has a large footprint on the DNA which completely covers the DNA lesion<sup>51</sup>, so the main substrate for the TCR machinery is the lesion arrested RNAP and not the lesion itself<sup>80</sup>. Arrested RNAP is initially recognized by Cockayne syndrome (CS) proteins CsA and CsB which are considered to be the main TCR-specific factors<sup>127-129</sup>. Recently, a couple of additional factors, UVSSA and Usp7, were discovered to be essential for TCR<sup>130-132</sup>.

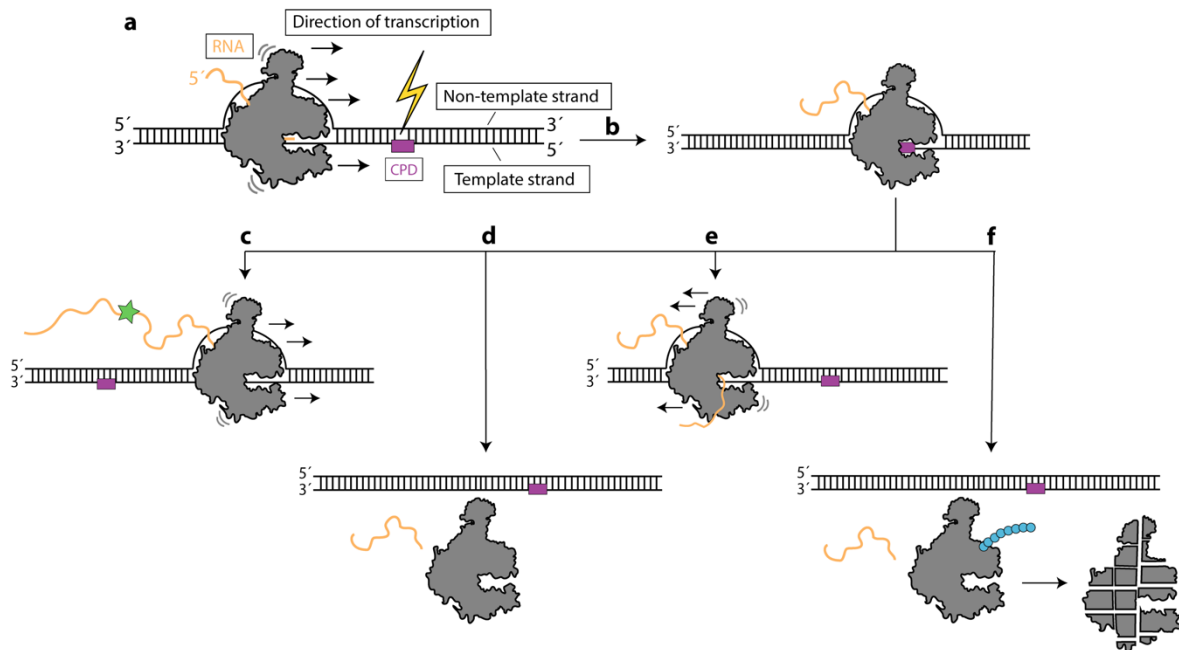


**CsB** is a member of SF2-family of DNA helicases<sup>133</sup>, however, its helicase or translocase activity has up to now not been demonstrated. CsB contains a central ATPase domain flanked by long N- and C-terminal regions<sup>134</sup>. The N-terminal region negatively regulates the ATPase function of CsB<sup>135</sup> and might be important for the recognition of RNAP<sup>135</sup>. The C-terminal region harbors a ubiquitin binding domain essential for TCR<sup>136</sup>, however, the function of this domain or the ubiquitinated binding partner have not been identified. CsB stimulates transcription elongation *in vitro*<sup>137</sup>, which helps the RNAP to transcribe over small DNA lesions<sup>138</sup>. Cryo-EM structure of the yeast CsB homologue Rad26 bound to the RNAP elongation complex<sup>138</sup> shows that Rad26 binds the upstream DNA, suggesting that Rad26 facilitates transcription by pushing RNAP forward. However, CsB cannot facilitate the RNAP passage over the transcription-blocking lesions<sup>137</sup> nor dissociate stalled elongation complexes from DNA to remove the RNAP from the lesion site<sup>50</sup>. Since CsB recruits downstream DNA repair factors to the arrested RNAP<sup>139</sup>, the role of CsB in TCR might be to probe the status of elongating RNAP and to initiate repair when RNAP forward translocation is blocked by a DNA lesion. CsB is also a member of Swi2/Snf2 family of chromatin remodelers and it was shown to remodel chromatin *in vitro*<sup>140</sup> and *in vivo*<sup>141</sup>. In addition, CsB recruits the histone acetyltransferase p300 and the nucleosome binding protein HMGN1 to the lesion site<sup>139</sup>. Thus, CsB might facilitate TCR by relaxing the chromatin environment around the lesion arrested RNAP. CsB is also important for the transcription restart after UV-irradiation, partially by counteracting the function of the transcription repressor ATF3<sup>142</sup>. The analysis of the CsB interactome by pull-down and mass-spectrometry<sup>143</sup> has revealed that CsB interacts with many components of the RNA splicing machinery, suggesting a role for CsB in the RNA metabolism outside DNA repair. Interestingly, mutations in CsB cause Cockayne syndrome in humans which is characterized by premature aging, mental retardation, dwarfism and other physical abnormalities<sup>144</sup>. These clinical features are more severe compared to diseases caused by defects in DNA repair, further emphasizing the important roles for CsB in other cellular functions. Overall, despite the large amount of accumulated knowledge on CsB, it is still not clear how and if CsB is directly involved in the DNA repair reaction<sup>80</sup>. It might be that the main CsB function in TCR is to stimulate transcription and to prepare the chromatin environment surrounding the arrested RNAP for efficient lesion processing by repair enzymes.

**CsA** is a small seven bladed-WD40 propeller protein that functions as a substrate receptor of a DDB1-Cul4-Rbx1 (CRL4) E3-ubiquitin ligase<sup>120,145,146</sup>. The ubiquitin ligase ubiquitinates CsB, thereby facilitating its degradation<sup>147</sup>, but does not modify the arrested RNAP<sup>148</sup>. Thus, CsA might be required in later stages of TCR to extract CsB from the complex with the RNAP and to regulate the TCR progression via post-translational modifications. **UVSSA** is an essential TCR factor<sup>130-132</sup> comprised of two domains, the C-terminal DUF2043 domain and the N-terminal VHS domain<sup>131</sup>. The VHS domain is necessary for the role UVSSA plays in TCR and it was implicated in ubiquitin binding, even

though the ubiquitinated binding substrate for UVSSA has not been identified<sup>131</sup>. UVSSA is involved in RNAP ubiquitination by potentially recruiting an E3-ubiquitin ligase to the lesion site, however, the ubiquitination does not result in degradation of the RNAP and likely has a signaling function<sup>131</sup>. Co-immunoprecipitation experiments have shown that UVSSA interacts with RNAP, CsB, CsA and several TFIIH subunits<sup>130-132</sup>. The interaction with the RNAP might be mediated by CsB and CsA<sup>149</sup>. Moreover, it was shown by NMR spectroscopy that the acidic amino acid stretch in UVSSA binds the PH domain in the p62 subunit of TFIIH<sup>150</sup>, which is very similar to the mode of interaction between XPC and TFIIH during the initial steps of GGR<sup>151</sup>. Thus, it was suggested that UVSSA recruits the core NER machinery to the lesion-arrested RNAP. Mutations in UVSSA cause the UV-sensitive syndrome which is accompanied by mild clinical features, mostly limited to the UV-hypersensitivity<sup>149</sup>. Since the sensitivity to UV-irradiation is a typical consequence of a defect in DNA repair, UVSSA might exclusively be a DNA repair factor. In addition, UVSSA recruits the **Usp7** ubiquitin protease to the lesion site<sup>130,132</sup>. Usp7 is an essential TCR factor because the cells lacking Usp7 exhibit a similar TCR defect as the cells lacking UVSSA<sup>130,132</sup>. Usp7 prevents the proteasomal degradation of CsB following DNA damage<sup>130</sup>, most likely by counteracting the CsB ubiquitination by CsA<sup>147</sup>, which may provide a longer time window for the successful completion of TCR.

Unlike GGR, TCR has not been successfully reconstituted *in vitro* with purified components and its mechanism is less well understood. After the arrest, the RNAP has to be removed to expose the lesion to repair enzymes<sup>152</sup>. How RNAP is removed remains the central open question in TCR but several scenarios can be envisioned (Fig. 3)<sup>152,153</sup>: RNAP can be dissociated from the DNA, actively pushed or pulled backwards far enough to reveal the lesion or degraded by the proteasome<sup>152,153</sup>. Up to now, a human DNA repair protein which could dissociate the lesion-arrested RNAP from the DNA has not been found<sup>80</sup>. Also, a case of active RNAP backtracking in eukaryotes has not been reported yet. The active RNAP backtracking would require an ATP-consuming motor protein which can push the arrested RNAP backwards – that is, in the direction opposite to the direction of transcription (Fig. 3a)<sup>154</sup>. During the backtracking, the 3' end of the RNA disengages from the RNAP active site and the RNA is threaded through a pore in RNAP which differs from the usual RNA trajectory during transcription elongation (Fig. 3e)<sup>155</sup>. Since the RNA 3' end is no longer accessible to be extended, the RNAP cannot resume transcription. However, a specialized transcription elongation factor IIS (TFIIS) can bind the backtracked RNAP and induce the intrinsic endonuclease activity of RNAP, which results in the cleavage of RNA<sup>156-159</sup>. This creates a new 3' RNA end and allows the restart of transcription. Thus, the backtracking strategy for the removal of RNAP from the lesion site would allow the transcription to continue after DNA repair<sup>152,153</sup>. Finally, the DNA-damage induced degradation of the arrested RNAP is considered to be the last resort when the transcription arrest persists for a long time<sup>160</sup>, and the process has been well described in yeast<sup>161</sup>.



**Figure 3 | The fate of lesion arrested RNAP.**

(a) The general architecture of the elongation complex. The transcribing RNAP is shown in gray and RNA in yellow. The template (TS) and non-template DNA strands (NTS) are indicated. RNAP translocates on the TS in 3'-5' direction. (b) The elongating RNAP is arrested by a CPD lesion (indicated by the purple square) in the TS. The arrested RNAP sequesters the access to the lesion and has to be removed by one of the following mechanisms; (c) Accessory factors chaperone transcription past the lesion. If an incorrect nucleotide was paired with the damaged DNA base, the RNA transcript will carry a mutation (indicated by the green star). (d) RNAP is actively dislodged from the DNA template, but can participate in the next round of transcription. (e) RNAP is actively pushed backwards (backtracked) to expose the DNA lesion to repair enzymes. (f) RNAP is marked with ubiquitin (indicated by the blue chain) and degraded by the proteasome. Adapted from<sup>152</sup>.

The RNAP is first marked with K63 polyubiquitin chains by the Rsp5 ubiquitin ligase<sup>162</sup> and these chains are subsequently hydrolyzed by Ubp2 to produce the monoubiquitinated RNAP<sup>163</sup>. The single ubiquitin can then be extended by the Elc1/Cul3 ubiquitin ligase to form the K48 polyubiquitin chains, which targets RNAP for degradation<sup>163</sup>. In human cells, the RNAP ubiquitination following the DNA damage is likely to play out in a similar cascade and may involve Nedd4 and the elongin complex<sup>148</sup>. After the removal of RNAP from the lesion site, the canonical DNA excision reaction occurs in the same way as in GGR<sup>153</sup>.

The mechanism of RNAP removal from the lesion site has been well characterized in bacteria<sup>154</sup>. The transcription-DNA repair coupling factor Mfd actively dislodges the lesion-arrested RNAP from the DNA, and immediately recruits the downstream repair machinery to the cleared lesion site<sup>164</sup>. In addition, the UvrD helicase can pull the arrested RNAP backwards to facilitate DNA repair<sup>165</sup>. It was suggested that the Mfd-mediated pathway is preferred during the normal growth conditions when DNA damage is sporadic and less frequent<sup>154</sup>. However, during the excessive DNA damage, bacterial cells upregulate the expression of the UvrD helicase which might favor the RNAP backtracking<sup>154</sup>. Thus, bacteria can adapt to different genotoxic conditions by switching between the two alternative TCR strategies<sup>154</sup>.

### 1.3 General transcription factor IIIH (TFIIH) is a central NER factor

Since the discovery of transcription factor IIIH (TFIIH) nearly three decades ago<sup>166</sup>, its structure, biochemical and cellular functions have been extensively studied<sup>64</sup>. TFIIH is comprised of ten polypeptides divided into two modules, the core and the kinase module (Table 1)<sup>64</sup>. The core is comprised of the DNA helicase XPD<sup>73</sup>, the DNA helicase/translocase XPB<sup>67,69</sup>, the ubiquitin ligase p44<sup>167</sup> and three non-enzymatic subunits: p62, p52 and TTDA (Fig. 4). The heterotrimeric kinase module contains the protein kinase CDK7, cyclin H and MAT1<sup>64</sup>. Mutations in various TFIIH subunits were linked to several human disorders, including Xeroderma pigmentosum, Cockayne syndrome, Trichothiodystrophy and Cerebro-oculo-facio-skeletal syndrome<sup>168,169</sup>. The range of symptoms, from mild UV-sensitivity to neurodegeneration, premature aging and a high risk of cancer development emphasizes the important roles TFIIH plays in cellular viability<sup>169</sup>.

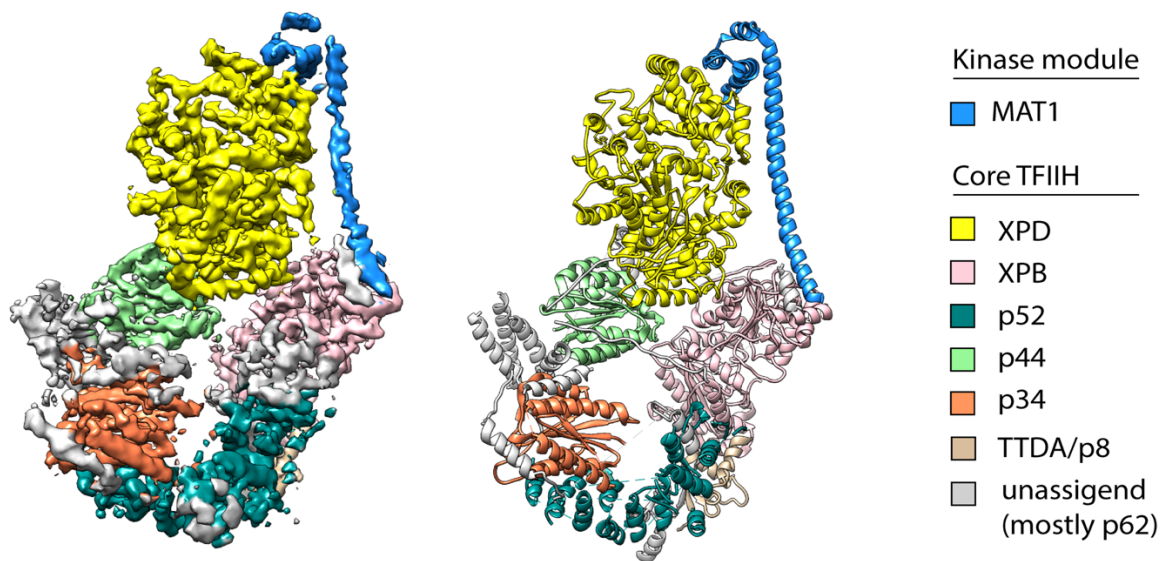
TFIIH has a central role in both transcription initiation and DNA repair<sup>64</sup> and it has also been implicated in the cell-cycle regulation and chromosome segregation<sup>170,171</sup>. Transcription initiation relies on six general transcription factors (GTF) to deliver the RNAP to the transcription start site and stimulate the early mRNA synthesis<sup>172</sup>. One of the GTFs is TFIH and its role is to facilitate the DNA melting necessary for the formation of the RNAP elongation complex<sup>64</sup>. This TFIH function is mediated by the enzymatic activity of the XPB subunit<sup>64</sup> but the underlying mechanism is not yet well understood. It was proposed that XPB acts as a molecular wrench which rotates while gripping the downstream DNA to twist and melt the DNA within the RNAP<sup>173-175</sup>. Another model suggests that XPB utilizes a DNA translocase activity to pump the DNA into the RNAP by tracking on the non-template promoter strand in the 5'-3' direction or on the template promoter strand in the 3'-5' direction (i.e. away from RNAP), which would result in DNA underwinding and promoter opening<sup>69,70</sup>. The TFIIH kinase module also has an important role during the transcription initiation<sup>176</sup>. Initiating RNAP is phosphorylated on the C-terminal domain (CTD) of the largest RNAP subunit comprised of 52 heptad repeats (Y<sub>1</sub>S<sub>2</sub>P<sub>3</sub>T<sub>4</sub>S<sub>5</sub>P<sub>6</sub>S<sub>7</sub>) in humans<sup>177</sup>. The CTD phosphorylation is not important for the RNA synthesis, but rather serves as a recruiting platform for the RNA processing machinery, including RNA capping, splicing and polyadenylation<sup>177</sup>. The CKD7 subunit of the TFIIH kinase module phosphorylates serine 5 and serine 7<sup>178</sup> on the CTD heptad repeats which is needed for mRNA capping<sup>179</sup> and processing spliceosomal small nucleolar RNAs<sup>180</sup>, respectively.

During the DNA repair TFIIH changes its composition and employs different enzymatic activities when compared to its role in the transcription initiation<sup>64</sup>. The kinase module is essential for transcription but inhibits DNA repair by suppressing the XPD helicase activity and possibly by phosphorylating DNA repair factors<sup>72,73</sup>. In vivo, the kinase module dissociates from the TFIIH core concomitantly to the recruitment of XPA and re-associates with the core TFIIH after the repair is completed<sup>84</sup>. Another major difference is

the role for the XPD helicase<sup>181</sup>. During the transcription initiation, XPD acts as a scaffold and stabilizes the interactions between TFIIH and the transcription pre-initiation complex (PIC)<sup>70</sup>. In contrast, the XPD helicase activity is essential for the repair bubble opening and the lesion verification during the DNA repair<sup>19</sup>. Only the XPB ATPase activity is a shared requirement for both transcription and DNA repair, and it might be employed for DNA melting which underlies both processes<sup>64</sup>. Cryo-EM structures of TFIIH<sup>182</sup> and TFIIH bound to the transcription pre-initiation complex (PIC)<sup>70</sup> were reported, as discussed in more detail below. However, TFIIH has not yet been visualized in the context of DNA repair.

## 1.4 Structural studies of NER

The core NER machinery is built around the multi-subunit TFIIH complex<sup>64</sup>, which in molecular mass is roughly equal to all of the other NER factors combined. The TFIIH architecture was investigated over the past decades with cross-linking mass-spectrometry, 2D-crystallography and negative stain electron microscopy<sup>64,183-186</sup>. The overall shape of the core TFIIH resembles a horse shoe with the core TFIIH ATPases located at its open ends. The kinase module docks between XPB and XPD but shares a larger interface with XPD. In addition, structures of several TFIIH subunits and domains from various organisms were solved by X-ray crystallography and NMR, including XPB<sup>67</sup>, XPD<sup>78</sup>, TTDA complexed with the p52 N-terminus<sup>187</sup>, p62 PH domain<sup>188</sup>, p34 vWA domain and p44 RING domain<sup>189</sup>. However, a major breakthrough in the TFIIH structural biology occurred very recently, when cryo-EM structures of the human TFIIH<sup>182</sup> (Fig. 4) and the yeast TFIIH bound to PIC<sup>70</sup> were reported at 4.4 Å and ~6 Å resolution, respectively. Combined with the previous structural and function studies on TFIIH and homology modelling, the pseudo-atomic structure of the almost complete TFIIH core could be assembled<sup>70,182</sup>. XPB and XPD are joined by a ring composed of subunits p52, p44 and p34<sup>70</sup>. Within the core, p52 interacts with XPB and p44 interacts with XPD<sup>70</sup>, which is consistent with the stimulatory effect of p52 and p44 on the respective ATPases *in vitro*<sup>68,70</sup>. Subunit p62 is not a part of the horse shoe core assembly, but wraps around it and acts as a molecular glue which stabilizes TFIIH<sup>182</sup>. In addition, the PH domain of p62 is important for TFIIH recruitment to PIC where it binds the general transcription factor IIE (TFIIE)<sup>70</sup>. The kinase module is poorly resolved in both reported structures, likely due to its high flexibility<sup>70,182</sup>. However, in the human TFIIH structure<sup>182</sup>, a long helix belonging to the MAT 1 subunit of the kinase module bridges the core TFIIH ATPases and a MAT1 helical bundle stacks on top of the XPD helicase (Fig. 4).



**Figure 4 | Cryo-EM structure of human TFIIH.**

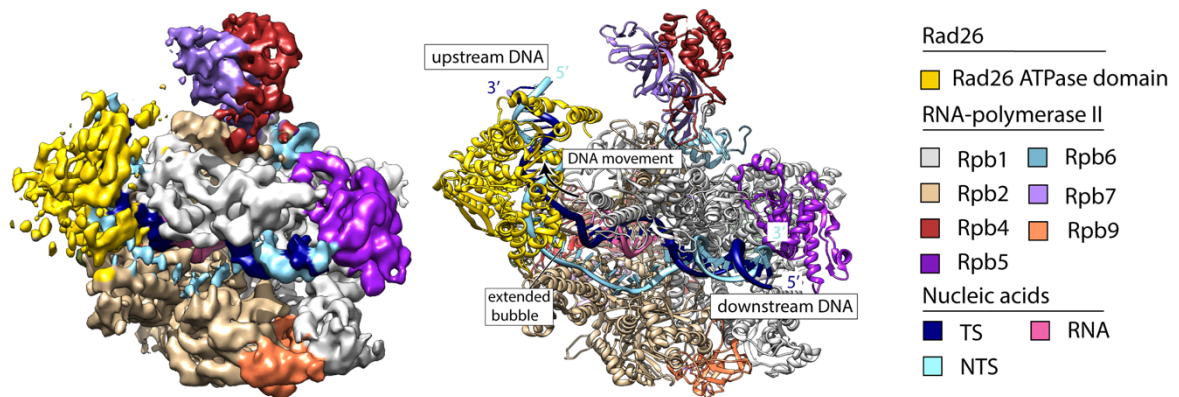
Cryo-EM density (left, EMDB:3802)<sup>182</sup> and the resulting pseudo-atomic model (center, PDB:5OF4)<sup>182</sup> are shown. TFIIH subunits are color-coded as indicated (right).

In addition to the structural characterization of TFIIH, the structure of the yeast XPC homologue Rad4 in a complex with the lesion containing DNA was reported<sup>26</sup>. Rad4 is comprised of a N-terminal  $\alpha/\beta$  domain, followed by three consecutive  $\alpha/\beta$  domains containing long  $\beta$ -hairpins (beta-hairpin domains, BHD1-3). Rad 4 extensively interacts with 11 base pairs of a DNA duplex 3' of the lesion-site and with the four base pairs of melted DNA surrounding the lesion. BHD2 and BHD3 domains tightly bind the non-damaged DNA strand and a  $\beta$ -hairpin from BHD3 inserts into the DNA and causes the lesion to flip out of the DNA duplex. Thus, Rad4 does not interact with the DNA lesion itself, but rather recognizes the DNA elements surrounding the helix-destabilizing DNA lesion<sup>26</sup>.

A solution structure of human XPA<sup>92</sup> and a crystal structure of a lesion bound yeast XPA homologue Rad14<sup>100</sup> were also described. XPA is a small protein comprised of a N-terminal zinc finger and a central DNA-binding domain<sup>83</sup>. In a structure with damaged DNA<sup>100</sup>, the XPA dimer binds a 13 base pair DNA duplex which is bent by 70° at the lesion site, and each XPA copy uses a  $\beta$ -hairpin to stack against the DNA bases at the ends of the duplex. Like XPC, XPA does not interact with the DNA lesion directly but the lesion remains within the DNA duplex<sup>26,100</sup>. The structure does not explain the preference of XPA for binding Y-shaped DNA junctions<sup>82</sup> and it is also hard to conceive how XPA integrates into the NER assembly as a dimer, so it is still unclear how relevant this structure is for the role XPA plays in the context of NER.

Structural insights into TCR are quite limited. The crystal structure of the yeast RNAP arrested on a CPD lesion<sup>51</sup> represents the initial substrate for TCR. Interestingly, the structure is identical to the RNAP elongation complex in the absence of a DNA lesion, arguing against an allosteric mechanism for the recruitment of the DNA repair machinery to the arrested RNAP<sup>51</sup>.

Recently, a cryo-EM structure of the yeast CsB homologue Rad26 bound to transcribing RNAP was solved (Fig. 5)<sup>138</sup>. The RNAP transcription bubble was artificially extended on the upstream end to facilitate the Rad26-RNAP complex formation. Rad26 binds the upstream DNA duplex of the elongation complex and uses a structural element in the ATPase lobe 2 to wedge between the re-annealing DNA strands<sup>138</sup>. Interestingly, Rad26 bends the upstream DNA by 70°<sup>138</sup>. The structure suggests that Rad26 stimulates transcription by translocating on the template strand in 3'-5' direction which would push RNAP forward<sup>138</sup> (Fig. 5). Structures of all the other TCR-specific or core NER factors in complex with lesion-arrested RNAP have not yet been solved, probably due to the lack of mechanistic knowledge on TCR which would enable identification of relevant and stable targets for structural studies.



**Figure 5 | Cryo-EM structure of Rad26 bound to RNA-polymerase II.**

Cryo-EM density (left, EMDB:8735)<sup>138</sup> and the resulting pseudo-atomic model (center, PDB:5VVR)<sup>138</sup> are shown. Rad26 and the RNA-polymerase II subunits visible in this view are color-coded as indicated (right). Rad26 was suggested to translocate on the template strand (TS) in the 3'-5' direction, which would pull the TS out of the RNA-polymerase II (indicated by the arrow) and consequently push the RNA-polymerase II forward<sup>138</sup>. NTS, non-template strand.

## 1.5 Aims and scope of this work

Human NER has been a subject of intense research for over a half of a century. Numerous *in vivo* and *in vitro* studies have identified factors driving NER and defined the molecular framework of this DNA repair pathway. However, the main difficulty in studying the NER reaction mechanism is to obtain large amounts of highly purified DNA repair factors, especially multi-subunit complexes such as RNA-polymerase II and TFIIH. As a result, we are currently lacking thorough biochemical analysis and structures of NER assembly intermediates, which hinders our mechanistic understanding of NER. The aim of this work was to reconstitute both NER sub-pathways in a highly purified *in vitro* system and to analyze the isolated reaction intermediates biochemically and structurally, using spectroscopic and FRET-based enzymatic assays, cross-linking mass-spectrometry and cryo-electron microscopy.

Here, we were able to:

- (i) Clone, express and purify all NER factors in milligram amounts and high purity, including XPC-Rad23b, RPA, XPA, XPG, XPF-ERCC1, core TFIIH, TFIIH kinase module, CsB, CsA-DDB1, UVSSA, Usp7 and TFIIIS. Complete heterodecameric TFIIH complex was reconstituted from the purified TFIIH core and the kinase module.
- (ii) Reconstitute GGR and TCR initiation. For GGR, XPC-Rad23b was bound to a DNA containing a bulky DNA modification, which served as a recruitment platform for TFIIH and the core TFIIH-XPA complex. For TCR, we synthesized a DNA scaffold containing a site-specific photolesion CPD. Purified pig RNAP was transcribed into a photolesion and decorated with TCR factors. Complexes were isolated and analyzed by crosslinking mass-spectrometry and negative stain EM.
- (iii) Biochemically characterize the core TFIIH ATPases XPB and XPD and dissect their regulation in the context of NER.
- (iv) Solve cryo-EM structure of the core TFIIH-DNA-XPA lesion scanning NER intermediate at 3.6 Å resolution. The structure reveals the topology of the NER machinery on the DNA, suggests how DNA lesions are recognized by the XPD helicase and explains many mutations found in human diseases. The structure also provides insights into the role and the positioning of the main NER scaffolding factor XPA.
- (v) Reveal the fate of the lesion-arrested RNAP. RNAP is actively backtracked by a machinery comprised of TFIIH core, XPA and XPG. The backtracking is mediated by the XPB translocase activity and it is sufficient to fully expose lesions to repair.

Work presented here expands and redefines our understanding of NER and provides the basis for unraveling the complete mechanism of this repair pathway.

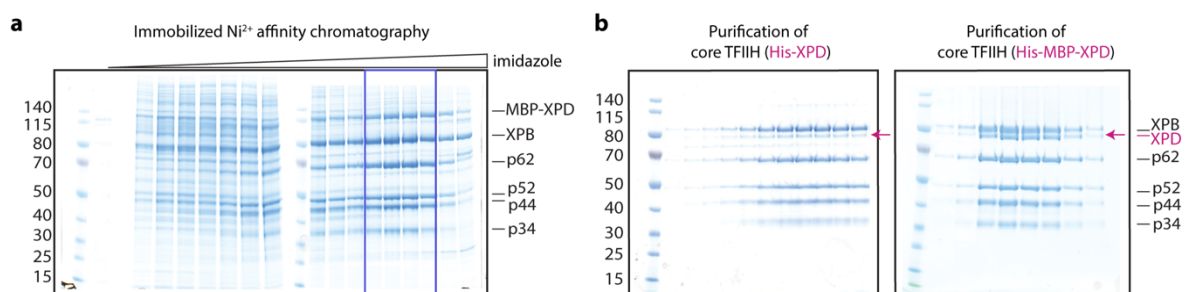


## 2. RESULTS

### 2.1 Reconstitution and analysis of GGR initiation

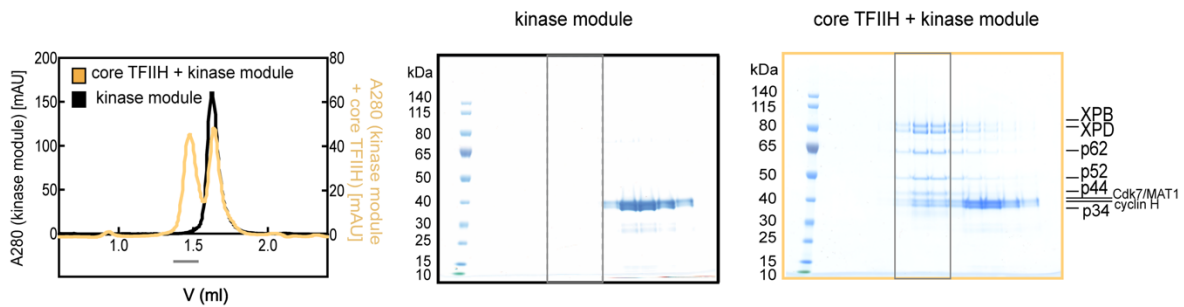
GGR starts with the recognition of damaged DNA by XPC, followed by the XPC-mediated recruitment of TFIIH and XPA<sup>46,56,73</sup>. We reconstituted this step of the NER pathway with highly purified human factors and a lesion-containing DNA scaffold. The NER intermediate was analyzed by crosslinking mass-spectrometry and negative staining electron microscopy.

A robust biochemical and structural characterization of human NER was thus far hampered because pure TFIIH was not available in large amounts. TFIIH is frequently obtained from endogenous sources such as HeLa cells<sup>182</sup>, which requires large amounts of starting material and restricts the manipulation of the TFIIH amino acid sequence. Recombinant expression of TFIIH in insect cells was previously reported<sup>73</sup>, however, the final protein yield was limited because many viruses were used to infect insect cells, and each virus encoded one or two TFIIH subunits. We used the MacroBac vector series<sup>190</sup> to combine 6 TFIIH subunits (XPB, p62, p52, p44, p34 and TTDA) into a single vector by sequence- and ligation-independent cloning<sup>190</sup> (Methods). The XPD subunit was left on a separate vector to facilitate mutagenesis of this subunit for later biochemical studies. The core TFIIH module was produced by co-infecting insect cells with only two viruses which dramatically increased the final protein yield and quality. Several different tagging strategies were tested to optimize the core TFIIH purification protocol. We tagged four different subunits with an N-terminal 6xHis tag (Table 8) which allowed us to separate the properly formed TFIIH core from the excess subunits and subassemblies by a shallow imidazole gradient on a nickel column (Fig. 6a). Initial purification attempts always resulted in a sub-stoichiometric amount of XPD, so we tagged this subunit with a maltose binding protein to enrich for the complete TFIIH core assembly by an additional affinity chromatography step (Fig. 6b, Methods).



**Figure 6 | Purification of the human TFIIH core.**

(a) Elution fractions from the nickel column were analyzed by SDS-PAGE. The full core TFIIH complex eluted at higher imidazole concentrations (indicated by the purple frame), while the sub-assemblies eluted at lower imidazole concentrations. (b) Purification of core TFIIH before (left) and after (right) an MBP-tag was added to the XPD subunit and an amylose affinity step was incorporated into the purification protocol. Gels show the fractions from the final purification step (size-exclusion chromatography), analyzed by SDS-PAGE. The improved protocol yielded the TFIIH core containing stoichiometric amounts of the XPD subunit (indicated by the arrow).



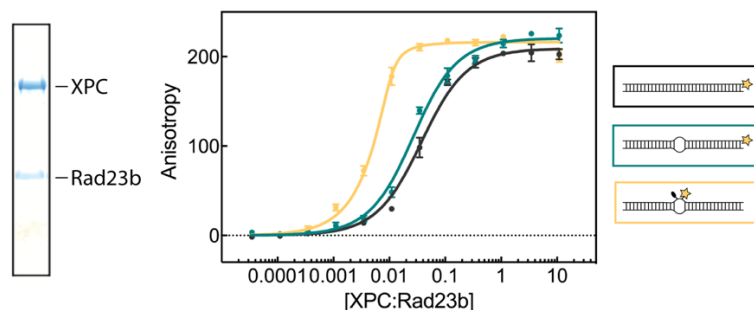
**Figure 7 | In vitro reconstitution of the TFIIH complex.**

(left) Chromatograms show analytical size-exclusion elution profiles for the kinase module in black and the mixture of the core TFIIH and the kinase module in yellow. (middle) Fractions corresponding to the elution of the kinase module were analyzed by SDS-PAGE and Coomassie staining. (right) Fractions corresponding to the elution of the kinase module and the core TFIIH were analyzed by SDS-PAGE. Comparison with the kinase module only fractions (gray frame) shows a shift of the kinase module to earlier fractions in the presence of the TFIIH core, indicating 10 subunit TFIIH complex formation.

All tags were cleaved off with tobacco etch virus protease to prevent potential interference of tags with the TFIIH enzymatic activities or the reconstitution of NER intermediates. The final purification scheme included five different affinity, ion-exchange and size-exclusion chromatography steps (Methods) and we could obtain milligram amounts of the TFIIH core at near homogeneity. The kinase module which is comprised of subunits MAT1, CDK7 and cyclin H<sup>64</sup> was purified separately. We could reconstitute the full 10 subunits human TFIIH complex from the TFIIH core and the TFIIH kinase module (Fig. 7). XPC is the primary lesion sensor in GGR and it is tightly associated with two additional factors, Rad23b and centrin 2<sup>191,192</sup>. In vitro repair assays have shown that only Rad23b is required for efficient NER<sup>192</sup>. Thus, we established the purification protocol for the XPC-Rad23b binary complex expressed in insect cells (Fig. 8, Methods). XPA could be purified from both *Escherichia coli* and insect cells, however, we used only the protein obtained from the insect cell expression system (Methods).

XPC recognizes bulky DNA lesions that kink the DNA duplex<sup>24,52,53</sup>. We designed a model DNA scaffold which contains fluorescein and biotin moieties coupled to consecutive thymine bases as a bulky lesion mimic (Table 12). To test if the lesion mimic can specifically recruit the XPC-Rad23b complex we performed fluorescence anisotropy and measured the DNA affinity of XPC-Rad23b for nondamaged and damaged DNA (Fig. 8). We also included a control in which a three base pair mismatch was introduced at the site of a DNA lesion to check if the impaired DNA base pairing affects XPC binding. We fitted the anisotropy data with the single-site binding equation to determine the apparent dissociation constant ( $K_{d,app}$ ) for different DNA substrates. XPC-Rad23b showed a high affinity for DNA duplex ( $K_{d,app} = 32.6 \pm 3.1$  nM) which additionally increased for mismatched DNA ( $K_{d,app} = 21.8 \pm 2.1$  nM). However, in the presence of the lesion mimic the affinity increased by two orders of magnitude ( $K_{d,app} \approx 0.6$  nM). This suggests that fluorescein and biotin efficiently

mimic bulky DNA lesions and facilitate the recruitment of the XPC-Rad23b complex to DNA.

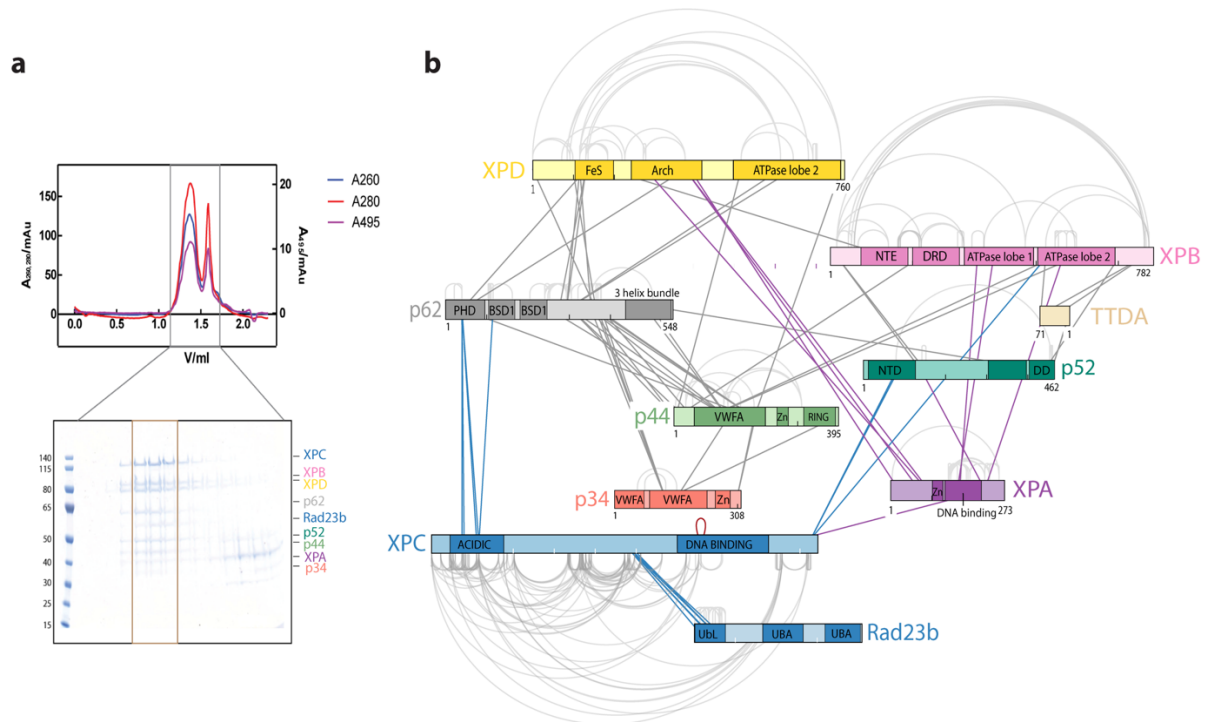


**Figure 8 | Binding of XPC-Rad23b to DNA as measured by fluorescence anisotropy.**

(left) Purified XPC-Rad23b complex was resolved by SDS-PAGE and visualized by Coomassie staining. (right) Binding of XPC-Rad23b to different DNA scaffolds is shown; fully complementary DNA is in black, DNA containing a mismatch bubble in green and DNA containing modified DNA bases that mimic bulky DNA lesions in yellow. Error bars represent s.d. from the mean values for 3 experimental replicates. Fluorescein is indicated as a yellow star and biotin as a black circle.

To investigate the architecture of the NER intermediate which forms after the lesion recognition and prior to the repair bubble opening, we assembled the core TFIIH-XPC-Rad23b-XPA-DNA complex (Fig. 9a) and analyzed the protein interactions by chemical crosslinking and mass-spectrometric detection of crosslinking sites (Fig. 9b). The NER intermediate was reconstituted by mixing the XPC-Rad23b lesion sensor with the damaged DNA, followed by the addition of the core TFIIH and XPA. The resulting complex was purified by size-exclusion chromatography (Fig. 9a) and crosslinked with BS3, which crosslinks lysine residues within a 30 Å distance. The kinase module was omitted because it was shown that the kinase module inhibits DNA repair and dissociates from the TFIIH core upon XPA recruitment<sup>84</sup>.

The crosslinking network between the core TFIIH subunits is very similar to the BS3 crosslinking network reported for TFIIH immunopurified from human cells<sup>186</sup> which validates our crosslinking results and suggests that the core TFIIH subunits are properly assembled during the expression in insect cells. The XPC-Rad23b complex does not extensively crosslink to TFIIH. The acidic region in the XPC N-terminus mainly crosslinks to the pleckstrin homology (PH) domain of the p62 subunit in TFIIH (Fig. 9b). This agrees with a coimmunoprecipitation study which showed that the XPC N-terminal region interacts with p62<sup>62,193</sup>. Moreover, an NMR structure of the acidic amino acid stretch in the XPC N-terminus bound to the PH domain of p62 was previously reported for human and yeast proteins<sup>151,194</sup>.



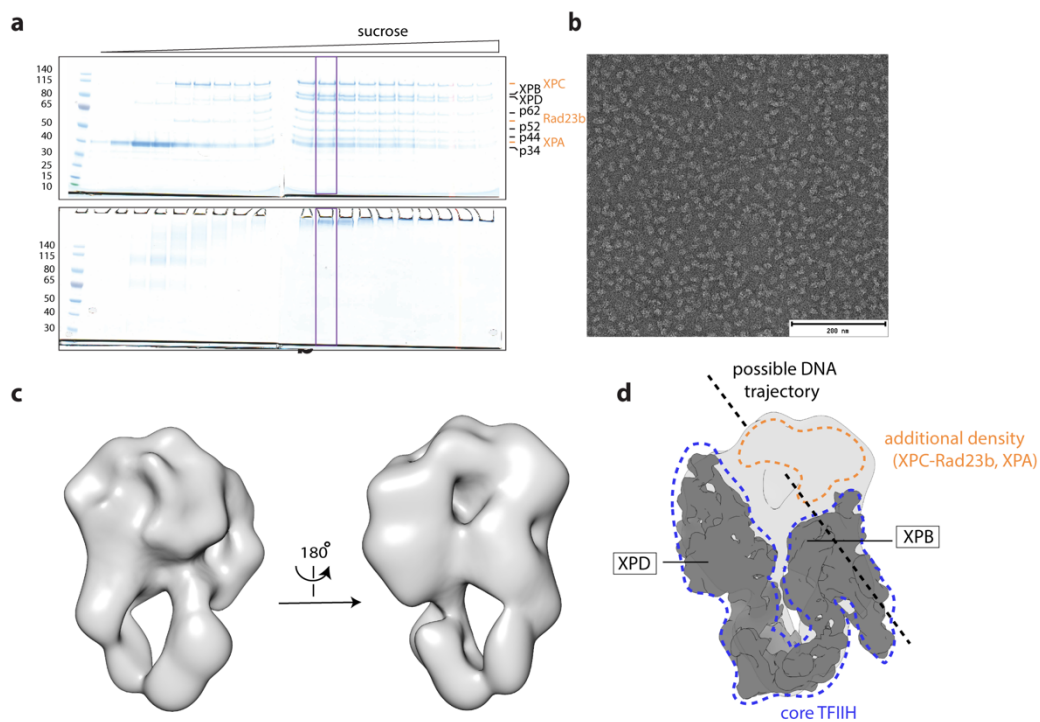
**Figure 9 | Reconstitution and the crosslinking mass-spectrometry network of the core TFIIH-XPA-XPC-Rad23b-DNA complex.**

(a) Chromatogram shows the size-exclusion elution profile of the core TFIIH-XPA-XPC-Rad23b-DNA complex. We monitored the elution of the fluorescein-containing DNA by absorption at 495 nm (purple). Peak fractions were analyzed by SDS-PAGE and Coomassie staining. All protein subunits coeluted in stoichiometric amounts in the presence of the lesion-containing DNA scaffold. (b) Crosslinking network of the complex. TFIIH subunits are colored as in Fig.4, XPC-Rad23b is shown in blue and XPA in purple. Crosslinks between the core TFIIH subunits are shown in gray, between XPC and core TFIIH in blue, and between XPA and core TFIIH in purple. Crosslinks with a score above 3 are shown. List of inter-subunit crosslinks is provided in the Supplemental Table 1. DRD, damage recognition domain; NTE, N-terminal extension; BSD, BTF2-like, synapse-associated and DOS2-like domains; VWFA, von Willebrand factor type A domain; UBL, ubiquitin like; UBA, ubiquitin-associated domain; FeS, iron-sulfur domain; Zn, zinc finger; DD, dimerization domain; NTD, N-terminal domain

Limited crosslinks between TFIIH and XPC suggest that the interaction is largely mediated by DNA. XPA mainly crosslinks to the core TFIIH ATPases. In particular, the N-terminal region of XPA crosslinks to XPD and the C-terminal region of XPA crosslinks to XPB. Since XPA is a DNA binding protein<sup>82,195,196</sup>, it may sit on the DNA between the core TFIIH ATPases.

We also analyzed the core TFIIH-XPC-Rad23b-XPA-DNA complex by negative-stain electron microscopy (EM) (Fig. 10). The complex was purified by sucrose gradient ultra-centrifugation in the presence of glutaraldehyde which fixes and stabilizes particles during the sample preparation for EM (Fig. 10a)<sup>197</sup>. The sample was stained with uranyl formate (Fig. 10b) and a small data set was acquired. The structure was solved to  $\sim 30\text{\AA}$  (Fig. 10c). The structure strongly resembles the previously reported structure of the human TFIIH bound to PIC<sup>198</sup> (Fig. 10d). After docking the TFIIH structure<sup>198</sup> into our negative stain envelope, an additional density between the TFIIH ATPases XPB and XPD was observed. The additional density might correspond to the XPC-Rad23b lesion sensor. Since XPC-

Rad23b binds the damaged DNA duplex with a very high affinity<sup>48</sup> (Fig. 8) and XPB binds the double-stranded DNA in the TFIIH-PIC structure<sup>198</sup>, it is tempting to speculate that in our structure XPC and XPB sit on the DNA (Fig. 10d). This would also be in agreement with a model in which XPC first recognizes the lesion, followed by the recruitment of TFIIH and the initial DNA melting by XPB<sup>199</sup>. Moreover, the position of XPD between XPB and XPC would be ideally suited for XPD to engage the DNA single-strand after the repair bubble opening and to start scanning for the lesion (further discussed in section 3.5 and Fig. 39). To verify this model, we attempted to solve the high-resolution structure of the complex with cryo-EM. However, despite many attempts at sample optimization and different strategies used to purify and crosslink the complex, we could never properly align particles in 2D or 3D, presumably due to the high flexibility of the complex.



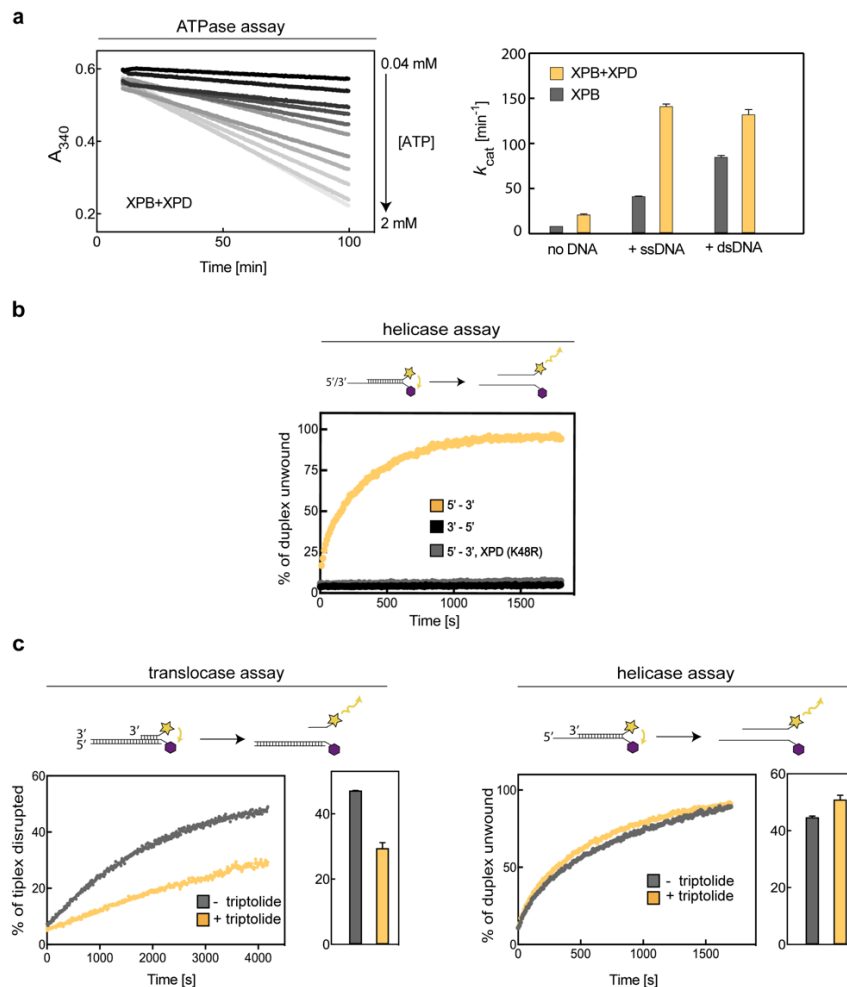
**Figure 10 | Negative stain reconstruction of the core TFIIH-XPA-XPC-Rad23b-DNA complex.**

(a) Complex formation on a 10–30% sucrose gradient in the absence (above) or the presence (below) of glutaraldehyde. The fraction of the crosslinked sample used for negative staining and electron microscopy is indicated by the purple frame. (b) Exemplary micrograph of the sample stained with uranyl formate. (c) Negative stain reconstruction of the core TFIIH-XPA-XPC-Rad23b-DNA complex. Two views are shown. (d) The cryo-EM density of the core TFIIH bound to PIC (blue dashed line outlines the core TFIIH; EMDB 8131)<sup>198</sup> was fitted into the negative stain reconstruction. The additional density observed in our reconstruction is indicated by a yellow dashed line and might correspond to XPC-Rad23b and XPA. Based on the DNA bound to the XPB subunit in the TFIIH-PIC structure and a strong affinity of XPC-Rad23b for the DNA duplex, we suggest a DNA trajectory through the complex (black dashed line).

## 2.2 Structural basis of TFIIH activation for NER

### 2.2.1 Biochemical characterization of core TFIIH ATPases

Systematic biochemical analysis of core TFIIH ATPases XPB and XPD in the context of NER is currently lacking due to difficulties in obtaining large amounts of recombinant TFIIH and other NER factors. Since we overcame this barrier (Methods), we first measured the ability of XPD and XPB to hydrolyze ATP (Fig. 11a).



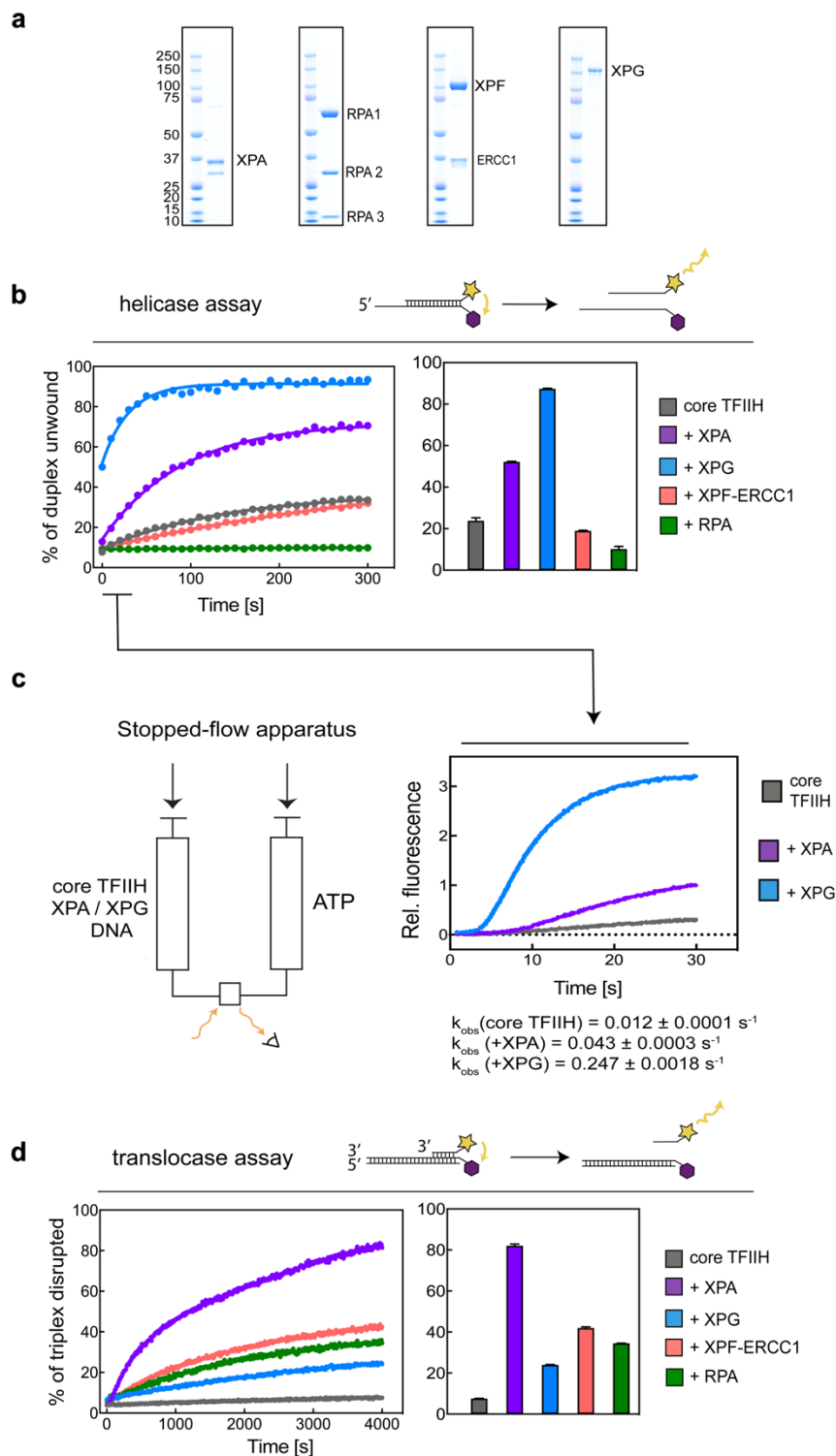
**Figure 11 | Biochemical characterization of the core TFIIH ATPases.**

(a) (left) Raw data of the ATPase assay for the TFIIH core. In the assay ADP recycling is coupled to NADH oxidation which can be monitored in real time by a decline in absorbance at 340 nm (Methods). The assay was performed at different ATP concentrations as indicated next to the plot. (right) The turnover number ( $k_{cat}$ ) of ATP consumption is shown for the core TFIIH (yellow) and the core TFIIH lacking XPD (gray), in absence of DNA or in the presence of ssDNA or dsDNA as indicated below the plot. Error bars represent the s.d. of the mean values for 2 experimental replicates (b) Polarity of the XPD helicase. DNA unwinding was monitored by a FRET-based assay (Methods). Fluorescence traces of the core TFIIH unwinding in the 5'–3' and the 3'–5' direction are shown in yellow and black, respectively. DNA unwinding of the core TFIIH containing the XPD K48R mutant is shown in gray. Shown traces are representative of 2 independent experiments. (c) Double-stranded DNA translocation by the core TFIIH is mediated by XPB. DNA unwinding and triplex disruption were monitored by a FRET-based assay (Methods). Translocase (left) and helicase (right) activities of the core TFIIH in the presence (yellow) or absence (gray) of triptolide (100  $\mu$ M final) are shown. Triptolide inhibits XPB<sup>200,201</sup> and specifically affects the core TFIIH translocase activity. Bar graphs show the percentage of disrupted triplex after 4000 s (left) and the percentage of unwound duplex after 300 s (right) for 2 independent replicates. Error bars represent s.d. from the mean values.

We additionally purified the core TFIIH lacking the XPD subunit, so that the ATP hydrolysis by XPB can be measured directly. An NADH-coupled photometric assay<sup>202</sup> was used to monitor ATP hydrolysis in real time (Fig. 11a). Since XPB and XPD are both DNA motor proteins we also tested if the addition of single (ssDNA) or double stranded DNA (dsDNA) will stimulate the ATPase rate of core TFIIH. In the absence of nucleic acids both core TFIIH ATPases are quite inefficient in hydrolyzing ATP (Fig. 11a), probably to prevent futile ATP consumption when their enzymatic function is not required. Upon the addition of excess ssDNA or dsDNA the turnover number of ATP consumption for both heptameric and hexameric core TFIIH assembly increased ~7 times (Fig. 11a). The stimulation of the TFIIH ATPase activity by DNA was also shown before<sup>73</sup>. However, the contribution of XPB and XPD to ATP hydrolysis is dependent on the type of DNA. XPD is more stimulated by ssDNA while XPB is more stimulated by dsDNA (Fig. 11a), which reflects the different substrate specificities of the two ATPases.

To further investigate XPB and XPD, we established fluorescence resonance energy transfer (FRET)-based assays to monitor their helicase and dsDNA translocase activities in real time. The dsDNA translocase activity (called just translocase activity throughout the thesis) of core TFIIH was measured by a triplex disruption assay, as previously described<sup>69,203</sup>. For the helicase activity, we also measured DNA unwinding in 5'-3' and 3'-5' direction. Core TFIIH exhibited a robust 5'-3' helicase activity (Fig. 11b) which could be attributed to XPD as this activity is abolished when XPD is replaced with the XPD point mutant which cannot hydrolyze ATP<sup>68</sup> (Fig. 11b). This agrees with previous reports which described XPD as a 5'-3' DNA helicase<sup>72,73</sup>. However, core TFIIH did not show any DNA unwinding in 3'-5' direction (Fig. 11b), the activity which has been demonstrated for archaeal XPB<sup>67</sup>. Core TFIIH also has a translocase activity which can be attributed to XPB because this activity is sensitive to the addition of triptolide, a drug that specifically inactivates XPB<sup>200,201</sup> (Fig. 11c). The translocase activity was also demonstrated for the yeast XPB homologue Ssl2<sup>69</sup>. As a control we showed that triptolide does not influence core TFIIH helicase activity which depends on XPD (Fig. 11c). A comparison of TFIIH enzymatic activities suggests that XPD is the dominant TFIIH motor, since DNA unwinding occurs on a shorter time scale and is more efficient compared to DNA translocation (Fig. 11). In summary, XPD is a 5'-3' DNA helicase and XPB is a dsDNA translocase.

Next we investigated the effect of NER factors on the core TFIIH ATPases. We purified XPA, RPA heterotrimeric complex, XPG and XPF-ERCC1 complex in large amounts and high purity (Fig. 12a). We first tested the effect of the NER factors on XPD helicase activity. XPD helicase activity was stimulated by both XPA and XPG (Fig. 12b). Helicase activity was not observed in the presence of RPA complex probably because RPA has a very high affinity for ssDNA<sup>104,204</sup> and masks the DNA overhang XPD uses to initiate DNA unwinding. The XPD helicase activity in the presence of XPG was too fast to be captured with a microplate reader which has a dead time of more than 10 s (Fig. 12b, blue trace).



**Figure 12 | Regulation of ATPases in core TFIIH.**

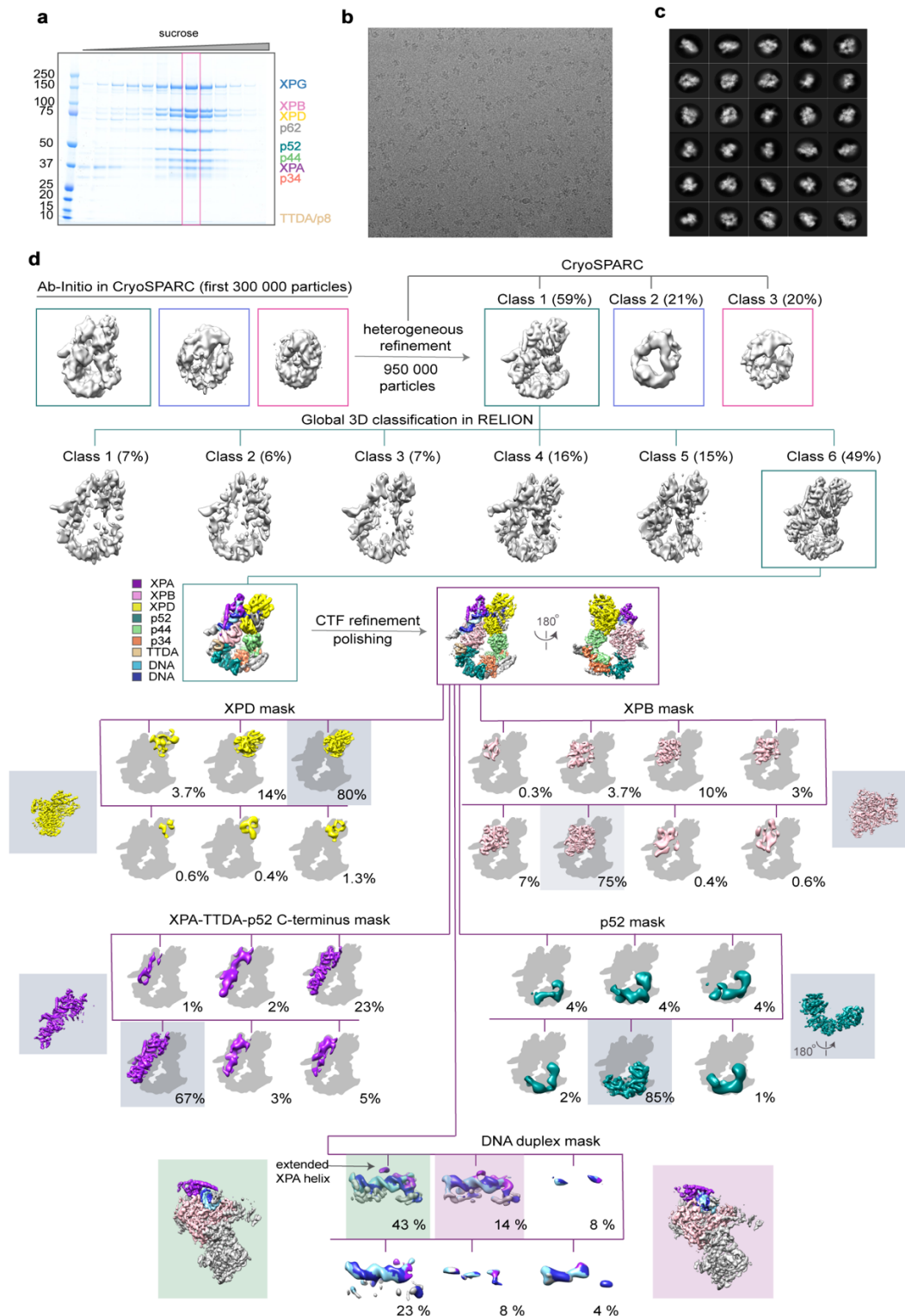
(a) Purified XPA, RPA, XPF-ERCC1 and XPG NER factors were resolved by SDS-PAGE and visualized by Coomassie staining. (b) Effect of NER proteins on XPD 5'-3' helicase activity. Real-time fluorescence measurement using a FRET assay as in Fig. 11b is shown. Bars show percentage of unwound product after 100s. Error bars provide s.d. of the mean of two replicates. (c) (left) Schematic representation of a stopped-flow apparatus. (right) Stopped-flow measurement of DNA unwinding. TFIIH core (gray) was pre-incubated with excess XPA (purple) or XPG (blue) and rapidly mixed with ATP in a stopped-flow apparatus. DNA unwinding was monitored by a FRET-based assay as in Fig. 11b. Fluorescence traces represent the average of 5 measurements. Initial linear parts of the traces were fitted with Prism software to obtain the initial rate of DNA unwinding, as indicated below the graph. (d) Effect of NER proteins on XPB translocase activity. Real-time fluorescence measurement of triplex disruption as in Fig. 11c is shown. Bars show percentage of disrupted triplex after 4000s. Error bars provide s. d. of the mean of two replicates.



Therefore, we coupled the helicase assay to a stopped-flow apparatus and resolved the entire DNA unwinding reaction in the presence of XPA and XPG (Fig. 12c). The unwinding was ~4 and ~20 times faster in the presence of XPA and XPG, respectively (Fig. 12c). Strong stimulation of XPD by XPG explains a longstanding observation that XPG is required for efficient repair bubble opening, which is independent of its endonuclease activity<sup>113</sup>. We also tested the effect of NER factors on the XPB translocase activity (Fig. 12d). The activity was stimulated by all NER factors tested, however, the stimulation by XPA was exceptionally strong. In the presence of XPA, the percentage of disrupted triplexes increased from ~10% to ~80% after 4000 s (Fig. 12d). Even though the positive effect of XPA on DNA unwinding by TFIIH was known<sup>56</sup>, the stimulation of the XPD helicase and the XPB translocase activity by XPG and XPA, respectively, has not been reported before. This may be important for the lesion scanning step of NER which is mediated by TFIIH migration on the DNA while searching for the lesion<sup>56</sup>.

### **2.2.2 Core TFIIH-XPA-XPG-DNA complex formation and cryo-EM structure determination**

In order to reveal the structural basis of XPB and XPD stimulation by XPA and XPG we formed the core TFIIH-XPA-XPG complex on a bifurcated DNA scaffold for cryo-EM. We chose bifurcated DNA as a nucleic acid scaffold because it mimics one half of a DNA repair bubble formed during NER<sup>65</sup> and it provides a binding substrate for both XPB and XPD. The complex was formed by mixing core TFIIH with excess DNA, XPA and XPG, and purified on a sucrose gradient (Fig. 13a). An XPG endonuclease mutant was used for the complex formation to prevent DNA cleavage during the sample preparation. All proteins comigrated on a sucrose gradient in apparently stoichiometric amounts, indicating that the complex was efficiently formed (Fig. 13a). For cryo-EM particles were crosslinked with glutaraldehyde which was added to the sucrose gradient<sup>197</sup>. After the gradient centrifugation, crosslinking was stopped with lysine and aspartate, which stochastically adds charges to the particles and improves the particle orientation distribution within the vitrified sample. The sample was visualized by a Titan Krios cryo-electron microscope and images showed nicely spaced individual particles of the correct size (Fig 13b). We collected a large data set of approximately 9000 micrographs. The micrographs were pre-processed in Warp<sup>205</sup> and particles were picked with a retrained BoxNet<sup>205</sup>. We used the first 300 000 particles auto-picked during the data collection for ab-initio reconstruction in CryoSPARC<sup>206</sup>. Interestingly, one ab-initio class already showed clear core TFIIH features and the reconstruction was not biased by using an already existing structure as input model (Fig. 13d). After further processing in CryoSPARC<sup>206</sup> and RELION<sup>207</sup> (Methods), we obtained a cryo-EM map at a global resolution of 3.6 Å according to the gold-standard Fourier shell correlation (FSC) 0.143 criterion with an applied B-factor of -110.02 Å<sup>-2</sup> (Fig 15b).



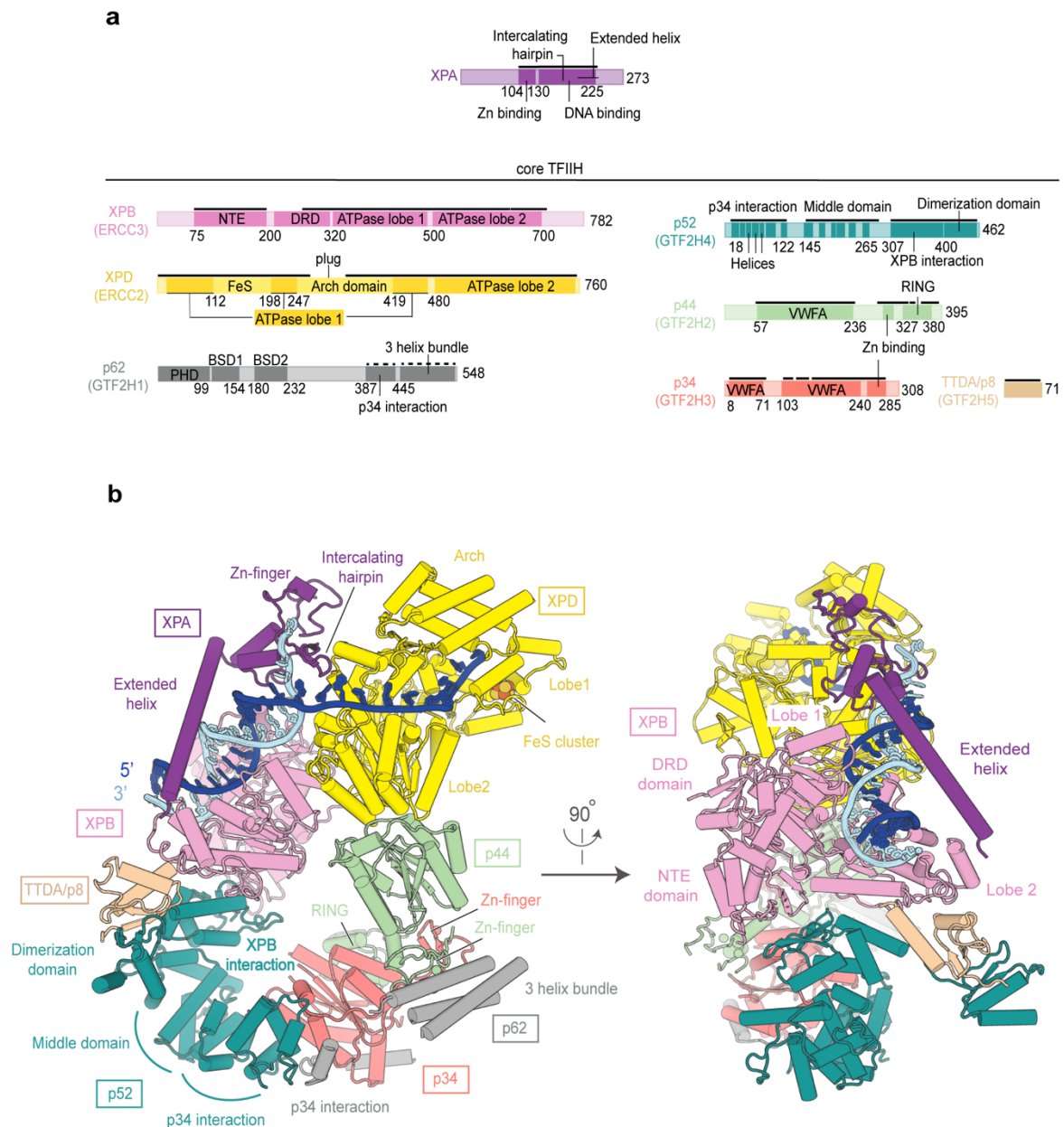
**Figure 13 | Cryo-EM data processing.**

(a) Core TFIIH-XPA-XPG-DNA complex formation for cryo-EM. Sucrose gradient fractions from the control without glutaraldehyde were analyzed by SDS-PAGE and Coomassie staining. The fraction marked by the pink frame, but in the presence of glutaraldehyde, was used for structure determination. (b) Representative micrograph shows nicely spaced individual particles. (c) Thirty representative 2D classes obtained after the final 3D classification. (d) Processing tree for structure determination of the TFIIH-XPA-DNA complex. Particle class-distribution after 3D classification is indicated next to the corresponding map. The final map obtained by global refinement is color coded by subunits as indicated. 5 focused classifications used to improve different parts of the structure are shown over the gray TFIIH silhouette for context. Classes used for the refinement are indicated with colored background and the refined structure is shown next to the corresponding 3D classes.

However, as seen from the 2D classes (Fig. 13c) and local resolution estimation (Fig. 15c) many peripheral parts of the complex were flexible and less well resolved. To improve the resolution of several TFIIH subunits, we performed focused classifications and refinements (Fig. 13d). This resulted in considerably improved maps for many regions in the TFIIH core, which facilitated model building (Methods).

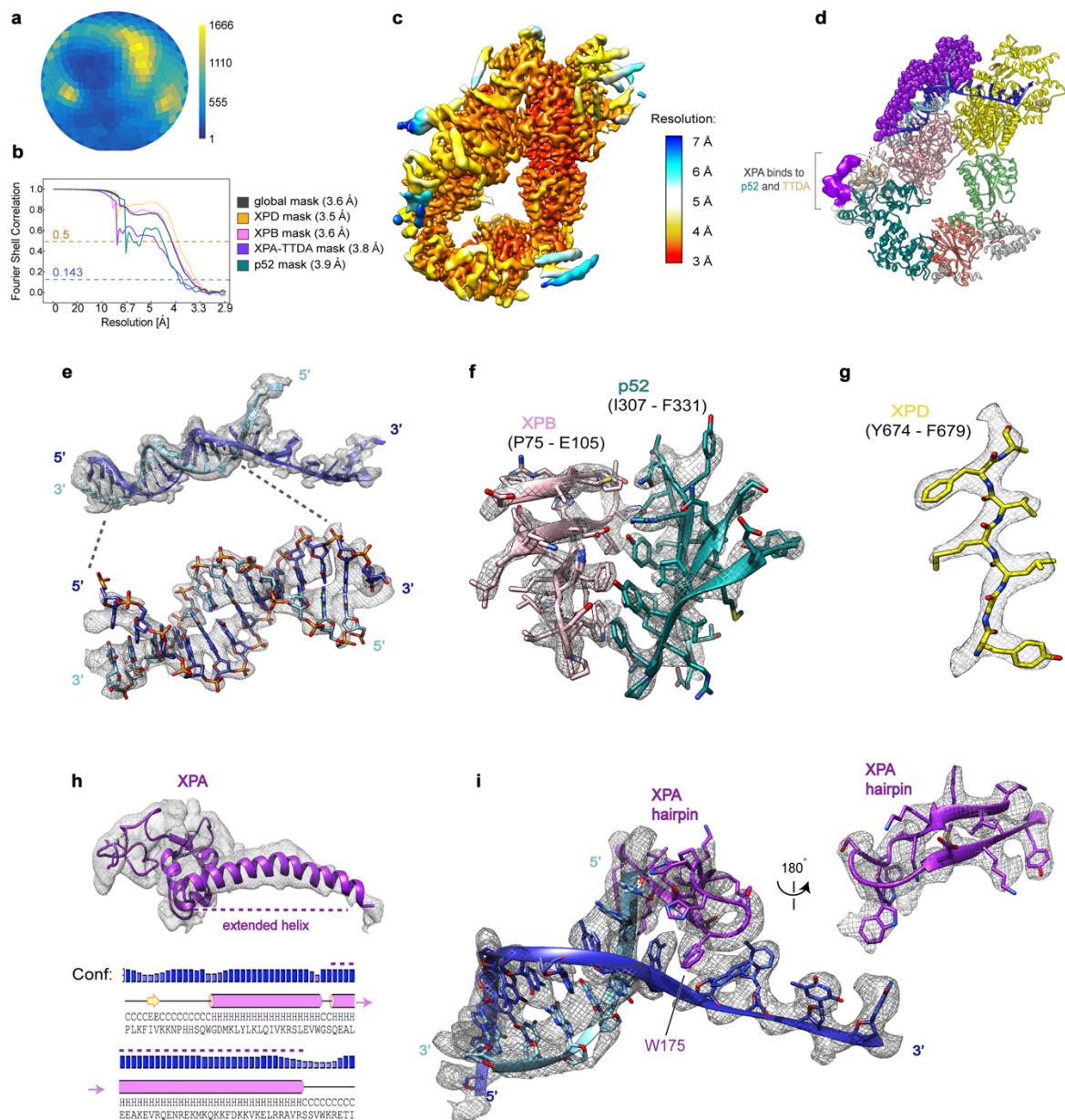
The final cryo-EM map revealed densities corresponding to all core TFIIH subunits except p62, for which only some secondary structure elements contacting p34 and XPD subunits were visible (Fig. 14). The p62 subunit was also not visible in the previously reported human TFIIH model<sup>182</sup> and it was proposed to act as a flexible molecular glue that spreads around the TFIIH core. However, due to the higher resolution obtained in our study, we were able to expand the previous model of the human TFIIH core<sup>182</sup> by modelling or *de novo* building the N-terminal extension (NTE) domain and part of the damage recognition domain (DRD) of XPB, most of the p52 subunit and the interface between the subunits p34 and 44 comprised of two zinc fingers and a RING domain (Figs. 14, 35). The newly built parts show how XPB is attached to the TFIIH core and explain previous biochemical observations (section 3.2).

We observed an additional EM density corresponding to the nucleic acid scaffold used for the complex formation (Figs. 13,14,15,16). We were able to trace the entire bifurcated DNA, which was coated with DNA repair proteins (Figs. 14, 16). XPB was bound to the DNA duplex, consistent with its role as a DNA translocase<sup>69</sup> (Fig. 11c), and XPD was bound to the DNA single strand, consistent with its role as a DNA helicase<sup>72,73</sup> (Fig. 11b). To my knowledge, this is the first structure of a DNA translocase and a DNA helicase sharing the DNA scaffold. Interestingly, XPB helicase motifs preferentially contact the 5'-3' DNA strand (Fig. 16c), suggesting that this is the strand XPB tracks on. The same DNA strand extends into the XPD helicase (Fig. 16a, c). Thus, both XPB and XPD use the same DNA strand for DNA migration, which has interesting implications for DNA repair bubble opening and XPD loading during the initial stages of NER (section 3.5). The EM density also revealed XPA which binds the duplex-single strand junction (Figs. 14, 18). However, XPG density was not observed and the protein probably dissociated during the sample vitrification for cryo-EM.



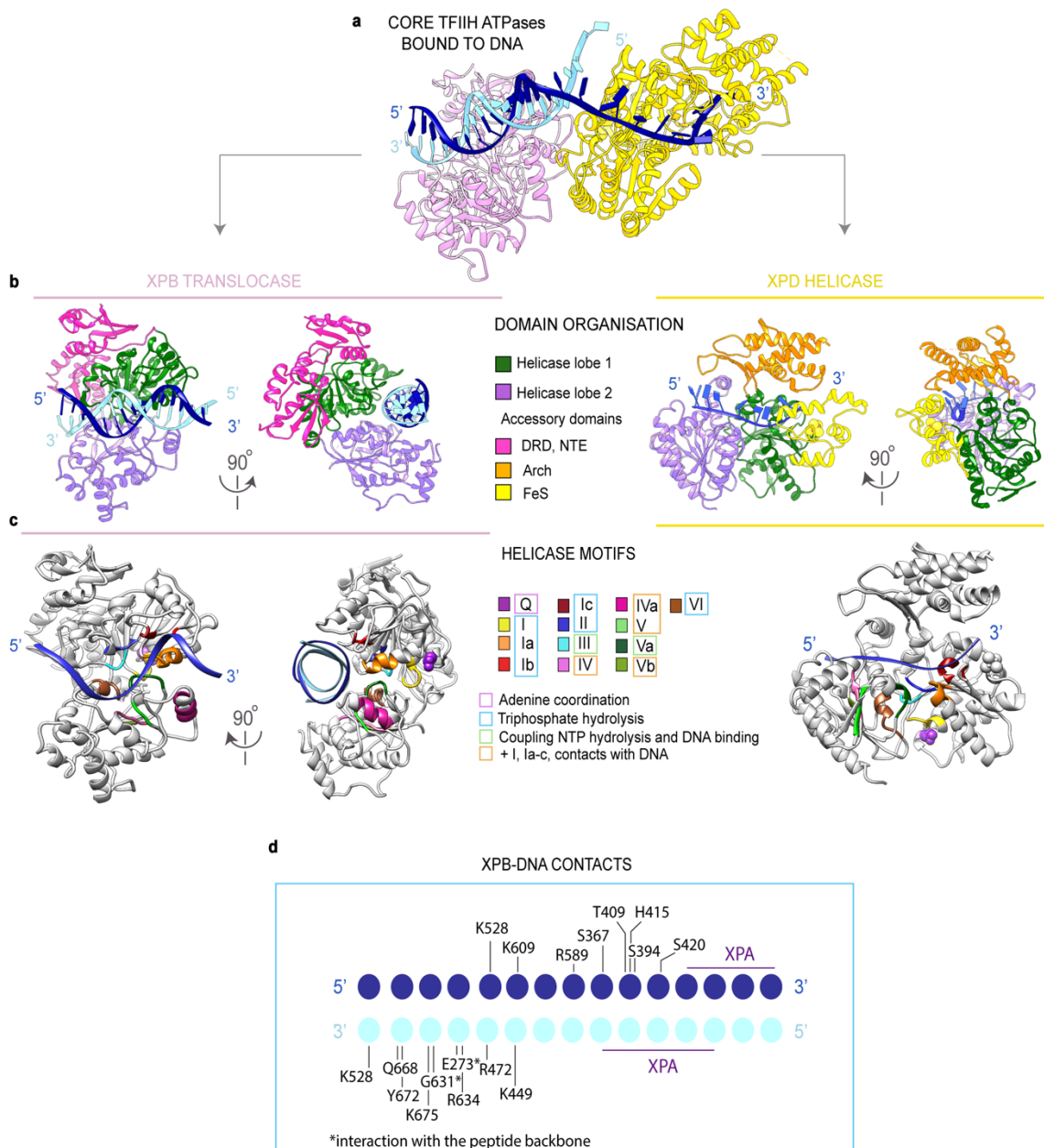
**Figure 14 | Structure of human TFIIH-XPA-DNA complex.**

(a) Domain organization of XPA and human TFIIH subunits. Residues at domain borders are indicated. Solid and dashed black lines mark residues modeled as atomic and backbone structures, respectively. DRD, damage recognition domain; NTE, N-terminal extension; BSD, BTF2-like, synapse-associated and DOS2-like domains; VWFA, von Willebrand factor type A domain. (b) Cylindrical representation of the structure. Proteins colored as in (a).



**Figure 15 | Cryo-EM map quality.**

(a) Angular distribution plot for particles contributing to the final reconstruction. Number of particles assigned to a particular orientation is color-coded as indicated. (b) Fourier shell correlation plots for the final and focused-classified maps. (c) Local resolution estimates for the final map. (d-i) Examples of the final map quality. (d) Additional XPA density (purple) contacting p52 dimerization domain and TTDA/p8 is observed at lower resolution and it was not used for model building. XPA is shown as purple spheres and TFIIF subunits are color coded as in Fig. 14. (e) Cryo-EM density corresponding to the bifurcated DNA scaffold. Close-up view of the double stranded DNA region shows a clear separation of the density corresponding to phosphates, ribose and DNA bases. (f) De novo built interface between the XPB and p52 subunits. Bulky side chains were used to assign the protein register. (g) Individual beta-strand from the second helicase lobe of XPD. The high map quality allows the protein backbone tracing and the placement of side-chains. (h) Cryo-EM density for the extended XPA helix directs the placement of the secondary structure element in the lower resolution map. PSIPRED tool predicts the formation of the long helix with high confidence. (i) Cryo-EM density for the XPA intercalating hairpin. W175 is indicated.

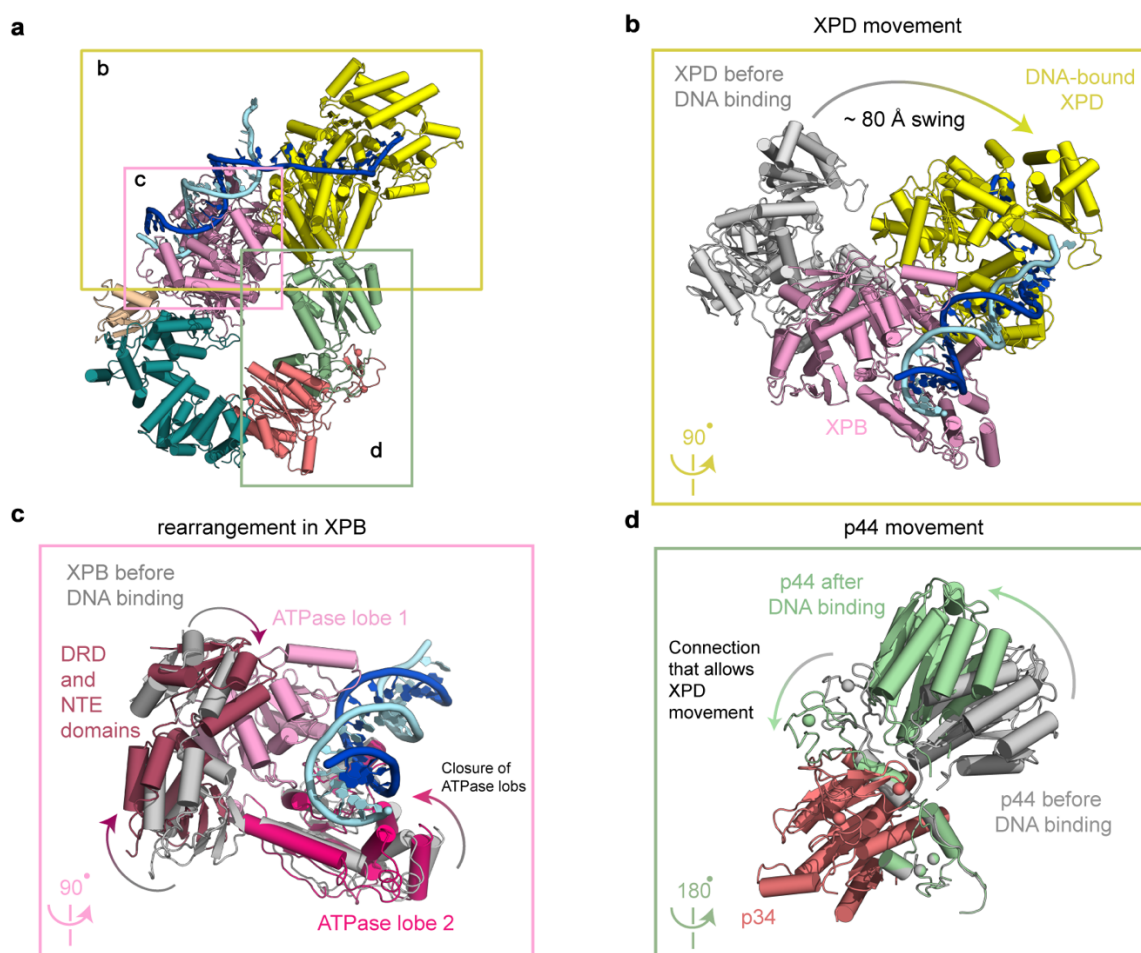


**Figure 16 | Detailed analysis of XPB and XPD bound to the DNA.**

(a) Ribbon representation of XPB (pink) and XPD (yellow) bound to the DNA. (b) Domain organization of XPB (left) and XPD (right). Domains are color coded as indicated. (c) Helicase motifs colored for XPB (left) and XPD (right) according to<sup>133</sup>. In the XPB structure it is clearly visible that the 5'-3' DNA strand interacts with the helicase motifs. (d) Schematic representation of XPB-DNA contacts. Schematic representation of XPD-DNA contacts is presented in the Fig. 19.

### 2.2.3 Structural rearrangements in TFIIH upon activation for DNA repair

A structure of TFIIH immunopurified from human cells was reported recently<sup>182</sup>. Aside from the core TFIIH subunits, the structure also revealed secondary-structure elements which belong to the kinase module subunit MAT1. In particular, a long MAT1 helix spans between XPB and XPD, and a helical bundle in MAT1 stacks onto XPD (Fig. 4, 22). The kinase module was shown to inhibit DNA repair by preventing TFIIH to engage DNA<sup>72,73</sup>. Thus, the comparison of our core TFIIH-XPA-DNA structure, which represents an NER intermediate, to that of the kinase bound TFIIH reveals the structural basis of TFIIH activation for DNA repair (Fig. 17).



**Figure 17 | Large-scale structural changes in TFIIH upon activation for repair.**

(a) The core TFIIH-DNA structure shown for context and color-coded as in Fig. 14. (b-d) Superpositions of different TFIIH subunits before and after DNA binding show inter- and intra-subunit flexibility. The structure in the absence of DNA was modeled by fitting TFIIH domains from our structure into the EM density for the kinase-bound TFIIH<sup>182</sup> for easier comparison, since our structure contains several domains not built in the previous TFIIH model<sup>182</sup>. TFIIH structure in the absence of DNA is always shown in gray. (b) The structures with and without the DNA were aligned on the ATPase lobe 1 of XPB. XPD swings  $\sim 80$  Å during DNA loading (R272 in corresponding structures was used for the distance measurement). (c) Internal rearrangements in XPB. XPB domains are color-coded as follows: NTE and DRD domains in dark pink, ATPase lobe 1 in pink and ATPase lobe 2 in hotpink. Binding to DNA duplex induces closure of XPB ATPase lobes and substantial rearrangement of accessory domains DRD and NTE. (d) Flexible interface between p44 and p34 comprised of several Zn-fingers supports the large-scale movement of XPD during the DNA loading.

The structure of the TFIIH core presented here is more complete compared to the structure of the TFIIH core bound to the kinase module<sup>182</sup> and contains several newly built domains (Fig. 35). Thus, for an easier comparison between the two structures we docked individual domains from our structure into the EM density of the kinase-bound TFIIH<sup>182</sup>, followed by real-space refinement in PHENIX<sup>208</sup> (Fig. 17). Remarkably, the core TFIIH undergoes a huge conformational rearrangement between the two states with a substantial intra- and inter-subunit reorganization (Fig. 17). The most prominent movement is the repositioning of the XPD helicase which has to swing for more than 80 Å to grasp the DNA (Fig. 17b). We also observed a rearrangement in the XPB translocase. The two XPB ATPase lobes close on the DNA duplex and the two accessory domains NTE and DRD rotate in opposite directions (Fig. 17c). The rotation of the DRD domain is important because this domain is one of the anchor points for the kinase on the TFIIH core<sup>182</sup>. A major source of the core TFIIH flexibility is the interface between the p34 and p44 subunits which is comprised of two intertwined zinc fingers (Fig. 17d). The p44 subunit latches on p34 subunit with the RING domain but the rest of its body is free to rotate. Since p44 couples XPD to the rest of the TFIIH core, the flexibility of p44 enables the high degree of mobility observed for the XPD helicase. The plasticity of the TFIIH ring which couples the enzymatically active subunits XPB and XPD might allow TFIIH to participate in very different biological roles<sup>64</sup>.

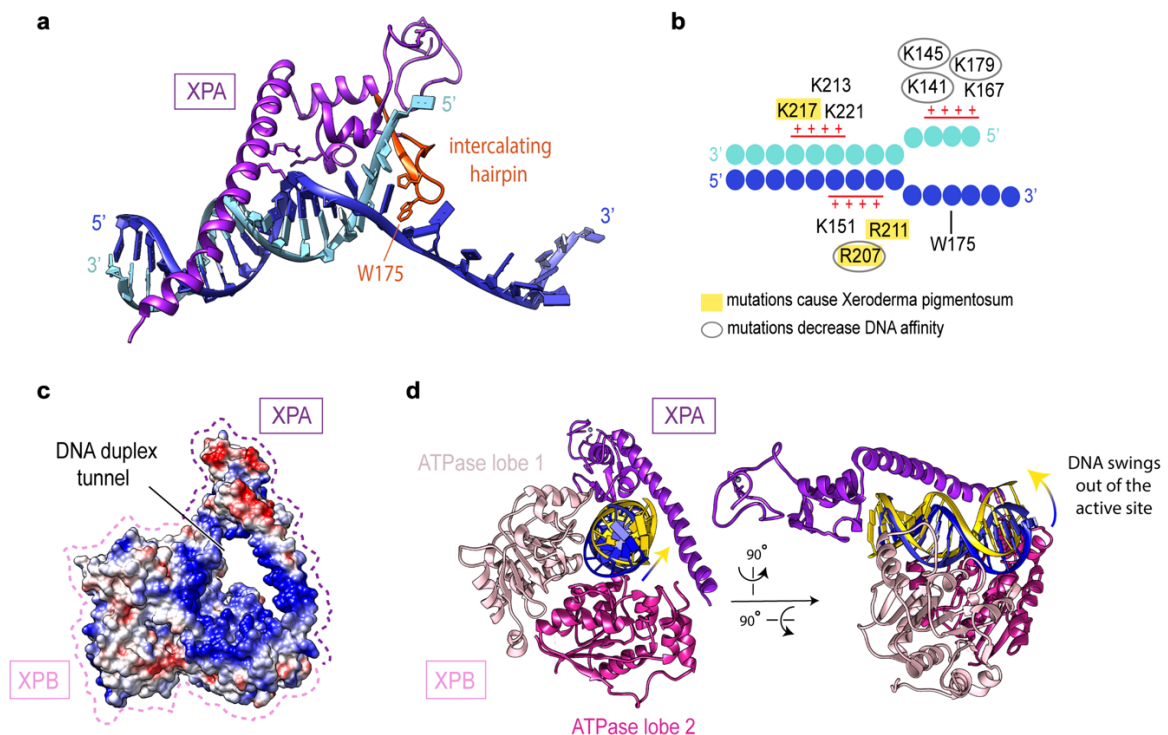
#### 2.2.4 XPA interactions with core TFIIH and DNA

XPA is the smallest NER protein, yet mutations in XPA cause the most severe symptoms of *Xeroderma pigmentosum*<sup>83</sup>. Despite decades of structural and biochemical investigation of XPA, its role in NER is still unclear. XPA is considered to be the main scaffolding protein in NER because it interacts with most of the other NER factors<sup>85-96</sup>. Our structure reveals the position of XPA within the NER machinery and the mode of its interaction with TFIIH and DNA. The structure also explains how XPA strongly stimulates XPB translocase activity as observed in our biochemical assays (Fig. 12d).

XPA contains an N-terminal zinc finger and a DNA-binding domain with an extended helix and an intercalating  $\beta$ -hairpin (Fig. 14). XPA forms an elongated arch over DNA that bridges between XPB and XPD (Fig. 14). XPA binds XPB and XPD with its extended helix and its intercalating hairpin, respectively. The XPA C-terminus further extends towards the p52 dimerization domain and TTDA. This could only be visualized when the EM-density was filtered to a lower resolution (Fig. 15d) due to the high flexibility of the structure in this region, but the interaction was additionally confirmed with crosslinking mass-spectrometry (Fig. 23c). Also, the interaction between XPA and TTDA explains why TTDA is required for XPA recruitment to the growing NER machinery *in vivo*<sup>209</sup>. Overall, the position of XPA between the core TFIIH ATPases and its elongated



shape might allow XPA to act as a scaffold which interacts with proteins at both junctions of a DNA repair bubble<sup>83</sup>.



**Figure 18 | XPA-DNA interactions.**

(a) Ribbon representation of the XPA-DNA structure. XPA intercalating hairpin is shown in orange and the rest of XPA is shown in purple. (b) Electrostatic interactions between XPA and the DNA junction. DNA nucleotides are indicated as circles. Patches of positively charged residues in proximity to the DNA backbone are indicated. Residues that are mutated in Xeroderma pigmentosum are highlighted in yellow. Mutation of encircled residues decreases DNA affinity. (c) DNA duplex tunnel formed by XPA and XPB. Blue, white and red color indicates positive, neutral and negative electrostatic surface potential, respectively. Created with UCSF Chimera<sup>210</sup>. (d) Two positions of DNA in the tunnel. Tightly bound DNA is in blue, dissociated DNA in yellow, ATPase lobe 1 of XPB in pink, ATPase lobe 2 in hotpink, and XPA in purple.

XPA binds to the duplex-single strand junction positioned between XPB and XPD which is consistent with previous biochemical data<sup>82</sup> (Fig. 14, 18a). During lesion scanning by XPD, this junction is located 5' of the lesion site and represents the 5' edge of the repair bubble<sup>71</sup>. The XPA intercalating hairpin inserts at the duplex-single strand junction and the conserved W175 located at the tip of the hairpin stacks against a DNA base of the DNA 3'-extension at the junction (Figs. 15i, 18a,b). Interaction between XPA and DNA is mostly electrostatic in nature. We observed several patches of positively charged residues near the sugar phosphate backbone which recognize different structural elements of the DNA junction (Fig. 18b). Mutations in several of the residues located in these patches were reported to cause Xeroderma pigmentosum<sup>83</sup>. In addition, many of the residues were previously shown to be important for DNA binding by XPA<sup>81</sup>. Precise positioning of XPA at the presumed 5' edge of the DNA repair bubble may provide a reference point for the organization of the NER assembly<sup>83</sup> as the assembly grows around the site of the DNA damage.

Our structure also suggests how XPA stimulates XPB translocation (Fig. 12d). XPB is not a processive enzyme and readily dissociates from DNA<sup>69,211</sup>. However, in our structure DNA is held in a positively charged DNA duplex tunnel that is formed between the extended helix of XPA and the XPB ATPase (Fig. 18c). XPA may thus retain DNA near the XPB active site. Indeed, our cryo-EM data revealed an alternative state of the complex with the DNA duplex disengaged from XPB but retained by the XPA extended helix (Figs. 13d, 18d). Thus, by trapping the DNA within the duplex tunnel, XPA may stimulate the XPB translocase activity and retain the NER machinery on the DNA during lesion scanning and processing.

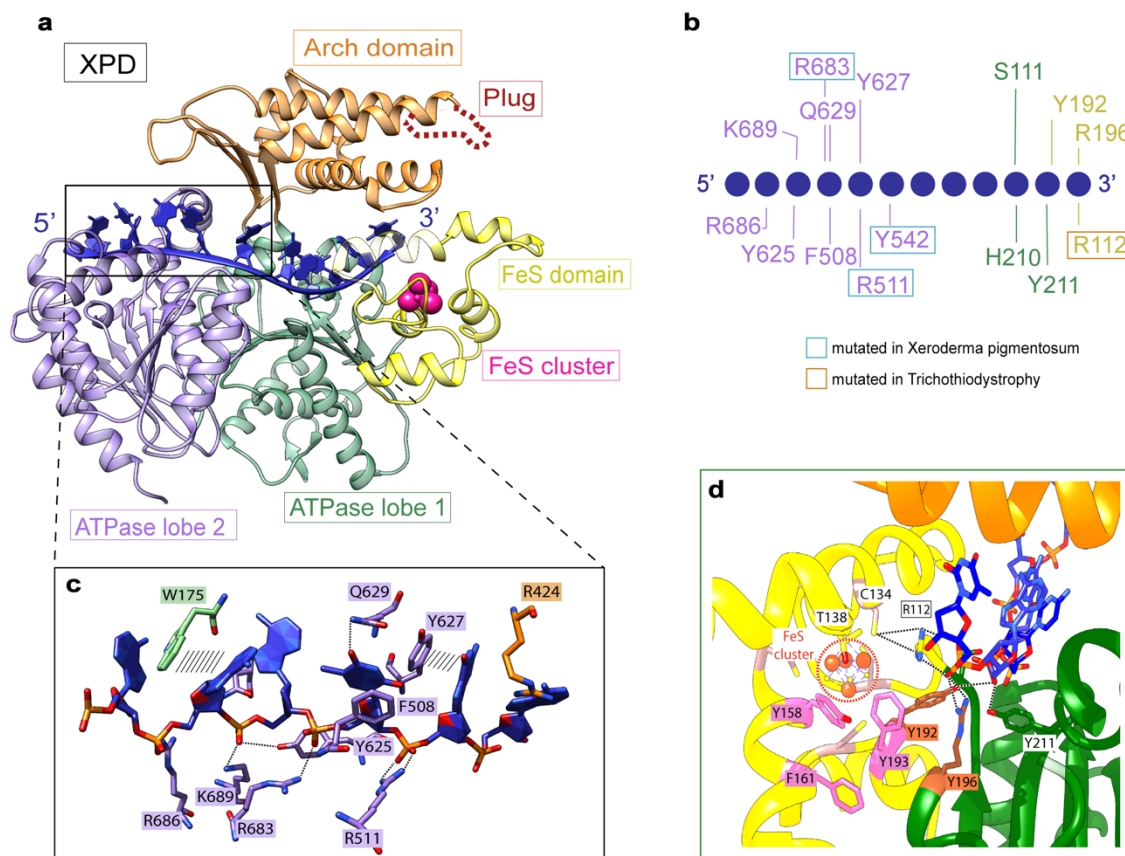
### 2.2.5 XPD-DNA interactions

The crystal structure of the archaeal XPD-DNA complex was previously reported, however, only a short DNA fragment bound to the ATPase lobe 2 was observed<sup>71</sup>. The 5' end of the short fragment was directed towards the ATPase lobe 2 and the 3' end towards the ATPase lobe 1, which suggested the polarity of DNA binding to XPD. Here we reveal the full trajectory of DNA through the human XPD helicase in the context of the TFIIF core (Figs. 14, 19). Focused classification and focused refinement with a mask around XPD significantly improved the cryo-EM map quality for this TFIIF subunit (Fig. 15b, 3.5 Å final resolution), which was important for visualization of DNA-protein contacts (Fig. 19b-d). We could resolve the density corresponding to individual bases and the sugar phosphate backbone in the areas where DNA is stabilized by interactions with the helicase. However, between the ATPase lobes, the density for DNA worsened presumably due to higher flexibility of the DNA.

The structure elucidates details on how XPD binds the DNA (Fig. 19). XPD consists of two ATPase lobes and the ATPase lobe 1 contains two inserted accessory domains, the Arch domain and the iron-sulphur (FeS) domain (Fig. 19a)<sup>78</sup>. The structure reveals extensive interactions between the 5' end of the DNA and the ATPase lobe 2 of XPD (Fig. 19c). Aside from multiple positively charged residues that engage the sugar-phosphate backbone, many contacts with the DNA bases are also formed, such as base stacking with the side chains of F508 and Y627 (Fig. 19c). In addition, the bases emerging from the helicase base stack with the side chain of W175 in XPA (Fig. 19c). This mode of interaction with the DNA is unusual for the SF2-family helicases which generally engage only the sugar-phosphate backbone<sup>212,213</sup>. It may be that extensive contacts with the DNA by XPA and XPD facilitate DNA bubble formation and XPD loading during the initial stages of NER.

The structure further suggests how XPD recognizes the lesion during DNA scanning. The DNA single strand extends into a pore formed by the ATPase lobe 1, the FeS domain, and the Arch domain (Fig. 19a). The sugar-phosphate backbone is bound by residues in the FeS domain, including R112, which is frequently mutated in *Trichothiodystrophy*<sup>214</sup>, and Y192 and R196, which are implicated in DNA lesion sensing<sup>215</sup> (Fig. 19b, d). These

interactions guide the DNA strand towards the FeS cluster which may be involved in the detection of DNA lesions<sup>74,216,217</sup>. Interestingly, R112 bridges the FeS cluster and DNA by interacting with the Fe-ligand C134 (Fig. 19d), which might be important for the lesion detection via DNA-mediated charge transfer<sup>74,216</sup>. Also, the FeS cluster is flanked by two protein pockets that are lined with aromatic residues Y158, F161 and F193 (Fig 19d), and may proof-read DNA bases, as observed for base excision DNA repair<sup>218</sup>.



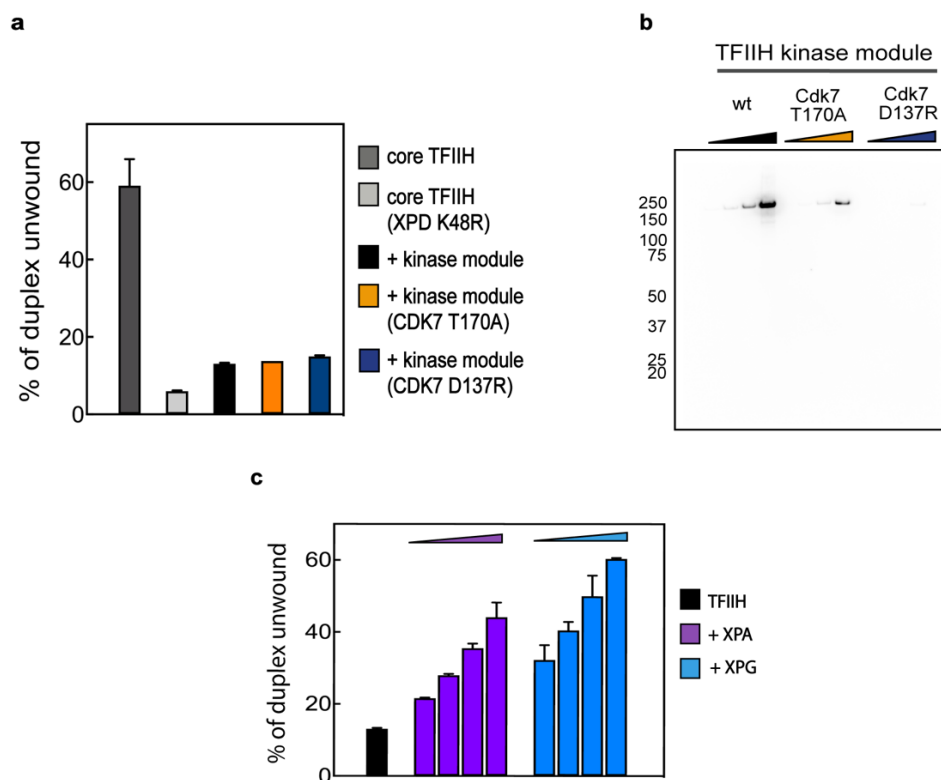
**Figure 19 | XPD-DNA interactions.**

(a) Ribbon representation of DNA-bound XPD. XPD domains ATPase lobe 1, FeS, Arch, and ATPase lobe 2 are in green, yellow, orange, and medium purple, respectively. DNA is in dark blue, plug element in red and FeS cluster in pink (b) Schematic representation of XPD-DNA interactions. The DNA-contacting residues mutated in human diseases are indicated. (c) Close-up view of the interactions between the DNA and the ATPase lobe 2. The amino acid residues are color-coded as in (a). XPA W175 is shown in green. (d) XPD-DNA contacts near the FeS cluster. The residues implicated in lesion sensing are shown in brown and the aromatic residues lining the FeS cluster in pink. The FeS cluster is indicated by the dotted red circle.

### 2.2.6 XPD inhibition by the kinase module and activation by XPA

We could reproduce the known repression of the XPD helicase activity by the TFIIH kinase module<sup>72,73,84</sup> in our helicase assay (Fig. 20a). To prevent that the ATPase activity of CDK7 inhibits XPD helicase by depleting the ATP levels, we performed the assays in the presence of an ATP regenerating system comprised of the phosphoenolpyruvate and the pyruvate kinase (Methods). However, it is still unclear if the repression is a result of the kinase module

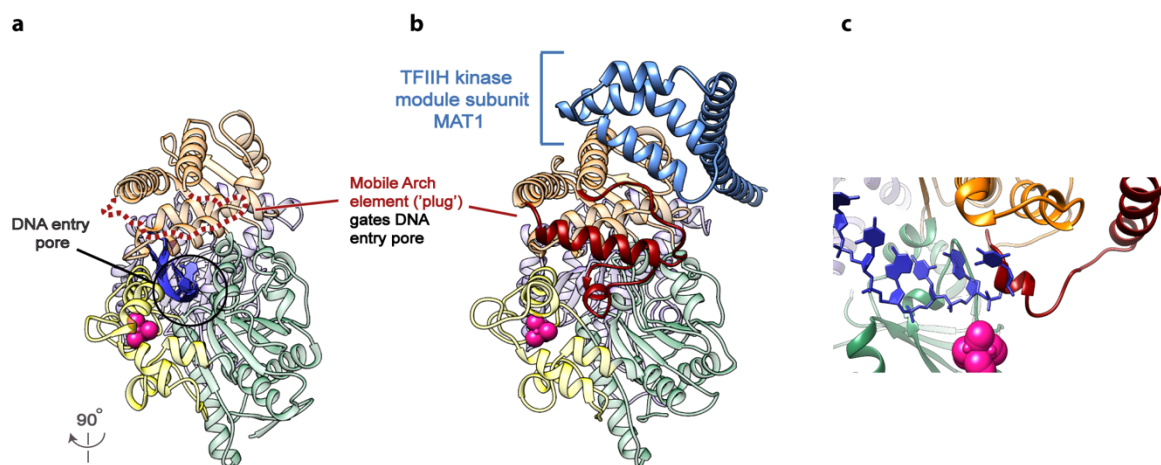
binding or XPD phosphorylation by CDK7. To test this, we designed and purified two variants of the kinase module which contain the catalytically impaired CDK7 kinase<sup>219</sup>. We confirmed the lack of kinase activity for the kinase module variants by measuring their ability to phosphorylate the C-terminal domain of the largest RNA polymerase II subunit (Fig. 20b), which is a natural CDK7 substrate during transcription initiation<sup>220,221</sup>. We found that catalytically inactive variants of the TFIIH kinase module could also repress XPD helicase activity, showing that repression does not require the kinase activity of CDK7 (Fig. 20a). In addition, we found that both XPA and XPG could relieve the kinase-mediated repression of XPD in a concentration-dependent manner (Fig. 20c). These observations show that XPA and XPG counteract the repressive effect of the kinase module on XPD, most likely by interrupting the kinase module binding to XPD.



**Figure 20 | XPD regulation by the kinase module.**

(a) Effect of kinase module variants on XPD helicase activity. Core TFIIH was incubated with a 2-fold excess of the kinase module and helicase activity was monitored as in Fig. 11b. The helicase activity of the ATPase deficient XPD-K48R mutant is shown as a control. Bars show the percentage of unwound product after 300s for two replicates. Error bars represent s.d. from the mean values. (b) Activity of kinase module variants. Different kinase module variants were incubated with the dephosphorylated yeast RNA-polymerase II and the phosphorylation status of the C-terminal domain of Rpb1 was probed by Western blotting. CDK7:T170A exhibits a much weaker kinase activity compared to the wild type CDK7, while the CDK7:D137R mutant is kinase dead. (c) Effect of increasing concentrations of XPA and XPG on XPD helicase activity in the presence of kinase module (0.25  $\mu$ M core TFIIH, 0.5  $\mu$ M kinase module, 0.375, 0.75, 1.5 or 3  $\mu$ M XPA or XPG). Otherwise as in (a).

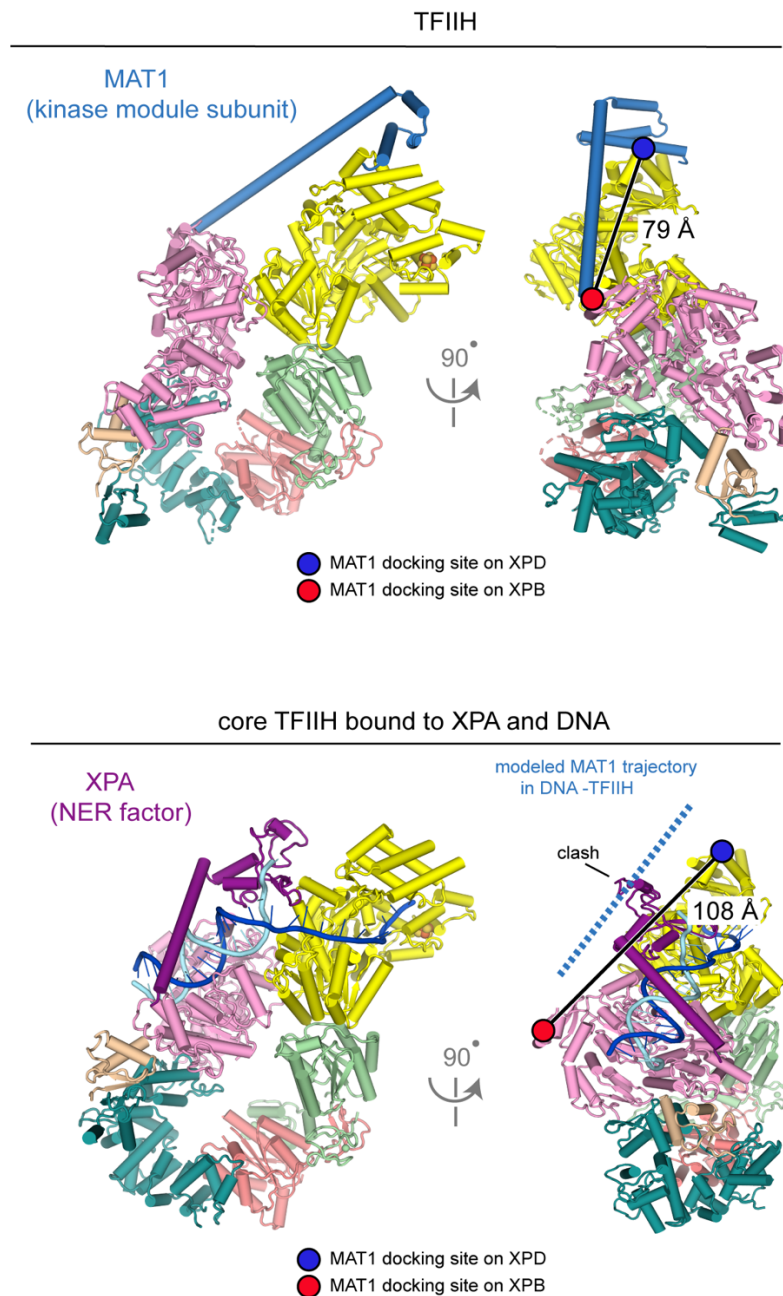
A comparison of our structure with the previous TFIIH structures shows how the kinase module represses XPD activity. In previous TFIIH structures<sup>70,182</sup>, a region in the XPD Arch domain forms a ‘plug’ (residues 273-325) which occupies the DNA pore of XPD (Fig. 21b). The plug would clash with DNA in the XPD pore (Fig. 21c), but is displaced and mobile in our structure (Fig. 21a). The kinase module subunit MAT1 contacts the plug and may stabilize it in the XPD pore (Fig. 21b), explaining how the kinase module impairs binding of core TFIIH to single stranded DNA<sup>73</sup> and inhibits XPD helicase activity. In addition, a loop in the yeast counterpart of p62 extends into the XPD active site during transcription initiation<sup>70</sup> and is predicted to interfere with the observed DNA trajectory through the helicase. Thus, multiple strategies are put into place to inhibit XPD when its helicase activity is not required.



**Figure 21 | Structural basis for XPD inhibition by the kinase module.**

(a) Side view of XPD bound to DNA. XPD domains ATPase lobe 1, FeS, Arch, and ATPase lobe 2 are in green, yellow, orange, and medium purple, respectively. DNA is in dark blue. A black circle depicts the DNA pore. (b) Side view of XPD bound to the kinase module (PDB code 5OF4)<sup>181</sup>. The plug in the Arch domain is in dark red, the kinase module subunit MAT1 in blue. (c) Overlay of XPD bound to the DNA and XPD bound to the kinase module with a zoom-in on the clash between the DNA and the plug.

Structural comparisons also suggest how XPA relieves XPD inhibition by the kinase module (Fig. 20c). XPA stabilizes TFIIH in a new conformation in which the two ATPases are drastically reoriented (Fig. 22). This conformation is incompatible with MAT1 binding as observed in the previous TFIIH structure<sup>181</sup>. The docking sites of MAT1 on XPB and XPD increase in distance for ~30 Å in the TFIIH conformation stabilized by XPA. In addition, the modeled trajectory of MAT1 in the TFIIH-XPA-DNA structure would clash with the XPA zinc finger (Fig. 22). This explains how XPA facilitates kinase module removal upon NER induction *in vivo*<sup>84</sup>. Taken together, MAT1 and XPA stabilize two entirely different conformations of TFIIH, which contain the repair helicase XPD in an inactive or an active state, respectively.

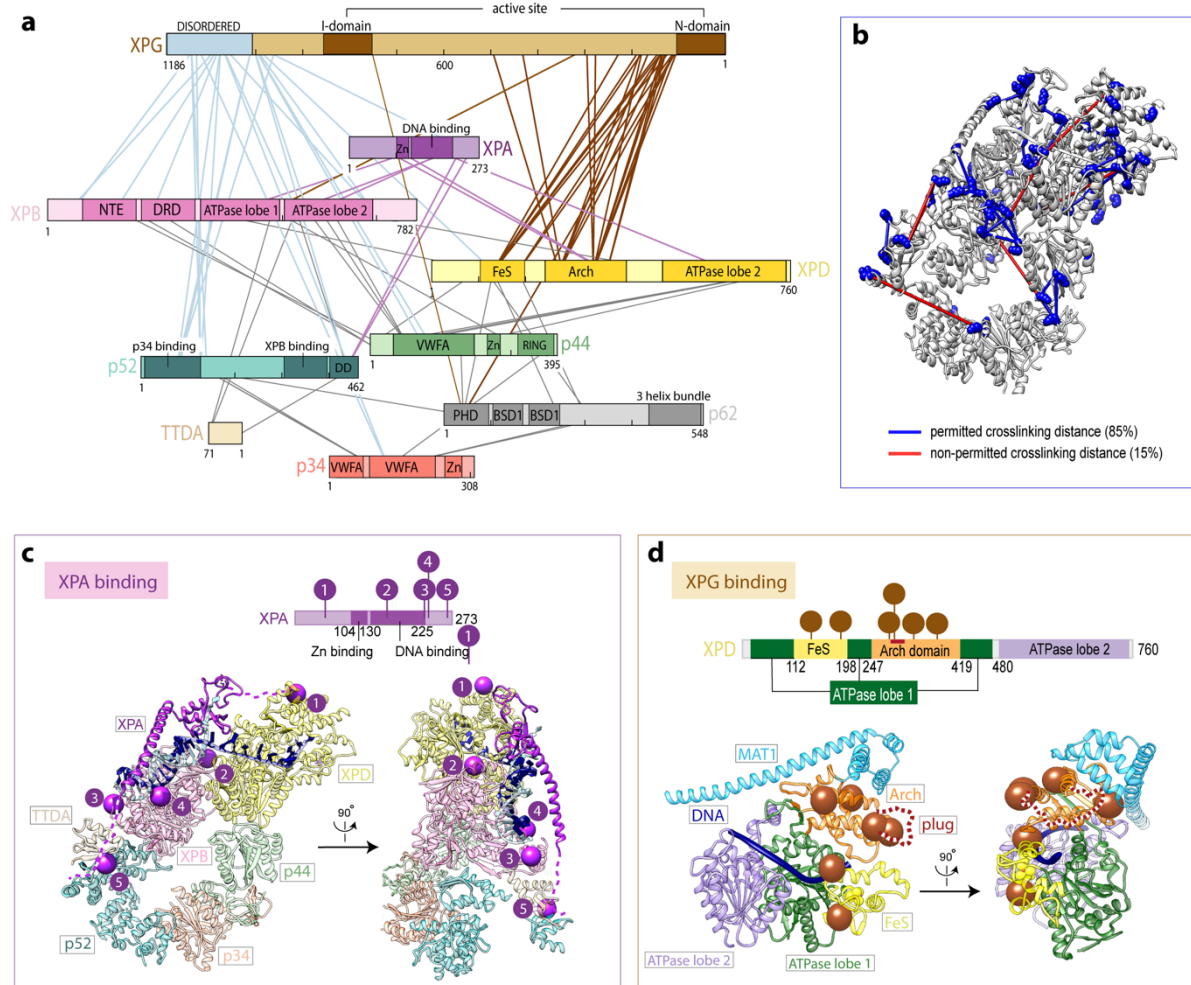


**Figure 22 | Structural basis for XPD activation by XPA.**

Kinase-bound core TFIIH (upper panel, produced from the previous structure<sup>182</sup> as described in Fig. 17) and DNA-bound core TFIIH-XPA (lower panel) structures are shown. The structures are color-coded as in Fig. 14, MAT1 kinase module subunit is shown in blue. Front views (left) show that MAT1 and XPA bridge the core TFIIH ATPases in the kinase-bound structure and in the DNA-bound structure, respectively. Side views (right) were aligned on XPD and show how XPB and XPD reorient during DNA binding. Distance between the docking site of MAT 1 on XPB (red circle, Q190 residue was used for distance measurement) and XPD (blue circle, V365 was used for distance measurement) on both structures is indicated. The distance increases significantly after DNA-binding which would induce the kinase module rearrangement on the core TFIIH and may facilitate kinase module dissociation *in vivo*<sup>84</sup>. The MAT1 helix trajectory that spans between XPB and XPD was modeled in the DNA-bound structure (dotted blue line) and suggests a potential clash with XPA.

### 2.2.7 XPG binding to the core TFIIH-XPA-DNA complex

Since XPG was not visible in our structure, we located it in a core TFIIH-XPA-XPG-DNA complex by chemical crosslinking (Fig. 23a). The quality of the obtained crosslinking network was validated by mapping the crosslink sites on the core TFIIH-XPA-DNA structure, which showed that 85% of crosslink sites fall within the permitted crosslinking distance of 30 Å (Fig. 23b). XPA-TFIIH crosslinks match the XPA position on the TFIIH core observed in our structure (Fig. 23c), further validating the crosslinking results.



**Figure 23 | Crosslinking mass-spectrometry network of the core TFIIH-XPA-XPG-DNA complex.**

(a) Schematic representation of the crosslinking network. Crosslinks between XPA and core TFIIH are shown in purple, the disordered XPG C-terminus and core TFIIH in blue, and XPG N-terminal region and core TFIIH in brown. Crosslinks between the core TFIIH subunits are in gray. Only crosslinks with a score above 6 are shown. The list of inter-subunit crosslinks is provided in the Supplemental Table 2. Domain abbreviations as in Fig. 9. (b) Validation of crosslinking sites. Crosslinks with a score above 6 were mapped onto core TFIIH-XPA-DNA structure. Colored rods connecting crosslinked residues represent permitted (blue) or non-permitted (red) crosslinking distances, and lysine residues are shown as blue spheres. 85% of mapped crosslinks fall within the permitted crosslinking distance. Created with UCSF Chimera<sup>210</sup> and Xlink Analyzer<sup>222</sup> (c) XPA crosslinks mapped onto core TFIIH-XPA-DNA structure. XPA crosslinks are shown as purple spheres. Schematic representation of XPA with mapped crosslinks is shown above the structure. (d) XPG crosslinks mapped onto XPD-DNA structure. XPD domains are color coded as in Fig. 19, the MAT1 subunit of the kinase module is shown in blue. The position of the MAT1 subunit was modeled based on the previous TFIIH structure<sup>182</sup>. XPG crosslinks are shown as brown spheres. Schematic representation of XPD with mapped crosslinks is shown above the structure.

The disordered XPG C-terminus crosslinks mostly to XPB and p52 subunits and the structured XPG N-terminus specifically crosslinks to the FeS and Arch domains in XPD, including the regulatory plug element (Fig. 23a, d). This is in agreement with the observation that XPG strongly stimulates XPD helicase activity (Fig. 12c), possibly by increasing the mobility of the interface between the Arch and FeS domains, which is required for the efficient DNA unwinding by XPD<sup>223</sup>. In addition, XPG crosslinks overlap with the binding site of the kinase module on XPD (Fig. 23d), suggesting that XPG competes with the kinase module for XPD binding to relieve the kinase inhibition of XPD (Fig. 20c). It is still uncertain if the kinase module dissociates before or concomitant to XPG recruitment to the NER machinery<sup>84</sup>. However, since XPG displaces XPC<sup>63</sup>, sequesters the docking site for the inhibitory kinase module on XPD (Fig. 23d) and strongly stimulates XPD helicase activity (Fig. 12c), XPG may aid XPA in initiating the lesion scanning step of the NER reaction pathway. Moreover, XPG binds XPD close to the presumed 3' repair bubble junction located near the FeS cluster in XPD, which is in agreement with the observation that XPG makes the 3' incision during NER<sup>107,224</sup>.



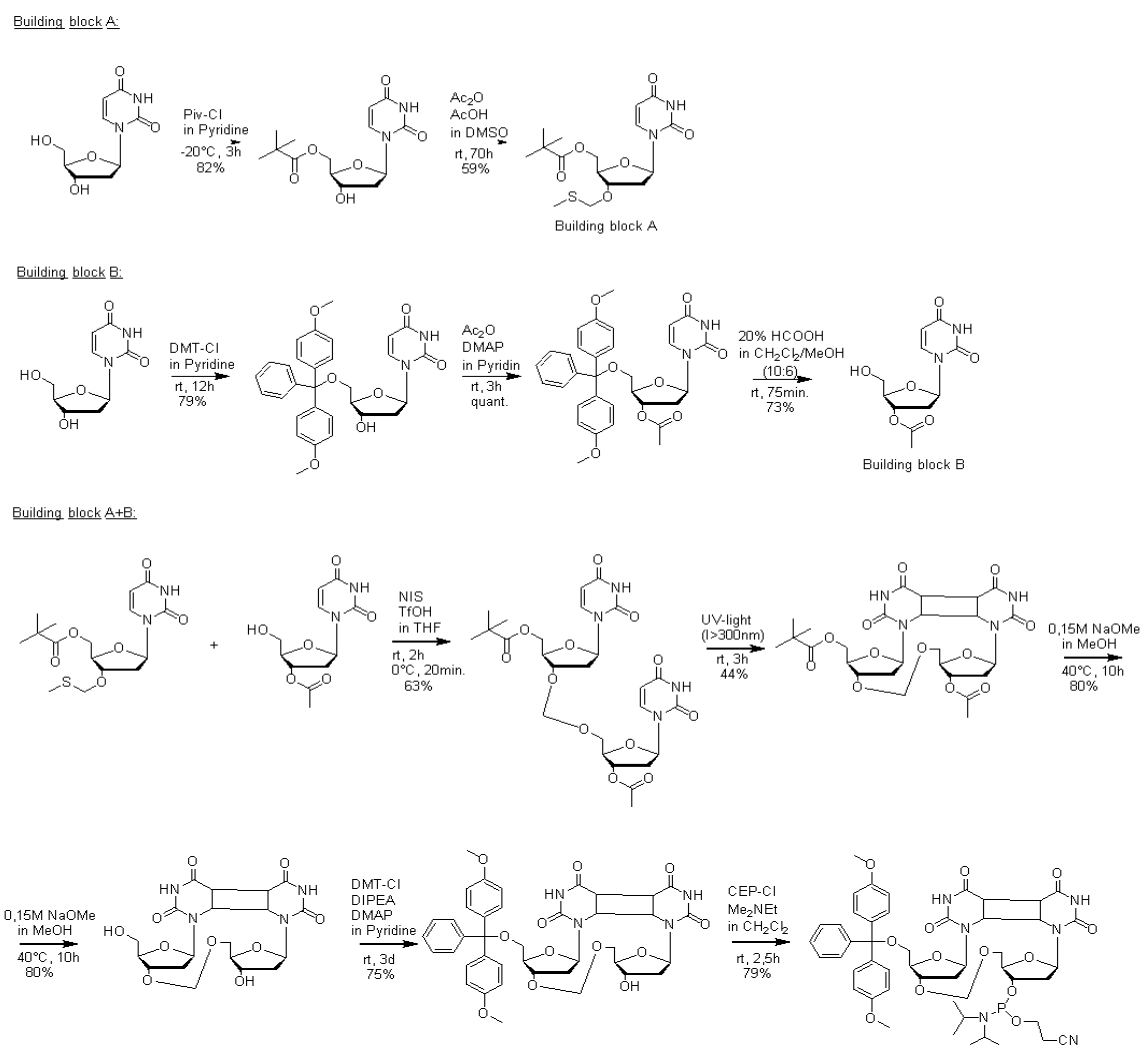
## 2.3 Reconstitution and analysis of TCR intermediates

### 2.3.1 Biochemical characterization of TCR initiation

TCR is initiated by an RNA-polymerase II (RNAP) arrested on a DNA lesion<sup>152,153</sup>. The most important group of DNA lesions removed by TCR are the lesions induced by UV-irradiation, such as cyclobutane pyrimidine dimers (CPD) and pyrimidine-pyrimidone 6–4 photoproducts (6-4PP), which accumulate in human skin after sun exposure<sup>225-230</sup>. Thus, in order to reconstitute TCR initiation *in vitro* we purified the mammalian RNAP from pig thymus and also produced a synthetic oligonucleotide that contained the physiological TCR substrate CPD (Methods).

A CPD contains a cyclobutane ring which forms between the two neighboring thymine bases upon UV-irradiation<sup>231</sup>. Organic synthesis of a chemical CPD analogue (called CPD throughout the thesis) was previously described<sup>232</sup>, however, we modified several synthesis steps in order to improve the final yield (Fig. 24, Methods). The lesion was synthesized in a form suitable for solid-phase DNA synthesis. Thus, we used a CPD building block to synthesize a 14-nucleotide long DNA strand with a centrally positioned CPD lesion (Table 12). A longer DNA scaffold was needed for the assembly of the RNAP elongation complex, so we expanded the short CPD-containing oligonucleotide on both the 3' and 5' end by enzymatic ligation with two additional oligonucleotides (Methods). The final DNA strand was purified on a sequencing gel to ensure DNA length homogeneity and to remove the shorter ligation byproducts.

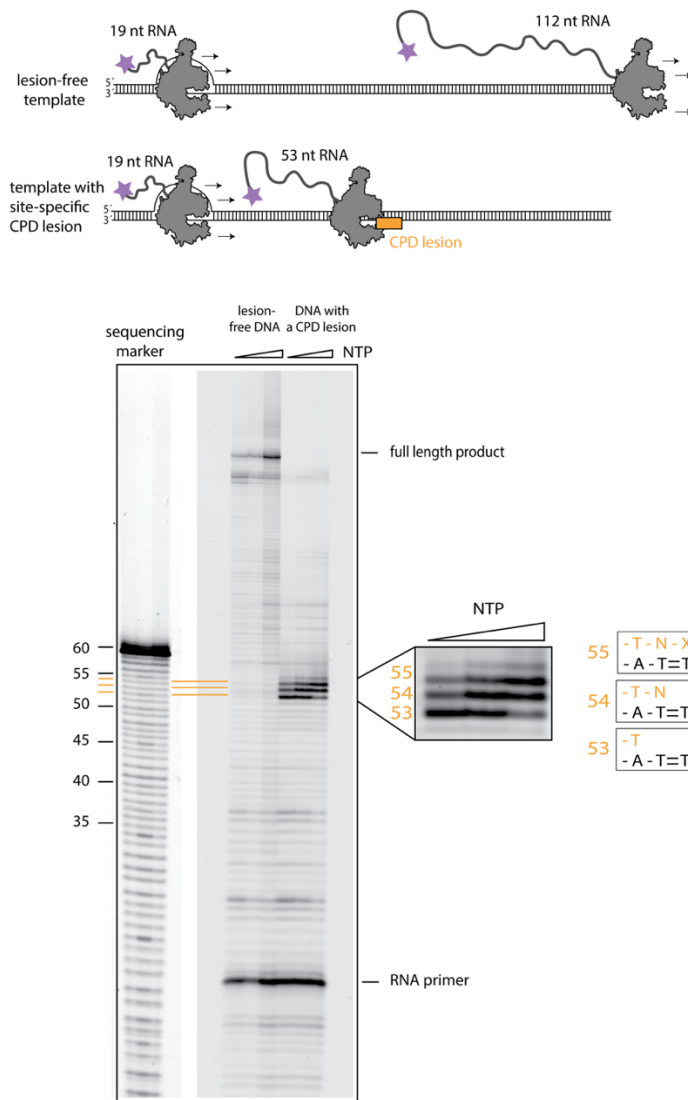
We verified that the CPD lesion arrests the RNAP progression<sup>51,233,234</sup> in our *in vitro* transcription system (Fig. 25). We assembled a ternary complex between the purified pig RNAP, the template DNA strand (TS) which contained a CPD lesion and the RNA primer which was 5' labeled with a Cy5 dye. As a control, we used an identical TS lacking the lesion. We finalized the formation of the elongation complex by the addition of the non-template DNA strand (NTS) and we monitored the extension of the RNA primer by RNAP in the presence of increasing concentrations of NTP (Fig. 25). The transcription products were resolved on a sequencing gel alongside a marker produced by alkaline hydrolysis of a synthetic RNA which was identical in sequence to the expected RNA transcript of our model DNA (Fig 25). Many bands corresponding to the abortive transcription products were observed, as was expected in the absence of elongation factors which chaperone RNAP or rescue backtracked RNAP complexes<sup>235,236</sup>.



**Figure 24 | Organic synthesis of a chemical CPD analogue.**

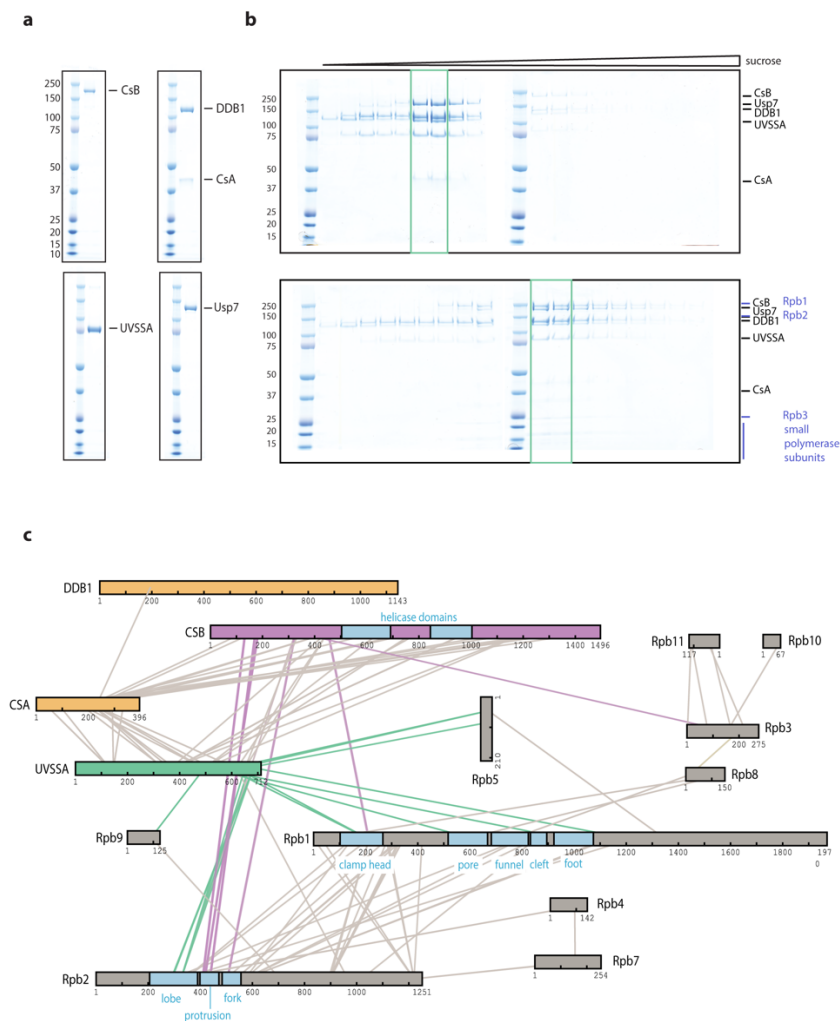
Detailed description in Methods.

Additionally, a strong signal from the RNA primer was observed because an excess of the nucleic acid scaffold with regard to RNAP had to be used to maximize the efficiency of the elongation complex formation (Fig. 25). In control reactions lacking the CPD lesion we observed the full-length RNA products made by run-off transcription of the DNA template (Fig 25). However, in the presence of a CPD lesion we detected the accumulation of shorter RNA products, 53-55 nucleotides in length. The 53 nucleotides long RNA is a product of an RNAP arrest in front of the CPD lesion, whereas further addition of one and two nucleotides extends the RNA transcript to the first and second thymidine in the CPD lesion (Fig. 25). However, no significant transcription was observed past the lesion site indicating that a CPD indeed poses a permanent block to the elongating RNAP, which is in agreement with previous studies<sup>51,233,234</sup> and validates out in vitro TCR system.



**Figure 25 | RNAP is arrested by a CPD lesion.** Schematic representation of the DNA scaffold with or without the site-specific CPD lesion is shown above the gel. RNA is labeled with a 5' Cy5 dye (purple star), and the CPD lesion is indicated by the yellow square. Lengths of the expected RNA products are indicated. The gel shows the products of the in vitro transcription reaction using the lesion-free or the CPD-containing DNA in the presence of increasing NTP concentrations. The RNA was separated on a denaturing sequencing gel alongside a sequencing marker. The lesion-free DNA was transcribed until the run-off (full length) product was made. In the presence of a CPD lesion, 3 discrete RNA products of 53, 54 and 55 nucleotides were observed. The zoom-in panel shows the magnified arrest bands. The schematic on the right shows where on the DNA scaffold RNAP did arrest: the 53 nucleotides long RNA corresponds to the arrest in front of the CPD lesion (indicated with T=T), the 54 nucleotides long RNA incorporated a nucleotide opposite to the first thymidine (T) in the CPD lesion and the 55 nucleotides long RNA opposite to the second T. With increasing NTP concentrations more of the RNA transcripts were elongated to 55 nucleotides.

The next step in reconstituting the TCR initiation was to demonstrate that the RNAP arrested on a CPD lesion can be recognized by TCR-specific factors, as shown in vivo by pull-down experiments<sup>139</sup>. We have established purification protocols for two major TCR-specific proteins CsB and CsA, as well as the two additional proteins UVSSA and Usp7 which were implicated in TCR only recently<sup>130-132</sup> (Methods, Fig. 26a). CsA alone is insoluble so we purified it in a binary complex with DDB1, its protein partner within the E3 ubiquitin ligase involved in TCR<sup>145</sup>. Interestingly, we observed that the TCR-specific factors form a stable complex even in the absence of RNAP, as all factors comigrated on a sucrose gradient in apparent stoichiometric amounts (Fig. 26b). In addition, in the presence of an RNAP that was transcribed into a CPD lesion, the TCR-specific factors shifted into lower fractions of the sucrose gradient, indicating that the factors formed a complex with the arrested RNAP (Fig. 26b).



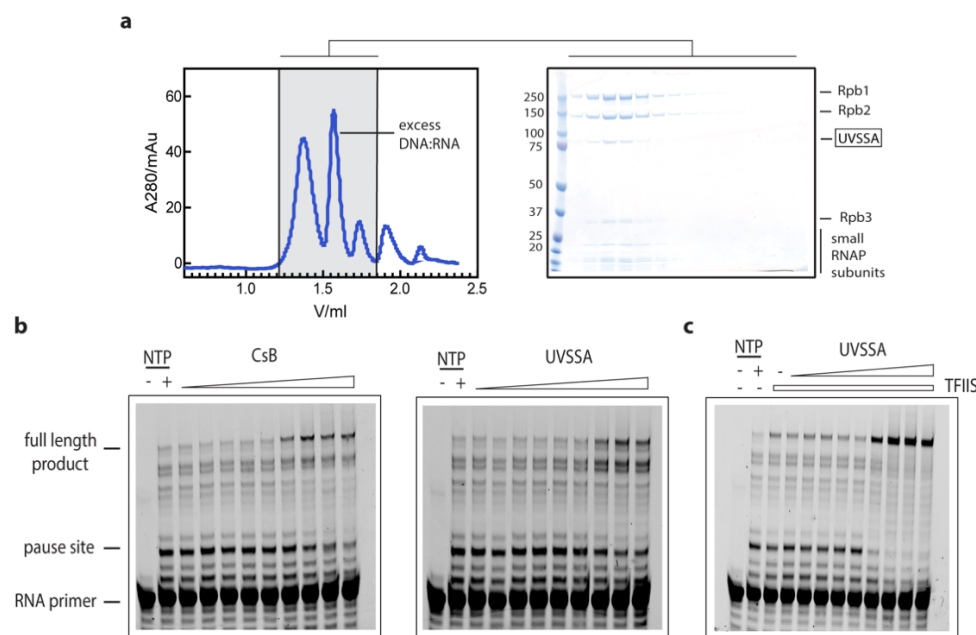
**Figure 26 | Complex formation between the RNAP and TCR-specific factors.**

(a) Purified TCR-specific factors (CsB, CsA-DDB1, UVSSA and Usp7) were analyzed by SDS-PAGE and visualized by Coomassie staining. (b) Binding of TCR-specific factors to RNAP. (above) TCR-specific factors were mixed in equimolar ratio and separated on a 10 – 30% sucrose gradient. The gradient was fractionated in 200  $\mu$ l fractions and each fraction was analyzed by SDS-PAGE followed by Coomassie staining. All factors comigrate in apparently stoichiometric amounts which indicates that they form a complex. (below) The same protein mixture was incubated with purified pig RNAP transcribed into a CPD lesion and analyzed by a 10-30% sucrose gradient as above. The peak fractions for the TCR-specific proteins were located in lower fractions of the gradient in the presence of RNAP (peak fractions are indicated by the green frames) which indicates that TCR-specific factors bind RNAP. (c) The crosslinking network between RNAP and the TCR-specific factors. The complex was formed as in (b) but purified by size exclusion chromatography and crosslinked with BS3. CsA-DDB1 is shown in yellow, CsB in pink, UVSSA in green and the RNAP subunits in gray. Selected domains are shown in blue. Crosslinks between RNAP and CsB or UVSSA are shown in pink or green, respectively. The list of inter-subunit crosslinks is provided in the Supplemental Table 3.

To determine which of the factors directly interact with RNAP, we chemically crosslinked the lesion arrested RNAP with the CsB-CsA-DDB1-UVSSA complex and analyzed the protein interaction sites with mass-spectrometry (Fig. 26c). The crosslinking network of the TCR initiation complex shows that CsB and UVSSA mediate the interaction with RNAP. The two factors are bridged by CsA, consistent with previous immunoprecipitation studies<sup>237,238</sup>. CsB primarily binds the second largest RNAP subunit Rpb2, as was observed for the yeast CsB homologue Rad26<sup>138</sup>. Most of the crosslinks were mapped between the RNAP protrusion and the N-terminus of CsB (Fig. 26c). Since the N-

terminal region of CsB autoinhibits the ATPase activity of the protein<sup>239</sup>, binding to RNAP might relieve this autoinhibition and activate the enzymatic activity of CsB.

Surprisingly, UVSSA occupies even more of the RNAP surface compared to CsB and strongly crosslinks to the lobe and the clamp regions in RNAP (Fig. 26c). To confirm that UVSSA directly interacts with RNAP we probed the interaction by analytical size-exclusion chromatography (Fig. 27a). UVSSA coelutes with RNAP despite being much smaller than RNAP, which indicates that UVSSA binds RNAP even in the absence of other TCR-specific factors. We have also identified UVSSA as a novel positive transcription elongation factor because we observed that UVSSA stimulates the RNA extension by RNAP in vitro, as previously reported only for CsB<sup>137</sup> (Fig. 27b). In addition, the crosslinking data showed that UVSSA binds near the RNAP pore, a known binding site for TFIIS<sup>240</sup> (Fig. 26c), which suggests that these factors might functionally interact. TFIIS is a well-studied elongation factor which stimulates intrinsic RNAP endonuclease activity<sup>156-159</sup> in order to rescue backtracked RNAP complexes and stimulate transcription elongation<sup>241</sup>. We have indeed observed that UVSSA strongly enhances the TFIIS effect on transcription elongation in vitro (Fig. 27c), which might be important during TCR for the transcription restart after lesion removal from the DNA.



**Figure 27 | The effect of CsB and UVSSA on transcription in vitro.**

(a) UVSSA forms a complex with RNAP. The chromatogram shows the size-exclusion elution profile of the RNAP-UVSSA mixture. Fractions were analyzed by SDS-PAGE and the proteins were visualized by Coomassie staining. UVSSA (~ 70 kDa) coelutes with the much larger RNAP (~ 500 kDa) indicating that UVSSA and RNAP form a gel-filtration stable complex. (b) In vitro transcription assay (0.8 pmol RNAP) was performed in the presence of increasing concentrations of CsB (left, the highest amount was 30 pmol of CsB, followed by a half-log dilution series) or UVSSA (right, the highest amount was 30 pmol of UVSSA, followed by a half-log dilution series). Control reactions without the addition of NTPs or corresponding factors are indicated. Both CsB and UVSSA stimulate transcription in vitro as seen by the increase in the amount of full-length RNA products and the decrease in the amount of shorter RNA products. (c) UVSSA stimulates TFIIS activity. We determined the TFIIS concentration which moderately stimulates RNAP transcription (100 nM; third lane which contains TFIIS but not UVSSA) and we performed the in vitro transcription reaction with increasing amounts of UVSSA (as in (b)). The observed transcription stimulation was much stronger than in the presence of UVSSA alone, likely due to the stimulatory effect of UVSSA on TFIIS.

Since we were able to reconstitute the initiation of TCR in vitro, we tried to solve the cryo-EM structure of the lesion arrested RNAP decorated with TCR-specific factors. We tried different combinations and truncations of TCR factors, different DNA scaffolds (longer upstream or downstream DNA, or both), post-translational modification of RNAP (phosphorylation and ubiquitination), different crosslinking strategies, detergents and support materials for particles on cryo-EM grids (carbon and graphene oxide), as all those factors can influence the complex formation and stability. However, we did not observe any additional density on RNAP after the 3D reconstruction of the complex. It could be that the complex falls apart during the sample vitrification or it is too flexible and heterogeneous for cryo-EM.

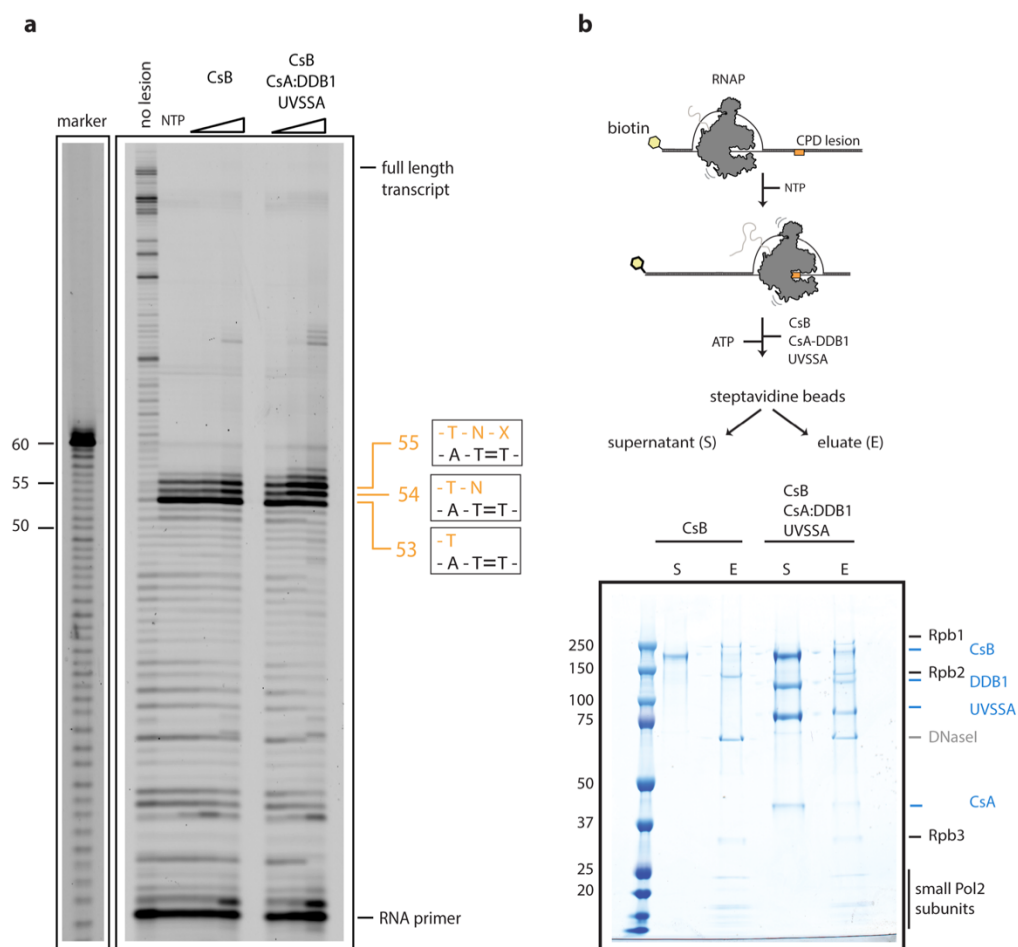
### 2.3.2 The fate of lesion-arrested RNAP

The lesion-arrested elongation complexes are highly cytotoxic because they lead to cell-cycle arrest and can ultimately trigger programmed cell death<sup>120,242-245</sup>. In addition, the large RNAP footprint on the DNA prevents the repair enzymes to access the lesion<sup>123,234</sup>. It is thus necessary to quickly identify and remove the arrested elongation complexes. Several fates for the arrested RNAP can be envisioned (Fig. 3). First, RNAP can efficiently transcribe over some smaller lesions which may lead to mutagenesis if a wrong nucleotide is paired with the modified DNA base<sup>246</sup>. Second, RNAP can be degraded by the proteasome. Several components of the TCR machinery undergo ubiquitination, which includes RNAP mono- and poly-ubiquitination by Nedd4 and the elongin complex<sup>148,160</sup>. Third, RNAP can be pushed forward to transcribe across the transcription-blocking lesion which destabilizes the elongation complex and induces RNAP dissociation from the DNA<sup>80</sup>. Fourth, RNAP can be pushed backwards on the DNA far enough to expose the lesion to repair<sup>80</sup>. This would also allow the restart of transcription after the lesion has been removed.

We confirmed that RNAP alone cannot transcribe past a CPD lesion<sup>51,233,234</sup> (Fig. 25). However, both CsB and UVSSA stimulate the RNAP elongation in vitro, so we tested if these factors could chaperone transcription across the lesion. We assembled the elongation complex on a CPD-containing DNA scaffold and added CsB or the CsB-CsA-DDB1-UVSSA complex to the reaction before starting the RNA extension (Fig. 28a). Even in the presence of TCR-specific factors we did not detect any significant transcription past the site of the CPD lesion. However, we observed a stronger signal for the 54 and 55 nucleotides long RNA products with increasing CsB concentrations, suggesting that RNAP was able to translocate onto the CPD lesion more efficiently in the presence of CsB (Fig. 28a), as shown before<sup>137</sup>. This effect was even stronger in the presence of UVSSA, likely due to the synergistic effect of CsB and UVSSA on transcription elongation.

Next, we tested if the TCR-specific factors can dislodge the RNAP arrested on a CPD lesion (Fig. 28b). We designed an assay which allows us to separate the DNA-bound RNAP from the dissociated RNAP. We formed the elongation complex with the CPD containing

TS and the biotinylated NTS. The biotin-tag on the NTS was used to immobilize the RNAP on the streptavidin-coated beads. An *in vitro* transcription reaction was performed to position the RNAP in front on the CPD lesion. The lesion-arrested RNAP was coupled to a solid support and the unbound RNAP was washed away. We added a large excess of TCR-specific factors to the reaction, followed by incubation at 37 °C for 3 h. The reaction was supplemented with ATP to support the CsB translocase activity which may be necessary to drive the RNAP dissociation from the DNA. After the removal of the supernatant, the DNA bound proteins were specifically eluted with DNase I. We visualized the protein composition in the supernatant and the eluate by SDS-PAGE and Coomassie staining (Fig. 28b). The supernatant contained the excess of TCR-specific factors.



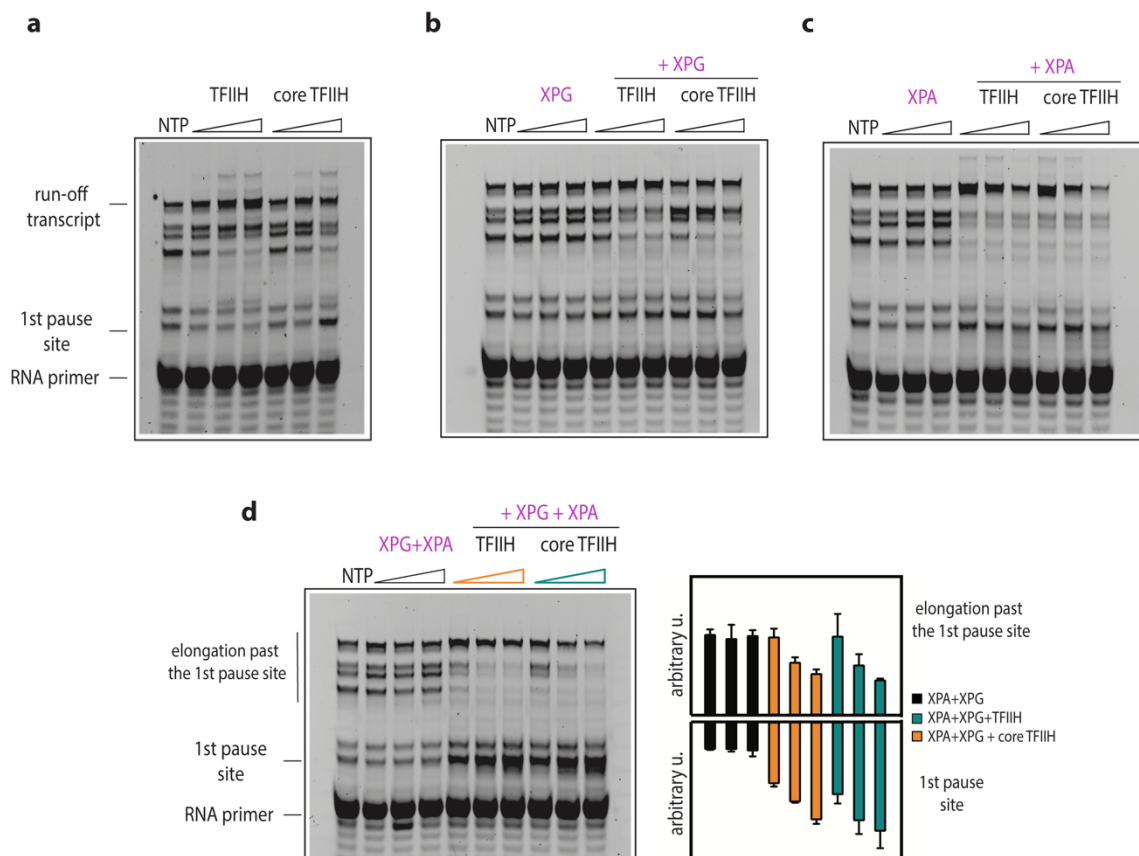
**Figure 28 | TCR-specific factors do not stimulate transcription past the CPD lesion nor the dissociation of the lesion-arrested RNAP from the DNA template.**

(a) *In vitro* transcription (2.8 pmol RNAP) of a CPD-containing DNA scaffold in the presence of increasing concentrations of CsB (5, 10 and 20 pmol) or the CsB-CsA-DDB1-UVSSA complex (5, 10 and 20 pmol). The RNA products were resolved on a denaturing sequencing gel and visualized by fluorescence scanning of the labeled RNA. RNAP arrested on a CPD lesion as observed by the appearance of 53-55 nucleotides long RNA products. No significant transcription past the CPD lesion was observed in the presence of the TCR factors. (b) Schematic representation of the pull-down assay used to determine if TCR-specific factors can dissociate the CPD-arrested RNAP from the DNA. RNAP was transcribed into a CPD lesion, coupled to streptavidin beads and incubated with ATP and 6x excess of TCR-specific factors. After a wash, DNA-bound proteins were specifically eluted with DNase I. Supernatant (S) and elution (E) fractions were analyzed by SDS-PAGE and Coomassie staining. RNAP subunits were only detected in the DNA-bound fractions, suggesting that TCR-specific factors cannot dissociate CPD arrested RNAP from the DNA.

However, the RNAP was only present in DNase elution fractions indicating that TCR-specific factors cannot dislodge the lesion-arrested RNAP from the DNA (Fig. 28b). Moreover, the TCR factors coeluted with the RNAP in apparently stoichiometric amounts suggesting that the factors did bind the RNAP in the assay.

Thus far we excluded RNAP removal from the lesion site by transcription read-through or dissociation from the DNA (Fig. 28). Next, we investigated if RNAP could be actively pushed backwards during TCR. Active RNAP backtracking would require an ATP-driven motor protein to push or pull the RNAP in the direction opposite to the direction of transcription<sup>154</sup>. Three ATPases from the SF2-superfamily of helicases suitable for such task are involved in TCR: CsB, XPB and XPD<sup>133</sup>. CsB can be excluded because CsB stimulates forward RNAP movement in vitro (Fig. 27b)<sup>137</sup>. Thus, we tested the effect of TFIIH, which contains both XPB and XPD, on transcription. We used a DNA scaffold that is transcribed with a medium efficiency in order to easily detect the RNAP stimulation or inhibition by accessory factors. After screening the several naturally occurring arrest sequences<sup>247</sup>, we decided to use the pML5-4NR sequence as we observed a ladder of 6 discrete bands after transcribing the sequence in our in vitro transcription system (Fig. 29a). In the presence of TFIIH or TFIIH core we observed a very subtle stimulation of transcription (Fig. 29a). However, our previous biochemical characterization of the core TFIIH ATPases showed that XPA and XPG NER factors are required for the optimal activity of XPB and XPD (Fig. 12). We showed that XPA and XPG alone do not affect the RNA extension (Fig. 29 b, c). In the presence of XPG, TFIIH again showed a modest stimulation of RNAP activity, suggesting that a strong stimulation of the XPD helicase activity by XPG (Fig. 12) does not change the effect of TFIIH on transcription (Fig. 29b). In contrast, in the presence of XPA which stimulates both XPD<sup>56</sup> and XPB (Fig. 12) we observed a much stronger transcription stimulation by TFIIH and almost all RNA products were extended to the length of the run-off transcript (Fig. 29c). It may be that the increased translocase activity of XPB on the DNA duplex in the presence of XPA drives RNAP forward when TFIIH binds the upstream DNA and translocates into the RNAP, which might reactivate stalled or backtracked elongation complexes. Surprisingly, in the presence of both XPA and XPG, TFIIH caused transcription inhibition, as observed by the accumulation of short RNA products (Fig. 29d). The in vitro transcription was specifically inhibited at the first pause site (Fig. 29d) indicating that paused polymerase may be a substrate for the TFIIH-XPA-XPG complex.



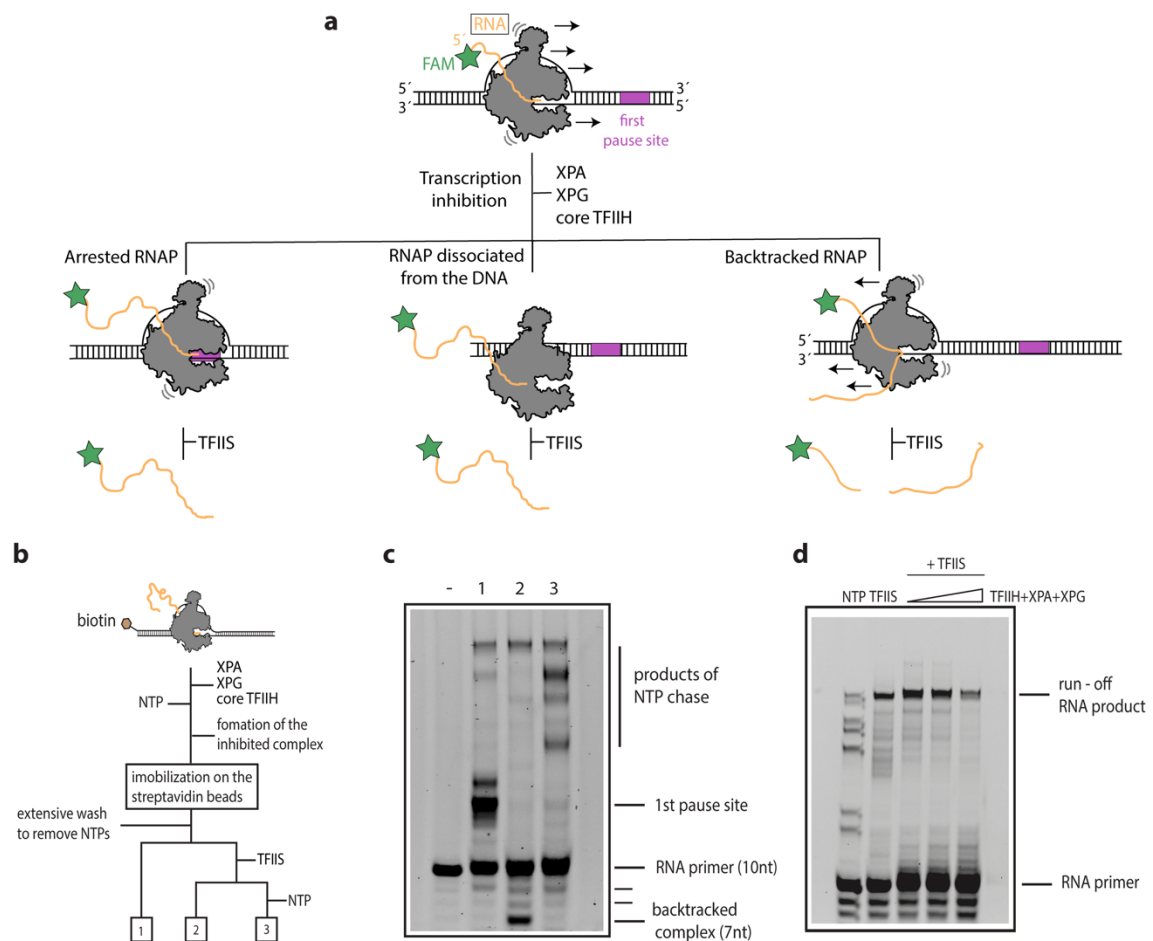


**Figure 29 | Core TFIIH inhibits transcription in the presence of XPG and XPA.**

(a-d) Naturally occurring pML5-4NR arrest DNA sequence<sup>247</sup> was used for the *in vitro* transcription reactions. Transcription of this sequence produces a ladder of 6 distinct bands we could monitor in the presence of additional factors. RNA products were resolved by denaturing PAGE and visualized by fluorescence scanning. Gels are representative of 2 experimental replicates. (a) Transcription reaction (0.8 pmol RNAP) was performed in the presence of increasing amounts of TFIIH or core TFIIH (2, 5 and 10 pmol). (b) The effect of XPG on transcription in the absence or presence of TFIIH (as (a) + 6, 15 and 30 pmol of XPG). (c) The effect of XPA on transcription in the absence or presence of TFIIH (as (a) + 6, 15 and 30 pmol of XPA). (d) The effect of XPA and XPG on transcription in the absence or presence of TFIIH (as (a) + 6, 15 and 30 pmol of XPA and XPG). Bar graph shows the quantification of a band corresponding to the first paused RNA product (below) and the quantification of bands corresponding to the elongation past the first pause site (above) for different concentrations of additional factors. The error bars show s.d. of the mean values for 2 experimental replicates.

From the previous assay it is unclear what the underlying mechanism of transcription inhibition by the TFIIH-XPA-XPG complex is. The complex could arrest RNAP progression, dissociate the RNAP from the DNA or push the RNAP backwards (Fig. 30a). We designed a biochemical assay to differentiate between those three scenarios (Fig. 30a,b). We performed the *in vitro* transcription assay using biotinylated DNA scaffold containing the arrest sequence in the presence of core TFIIH, XPA and XPG. The RNAP inhibition was efficiently reproduced because we observed a large amount of short RNA products and almost no run-off transcripts (Fig. 30c, lane 1). We purified the inhibited elongation complexes by coupling the DNA to streptavidin beads and extensively washing the beads to remove NTPs and NER factors. We briefly incubated the inhibited elongation complexes with TFIIS. TFIIS induces the intrinsic RNAP endonuclease activity<sup>156-159</sup> and can be used to reveal the nature of the RNAP inhibition (Fig. 30a). If the inhibited complexes are arrested

or dislodged from the DNA, RNA cannot be cleaved by RNAP because the active site of the RNAP accommodates the RNA 3' end or no RNA, respectively. However, if the inhibited elongation complexes are backtracked, RNAP will cleave the RNA and produce shorter RNA products (Fig. 30a). Indeed, after incubation with TFIIS, we observed RNA shortening (Fig. 30c, lane 2) suggesting that RNAP was in a backtracked state. The backtracking was substantial because the cleaved RNA was 3 nucleotides shorter than the RNA primer used for the elongation complex assembly. Also, as the initial RNA primer produced a 10 base pair DNA-RNA hybrid, the backtracking proceeded until the minimal length of a stable DNA-RNA hybrid of 7 base pairs<sup>248,249</sup> was reached which could limit the extent of backtracking in the experimental system.

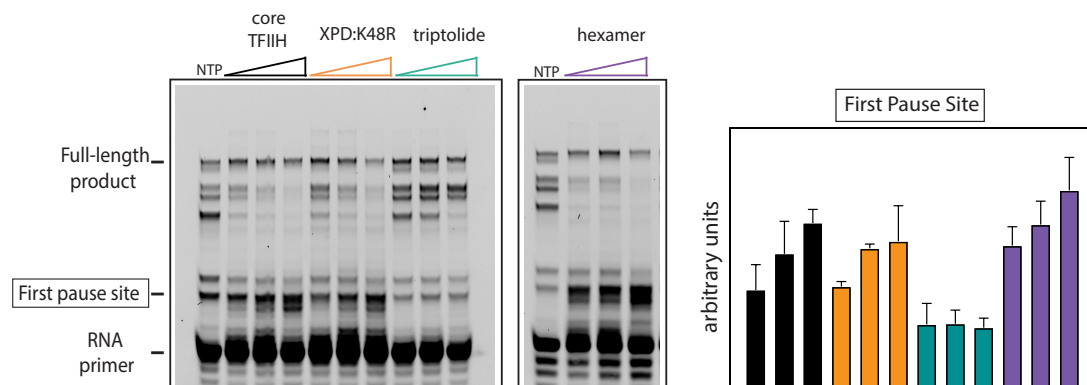


**Figure 30 | Core TFIH induces RNAP backtracking in the presence of XPG and XPA.**

(a) Possible mechanisms of RNAP inhibition by the core TFIH-XPA-XPG complex. RNAP could be arrested, dissociated from the DNA or backtracked (pushed backwards). During backtracking the RNA 3' end is threaded through a pore in RNAP but the RNA-DNA hybrid is still held by the RNAP active site<sup>155</sup>. Thus, only the backtracked RNAP is sensitive to the addition of TFIIS which induces the RNAP endonuclease activity<sup>156-159</sup> and shortens the bound RNA. (b) Schematic of the experimental setup. (c) RNAP was preincubated with an excess of core TFIH, XPA and XPG, followed by the addition of NTPs. After the immobilization of elongation complexes on streptavidin beads and extensive washing, we observed a strong accumulation of a short RNA product corresponding to the first pause site, indicating efficient transcription inhibition (lane 1). The addition of TFIIS induced the cleavage of RNA, suggesting that the RNAP was in the backtracked state (lane 2). When the reaction was again supplemented with NTPs, the cleaved RNA was extended (lane 3). (d) In vitro transcription reaction performed in the presence of the increasing amounts of core TFIH, XPA and XPG (as in Fig. 29d), and constant amounts of TFIIS (0.5 pmol). The expected transcription inhibition by the core TFIH-XPA-XPG complex was not observed in the presence of TFIIS probably because TFIIS reactivated the backtracked elongation complexes which resumed transcription in the presence of NTPs.

To determine if the backtracked RNAPs are still transcription-competent, we supplemented the TFIIS treated complexes with NTPs. We observed the elongation of cleaved RNA upon the addition of NTPs (Fig. 30c, lane 3) suggesting that the backtracked RNAP can be reactivated by TFIIS to continue transcription. To further demonstrate that active RNAP backtracking mediates transcription inhibition, we performed the transcription assay in the presence of backtracking factors (core TFIIH, XPA and XPG) and TFIIS. In the presence of TFIIS alone we observed the accumulation of the full-length RNA and the disappearance of bands corresponding to aborted transcription, consistent with the role of TFIIS in transcription elongation<sup>241</sup> (Fig. 30d). Moreover, after the addition of increasing amounts of backtracking factors, the transcription was no longer inhibited (Fig. 30d). Thus, in the presence of TFIIS and NTPs the backtracked complexes are rescued and can continue transcription until the run-off RNA products are made.

We showed that core TFIIH actively backtracks stalled elongation complexes in the presence of XPA and XPG (Fig. 30c), however, it is unclear which of the two core TFIIH ATPases pushes RNAP backwards. To differentiate between the activity of the XPD helicase and the XPB translocase we performed the *in vitro* transcription assay with the core TFIIH containing the ATPase deficient XPD point mutant<sup>73</sup> or in the presence of the XPB inhibitor triptolide<sup>200,201</sup> (Fig. 31). The loss of XPD helicase activity did not affect the transcription inhibition but the XPB inactivation abolished the ability of TFIIH to backtrack RNAP. To test if the XPD presence is required for the TFIIH backtracking function, we tested the core TFIIH hexamer which lacks the XPD subunit in the transcription assay (Fig. 31). Transcription inhibition was not affected by the absence of XPD further confirming that XPB is the TFIIH subunit responsible for RNAP backtracking.



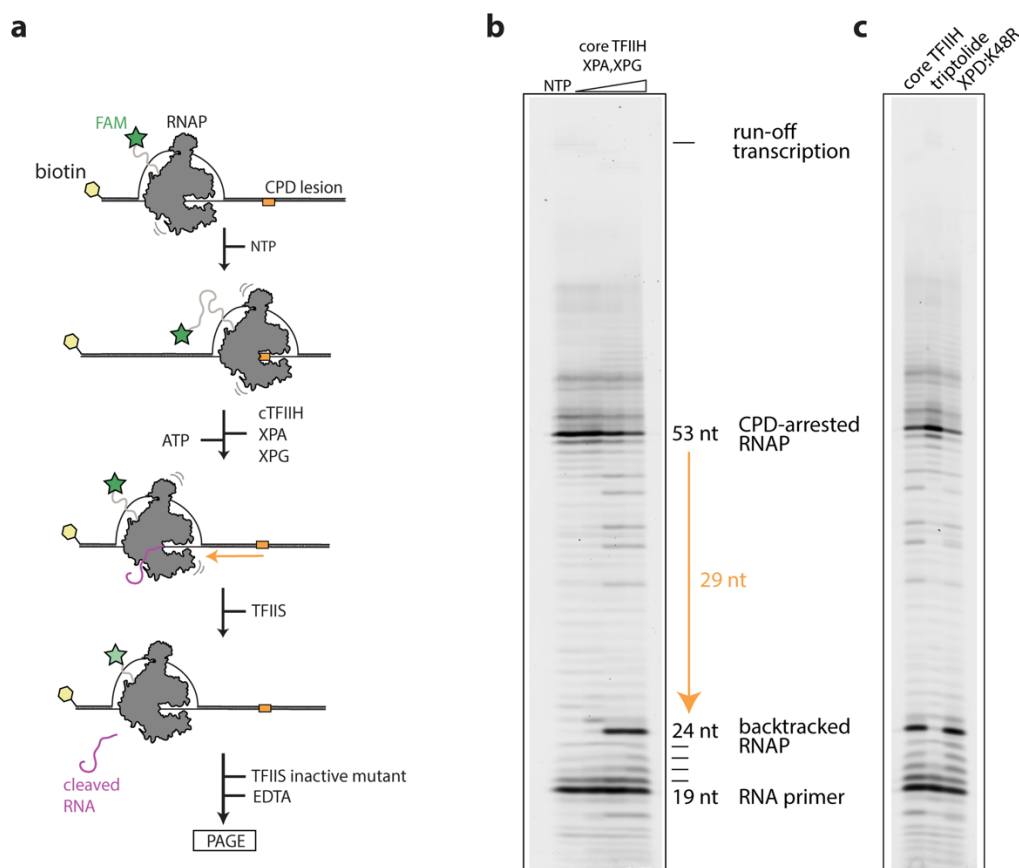
**Figure 31 | XPB translocase pushes RNAP backwards.**

*In vitro* transcription reaction was performed in the presence of increasing concentrations of core TFIIH, XPA and XPG as in Fig. 29d (black triangle) which inhibited transcription as observed by the RNAP arrest at the first pause site. The same reaction was performed with the core TFIIH containing XPD:K48R inactive mutant<sup>73</sup> (yellow triangle) but no difference to the wild type core was observed. However, in the presence of the drug triptolide which inhibits XPB<sup>200,201</sup>, we did not observe the transcription inhibition (green triangle). When the reaction was performed with the hexameric TFIIH assembly which lacks the XPD subunit, the transcription inhibition was again observed (purple triangle), suggesting that XPB is necessary and sufficient for the RNAP backtracking. Bar graph shows the quantification of the RNA corresponding to the first pause site and the bars are color-coded as the triangles above the gels. Error bars represent the s.d. of the mean values for two experimental replicates.

### 2.3.3 Active RNAP backtracking exposes DNA lesions to repair

We have shown that the core TFIIH can push the stalled RNAP backwards in the presence of XPA and XPG (Fig. 30c). However, it is unknown if the backtracking is efficient enough to expose DNA lesions to repair. RNAP arrested by a CPD sequesters ~18 nucleotides of the DNA downstream of the lesion<sup>51,233</sup>. In addition, XPG cleaves the DNA 5 nucleotides 3' of the lesion during NER<sup>107,111</sup>. Thus, RNAP has to be pushed backwards for more than 23 nucleotides to fully expose the lesion to repair enzymes.

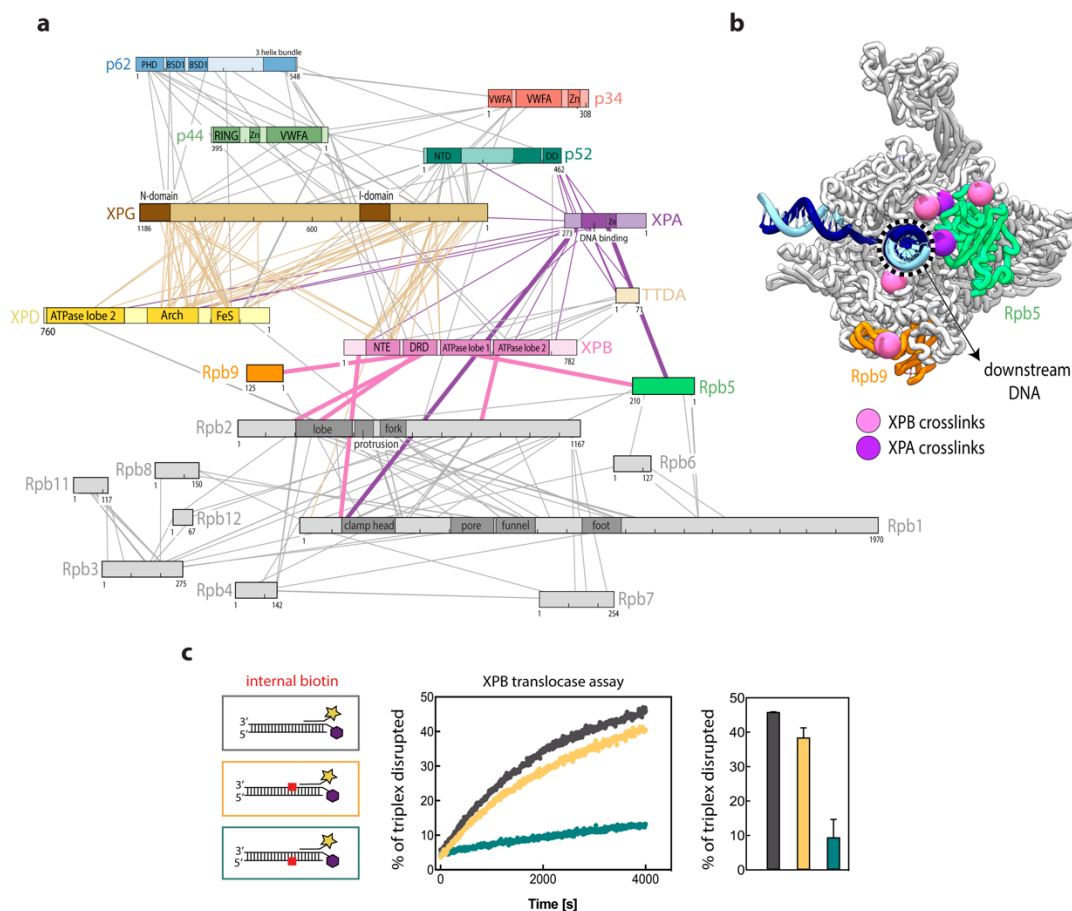
We designed a biochemical assay which allowed us to determine the RNAP position on the DNA after backtracking at single-nucleotide precision (Fig 32a). We assembled the elongation complex using a CPD-containing DNA scaffold and a fluorescently labeled RNA. Upon the addition of NTPs, the RNAP transcribed into the CPD lesion, as we observed the appearance of the 53 nucleotides long RNA products (Fig 32b).



**Figure 32 | Active RNAP backtracking by XPB exposes CPD lesions to repair.**

(a) Schematic representation of the experimental design used for measuring the extent of RNAP backtracking at a single nucleotide resolution. (b) In vitro transcription reaction (2.8 pmol RNAP) on a CPD containing DNA scaffold produced 53 nucleotides long RNA products, as in Fig.25. After the incubation with increasing amounts of the core TFIIH, XPA and XPG (2.2, 4.4 and 8.8 pmol) and RNA cleavage induction by TFIIIS (0.5 pmol) we observed the 24 nucleotides long cleavage products, suggesting that RNAP was backtracked for 29 nucleotides. (c) Backtracking was also observed when core TFIIH containing XPD inactive mutant<sup>73</sup> was used, but not in the presence of triptolide which inhibits XPB<sup>200,201</sup>. Thus, XPB is responsible for pushing the RNAP backwards.

In the next step, we added the backtracking machinery composed of the core TFIIH, XPA and XPG to the lesion-arrested RNAP and incubated the reaction with ATP to enable backtracking. Finally, after the RNAP endonuclease activity was triggered by TFIIIS, we observed the appearance of 24 nucleotides long RNA cleavage products. Since the RNA cleavage is mediated by RNAP, we can conclude that the RNAP active site was pushed backwards for 29 nucleotides with increasing amount of the backtracking machinery (Fig. 32b). Thus, the backtracking is sufficient for the complete removal of RNAP from the lesion site. The backtracking was also observed when the XPD inactive mutant was used, but the effect was abolished when XPB translocase was inhibited by triptolide (Fig 32c).



**Figure 33 | Crosslinking mass-spectrometry network of the RNAP-core TFIIH–XPA–XPG complex and the effect of DNA lesions on XPB translocase activity.**

(a) Core TFIIH subunits are color-coded as in Fig. 14, XPG is in brown, XPA in purple and RNAP subunits in gray, except Rpb5 is in green and Rpb9 in orange. XPA and XPG crosslinks to the core TFIIH are shown with purple and brown lines, respectively. Crosslinks between XPA and RNAP are shown with thick purple lines and between XPB and RNAP with thick pink lines. The list of inter-subunit crosslinks is provided in the Supplemental Table 4. Domain abbreviations as in Fig. 9. (b) Selected crosslink sites for XPA (purple spheres) and XPB (pink spheres) were mapped onto the RNAP structure (PDB code: 5FLM)<sup>123</sup> and cluster around the downstream DNA (dashed black circle). RNAP subunits are shown in gray, except Rpb5 is in green and Rpb9 in orange, DNA is in blue. (c) XPB translocase assay performed as in Fig. 11c. (left) Schematic representation of scaffolds used to check if a lesion mimic (internal biotin) placed in the 5′-3′ or the 3′-5′ DNA strand can inhibit the XPB translocase activity. Position of the internal biotin is indicated with red squares. (middle) Representative fluorescence traces of the core TFIIH translocation on corresponding scaffolds (the color of the trace matches the color of the frame around the corresponding scaffold). XPB cannot translocate efficiently when biotin is installed in the 5′-3′ DNA strand and the biotin located in the 3′-5′ DNA strand has no effect on XPB translocation. (right) Bar graph shows the percentage of triplex disrupted after 4000 s for 2 independent replicates. Error bars represent s.d. from the mean values.

To gain insight into the architecture of the RNAP backtracking complex we obtained a protein interaction network within the purified RNAP-core TFIIH-XPA-XPG complex by chemical crosslinking and mass-spectrometry (Fig. 33a). The complex was assembled by a large-scale *in vitro* transcription and backtracking reaction on a CPD containing DNA scaffold, followed by the removal of excess NER factors and NTPs via size-exclusion chromatography (Methods). Many crosslinks were detected within the core TFIIH–XPA–XPG complex and between the RNAP subunits, however, crosslinks between the repair proteins and RNAP were rather scarce. It may be that the interface between RNAP and the backtracking machinery is quite small and largely mediated by DNA. Also, the complex with RNAP might be unstable and accompanied by a high rate of TFIIH dissociation, which could lead to more inefficient crosslinking. XPA crosslinks to the core TFIIH ATPases, TTDA and the N-terminus of p52 subunit, which is consistent with the structure we presented here (Fig. 14). XPG crosslinks to XPB and XPD, which is similar to other crosslinking data we acquired (Fig. 23). Interestingly, the XPB and XPA crosslink sites map to Rpb1, Rpb2, Rpb5 and Rpb9 RNAP subunits, mostly around the downstream DNA (Fig. 33b). This suggests that XPB binds the downstream DNA during TCR and pushes the lesion-arrested RNAP backwards by translocating on the DNA duplex towards RNAP.

Since we observed that the backtracking of lesion arrested RNAP proceeds for exactly 29 nucleotides (Fig. 32b), it is tempting to speculate that XPB pushes RNAP backwards until it encounters the lesion. This would suggest that a lesion in the 5'-3' DNA strand inhibits the XPB translocase activity while XPB migrates on the downstream DNA towards RNAP. To test this, we performed a double-strand DNA translocase assay using DNA templates which contain an internal biotin as a DNA lesion mimic placed in the 5'-3' or the 3'-5' DNA strand (Fig. 33c). Indeed, we observed a strong inhibition of XPB translocase activity when the lesion mimic was located in the 5'-3' DNA strand but not when it was located in the 3'-5' DNA strand, as also shown for the yeast XPB homologue Ssl2<sup>69</sup>. Thus, during TCR, the XPB translocase activity might be employed for the removal of RNAP from the lesion site and for the subsequent verification of the lesion in the template strand. After the lesion verification, XPB may dissociate from the DNA and provide the space for lesion processing by the canonical NER pathway.

## 3. DISCUSSION AND OUTLOOK

### 3.1 First structure of an NER intermediate

A mechanistic framework for the NER pathway was established through decades of fruitful biochemical, functional and genetic research, which identified the major protein players and defined the main steps of the NER reaction pathway (Fig. 2)<sup>20,106</sup>. However, despite the large amount of existing data, we do not know how NER works on a molecular level. It is unknown, for instance, how TFIIH opens the DNA repair bubble or how XPA facilitates the kinase module dissociation from the TFIIH core. Even more fundamental features of NER, such as how TFIIH binds DNA or where XPA is located in the NER assembly, are a matter of debate in the current literature<sup>20,83,106</sup>. The lack of structural insight became an obvious bottleneck in NER research which could not be surpassed due to difficulties in the preparation of NER complexes for structural analysis. To overcome this barrier, we have established and optimized the production of recombinant human NER factors and performed a systematic biochemical analysis of the central NER component TFIIH. This allowed us to trap and, for the first time, solve the structure of an NER intermediate.

We focused on a lesion scanning step of the NER pathway which is mediated by the migration of the TFIIH core on DNA while searching for a lesion. Our biochemical characterization identified two additional NER factors XPA and XPG to be important for this process by stimulating the XPB translocase and the XPD helicase activity (Fig. 12). We solved the cryo-EM structure of the core TFIIH-XPA complex bound to the edge of a repair bubble at 3.6Å (Fig. 14), which represents the minimal lesion scanning NER intermediate, and mapped the position of XPG by crosslinking mass-spectrometry (Fig. 23). The newly gained structural and functional knowledge greatly extends our understanding of the NER pathway (Fig. 34), as discussed in the following sections. We improved the existing structure of the TFIIH core (section 3.2) and explained human disease mutations (section 3.3). The structure elucidates the long-debated architecture of XPB, XPD and XPA on DNA and explains how XPA anchors the NER machinery to DNA, how the kinase module inhibits TFIIH helicase activity and how XPA facilitates kinase module dissociation and commits TFIIH to repair (sections 3.4 and 3.6). Our results redefine the previously assigned roles for XPA and open new questions for the role of XPB (section 3.5). Finally, our protocols for the purification of NER components and the preparation of NER complexes for cryo-EM can now be applied, with only minor adjustments, for the characterization of other NER intermediates which will hopefully lead to a complete molecular mechanism of the NER pathway (section 3.8).

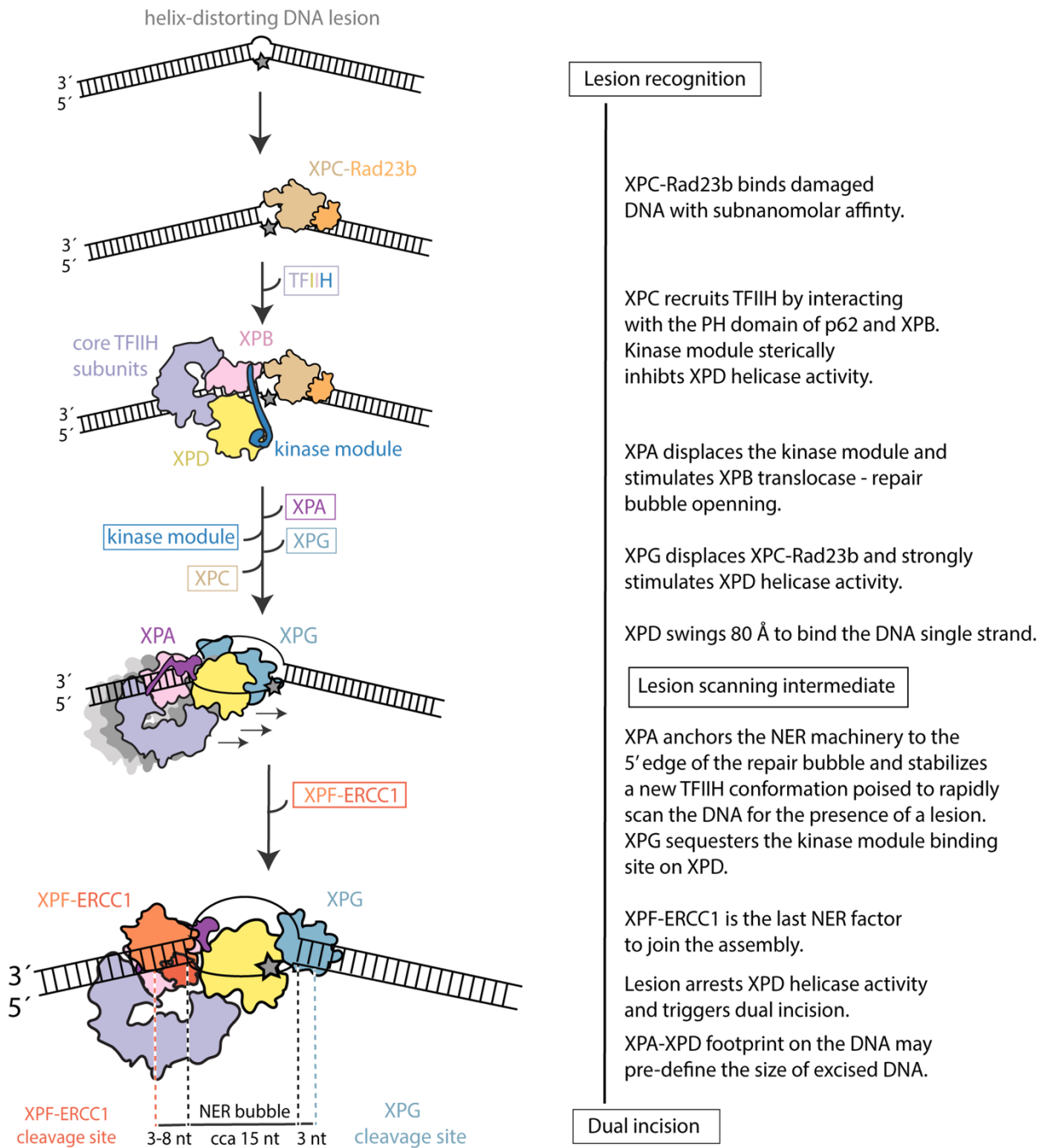


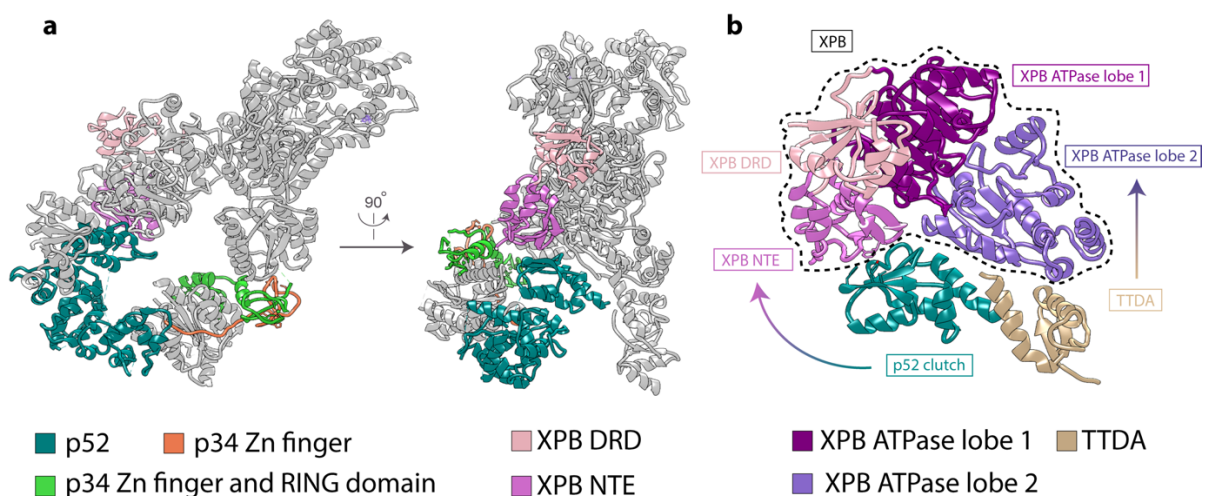
Figure 34 | NER model which incorporates novel insights obtained in this study.



### 3.2 Extended model of the human TFIIH core and comparison to other TFIIH structures

A breakthrough in TFIIH structural biology occurred a year ago, when cryo-EM structures of human TFIIH<sup>182</sup> and yeast TFIIH bound to the pre-initiation complex (PIC)<sup>70</sup> were reported. However, due to resolution limitations (4.4Å for the human and ~6Å for the yeast TFIIH), most parts of the structures were created by homology modelling and some parts of TFIIH could not be built at all. We solved the core TFIIH structure at higher resolution (3.6 Å) which allowed us to correct sequence registers for the previous human TFIIH model<sup>182</sup> and to additionally build most of the p52 subunit, the DRD (damage recognition domain) and the NTE (N-terminal extension) of XPB and the p34-p44 interface comprised of two zinc fingers and a ring domain (Fig. 35a).

These expansions reveal how XPB is recruited to the TFIIH core. The main XPB-core TFIIH interaction surface is formed between the NTE of XPB and the clutch domain in p52 (Fig. 35b). Due to high resolution of our structure in this region, we were able to build both domains *de novo*. Interestingly, the two domains are strikingly similar in structure and consist of four  $\alpha$ -helices and a  $\beta$ -sheet which is formed by four antiparallel  $\beta$ -strands (Fig. 35b). The two domains are oriented sheet-to-sheet and form an extensive hydrophobic interface. The interface is functionally important because it was shown that structure-destabilizing mutations in the NTE of XPB weaken the XPB-p52 interaction and impair TFIIH repair activity<sup>68</sup>. In addition, p52 stimulates XPB to hydrolyze ATP and this activity is dependent on the presence of the clutch domain<sup>68</sup>. Since we did not observe any p52 density extending towards the active site of XPB, the interaction with p52 likely stabilizes a more active XPB conformation.

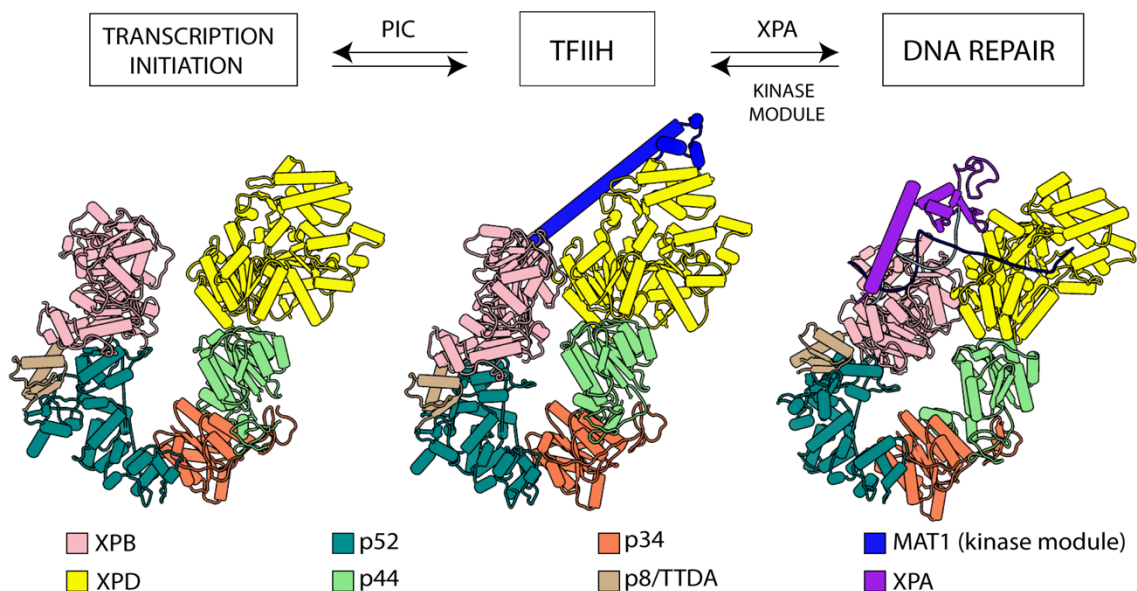


**Figure 35 | Extension of the previous model for the human TFIIH core.**

(a) Parts of the TFIIH core present in the previous TFIIH model<sup>182</sup> are shown in gray and structural elements additionally built or modeled in our study are colored as indicated. (b) XPB is anchored to the TFIIH core via interactions with p52 and TTDA subunits. XPB NTE is shown in pink, DRD in light pink, ATPase lobe 1 in purple, ATPase lobe 2 in medium purple, p52 clutch in dark cyan and TTDA in light brown. XPB is outlined by the black dashed line. Bulky hydrophobic side chains at the XPB-p52 interface are shown as sticks.

The second XPB attachment point to the TFIIH core is the interaction between the ATPase lobe 2 in XPB and the smallest TFIIH subunit TTDA (Fig. 35b). TTDA is essential for DNA repair and mutations in this TFIIH subunit lead to severe disorders in humans<sup>250-252</sup>. Since TTDA has no enzymatic activity, mutations in this subunit most likely destabilize the core TFIIH-XPB interface which may result in a decreased enzymatic activity of XPB, as shown *in vitro*<sup>68,252</sup>.

Interestingly, the conformation of the TFIIH core observed in the context of DNA repair differs significantly from conformations of free TFIIH<sup>182</sup> or TFIIH engaged in transcription initiation<sup>70</sup> (Fig. 36). The core TFIIH seems to be very flexible and can adapt to different requirements for the complex in transcription and DNA repair<sup>64</sup>. Different TFIIH conformations are trapped by external factors which mostly bind the core ATPases XPD and XPB<sup>70,182</sup> (Fig. 36). During transcription initiation, the XPB and p62 subunits interact with TFIIIE directly, while XPD interacts with the RNAP stalk and TFIIIE via the kinase module<sup>70</sup>. These interactions force the TFIIH core to widen<sup>70</sup>, which breaks the contact between XPB and XPD and destabilizes the long MAT1 helix observed in free TFIIH<sup>182</sup> (Fig. 36). Aside from this core widening, conformations of the free and the PIC-bound TFIIH are very similar. In contrast, for DNA repair TFIIH has to dramatically rearrange (Fig. 17) and change its composition<sup>84</sup>. Both XPB and XPD have to reorient to simultaneously bind the edge of a repair bubble (Fig. 14) and this process is facilitated by XPA which wraps around the DNA and contacts both core ATPases (Fig. 36).



**Figure 36 | Different conformations of the TFIIH core in transcription and DNA repair.**

(left) The human TFIIH structure during transcription initiation was modeled by aligning human subunits with the corresponding subunits of the yeast TFIIH bound to PIC (PDB:5OQJ)<sup>70</sup>. (middle) Human TFIIH was modeled by fitting individual domain from our structure into the EM density of the kinase-bound TFIIH (EMDB:3802)<sup>182</sup>. TFIIH subunits and accessory factors are color-coded as indicated.

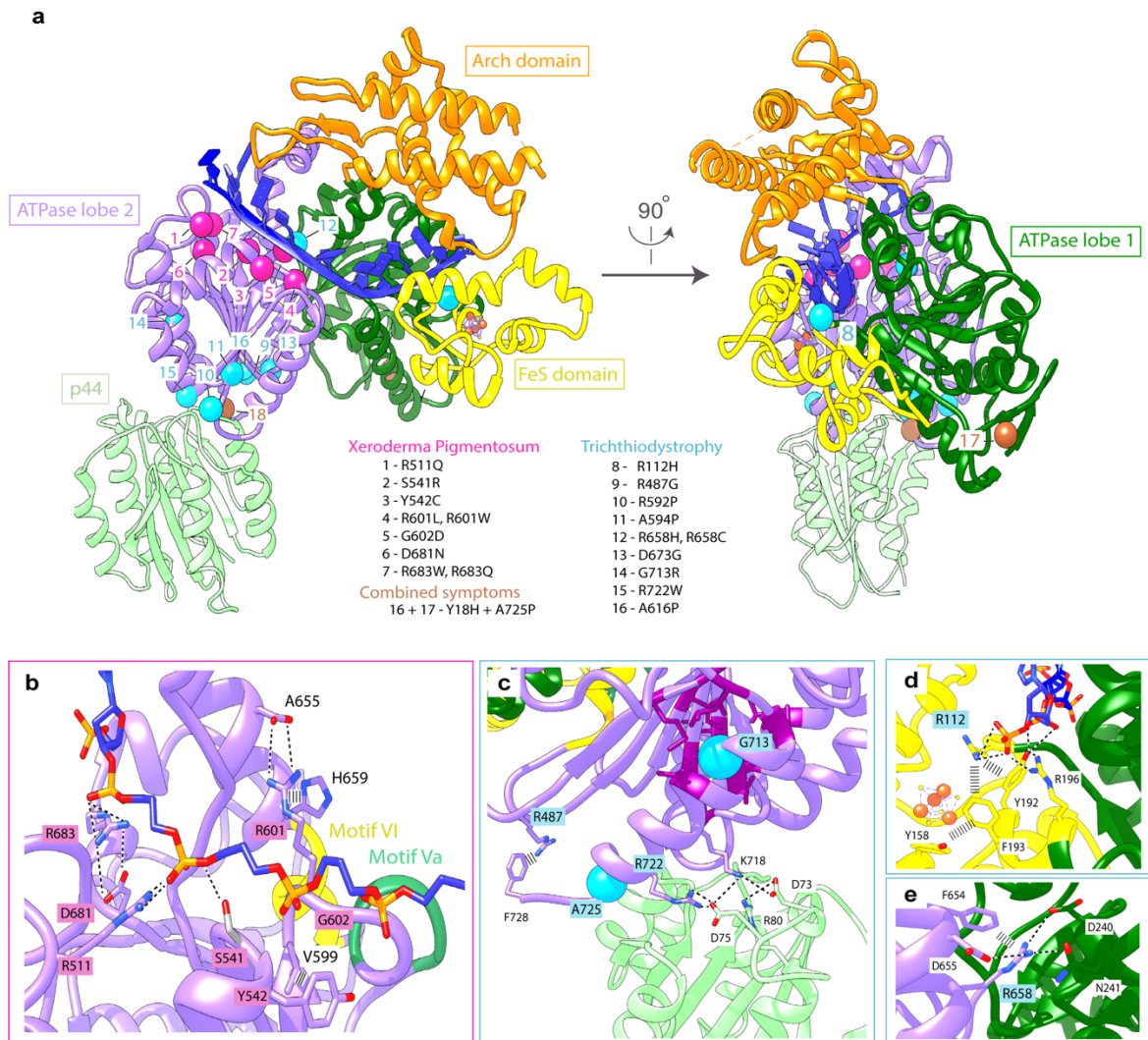
Due to this rearrangement, the kinase module dissociates from the TFIIH core<sup>84</sup>, which in turn de-represses XPD helicase activity and initiates lesion scanning<sup>56,73</sup>. The reorientation of XPB and XPD is enabled by substantial inter- and intra-subunit rearrangements in p34, p44 and p52 (Fig. 17). It may be that the modular architecture of the TFIIH complex emerged to support, restrict and guide the rearrangement of enzymatically active subunits for the essential, yet very different roles of TFIIH in transcription and DNA repair<sup>64</sup>.

### 3.3 Structural etiology of Xeroderma pigmentosum and Trichothiodystrophy

Mutations in the XPD helicase mainly cause two human diseases with markedly different symptoms: Xeroderma pigmentosum (XP) and Trichothiodystrophy (TTD)<sup>40,214</sup>, and in rare cases XP is additionally coupled with the Cockayne syndrome (CS)<sup>77</sup>. XP patients are extremely sensitive to the UV-irradiation and readily develop skin cancer, while TTD patients suffer from mental and growth retardation, accelerated aging and decreased fertility<sup>214</sup>. This can be explained by the dual role for XPD in DNA repair and transcription<sup>214</sup>. During DNA repair, XPD helicase activity is required for the repair bubble opening and lesion scanning<sup>65,73,224,253-255</sup>, while in transcription XPD has a structural role and couples TFIIH to the PIC<sup>70,198</sup>. Milder XP symptoms are likely caused by a defect in DNA repair and the more severe TTD symptoms by a defect in transcription<sup>214</sup>. Since we solved the core TFIIH structure at significantly higher resolution compared to previous models (Fig. 15)<sup>70,182</sup> and XPD helicase in our structure is bound to DNA (Figs. 16, 19), we can precisely map XP and TTD disease mutations and discuss the structural basis for disease development.

Interestingly, the XP mutations<sup>214,256,257</sup> cluster around DNA in the ATPase lobe 2 of XPD (Fig. 37a). Several of the affected residues (R683, R511 and S541) directly interact with the DNA phosphate backbone (Fig. 37b). It is worth noting that mutations in R683, which binds the third to the last nucleotide exiting XPD (Fig. 37b), account for 64% of all XP mutations in XPD and for 55% of mutations in TFIIH overall<sup>256</sup>. An additional two mutations, D681N and Y542C, are found directly adjacent to DNA contacting residues: D681 forms a salt bridge with R683 and Y542 is located next to S541. Y542 stabilizes the loop containing S541 by interacting with V599, which is not possible when the bulky Y543 residue is replaced with a smaller cysteine in XP (Fig. 37b). Overall, the direct disruption of XPD-DNA contacts by these mutations, as seen in our structure, would impair the XPD helicase activity and NER, as also shown *in vitro* for D234N, S541R and R683W mutations<sup>256</sup>. Interestingly, another residue mutated in XP, R601, is located next to DNA but the side chain of the residue points away from the DNA backbone (Fig. 37b). However, R601 is placed next to the helicase motif Va<sup>133</sup> and the long side chain of R601 forms a salt bridge with D655 and a cationic pair with H659, both of which are located in a helix carrying

the helicase motif VI<sup>133</sup>. Thus, substitution of R601 with a leucine or a much bulkier tryptophan in XP would disrupt these interactions and misalign helicase motifs, leading to XPD helicase impairment and XP phenotype. G602 is a direct neighbor of R601 and the substitution to a less flexible aspartate found in XP patients would probably have similar disruptive effect as the R601 mutation (Fig. 37b).



**Figure 37 | DNA-bound TFIIH core at 3.6Å reveals structural etiology of Xeroderma pigmentosum (XP) and Trichothiodystrophy (TTD).**

(a) XP and TTD mutations<sup>214,256,257</sup> mapped onto XPD-DNA structure are shown as pink and blue spheres, respectively. XP mutations cluster around DNA in the ATPase lobe 2 of XPD and TTD mutations cluster at the XPD-p44 interface. A XP-TTD overlap patient carries 2 mutations<sup>258</sup>, shown as brown spheres. (b) Zoom-in on XP mutations in the direct vicinity of DNA. Residues mutated in XP are highlighted in pink and their interaction partners in white. Helicase motifs VI and Va are shown in yellow and green, respectively. (c) XPD-p44 interface. A single helix in XPD is affected by four different TTD mutations. Mutated residues are highlighted in blue. G713 and A725 are depicted as blue spheres. The helix carrying TTD mutations interacts with p44 via two salt bridges, R722 (XPD) - D75 (p44), and K718 (XPD) - D73 (p44). (d) R112 residue frequently mutated in TTD interacts with the DNA backbone and residues surrounding the iron sulfur cluster. (e) R658 residue mutated in TTD engages in the interaction network between the ATPase lobes.

In addition to affecting the structural role of R601, the G602D mutation also introduces a negative charge into the DNA-binding groove of XPD, which may be particularly disruptive for the XPD structure and function, and could explain why this mutation also leads to severe CS symptoms<sup>214</sup>.

In contrast to XP, all but one TTD mutations<sup>214</sup> map to regions of our structure further away from the DNA trajectory (Fig. 37a). Most of the mutations are located at the XPD-p44 interface. This interface is the main attachment point of XPD to the rest of the TFIIH core<sup>259,260</sup>, but it is surprisingly small and could be particularly vulnerable to destabilizing mutations. Since XPD plays an important structural role during transcription initiation<sup>70</sup>, a disruption of the XPD-p44 interface and the dissociation of XPD may cause a transcription defect and a TTD disease phenotype<sup>256</sup>. Interestingly, a single XPD helix (residues 711-725) is affected by four different TTD mutations (G713R, R722W, A725P and R487G) (Fig. 37c). This helix is an important contact point between XPD and p44 because R722 and K718 in the helix form salt bridges with D75 and D73 located in a p44 loop (residues 68-76). The R722W mutation directly disrupts one of the two salt bridges. The G713R mutation likely destabilizes the helix because G713 is located in a hydrophobic pocket where a long arginine side chain cannot be accommodated nor neutralized (Fig. 37c). The R487G mutation can also be explained based on our structure. Even though R487 is distant from the helix in the primary structure, it engages in cation- $\pi$  interaction with F728 located two residues away from the helix (Fig. 37c). This R487-F738 interaction might keep the helix stacked against XPD and disruption of this interaction would likely displace the helix and weaken the XPD-p44 interface. An A725P substitution might have a disruptive effect on the helix because prolines have poor helix-forming propensities, and this substitution may also affect the neighboring R487-F738 interaction (Fig. 37c). Several other destabilizing substitutions to proline residues (R616P, A594P and A592P) are located at the XPD-p44 interface (Fig. 37a).

The R658H and R658C TTD mutations do not map to the XPD-p44 interface but to the site between the ATPase lobes of XPD (Fig. 37e). R658 participates in the extensive interaction network of polar residues between the lobes, which includes salt bridges with D240 located in the opposite lobe and D655 located in the same lobe. Thus, a mutation in R658 might destabilize the structure of XPD by disrupting the contacts between the lobes which may perturb the scaffolding role of XPD in transcription initiation and lead to TTD. R112H is the only TTD mutation in the vicinity of DNA. It was shown that TFIIH carrying the R112H mutation is helicase dead and completely inactive in NER<sup>256,261</sup>, and patient-derived cell lines are extremely UV-sensitive, which is a typical feature of XP. The helicase impairment of XPD by a R112H mutation is supported by our structure which shows that R112 directly binds the DNA phosphate backbone (Fig 37d). However, the same mutation causes a mild defect in basal transcription *in vitro* and mild TTD symptoms in patients<sup>256,261</sup>. Since R112 is located in front of the iron-sulfur cluster, the long side chain of the residue may shield the cluster from the reactive species in the solvent. In addition, R112 interacts with F193 and Y192 and contributes to the structural stability of the FeS domain (Fig. 37d).

It was indeed shown in yeast that this mutation results in the loss of the iron-sulfur cluster<sup>262</sup>, which may destabilize the entire protein<sup>76</sup>. Structural destabilization of XPD would explain the effect of this mutation on transcription and the TTD phenotype in patients.

An interesting case of a patient with a hybrid XP/TTD phenotype was reported relatively recently<sup>258</sup>. Aside from the typical TTD clinical features such as brittle hair, the patient also developed skin cancer, which is specific to XP<sup>258</sup>. The sequence analysis of the XPD gene revealed two different mutations: A725P and Y18H (Fig. 37a). We already described how the A725P mutation leads to destabilization of the XPD-p44 interface and a TTD phenotype. However, the side chain of the Y18 residue is located only ~ 4Å away from the Q-motif/Q21 which was predicted to bind ATP<sup>133</sup>. The Y18H mutation might thus affect the ATP hydrolysis and the XPD helicase activity, which could explain the XP phenotype.

In summary, we could map most of the XP and TTD mutations onto our XPD-DNA structure and we described the molecular interactions surrounding the affected residues. We provided the structural basis for how XP and TTD mutations impair the XPD structure and function, and TFIIH integrity, thus leading to human disease development.

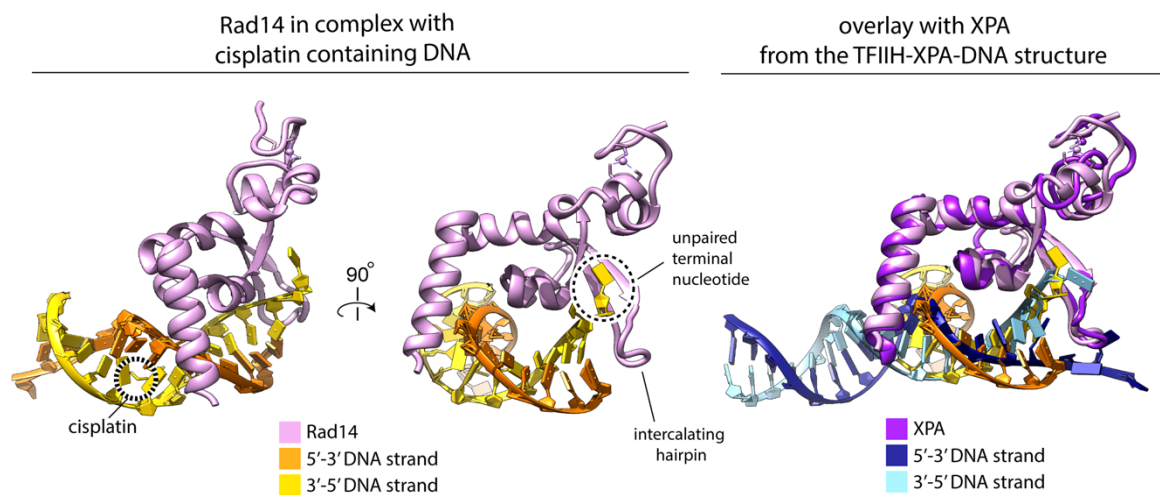
### 3.4 Role of XPA in NER and activation of TFIIH for repair

XPA is an essential NER factor comprised of an N-terminal zinc finger and a C-terminal DNA-binding domain<sup>83</sup>. The DNA-binding activity of XPA has been extensively studied over the past decades<sup>83</sup>. It was initially shown that XPA binds damaged DNA duplexes with a higher affinity than non-damaged DNA<sup>93,263-267</sup>, which led to the conclusion that XPA may be involved in the initial lesion recognition. A crystal structure of the yeast XPA homologue Rad14 bound to a damaged DNA duplex<sup>100</sup> further suggested that XPA binds DNA lesions as a dimer and recognizes the lesion-induced DNA bending rather than the lesion itself. However, the role for XPA in the initial DNA recognition was challenged by later observations: (i) XPC binds the damaged DNA with a much higher affinity compared to XPA<sup>26,46,62,268</sup> which makes XPC a likelier candidate for the initial lesion sensing, and (ii) XPA joins the NER assembly only after TFIIH is recruited and the DNA opening takes place<sup>61,63,87</sup>. Moreover, it was shown that XPA strongly binds duplex-single strand junctions<sup>82</sup> and kinked DNA structures<sup>81</sup>. This large amount of somewhat conflicting data is difficult to reconcile because no structural information on XPA in the context of NER is available up to date. Thus, it is unknown how XPA interacts with the DNA repair bubble and how this interaction is influenced by other NER factors, such as TFIIH.

Our cryo-EM structure of TFIIH-XPA bound to a bifurcated DNA which mimics one half of a repair bubble (Fig. 14), helps to explain the above mentioned observations. The structure shows that XPA interacts with several TFIIH subunits and wraps around the duplex-single strand junction formed between the XPB translocase which binds the DNA duplex and the XPD helicase which binds the single-stranded DNA overhang (Fig. 14).

Thus, DNA-bound TFIIH is the actual XPA substrate during NER. This explains the high binding affinity of XPA for duplex-single strand junctions<sup>82</sup>, as well as why the stable XPA integration into the NER complex occurs only after TFIIH initiates the repair bubble opening<sup>63</sup>. Assuming that a DNA lesion is located in the portion of DNA scanned by the XPD helicase, as suggested by previous biochemical work<sup>56,73</sup>, our structure also reveals that XPA demarcates the 5' edge of the repair bubble. Furthermore, it was shown that XPA recruits the XPF-ERCC1 endonuclease to the growing NER assembly<sup>63,96</sup>. This is in good agreement with the position of XPA at the 5' edge of the bubble because XPF-ERCC1 is known to cleave DNA 5' of the lesion, only 3-8 nucleotides away from the edge of the repair bubble<sup>109,110</sup>.

Interestingly, the DNA binding mode of XPA during NER closely resembles the binding mode observed in the crystal structure of the Rad14 dimer (yeast XPA homologue) bound to the damaged DNA duplex<sup>100</sup> (Fig. 38). In our structure XPA inserts the intercalating  $\beta$ -hairpin at the duplex-single strand junction and uses W175 at the tip of the hairpin to stack against a base in the single-stranded overhang (Fig. 18a, b). Rad14 similarly uses F262 (which corresponds to W175 in XPA) located in the  $\beta$ -hairpin to stack against the last base pair of the DNA duplex. However, even though a fully complementary DNA duplex was used for crystallization, Rad14 unpairs the terminal bases with the hairpin which creates a minimal duplex-single strand junction<sup>100</sup> (Fig. 38).



**Figure 38 | Comparison between Rad14 bound to the damaged DNA duplex and XPA bound to the duplex-single strand junction.**

(left) The crystal structure of Rad14 (yeast XPA homologue) bound to the DNA duplex damaged with cisplatin (PDB-5A39)<sup>100</sup>. Rad14 is shown in pink, the 5'-3' DNA strand in orange and the 3'-5' DNA strand in yellow. The DNA damage (cisplatin) and the unpaired terminal nucleotide are indicated by dashed circles. Rad14 does not interact with the DNA damage but rather binds the edge of the DNA duplex and unpairs the terminal bases. (right) The XPA structure presented here is superimposed onto the Rad14 structure. XPA is shown in purple, the 5'-3' DNA strand in dark blue and the 3'-5' DNA strand in light blue. The DNA binding mode, including the position of the unpaired terminal nucleotide and the intercalating hairpin, between the two structures is strikingly similar. This suggests that the Rad14 structure bound to the damaged DNA reflects the tendency of the protein to bind duplex-single strand junctions, which may be the primary role for XPA during the DNA repair, rather than lesion recognition.

Thus, the crystal structure likely reflects XPA binding to the edge of a repair bubble rather than the lesion recognition by XPA. Also, in our structure a single copy of XPA makes specific contacts to several TFIIH subunits: the XPA hairpin interacts with XPD, the extended helix bridges XPB ATPase lobes and the flexible C-terminus extends towards TTDA and p52 dimerization domain (Fig. 14, 15d). Thus, it is very unlikely that XPA could be integrated into the NER complex as a dimer.

Aside from providing a topological information on XPA binding to the repair bubble, our structure also implies how XPA activates TFIIH for repair. A decade ago it was shown by pull-downs and Western blotting that XPA facilitates dissociation of the kinase module from the TFIIH core during the course of NER<sup>84</sup>. However, no mechanistic explanation could be provided for this observation. Our structure has revealed that XPA, together with the bound DNA, stabilizes an entirely novel TFIIH conformation which markedly differs from the conformation of the kinase-bound TFIIH (Fig. 17). Moreover, the XPA-induced re-orientation of XPB and XPD would break the contacts between the TFIIH core and the kinase module, thus resulting in kinase module dissociation<sup>84</sup> (Fig. 22). This marks a major checkpoint in NER because the kinase module dissociation de-represses XPD helicase activity<sup>73</sup> and initiates the lesion scanning step of the NER reaction pathway, thereby committing TFIIH to repair.

XPA binds kinked DNA structures, such as 3-way and 4-way junctions with a very high affinity<sup>81</sup>. Since injured nucleotides distort the DNA backbone, it was suggested that XPA senses DNA kinks during lesion scanning and verifies the presence of a lesion<sup>20</sup>. Residues involved in such a verification process were mapped to the DNA-binding domain of XPA<sup>81</sup>. However, our structure shows that XPA binds the 5' edge of the repair bubble, while the lesion recognition by XPD helicase presumably occurs at the 3' edge<sup>73,253-255</sup>. It is therefore very unlikely that XPA can reach the modified DNA backbone during lesion verification. In addition, it was never experimentally shown that XPA binds damaged DNA single-strands with a higher affinity compared to unmodified DNA, which also argues against the lesion verification role for XPA in NER.

In summary, based on our functional and structural data, and the existing literature, we can propose a model for how XPA functions in NER. XPA joins the NER assembly after TFIIH is recruited to the lesion site by XPC<sup>61,63,87</sup>. XPA anchors TFIIH to the DNA by trapping the DNA duplex between the ATPase lobes of XPB (Fig. 18). XPA strongly stimulates the XPB translocase activity (Fig. 12) required for repair bubble opening<sup>65,252</sup>. The bubble opening creates the optimal DNA binding substrate for XPA which inserts the  $\beta$ -hairpin at the 5' edge of the bubble (Fig. 18). By chaperoning TFIIH-DNA interactions and bridging the re-oriented core TFIIH ATPases (Fig. 14), XPA stabilizes a novel TFIIH conformation which is no longer compatible with kinase module binding (Fig. 22). The kinase module dissociation removes a plug element from the DNA-entry pore in XPD (Fig. 21) which enables XPD to bind the DNA single-strand and start unwinding the DNA while searching for a lesion (Fig. 20). XPA facilitates lesion scanning by stimulating the XPD



helicase activity<sup>56</sup> (Fig. 12), most likely by preventing TFIIH dissociation from the DNA. Upon lesion recognition by the XPD helicase<sup>65,73,224,253-255</sup>, XPA is ideally located at the 5' edge of a repair bubble to recruit XPF-ERCC1 endonuclease<sup>63,96</sup> and to complete the assembly of the dual-incision complex.

### 3.5 Role of XPB in NER and the problem of repair bubble opening

The TFIIH core contains two ATPases, XPB and XPD<sup>64</sup>, that belong to the SF2-type family of DNA helicases<sup>133</sup>. XPD has been extensively studied and its robust 5'-3' helicase activity was demonstrated many times<sup>71-73,78</sup>, including here (Figs. 11, 12). In contrast, the enzymatic activity of XPB is still a matter of debate. A 3'-5' helicase activity was shown for archaeal XPB in isolation<sup>67</sup> and the same, although much weaker activity was demonstrated for the human XPB within the TFIIH complex<sup>73</sup>. Recently it was shown that the yeast XPB homologue Ssl2 acts as a DNA translocase<sup>140</sup>. We observed that the XPB ATPase activity is stimulated by double-stranded DNA rather than DNA single-strands. Therefore, XPB most likely acts as a translocase (Fig. 11). We confirmed the XPB translocase activity in our FRET-based translocase assay, but we could not detect any significant 3'-5' helicase activity even at high core TFIIH concentrations (Fig. 11). Moreover, we showed that the XPB translocase activity is modulated by XPA, an additional NER factor, which strongly stimulates XPB translocation in vitro (Fig. 12). Our TFIIH-XPA-DNA structure provides the molecular basis for such stimulation: XPA interacts with both ATPase lobes of XPB and wraps around the DNA, thereby trapping DNA between XPA and XPB (Fig. 18). This may prevent XPB dissociation from the DNA and increase XPB processivity. In addition, XPA physically links the XPB ATPase lobes with TTDA and p52 which may stabilize the XPB-core TFIIH interface (Figs. 14, 15d, 23c, 35b) and stimulate the XPB enzymatic activity<sup>68</sup>.

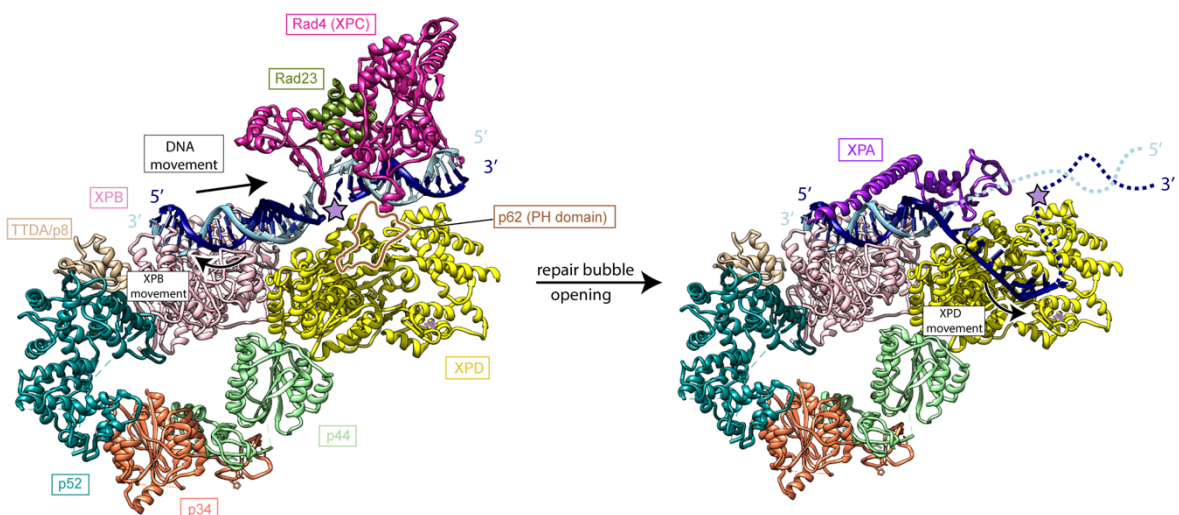
The role for XPB in NER is not fully understood. It is unlikely that XPB translocase activity plays a major role in lesion scanning, as it was shown that the XPD helicase activity drives this step of NER<sup>71-73,78</sup>. Also, as we monitored the XPB translocase and the XPD helicase activity in real time and in the same conditions, we observed that XPD is the dominant motor which migrates on DNA much faster and with higher efficiency compared to XPB (Fig. 12). Thus, migration of the NER complex on DNA during lesion scanning would be dictated by XPD unwinding in 5'-3' direction, as previously suggested<sup>56</sup>. However, XPB might be important for the processivity of lesion scanning because XPB and XPA form a positively charged DNA duplex tunnel (Fig. 18c) which could retain the NER machinery on DNA.

XPB might play an important role during the initial repair bubble opening. It was shown that the ATPase activity of XPB is essential for this process since ATPase deficient XPB mutants do not support the repair bubble opening in vitro<sup>65,252</sup>. Moreover, it was demonstrated that only the ATPase and not the helicase activity of XPB is required for

NER<sup>68</sup>. This led to the current model for the repair bubble opening in which XPB acts as an “unusual DNA helicase” which undergoes large scale conformational changes during ATP hydrolysis to wrap around DNA and induce DNA melting<sup>67,68</sup>. This model can be challenged by several observations made in our study. First, the conclusion that the helicase activity of XPB is not required for the bubble opening, therefore XPB works as an “unusual helicase”, is most likely invalid because XPB in general does not unwind the DNA (Figs. 11, 12). It may be that the helicase activity of XPB detected in some of the previous studies was a result of a small amount of contaminating proteins which co-purified with TFIIH. Second, XPB does translocate on DNA duplexes and this activity was not assayed in the context of the repair bubble opening. Third, we observed that XPA strongly stimulates XPB translocase activity in vitro. It was shown that XPA facilitates and stabilizes bubble opening<sup>252</sup>, further suggesting that XPB translocase activity might govern the process. Finally, the DNA wrapping function of XPB is based on a structure-function analysis of the archaeal XPB homologue<sup>67</sup>. Based on the biochemical observations it was suggested that XPB first binds destabilized DNA and subsequently closes its ATPase lobes via a large conformational change which could further melt the DNA<sup>67</sup>. However, comparison between free and DNA-bound XPB in the context of TFIIH (Fig. 17c) shows that XPB undergoes a very minimal conformational change while binding DNA, which involves a subtle closure of the ATPase lobes around the DNA and a rotation of the DRD and NTE domains. It could be that p52 and TTDA stabilize the closed XPB conformation (Fig. 35b) and prevent the large conformational change observed in the crystal structure of isolated XPB<sup>67</sup>.

Based on the data presented here, we can speculate that XPB uses its double-stranded DNA translocase activity (Fig. 11c) to open DNA in analogy to the transcription bubble opening during transcription initiation<sup>69,70,211</sup> (Fig. 39). XPC tightly binds 11 base pairs of the DNA duplex 3' of the lesion site<sup>26</sup> and recruits TFIIH 5' of the lesion<sup>56</sup> (Fig. 39). TFIIH is therefore correctly positioned for XPD to engage the damaged DNA strand after DNA opening and to unwind the DNA towards the lesion<sup>56</sup>. Moreover, our structure shows that XPD and XPB share the tracking strand (Fig. 16c) which suggests that XPB also tracks on the damaged DNA strand. Upon TFIIH recruitment to the lesion site by XPC, XPB may bind the DNA duplex 5' of the lesion (Fig. 39). Since XPC rotationally fixes the DNA duplex 3' of the lesion<sup>26</sup> and physically interacts with TFIIH<sup>151,194</sup> (Fig. 9), XPB would need to translocate in the 3'-5' direction on the damaged DNA strand to pump the DNA between XPC and XPB, which would lead to underwinding of DNA and DNA opening (Fig. 39). This is very similar to the proposed model for DNA opening by TFIIH in transcription<sup>70</sup>, only the position of XPC is replaced by the PIC. It might be that the DNA opening function was first employed by TFIIH to facilitate DNA repair, but was later adopted by the transcription machinery as an additional gene expression regulatory mechanism, since not all promoters depend on TFIIH for activation<sup>269</sup>. However, the main uncertainty of the proposed model is the polarity of XPB translocation, which is very difficult to assay and is still a matter of debate<sup>69,70</sup>. Structural comparison to the Chd1 translocase and the NS3

helicase indeed suggests a 3'-5' polarity of XPB<sup>70</sup>. In contrast, biochemical analysis of yeast Ssl2 suggests a 5'-3' polarity because XPB translocation is inhibited when an obstacle is inserted in the 5'-3' DNA strand<sup>69</sup>. However, we solved the XPB-DNA structure at high enough resolution to visualize protein DNA contacts and we observed that XPB tightly interacts with both DNA strands of the duplex (Fig. 16d). Thus, it is uncertain which contact would be affected by the obstacle in DNA and if the results of the polarity assay were accurately interpreted. The problem of the repair bubble opening could be answered by solving XPB-DNA structures in different stages of the translocation cycle. This would provide a molecular movie of XPB migration on DNA and reveal the XPB polarity. Repair bubble opening could be further investigated *in vitro* by using single molecule fluorescence assays and magnetic tweezers to monitor the dynamics of NER proteins and the dynamics of the DNA scaffold, as previously shown for other biochemical systems<sup>211,270</sup>.



**Figure 39 | A possible role for XPB in DNA opening.**

(left) Possible architecture of the core-TFIIH-XPC-Rad23b complex after the TFIIH recruitment to the lesion site (similar model for this complex was proposed in a recent review<sup>199</sup>). The crystal structure of the yeast XPC-Rad23 homologue (Rad4-Rad23, PDB-2QSG)<sup>26</sup> and the core TFIIH structure solved in this study were modeled on the same DNA, which was possible because both structures were determined in the presence of DNA. TFIIH subunits are color coded as in Fig. 14, XPC (Rad4) is shown in violet red and Rad23 in olive green. The CPD lesion that was present in the DNA scaffold bound to XPC<sup>26</sup> is indicated by the purple star. TFIIH interacts with XPC via the PH domain in the p62 subunit<sup>151,194</sup> (Fig. 9), which is indicated in brown. The position of the PH domain was modeled based on the TFIIH-PIC structure<sup>70</sup>. XPC binds the DNA 3' of the lesion site<sup>26</sup> and recruits TFIIH 5' of the lesion site. In this model, XPB translocates on the damaged DNA strand in 3'-5' direction (away from XPC). However, since TFIIH interacts with XPC, and XPC is anchored to DNA, XPB does not move forward but rather pushes the DNA duplex towards XPC. As a result, a DNA underwinding occurs between XPB and XPC which is resolved by DNA melting and the repair bubble formation. Interestingly, in this model there is enough space between XPB and XPC to accommodate XPA. XPA strongly stimulates the XPB translocase activity and might facilitate the repair bubble opening (Fig. 12d). (right) Upon DNA melting, XPD is ideally positioned to engage the damaged DNA strand and to start unwinding the DNA duplex in the 5'-3' direction towards the lesion. Since XPD is the dominant TFIIH motor (Fig. 12), the NER complex will migrate in the 5'-3' direction independently of the XPB translocase activity.

### 3.6 Positive and negative regulation of XPD helicase activity

XPD is the only TFIIH subunit which undergoes both negative and positive regulation<sup>56,73</sup>. During the transcription initiation, XPD helicase activity is suppressed by the kinase module<sup>56,73</sup>. We have shown that this effect is mediated by the kinase module binding since kinase deficient variants of the module suppress the XPD helicase activity equally well as the kinase active variant (Fig. 20). Comparison between XPD structures in the repressed<sup>182</sup> and the activated state (Fig. 21) reveals how kinase module binding suppresses XPD helicase activity. The kinase module subunit MAT1 binds the XPD Arch domain<sup>182</sup> and stabilizes a structural element we named the plug in the DNA-entry pore of XPD. The plug would prevent XPD to load onto DNA because it clashes with the DNA trajectory through the helicase (Fig. 21). In addition, a loop in the p62 subunit enters the active site of repressed XPD, as seen in the TFIIH-PIC structure<sup>70</sup>, where it might mimic DNA and additionally prevent XPD from binding to DNA. However, since the helicase activity of the core TFIIH is still quite robust (Fig. 11), the contribution of p62 is likely minor.

In contrast to transcription initiation, the XPD enzymatic activity is essential for repair<sup>181</sup>. We systematically investigated the effect of known NER factors on XPD helicase activity and we identified two factors with a stimulatory effect, XPA and XPG. The role for XPA in XPD stimulation was previously observed<sup>56,73</sup>. However, XPG is a newly identified XPD regulator. In the presence of XPG, DNA unwinding by XPD is much faster and more efficient, even compared to the unwinding facilitated by XPA (Fig. 12). It is likely that the mechanism of stimulation by XPA and XPG is quite different. XPA has a very small interaction surface with XPD (Fig. 14) and might facilitate the DNA unwinding by stabilizing TFIIH-DNA interactions and anchoring the complex to DNA (Fig. 18). However, as seen in our crosslinking network (Fig. 23), XPG strongly interacts with the ATPase lobe 1 of XPD and might directly affect the XPD helicase cycle, in analogy to other helicases<sup>271,272</sup>. This XPG function in NER has gone undetected probably because it affects the kinetics of the NER reaction which is difficult to assay *in vitro* unless the step is isolated, as done in our study. Thus, by displacing the XPC lesion sensor from the DNA<sup>63,194</sup> and strongly stimulating XPD helicase activity (Fig. 12), XPG may facilitate efficient transition from the repair bubble opening to lesion scanning.

### 3.7 Novel insights into the mechanism of TCR and a first case of active RNA-polymerase backtracking in eukaryotes

TCR is an NER sub-pathway which relies on actively transcribing RNAP for the initial lesion detection<sup>120,152,153</sup>. Despite decades of TCR research surprisingly little is known about the underlying mechanisms<sup>80</sup>. Several protein factors were identified which are essential for TCR but are not required for GGR<sup>120,152,153</sup>. Thus, it was generally accepted that these TCR-

specific factors orchestrate the initial steps of TCR until the canonical NER machinery takes over to perform the DNA excision. However, there is currently very little experimental data showing how or if TCR-specific factors participate in the repair reaction<sup>80</sup>. The main bottleneck in TCR research is the difficulty to isolate numerous protein factors involved, which includes TCR-specific factors, RNAP and canonical NER factors, as well as the lack of mechanistic understanding of the process to guide experimental designs. In this study we successfully purified all components involved in the core TCR reaction and we established an experimental system that allows us to monitor the fate of the lesion-arrested RNAP.

Arrested elongation complexes sequester the lesion site<sup>233</sup> and need to be removed for the repair to take place. This seems to be the central, but still unresolved mechanistic problem in TCR. In bacteria, a dedicated protein Mfd recognizes arrested elongation complexes and dislodges RNAP from the site of damage by pushing the RNAP forward<sup>273,274</sup>. Since CsB is also a translocase and pushes the RNAP forward<sup>137</sup> (Fig. 27), it might act in TCR in a similar way. It was shown that CsB cannot dislodge nucleotide starved elongation complexes<sup>50</sup>, but it is unknown if the same conclusion can be applied to the lesion-arrested RNAP. In addition, we discovered that an additional TCR-specific factor UVSSA directly binds RNAP and stimulates transcription *in vitro* (Fig. 27). However, we found that neither CsB nor a CsB-UVSSA complex can dissociate the lesion-arrested RNAP from DNA, even after hours of incubation and a constant supply of ATP (Fig 28, 40). In addition, no transcriptional bypass of the lesion was observed, so CsB and UVSSA also cannot rescue arrested complexes by facilitating transcription past the lesion (Fig. 28, 40). It seems that the TCR mechanism is not conserved from bacteria to humans, because the Mfd-mediated dissociation of arrested elongation complexes is the dominant TCR pathway in bacteria<sup>275</sup>.

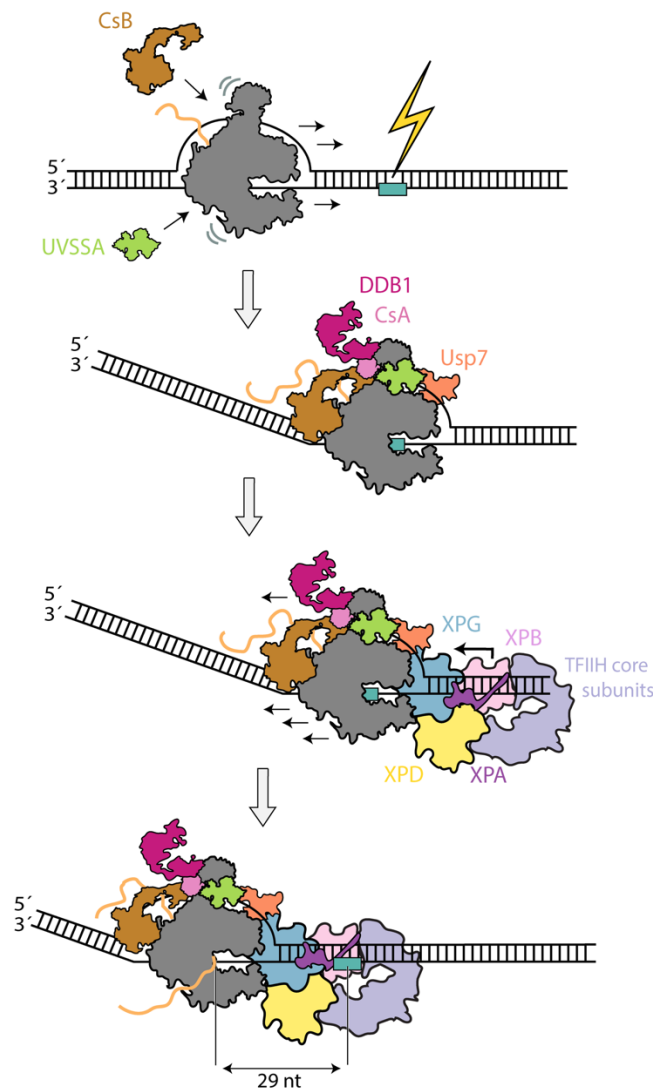
Another possible scenario is that the lesion-arrested RNAP backtracks far enough to expose the lesion to repair enzymes<sup>80,120,153</sup>. A big benefit of such a strategy is that transcription can resume after the lesion has been removed. This would be particularly important in mammals because they have longer genes which require more time and energy to be transcribed<sup>276</sup>. Even though a mechanism of RNAP removal via backtracking was proposed in all recent reviews on TCR<sup>80,120,153</sup>, up to now there is no experimental data to support this model, probably due to the complexity of this multicomponent biochemical system. Furthermore, an active RNAP backtracking, i.e. pushing or pulling the RNAP backwards by an ATP-consuming motor protein, has to my knowledge never been demonstrated in eukaryotes. Here we show that TFIIH can push the lesion-arrested RNAP backwards (Figs. 30, 40). Surprisingly, this activity is mediated by the XPB translocase and further requires the presence of two additional NER factors, XPA and XPG (Figs. 29, 31). The requirement for XPA is easy to rationalize because XPA strongly stimulates the XPB translocase activity (Fig. 12), which governs RNAP removal (Fig. 31). However, the role for XPG is still not clear. XPG strongly stimulates the XPD helicase activity (Fig. 12), however, neither XPD helicase activity nor XPD presence was required for efficient RNAP

backtracking (Fig. 31). We observed that in the absence of XPG, the TFIIH-XPA complex pushes RNAP forward, instead of backwards (Fig. 29c). This could be explained by a random XPB loading onto the DNA scaffold in our experimental setup. If XPB is loaded on the downstream DNA it would collide with the RNAP from the front and induce RNAP backtracking. However, if XPB is loaded on the upstream DNA, XPB would collide with RNAP from the back, which would push RNAP forward and rescue backtracked elongation complexes, resulting in an overall transcription stimulation. Since XPG presence specifically induces RNAP backtracking, we could speculate that XPG recruits TFIIH-XPA on the downstream DNA, potentially by interacting with RNAP. Thus, XPG may couple the backtracking machinery to RNAP. This would be in agreement with previous reports showing that XPG participates in transcription<sup>277</sup> and directly interacts with the elongation complex<sup>278</sup>. In addition, after the backtracking XPG would be ideally positioned to cleave DNA 3' of then lesion, as observed during the canonical NER reaction<sup>107,224</sup>.

We also demonstrated that the core TFIIH-XPA-XPG complex pushes the lesion-arrested RNAP backwards for 29 nucleotides (Figs. 32, 40), which is sufficient for a full clearance of the lesion site. It is still unclear why we observed such a discrete backtracking distance. It would be tempting to speculate that XPB stops pushing RNAP after reaching the lesion site. This would suggest a lesion verification role for XPB during TCR. Interestingly, during transcription initiation XPB is located ~30 nucleotides downstream of the RNAP active site<sup>70,198</sup>. If XPB is similarly positioned during DNA repair, the backtracking distance would suggest that XPB stops upon lesion encounter. Also, we showed that artificial DNA modifications can indeed arrest XPB translocation (Fig. 33c), which is in agreement with previous biochemical studies<sup>69</sup>. In summary, we have identified a novel step of the TCR reaction pathway which includes the active backtracking of lesion-arrested RNAP by a machinery comprised of core TFIIH, XPA and XPG. The process depends on the translocase activity of XPB and completely exposes DNA lesions to repair. Our results provide important mechanistic insights into the fate of the lesion arrested RNAP in eukaryotes.

Interestingly, the TCR mechanism differs significantly between bacteria and humans. In bacteria arrested elongation complexes are efficiently removed by Mfd protein which actively dislodges RNAP from DNA<sup>273,274</sup>. Even though it was demonstrated that the UvrD helicase can also facilitate repair by pulling the RNAP backwards<sup>165</sup>, the Mfd-dependent TCR was shown to be the prevalent TCR strategy<sup>275</sup>. In contrast, a protein with an RNAP dislodging function has not been discovered in humans. In our study we have demonstrated that the canonical NER factors can recognize arrested elongations complexes and actively push the RNAP backwards to clear the lesion site for the repair enzymes (Fig. 40). The backtracking strategy might have been disfavored in bacteria due to transcription-translation coupling<sup>279</sup>. The pioneering ribosome might even be physically linked to the RNAP<sup>280</sup> which would make backtracking very challenging. In contrast, the human TCR machinery might have evolved to preserve RNA transcripts, as genes are much longer in humans compared to bacteria. For example, the human dystrophin gene is 2.8 mega-bases

long and it takes almost 16 h to transcribe<sup>281</sup>, thus it would be very energetically expensive to abort the transcription every time a lesion is encountered. Repurposing already existing NER factors for the RNAP backtracking, as observed in our biochemical analysis, would offer further advantages: (i) the repair machinery is already present when the lesion is uncovered and there is no need for additional recruitment steps, (ii) specific positioning of the TCR machinery on the downstream DNA might establish which DNA strand carries the lesion and (iii) XPB translocase activity could be employed for lesion verification.



**Figure 40 | New insights into the mechanism of human TCR.**

TCR-specific factors CsB and UVSSA stimulate transcription elongation<sup>137</sup> (Fig. 27). A UV-induced CPD lesion (indicated by the blue square) arrests RNAP migration on the DNA<sup>51,233,234</sup> (Fig. 25). TCR-specific factors cannot facilitate transcription past the lesion nor dissociate the arrested RNAP from the DNA (Fig. 28a, b), however, the factors favor the addition of nucleotides opposite to the lesion (Fig. 28a) which may change the RNAP conformation and favor the recruitment of downstream repair factors. The backtracking machinery comprised of the TFIID core, XPA and XPG binds the arrested RNAP and pushes it backwards (Figs. 29, 30). The backward translocation of RNAP is mediated by the XPB translocase activity (Fig. 31) which is strongly stimulated by XPA (Fig. 12). XPG most likely couples the core TFIID-XPA complex to RNAP. The RNAP is pushed backwards for 29 nucleotides (Fig. 32) which is sufficient to fully expose the lesion site to repair enzymes. The active RNAP backtracking might stop when XPB translocase encounters the lesion (Fig. 33c), which might verify the presence of the lesion and initiate the canonical NER pathway.

Finally, it is important to note that we characterized RNAP backtracking *in vitro* and, even though important basic principles of the reaction were uncovered, it is still unclear to which extent this process occurs *in vivo*. Moreover, backtracked RNAP could still be a target for proteasomal degradation and might not resume with transcription after the repair is completed. Exciting future research, driven by advancement in structural<sup>282</sup> and systems biology approaches<sup>283</sup>, will further elucidate the mechanism of TCR.

### **3.8 Towards a complete molecular mechanism of human NER**

NER is one of the most important DNA repair pathways because it has a very broad substrate specificity and it can eliminate many different types of DNA lesions<sup>19,106</sup>. Unfortunately, DNA-modifying chemotherapeutics, such as the frequently used cisplatin<sup>284,285</sup>, are also very good NER substrates<sup>286</sup>. Moreover, cancer cells can intensify NER as a mean to resist chemotherapy<sup>287</sup>. Understanding the molecular mechanism of NER and gaining a structural insight into reaction intermediates will be instrumental for the design of potent NER inhibitors which can boost the efficiency of existing anti-cancer medications<sup>288</sup>. NER can also serve as a paradigm for the investigation of diseases caused by a dysregulation in the response to DNA damage because defects in NER cause a striking variety of clinical features<sup>20</sup>. The work described here provides a basis for the complete structural characterization of the core NER reaction. In addition, we described several structure-based mechanisms underlying NER and we biochemically characterized completely novel step of the NER reaction pathway. However, future research will address several pending open questions, as follows.

Extensive functional and biochemical analysis of GGR elucidated the order of assembly of NER factors and outlined the general sequence of events on the path to lesion excision<sup>20,106</sup>. However, a structure-based molecular mechanism of the NER reaction is lacking. Since we optimized the purification of all NER components and provided a general protocol for structure determination of TFIIH-bound repair factors (Methods), future structural work will mainly involve minor modifications, such as redesigning the DNA scaffold to capture different NER intermediates. Three structures in particular would be important to solve in order to complete the mechanism of GGR. First, a structure of XPC-Rad23 bound to TFIIH. This structure would provide insight into the repair bubble opening and explain how TFIIH is properly oriented to initiate scanning on the damaged DNA strand<sup>199</sup>. Second, a structure of the lesion-arrested XPD helicase would provide the mechanism of lesion verification during NER<sup>23,73-75</sup>. It may be useful to test different types of lesions, such as bulky DNA modifications, intra-strand crosslinks and photo-lesions, since different lesion-recognition strategies might be employed. Third, a structure of the pre-precision complex would uncover how the size of the excised DNA fragment is defined and how the structure-specific endonucleases XPG and XPF-ERCC1 are properly positioned to



cleave the damaged DNA strand at the positions observed in dual-incision assays<sup>65,224</sup>. This structure might also shed light on the role of RPA, as it was suggested that the specific polarity of RPA binding to the non-damaged DNA strand orients endonucleases on the repair bubble<sup>105</sup>. In addition, reparatory DNA synthesis has not been extensively researched, even though DNA replication and DNA ligation are in general well understood<sup>289</sup>. However, it would be interesting to investigate how 3' and 5' incision events are coordinated during the gap-filling and how NER factors are displaced from the site of damage.

In contrast to GGR, we are only beginning to understand the mechanism of TCR and many basic features of this repair pathway still need to be unraveled<sup>80</sup>. CsB has been at the focus of TCR research<sup>134</sup> but its role still remains elusive. Here we showed that CsB does not directly participate in the TCR reaction since CsB cannot push RNAP over a photolesion CPD<sup>137</sup> or dislodge lesion-arrested elongation complexes from the DNA, even in the presence of other TCR-specific factors (Fig. 28). However, a recent structure of the yeast CsB homologue Rad26 bound to the elongation complex revealed that CsB binds the upstream DNA<sup>138</sup>, which is usually occupied by the core transcription elongation factor DSIF<sup>290</sup>. The structure thus raises an interesting possibility that the function of CsB is to strip away elongation factors from the RNAP surface<sup>80</sup>, which would facilitate RNAP processing during repair. This could be easily tested in a reconstituted system, as the one described here, by a competition experiment between CsB and DSIF for binding the arrested RNAP. We also showed that core TFIIH, in complex with XPA and XPG, pushes the arrested RNAP backwards (Figs. 30, 32). Since all components are available in large amounts and high purity, it would be interesting to solve the cryo-EM structure of the RNAP backtracking complex. This would provide insights into how XPB translocase pushes RNAP backwards and how XPG can couple the backtracking-inducing machinery to the elongation complex. On a larger scale, it might be beneficial to make an attempt at reconstituting the entire TCR reaction *in vitro* and to investigate if RNAP can continue transcribing after the lesion is removed by repair enzymes. This would require isolation of the gap-filling machinery and the construction of a biochemical system that would allow simultaneous monitoring of transcription and DNA excision. Due to advancement of the genomic systems biology toolkit<sup>283</sup>, it would also be beneficial to develop strategies for monitoring TCR in a cellular context. Although a challenging task, such a system would give us a better overview on the dynamics of arrested elongation complexes *in vivo*, when other factors, such as chromatin, post-translational modifications and proteasomal degradation come into play.

Overall, the first structure of an NER intermediate described here and protocols for the reconstitution of further steps along the NER pathway mark a milestone in structural biology of this important DNA repair pathway. It seems likely that several other structural snapshots of the NER machinery in action will be obtained over the next several years which will hopefully complete our understanding of NER.

## 4. MATERIALS AND METHODS

### 4.1 Chemicals and consumables

**Table 2** | List of chemicals and consumables.

Chemical/Consumable	Manufacturer
Acetic acid	Merck kGaA, Darmstadt, Germany
Acrylamide/Bis Solution 19:1	ThermoFischer Scientific, Massachusetts, US
Agarose, UltraPure	ThermoFischer Scientific, Massachusetts, US
Ammonium chloride	Merck kGaA, Darmstadt, Germany
Ammonium peroxodisulfate (APS)	Merck kGaA, Darmstadt, Germany
Ampicillin	Sigma-Aldrich Chemie GmbH, Steinheim, Germany
Amylose Resin	New England BioLabs GmbH, Frankfurt on Main, Germany
ATP	Jena Biosciences GmbH, Jena, Germany
Bromphenol Blue sodium salt	Merck kGaA, Darmstadt, Germany
Calf Intestinal Alkaline Phosphatase	New England BioLabs GmbH, Frankfurt on Main, Germany
Chloramphenicol	Sigma-Aldrich Chemie GmbH, Steinheim, Germany
Deoxyribonucleotide (dNTP) mix	New England BioLabs GmbH, Frankfurt on Main, Germany
Dimethylformamide (DMF)	Merck kGaA, Darmstadt, Germany
Disodium hydrogen phosphate dihydrate	Merck kGaA, Darmstadt, Germany
Dimethyl sulfoxide (DMSO)	Merck kGaA, Darmstadt, Germany
DNase I	ThermoFischer Scientific, Massachusetts, US
DpnI	New England BioLabs GmbH, Frankfurt on Main, Germany
DTT	Sigma-Aldrich Chemie GmbH, Steinheim, Germany
Dynabeads MyOne Streptavidin T1	ThermoFischer Scientific, Massachusetts, US
Econo-Pac Chromatography columns	Bio-Rad Laboratories GmbH, Germany
EcoRV-HF	New England BioLabs GmbH, Frankfurt on Main, Germany
Ethanol	Merck kGaA, Darmstadt, Germany
Ethylenediaminetetraacetic acid (EDTA)	Merck kGaA, Darmstadt, Germany
Gentamicin sulfate salt	Sigma-Aldrich Chemie GmbH, Steinheim, Germany
Glycerol	Merck kGaA, Darmstadt, Germany
HEPES	Sigma-Aldrich Chemie GmbH, Steinheim, Germany
HiTrap Q column	GE Healthcare, Germany
Hydrochloric acid	Merck kGaA, Darmstadt, Germany
Imidazole	Merck kGaA, Darmstadt, Germany

IPTG	Carl Roth GmbH, Karlsruhe, Germany
Kanamycin sulfate	ThermoFischer Scientific, Massachusetts, US
L-lactic Dehydrogenase	Sigma-Aldrich Chemie GmbH, Steinheim, Germany
Magnesium chloride hexahydrate	Merck kGaA, Darmstadt, Germany
NADH	Sigma-Aldrich Chemie GmbH, Steinheim, Germany
NTP solution	ThermoFischer Scientific, Massachusetts, US
Phosphoenolpyruvate (PEP)	Sigma-Aldrich Chemie GmbH, Steinheim, Germany
Phusion High Fidelity DNA Pol.	New England BioLabs GmbH, Frankfurt on Main, Germany
PmeI	New England BioLabs GmbH, Frankfurt on Main, Germany
Potassium chloride	Merck kGaA, Darmstadt, Germany
Potassium hydroxide	Merck kGaA, Darmstadt, Germany
Proteinase K	New England BioLabs GmbH, Frankfurt on Main, Germany
Pyruvate kinase from rabbit muscle	Roche Diagnostics Deutschland GmbH, Mannheim, Germany
QIAprep Spin Miniprep Kit	Qiagen, Hilden, Germany
QIAquick Gel Extraction Kit	Qiagen, Hilden, Germany
T4 DNA ligase	New England BioLabs GmbH, Frankfurt on Main, Germany
Slide-A-Lyzer Dialysis Cassettes	ThermoFischer Scientific, Massachusetts, US
Sodium acetate	Merck kGaA, Darmstadt, Germany
Sodium bicarbonate	Merck kGaA, Darmstadt, Germany
Sodium chloride	Merck kGaA, Darmstadt, Germany
Sodium hydroxide	Merck kGaA, Darmstadt, Germany
SspI	New England BioLabs GmbH, Frankfurt on Main, Germany
Sucrose	Merck kGaA, Darmstadt, Germany
SwaI	New England BioLabs GmbH, Frankfurt on Main, Germany
TBE Buffer, 10x	ThermoFischer Scientific, Massachusetts, US
TEMED	Sigma-Aldrich Chemie GmbH, Steinheim, Germany
UltraPure DNase/RNase -Free	ThermoFischer Scientific, Massachusetts, US
Distilled Water	
Urea	Merck kGaA, Darmstadt, Germany
X-Gal	Sigma-Aldrich Chemie GmbH, Steinheim, Germany
Xylene cyanol FF	Merck kGaA, Darmstadt, Germany
$\beta$ -mercaptoethanol	Sigma-Aldrich Chemie GmbH, Steinheim, Germany

## 4.2 Buffers and growth media

**Table 3** | List of buffers.

Buffer	Composition	Application
PBS	137 mM NaCl, 2.7 mM KCl, 10 mM Na <sub>2</sub> HPO <sub>4</sub> pH 7.4, 1.76 mM KH <sub>2</sub> PO <sub>4</sub> pH 7.4	Western blotting
TBE	89 mM Tris, 89 mM boric acid, 2 mM EDTA	denaturing urea PAGE
TAE	40 mM Tris, 20 mM acetic acid, 1 mM EDTA	agarose gel electrophoresis
CutSmart (NEB)	50 mM potassium acetate, 30 mM Tris-acetate, 10 mM magnesium acetate, 100µg/ml BSA, pH 7.9	LIC cloning
NEB 3.1	100 mM NaCl, 50 mM Tris-HCl, 10 mM MgCl <sub>2</sub> , 100µg/ml BSA, pH 7.9	LIC cloning
T4 ligase (NEB)	50 mM Tris-HCl, 10 mM MgCl <sub>2</sub> , 10 mM DTT, 1 mM ATP, pH 7.5	Cloning, DNA ligation
CIP phosphatase (NEB)	50 mM potassium acetate, 30 mM Tris-acetate, 10 mM magnesium acetate, 100µg/ml BSA, pH 7.9	Cloning, DNA dephosphorylation
NuPAGE LDS Sample buffer	141 mM Tris, 106 mM Tris-HCl, 2% LDS, 10 % glycerol, 0.51 mM EDTA, 0.22 mM SERVA Blue G, 0.127 mM Phenol Red, pH 8.5	SDS-PAGE

**Table 4** | List of growth media for *E. coli* and insect cells.

Media	Composition	Supplier	Application
LB	1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl	in house made	<i>E. coli</i> cultures
2X YT	1.6 % (w/v) Bacto tryptone, 1 % (w/v) Bacto yeast extract, 0.5 % (w/v) NaCl	in house made	<i>E. coli</i> protein expression
Gibco® Sf-900™ III SFM	low-hydrolysate, serum-free, protein-free, animal origin-free	Life Technologies	Sf9 cultures
ESF921™	serum-free, protein-free, supplemented with L-glutamine and Kolliphor® P188	Expression Systems	Sf21 and Hi5 cultures

**Table 5** | List of antibiotics and supplements for E. coli cultures.

Antibiotics/Supplements	Stock concentration	Final concentration
Kanamycin	50 mg/ml	50 µg/ml
Ampicillin	100 mg/ml	100 µg/ml
Chloramphenicol	34 mg/ml	34 µg/ml
Gentamycin	10 mg/ml	10 µg/ml
X-Gal	150 mg/ml	150 µg/ml
IPTG	1 M	1 mM

### 4.3 Bacterial and insect cell strains

**Table 6** | Bacterial strains used in this study.

Bacterial strain	Genotype	Supplier	Application
XL1-Blue	<i>recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacI<sup>q</sup>ZΔM15 Tn10 (Tet<sup>r</sup>)]</i>	Agilent	cloning
BL21-Codon Plus(DE3)-RIL	<i>E. coli B F ompT hsdS(r<sub>B</sub><sup>-</sup> m<sub>B</sub><sup>-</sup>) dcm<sup>+</sup> Tet<sup>r</sup> E. coli gal λ (DE3) endA Hte [argU ileY leuW Cam<sup>r</sup>]</i>	Agilent	protein expression
Rosetta B834(DE3)	<i>E. coli B F- ompT hsdS(r<sub>B</sub>- m<sub>B</sub>-) dcm<sup>+</sup> metB (DE3)</i>	Novagen	protein expression
DH10EMBacY	<i>F mcrA Δ(mrr-hsdRMS-mcrBC) φ80dlacZΔM15 ΔlacX74 endA1 recA1 deoR Δ(ara, leu)7697 araD139 galU galK λ- rpsL nupG / bMON14272<sup>+</sup> yfp+/ pMON7124</i>	Geneva Biotech	bacmid preparation

**Table 7** | Insect cell strains used in this study.

Strain	Species	Source	Application
Sf9	<i>Spodoptera frugiperda</i>	Life Technologies	Vo and V1 production
Sf21 (IPLB-Sf-21-AE)	<i>Spodoptera frugiperda</i>	Expression Systems, LLC	V1 production
Hi5	<i>Trichoplusia ni</i>	Expression Systems, LLC	Protein expression

## 4.4 Plasmids

**Table 8** | Vectors produced for protein expression in this study.

Insert gene	Vector	Tag	Antibiotic resistance	Expression system
CsB	438B	N-terminal 6X His tag, TEV cleavage site	Ampicillin	insect cells
<b>DDB1</b>	438B	N-terminal 6X His tag, TEV cleavage site	Ampicillin	insect cells
<b>CsA</b>	438A	no tag	Ampicillin	insect cells
UVSSA	438B	N-terminal 6X His tag, TEV cleavage site	Ampicillin	insect cells
Usp7	438B	N-terminal 6X His tag, TEV cleavage site	Ampicillin	insect cells
<b>TFIIS*</b>	1C	N-terminal 6X His tag, MBP tag, TEV cleavage site	Kanamycin	E. coli
<b>TFIIS (D282AE283A)</b>	1C	N-terminal 6X His tag, MBP tag, TEV cleavage site	Kanamycin	E. coli
<b>XPD</b>	438C	N-terminal 6X His tag, MBP tag, TEV cleavage site	Ampicillin	insect cells
<b>XPD (K48R)</b>	438C	N-terminal 6X His tag, MBP tag, TEV cleavage site	Ampicillin	insect cells
<b>XPB</b>	438B	N-terminal 6X His tag, TEV cleavage site	Ampicillin	insect cells
<b>p62</b>	438B	N-terminal 6X His tag, TEV cleavage site	Ampicillin	insect cells
<b>p52</b>	438A	no tag	Ampicillin	insect cells
<b>p44</b>	438B	N-terminal 6X His tag, TEV cleavage site	Ampicillin	insect cells
<b>p34</b>	438A	no tag	Ampicillin	insect cells
<b>TTDA</b>	438A	no tag	Ampicillin	insect cells
<b>CKD7</b>	438B	N-terminal 6X His tag, TEV cleavage site	Ampicillin	insect cells
<b>CDK7 (D137R)</b>	438B	N-terminal 6X His tag, TEV cleavage site	Ampicillin	insect cells
<b>CDK7 (T170A)</b>	438B	N-terminal 6X His tag, TEV cleavage site	Ampicillin	insect cells
<b>MAT1</b>	438B	N-terminal 6X His tag, TEV cleavage site	Ampicillin	insect cells
<b>Cyclin H</b>	438B	N-terminal 6X His tag, TEV cleavage site	Ampicillin	insect cells
<b>XPG</b>	438C	N-terminal 6X His tag, MBP tag, TEV cleavage site	Ampicillin	insect cells
<b>XPG (E791A)</b>	438C	N-terminal 6X His tag, MBP tag, TEV cleavage site	Ampicillin	insect cells
<b>XPC</b>	438C	N-terminal 6X His tag, MBP tag, TEV cleavage site	Ampicillin	insect cells
<b>Rad23b</b>	438A	no tag	Ampicillin	insect cells
<b>XPF</b>	438B	N-terminal 6X His tag, TEV cleavage site	Ampicillin	insect cells
<b>ERCC1</b>	438A	no tag	Ampicillin	insect cells
<b>RPA1</b>	14 B	N-terminal 6X His tag, TEV cleavage site	Ampicillin	E. coli
<b>RPA2</b>	14 A	no tag	Ampicillin	E. coli
<b>RPA3</b>	14 A	no tag	Ampicillin	E. coli
<b>XPA</b>	438B	N-terminal 6X His tag, TEV cleavage site	Ampicillin	insect cells

Genes (or their mutated counterparts) on the same colored background were combined into single vectors.

\*cloned by Seychelle M. Vos

## 4.5 Gel electrophoresis

Gel electrophoresis was used for the analytical or preparative separation of proteins or nucleic acids.

### 4.5.1 Agarose gel electrophoresis

Agarose gel electrophoresis was used during cloning for the identification of vectors containing the insert of choice and for the purification of vectors and their digested products depending on their subsequent application. Agarose gel was prepared by mixing agarose with TAE buffer; 1% (w/v) of agarose was used in most cases but the percentage of agarose was increased up to 2% (w/v) for the separation of shorter DNA fragments. The agarose suspension was boiled until the agarose was fully dissolved. Agarose solution was then cooled to approximately 50 °C and supplemented with 1µl/ml of SYBR<sup>TM</sup>Safe DNA Gel stain (Invitrogen) to allow for a later visualisation of DNA under the ultraviolet light. Gels were left to slowly polymerise at RT. Samples for electrophoresis were mixed with the 6x DNA Loading Dye (NEB) and directly used for gel loading. The 1kb DNA Ladder size standard (NEB) was run alongside the samples. Samples were run at 100 – 150 V until the desired separation of DNA products was achieved. GEL iX20 Imager system (Intas) was used for the final DNA visualization.

### 4.5.2 Sodium-dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was used for the analytical separation of proteins according to their size during protein expression and purification, and many biochemical assays. Samples were mixed with the NuPAGE<sup>TM</sup> LDS Sample Buffer (4x, Thermo Fischer Scientific), heated to 95 °C for 3 min and loaded onto the pre-cast gradient NuPAGE 4-12% Bis-Tris Protein Gels (Invitrogen). The PageRuler<sup>TM</sup> Prestained Protein Ladder (Thermo Fischer Scientific) or Precision Plus Protein<sup>TM</sup> Dual Color Standard (Bio-Rad) molecular weight standards were run alongside the sample, depending on the expected sizes of proteins of interest. Gels were run at 150-200 V in 1x MOPS or 1x MES buffer (Invitrogen) for a better separation of proteins in the higher or lower molecular weight range, respectively. Proteins were stained with Instant Blue (Invitrogen) for several hours while shaking and destained with ddH<sub>2</sub>O until the background staining was no longer visible. Gels were scanned with the Epson Perfection V800 flatbed scanner (Epson) for documentation.

### 4.5.3 Denaturing urea polyacrylamide gel electrophoresis

Denaturing urea PAGE was used for the analytical separation of RNA after transcription and backtracking assays. Gel solution was prepared by mixing urea (8M final) and the 40% Acrylamide/Bis solution, 19:1 (Thermo Fischer Scientific, 20% final) in 1X TBE buffer (Carl Roth). The solution was incubated in a water bath operating at 50 °C and occasionally mixed until the urea was dissolved. The gel solution was supplemented with ammonium persulphate and TEMED, and poured into the gel gaskets compatible with the XCell SureLock™ Mini-Cell Electrophoresis System (Thermo Fisher Scientific). Gels were left to polymerize for 30 min at RT. Gels were pre-run at 300V for 15 min to remove the polymerization agents and to heat up the gel which facilitates the RNA separation. Samples were heated to 95 °C for 3 min and 6 µl of sample was loaded per gel lane. Gels were run in 1x TBE buffer at 300V for 90 min. RNA products were visualized using the 6-FAM or the Cy5 label and a Typhoon 9500 FLA Imager (GE Healthcare Life Sciences).

## 4.6 Cloning

Vectors encoding the full-length human CsB, CsA, DDB1, UVSSA, Usp7, XPA, XPC, Rad23b, XPG, XPF, ERCC1, RPA1, RPA2, RPA3, XPB, XPD, p62, p52, p44, p34, MAT1, Cdk7, Cych and TTDA were ordered from the Harvard Medical School PlasmID Repository and used as a template for gene amplification by PCR. The human TFIIS gene optimized for expression in *E. coli* was ordered from IDT. Cloning into the MacroBac vector series (438A, B and C, 14A, 14B and 1C) was performed as previously described (including sections 4.6.1-3, and Tables 9-11)<sup>190</sup>.

### 4.6.1 Polymerase chain reaction (PCR)

Standard PCR reactions were used for the amplification of genes received from the PlasmID Repository. This step was also used to introduce overhangs around the amplified genes (Table 9)<sup>190</sup> for the subsequent ligation-independent cloning of genes into vectors for protein expression in *E. coli* or insect cells (sections 4.6.2 and 4.6.3). The Phusion High-Fidelity DNA polymerase (NEB) was used for the DNA amplification. A typical PCR reaction of 50 µl contained dNTPs (200 µM), forward primer (0.5 µM), reverse primer (0.5 µM), template DNA (50 ng) and Phusion DNA polymerase (1 unit) in Phusion HF buffer (NEB). The PCR reaction was performed in a Professional TRIO Thermocycler® (Biometra) according to a program in Table 10. After the PCR reaction was completed, the reaction was supplemented with the 6x DNA Loading Dye (NEB) and the reaction products were separated by agarose gel electrophoresis alongside a molecular weight marker to confirm that a PCR product of the correct size was produced. If the PCR reaction was successful, the PCR product was



extracted from the gel for subsequent cloning steps. The DNA bands were visualized by a BST-20G-D2R BlueLED BioTransilluminator (Biostep) and a gel piece containing the DNA band of interest was cut out of the gel, placed into a tube and weighted. DNA was purified from the gel using the QIAquick Gel Extraction kit (Qiagen) according to the manufacturer's instructions.

**Table 9** | Overhang sequences for the ligation-independent cloning.

<b>Tag</b>	<b>Sequence</b>	<b>Vector</b>	<b>Directionality</b>
<b>v1 F</b>	TACTTCCAATCCAATGCA	1C, 14B	forward
<b>vBac</b>	TACTTCCAATCCAATCG	438A,B,C	forward
<b>v1 rv</b>	TTATCCACTTCCAATGTTATTA	1C; 438A,B,C, 14B	reverse
<b>v2 F</b>	TTAAGAAGGAGATATAGATC	14A	forward
<b>v2 rv</b>	TTATGGAGTTGGGATCTTATTA	14A	reverse

**Table 10** | PCR program used for gene amplification.

<b>Step</b>	<b>Temperature</b>	<b>Time</b>
<b>Initial denaturation</b>	98 °C	30 s
<b>Amplification (30 cycles)</b>	98 °C	10 s
	65-70 °C	30 s
	72 °C	20s / kb
<b>Final extension</b>	72 °C	10 min
<b>Hold</b>	4 °C	-

#### 4.6.2 Ligation-independent cloning (LIC)

Ligation-independent cloning (LIC) was performed as previously described<sup>190</sup>. LIC relies on the complementary sequences which flank the insert and the targeted site in the destination vector. The sequences allow specific annealing of the insert and the linearized targeted vector after each component was treated with the T4 polymerase exonuclease activity which produces single stranded overhangs driving the annealing reaction. Bacteria transformed with such an annealed product will seal the gap remaining after the annealing, thereby removing the need for the in vitro ligation step.

In the previous section we described the preparation of the insert for LIC by tagging the genes of interest with specific sequences in a PCR reaction. Targeted vectors were linearized with restriction enzymes<sup>190</sup>. 438A, B and C, 1C and 14B vectors were digested with SspI-HF (NEB) and the 14A vector with EcoRV-HF (NEB). A typical digestion reaction of 50  $\mu$ l contained 2  $\mu$ g of DNA and 100 units of restriction enzyme in the CutSmart buffer (NEB). The reaction was incubated at 37 °C for 5 h. After the digestion, the linearized vector was purified by agarose gel electrophoresis as described above. Both purified inserts and linearized vectors were treated with T4 polymerase (purified in house) in a LIC reaction. A LIC reaction was assembled as described in Table 11<sup>190</sup> and incubated at 22 °C for 30 min. In case of vectors linearized by SspI, the LIC reaction was performed in the presence of dGTP for vectors and dCTP for inserts. In case of vectors linearized by EcoRV, the LIC reaction was performed in the presence of dCTP for vectors and dGTP for inserts. The T4 polymerase was inactivated by elevating the temperature to 75 °C for 10 min. After the LIC reaction, the vector (2  $\mu$ l) and the insert (2  $\mu$ l) were combined and incubated for 1h at RT. The mixture was directly used for the transformation of chemically competent XL1-Blue cells (see section 4.6.5).

**Table 11** | LIC reaction for purified inserts and target vectors.

Component	Stock concentration	Volume
Purified vector/insert	10 ng/ $\mu$ l	10 $\mu$ l
dGTP/dCTP	25 mM	2 $\mu$ l
T4 DNA pol buffer	10x	2 $\mu$ l
DTT	100 mM	1 $\mu$ l
T4 DNA polymerase		0.4 $\mu$ l
water		4.6 $\mu$ l

### 4.6.3 Assembly of multiple genes into a single vector by LIC reactions

For the efficient expression of multi-subunit protein complexes, we cloned all subunits of a complex into a single vector, as previously described<sup>190</sup>. Different strategies were used to produce vectors for the protein expression in insect cells and in *E. coli*<sup>190</sup>. For insect cells we used the 438 vector series<sup>190</sup>. We combined 2 genes by digesting the acceptor plasmid with *Swa*I (NEB) to create the “vector” DNA, and the donor plasmid with *Pme*I (NEB) to create the “insert” DNA. A typical digestion reaction of 15  $\mu$ l contained 0.5  $\mu$ g of DNA, 0.5  $\mu$ l of enzyme in CutSmart Buffer (NEB) for *Pme*I or Buffer 3.1 (NEB) for *Swa*I. *Pme*I digestion was performed for 2 h at 37 °C and *Swa*I digestion for 2 h at 25 °C. Reaction products were separated by agarose gel electrophoresis and the desired DNA products were gel extracted with the QIAquick Gel Extraction kit (Qiagen). The purified vector and the insert DNA were combined (2  $\mu$ l each at 10 ng/ $\mu$ l), incubated at RT for 30 min and used for the transformation of chemically competent XL1-Blue cells. The vector containing 2 genes was then used for a subsequent round of gene joining by LIC and so on, until all genes were combined on a single vector.

For protein expression in *E. coli*, we used the 14A and 14B vectors. The acceptor plasmid (0.5  $\mu$ g) was digested with *Not*I (NEB) and *Pac*I (NEB) in CutSmart Buffer for 2.5 h to at 37 °C create the “vector” DNA. After the digestion, the 20  $\mu$ l reaction was supplemented with Calf Intestinal Alkaline Phosphatase (CIP, NEB, 5 units) and incubated for 5 min at 37 °C. The donor plasmid was digested with *Not*I (NEB) and *Asi*SI (NEB) in CutSmart Buffer for 2.5 h at 37 °C to create the “insert” DNA. Both digestion reactions were separated by agarose gel electrophoresis and the DNA products of interest were gel extracted with the QIAquick Gel Extraction kit (Qiagen). 10  $\mu$ l of the purified insert DNA (10 ng/ $\mu$ l) and 4  $\mu$ l of the vector DNA (10 ng/ $\mu$ l) were mixed with 400 units of T4 ligase (NEB) in a T4 ligase buffer (NEB) and incubated for 30 min at RT. 6  $\mu$ l of the ligated vector and the insert DNA were used for the transformation of chemically competent XL1-Blue cells (see section 4.6.5). The vector containing 2 genes was then used as donor vector and combined with additional genes in the next round of cloning.

### 4.6.4 Round-the-horn side-directed mutagenesis

Round-the-horn side-directed mutagenesis is an easy and convenient way to introduce mutations, insertions and deletions into any position in the vector of choice. Mutagenesis relies on 2 primers which amplify the entire plasmid in a linear form. The 3' end of each primer has to be complementary to the target sequence on the vector, however, the 5' ends can be modified to carry a single nucleotide mutation, several mutations or an insertion. Since the plasmid is linearized during the reaction, the recircularization by DNA ligation is required prior to transformation of competent bacterial cells.

Here we used the round-the-horn mutagenesis to create various point mutants; TFIIS:D282AE283A, XPD:K48R, CDK7:D137R, CDK7:T170A and XPG:E791A. The primers were created to have annealing temperature around 60 °C with IDT OligoAnalyzer Tool. The primers were 5' phosphorylated to facilitate the final ligation step. A standard PCR reaction (Table 10) was performed for the vector amplification, only the extension step was prolonged to allow enough time for the synthesis of longer DNA fragments. After the vector amplification by PCR, the reaction was supplemented with DpnI (NEB) and incubated for 3 h at 37 °C to remove the template DNA. Amplified vectors were purified by agarose gel electrophoresis and gel extraction. Extracted DNA (50 ng) was mixed with 40 units of T4 ligase (NEB) in a T4 ligase buffer (NEB) and incubated at 16 °C for 2 h. The ligated product was directly used for the transformation of chemically competent XL1-Blue cells.

#### **4.6.5 Transformation of chemically competent cells**

100 µl of chemically competent XL1-Blue, BL-21 (DE3)-RIL or Rosetta (DE3) cells were thawed on ice, supplemented with DNA (200 ng when quantification was possible) and gently mixed. The cells were incubated on ice for 30 min, followed by a heat shock at 42 °C for 45s and a recovery on ice for 2 min. The bacterial suspension was supplemented with 900 µl of LB medium (Table 4) at RT and shaken at 400 rpm and 37 °C in a ThermoMixer® C (Eppendorf). After 1-2 h of shaking the suspension was spun down and most of the supernatant was discarded, except for 100 µl which was used to resuspend the bacterial pellet. Bacterial suspension was plated on a pre-warmed LB agar plates supplemented with the appropriate antibiotics. The plates were incubated at 37 °C overnight.

#### **4.6.6 Isolation of vector DNA from bacterial cultures**

After transformation of bacteria with a vector of interest and selection on LB agar plates containing antibiotics, single colonies were picked and used to inoculate 30 ml of LB medium (Table 4) supplemented with antibiotics. The bacterial suspension was shaken at 37 °C overnight. Next, bacterial cells were harvested by centrifugation (2500g, 10 min, RT) and plasmid DNA was isolated by using the QIAprep Spin Miniprep kit (Qiagen) according to the manufacturer's protocol. DNA was quantified with a Nanodrop-2000 spectrophotometer (Thermo Scientific), flash frozen in liquid nitrogen and stored at -80 °C. DNA inserts were verified by Sanger sequencing (Seqlab, Göttingen).

## 4.7 Protein Expression

### 4.7.1 Protein expression in *E. coli*

RPA heterotrimeric complex was expressed and purified from Rosetta cells (Table 6), as described previously<sup>291</sup>. TFIIS and inactive TFIIS:D282AE283A mutant were expressed as follows. After the transformation of BL21 (DE3) RIL (Table 6) and selection on LB agar plates, a colony was used for the inoculation of 500 ml LB medium (Table 4) supplemented with kanamycin (30 µg/ml) and chloramphenicol (34 µg/ml). The culture was grown overnight at 37 °C while shaking. The next day 6L of 2x YF (Table 4) medium supplemented with kanamycin (30 µg/ml) and chloramphenicol (34 µg/ml) was inoculated with 180 ml of the overnight culture and grown at 37 °C until the OD<sub>600</sub> of 0.4 – 0.5 was reached. Next, the temperature was lowered to 18 °C and the cells were induced with IPTG (0.5 mM final). The cells were shaken at 18 °C overnight. Cells were harvested by centrifugation (7000g at 4°C for 20 min) and the supernatant was discarded. Cells were resuspended in the lysis buffer (400 mM NaCl, 20 mM Tris-HCl pH 7.9, 10 % glycerol (v/v), 1 mM DTT, 30 mM imidazole pH 8.0, 0.284 µg/ml leupeptin, 1.37 µg/ml pepstatin A, 0.17 mg/ml PMSF and 0.33 mg/ml benzamidine), flash-frozen in liquid nitrogen and stored at -80 °C.

### 4.7.2 Protein expression in insect cells

Protein expression in insect cells included several steps after the genes of interest were cloned into the respective 438 vectors: transformation of electrocompetent DH10EMBacY cells, isolation of the bacmid DNA, production of V<sub>0</sub> and V<sub>1</sub> viruses and protein expression in Hi5 cells, as previously described<sup>292,293</sup>.

### 4.7.3 Transformation of electrocompetent DH10EMBacY cells

100 µl of electrocompetent DH10EMBacY cells were thawed on ice and gently mixed with the 500 ng of DNA dissolved in ddH<sub>2</sub>O. Cells were incubated with the DNA for 15 min on ice and then transferred to a pre-chilled BIORAD Gene®/Micro™ Pulser cuvette (0.1 cm width). Electroporation was performed with a single pulse (25µF, 1.8 kV). Cells were immediately supplemented with 900 µl of LB medium (Table 4) at RT and transferred to a 15 ml culture tube. Cells were incubated at 37 °C for 5 h while shaking. After the incubation, 50 µl of transformed cells was plated on the LB agar supplemented with gentamycin (10 µg/ml), X-Gal (150 µg/ml) and IPTG (1mM) for blue-white selection. If the integration of the desired DNA into the bacmid DNA was successful, the colonies would appear white, and if the integration was unsuccessful the colonies would turn blue. The plates were incubated for 36 h at 37 °C for a reliable blue-white selection. Single white colonies were

picked and restreaked on the LB agar supplemented with gentamycin (10 µg/ml), X-Gal (150 µg/ml) and IPTG (1mM). If the colonies were still white after incubation at 37 °C for 36 h, we proceeded with the isolation of bacmid DNA.

#### **4.7.4 Isolation of bacmid DNA**

White colonies were used to inoculate 5 ml of LB medium (Table 4) supplemented with gentamycin (10 µg/ml). The bacterial culture was grown overnight at 37 °C while shaking. After the incubation, the cells were harvested by centrifugation (3000g at 4 °C for 10 min). The bacterial pellet was resuspended in 250 µl of P1 buffer (QIAprep Spin Miniprep kit, Quiagen), followed by the addition of 250 µl of P2 buffer (QIAprep Spin Miniprep kit, Quiagen) and incubation at RT for 5 min. After the incubation, 350 µl of N3 buffer (QIAprep Spin Miniprep kit, Quiagen) was added and the mixture was centrifuged at 21 000g and 4 °C for 10 min. The supernatant was transferred to a fresh 1.5 ml tube and centrifuged for another 10 min to remove the remaining precipitates. The supernatant was supplemented with 700 µl of cold isopropanol and gently mixed. The bacmid DNA was precipitated at -20 °C overnight. Next day, the DNA was pelleted by centrifugation at 21 000g and 4 °C for 15 min. The pellet was washed with 500 µl of cold 70% ethanol. After the ethanol removal, the pellets were covered with 30 µl of fresh 70% ethanol and stored at -20 °C until the transfection of insect cells.

#### **4.7.5 Transfection of Sf9 cells with bacmid DNA and V<sub>0</sub> virus production**

All consecutive steps were performed under a sterile hood (Biowizard Golden Line, Kojair) because the insect cell cultures are extremely susceptible to fungal infections. The precipitated bacmid DNA was transferred into a sterile hood and the ethanol was removed. The DNA pellets were left to dry for 10-15 min followed by the addition of 20 µl of water without resuspending the DNA to prevent DNA fragmentation. For transfection, the X-treme GENE™ 9 transfection agent (10 µl) was mixed with Sf-900™ III SFM medium (100 µl). The transfection agent solution was added to the dissolved bacmid DNA and the solution was further diluted with 200 µl of the Sf-900™ III SFM medium. The solution was incubated at RT for 1h. In the meantime, the Sf9 cells were prepared for the transfection. Density of the Sf9 cell culture and the cell viability was determined with the CASY® Modell TT Cell Counter and Analyzer System equipped with a 150 mm capillary (OMNI Life Science) according to manufacturer's instructions. The cell density was adjusted to 1.0x10<sup>6</sup> cells/ml and 3 ml of cells was added to each of the 6 wells in a 6-well plate (Greiner). Cells were incubated at 27 °C until the bacmid DNA was ready for transfection. The 300 µl of the bacmid solution was then added to 2 wells containing Sf9 cells in a dropwise manner (150 µl of solution was added per well). Typically, two bacmid solutions were prepared for a

particular DNA construct. As a control, one well contained the non-transfected cells and one well contained only the medium. Transfected cells were incubated at 27 °C in the dark without agitation. Efficiency of transfection was monitored with a fluorescence microscope (Leica). Since the bacmid DNA also contains the EYFP gene driven by the same viral promoter that drives the expression of the gene of interest, increase in fluorescence was indicative of a successful transfection. Usually, the cells were incubated for 2-3 days before the fluorescence increased. If a reasonable number of fluorescent cells was observed (at least 50 cells per well) before the fourth day, the supernatant containing the  $V_0$  virus was removed in the sterile hood, transferred to a 10 ml tube and stored at 4 °C in the dark.

#### 4.7.6 $V_1$ virus production

$V_0$  virus is still too weak to be used for a protein expression, so the virus is further propagated to produce the stronger  $V_1$  virus. 25 ml of the Sf9 or Sf21 insect cell culture (Table 7) at a density of  $1.0 \times 10^6$  cells/ml was infected with 0.5-3 ml of the corresponding  $V_0$  virus. After the infection, the density, the cell viability and the YFP fluorescence was monitored daily. The first day at which the cells do not divide is considered the day of proliferative arrest (DPA). Ideally, the cells divided once after the infection with the  $V_0$  virus, followed by DPA. However, if the initial virus is weak, the DPA is reached only after several days. In the meantime, cells were kept at the density of  $1.0 \times 10^6$  cells/ml. When DPA was reached, cells were usually grown for additional 48-72 h which was accompanied by an increase in cell diameter (from 25 to 28  $\mu\text{m}$  for Sf21 cells and from 18 to 20  $\mu\text{m}$  for Sf9 cells), a decline in cell viability and an increase in YFP fluorescence. The level of YFP was monitored with an Infinite® M1000 Pro micro-plate reader (Tecan) using excitation and emission wavelengths of 514 nm and 527 nm, respectively. When the viability dropped below 90 % (usually around 80 – 85 %), the cells were harvested by centrifugation (320g, 15 min, 4 °C). The supernatant ( $V_1$  virus) was decanted in the hood into a 50 ml falcon tube and stored at 4 °C in the dark.

#### 4.7.7 Protein expression ( $V_2$ virus production)

$V_1$  virus was used for the large-scale protein expression. Here we used primarily Hi5 cells for protein expression as Hi5 cells tend to produce larger amounts of protein. 600 ml of Hi5 cell culture at the cell density of  $1.0 \times 10^6$  cells/ml was infected with 0.3 – 2 ml of  $V_1$  virus, depending of the strength and the age of the virus. Usually the amount of the  $V_1$  virus added to Hi5 cells was optimized so that the cells divided exactly once before the DPA. Here, we used only one virus for infecting Hi5 cells, except for the expression of the core TFIIH when two viruses were used, one encoding subunits XPB, p62, p52, p44, p34 and TTDA, and one encoding XPD. The cell density, the cell viability and the YFP levels of infected cells were monitored daily. The cell density was kept at  $1.0 \times 10^6$  cells/ml throughout the expression

period. After the DPA was reached, cells were propagated for additional 48 – 72 h. When the cell viability dropped between 80 and 85 %, cells were harvested by centrifugation (238g, 20 min, 4 °C), resuspended in lysis buffer, flash frozen in liquid nitrogen and stored at -80 °C.

## 4.8 Protein purification

The RPA complex<sup>291</sup>, the TFIIF kinase module<sup>297</sup> and the pig RNAP<sup>290</sup> were purified as described before. Here we provide detailed purification protocols for the human TFIIS, the core TFIIF, XPC-Rad23b, XPA, XPG, XPF-ERCC1, CsB, CsA-DDB1, UVSSA and Usp7. All purification steps were performed at 4 °C and all buffers were filtered and thoroughly degassed immediately before use.

### 4.8.1 TFIIS and TFIIS:D282AE283A inactive mutant

Frozen bacterial cell pellets were thawed in a water bath operating at 30 °C. Cells were opened by sonication: 2x 2.5 min at 60% power (5s on, 5s off) with a cooling step on ice between the sonication rounds of 5 min. The lysate was centrifuged for 45 min at 18 000xg and 4 °C. The clarified lysate was loaded onto GE HisTrap HP 5 mL column (GE Healthcare, Little Chalfont, United Kingdom) pre-equilibrated in lysis buffer (400 mM NaCl, 10 % glycerol (v/v), 20 mM Tris-Cl pH 7.9, 1 mM DTT, 30 mM imidazole pH 8.0, 0.284 µg/ml leupeptin, 1.37 µg/ml pepstatin A, 0.17 mg/ml PMSF and 0.33 mg/ml benzamidine). The column was washed with 10 CV of a high salt buffer (as lysis buffer but with 1M NaCl), followed by a wash with the lysis buffer for additional 5 CV. Protein was eluted with a gradient from 0-100 % elution buffer (400 mM NaCl, 10 % glycerol (v/v), 20 mM Tris-Cl pH 7.9, 1 mM DTT, 500 mM imidazole pH 8.0, 0.284 µg/ml leupeptin, 1.37 µg/ml pepstatin A, 0.17 mg/ml PMSF and 0.33 mg/ml benzamidine) in 6 CV. Elution fractions were analysed on NuPAGE 4-12 % Bis – Tris Protein Gels (Invitrogen) run in MES buffer (Invitrogen) and the TFIIS containing fractions were pulled. The protein solution was supplemented with 2 mg of TEV protease and dialysed against 1L of the lysis buffer lacking imidazole overnight. Dialysed solution was passed through a freshly charged GE HisTrap HP 5 mL column (GE Healthcare, Little Chalfont, United Kingdom) pre-equilibrated in the lysis buffer without the imidazole and elution fractions were collected. Elution fractions were again analysed by SDS-PAGE and appropriate fractions were pulled and concentrated to a final volume of 2 ml with Amicon Millipore 15 ml concentrator (10,000 MWCO). Concentrated solution was loaded onto a HiLoad 16/600 Superdex 75 PG (GE Healthcare) pre-equilibrated in the size-exclusion buffer (400 mM NaCl, 10 % glycerol (v/v), 20 mM Tris-Cl pH 7.9 and 1 mM DTT) and 500 µl fractions were collected. Fractions containing



the pure TFIIS, as visualised by SDS-PAGE and staining with Instant Blue (Invitrogen), were pulled and concentrated with the Amicon Millipore 15 ml concentrator (10,000 MWCO). The concentrated protein solution was supplemented with glycerol to reach final 30% of glycerol (v/v), aliquoted, flash frozen in liquid nitrogen and stored at -80 °C.

#### 4.8.2 Core TFIIH and core TFIIH containing XPD:K48R point mutant

Cells were thawed in a water bath operating at 30 °C and opened by sonication. The lysate was clarified by centrifugation (18,000g, 30 min), followed by ultracentrifugation (235,000g, 60 min). The clarified lysate was filtered using 0.8 µm syringe filters (Millipore) and loaded onto HisTrap HP 5 mL (GE Healthcare, Little Chalfont, United Kingdom). The column was washed with 10 CV of lysis buffer, followed by 20 CV of high salt wash (800 mM KCl, 20 % glycerol (v/v), 20 mM KOH-HEPES pH 7, 5 mM β-mercaptoethanol, 30 mM imidazole pH 8.0, 0.284 µg/ml leupeptin, 1.37 µg/ml pepstatin A, 0.17 mg/ml PMSF and 0.33 mg/ml benzamidine). Column was washed again with 5 CV of lysis buffer and protein was subsequently eluted with a gradient of 0–80% elution buffer (400 mM KCl, 20 % glycerol (v/v), 20 mM KOH-HEPES pH 7, 5 mM β-mercaptoethanol, 500 mM imidazole pH 8.0, 0.284 µg/ml leupeptin, 1.37 µg/ml pepstatin A, 0.17 mg/ml PMSF and 0.33 mg/ml benzamidine). Fractions were checked on NuPAGE 4-12 % Bis – Tris Protein Gels (Invitrogen) for the presence of all core TFIIH subunits and appropriate fractions were pulled and mixed with 10 ml of amylose resin (New England BioLabs) pre-equilibrated in washing buffer (400 mM KCl, 20 % glycerol (v/v), 20 mM KOH-HEPES pH 7, 5 mM β-mercaptoethanol, 2 mM MgCl<sub>2</sub> and 10 µM ZnCl<sub>2</sub>). The protein solution was incubated with the beads for 1h while rotating. The amylose resin was poured into Econo-Pac Chromatography columns (BioRad) and washed with 5 CV of washing buffer. The protein was eluted with the washing buffer containing 100 mM maltose. The protein-containing fractions were pooled, mixed with 2 mg of TEV protease and dialysed against 2 L of dialysis buffer overnight (250 mM KCl, 20 % glycerol (v/v), 20 mM KOH-HEPES pH 7, 5 mM β-mercaptoethanol, 2 mM MgCl<sub>2</sub> and 10 µM ZnCl<sub>2</sub>). The dialysed sample was applied to DEAE (GE Healthcare) and heparin columns (GE Healthcare) in tandem and washed with 20 CV of dialysis buffer. After the removal of DEAE column, protein was eluted with a gradient of elution buffer 0-100% (1M KCl, 20 % glycerol (v/v), 20 mM KOH-HEPES pH 7, 5 mM β-mercaptoethanol, 2 mM MgCl<sub>2</sub> and 10 µM ZnCl<sub>2</sub>). Peak fractions were pooled, concentrated with the Amicon Millipore 15 ml 100,000 MWCO centrifugal concentrator and applied to the Superose 6 increase 10/300 GL column (GE Healthcare) equilibrated in a storage buffer (400M KCl, 20 % glycerol (v/v), 20 mM KOH-HEPES pH 7.0, 5 mM β-mercaptoethanol, 2 mM MgCl<sub>2</sub>). Peak fractions were again concentrated, aliquoted, flash frozen and stored at -80 °C. The TFIIH hexamer lacking the XPD subunit was purified as the full TFIIH core, only the amylose affinity step was omitted.

#### 4.8.3 CsB, UVSSA, Usp7, XPA, XPG, XPF-ERCC1, XPC-Rad23b

Cells were thawed in a water bath operating at 30 °C and opened by sonication. The CsB, UVSSA, Usp7, XPF:ERCC1, XPG, XPC-Rad23b and XPA containing lysate was applied onto GE HisTrap HP 5 mL (GE Healthcare, Little Chalfont, United Kingdom) equilibrated in lysis buffer (400 mM NaCl, 20 mM Tris-HCl pH 7.9, 10 % glycerol (v/v), 1 mM DTT, 30 mM imidazole pH 8.0, 0.284 µg/ml leupeptin, 1.37 µg/ml pepstatin A, 0.17 mg/ml PMSF and 0.33 mg/ml benzamidine). In case of XPA all downstream steps were performed in the presence of 5 mM β-mercaptoethanol and 10 µM ZnCl<sub>2</sub> instead of 1 mM DTT. The column was washed with 20 CV of high salt buffer (800 mM NaCl, 20 mM Tris-HCl pH 7.9, 10 % glycerol (v/v), 1 mM DTT, 30 mM imidazole pH 8.0, 0.284 µg/ml leupeptin, 1.37 µg/ml pepstatin A, 0.17 mg/ml PMSF and 0.33 mg/ml benzamidine). After the 5 CV wash with lysis buffer, proteins were eluted with the elution buffer gradient 0-80% (400 mM NaCl, 20 mM Tris-HCl pH 7.9, 10 % glycerol (v/v), 1 mM DTT, 500 mM imidazole pH 8.0, 0.284 µg/ml leupeptin, 1.37 µg/ml pepstatin A, 0.17 mg/ml PMSF and 0.33 mg/ml benzamidine). Pulled peak fractions were processed differently for different proteins. CsB, XPF-ERCC1 and XPA containing protein solutions were directly mixed with 2 mg of TEV protease and dialysed against 1L of dialysis buffer (400 mM NaCl, 20 mM Tris-HCl pH 7.9, 10 % glycerol (v/v), 1 mM DTT). XPG, XPC-Rad23b and UVSSA solutions were mixed with 10 ml of amylose resin (New England BioLabs) pre-equilibrated in dialysis buffer. Solutions were mixed for 1 hr and subsequently poured into Econo-Pac Chromatography columns (BioRad). Columns were washed with 10 CV of dialysis buffer followed by the elution with dialysis buffer containing 100 mM maltose. Protein containing fractions were pulled, mixed with 2 mg of TEV protease and dialysed against 1L of dialysis buffer. Dialysed solutions containing CsB, XPF-ERCC1, XPC-Rad23b, XPA, XPG and UVSSA were loaded on GE HisTrap HP 5 mL and flow through fractions were collected. Protein containing fractions were checked for contaminants on NuPAGE 4-12 % Bis – Tris Protein Gels (Invitrogen), pulled, concentrated with appropriate Amicon Millipore 15 ml centrifugal concentrator and applied onto Superdex 75 10/300 equilibrated in storage buffer (400 mM NaCl, 20 mM NaOH:HEPES pH 7.5, 10% glycerol (v/v), 10 µM ZnCl<sub>2</sub> and 5 mM β-mercaptoethanol) for XPA and Superdex 200 10/300 increase (GE Healthcare) equilibrated in a different storage buffer (400 mM NaCl, 20 mM NaOH:HEPES pH 7.5, 10% glycerol (v/v), 1mM DTT) for the rest of the proteins. Peak fractions were pulled, concentrated, aliquoted, flash frozen and stored at -80 °C.

#### 4.8.4 CsA-DDB1

Many attempts were made at purifying CsA. We tried to optimise protein expression in various *E. coli* strains at various conditions, as well as in insect cells. However, protein was always insoluble. The solubility was increased by fusing CsA to the MBP tag, however, the protein precipitated when the tag was cleaved off. Thus, we tried co-expression of CsA with DDB1, which dramatically improved protein yield and solubility.

Insect cells expressing CsA-DDB1 were thawed in a water bath operating at 30 °C and opened by sonication. CsA-DDB1 containing lysate was applied onto GE HisTrap HP 5 mL (GE Healthcare, Little Chalfont, United Kingdom) equilibrated in lysis buffer (400 mM NaCl, 20 mM Tris-HCl pH 7.9, 10% glycerol (v/v), 1 mM DTT, 30 mM imidazole pH 8.0, 0.284 µg/ml leupeptin, 1.37 µg/ml pepstatin A, 0.17 mg/ml PMSF and 0.33 mg/ml benzamidine). The column was washed with 5 CV of high salt buffer (800 mM NaCl, 20 mM Tris-HCl pH 7.9, 10% glycerol (v/v), 1 mM DTT, 30 mM imidazole pH 8.0, 0.284 µg/ml leupeptin, 1.37 µg/ml pepstatin A, 0.17 mg/ml PMSF and 0.33 mg/ml benzamidine), followed by 5 CV of low salt buffer (150 mM NaCl, 20 mM Tris-HCl pH 7.9, 10% glycerol (v/v), 1 mM DTT, 30 mM imidazole pH 8.0, 0.284 µg/ml leupeptin, 1.37 µg/ml pepstatin A, 0.17 mg/ml PMSF and 0.33 mg/ml benzamidine). HisTrap column was then coupled to a 5 mL HiTrapQ HP column (GE Healthcare) and protein was eluted with a 0-100% gradient of elution buffer (150 mM NaCl, 20 mM Tris-HCl pH 7.9, 10% glycerol (v/v), 1 mM DTT, 500 mM imidazole pH 8.0, 0.284 µg/ml leupeptin, 1.37 µg/ml pepstatin A, 0.17 mg/ml PMSF and 0.33 mg/ml benzamidine). HiTrapQ column was washed with 10 CV of low salt buffer and eluted with a 0 – 100% gradient of monoQ elution buffer (1M NaCl, 20 mM Tris-HCl pH 7.9, 10 % glycerol (v/v), 1 mM DTT, 500 mM imidazole pH 8.0, 0.284 µg/ml leupeptin, 1.37 µg/ml pepstatin A, 0.17 mg/ml PMSF and 0.33 mg/ml benzamidine). MonoQ step was necessary to separate the CsA-DDB1 binary complex from the excess of DDB1. Subsequent steps including the dialysis in the presence of TEV protease, the reverse HisTrap and the size-exclusion chromatography via Superdex 200 10/300 increase was done as described for CsB above.

#### 4.8.5 Protein quantification

Absorption of protein solution after the final purification step and concentration was measured at 260 nm and 280 nm with a NanoDrop-200 spectrophotometer (Thermo Scientific). The ratio of absorption at 260 and 280 nm informed on the contamination of protein preparations with nucleic acids. Typically, a ratio of 0.5 – 0.7 was achieved for our protein preparations and was considered DNA- and RNA-free. Extinction coefficient for the particular protein or multi-protein complexes was calculated using the Expasy ProtParam software. Final concentration was calculated as a ratio of the measured absorbance at 280 nm and the extinction coefficient.

#### 4.9 DNA and RNA constructs

All DNA and RNA constructs except from CPD-containing oligo (DNA22) were purchased from IDT, dissolved in ddH<sub>2</sub>O to final concentration of 100 μM, aliquoted, flash frozen in liquid nitrogen and stored at -80 °C. The CPD containing oligo was synthesised by Jan Seikowski in a collaboration with Prof. Claudia Höbatner (section 4.11.2).

**Table 12** | List of DNA constructs used in this study.

DNA	Sequence	Application
DNA1	GAAGCTACCATGAGACAGACAATCGCAGAGCACTCA ATCGCAGTGCCTGGATGTACCACAGTCATGCACGC GCGTGCATGACTGTGGTACATCCAGGACACTGCGAT	Florescence anisotropy
DNA2	TGAGTGCTCTGCGATTGTCTGTCTCATGGTAGCTTC/3 6-FAM/	Florescence anisotropy
DNA3	GAAGCTACCATGAGACAGACAATCGCAGAGCACTCT TACGCAGTGCCTGGATGTACCACAGTCAT GCA CGC GAAGCTACCATGAGACAGACAATCGCAGAGCACTC/i	Florescence anisotropy
DNA4	BiodT//iFluorT/ACGCAGTGCCTGGATGTACCACAGTC ATGCACGC	Florescence anisotropy
DNA5	GCGTGCATGACTGTGGTACATCCAGGACACTGCGAT TGAGTGCTCTGCGATTGTCTGTCTCATGGTAGCTTC	Florescence anisotropy
DNA6	AAGCTCAAGTACTTAAGCCTGGTCATTACTAGTACTG CC	Analytical gel-filtration
DNA7	GGCAGTACTAGTAAACTAGTATTGAAAGTACTTGGAG CTT	Analytical gel-filtration
DNA8	/56-FAM/TTCACCAGTGAGACGGGCAACAGC AAATCGCGCTCTAGACTCA	Helicase assay
DNA9	GCTCTGGCTGTTGCCCGTCTCACTGGTGAA/3BHQ_1/ CGACAACGGGCAGAGTGACCACTT/36-FAM/	Helicase assay
DNA10	/5IABkFQ/AAGTGGTCACTCTGCCCCGTTGTCGGTCTCG ACTCAGATCTGCGCGCTAAA	Helicase assay
DNA11	GTCTTCTTTTAAACACTATCTTCCTGCTCATTCTTTT TTCTTTCTTTTCTT	Translocase assay
DNA12	/5IABkFQ/AAGAAAAGAAAGAAGAAAGAAATGAGCA GGAAGATAGTGTAAAAGAAGAC	Translocase assay
DNA13	56-FAM/TTCTTTTCTTTCTTTCTTTCTTT	Translocase assay
DNA14	TCTTCTTTTAAACACTATCTTCCTGCT	Translocase assay
DNA15	/5BiosG/CATTCTTTCTTTCTTTCTTTCTTTCTT	Translocase assay
DNA16		

DNA17	/5IABkFQ/AAGAAAAGAAAGAAGAAAGAAAT	Translocase assay
DNA18	/5BiosG/GAGCAGGAAGATAGTGTTTAAAAGAAGAC	Translocase assay
DNA19	TATATCAACGGTGGTATATCCAGTGATTTTTCTCCA TTTAGCTTCCCTACCCTACTCCATAACTTCAACAACCA	Transcription assay
DNA20	TGGTTGTTGAAGTTATGGAGTAGGGTAGGAAGCTAA ATGGAGAAAAAATCACTGGATATACCACCGTTGATA	Transcription assay
DNA21	TATCTATGAATCATAATTCGACAAAGCTAAAATGAG AAAGACAATCGCAG	Transcription assay
DNA22	AGCACTG(CPD)ACGCAGT	Transcription assay
DNA23	GTCCTGTAGATACCACAGTTGTTACGTACCCTACCC ATAACTTCAACAACC	Transcription assay
DNA24	GTGGTATCTACAGGACACTGCGTAAACAGTGCTCTGC GATTGTCTTTCTCA	Transcription assay
DNA25	TATCTATGAATCATAATTCGACAAAGCTAAAATGAG AAAGACAATCGCAGAGCACTG(CPD)ACGCAGTGTCC TGTAGATACCACAGTTGTTACGTACCCTACTCCATA ACTTCA CAACC	Transcription assay
DNA26	TATCTATGAATCATAATTCGACAAAGCTAAAATGAG AAAGACAATCGCAGAGCACTGTTACGCAGTGTCTG TAGATACCACAGTTGTTACGTACCCTACTCCATAAC TTCA CAACC	Transcription assay
DNA27	biotin/GGTTGTTGAAGTTATGGAGTAGGGTACGTGAA CAACTGTGGTATCTACAGGACACTGCGTAAACAGTGC TCTGCGATTGTCTTTCTCATTTTAGCTTTGTCGAATTA TGATTCATAGATA	Transcription assay
DNA28	TGATCCATACACGAGTAACGACGCTCGAGCTTAAGT GACCTCACTGGAG	Cryo-EM
DNA29	TGATCCATACACGAGTAACGACGCTCGAGCTTAAGT GACCTCACTGGAG	Cryo-EM

**Table 13** | List of RNA constructs used in this study.

RNA	Sequence	Application
RNA1	UAUAUGCAUAAAGACCAGGC	Analytical gel-filtration
RNA2	FAM5/UUA UUA UAA UUA GGA GUA GGG U	Transcription assay
RNA3	Cy5/UUA UAA UUA GGA GUA GGG U	Transcription assay

## 4.10 Biochemical assays

### 4.10.1 Analytical gel filtration

Analytical gel filtration was used to demonstrate the reconstitution of decameric TFIIH complex (Fig. 7) and to show that UVSSA binds the RNAP elongation complex (Fig. 27). For the TFIIH reconstitution we mixed 30 pmol of purified TFIIH core with 60 pmol of purified kinase module in a final reaction of 50  $\mu$ l containing 200 mM KCl, 20 mM KOH:HEPES pH 7, 2 mM MgCl<sub>2</sub>, 10% glycerol and 5 mM  $\beta$ -mercaptoethanol. The reaction was loaded onto Superose 6 increase 3.2/300 column (GE Healthcare) pre-equilibrated in a buffer containing 150 mM KCl, 20 mM KOH:HEPES pH 7, 2 mM MgCl<sub>2</sub>, 5% glycerol and 5 mM  $\beta$ -mercaptoethanol, and 50  $\mu$ l fractions were collected. Peak fractions were analyzed on a pre-cast NuPAGE 4-12 % Bis-Tris gel (Invitrogen) in MOPS buffer (Invitrogen) and visualized by Coomassie staining.

For the analysis of UVSSA binding to RNAP, we first assembled the RNAP elongation complex. We annealed DNA6 (50 $\mu$ M final; Table 12) and RNA1 (50 $\mu$ M final; Table 13) in water by heating the reaction to 95 °C and slowly cooling (1°C/min) to 4°C in a thermocycler. We mixed 25 pmol of purified pig RNAP with 50 pmol of annealed DNA6:RNA2 and incubated at 30 °C for 10 min. Then we added 50 pmol of DNA7 (Table 12) and incubated for additional 10 min at 30 °C. We mixed 75 pmol of purified UVSSA with elongation complexes in a final reaction of 50  $\mu$ l containing 100 mM NaCl, 20 mM NaOH:HEPES pH 7.5, 10% glycerol and 1 mM DTT. The reaction was loaded onto Superose 6 increase 3.2/300 column (GE Healthcare) pre-equilibrated in buffer containing 100 mM NaCl, 20 mM NaOH:HEPES pH 7.5, 2 mM MgCl<sub>2</sub>, 5% glycerol and 1 mM DTT, and 50  $\mu$ l fractions were collected. Peak fractions were analyzed on a pre-cast NuPAGE 4-12 % Bis-Tris gel (Invitrogen) in MOPS buffer (Invitrogen) and visualized by Coomassie staining.

### 4.10.2 Fluorescence anisotropy

The fluorescence anisotropy experiment was performed by my former master student Kristina Stakyte under my supervision. Fluorescence anisotropy was used for measuring the affinity of XPC-Rad23b complex for different DNA constructs. The fully complementary DNA was made by annealing DNA1 and DNA2 (Table 12), the DNA containing a central 3 nucleotide mismatch by annealing DNA2 and DNA3 (Table 12), and the DNA containing a central bulky lesion mimic by annealing DNA4 and DNA5 (Table 12). A typical reaction of 25  $\mu$ l contained a corresponding DNA scaffold (10 nM) and XPC-Rad23b (10  $\mu$ M highest concentration followed by a half-log dilution series) in a final buffer containing 100 mM NaCl, 20 mM NaOH:HEPES pH 7.5, 5% glycerol, 100  $\mu$ g/ml BSA, 1 mM MgCl<sub>2</sub> and 0.5

mM DTT. The binding reaction was incubated at RT for 10 min followed by the anisotropy measurement with the Infinite M1000 PRO plate reader (Tecan) with the following settings: excitation wavelength 470 nm, emission wavelength 518 nm and gain of 72. Fluorescence anisotropy binding data was analysed using GraphPad Prism Version 7. Curve fitting was performed on the average of three independent experiments by a single-site binding equation, as described before<sup>294</sup>:

$$y = B_{max} \left( \frac{([x] + [L] + K_{d,app}) - \sqrt{([x] + [L] + K_{d,app})^2 - 4([x] \times [L])}}{2 \times [L]} \right)$$

where  $B_{max}$  is the maximum specific binding,  $L$  is the DNA concentration,  $x$  is XPC-Rad23b concentration, and  $K_{d,app}$  is the apparent dissociation constant for XPC-Rad23b and DNA.

#### 4.10.3 Bulk ATPase assay

The enzyme-coupled ATPase assay was performed in a similar way as previously described<sup>202</sup>. The assay uses two separate fast enzymatic reactions to couple ATP regeneration to NADH oxidation<sup>202</sup>. A typical reaction contained the protein of interest (100 nM) resuspended in 30  $\mu$ l of 80 mM potassium acetate, 20 mM KOH:HEPES pH 7, 5 mM magnesium acetate, 5 % glycerol (v/v), 0.2 mg/ml BSA, 3 mM phosphoenolpyruvate (PEP), 0.3 mM NADH and the excess of the pyruvate kinase and lactate dehydrogenase enzyme mix (Sigma). The mixture was incubated for 10 min at 30 °C and the reaction was started by the addition of ATP solution (10  $\mu$ l). ATP hydrolysis was monitored by measuring a decrease in the absorption at 340 nm with the Infinite M1000Pro reader (Tecan). Resulting curves were fitted to a linear model using the GraphPad Prism Version 6 to obtain the rates of ATP hydrolysis. The rates of ATP hydrolysis were plotted against the ATP concentration and fitted with a Michaelis-Menten model to extract the  $V_m$ ,  $K_m$  and  $k_{cat}$  reaction parameters.

#### 4.10.4 Helicase assay

The helicase assay was performed in a similar way as previously described<sup>73</sup>. DNA8 and DNA9 (Table 12) oligonucleotides were used for monitoring the helicase activity in the 5-3' direction and DNA10 and DNA11 (Table 12) were used for monitoring the helicase activity in the 3'-5' direction. The DNA annealing reaction contained the fluorescent DNA primer (25  $\mu$ M; DNA8, DNA10) and the quenching DNA oligo (37.5  $\mu$ M; DNA9, DNA11) dissolved in water. The annealing was performed in a thermocycler by heating the DNA solution to 95 °C for 5 min, followed by slow cooling (1 °C/min) to 4 °C. A typical unwinding reaction (20  $\mu$ l) contained 0.4 pmol of DNA duplex and 8 pmol of core TFIID in 100 mM KCl, 20 mM KOH:HEPES pH 7, 5% glycerol, 0.2 mg/ml BSA, 3 mM

phosphoenolpyruvate, 10 mM MgCl<sub>2</sub>, 1 mM DTT and the excess amount of pyruvate kinase (Sigma). When the effect of DNA repair factors on the unwinding was measured we supplemented the reaction with 24 pmol of the corresponding factor. The reaction mixture was preincubated at 26 °C for 10 min. The reaction was started by the addition of ATP (2 mM final) and the unwinding was monitored at 26 °C by using the Infinite M1000Pro reader with the excitation wavelength 495 nm, the emission wavelength 520 nm and the gain of 150. Percentage of unwound product was calculated by dividing the observed fluorescence intensity by the intensity of the fluorescent primer in the reaction buffer (mimicking the fully unwound DNA).

The DNA unwinding monitored by stopped-flow was performed in the same buffer conditions and with the same final protein and DNA concentrations as above. The core TFIIH preincubated with XPA or XPG was rapidly mixed with equal volume of ATP (2mM final) in the SX-20MV stopped-flow apparatus (Applied Photophysics). FAM fluorescence was monitored upon the excitation at 465 nm after passing through a KV500 cut-off filter (Schott). All time courses shown represent average of 5 technical replicates. The initial rate of DNA unwinding was calculated using Prism 7 (Graphpad software) by fitting the initial linear part of the fluorescence trace.

#### **4.10.5 Translocase assay**

The translocase assay was performed in a similar way as previously described<sup>69</sup>. The annealing reaction (10 µl) contained DNA12 (30 µM) and DNA13 (25 µM) (Table 12) in 25 mM MES pH 5.5 and 10 mM MgCl<sub>2</sub>. The reaction was heated to 95 °C for 5 min followed by slow cooling (1 °C/min) to 4 °C. After cooling, the reaction was supplemented with 1 µl of florescent DNA14 (9 µM final) (Table 12), heated to 57 °C and cooled down to 20 °C at the speed of 1 °C/min. Translocation reactions were performed exactly as described for the helicase assay (section 4.10.4), only with the triplex DNA as a substrate. A higher core TFIIH input (75 pmol in 20 µl reactions) was used when the effect of triptolide on DNA translocation was tested. When the effect of a biotin as a lesion mimic on XPB translocase activity was tested, the roadblock in the 5'-3' direction was made was mixing DNA13, DNA15 and DNA16, and the roadblock in the 3'-5' direction by mixing DNA12, DNA17 and DNA 18.

#### **4.10.6 Transcription assay**

The transcription assay was performed in a similar was as previously described<sup>123,295</sup>. The template strand (DNA19, 25 µM final) (Table 12) and RNA2 (25 µM final) (Table 13) were annealed in buffer containing 50 mM NaCl, 5 mM NaOH:HEPES pH 7.5 and 1 U/µl RNaseOUT Ribonuclease Inhibitor (Invitrogen) by heating the mixture to 95 °C for 5 min



followed by slow cooling (1 °C/min) to 4 °C. Elongation complexes were assembled by incubating a double molar excess of DNA19-RNA2 over the RNAP for 20 min at 30 °C, followed by the addition on the non-template strand (DNA20, Table 12), again at double molar excess over the RNAP, and additional incubation for 20 min at 30 °C. A typical transcription reaction (10 µl) contained 0.8 pmol of elongation complex and different amounts of factors of interest in 125 mM NaCl, 30 mM HEPES pH 7.5, 10 % glycerol, 3 mM MgCl<sub>2</sub>, 20 µg/ml BSA and 1 mM DTT. Transcription reactions were pre-warmed at 37 °C for 2 min followed by the addition of NTPs (0.2 mM final). The reaction was incubated for 10 min at 37 °C and quenched by the addition of 2x stop buffer (50 mM EDTA pH 8.0 and 7M urea in 1x TBE buffer). Proteins in the reactions were digested with Proteinase K (2 U/ml final, New England Biolabs) for 3 hrs at 37 °C. After the digestion, the reaction was heated up to 98 °C for 3 min and placed on ice. Reaction products were separated on 20 % urea denaturing PAGE (section 4.5.3) and visualized with a Typhoon FLA 9500 (GE healthcare Life Sciences). Assays were performed in duplicates or triplicates.

#### 4.10.7 Backtracking assay

RNAP elongation complexes were assembled as described in the section 4.10.6 by using the following oligos: DNA25 (Table 12) as the CPD-containing template strand, DNA26 (Table 12) as a control template strand without the lesion, DNA27 (Table 12) as the non-template strand and RNA3 (Table 13). Transcription was initiated by the addition of NTPs (0.25 mM final) and incubated for 30 min at 37 °C in the same buffer conditions as described in the section 4.10.6. In parallel, the backtracking-inducing complex was formed by mixing the core TFIID with 3-fold molar excess of XPA and XPG. Backtracking reaction (10µl) was assembled by mixing transcribed RNAP complexes (2.8 pmol) with the backtracking inducing complex (2.2, 4.4 or 8.8 pmol) or TFIID storage buffer in final buffer conditions as follows: 20 mM NaCl, 80 mM KCl, 10 % glycerol, 10 mM NaOH:HEPES pH 7.5, 6 mM MgCl<sub>2</sub>, 1 mM DTT, 3 mM PEP, 1 mM ATP and excess pyruvate kinase (Sigma). The reaction was incubated for 20 min at 37 °C, followed by the addition of TFIIS (0.5 pmol) and additional incubation of 1 min at 37 °C. The reaction was quenched with 10 µl of final reaction buffer supplemented with 50 mM EDTA and the TFIIS inactive mutant (50 pmol). Proteins in the reactions were digested with the Proteinase K (2 U/ml final, New England Biolabs) for 3 hrs at 37 °C. Sample was ethanol precipitated and resuspended in 1x TBE buffer supplemented with 7M urea. The resuspended sample was heated to 98 °C for 2 min and placed on ice. Reaction products were separated on a 9.6 % denaturing sequencing gel and the fluorescent signal was visualised with a Typhoon FLA 9500 (GE healthcare Life Sciences). Assays were performed in triplicates.

#### 4.10.8 Bead-based RNAP assays

Bead-based RNAP assays were used to monitor the RNAP polymerase dissociation from the CPD-containing DNA in the presence of TCR-specific factors (Dislodging assay) and to determine the fate of elongation complexes inhibited by the core TFIIH-XPA-XPG complex (Bead-based backtracking assay).

##### **Dislodging assay**

RNAP was transcribed into a CPD lesion as described in the section 4.10.7, only the DNA27 (Table 12) contained a biotin tag on the 5' terminus. The elongation complexes (15 pmol) were mixed with CsB (75 pmol) or CsB-CsA-DDB1-UVSSA complex (75 pmol) in a 70  $\mu$ l reaction volume and final reaction conditions as follows: 100 mM NaCl, 20 mM NaOH-HEPES pH 7.5, 6mM MgCl<sub>2</sub>, 10 % glycerol, 1 mM DTT, 0.2 mg/ml BSA, 2 mM PEP, 1 mM ATP and pyruvate kinase (1 U/ml, Sigma). Reaction was incubated for 3 hrs at 37 °C, followed by the addition of 20  $\mu$ l Dynabeads MyOne Streptavidine T1 (Invitrogen), pre-blocked with 10 mg/ml BSA in the reaction buffer for 1 hr at 37 °C. The reaction mixture was incubated with the beads for 20 min at 37 °C while shaking. Beads were separated from the supernatant on a magnetic rack and resuspended in 20  $\mu$ l reaction buffer supplemented with 5 mM CaCl<sub>2</sub> and DNaseI (200 U/ml, New England Biolabs). The elution was performed at 37 °C for 30 min while shaking. The eluate was again separated from the beads on the magnetic rack. The supernatant and the eluate were analyzed on NuPAGE 4-12 % Bis – Tris Protein Gels (Invitrogen) and stained with InstantBlue (Sigma).

##### **Bead-based backtracking assay**

Elongation complexes were assembled as described in the section 4.10.6, only the DNA20 (Table 12) was labelled with a 5'-terminal biotin. The elongation complexes (12 pmol) were mixed with 100 pmol of core TFIIH, 189 pmol XPG and 300 pmol XPA in a 70  $\mu$ l reaction volume and final reaction conditions as follows: 20 mM NaCl, 80 nM KCl, 20 mM NaOH:HEPES pH 7.5, 6 mM MgCl<sub>2</sub>, 10 % glycerol, 1 mM DTT. Reaction was incubated for 10 min at 30 °C, followed by a 2 min incubation at 37 °C. Transcription was initiated by the addition of NTPs (final 0.2 mM GTP, CTP and UTP, 1 mM ATP) and the reaction was incubated for 20 min at 37 °C and stopped by placing on ice. Reaction was split in 3x 20  $\mu$ l and each aliquot was mixed with 20  $\mu$ l of Dynabeads MyOne Streptavidine T1 (Invitrogen), which were blocked with 10 mg/ml BSA in the reaction buffer for 1 hr at 37 °C. Beads were shaken for 20 min at 37 °C and placed in a magnetic rack to separate the beads from the supernatant. The beads were washed 3 times with 100  $\mu$ l of reaction buffer. After the wash, the first reaction was supplemented with 30  $\mu$ l of washing buffer, the second reaction with 30  $\mu$ l of washing buffer containing 40 nM TFIIS and the third reaction with 30

µl of washing buffer containing 40 nM TFIIS and 0.2 mM NTPs. Beads were incubated for 2 min at 37 °C followed by supernatant removal on the magnetic rack. RNA was eluted by resuspending the beads in the washing buffer supplemented with 5 mM CaCl<sub>2</sub> and DNaseI (200 U/ml, New England Biolabs) and incubation for 5 min at 37 °C. Supernatant was separated from the beads with the magnetic rack and mixed with 2x stop buffer (50 mM EDTA pH 8.0 and 7M urea in 1x TBE buffer). Reactions were digested with the Proteinase K (2 U/ml final, New England Biolabs) for 3 h at 37 °C. After the digestion, the reaction was heated up to 98 °C for 3 min and placed on ice. Reaction products were separated by 20 % urea denaturing PAGE (section 4.5.3) and visualised with a Typhoon FLA 9500 (GE healthcare Life Sciences).

#### **4.10.9 Kinase activity assay**

We used the kinase activity assay to assess the activity of the kinase module variants containing CDK7:D137R or CDK7:T170A mutants. As CDK7 phosphorylates the C-terminal domain of RNA polymerase II during transcription initiation, we used purified yeast RNA polymerase II<sup>296</sup> dephosphorylated with lambda phosphatase during purification<sup>297</sup> as a substrate in the assay. RNA polymerase II (50 nM final) was mixed with increasing concentrations of kinase module variants (30, 100, 220 and 500 nM final) in a final buffer conditions containing 100 mM KCl, 20 mM KOH:HEPES pH 7.5, 3 mM MgCl<sub>2</sub>, 5% glycerol and 5 mM β-mercaptoethanol, and preincubated for 2 min at 30 °C. Reactions were started by the addition of ATP (0.5 mM final) and quenched after 2 min at 30 °C with EDTA (100 mM) and 4X LDS buffer (Invitrogen). Reactions were run on 4-12% Bis-Tris gel in MOPS buffer (ThermoFisher Scientific) and transferred to nitrocellulose membranes (GE Healthcare Life Sciences). The membranes were blocked with 5% (w/v) milk in PBS buffer supplemented with 0.1% Tween 20 for 1h at room temperature. The membranes were treated with primary antibody (3E8, 1:25 dilution) in 0.25 % (w/v) milk in PBS supplemented with 0.1 % Tween 20 and incubated at room temperature for 1 h. After several rounds of washing with PBS buffer supplemented with 0.1 % Tween 20, the membranes were incubated with HRP-conjugated anti-rat secondary antibody (1:5000 dilution, Sigma-Aldrich A9037) in 0.1 % (w/v) milk in PBS supplemented with 0.1 % Tween 20 and incubated at room temperature for 1 h. Antibodies were detected with SuperSignal West Pico Chemiluminescent Substrate (ThermoFisher) and the membranes were scanned with ChemoCam Advanced Fluorescence imaging system (Intas Science Imaging).

## 4.11 Synthesis of a CPD lesion and preparation of the CPD-containing DNA

### 4.11.1 Organic synthesis of a CPD formacetal analogue

Organic synthesis was performed by Jan Seikowski under the supervision of Dr. Vladimir Belov at the Max Planck Institute for Biophysical Chemistry. The synthesis of a CPD analogue in which the central phosphate is replaced with a formacetal bond (called CPD throughout the thesis) was synthesized as previously described<sup>298</sup>, with the following modifications: (i) for the detritylation of the building block B (Fig. 24) milder conditions were used – 20% HCOOH in dichloromethane/MeOH (10:6) at RT for 3h, (ii) for the cleavage of the 5'-O-pivaloyl and 3'-O-acetyl of the dimer, 0.15M sodium methoxide in MeOH (40°C for 10h) was used, followed by H<sup>+</sup> ion exchange by Amberlite, (iii) for the tritylation of the deprotected dimer the reaction time was prolonged to 26h and 0.18eq DMAP was added as a catalyst and finally, (iv) for the preparation of the final phosphoramidite, dichloromethane was used as a solvent, instead of THF. The CPD analogue was incorporated into a short DNA fragment (DNA22, Table 12) by Jan Seikowski under the supervision of Prof. Claudia Höbartner (currently at the Institute for Organic Chemistry, University of Würzburg).

### 4.11.2 Preparation of the CPD - containing DNA

The short CPD-containing DNA (DNA22, Table 12) was extended on both sides by enzymatic ligation with two additional oligonucleotides for later application in transcription and backtracking assays. Enzymatic ligation was performed by annealing three DNA pieces to a complementary DNA split which brings DNA ends to be ligated in a close proximity and results in a formation of partial DNA duplex needed for the activity of the T4 DNA ligase. We annealed DNA21 (2 nmol), DNA22 (2 nmol), DNA23 (2 nmol) and DNA24 (2 nmol) (Table 12) in water by heating the solution to 95 °C for 15 min followed by slow cooling to RT. Solution was supplemented with the T4 ligase buffer (NEB) and the T4 DNA ligase (5 units) and incubated at RT for 3 h. The reaction was quenched by the addition of 2x STOP buffer (8 M urea in TBE buffer) and heated to 95 °C for 3 min. Ligation products were separated by denaturing urea PAGE (section 4.5.3) and visualized under UV-light. The full-length ligation product was excised from the gel with a scalpel. The gel slice was crushed by passing through a syringe, and the gel pieces were incubated in water overnight at 4 °C while shaking to extract the DNA. After the overnight incubation, the gel suspension was centrifuged and the supernatant was supplemented with potassium acetate (0.3 M, pH 5.2) and 2 volumes of ice-cold ethanol. Solution was incubated at -20 °C for 2 h and the precipitated DNA was pelleted by centrifugation. The DNA pellet was washed with 500 ul

of 70% ice-cold ethanol, resuspended in ddH<sub>2</sub>O, flash-frozen in liquid nitrogen and stored at -80 °C.

## 4.12 Mass-spectrometric identification of crosslinking sites

### 4.12.1 Sample preparation

The XPC-Rad23b-core TFIIH-XPA-DNA complex (Fig. 9) was assembled as described in the section 4.13.1, only the complex was purified by the Superose 6 increase 3.2/300 column (GE Healthcare). The core TFIIH-XPA-XPG-DNA complex (Fig. 23) was assembled as described in the section 4.14.1, only the complex was purified by the Superose 6 increase 3.2/300 column (GE Healthcare). The RNAP-UVSSA-CsB-CsA-DDB1 complex (Fig. 26) was assembled by transcribing the RNAP into a CPD lesion as described in the section 4.10.7, followed by the addition of the 1.5 molar excess of UVSSA, CsB and CsA-DDB1 over the RNAP in the assembly buffer containing 100 mM NaCl, 20 mM KOH:HEPES pH 7.5, 2 mM MgCl<sub>2</sub>, 5% glycerol and 5 mM β-mercaptoethanol. The complex was purified by Superose 6 increase 3.2/300 column (GE Healthcare) equilibrated in the complex assembly buffer. The RNAP-core TFIIH-XPA-XPG complex (Fig. 33) was formed by first transcribing RNAP (100 pmol) into a CPD lesion as described in the section 4.10.7, followed by incubation with the core TFIIH (250 pmol), XPG (250 pmol) and XPA (250 pmol) at 30 °C for 30 min in the final buffer containing 200 mM KCl, 20 mM KOH-HEPES pH 7.5, 10% glycerol, 6 mM MgCl<sub>2</sub>, 1 mM DTT and 2 mM ATP. The complex was purified by the Superose 6 increase 3.2/300 column equilibrated in buffer containing 200 mM KCl, 20 mM KOH-HEPES pH 7.5, 10% glycerol, 2 mM MgCl<sub>2</sub> and 1 mM DTT. For all complexes, the fractions from the size-exclusion chromatography were analyzed by SDS-PAGE and the fractions containing the complex of interest were pulled and crosslinked with BS3 (Thermo Fisher Scientific; 1 mM final, 30 min at 30 °C). The crosslinking reaction was quenched with ammonium bicarbonate (100 mM final, 10 min at 30 °C) and the sample was processed for the mass-spectrometric analysis as describe below.

### 4.12.2 Mass spectrometry

All mass-spectrometry experiments were performed by Aleksandar Chernev under the supervision of Prof. Henning Urlaub at the Max Planck Institute for Biophysical Chemistry. Crosslinked proteins were in-solution digested, fractionated and analyzed as previously described<sup>295</sup>. Shortly – the sample was reduced with 10 mM dithiothreitol and alkylated with 40 mM iodoacetamide, followed by overnight digestion with trypsin (1:20 w/w) at 37 °C in the presence of 1M urea. The digested peptides were purified by reverse phase C18 chromatography and fractionated on a Superdex Peptide PC3.2/30 column (GE Healthcare)

with a flow rate of 50  $\mu\text{l}/\text{min}$  30% ACN, 0.1% trifluoroacetic acid (TFA) (v/v). 100  $\mu\text{l}$  fractions were collected and analyzed over two hour gradient on Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Scientific) with 120,000 resolution for survey scans and 30,000 resolution of fragment spectra. Protein-protein crosslinks were identified with pLink 1.23<sup>299</sup>. Cross-linking figures were created with XiNet<sup>300</sup> and Xlink Analyzer<sup>222</sup>.

### **4.13 Negative stain electron microscopy of the XPC-Rad23b-core TFIIH-XPA-DNA complex**

#### **4.13.1 Complex formation**

The complex formation and the negative stain electron microscopy (section 4.13.2) were performed by my former master student Kristina Stakyte, under my supervision. DNA scaffold was prepared by annealing DNA4 (Table 12; 50  $\mu\text{M}$  final) and DNA5 (Table 12; 50  $\mu\text{M}$  final) in water by heating the solution to 95  $^{\circ}\text{C}$  and slowly cooling to 4  $^{\circ}\text{C}$  in a thermocycler. Annealed DNA (250 pmol) was mixed with XPC-Rad23b (200 pmol) and incubated on ice for 20 min and at RT for 10 min. XPC-Rad23b-DNA complex was mixed with the core TFIIH (160 pmol) and XPA (400 pmol) in a final buffer containing 200 mM KCl, 20 mM KOH:HEPES pH 7.5, 2 mM  $\text{MgCl}_2$ , 5 mM  $\beta$ -mercaptoethanol and 10 % glycerol (v/v). The reaction was incubated for 30 min on ice before in was loaded onto a sucrose gradient supplemented with glutaraldehyde, as previously described<sup>197</sup>. The sucrose gradient was prepared with BioComp Gradient Master 108 (BioComp Instruments) by mixing equal volume of heavy (30 % (w/v) sucrose, 150 mM KCl, 2 mM  $\text{MgCl}_2$ , 5 mM  $\beta$ -mercaptoethanol, 20 mM KOH-HEPES pH 7.5 and 0.1 % glutaraldehyde) and light solutions (10 % (w/v) sucrose, 150 mM KCl, 2 mM  $\text{MgCl}_2$ , 5 mM  $\beta$ -mercaptoethanol and 20 mM KOH-HEPES pH 7.5) in 5 ml centrifugation tubes. After 16 h of centrifugation at 4 $^{\circ}\text{C}$  and 175,000  $\times g$  the gradient was fractionated, and glutaraldehyde was quenched with lysine (50 mM final). Fractions were analysed by SDS-PAGE and Coomassie staining. In parallel, the same sample was subjected to sucrose gradient centrifugation but in the absence of glutaraldehyde, to monitor the migration of the complex in the gradient.

#### **4.13.2 Negative staining and electron microscopy**

Quenched fractions from the sucrose gradient were directly used for negative staining. The staining solution was prepared by dissolving uranyl formate (2 % w/v) in double distilled water. Solution was vortexed for 15 min and centrifugated for 30 min at full-speed and 4  $^{\circ}\text{C}$  to remove any remaining uranyl formate precipitates. The staining solution was kept on ice and protected from light. Carbon coated copper mesh grids (S160-4, Plano) were glow-discharged (30 s, 25 mA, 0.39 mbar) by using a PELCO easiGlow System (Ted Pella). 4  $\mu\text{l}$

of sample was incubated on a grid for various amounts of time to allow the particle adsorption on the carbon foil. Grids were first washed by floating on top of a double-distilled water droplet to remove excess sucrose, followed by incubation on 3 consecutive droplets containing uranyl formate solution for 30 s. The excess stain was removed by blotting the EM grid with a filter paper and the sample was left to dry at RT protected from light. Images of the grids were acquired at RT in a CM200 FEG microscope (Philips) operated at 160 kV using a 4k x 4k CCD camera (TVIPS) with a two-fold binning of pixels. Images were acquired at a nominal magnification of 88 000x (2.51 Å / pix) with defocus values ranging from -0.5 to -2.5 µm. Particles were picked semi-automatically using the EMAN2 software<sup>301</sup>. Subsequent image processing steps, including CTF correction, 2D and 3D-classifications and 3D refinements were performed with the RELION 2.0 package<sup>302,303</sup>.

## 4.14 TFIIH-XPA-DNA structure determination

### 4.14.1 Sample preparation

Sample was prepared by mixing the pre-annealed DNA scaffold (DNA28 and DNA29; Table 12) with core TFIIH, XPA and XPG(E791A) in 1.5:1:3:3 molar ratio in final buffer containing 150 mM KCl, 20 mM KOH:HEPES pH 7, 10 % glycerol, 2 mM MgCl<sub>2</sub> and 5 mM β-mercaptoethanol. XPG:E791A endonuclease mutant<sup>107</sup> was used to prevent DNA cleavage during the sample preparation. The sample mixture was applied to a sucrose gradient in order to purify the complex from excess factors and fix it with glutaraldehyde, as previously described<sup>197</sup>. The sucrose gradient was prepared with BioComp Gradient Master 108 (BioComp Instruments) by mixing equal volume of heavy (30 % (w/v) sucrose, 150 mM KCl, 2 mM MgCl<sub>2</sub>, 5 mM β-mercaptoethanol, 20 mM KOH-HEPES pH 7.5 and 0.1 % glutaraldehyde) and light solutions (10 % (w/v) sucrose, 150 mM KCl, 2 mM MgCl<sub>2</sub>, 5 mM β-mercaptoethanol and 20 mM KOH-HEPES pH 7.5) in 5 ml centrifugation tubes. After 16 h of centrifugation at 4°C and 175, 000 xg the gradient was fractionated and glutaraldehyde was quenched with lysine (50 mM final) and aspartate (20 mM final). Fractions were dialysed in Slide-A-Lyzer MINI Dialysis Devices (2ml and 20 kDa cut-off) (ThermoFisher Scientific) for 10 h against buffer containing 100 mM KCl, 20 mM KOH-HEPES pH 7.5, 1 mM MgCl<sub>2</sub>, 1 mM DTT, 0.5% glycerol (v/v) and 0.004% n-octyl glucoside (w/v). Dialysed samples were immediately used for cryo-grid preparation. 4 µl of sample was applied to glow-discharged R2/2 gold grids (Quantifoil) which were blotted for 5 s and plunge-frozen in liquid ethane with a Vitrobot Mark IV (FEI) operated at 4 °C and 100 % humidity.

#### 4.14.2 Cryo-electron microscopy and image processing

Micrographs of the sample were acquired on a FEI Titan Krios G2 transmission electron microscope with a K2 summit direct electron detector (Gatan). Data acquisition was automated with FEI EPU software package. Micrographs were acquired at a nominal magnification of 130,000x (1.05 Å per pixel) using a dose rate of 4.55 e<sup>-</sup> per Å<sup>2</sup> per s over the time of 9 s which resulted in a total dose of 41 e<sup>-</sup> per Å<sup>2</sup> fractionated over 40 frames. CTF correction, motion correction and particle picking was done on-the-fly using Warp<sup>205</sup>. Automated picking in retrained BoxNet implemented in Warp<sup>205</sup> yielded a total of 1,354,997 particles from 8,993 micrographs, which were further subjected to 2D classification in CryoSPARC<sup>206</sup>. After 2D cleaning, 950,000 particles were used for heterogeneous refinement in CryoSPARC. Three ab-initio classes obtained from the first 300,000 particles picked during data acquisition were used as an input for the refinement. The class showing clear core TFIIH features was further 3D classified into 6 classes using RELION-3<sup>207</sup>. Particles corresponding to the best 3D class were subjected to CTF refinement and Bayesian polishing. Particles were 3D refined and post-processed with automatic B-factor determination in RELION. Final map showed an overall resolution of 3.6 Å according to the gold-standard Fourier shell correlation (FSC) 0.143 criterion with an applied B-factor of -110.02 Å<sup>-2</sup>. Due to flexibility of peripheral regions of core TFIIH, we improved the map quality for 5 different regions of the complex by focused 3D classification and refinement. The classifications were performed with particles contributing to the final map without image alignment to speed up the calculations. Masks encompassing the regions of interest were created with UCSF Chimera<sup>210</sup> and RELION. 3D classification of the DNA duplex revealed two alternative DNA conformations within the complex.

#### 4.14.3 Model building

The final cryo-EM map and focused classified maps were used for model building. The final map was denoised in Warp 1.0.6<sup>205</sup>. Structures of ATPase lobe 1 and 2 of XPB, XPD, p44 vWA-like domain and p52 C-terminus (residues 383-458) from the TFIIH structure (PDB core 5OF4)<sup>182</sup>, as well as the crystal structure of p34 vWA-like domain bound to p44 RING domain (PDB code 5O85)<sup>189</sup> were rigid-body fitted into our cryo-EM density in UCSF Chimera<sup>210</sup> and manually adjusted in COOT<sup>304</sup>. Due to high quality of the EM density, the NTE domain and part of the DRD domain (residues 71-199 and 266-300), as well as the p52 region that interacts with XPB (residues 290-382) were built *de novo* guided by secondary structure prediction in PSIPRED<sup>305</sup> and bulky amino acid side chains as sequence registers. In case of XPD we did not observe EM density corresponding to residues 273-325, so we removed this part of the structure. We observed a very strong density for the iron-sulphur cluster indicating that the ligand was not damaged or dissociated during protein expression



and purification, as well as sample preparation for cryo-EM. The N-terminal region of p52 (residues 18-289) and zinc-fingers belonging to subunits p34 and p44 were modelled with SWISS-Model<sup>306,307</sup> based on the yeast p52 counterpart (PDB code 5OQJ)<sup>70</sup> and manually adjusted in COOT. Interestingly, the p34 zinc finger region in human contains additional cysteine (C257) and histidine (H258) residues not present in the yeast counterpart which allows binding of an additional zinc ion. The smallest TFIIH subunit TTDA (p8) was generated in Modeller<sup>308</sup> with the yeast TTDA structure as a reference (PDB code 5OQJ)<sup>70</sup>, rigid-body fitted in our density using UCSF Chimera and manually adjusted in COOT. The NMR structure of truncated human XPA (PDB code 1XPA)<sup>92</sup> was also docked in our density as a rigid body and adjusted in COOT. We observed additional helical density that extends from the C-terminus of the docked structure towards the ATPase lobe 2 of XPB when the map is filtered to lower resolution. Secondary structure prediction with PSIPRED shows that the 22 residues that follow the docked XPA C-terminus form a helix, so we extended the C-terminal helix in COOT guided by the cryo-EM density.

DNA sequence was assigned based on the position of the DNA duplex-single strand junction, however protein binding to the junction could induce additional DNA melting so register shifts cannot be excluded. DNA duplex was built by docking ideal B-DNA into the density, followed by manual adjustments in COOT. Several rounds of real space refinement and geometry optimization with secondary structure restraints (including base pairing and base stacking restraints) were performed in PHENIX<sup>208</sup>. The DNA duplex-single strand junction and single-strand extensions were manually built in COOT. The EM density for the 5'-3' DNA single strand showed clear separation of sugar, phosphate and DNA bases for nucleotides A30-G36 and for C40-A41. The decreased quality of EM map for nucleotides G37-A39 and T42, presumably due to increased flexibility of DNA between XPD helicase lobes, allowed the trajectory of DNA to be determined, but the nucleotides were positioned manually in COOT guided by the structure of NS3 helicase in complex with DNA<sup>309</sup> and real space refined in PHENIX. All core TFIIH subunits, XPA and DNA were first real-space refined in PHENIX separately in their corresponding focused classified maps. Then, all components were combined and real-space refined together in the global map. The final model was validated using Molprobity<sup>310</sup> (Supplemental Table 5) and showed a score of 2.24.

## 5. SUPPLEMENTARY MATERIALS

The first four Supplementary Tables contain the crosslinking data. Only the inter-subunit crosslinking sites with the score above three are provided and the complete list of sites is available upon request. Also, if a crosslinking site was detected more than once, the highest score for the site is provided in the table. P1, protein1; P2, protein 2; R1, residue in protein 1; R2, residue in protein 2, S, score.

**Supplemental Table 1 | Core TFIIH-XPA-XPC-Rad23b crosslinking data.**

P1	P2	R1	R2	S	P1	P2	R1	R2	S
<b>XPD</b>	p44	751	190	6,94	<b>Rad23</b>	XPC	51	469	11,17
<b>XPD</b>	p62	634	293	3,87	<b>Rad23</b>	XPC	14	469	14,29
<b>XPD</b>	p62	96	365	5,1	<b>p62</b>	p34	378	125	5,74
<b>XPD</b>	p62	370	38	8,45	<b>p62</b>	p34	365	125	7,52
<b>XPD</b>	p62	113	293	8,84	<b>p62</b>	p34	370	125	10,59
<b>XPD</b>	p62	128	38	10,23	<b>p62</b>	p34	353	125	13,75
<b>XPD</b>	p62	603	293	13,31	<b>p62</b>	p44	378	40	5,18
<b>XPD</b>	XPA	370	63	6,38	<b>p62</b>	p44	252	126	5,48
<b>XPD</b>	XPB	181	145	5,57	<b>p62</b>	p44	370	190	5,88
<b>XPD</b>	XPC	634	913	5	<b>p62</b>	p44	365	190	6,65
<b>XPC</b>	p52	917	77	7,83	<b>p62</b>	p44	38	190	7,37
<b>XPC</b>	p62	72	38	5,44	<b>p62</b>	p44	504	126	11,68
<b>XPC</b>	p62	77	38	7,11	<b>p62</b>	p44	489	126	13,28
<b>XPC</b>	p62	113	38	7,66	<b>p62</b>	p52	547	552	4,28
<b>XPC</b>	p62	108	38	11,98	<b>p62</b>	XPC	38	72	5,44
<b>XPC</b>	Rad23	477	51	4,42	<b>p62</b>	XPC	114	113	6,1
<b>XPC</b>	Rad23	469	14	14,29	<b>p62</b>	XPD	275	128	6,58
<b>XPC</b>	XPA	853	236	9,79	<b>p62</b>	XPD	293	113	8,84
<b>XPC</b>	XPB	913	529	6,89	<b>p62</b>	XPD	38	128	10,23
<b>XPC</b>	XPD	913	634	5	<b>p62</b>	XPD	353	113	15,13
<b>XPB</b>	p44	549	77	8,37	<b>p52</b>	p62	452	547	4,28
<b>XPB</b>	p44	601	77	10,36	<b>p52</b>	XPA	140	224	7,32
<b>XPB</b>	p52	16	140	9,73	<b>p52</b>	XPB	150	16	4,37
<b>XPB</b>	p52	706	438	11,53	<b>p52</b>	XPB	140	16	9,73
<b>XPB</b>	TTDA	767	31	5,1	<b>p52</b>	XPB	438	706	11,53
<b>XPB</b>	TTDA	526	71	9,25	<b>p52</b>	XPC	77	917	7,83
<b>XPB</b>	XPA	399	157	6,45	<b>p52</b>	XPC	81	917	14,59
<b>XPB</b>	XPA	569	224	7,59	<b>p44</b>	p34	190	198	7,63
<b>XPB</b>	XPC	529	913	6,89	<b>p44</b>	p34	126	149	8,79
<b>XPB</b>	XPD	145	181	5,57	<b>p44</b>	p34	193	298	9,36
<b>XPA</b>	p52	224	140	7,32	<b>p44</b>	p34	388	125	14,61
<b>XPA</b>	XPB	157	399	6,45	<b>p44</b>	p62	40	378	5,18
<b>XPA</b>	XPB	224	569	7,59	<b>p44</b>	p62	23	305	5,3
<b>XPA</b>	XPC	236	853	9,79	<b>p44</b>	p62	126	252	5,48

<b>XPA</b>	XPD	86	370	6,06	<b>p44</b>	p62	190	38	7,37
<b>XPA</b>	XPD	63	370	6,38	<b>p44</b>	XPB	77	549	8,37
<b>TTDA</b>	p52	1	453	4,8	<b>p44</b>	XPD	190	751	6,94
<b>TTDA</b>	p52	2	453	9,8	<b>p34</b>	p44	298	190	7,63
<b>TTDA</b>	XPB	31	767	5,1	<b>p34</b>	p44	149	126	8,79
<b>TTDA</b>	XPB	71	526	9,25	<b>p34</b>	p44	125	388	14,61
<b>Rad23</b>	XPC	36	469	5,26	<b>p34</b>	p62	125	353	13,75
<b>Rad23</b>	XPC	63	469	8,13	<b>Rad23</b>	XPC	34	469	8,37

**Supplemental Table 2 | Core TFIIDH-XPA-XPG crosslinking data.**

<b>P1</b>	<b>P 2</b>	<b>R1</b>	<b>R2</b>	<b>S</b>	<b>P1</b>	<b>P 2</b>	<b>R1</b>	<b>R2</b>	<b>S</b>
<b>XPG</b>	XPD	1120	268	3,42	<b>XPD</b>	p62	128	38	6,69
<b>XPG</b>	XPD	1011	128	3,47	<b>XPD</b>	p62	181	296	7,74
<b>XPG</b>	XPD	154	603	3,71	<b>XPD</b>	p62	113	305	9,64
<b>XPG</b>	XPD	1011	268	3,77	<b>XPD</b>	p44	751	193	4,84
<b>XPG</b>	XPD	433	370	3,78	<b>XPD</b>	p44	751	40	7,15
<b>XPG</b>	XPD	64	370	3,93	<b>XPD</b>	p44	751	190	9,05
<b>XPG</b>	XPD	274	634	4,02	<b>XPD</b>	p44	751	23	11,78
<b>XPG</b>	XPD	498	370	4,25	<b>XPB</b>	XPG	715	1011	3,19
<b>XPG</b>	XPD	120	268	4,28	<b>XPB</b>	XPG	549	1011	3,39
<b>XPG</b>	XPD	274	268	4,64	<b>XPB</b>	XPG	688	1011	3,73
<b>XPG</b>	XPD	163	113	4,72	<b>XPB</b>	XPG	569	1011	4,16
<b>XPG</b>	XPD	1071	370	4,72	<b>XPB</b>	XPG	52	1071	4,20
<b>XPG</b>	XPD	438	268	5,06	<b>XPB</b>	XPG	142	1120	4,97
<b>XPG</b>	XPD	120	113	5,30	<b>XPB</b>	XPG	52	1120	6,21
<b>XPG</b>	XPD	433	268	5,33	<b>XPB</b>	XPG	142	1071	6,53
<b>XPG</b>	XPD	392	370	5,44	<b>XPB</b>	XPG	44	1071	6,92
<b>XPG</b>	XPD	913	128	5,56	<b>XPB</b>	XPG	688	1004	8,04
<b>XPG</b>	XPD	293	370	5,78	<b>XPB</b>	XPG	549	1004	8,41
<b>XPG</b>	XPD	154	128	5,81	<b>XPB</b>	XPG	476	274	8,61
<b>XPG</b>	XPD	274	128	5,88	<b>XPB</b>	XPG	222	1071	10,08
<b>XPG</b>	XPD	172	341	6,13	<b>XPB</b>	XPG	526	1071	11,02
<b>XPG</b>	XPD	126	341	6,13	<b>XPB</b>	XPG	688	1071	18,48
<b>XPG</b>	XPD	126	370	6,15	<b>XPB</b>	XPD	44	89	3,10
<b>XPG</b>	XPD	120	341	6,82	<b>XPB</b>	XPD	569	82	4,83
<b>XPG</b>	XPD	120	370	6,99	<b>XPB</b>	XPD	601	507	8,06
<b>XPG</b>	XPD	172	128	7,08	<b>XPB</b>	XPD	311	113	12,49
<b>XPG</b>	XPD	163	128	7,29	<b>XPB</b>	XPA	715	224	3,04
<b>XPG</b>	XPD	95	181	7,46	<b>XPB</b>	XPA	222	67	3,34
<b>XPG</b>	XPD	433	274	7,69	<b>XPB</b>	XPA	569	236	3,75
<b>XPG</b>	XPD	104	128	7,97	<b>XPB</b>	XPA	217	188	5,14
<b>XPG</b>	XPD	213	128	8,09	<b>XPB</b>	XPA	569	224	6,06
<b>XPG</b>	XPD	120	181	8,51	<b>XPB</b>	XPA	526	236	7,78

SUPPLEMENTARY MATERIALS

<b>XPG</b>	XPD	154	268	9,57	<b>XPB</b>	XPA	399	157	9,96
<b>XPG</b>	XPD	126	128	9,61	<b>XPB</b>	TTDA	44	71	3,01
<b>XPG</b>	XPD	172	268	9,97	<b>XPB</b>	TTDA	526	31	3,15
<b>XPG</b>	XPD	1004	128	10,14	<b>XPB</b>	TTDA	526	71	12,42
<b>XPG</b>	XPD	293	341	10,57	<b>XPB</b>	TTDA	476	71	13,40
<b>XPG</b>	XPD	498	268	10,66	<b>XPB</b>	p52	16	140	4,73
<b>XPG</b>	XPD	95	341	11,67	<b>XPB</b>	p52	706	438	4,97
<b>XPG</b>	XPD	120	128	12,57	<b>XPB</b>	p52	16	150	5,93
<b>XPG</b>	XPD	274	341	13,02	<b>XPB</b>	p52	526	438	8,06
<b>XPG</b>	XPD	95	268	13,53	<b>XPB</b>	p44	142	23	7,03
<b>XPG</b>	XPD	104	341	16,01	<b>XPB</b>	p44	476	306	10,74
<b>XPG</b>	XPD	126	268	17,74	<b>XPB</b>	p44	601	77	12,53
<b>XPG</b>	XPB	1011	715	3,19	<b>XPB</b>	p44	577	77	12,86
<b>XPG</b>	XPB	1011	549	3,39	<b>XPB</b>	p44	549	77	16,85
<b>XPG</b>	XPB	1011	688	3,73	<b>XPA</b>	XPG	236	1011	3,34
<b>XPG</b>	XPB	1011	569	4,16	<b>XPA</b>	XPG	236	1004	7,63
<b>XPG</b>	XPB	1071	52	4,20	<b>XPA</b>	XPD	86	692	3,92
<b>XPG</b>	XPB	1120	142	4,97	<b>XPA</b>	XPD	67	370	4,37
<b>XPG</b>	XPB	1120	52	6,21	<b>XPA</b>	XPD	86	370	5,25
<b>XPG</b>	XPB	1071	142	6,53	<b>XPA</b>	XPD	89	370	6,02
<b>XPG</b>	XPB	1071	44	6,92	<b>XPA</b>	XPD	63	370	6,17
<b>XPG</b>	XPB	1004	688	8,04	<b>XPA</b>	XPD	167	634	6,94
<b>XPG</b>	XPB	1004	549	8,41	<b>XPA</b>	XPB	224	715	3,04
<b>XPG</b>	XPB	274	476	8,61	<b>XPA</b>	XPB	67	222	3,34
<b>XPG</b>	XPB	1071	222	10,08	<b>XPA</b>	XPB	236	569	3,75
<b>XPG</b>	XPB	1071	526	11,02	<b>XPA</b>	XPB	188	217	5,14
<b>XPG</b>	XPB	1071	688	18,48	<b>XPA</b>	XPB	224	569	6,06
<b>XPG</b>	XPA	1011	236	3,34	<b>XPA</b>	XPB	236	526	7,78
<b>XPG</b>	XPA	1004	236	7,63	<b>XPA</b>	XPB	157	399	9,96
<b>XPG</b>	TTDA	1071	31	4,76	<b>XPA</b>	p52	224	140	4,59
<b>XPG</b>	p62	1011	535	3,17	<b>XPA</b>	p52	236	438	6,89
<b>XPG</b>	p62	120	504	3,83	<b>XPA</b>	p52	259	438	25,74
<b>XPG</b>	p62	1004	114	4,02	<b>TTDA</b>	XPG	31	1071	4,76
<b>XPG</b>	p62	952	504	4,58	<b>TTDA</b>	XPB	71	44	3,01
<b>XPG</b>	p62	498	38	5,70	<b>TTDA</b>	XPB	31	526	3,15
<b>XPG</b>	p62	938	38	5,76	<b>TTDA</b>	XPB	71	526	12,42
<b>XPG</b>	p62	754	38	7,49	<b>TTDA</b>	XPB	71	476	13,40
<b>XPG</b>	p62	95	38	9,74	<b>TTDA</b>	p52	6	456	3,94
<b>XPG</b>	p52	1011	81	3,11	<b>TTDA</b>	p52	2	456	7,41
<b>XPG</b>	p52	1071	459	3,22	<b>p62</b>	XPG	535	1011	3,17
<b>XPG</b>	p52	392	459	4,10	<b>p62</b>	XPG	504	120	3,83
<b>XPG</b>	p52	1120	127	4,66	<b>p62</b>	XPG	114	1004	4,02
<b>XPG</b>	p52	1011	438	4,93	<b>p62</b>	XPG	504	952	4,58
<b>XPG</b>	p52	1120	140	5,10	<b>p62</b>	XPG	38	498	5,70

SUPPLEMENTARY MATERIALS

<b>XPG</b>	p52	1071	127	6,27	<b>p62</b>	XPG	38	938	5,76
<b>XPG</b>	p52	952	77	6,74	<b>p62</b>	XPG	38	754	7,49
<b>XPG</b>	p52	1134	127	7,18	<b>p62</b>	XPG	38	95	9,74
<b>XPG</b>	p52	1123	127	8,84	<b>p62</b>	XPD	305	181	4,10
<b>XPG</b>	p52	1004	81	8,87	<b>p62</b>	XPD	240	128	6,47
<b>XPG</b>	p52	1004	77	9,71	<b>p62</b>	XPD	38	128	6,69
<b>XPG</b>	p52	1004	438	9,92	<b>p62</b>	XPD	296	181	7,74
<b>XPG</b>	p52	1134	140	12,87	<b>p62</b>	XPD	305	113	9,64
<b>XPG</b>	p52	1071	77	14,39	<b>p62</b>	p52	38	77	12,85
<b>XPG</b>	p44	95	110	4,24	<b>p62</b>	p44	353	190	3,35
<b>XPG</b>	p44	1011	126	4,27	<b>p62</b>	p44	38	190	7,16
<b>XPG</b>	p44	1004	110	6,68	<b>p62</b>	p34	365	125	4,66
<b>XPG</b>	p44	1004	126	8,01	<b>p62</b>	p34	370	125	7,44
<b>XPG</b>	p34	1011	125	3,44	<b>p62</b>	p34	353	125	17,68
<b>XPG</b>	p34	1120	125	5,24	<b>p52</b>	XPG	81	1011	3,11
<b>XPG</b>	p34	1120	74	5,60	<b>p52</b>	XPG	459	1071	3,22
<b>XPG</b>	p34	1071	125	6,21	<b>p52</b>	XPG	459	392	4,10
<b>XPG</b>	p34	1123	125	9,59	<b>p52</b>	XPG	127	1120	4,66
<b>XPD</b>	XPG	268	1120	3,42	<b>p52</b>	XPG	438	1011	4,93
<b>XPD</b>	XPG	128	1011	3,47	<b>p52</b>	XPG	140	1120	5,10
<b>XPD</b>	XPG	603	154	3,71	<b>p52</b>	XPG	127	1071	6,27
<b>XPD</b>	XPG	268	1011	3,77	<b>p52</b>	XPG	77	952	6,74
<b>XPD</b>	XPG	370	433	3,78	<b>p52</b>	XPG	127	1134	7,18
<b>XPD</b>	XPG	370	64	3,93	<b>p52</b>	XPG	127	1123	8,84
<b>XPD</b>	XPG	634	274	4,02	<b>p52</b>	XPG	81	1004	8,87
<b>XPD</b>	XPG	370	498	4,25	<b>p52</b>	XPG	77	1004	9,71
<b>XPD</b>	XPG	268	120	4,28	<b>p52</b>	XPG	438	1004	9,92
<b>XPD</b>	XPG	268	274	4,64	<b>p52</b>	XPG	140	1134	12,87
<b>XPD</b>	XPG	113	163	4,72	<b>p52</b>	XPG	77	1071	14,39
<b>XPD</b>	XPG	370	1071	4,72	<b>p52</b>	XPB	140	16	4,73
<b>XPD</b>	XPG	268	438	5,06	<b>p52</b>	XPB	438	706	4,97
<b>XPD</b>	XPG	113	120	5,30	<b>p52</b>	XPB	150	16	5,93
<b>XPD</b>	XPG	268	433	5,33	<b>p52</b>	XPB	438	526	8,06
<b>XPD</b>	XPG	370	392	5,44	<b>p52</b>	XPA	140	224	4,59
<b>XPD</b>	XPG	128	913	5,56	<b>p52</b>	XPA	438	236	6,89
<b>XPD</b>	XPG	370	293	5,78	<b>p52</b>	XPA	438	259	25,74
<b>XPD</b>	XPG	128	154	5,81	<b>p52</b>	TTDA	456	6	3,94
<b>XPD</b>	XPG	128	274	5,88	<b>p52</b>	TTDA	456	2	7,41
<b>XPD</b>	XPG	341	172	6,13	<b>p52</b>	p62	77	38	12,85
<b>XPD</b>	XPG	341	126	6,13	<b>p52</b>	p34	140	100	3,04
<b>XPD</b>	XPG	370	126	6,15	<b>p52</b>	p34	127	100	6,56
<b>XPD</b>	XPG	341	120	6,82	<b>p52</b>	p34	140	97	8,67
<b>XPD</b>	XPG	370	120	6,99	<b>p52</b>	p34	127	97	22,96
<b>XPD</b>	XPG	128	172	7,08	<b>p44</b>	XPG	110	95	4,24

SUPPLEMENTARY MATERIALS

<b>XPD</b>	XPG	128	163	7,29	<b>p44</b>	XPG	126	1011	4,27
<b>XPD</b>	XPG	181	95	7,46	<b>p44</b>	XPG	110	1004	6,68
<b>XPD</b>	XPG	274	433	7,69	<b>p44</b>	XPG	126	1004	8,01
<b>XPD</b>	XPG	128	104	7,97	<b>p44</b>	XPD	193	751	4,84
<b>XPD</b>	XPG	128	213	8,09	<b>p44</b>	XPD	40	751	7,15
<b>XPD</b>	XPG	181	120	8,51	<b>p44</b>	XPD	190	751	9,05
<b>XPD</b>	XPG	268	154	9,57	<b>p44</b>	XPD	23	751	11,78
<b>XPD</b>	XPG	128	126	9,61	<b>p44</b>	XPB	23	142	7,03
<b>XPD</b>	XPG	268	172	9,97	<b>p44</b>	XPB	306	476	10,74
<b>XPD</b>	XPG	128	1004	10,14	<b>p44</b>	XPB	77	601	12,53
<b>XPD</b>	XPG	341	293	10,57	<b>p44</b>	XPB	77	577	12,86
<b>XPD</b>	XPG	268	498	10,66	<b>p44</b>	XPB	77	549	16,85
<b>XPD</b>	XPG	341	95	11,67	<b>p44</b>	p62	190	353	3,35
<b>XPD</b>	XPG	128	120	12,57	<b>p44</b>	p62	190	295	4,47
<b>XPD</b>	XPG	341	274	13,02	<b>p44</b>	p62	190	38	7,16
<b>XPD</b>	XPG	268	95	13,53	<b>p44</b>	p62	388	122	14,92
<b>XPD</b>	XPG	341	104	16,01	<b>p34</b>	XPG	125	1011	3,44
<b>XPD</b>	XPG	268	126	17,74	<b>p34</b>	XPG	125	1120	5,24
<b>XPD</b>	XPB	89	44	3,10	<b>p34</b>	XPG	74	1120	5,60
<b>XPD</b>	XPB	82	569	4,83	<b>p34</b>	XPG	125	1071	6,21
<b>XPD</b>	XPB	507	601	8,06	<b>p34</b>	XPG	125	1123	9,59
<b>XPD</b>	XPB	113	311	12,49	<b>p34</b>	p62	125	365	4,66
<b>XPD</b>	XPA	692	86	3,92	<b>p34</b>	p62	125	370	7,44
<b>XPD</b>	XPA	370	67	4,37	<b>p34</b>	p62	125	353	17,68
<b>XPD</b>	XPA	370	86	5,25	<b>p34</b>	p52	100	140	3,04
<b>XPD</b>	XPA	370	89	6,02	<b>p34</b>	p52	100	127	6,56
<b>XPD</b>	XPA	370	63	6,17	<b>p34</b>	p52	97	140	8,67
<b>XPD</b>	XPA	634	167	6,94	<b>p34</b>	p52	97	127	22,96
<b>XPD</b>	p62	181	305	4,10	<b>p34</b>	p44	298	190	4,47
<b>XPD</b>	p62	128	240	6,47	<b>p34</b>	p44	125	388	14,92

Supplemental Table 3 | RNAP-CsB-CsA-DDB1-UVSSA crosslinking data.

<b>P1</b>	<b>P2</b>	<b>R1</b>	<b>R2</b>	<b>S</b>	<b>P1</b>	<b>P2</b>	<b>R1</b>	<b>R2</b>	<b>S</b>
<b>CsA</b>	CsB	85	1327	9,19	<b>Rpb10</b>	Rpb3	37	171	7,10
<b>CsA</b>	CsB	167	650	10,54	<b>Rpb10</b>	Rpb3	47	225	9,94
<b>CsA</b>	CsB	167	804	9,05	<b>Rpb10</b>	Rpb3	110	2	6,94
<b>CsA</b>	CsB	167	1216	6,38	<b>Rpb10</b>	Rpb3	110	81	10,61
<b>CsA</b>	CsB	167	1363	7,07	<b>Rpb10</b>	Rpb3	112	20	5,64
<b>CsA</b>	CsB	167	1487	12,71	<b>Rpb10</b>	UVSSA	1	652	1,62
<b>CsA</b>	CsB	174	1392	14,80	<b>Rpb2</b>	CsB	326	324	7,95
<b>CsA</b>	CsB	212	131	11,49	<b>Rpb2</b>	CsB	409	131	11,19

SUPPLEMENTARY MATERIALS

<b>CsA</b>	CsB	212	462	14,01	<b>Rpb2</b>	CsB	417	181	9,55
<b>CsA</b>	CsB	212	783	14,00	<b>Rpb2</b>	CsB	437	170	10,70
<b>CsA</b>	CsB	212	804	7,22	<b>Rpb2</b>	CsB	504	324	6,27
<b>CsA</b>	CsB	212	1156	9,57	<b>Rpb2</b>	Rpb1	288	796	6,25
<b>CsA</b>	CsB	212	1360	8,37	<b>Rpb2</b>	Rpb1	288	1133	4,15
<b>CsA</b>	CsB	212	1363	17,46	<b>Rpb2</b>	Rpb1	288	1135	4,31
<b>CsA</b>	CsB	212	1489	5,49	<b>Rpb2</b>	Rpb1	326	1219	10,64
<b>CsA</b>	CsB	233	131	4,59	<b>Rpb2</b>	Rpb1	341	149	11,20
<b>CsA</b>	CsB	233	170	12,12	<b>Rpb2</b>	Rpb1	341	213	11,98
<b>CsA</b>	CsB	233	462	6,86	<b>Rpb2</b>	Rpb1	341	1008	4,61
<b>CsA</b>	CsB	233	804	5,04	<b>Rpb2</b>	Rpb1	409	550	5,62
<b>CsA</b>	DDB1	233	191	14,18	<b>Rpb2</b>	Rpb1	498	285	11,91
<b>CsA</b>	DDB1	233	204	4,92	<b>Rpb2</b>	Rpb1	535	331	9,81
<b>CsA</b>	UVSSA	34	129	15,85	<b>Rpb2</b>	Rpb1	535	337	14,53
<b>CsA</b>	UVSSA	34	130	7,01	<b>Rpb2</b>	Rpb1	571	149	13,99
<b>CsA</b>	UVSSA	85	129	7,52	<b>Rpb2</b>	Rpb1	571	346	10,94
<b>CsA</b>	UVSSA	167	339	11,71	<b>Rpb2</b>	Rpb1	571	796	10,10
<b>CsA</b>	UVSSA	167	491	9,88	<b>Rpb2</b>	Rpb1	571	853	7,97
<b>CsA</b>	UVSSA	212	334	10,08	<b>Rpb2</b>	Rpb1	571	1125	17,50
<b>CsA</b>	UVSSA	212	362	13,99	<b>Rpb2</b>	Rpb1	571	1132	10,66
<b>CsA</b>	UVSSA	212	379	9,39	<b>Rpb2</b>	Rpb1	571	1133	13,41
<b>CsA</b>	UVSSA	212	491	15,47	<b>Rpb2</b>	Rpb1	571	1135	11,24
<b>CsA</b>	UVSSA	212	523	6,55	<b>Rpb2</b>	Rpb1	574	1135	4,87
<b>CsA</b>	UVSSA	212	616	6,18	<b>Rpb2</b>	Rpb1	892	213	7,69
<b>CsA</b>	UVSSA	233	491	4,86	<b>Rpb2</b>	Rpb1	892	346	8,23
<b>CsA</b>	UVSSA	292	143	14,79	<b>Rpb2</b>	Rpb1	898	213	7,53
<b>CsA</b>	UVSSA	335	143	20,44	<b>Rpb2</b>	Rpb1	898	279	5,18
<b>CsB</b>	CsA	131	212	11,49	<b>Rpb2</b>	Rpb1	898	285	11,59
<b>CsB</b>	CsA	131	233	4,59	<b>Rpb2</b>	Rpb1	898	331	12,98
<b>CsB</b>	CsA	170	233	12,12	<b>Rpb2</b>	Rpb1	898	337	6,14
<b>CsB</b>	CsA	462	212	14,01	<b>Rpb2</b>	Rpb1	898	346	5,99
<b>CsB</b>	CsA	462	233	6,86	<b>Rpb2</b>	Rpb1	1019	853	10,36
<b>CsB</b>	CsA	650	167	10,54	<b>Rpb2</b>	Rpb1	1220	1	10,21
<b>CsB</b>	CsA	783	212	14,00	<b>Rpb2</b>	Rpb1	1220	32	12,19
<b>CsB</b>	CsA	804	167	9,05	<b>Rpb2</b>	Rpb1	1220	279	6,99
<b>CsB</b>	CsA	804	212	7,22	<b>Rpb2</b>	Rpb1	1220	285	8,66
<b>CsB</b>	CsA	804	233	5,04	<b>Rpb2</b>	Rpb1	1220	417	11,79
<b>CsB</b>	CsA	1156	212	9,57	<b>Rpb2</b>	Rpb1	1220	445	4,73
<b>CsB</b>	CsA	1216	167	6,38	<b>Rpb2</b>	Rpb3	966	166	3,92
<b>CsB</b>	CsA	1327	85	9,19	<b>Rpb2</b>	Rpb3	1019	199	1,51
<b>CsB</b>	CsA	1360	212	8,37	<b>Rpb2</b>	Rpb4	287	137	8,21
<b>CsB</b>	CsA	1363	167	7,07	<b>Rpb2</b>	Rpb4	291	137	17,23
<b>CsB</b>	CsA	1363	212	17,46	<b>Rpb2</b>	Rpb4	535	75	4,47
<b>CsB</b>	CsA	1392	174	14,80	<b>Rpb2</b>	Rpb7	1220	153	9,74

SUPPLEMENTARY MATERIALS

<b>CsB</b>	CsA	1487	167	12,71	<b>Rpb2</b>	Rpb7	1220	163	7,97
<b>CsB</b>	CsA	1489	212	5,49	<b>Rpb2</b>	Rpb9	707	92	10,99
<b>CsB</b>	DDB1	1360	1060	3,57	<b>Rpb2</b>	UVSSA	288	679	11,00
<b>CsB</b>	DDB1	1363	150	3,20	<b>Rpb2</b>	UVSSA	291	679	12,07
<b>CsB</b>	DDB1	1363	1060	3,26	<b>Rpb2</b>	UVSSA	325	662	4,62
<b>CsB</b>	Rpb1	311	139	4,26	<b>Rpb2</b>	UVSSA	326	647	6,98
<b>CsB</b>	Rpb1	448	212	6,56	<b>Rpb2</b>	UVSSA	326	662	15,37
<b>CsB</b>	Rpb1	470	1132	0,87	<b>Rpb2</b>	UVSSA	326	679	11,21
<b>CsB</b>	Rpb11	609	62	4,41	<b>Rpb2</b>	UVSSA	966	616	6,69
<b>CsB</b>	Rpb2	131	409	11,19	<b>Rpb3</b>	CsB	155	292	4,49
<b>CsB</b>	Rpb2	170	437	10,70	<b>Rpb3</b>	CsB	155	316	6,83
<b>CsB</b>	Rpb2	181	417	9,55	<b>Rpb3</b>	Rpb10	155	67	14,09
<b>CsB</b>	Rpb2	324	326	7,95	<b>Rpb3</b>	Rpb11	2	110	6,94
<b>CsB</b>	Rpb2	324	504	6,27	<b>Rpb3</b>	Rpb11	20	112	5,64
<b>CsB</b>	Rpb3	292	155	4,49	<b>Rpb3</b>	Rpb11	81	110	10,61
<b>CsB</b>	Rpb3	316	155	6,83	<b>Rpb3</b>	Rpb11	167	37	5,28
<b>CsB</b>	UVSSA	333	647	4,96	<b>Rpb3</b>	Rpb11	171	37	7,10
<b>CsB</b>	UVSSA	337	647	6,94	<b>Rpb3</b>	Rpb11	225	47	9,94
<b>CsB</b>	UVSSA	448	362	10,58	<b>Rpb3</b>	Rpb2	166	966	3,92
<b>CsB</b>	UVSSA	448	616	7,24	<b>Rpb3</b>	Rpb2	199	1019	1,51
<b>CsB</b>	UVSSA	448	647	6,10	<b>Rpb3</b>	Rpb8	199	20	11,63
<b>CsB</b>	UVSSA	650	334	5,62	<b>Rpb4</b>	DDB1	137	570	6,56
<b>CsB</b>	UVSSA	650	491	6,34	<b>Rpb4</b>	Rpb1	91	1132	5,07
<b>CsB</b>	UVSSA	783	362	5,83	<b>Rpb4</b>	Rpb2	75	535	4,47
<b>CsB</b>	UVSSA	783	491	6,69	<b>Rpb4</b>	Rpb2	137	287	8,21
<b>CsB</b>	UVSSA	804	377	16,58	<b>Rpb4</b>	Rpb2	137	291	17,23
<b>CsB</b>	UVSSA	804	379	16,57	<b>Rpb4</b>	Rpb7	94	155	11,23
<b>CsB</b>	UVSSA	804	491	4,93	<b>Rpb4</b>	UVSSA	75	142	3,32
<b>CsB</b>	UVSSA	804	523	4,83	<b>Rpb5</b>	Rpb1	12	125	3,83
<b>CsB</b>	UVSSA	804	616	5,79	<b>Rpb5</b>	Rpb1	41	1350	10,33
<b>CsB</b>	UVSSA	988	362	8,95	<b>Rpb5</b>	UVSSA	47	584	9,25
<b>CsB</b>	UVSSA	991	362	7,12	<b>Rpb5</b>	UVSSA	47	590	8,92
<b>CsB</b>	UVSSA	1216	362	6,62	<b>Rpb5</b>	UVSSA	47	616	8,87
<b>CsB</b>	UVSSA	1216	491	8,07	<b>Rpb5</b>	UVSSA	85	647	7,77
<b>CsB</b>	UVSSA	1328	129	4,64	<b>Rpb7</b>	Rpb2	153	1220	9,74
<b>DDB1</b>	CsA	194	233	14,18	<b>Rpb7</b>	Rpb2	163	1220	7,97
<b>DDB1</b>	CsA	207	233	4,92	<b>Rpb7</b>	Rpb4	155	94	11,23
<b>DDB1</b>	CsB	153	1363	3,20	<b>Rpb8</b>	Rpb1	21	758	8,09
<b>DDB1</b>	CsB	1063	1360	3,57	<b>Rpb8</b>	Rpb1	146	32	6,23
<b>DDB1</b>	CsB	1063	1363	3,26	<b>Rpb8</b>	Rpb1	146	581	16,29
<b>DDB1</b>	Rpb1	726	213	7,92	<b>Rpb8</b>	Rpb3	20	199	11,63
<b>DDB1</b>	Rpb4	573	137	6,56	<b>Rpb9</b>	Rpb2	92	707	10,99
<b>DDB1</b>	UVSSA	63	414	1,46	<b>Rpb9</b>	UVSSA	92	491	6,16
<b>Rpb1</b>	CsB	139	311	4,26	<b>UVSSA</b>	CsA	129	34	15,85



SUPPLEMENTARY MATERIALS

<b>Rpb1</b>	CsB	212	448	6,56	<b>UVSSA</b>	CsA	129	85	7,52
<b>Rpb1</b>	CsB	1132	470	0,87	<b>UVSSA</b>	CsA	130	34	7,01
<b>Rpb1</b>	DDB1	213	723	7,92	<b>UVSSA</b>	CsA	143	292	14,79
<b>Rpb1</b>	Rpb10	279	1220	6,99	<b>UVSSA</b>	CsA	143	335	20,44
<b>Rpb1</b>	Rpb11	285	498	11,91	<b>UVSSA</b>	CsA	334	212	10,08
<b>Rpb1</b>	Rpb12	285	898	11,59	<b>UVSSA</b>	CsA	339	167	11,71
<b>Rpb1</b>	Rpb13	285	1220	8,66	<b>UVSSA</b>	CsA	362	212	13,99
<b>Rpb1</b>	Rpb14	331	535	9,81	<b>UVSSA</b>	CsA	379	212	9,39
<b>Rpb1</b>	Rpb15	331	898	12,98	<b>UVSSA</b>	CsA	491	167	9,88
<b>Rpb1</b>	Rpb16	337	535	14,53	<b>UVSSA</b>	CsA	491	212	15,47
<b>Rpb1</b>	Rpb17	337	898	6,14	<b>UVSSA</b>	CsA	491	233	4,86
<b>Rpb1</b>	Rpb18	346	571	10,94	<b>UVSSA</b>	CsA	523	212	6,55
<b>Rpb1</b>	Rpb19	346	892	8,23	<b>UVSSA</b>	CsA	616	212	6,18
<b>Rpb1</b>	Rpb2	1	1220	10,21	<b>UVSSA</b>	CsB	129	1328	4,64
<b>Rpb1</b>	Rpb20	346	898	5,99	<b>UVSSA</b>	CsB	334	650	5,62
<b>Rpb1</b>	Rpb21	417	1220	11,79	<b>UVSSA</b>	CsB	362	448	10,58
<b>Rpb1</b>	Rpb22	445	1220	4,73	<b>UVSSA</b>	CsB	362	783	5,83
<b>Rpb1</b>	Rpb23	550	409	5,62	<b>UVSSA</b>	CsB	362	988	8,95
<b>Rpb1</b>	Rpb24	796	288	6,25	<b>UVSSA</b>	CsB	362	991	7,12
<b>Rpb1</b>	Rpb25	796	571	10,10	<b>UVSSA</b>	CsB	362	1216	6,62
<b>Rpb1</b>	Rpb26	853	571	7,97	<b>UVSSA</b>	CsB	377	804	16,58
<b>Rpb1</b>	Rpb27	853	1019	10,36	<b>UVSSA</b>	CsB	379	804	16,57
<b>Rpb1</b>	Rpb28	1008	341	4,61	<b>UVSSA</b>	CsB	491	650	6,34
<b>Rpb1</b>	Rpb29	1125	571	17,50	<b>UVSSA</b>	CsB	491	783	6,69
<b>Rpb1</b>	Rpb3	32	1220	12,19	<b>UVSSA</b>	CsB	491	804	4,93
<b>Rpb1</b>	Rpb30	1132	571	10,66	<b>UVSSA</b>	CsB	491	1216	8,07
<b>Rpb1</b>	Rpb31	1133	288	4,15	<b>UVSSA</b>	CsB	523	804	4,83
<b>Rpb1</b>	Rpb32	1133	571	13,41	<b>UVSSA</b>	CsB	616	448	7,24
<b>Rpb1</b>	Rpb33	1135	288	4,31	<b>UVSSA</b>	CsB	616	804	5,79
<b>Rpb1</b>	Rpb34	1135	571	11,24	<b>UVSSA</b>	CsB	647	333	4,96
<b>Rpb1</b>	Rpb35	1135	574	4,87	<b>UVSSA</b>	CsB	647	337	6,94
<b>Rpb1</b>	Rpb36	1219	326	10,64	<b>UVSSA</b>	CsB	647	448	6,10
<b>Rpb1</b>	Rpb4	149	341	11,20	<b>UVSSA</b>	DDB1	414	60	1,46
<b>Rpb1</b>	Rpb4	1132	91	5,07	<b>UVSSA</b>	Rpb1	523	619	12,87
<b>Rpb1</b>	Rpb5	149	571	13,99	<b>UVSSA</b>	Rpb1	523	976	7,53
<b>Rpb1</b>	Rpb5	125	12	3,83	<b>UVSSA</b>	Rpb1	584	212	6,75
<b>Rpb1</b>	Rpb5	1350	41	10,33	<b>UVSSA</b>	Rpb1	616	212	11,26
<b>Rpb1</b>	Rpb6	213	341	11,98	<b>UVSSA</b>	Rpb1	695	1219	7,42
<b>Rpb1</b>	Rpb7	213	892	7,69	<b>UVSSA</b>	Rpb12	652	1	1,62
<b>Rpb1</b>	Rpb8	213	898	7,53	<b>UVSSA</b>	Rpb2	616	966	6,69
<b>Rpb1</b>	Rpb8	32	146	6,23	<b>UVSSA</b>	Rpb2	647	326	6,98
<b>Rpb1</b>	Rpb8	581	146	16,29	<b>UVSSA</b>	Rpb2	662	325	4,62
<b>Rpb1</b>	Rpb8	758	21	8,09	<b>UVSSA</b>	Rpb2	662	326	15,37
<b>Rpb1</b>	Rpb9	279	898	5,18	<b>UVSSA</b>	Rpb2	679	288	11,00

<b>Rpb1</b>	UVSSA	212	584	6,75	<b>UVSSA</b>	Rpb2	679	291	12,07
<b>Rpb1</b>	UVSSA	212	616	11,26	<b>UVSSA</b>	Rpb2	679	326	11,21
<b>Rpb1</b>	UVSSA	619	523	12,87	<b>UVSSA</b>	Rpb4	142	75	3,32
<b>Rpb1</b>	UVSSA	976	523	7,53	<b>UVSSA</b>	Rpb5	584	47	9,25
<b>Rpb1</b>	UVSSA	1219	695	7,42	<b>UVSSA</b>	Rpb5	590	47	8,92
<b>Rpb10</b>	CsB	62	609	4,41	<b>UVSSA</b>	Rpb5	616	47	8,87
<b>Rpb10</b>	Rpb3	67	155	14,09	<b>UVSSA</b>	Rpb5	647	85	7,77
<b>Rpb10</b>	Rpb3	37	167	5,28	<b>UVSSA</b>	Rpb9	491	92	6,16

Supplemental Table 4 | RNAP-core TFIIH-XPA-XPG crosslinking data.

	<b>P1</b>	<b>P2</b>	<b>R1</b>	<b>R2</b>	<b>S</b>	<b>P1</b>	<b>P2</b>	<b>R1</b>	<b>R2</b>	<b>S</b>
<b>p34</b>	p44	149	126	8,71	<b>Rpb9</b>	Rpb2	27	242	13,43	
<b>p34</b>	p52	97	127	16,41	<b>Rpb9</b>	Rpb2	30	242	4,42	
<b>p34</b>	p52	97	140	8,12	<b>Rpb9</b>	XPB	30	418	6,85	
<b>p34</b>	p62	125	353	10,78	<b>TTDA</b>	p52	2	456	10,05	
<b>p34</b>	p62	125	378	4,39	<b>TTDA</b>	p52	6	456	3,84	
<b>p34</b>	XPG	74	1120	6,67	<b>TTDA</b>	XPA	2	259	13,08	
<b>p34</b>	XPG	125	95	5,58	<b>TTDA</b>	XPA	71	224	8,30	
<b>p34</b>	XPG	152	913	6,61	<b>TTDA</b>	XPB	17	282	3,77	
<b>p44</b>	p34	126	149	8,71	<b>TTDA</b>	XPB	71	59	3,74	
<b>p44</b>	p62	31	504	6,98	<b>TTDA</b>	XPB	71	476	20,57	
<b>p44</b>	p62	110	504	7,05	<b>TTDA</b>	XPB	71	526	15,73	
<b>p44</b>	p62	126	38	6,52	<b>TTDA</b>	XPB	71	609	5,31	
<b>p44</b>	p62	190	38	7,18	<b>TTDA</b>	XPB	71	767	7,44	
<b>p44</b>	XPB	23	142	4,19	<b>XPA</b>	p52	151	438	8,73	
<b>p44</b>	XPB	77	549	18,61	<b>XPA</b>	p52	224	140	6,37	
<b>p44</b>	XPB	77	577	7,30	<b>XPA</b>	p52	224	438	7,07	
<b>p44</b>	XPG	126	95	14,25	<b>XPA</b>	p52	236	438	5,07	
<b>p44</b>	XPG	126	1004	9,66	<b>XPA</b>	p52	259	438	30,91	
<b>p44</b>	XPG	126	1011	5,23	<b>XPA</b>	Rpb1	217	132	4,99	
<b>p44</b>	XPG	190	913	10,43	<b>XPA</b>	Rpb5	135	85	3,92	
<b>p52</b>	p34	127	97	16,41	<b>XPA</b>	TTDA	224	71	8,30	
<b>p52</b>	p34	140	97	8,12	<b>XPA</b>	TTDA	259	2	13,08	
<b>p52</b>	p62	452	547	4,57	<b>XPA</b>	XPB	157	399	10,97	
<b>p52</b>	TTDA	456	2	10,05	<b>XPA</b>	XPB	224	52	4,50	
<b>p52</b>	TTDA	456	6	3,84	<b>XPA</b>	XPB	224	569	6,76	
<b>p52</b>	XPA	140	224	6,37	<b>XPA</b>	XPB	224	715	4,94	
<b>p52</b>	XPA	438	151	8,73	<b>XPA</b>	XPB	236	526	9,07	
<b>p52</b>	XPA	438	224	7,07	<b>XPA</b>	XPD	63	370	6,08	
<b>p52</b>	XPA	438	236	5,07	<b>XPA</b>	XPD	137	370	4,87	
<b>p52</b>	XPA	438	259	30,91	<b>XPA</b>	XPD	167	634	7,88	
<b>p52</b>	XPB	438	706	9,76	<b>XPA</b>	XPG	213	213	3,24	
<b>p52</b>	XPD	81	101	13,74	<b>XPA</b>	XPG	236	1004	6,38	

SUPPLEMENTARY MATERIALS

<b>p52</b>	XPG	77	438	9,66	<b>XPA</b>	XPG	236	1011	6,25
<b>p52</b>	XPG	77	754	12,07	<b>XPB</b>	p44	142	23	4,19
<b>p52</b>	XPG	77	904	5,53	<b>XPB</b>	p44	549	77	18,61
<b>p52</b>	XPG	77	952	5,80	<b>XPB</b>	p44	577	77	7,30
<b>p52</b>	XPG	77	1004	9,04	<b>XPB</b>	p52	706	438	9,76
<b>p52</b>	XPG	77	1011	6,00	<b>XPB</b>	p62	311	255	5,09
<b>p52</b>	XPG	77	1071	9,05	<b>XPB</b>	Rpb1	52	127	5,64
<b>p52</b>	XPG	77	1123	4,73	<b>XPB</b>	Rpb2	282	242	3,39
<b>p52</b>	XPG	127	1123	5,31	<b>XPB</b>	Rpb2	311	144	8,71
<b>p52</b>	XPG	140	1120	7,80	<b>XPB</b>	Rpb2	528	824	6,85
<b>p52</b>	XPG	140	1123	9,02	<b>XPB</b>	Rpb5	311	186	3,39
<b>p52</b>	XPG	140	1134	16,24	<b>XPB</b>	Rpb9	418	30	6,85
<b>p52</b>	XPG	459	392	6,20	<b>XPB</b>	TTDA	59	71	3,74
<b>p62</b>	p34	353	125	10,78	<b>XPB</b>	TTDA	282	17	3,77
<b>p62</b>	p34	378	125	4,39	<b>XPB</b>	TTDA	476	71	20,57
<b>p62</b>	p44	38	126	6,52	<b>XPB</b>	TTDA	526	71	15,73
<b>p62</b>	p44	38	190	7,18	<b>XPB</b>	TTDA	609	71	5,31
<b>p62</b>	p44	504	31	6,98	<b>XPB</b>	TTDA	767	71	7,44
<b>p62</b>	p44	504	110	7,05	<b>XPB</b>	XPA	52	224	4,50
<b>p62</b>	p52	547	452	4,57	<b>XPB</b>	XPA	399	157	10,97
<b>p62</b>	Rpb1	492	1008	5,20	<b>XPB</b>	XPA	526	236	9,07
<b>p62</b>	Rpb2	252	824	4,55	<b>XPB</b>	XPA	569	224	6,76
<b>p62</b>	Rpb6	114	98	3,71	<b>XPB</b>	XPA	715	224	4,94
<b>p62</b>	Rpb7	365	1	5,84	<b>XPB</b>	XPG	44	1120	3,68
<b>p62</b>	XPB	255	311	5,09	<b>XPB</b>	XPG	52	1071	6,75
<b>p62</b>	XPD	38	128	9,65	<b>XPB</b>	XPG	52	1120	8,90
<b>p62</b>	XPD	305	128	4,88	<b>XPB</b>	XPG	52	1123	12,15
<b>p62</b>	XPG	38	95	10,41	<b>XPB</b>	XPG	52	1134	5,09
<b>p62</b>	XPG	38	498	6,01	<b>XPB</b>	XPG	52	1171	5,50
<b>p62</b>	XPG	38	754	8,99	<b>XPB</b>	XPG	142	95	5,37
<b>p62</b>	XPG	114	95	5,24	<b>XPB</b>	XPG	142	913	6,89
<b>p62</b>	XPG	114	120	5,55	<b>XPB</b>	XPG	142	1004	11,89
<b>p62</b>	XPG	114	754	9,90	<b>XPB</b>	XPG	142	1008	18,55
<b>p62</b>	XPG	240	95	4,27	<b>XPB</b>	XPG	142	1011	5,37
<b>Rpb1</b>	p62	1008	492	5,20	<b>XPB</b>	XPG	142	1050	11,16
<b>Rpb1</b>	Rpb2	1	1136	19,61	<b>XPB</b>	XPG	142	1071	13,26
<b>Rpb1</b>	Rpb2	139	333	7,94	<b>XPB</b>	XPG	142	1123	4,71
<b>Rpb1</b>	Rpb2	213	257	10,66	<b>XPB</b>	XPG	142	1134	6,52
<b>Rpb1</b>	Rpb2	331	450	4,06	<b>XPB</b>	XPG	167	1004	12,05
<b>Rpb1</b>	Rpb2	331	451	14,29	<b>XPB</b>	XPG	222	64	7,67
<b>Rpb1</b>	Rpb2	331	813	5,26	<b>XPB</b>	XPG	222	95	9,16
<b>Rpb1</b>	Rpb2	331	814	12,29	<b>XPB</b>	XPG	222	1004	10,78
<b>Rpb1</b>	Rpb2	337	451	10,41	<b>XPB</b>	XPG	311	913	5,61
<b>Rpb1</b>	Rpb2	344	487	7,47	<b>XPB</b>	XPG	326	1004	9,35

SUPPLEMENTARY MATERIALS

<b>Rpb1</b>	Rpb2	346	487	11,61	<b>XPB</b>	XPG	326	1011	5,41
<b>Rpb1</b>	Rpb2	346	808	5,45	<b>XPD</b>	p52	101	81	13,74
<b>Rpb1</b>	Rpb2	346	814	7,31	<b>XPD</b>	p62	128	38	9,65
<b>Rpb1</b>	Rpb2	417	1136	7,33	<b>XPD</b>	p62	128	305	4,88
<b>Rpb1</b>	Rpb2	550	325	5,89	<b>XPD</b>	Rpb1	89	707	5,54
<b>Rpb1</b>	Rpb2	796	204	4,06	<b>XPD</b>	Rpb1	89	708	7,67
<b>Rpb1</b>	Rpb2	796	487	14,31	<b>XPD</b>	Rpb2	689	320	5,02
<b>Rpb1</b>	Rpb2	796	490	4,83	<b>XPD</b>	Rpb2	689	330	3,22
<b>Rpb1</b>	Rpb2	803	242	10,09	<b>XPD</b>	XPA	370	63	6,08
<b>Rpb1</b>	Rpb2	853	487	7,90	<b>XPD</b>	XPA	370	137	4,87
<b>Rpb1</b>	Rpb2	853	935	9,13	<b>XPD</b>	XPA	634	167	7,88
<b>Rpb1</b>	Rpb2	1125	487	17,41	<b>XPD</b>	XPG	35	402	4,86
<b>Rpb1</b>	Rpb2	1132	487	4,11	<b>XPD</b>	XPG	113	163	14,33
<b>Rpb1</b>	Rpb2	1133	487	10,94	<b>XPD</b>	XPG	128	95	13,91
<b>Rpb1</b>	Rpb2	1133	490	4,35	<b>XPD</b>	XPG	128	104	12,30
<b>Rpb1</b>	Rpb2	1135	204	4,50	<b>XPD</b>	XPG	128	108	8,80
<b>Rpb1</b>	Rpb2	1135	487	13,82	<b>XPD</b>	XPG	128	120	12,47
<b>Rpb1</b>	Rpb2	1219	242	11,19	<b>XPD</b>	XPG	128	163	13,51
<b>Rpb1</b>	Rpb2	1234	242	9,46	<b>XPD</b>	XPG	128	172	10,89
<b>Rpb1</b>	Rpb3	581	225	9,69	<b>XPD</b>	XPG	128	213	9,84
<b>Rpb1</b>	Rpb3	697	199	7,69	<b>XPD</b>	XPG	128	754	8,64
<b>Rpb1</b>	Rpb4	1132	91	6,68	<b>XPD</b>	XPG	128	913	10,21
<b>Rpb1</b>	Rpb4	1133	91	4,86	<b>XPD</b>	XPG	128	1004	10,84
<b>Rpb1</b>	Rpb5	125	186	8,04	<b>XPD</b>	XPG	128	1011	5,46
<b>Rpb1</b>	Rpb5	127	186	5,37	<b>XPD</b>	XPG	133	163	8,97
<b>Rpb1</b>	Rpb5	132	186	10,45	<b>XPD</b>	XPG	181	172	6,98
<b>Rpb1</b>	Rpb5	134	186	11,30	<b>XPD</b>	XPG	268	64	20,66
<b>Rpb1</b>	Rpb5	1350	12	4,17	<b>XPD</b>	XPG	268	95	23,69
<b>Rpb1</b>	Rpb5	1350	15	9,38	<b>XPD</b>	XPG	268	120	4,64
<b>Rpb1</b>	Rpb5	1350	41	6,97	<b>XPD</b>	XPG	268	172	8,07
<b>Rpb1</b>	Rpb6	1376	59	4,57	<b>XPD</b>	XPG	268	274	6,86
<b>Rpb1</b>	Rpb6	1479	101	9,22	<b>XPD</b>	XPG	268	293	15,25
<b>Rpb1</b>	Rpb7	1	111	16,95	<b>XPD</b>	XPG	268	333	6,04
<b>Rpb1</b>	Rpb7	1	236	3,05	<b>XPD</b>	XPG	268	1071	20,74
<b>Rpb1</b>	Rpb8	32	146	3,53	<b>XPD</b>	XPG	274	95	12,02
<b>Rpb1</b>	Rpb8	581	146	11,50	<b>XPD</b>	XPG	274	293	5,95
<b>Rpb1</b>	Rpb8	758	21	10,53	<b>XPD</b>	XPG	274	433	9,23
<b>Rpb1</b>	Rpb8	767	21	6,91	<b>XPD</b>	XPG	274	913	7,12
<b>Rpb1</b>	XPA	132	217	4,99	<b>XPD</b>	XPG	274	1011	5,57
<b>Rpb1</b>	XPB	127	52	5,64	<b>XPD</b>	XPG	274	1071	7,91
<b>Rpb1</b>	XPD	707	89	5,54	<b>XPD</b>	XPG	341	95	13,69
<b>Rpb1</b>	XPD	708	89	7,67	<b>XPD</b>	XPG	341	104	12,37
<b>Rpb1</b>	XPG	19	938	3,85	<b>XPD</b>	XPG	341	120	6,47
<b>Rpb10</b>	Rpb2	1	808	9,24	<b>XPD</b>	XPG	341	172	7,54

SUPPLEMENTARY MATERIALS

<b>Rpb10</b>	Rpb2	2	204	6,89	<b>XPD</b>	XPG	341	274	14,22
<b>Rpb10</b>	Rpb2	58	749	17,24	<b>XPD</b>	XPG	341	1011	6,32
<b>Rpb10</b>	Rpb3	1	152	19,43	<b>XPD</b>	XPG	370	95	7,79
<b>Rpb10</b>	Rpb3	67	155	13,60	<b>XPD</b>	XPG	370	120	7,86
<b>Rpb11</b>	Rpb3	37	166	4,60	<b>XPD</b>	XPG	370	293	5,09
<b>Rpb11</b>	Rpb3	37	171	6,37	<b>XPD</b>	XPG	370	392	6,99
<b>Rpb11</b>	Rpb3	37	225	4,02	<b>XPD</b>	XPG	370	433	6,04
<b>Rpb11</b>	Rpb3	47	225	15,55	<b>XPD</b>	XPG	370	913	6,16
<b>Rpb11</b>	Rpb3	52	225	15,57	<b>XPD</b>	XPG	370	1004	4,59
<b>Rpb11</b>	Rpb3	110	81	7,57	<b>XPD</b>	XPG	370	1008	3,99
<b>Rpb11</b>	Rpb3	112	20	8,61	<b>XPD</b>	XPG	370	1011	4,27
<b>Rpb12</b>	XPG	2	1079	3,26	<b>XPD</b>	XPG	370	1120	5,24
<b>Rpb2</b>	p62	824	252	4,55	<b>XPD</b>	XPG	634	274	8,44
<b>Rpb2</b>	Rpb1	204	796	4,06	<b>XPD</b>	XPG	634	293	5,27
<b>Rpb2</b>	Rpb1	204	1135	4,50	<b>XPD</b>	XPG	634	913	9,36
<b>Rpb2</b>	Rpb1	242	803	10,09	<b>XPG</b>	p34	95	125	5,58
<b>Rpb2</b>	Rpb1	242	1219	11,19	<b>XPG</b>	p34	913	152	6,61
<b>Rpb2</b>	Rpb1	242	1234	9,46	<b>XPG</b>	p34	1120	74	6,67
<b>Rpb2</b>	Rpb1	257	213	10,66	<b>XPG</b>	p44	95	126	14,25
<b>Rpb2</b>	Rpb1	325	550	5,89	<b>XPG</b>	p44	913	190	10,43
<b>Rpb2</b>	Rpb1	333	139	7,94	<b>XPG</b>	p44	1004	126	9,66
<b>Rpb2</b>	Rpb1	450	331	4,06	<b>XPG</b>	p44	1011	126	5,23
<b>Rpb2</b>	Rpb1	451	331	14,29	<b>XPG</b>	p52	392	459	6,20
<b>Rpb2</b>	Rpb1	451	337	10,41	<b>XPG</b>	p52	438	77	9,66
<b>Rpb2</b>	Rpb1	487	344	7,47	<b>XPG</b>	p52	754	77	12,07
<b>Rpb2</b>	Rpb1	487	346	11,61	<b>XPG</b>	p52	904	77	5,53
<b>Rpb2</b>	Rpb1	487	796	14,31	<b>XPG</b>	p52	952	77	5,80
<b>Rpb2</b>	Rpb1	487	853	7,90	<b>XPG</b>	p52	1004	77	9,04
<b>Rpb2</b>	Rpb1	487	1125	17,41	<b>XPG</b>	p52	1011	77	6,00
<b>Rpb2</b>	Rpb1	487	1132	4,11	<b>XPG</b>	p52	1071	77	9,05
<b>Rpb2</b>	Rpb1	487	1133	10,94	<b>XPG</b>	p52	1120	140	7,80
<b>Rpb2</b>	Rpb1	487	1135	13,82	<b>XPG</b>	p52	1123	77	4,73
<b>Rpb2</b>	Rpb1	490	796	4,83	<b>XPG</b>	p52	1123	127	5,31
<b>Rpb2</b>	Rpb1	490	1133	4,35	<b>XPG</b>	p52	1123	140	9,02
<b>Rpb2</b>	Rpb1	808	346	5,45	<b>XPG</b>	p52	1134	140	16,24
<b>Rpb2</b>	Rpb1	813	331	5,26	<b>XPG</b>	p62	95	38	10,41
<b>Rpb2</b>	Rpb1	814	331	12,29	<b>XPG</b>	p62	95	114	5,24
<b>Rpb2</b>	Rpb1	814	346	7,31	<b>XPG</b>	p62	95	240	4,27
<b>Rpb2</b>	Rpb1	935	853	9,13	<b>XPG</b>	p62	120	114	5,55
<b>Rpb2</b>	Rpb1	1136	1	19,61	<b>XPG</b>	p62	498	38	6,01
<b>Rpb2</b>	Rpb1	1136	417	7,33	<b>XPG</b>	p62	754	38	8,99
<b>Rpb2</b>	Rpb10	204	2	6,89	<b>XPG</b>	p62	754	114	9,90
<b>Rpb2</b>	Rpb10	749	58	17,24	<b>XPG</b>	Rpb1	938	19	3,85
<b>Rpb2</b>	Rpb10	808	1	9,24	<b>XPG</b>	Rpb12	1079	2	3,26

SUPPLEMENTARY MATERIALS

<b>Rpb2</b>	Rpb3	203	137	12,47	<b>XPG</b>	Rpb2	1004	487	4,50
<b>Rpb2</b>	Rpb3	207	137	23,61	<b>XPG</b>	Rpb2	1175	623	6,98
<b>Rpb2</b>	Rpb3	257	137	19,61	<b>XPG</b>	XPA	213	213	3,24
<b>Rpb2</b>	Rpb3	402	75	3,67	<b>XPG</b>	XPA	1004	236	6,38
<b>Rpb2</b>	Rpb3	808	155	11,02	<b>XPG</b>	XPA	1011	236	6,25
<b>Rpb2</b>	Rpb3	814	155	9,67	<b>XPG</b>	XPB	64	222	7,67
<b>Rpb2</b>	Rpb3	935	171	7,93	<b>XPG</b>	XPB	95	142	5,37
<b>Rpb2</b>	Rpb5	204	15	4,31	<b>XPG</b>	XPB	95	222	9,16
<b>Rpb2</b>	Rpb7	1136	123	10,09	<b>XPG</b>	XPB	913	142	6,89
<b>Rpb2</b>	Rpb7	1136	163	19,48	<b>XPG</b>	XPB	913	311	5,61
<b>Rpb2</b>	Rpb7	1136	236	8,30	<b>XPG</b>	XPB	1004	142	11,89
<b>Rpb2</b>	Rpb8	1136	24	10,97	<b>XPG</b>	XPB	1004	167	12,05
<b>Rpb2</b>	Rpb9	242	27	13,43	<b>XPG</b>	XPB	1004	222	10,78
<b>Rpb2</b>	Rpb9	242	30	4,42	<b>XPG</b>	XPB	1004	326	9,35
<b>Rpb2</b>	XPB	144	311	8,71	<b>XPG</b>	XPB	1008	142	18,55
<b>Rpb2</b>	XPB	242	282	3,39	<b>XPG</b>	XPB	1011	142	5,37
<b>Rpb2</b>	XPB	824	528	6,85	<b>XPG</b>	XPB	1011	326	5,41
<b>Rpb2</b>	XPD	320	689	5,02	<b>XPG</b>	XPB	1050	142	11,16
<b>Rpb2</b>	XPD	330	689	3,22	<b>XPG</b>	XPB	1071	52	6,75
<b>Rpb2</b>	XPG	487	1004	4,50	<b>XPG</b>	XPB	1071	142	13,26
<b>Rpb2</b>	XPG	623	1175	6,98	<b>XPG</b>	XPB	1120	44	3,68
<b>Rpb3</b>	Rpb1	199	697	7,69	<b>XPG</b>	XPB	1120	52	8,90
<b>Rpb3</b>	Rpb1	225	581	9,69	<b>XPG</b>	XPB	1123	52	12,15
<b>Rpb3</b>	Rpb10	152	1	19,43	<b>XPG</b>	XPB	1123	142	4,71
<b>Rpb3</b>	Rpb10	155	67	13,60	<b>XPG</b>	XPB	1134	52	5,09
<b>Rpb3</b>	Rpb11	20	112	8,61	<b>XPG</b>	XPB	1134	142	6,52
<b>Rpb3</b>	Rpb11	81	110	7,57	<b>XPG</b>	XPB	1171	52	5,50
<b>Rpb3</b>	Rpb11	166	37	4,60	<b>XPG</b>	XPD	64	268	20,66
<b>Rpb3</b>	Rpb11	171	37	6,37	<b>XPG</b>	XPD	95	128	13,91
<b>Rpb3</b>	Rpb11	225	37	4,02	<b>XPG</b>	XPD	95	268	23,69
<b>Rpb3</b>	Rpb11	225	47	15,55	<b>XPG</b>	XPD	95	274	12,02
<b>Rpb3</b>	Rpb11	225	52	15,57	<b>XPG</b>	XPD	95	341	13,69
<b>Rpb3</b>	Rpb2	155	808	11,02	<b>XPG</b>	XPD	95	370	7,79
<b>Rpb3</b>	Rpb2	155	814	9,67	<b>XPG</b>	XPD	104	128	12,30
<b>Rpb3</b>	Rpb2	171	935	7,93	<b>XPG</b>	XPD	104	341	12,37
<b>Rpb3</b>	Rpb6	166	50	5,82	<b>XPG</b>	XPD	108	128	8,80
<b>Rpb3</b>	Rpb8	199	20	12,83	<b>XPG</b>	XPD	120	128	12,47
<b>Rpb4</b>	Rpb1	91	1132	6,68	<b>XPG</b>	XPD	120	268	4,64
<b>Rpb4</b>	Rpb1	91	1133	4,86	<b>XPG</b>	XPD	120	341	6,47
<b>Rpb4</b>	Rpb2	24	1136	10,97	<b>XPG</b>	XPD	120	370	7,86
<b>Rpb4</b>	Rpb2	75	402	3,67	<b>XPG</b>	XPD	163	113	14,33
<b>Rpb4</b>	Rpb2	137	203	12,47	<b>XPG</b>	XPD	163	128	13,51
<b>Rpb4</b>	Rpb2	137	207	23,61	<b>XPG</b>	XPD	163	133	8,97
<b>Rpb4</b>	Rpb2	137	257	19,61	<b>XPG</b>	XPD	172	128	10,89

SUPPLEMENTARY MATERIALS

<b>Rpb4</b>	Rpb7	24	153	5,83	<b>XPG</b>	XPD	172	181	6,98
<b>Rpb4</b>	Rpb7	94	155	20,01	<b>XPG</b>	XPD	172	268	8,07
<b>Rpb5</b>	Rpb1	12	1350	4,17	<b>XPG</b>	XPD	172	341	7,54
<b>Rpb5</b>	Rpb1	15	1350	9,38	<b>XPG</b>	XPD	213	128	9,84
<b>Rpb5</b>	Rpb1	41	1350	6,97	<b>XPG</b>	XPD	274	268	6,86
<b>Rpb5</b>	Rpb1	186	125	8,04	<b>XPG</b>	XPD	274	341	14,22
<b>Rpb5</b>	Rpb1	186	127	5,37	<b>XPG</b>	XPD	274	634	8,44
<b>Rpb5</b>	Rpb1	186	132	10,45	<b>XPG</b>	XPD	293	268	15,25
<b>Rpb5</b>	Rpb1	186	134	11,30	<b>XPG</b>	XPD	293	274	5,95
<b>Rpb5</b>	Rpb2	15	204	4,31	<b>XPG</b>	XPD	293	370	5,09
<b>Rpb5</b>	Rpb6	164	50	6,38	<b>XPG</b>	XPD	293	634	5,27
<b>Rpb5</b>	XPA	85	135	3,92	<b>XPG</b>	XPD	333	268	6,04
<b>Rpb5</b>	XPB	186	311	3,39	<b>XPG</b>	XPD	392	370	6,99
<b>Rpb6</b>	p62	98	114	3,71	<b>XPG</b>	XPD	402	35	4,86
<b>Rpb6</b>	Rpb1	59	1376	4,57	<b>XPG</b>	XPD	433	274	9,23
<b>Rpb6</b>	Rpb1	101	1479	9,22	<b>XPG</b>	XPD	433	370	6,04
<b>Rpb6</b>	Rpb3	50	166	5,82	<b>XPG</b>	XPD	754	128	8,64
<b>Rpb6</b>	Rpb5	50	164	6,38	<b>XPG</b>	XPD	913	128	10,21
<b>Rpb7</b>	p62	1	365	5,84	<b>XPG</b>	XPD	913	274	7,12
<b>Rpb7</b>	Rpb1	111	1	16,95	<b>XPG</b>	XPD	913	370	6,16
<b>Rpb7</b>	Rpb1	236	1	3,05	<b>XPG</b>	XPD	913	634	9,36
<b>Rpb7</b>	Rpb2	123	1136	10,09	<b>XPG</b>	XPD	1004	128	10,84
<b>Rpb7</b>	Rpb2	163	1136	19,48	<b>XPG</b>	XPD	1004	370	4,59
<b>Rpb7</b>	Rpb2	236	1136	8,30	<b>XPG</b>	XPD	1008	370	3,99
<b>Rpb7</b>	Rpb4	153	24	5,83	<b>XPG</b>	XPD	1011	128	5,46
<b>Rpb7</b>	Rpb4	155	94	20,01	<b>XPG</b>	XPD	1011	274	5,57
<b>Rpb8</b>	Rpb1	21	758	10,53	<b>XPG</b>	XPD	1011	341	6,32
<b>Rpb8</b>	Rpb1	21	767	6,91	<b>XPG</b>	XPD	1011	370	4,27
<b>Rpb8</b>	Rpb1	146	32	3,53	<b>XPG</b>	XPD	1071	268	20,74
<b>Rpb8</b>	Rpb1	146	581	11,50	<b>XPG</b>	XPD	1071	274	7,91
<b>Rpb8</b>	Rpb3	20	199	12,83	<b>XPG</b>	XPD	1120	370	5,24

Supplemental Table 5 | Cryo-EM data collection, refinement and validation statistics.

<b>TFIIH-XPA-DNA</b>	
<b>Data collection and processing</b>	
Magnification	130 000
Voltage (kV)	300
Electron exposure (e-/Å <sup>2</sup> )	41
Defocus range (µm)	0.6-4.3
Pixel size (Å/pix)	1.05
Symmetry imposed	C1
Initial particle images (no.)	1354997
Final particle images (no.)	227776
Map resolution (Å)	3.6
FSC threshold	0.143
Map resolution range (Å)	3.1-7.7
<b>Refinement</b>	
Initial model used (PDB code)	Ab-initio in CryoSPARC
Model resolution (Å)	3.4
FSC threshold	0.5
Map sharpening <i>B</i> factor (Å <sup>2</sup> )	denoised map with flattened amplitude spectrum
<b>Model composition</b>	
Non-hydrogen atoms	19804
Protein residues	2349
Ligands	SF4, 6xZn
<b><i>B</i> factors (Å<sup>2</sup>)</b>	
Protein	96.81
Ligand	202.97
<b>R.m.s. deviations</b>	
Bond lengths (Å)	0.017
Bond angles (°)	1.722
<b>Validation</b>	
MolProbity score	2.23
Clashscore	11.94
Poor rotamers (%)	1.5
<b>Ramachandran plot</b>	
Favored (%)	91.95
Allowed (%)	7.93
Disallowed (%)	0.12



## 6. REFERENCES

- 1 Lindahl, T. & Barnes, D. E. Repair of endogenous DNA damage. *Cold Spring Harbor symposia on quantitative biology* **65**, 127-133, (2000).
- 2 Sancar, A., Lindsey-Boltz, L. A., Ünsal-Kaçmaz, K. & Linn, S. Molecular Mechanisms of Mammalian DNA Repair and the DNA Damage Checkpoints. *Annual review of biochemistry* **73**, 39-85, (2004).
- 3 Wood, R. D. Fifty years since DNA repair was linked to cancer. *Nature* **557**, 648-649, (2018).
- 4 Jeggo, P. A., Pearl, L. H. & Carr, A. M. DNA repair, genome stability and cancer: a historical perspective. *Nature Reviews Cancer* **16**, 35, (2015).
- 5 Torgovnick, A. & Schumacher, B. DNA repair mechanisms in cancer development and therapy. *Frontiers in genetics* **6**, 157-157, (2015).
- 6 Yi, C. & He, C. DNA repair by reversal of DNA damage. *Cold Spring Harbor perspectives in biology* **5**, a012575-a012575.
- 7 Liu, Z., Wang, L. & Zhong, D. Dynamics and mechanisms of DNA repair by photolyase. *Physical chemistry chemical physics : PCCP* **17**, 11933-11949, (2015).
- 8 Warren, J. J., Forsberg, L. J. & Beese, L. S. The structural basis for the mutagenicity of O6-methyl-guanine lesions. *Proceedings of the National Academy of Sciences* **103**, 19701, (2006).
- 9 Lindahl, T., Sedgwick, B., Sekiguchi, M. & Nakabeppu, Y. Regulation and expression of the adaptive response to alkylating agents. *Annual review of biochemistry* **57**, 133-157, (1988).
- 10 Kunkel, T. A. & Erie, D. A. DNA MISMATCH REPAIR. *Annual review of biochemistry* **74**, 681-710, (2005).
- 11 Li, G.-M. Mechanisms and functions of DNA mismatch repair. *Cell Research* **18**, 85, (2007).
- 12 Wion, D. & Casadesus, J. N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nature reviews. Microbiology* **4**, 183-192, (2006).
- 13 Obmolova, G., Ban, C., Hsieh, P. & Yang, W. Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature* **407**, 703-710, (2000).
- 14 Ban, C. & Yang, W. Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis. *Cell* **95**, 541-552, (1998).
- 15 Junop, M. S., Obmolova, G., Rausch, K., Hsieh, P. & Yang, W. Composite active site of an ABC ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair. *Molecular cell* **7**, 1-12, (2001).
- 16 Krokan, H. E. & Bjørås, M. Base excision repair. *Cold Spring Harbor perspectives in biology* **5**, a012583-a012583.
- 17 Jacobs, A. L. & Schar, P. DNA glycosylases: in DNA repair and beyond. *Chromosoma* **121**, 1-20, (2012).
- 18 Kavli, B., Otterlei, M., Slupphaug, G. & Krokan, H. E. Uracil in DNA--general mutagen, but normal intermediate in acquired immunity. *DNA Repair (Amst)* **6**, 505-516, (2007).
- 19 Schärer, O. D. Nucleotide excision repair in eukaryotes. *Cold Spring Harbor perspectives in biology* **5**, a012609-a012609.
- 20 Marteijn, J. A., Lans, H., Vermeulen, W. & Hoeijmakers, J. H. J. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nature Reviews Molecular Cell Biology* **15**, 465, (2014).

- 21 Geacintov, N. E. *et al.* Thermodynamic and structural factors in the removal of bulky DNA adducts by the nucleotide excision repair machinery. *Biopolymers* **65**, 202-210, (2002).
- 22 Huang, J. C. & Sancar, A. Determination of minimum substrate size for human excinuclease. *The Journal of biological chemistry* **269**, 19034-19040, (1994).
- 23 Dip, R., Camenisch, U. & Naegeli, H. Mechanisms of DNA damage recognition and strand discrimination in human nucleotide excision repair. *DNA Repair (Amst)* **3**, 1409-1423, (2004).
- 24 Sugasawa, K. *et al.* A multistep damage recognition mechanism for global genomic nucleotide excision repair. *Genes & development* **15**, 507-521, (2001).
- 25 Mocquet, V. *et al.* The human DNA repair factor XPC-HR23B distinguishes stereoisomeric benzo[a]pyrenyl-DNA lesions. *The EMBO journal* **26**, 2923-2932, (2007).
- 26 Min, J. H. & Pavletich, N. P. Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* **449**, 570-575, (2007).
- 27 Kisker, C., Kuper, J. & Van Houten, B. Prokaryotic nucleotide excision repair. *Cold Spring Harbor perspectives in biology* **5**, a012591-a012591, (2013).
- 28 Chapman, J. R., Taylor, Martin R. G. & Boulton, Simon J. Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. *Molecular cell* **47**, 497-510, (2012).
- 29 Li, J. & Xu, X. DNA double-strand break repair: a tale of pathway choices. *Acta biochimica et biophysica Sinica* **48**, 641-646, (2016).
- 30 Wyman, C. & Kanaar, R. DNA Double-Strand Break Repair: All's Well that Ends Well. *Annual Review of Genetics* **40**, 363-383, (2006).
- 31 Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual review of biochemistry* **79**, 181-211, (2010).
- 32 Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nature reviews. Molecular cell biology* **18**, 495-506, (2017).
- 33 Ranjha, L., Howard, S. M. & Cejka, P. Main steps in DNA double-strand break repair: an introduction to homologous recombination and related processes. *Chromosoma* **127**, 187-214, (2018).
- 34 Jasin, M. & Rothstein, R. Repair of strand breaks by homologous recombination. *Cold Spring Harbor perspectives in biology* **5**, a012740-a012740.
- 35 Boyce, R. P. & Howard-Flanders, P. RELEASE OF ULTRAVIOLET LIGHT-INDUCED THYMINE DIMERS FROM DNA IN E. COLI K-12. *Proceedings of the National Academy of Sciences of the United States of America* **51**, 293-300, (1964).
- 36 Pettijohn, D. & Hanawalt, P. EVIDENCE FOR REPAIR-REPLICATION OF ULTRAVIOLET DAMAGED DNA IN BACTERIA. *Journal of molecular biology* **9**, 395-410, (1964).
- 37 Rasmussen, R. E. & Painter, R. B. EVIDENCE FOR REPAIR OF ULTRA-VIOLET DAMAGED DEOXYRIBONUCLEIC ACID IN CULTURED MAMMALIAN CELLS. *Nature* **203**, 1360-1362, (1964).
- 38 Setlow, R. B. & Carrier, W. L. THE DISAPPEARANCE OF THYMINE DIMERS FROM DNA: AN ERROR-CORRECTING MECHANISM. *Proceedings of the National Academy of Sciences of the United States of America* **51**, 226-231, (1964).
- 39 Cleaver, J. E. Defective repair replication of DNA in xeroderma pigmentosum. *Nature* **218**, 652-656, (1968).

- 40 DiGiovanna, J. J. & Kraemer, K. H. Shining a light on xeroderma pigmentosum. *The Journal of investigative dermatology* **132**, 785-796, (2012).
- 41 De Weerd-Kastelein, E. A., Keijzer, W. & Bootsma, D. Genetic heterogeneity of xeroderma pigmentosum demonstrated by somatic cell hybridization. *Nature: New biology* **238**, 80-83, (1972).
- 42 Thompson, L. H. Chinese hamster cells meet DNA repair: an entirely acceptable affair. *BioEssays : news and reviews in molecular, cellular and developmental biology* **20**, 589-597, (1998).
- 43 Araújo, S. J. *et al.* Nucleotide excision repair of DNA with recombinant human proteins: definition of the minimal set of factors, active forms of TFIIH, and modulation by CAK. *Genes & development* **14**, 349-359, (2000).
- 44 Aboussekhra, A. *et al.* Mammalian DNA nucleotide excision repair reconstituted with purified protein components. *Cell* **80**, 859-868, (1995).
- 45 Mu, D. *et al.* Reconstitution of human DNA repair excision nuclease in a highly defined system. *The Journal of biological chemistry* **270**, 2415-2418, (1995).
- 46 Sugawara, K. *et al.* Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Molecular cell* **2**, 223-232, (1998).
- 47 Chen, X. *et al.* Kinetic gating mechanism of DNA damage recognition by Rad4/XPC. *Nature communications* **6**, 5849-5849, (2015).
- 48 Trego, K. S. & Turchi, J. J. Pre-steady-state binding of damaged DNA by XPC-hHR23B reveals a kinetic mechanism for damage discrimination. *Biochemistry* **45**, 1961-1969, (2006).
- 49 Donahue, B. A., Yin, S., Taylor, J. S., Reines, D. & Hanawalt, P. C. Transcript cleavage by RNA polymerase II arrested by a cyclobutane pyrimidine dimer in the DNA template. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 8502-8506, (1994).
- 50 Selby, C. P. & Sancar, A. Human transcription-repair coupling factor CSB/ERCC6 is a DNA-stimulated ATPase but is not a helicase and does not disrupt the ternary transcription complex of stalled RNA polymerase II. *The Journal of biological chemistry* **272**, 1885-1890, (1997).
- 51 Brueckner, F., Hennecke, U., Carell, T. & Cramer, P. CPD damage recognition by transcribing RNA polymerase II. *Science (New York, N.Y.)* **315**, 859-862, (2007).
- 52 Buschta-Hedayat, N., Buterin, T., Hess, M. T., Missura, M. & Naegeli, H. Recognition of nonhybridizing base pairs during nucleotide excision repair of DNA. *Proceedings of the National Academy of Sciences* **96**, 6090, (1999).
- 53 Hess, M. T., Schwitter, U., Petretta, M., Giese, B. & Naegeli, H. Bipartite substrate discrimination by human nucleotide excision repair. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 6664-6669, (1997).
- 54 Masutani, C. *et al.* Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast RAD23. *The EMBO journal* **13**, 1831-1843, (1994).
- 55 Nishi, R. *et al.* Centrin 2 stimulates nucleotide excision repair by interacting with xeroderma pigmentosum group C protein. *Mol Cell Biol* **25**, 5664-5674, (2005).
- 56 Sugawara, K., Akagi, J., Nishi, R., Iwai, S. & Hanaoka, F. Two-step recognition of DNA damage for mammalian nucleotide excision repair: Directional binding of the XPC complex and DNA strand scanning. *Molecular cell* **36**, 642-653, (2009).
- 57 Chu, G. & Chang, E. Xeroderma pigmentosum group E cells lack a nuclear factor that binds to damaged DNA. *Science (New York, N.Y.)* **242**, 564-567, (1988).
- 58 Reardon, J. T. & Sancar, A. Recognition and repair of the cyclobutane thymine dimer, a major cause of skin cancers, by the human excision nuclease. *Genes & development* **17**, 2539-2551, (2003).

- 59 Wakasugi, M. *et al.* DDB accumulates at DNA damage sites immediately after UV irradiation and directly stimulates nucleotide excision repair. *The Journal of biological chemistry* **277**, 1637-1640, (2002).
- 60 Scrima, A. *et al.* Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex. *Cell* **135**, 1213-1223, (2008).
- 61 Volker, M. *et al.* Sequential assembly of the nucleotide excision repair factors in vivo. *Molecular cell* **8**, 213-224, (2001).
- 62 Yokoi, M. *et al.* The xeroderma pigmentosum group C protein complex XPC-HR23B plays an important role in the recruitment of transcription factor IIH to damaged DNA. *The Journal of biological chemistry* **275**, 9870-9875, (2000).
- 63 Riedl, T., Hanaoka, F. & Egly, J. M. The comings and goings of nucleotide excision repair factors on damaged DNA. *The EMBO journal* **22**, 5293-5303, (2003).
- 64 Compe, E. & Egly, J. M. Nucleotide Excision Repair and Transcriptional Regulation: TFIIH and Beyond. *Annual review of biochemistry* **85**, 265-290, (2016).
- 65 Evans, E., Moggs, J. G., Hwang, J. R., Egly, J. M. & Wood, R. D. Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. *The EMBO journal* **16**, 6559-6573, (1997).
- 66 Fuss, J. O. & Tainer, J. A. XPB and XPD helicases in TFIIH orchestrate DNA duplex opening and damage verification to coordinate repair with transcription and cell cycle via CAK kinase. *DNA repair* **10**, 697-713, (2011).
- 67 Fan, L. *et al.* Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair. *Molecular cell* **22**, 27-37, (2006).
- 68 Coin, F., Oksenyshyn, 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* **26**, 245-256, (2007).
- 69 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 of the United States of America* **112**, 3961-3966, (2015).
- 70 Schilbach, S. *et al.* Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature* **551**, 204, (2017).
- 71 Kuper, J., Wolski, S. C., Michels, G. & Kisker, C. Functional and structural studies of the nucleotide excision repair helicase XPD suggest a polarity for DNA translocation. *The EMBO journal* **31**, 494-502, (2012).
- 72 Abdulrahman, W. *et al.* ARCH domain of XPD, an anchoring platform for CAK that conditions TFIIH DNA repair and transcription activities. *Proceedings of the National Academy of Sciences of the United States of America* **110**, E633-642, (2013).
- 73 Li, C. L. *et al.* Tripartite DNA Lesion Recognition and Verification by XPC, TFIIH, and XPA in Nucleotide Excision Repair. *Molecular cell* **59**, 1025-1034, (2015).
- 74 Sontz, P. A., Mui, T. P., Fuss, J. O., Tainer, J. A. & Barton, J. K. DNA charge transport as a first step in coordinating the detection of lesions by repair proteins. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 1856-1861, (2012).
- 75 Mathieu, N., Kaczmarek, N. & Naegeli, H. Strand- and site-specific DNA lesion demarcation by the xeroderma pigmentosum group D helicase. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 17545-17550, (2010).
- 76 Fan, L. *et al.* XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. *Cell* **133**, 789-800, (2008).

- 77 Lehmann, A. R. XPD structure reveals its secrets. *DNA Repair (Amst)* **7**, 1912-1915, (2008).
- 78 Liu, H. *et al.* Structure of the DNA repair helicase XPD. *Cell* **133**, 801-812, (2008).
- 79 Lukianova, O. A. & David, S. S. A role for iron-sulfur clusters in DNA repair. *Current opinion in chemical biology* **9**, 145-151, (2005).
- 80 Gregersen, L. H. & Svejstrup, J. Q. The Cellular Response to Transcription-Blocking DNA Damage. *Trends in biochemical sciences* **43**, 327-341, (2018).
- 81 Camenisch, U., Dip, R., Schumacher, S. B., Schuler, B. & Naegeli, H. Recognition of helical kinks by xeroderma pigmentosum group A protein triggers DNA excision repair. *Nature structural & molecular biology* **13**, 278-284, (2006).
- 82 Yang, Z. *et al.* Specific and efficient binding of xeroderma pigmentosum complementation group A to double-strand/single-strand DNA junctions with 3'-and/or 5'-ssDNA branches. *Biochemistry* **45**, 15921-15930, (2006).
- 83 Sugitani, N., Sivley, R. M., Perry, K. E., Capra, J. A. & Chazin, W. J. XPA: A key scaffold for human nucleotide excision repair. *DNA repair* **44**, 123-135, (2016).
- 84 Coin, F. *et al.* Nucleotide Excision Repair Driven by the Dissociation of CAK from TFIIH. *Molecular cell* **31**, 9-20, (2008).
- 85 Bunick, C. G., Miller, M. R., Fuller, B. E., Fanning, E. & Chazin, W. J. Biochemical and structural domain analysis of xeroderma pigmentosum complementation group C protein. *Biochemistry* **45**, 14965-14979, (2006).
- 86 Giglia-Mari, G. *et al.* Dynamic interaction of TTDA with TFIIH is stabilized by nucleotide excision repair in living cells. *PLoS biology* **4**, e156, (2006).
- 87 Ziani, S. *et al.* Sequential and ordered assembly of a large DNA repair complex on undamaged chromatin. *J Cell Biol* **206**, 589-598, (2014).
- 88 Buchko, G. W. *et al.* Interactions of human nucleotide excision repair protein XPA with DNA and RPA70 Delta C327: chemical shift mapping and 15N NMR relaxation studies. *Biochemistry* **38**, 15116-15128, (1999).
- 89 Mer, G. *et al.* Structural basis for the recognition of DNA repair proteins UNG2, XPA, and RAD52 by replication factor RPA. *Cell* **103**, 449-456, (2000).
- 90 Lee, S. H., Kim, D. K. & Drissi, R. Human xeroderma pigmentosum group A protein interacts with human replication protein A and inhibits DNA replication. *The Journal of biological chemistry* **270**, 21800-21805, (1995).
- 91 Daughdrill, G. W. *et al.* Chemical shift changes provide evidence for overlapping single-stranded DNA- and XPA-binding sites on the 70 kDa subunit of human replication protein A. *Nucleic acids research* **31**, 4176-4183, (2003).
- 92 Ikegami, T. *et al.* Solution structure of the DNA- and RPA-binding domain of the human repair factor XPA. *Nature structural biology* **5**, 701-706, (1998).
- 93 Li, L., Lu, X., Peterson, C. A. & Legerski, R. J. An interaction between the DNA repair factor XPA and replication protein A appears essential for nucleotide excision repair. *Mol Cell Biol* **15**, 5396-5402, (1995).
- 94 Saijo, M., Takedachi, A. & Tanaka, K. Nucleotide excision repair by mutant xeroderma pigmentosum group A (XPA) proteins with deficiency in interaction with RPA. *The Journal of biological chemistry* **286**, 5476-5483, (2011).
- 95 Li, L., Elledge, S. J., Peterson, C. A., Bales, E. S. & Legerski, R. J. Specific association between the human DNA repair proteins XPA and ERCC1. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 5012-5016, (1994).
- 96 Tsodikov, O. V. *et al.* Structural basis for the recruitment of ERCC1-XPF to nucleotide excision repair complexes by XPA. *The EMBO journal* **26**, 4768-4776, (2007).

- 97 Wakasugi, M. *et al.* Physical and functional interaction between DDB and XPA in nucleotide excision repair. *Nucleic acids research* **37**, 516-525, (2009).
- 98 Gilljam, K. M., Muller, R., Liabakk, N. B. & Otterlei, M. Nucleotide excision repair is associated with the replisome and its efficiency depends on a direct interaction between XPA and PCNA. *PloS one* **7**, e49199, (2012).
- 99 Krasikova, Y. S., Rechkunova, N. I., Maltseva, E. A., Petruseva, I. O. & Lavrik, O. I. Localization of xeroderma pigmentosum group A protein and replication protein A on damaged DNA in nucleotide excision repair. *Nucleic acids research* **38**, 8083-8094, (2010).
- 100 Koch, S. C. *et al.* Structural insights into the recognition of cisplatin and AAF-dG lesion by Rad14 (XPA). *Proceedings of the National Academy of Sciences of the United States of America* **112**, 8272-8277, (2015).
- 101 Krasikova, Y. S. *et al.* Interaction of nucleotide excision repair factors XPC-HR23B, XPA, and RPA with damaged DNA. *Biochemistry. Biokhimiia* **73**, 886-896, (2008).
- 102 Liu, Y. *et al.* Cooperative interaction of human XPA stabilizes and enhances specific binding of XPA to DNA damage. *Biochemistry* **44**, 7361-7368, (2005).
- 103 Chen, R. & Wold, M. S. Replication protein A: single-stranded DNA's first responder: dynamic DNA-interactions allow replication protein A to direct single-strand DNA intermediates into different pathways for synthesis or repair. *BioEssays : news and reviews in molecular, cellular and developmental biology* **36**, 1156-1161, (2014).
- 104 Kim, C., Paulus, B. F. & Wold, M. S. Interactions of human replication protein A with oligonucleotides. *Biochemistry* **33**, 14197-14206, (1994).
- 105 de Laat, W. L. *et al.* DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair. *Genes & development* **12**, 2598-2609, (1998).
- 106 Spivak, G. Nucleotide excision repair in humans. *DNA Repair (Amst)* **36**, 13-18, (2015).
- 107 Hohl, M., Thorel, F., Clarkson, S. G. & Scharer, O. D. Structural determinants for substrate binding and catalysis by the structure-specific endonuclease XPG. *The Journal of biological chemistry* **278**, 19500-19508, (2003).
- 108 Schärer, O. D. XPG: its products and biological roles. *Advances in experimental medicine and biology* **637**, 83-92, (2008).
- 109 Enzlin, J. H. & Schärer, O. D. The active site of the DNA repair endonuclease XPF-ERCC1 forms a highly conserved nuclease motif. *The EMBO journal* **21**, 2045-2053, (2002).
- 110 de Laat, W. L., Appeldoorn, E., Jaspers, N. G. & Hoeijmakers, J. H. DNA structural elements required for ERCC1-XPF endonuclease activity. *The Journal of biological chemistry* **273**, 7835-7842, (1998).
- 111 O'Donovan, A., Davies, A. A., Moggs, J. G., West, S. C. & Wood, R. D. XPG endonuclease makes the 3' incision in human DNA nucleotide excision repair. *Nature* **371**, 432-435, (1994).
- 112 Ito, S. *et al.* XPG Stabilizes TFIIH, Allowing Transactivation of Nuclear Receptors: Implications for Cockayne Syndrome in XP-G/CS Patients. *Molecular cell* **26**, 231-243, (2007).
- 113 Mu, D., Wakasugi, M., Hsu, D. S. & Sancar, A. Characterization of reaction intermediates of human excision repair nuclease. *The Journal of biological chemistry* **272**, 28971-28979, (1997).
- 114 Staresinic, L. *et al.* Coordination of dual incision and repair synthesis in human nucleotide excision repair. *The EMBO journal* **28**, 1111-1120, (2009).

- 115 Mocquet, V. *et al.* Sequential recruitment of the repair factors during NER: the role of XPG in initiating the resynthesis step. *The EMBO journal* **27**, 155-167, (2008).
- 116 Ogi, T. *et al.* Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells. *Molecular cell* **37**, 714-727, (2010).
- 117 Moser, J. *et al.* Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase III alpha in a cell-cycle-specific manner. *Molecular cell* **27**, 311-323, (2007).
- 118 Nudler, E. RNA polymerase active center: the molecular engine of transcription. *Annual review of biochemistry* **78**, 335-361, (2009).
- 119 Archambault, J. & Friesen, J. D. Genetics of eukaryotic RNA polymerases I, II, and III. *Microbiological reviews* **57**, 703-724, (1993).
- 120 Vermeulen, W. & Fousteri, M. Mammalian transcription-coupled excision repair. *Cold Spring Harb Perspect Biol* **5**, a012625, (2013).
- 121 Svejstrup, J. Q. The RNA polymerase II transcription cycle: cycling through chromatin. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1677**, 64-73, (2004).
- 122 Hantsche, M. & Cramer, P. The Structural Basis of Transcription: 10 Years After the Nobel Prize in Chemistry. *Angewandte Chemie (International ed. in English)* **55**, 15972-15981, (2016).
- 123 Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J. M. & Cramer, P. Structure of transcribing mammalian RNA polymerase II. *Nature* **529**, 551-554, (2016).
- 124 Mellon, I., Spivak, G. & Hanawalt, P. C. Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell* **51**, 241-249, (1987).
- 125 Bohr, V. A., Smith, C. A., Okumoto, D. S. & Hanawalt, P. C. DNA repair in an active gene: removal of pyrimidine dimers from the DHFR gene of CHO cells is much more efficient than in the genome overall. *Cell* **40**, 359-369, (1985).
- 126 Sweder, K. S. & Hanawalt, P. C. Preferential repair of cyclobutane pyrimidine dimers in the transcribed strand of a gene in yeast chromosomes and plasmids is dependent on transcription. *Proceedings of the National Academy of Sciences* **89**, 10696, (1992).
- 127 Troelstra, C. *et al.* ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. *Cell* **71**, 939-953, (1992).
- 128 van den Boom, V. *et al.* DNA damage stabilizes interaction of CSB with the transcription elongation machinery. *J Cell Biol* **166**, 27-36, (2004).
- 129 Henning, K. A. *et al.* The Cockayne syndrome group A gene encodes a WD repeat protein that interacts with CSB protein and a subunit of RNA polymerase II TFIIH. *Cell* **82**, 555-564, (1995).
- 130 Zhang, X. *et al.* Mutations in UVSSA cause UV-sensitive syndrome and destabilize ERCC6 in transcription-coupled DNA repair. *Nature genetics* **44**, 593-597, (2012).
- 131 Nakazawa, Y. *et al.* Mutations in UVSSA cause UV-sensitive syndrome and impair RNA polymerase IIo processing in transcription-coupled nucleotide-excision repair. *Nature genetics* **44**, 586-592, (2012).
- 132 Schwertman, P. *et al.* UV-sensitive syndrome protein UVSSA recruits USP7 to regulate transcription-coupled repair. *Nature genetics* **44**, 598-602, (2012).
- 133 Fairman-Williams, M. E., Guenther, U.-P. & Jankowsky, E. SF1 and SF2 helicases: family matters. *Current opinion in structural biology* **20**, 313-324, (2010).
- 134 Lake, R. J. & Fan, H.-Y. Structure, function and regulation of CSB: a multi-talented gymnast. *Mechanisms of ageing and development* **134**, 202-211, (2013).

- 135 Wang, L. *et al.* Regulation of the Rhp26ERCC6/CSB chromatin remodeler by a novel conserved leucine latch motif. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 18566-18571, (2014).
- 136 Anindya, R. *et al.* A ubiquitin-binding domain in Cockayne syndrome B required for transcription-coupled nucleotide excision repair. *Molecular cell* **38**, 637-648, (2010).
- 137 Selby, C. P. & Sancar, A. Cockayne syndrome group B protein enhances elongation by RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 11205-11209, (1997).
- 138 Xu, J. *et al.* Structural basis for the initiation of eukaryotic transcription-coupled DNA repair. *Nature* **551**, 653-657, (2017).
- 139 Foustieri, M., Vermeulen, W., van Zeeland, A. A. & Mullenders, L. H. F. Cockayne Syndrome A and B Proteins Differentially Regulate Recruitment of Chromatin Remodeling and Repair Factors to Stalled RNA Polymerase II In Vivo. *Molecular cell* **23**, 471-482, (2006).
- 140 Citterio, E. *et al.* ATP-dependent chromatin remodeling by the Cockayne syndrome B DNA repair-transcription-coupling factor. *Mol Cell Biol* **20**, 7643-7653, (2000).
- 141 Cho, I., Tsai, P.-F., Lake, R. J., Basheer, A. & Fan, H.-Y. ATP-dependent chromatin remodeling by Cockayne syndrome protein B and NAP1-like histone chaperones is required for efficient transcription-coupled DNA repair. *PLoS genetics* **9**, e1003407-e1003407, (2013).
- 142 Velez-Cruz, R., Zadorin, A. S., Coin, F. & Egly, J. M. Sirt1 suppresses RNA synthesis after UV irradiation in combined xeroderma pigmentosum group D/Cockayne syndrome (XP-D/CS) cells. *Proceedings of the National Academy of Sciences of the United States of America* **110**, E212-220, (2013).
- 143 Nicolai, S. *et al.* Identification of Novel Proteins Co-Purifying with Cockayne Syndrome Group B (CSB) Reveals Potential Roles for CSB in RNA Metabolism and Chromatin Dynamics. *PloS one* **10**, e0128558, (2015).
- 144 Spivak, G. The many faces of Cockayne syndrome. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 15273-15274, (2004).
- 145 Fischer, Eric S. *et al.* The Molecular Basis of CRL4DDB2/CSA Ubiquitin Ligase Architecture, Targeting, and Activation. *Cell* **147**, 1024-1039, (2011).
- 146 Saijo, M. The role of Cockayne syndrome group A (CSA) protein in transcription-coupled nucleotide excision repair. *Mech Ageing Dev* **134**, 196-201, (2013).
- 147 Groisman, R. *et al.* CSA-dependent degradation of CSB by the ubiquitin-proteasome pathway establishes a link between complementation factors of the Cockayne syndrome. *Genes & development* **20**, 1429-1434, (2006).
- 148 Anindya, R., Aygün, O. & Svejstrup, J. Q. Damage-Induced Ubiquitylation of Human RNA Polymerase II by the Ubiquitin Ligase Nedd4, but Not Cockayne Syndrome Proteins or BRCA1. *Molecular cell* **28**, 386-397, (2007).
- 149 Schwertman, P., Vermeulen, W. & Marteijn, J. A. UVSSA and USP7, a new couple in transcription-coupled DNA repair. *Chromosoma* **122**, 275-284, (2013).
- 150 Okuda, M., Nakazawa, Y., Guo, C., Ogi, T. & Nishimura, Y. Common TFIIH recruitment mechanism in global genome and transcription-coupled repair subpathways. *Nucleic acids research* **45**, 13043-13055, (2017).
- 151 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 (London, England : 1993)* **23**, 1827-1837, (2015).
- 152 Svejstrup, J. Q. Mechanisms of transcription-coupled DNA repair. *Nature Reviews Molecular Cell Biology* **3**, 21, (2002).



- 153 Hanawalt, P. C. & Spivak, G. Transcription-coupled DNA repair: two decades of progress and surprises. *Nature Reviews Molecular Cell Biology* **9**, 958, (2008).
- 154 Kamarthapu, V. & Nudler, E. Rethinking transcription coupled DNA repair. *Current opinion in microbiology* **24**, 15-20, (2015).
- 155 Cheung, A. C. & Cramer, P. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* **471**, 249-253, (2011).
- 156 Reines, D. Elongation factor-dependent transcript shortening by template-engaged RNA polymerase II. *The Journal of biological chemistry* **267**, 3795-3800, (1992).
- 157 Izban, M. G. & Luse, D. S. The increment of SII-facilitated transcript cleavage varies dramatically between elongation competent and incompetent RNA polymerase II ternary complexes. *The Journal of biological chemistry* **268**, 12874-12885, (1993).
- 158 Wang, D. & Hawley, D. K. Identification of a 3'-->5' exonuclease activity associated with human RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 843-847, (1993).
- 159 Jeon, C., Yoon, H. & Agarwal, K. The transcription factor TFIIS zinc ribbon dipeptide Asp-Glu is critical for stimulation of elongation and RNA cleavage by RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 9106-9110, (1994).
- 160 Wilson, M. D., Harreman, M. & Svejstrup, J. Q. Ubiquitylation and degradation of elongating RNA polymerase II: the last resort. *Biochim Biophys Acta* **1829**, 151-157, (2013).
- 161 Steurer, B. & Marteijn, J. A. Traveling Rocky Roads: The Consequences of Transcription-Blocking DNA Lesions on RNA Polymerase II. *Journal of molecular biology* **429**, 3146-3155, (2017).
- 162 Huibregtse, J. M., Yang, J. C. & Beaudenon, S. L. The large subunit of RNA polymerase II is a substrate of the Rsp5 ubiquitin-protein ligase. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 3656-3661, (1997).
- 163 Harreman, M. *et al.* Distinct ubiquitin ligases act sequentially for RNA polymerase II polyubiquitylation. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 20705-20710, (2009).
- 164 Ganesan, A., Spivak, G. & Hanawalt, P. C. in *Progress in Molecular Biology and Translational Science* Vol. 110 (ed Paul W. Doetsch) 25-40 (Academic Press, 2012).
- 165 Epshtein, V. *et al.* UvrD facilitates DNA repair by pulling RNA polymerase backwards. *Nature* **505**, 372-377, (2014).
- 166 Gerard, M. *et al.* Purification and interaction properties of the human RNA polymerase B(II) general transcription factor BTF2. *The Journal of biological chemistry* **266**, 20940-20945, (1991).
- 167 Takagi, Y. *et al.* Ubiquitin ligase activity of TFIIF and the transcriptional response to DNA damage. *Molecular cell* **18**, 237-243, (2005).
- 168 Graham, J. M., Jr. *et al.* Cerebro-oculo-facio-skeletal syndrome with a nucleotide excision-repair defect and a mutated XPD gene, with prenatal diagnosis in a triplet pregnancy. *American journal of human genetics* **69**, 291-300, (2001).
- 169 Scharer, O. D. Hot topics in DNA repair: the molecular basis for different disease states caused by mutations in TFIIF and XPG. *DNA Repair (Amst)* **7**, 339-344, (2008).
- 170 Fisher, R. P. Secrets of a double agent: CDK7 in cell-cycle control and transcription. *Journal of cell science* **118**, 5171-5180, (2005).
- 171 Ito, S. *et al.* MMXD, a TFIIF-independent XPD-MMS19 protein complex involved in chromosome segregation. *Molecular cell* **39**, 632-640, (2010).

- 172 Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation  
by RNA polymerase II. *Nature Reviews Molecular Cell Biology* **16**, 129, (2015).
- 173 Douziech, M. *et al.* Mechanism of promoter melting by the xeroderma pigmentosum  
complementation group B helicase of transcription factor IIIH revealed by protein-  
DNA photo-cross-linking. *Mol Cell Biol* **20**, 8168-8177, (2000).
- 174 Kim, T. K., Ebright, R. H. & Reinberg, D. Mechanism of ATP-dependent promoter  
melting by transcription factor IIIH. *Science (New York, N.Y.)* **288**, 1418-1422,  
(2000).
- 175 Lin, Y. C., Choi, W. S. & Gralla, J. D. TFIIH XPB mutants suggest a unified  
bacterial-like mechanism for promoter opening but not escape. *Nature structural &  
molecular biology* **12**, 603-607, (2005).
- 176 Lu, H., Zawel, L., Fisher, L., Egly, J. M. & Reinberg, D. Human general transcription  
factor IIIH phosphorylates the C-terminal domain of RNA polymerase II. *Nature* **358**,  
641-645, (1992).
- 177 Phatnani, H. P. & Greenleaf, A. L. Phosphorylation and functions of the RNA  
polymerase II CTD. *Genes & development* **20**, 2922-2936, (2006).
- 178 Glover-Cutter, K. *et al.* TFIIH-associated Cdk7 kinase functions in phosphorylation  
of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by  
RNA polymerase II. *Mol Cell Biol* **29**, 5455-5464, (2009).
- 179 Cho, E. J., Takagi, T., Moore, C. R. & Buratowski, S. mRNA capping enzyme is  
recruited to the transcription complex by phosphorylation of the RNA polymerase II  
carboxy-terminal domain. *Genes & development* **11**, 3319-3326, (1997).
- 180 Egloff, S. *et al.* Serine-7 of the RNA polymerase II CTD is specifically required for  
snRNA gene expression. *Science (New York, N.Y.)* **318**, 1777-1779, (2007).
- 181 Winkler, G. S. *et al.* TFIIH with inactive XPD helicase functions in transcription  
initiation but is defective in DNA repair. *The Journal of biological chemistry* **275**,  
4258-4266, (2000).
- 182 Greber, B. J. *et al.* The cryo-electron microscopy structure of human transcription  
factor IIIH. *Nature* **549**, 414-417, (2017).
- 183 Schultz, P. *et al.* Molecular Structure of Human TFIIH. *Cell* **102**, 599-607, (2000).
- 184 Gibbons, B. J. *et al.* Subunit architecture of general transcription factor TFIIH.  
*Proceedings of the National Academy of Sciences of the United States of America*  
**109**, 1949-1954, (2012).
- 185 Chang, W.-H. & Kornberg, R. D. Electron Crystal Structure of the Transcription  
Factor and DNA Repair Complex, Core TFIIH. *Cell* **102**, 609-613, (2000).
- 186 Luo, J. *et al.* Architecture of the Human and Yeast General Transcription and DNA  
Repair Factor TFIIH. *Molecular cell* **59**, 794-806, (2015).
- 187 Kainov, D. E. *et al.* Structure determination of the minimal complex between Tfb5  
and Tfb2, two subunits of the yeast transcription/DNA-repair factor TFIIH: a  
retrospective study. *Acta crystallographica. Section D, Biological crystallography*  
**66**, 745-755, (2010).
- 188 Gervais, V. *et al.* TFIIH contains a PH domain involved in DNA nucleotide excision  
repair. *Nature Structural & Molecular Biology* **11**, 616, (2004).
- 189 Radu, L. *et al.* The intricate network between the p34 and p44 subunits is central to  
the activity of the transcription/DNA repair factor TFIIH. *Nucleic acids research* **45**,  
10872-10883, (2017).
- 190 Gradia, S. D. *et al.* MacroBac: New Technologies for Robust and Efficient Large-  
Scale Production of Recombinant Multiprotein Complexes. *Methods in enzymology*  
**592**, 1-26, (2017).

- 191 Araki, M. *et al.* Centrosome protein centrin 2/caltractin 1 is part of the xeroderma pigmentosum group C complex that initiates global genome nucleotide excision repair. *The Journal of biological chemistry* **276**, 18665-18672, (2001).
- 192 Sugasawa, K. *et al.* HHR23B, a human Rad23 homolog, stimulates XPC protein in nucleotide excision repair in vitro. *Molecular and cellular biology* **16**, 4852-4861, (1996).
- 193 Bernardes de Jesus, B. M., Bjoras, M., Coin, F. & Egly, J. M. Dissection of the molecular defects caused by pathogenic mutations in the DNA repair factor XPC. *Mol Cell Biol* **28**, 7225-7235, (2008).
- 194 Lafrance-Vanasse, J., Arseneault, G., Cappadocia, L., Legault, P. & Omichinski, J. G. Structural and functional evidence that Rad4 competes with Rad2 for binding to the Tfb1 subunit of TFIIH in NER. *Nucleic acids research* **41**, 2736-2745, (2013).
- 195 Hey, T., Lipps, G. & Krauss, G. Binding of XPA and RPA to damaged DNA investigated by fluorescence anisotropy. *Biochemistry* **40**, 2901-2910, (2001).
- 196 Sugitani, N., Shell, S. M., Soss, S. E. & Chazin, W. J. Redefining the DNA-binding domain of human XPA. *Journal of the American Chemical Society* **136**, 10830-10833, (2014).
- 197 Kastner, B. *et al.* GraFix: sample preparation for single-particle electron cryomicroscopy. *Nature methods* **5**, 53-55, (2008).
- 198 He, Y. *et al.* Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533**, 359-365, (2016).
- 199 Mu, H., Geacintov, N. E., Broyde, S., Yeo, J.-E. & Schärer, O. D. Molecular basis for damage recognition and verification by XPC-RAD23B and TFIIH in nucleotide excision repair. *DNA Repair* **71**, 33-42, (2018).
- 200 Titov, D. V. *et al.* XPB, a subunit of TFIIH, is a target of the natural product triptolide. *Nature Chemical Biology* **7**, 182, (2011).
- 201 He, Q. L. *et al.* Covalent modification of a cysteine residue in the XPB subunit of the general transcription factor TFIIH through single epoxide cleavage of the transcription inhibitor triptolide. *Angewandte Chemie (International ed. in English)* **54**, 1859-1863, (2015).
- 202 Kiiianitsa, K., Solinger, J. A. & Heyer, W. D. NADH-coupled microplate photometric assay for kinetic studies of ATP-hydrolyzing enzymes with low and high specific activities. *Analytical biochemistry* **321**, 266-271, (2003).
- 203 Firman, K. & Szczelkun, M. D. Measuring motion on DNA by the type I restriction endonuclease EcoR124I using triplex displacement. *The EMBO journal* **19**, 2094-2102, (2000).
- 204 Kumaran, S., Kozlov, A. G. & Lohman, T. M. *Saccharomyces cerevisiae* replication protein A binds to single-stranded DNA in multiple salt-dependent modes. *Biochemistry* **45**, 11958-11973, (2006).
- 205 Tegunov, D. & Cramer, P. Real-time cryo-EM data pre-processing with Warp. *bioRxiv*, (2018).
- 206 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature methods* **14**, 290-296, (2017).
- 207 Zivanov, J. *et al.* RELION-3: new tools for automated high-resolution cryo-EM structure determination. *bioRxiv*, (2018).
- 208 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* **66**, 213-221, (2010).
- 209 Ziani, S. *et al.* Sequential and ordered assembly of a large DNA repair complex on undamaged chromatin. *The Journal of cell biology* **206**, 589-598, (2014).

- 210 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25**, 1605-1612, (2004).
- 211 Tomko, E. J., Fishburn, J., Hahn, S. & Galburt, E. A. TFIIH generates a six-base-pair open complex during RNAP II transcription initiation and start-site scanning. *Nature structural & molecular biology* **24**, 1139-1145, (2017).
- 212 Kawaoka, J., Jankowsky, E. & Pyle, A. M. Backbone tracking by the SF2 helicase NPH-II. *Nature structural & molecular biology* **11**, 526-530, (2004).
- 213 Pyle, A. M. Translocation and unwinding mechanisms of RNA and DNA helicases. *Annual review of biophysics* **37**, 317-336, (2008).
- 214 Lehmann, A. R. The xeroderma pigmentosum group D (XPD) gene: one gene, two functions, three diseases. *Genes & development* **15**, 15-23, (2001).
- 215 Mathieu, N., Kaczmarek, N., Ruthemann, P., Luch, A. & Naegeli, H. DNA quality control by a lesion sensor pocket of the xeroderma pigmentosum group D helicase subunit of TFIIH. *Current biology : CB* **23**, 204-212, (2013).
- 216 Fuss, J. O., Tsai, C.-L., Ishida, J. P. & Tainer, J. A. Emerging critical roles of Fe-S clusters in DNA replication and repair. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1853**, 1253-1271, (2015).
- 217 Paul, V. D. & Lill, R. Biogenesis of cytosolic and nuclear iron-sulfur proteins and their role in genome stability. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1853**, 1528-1539, (2015).
- 218 Brooks, S. C., Adhikary, S., Rubinson, E. H. & Eichman, B. F. Recent advances in the structural mechanisms of DNA glycosylases. *Biochimica et biophysica acta* **1834**, 247-271, (2013).
- 219 Leclerc, V., Raisin, S. & Léopold, P. Dominant-negative mutants reveal a role for the Cdk7 kinase at the mid-blastula transition in Drosophila embryos. *The EMBO journal* **19**, 1567-1575, (2000).
- 220 Feaver, W. J., Gileadi, O., Li, Y. & Kornberg, R. D. CTD kinase associated with yeast RNA polymerase II initiation factor b. *Cell* **67**, 1223-1230, (1991).
- 221 Svejstrup, J. Q., Feaver, W. J., LaPointe, J. & Kornberg, R. D. RNA polymerase transcription factor IIH holoenzyme from yeast. *The Journal of biological chemistry* **269**, 28044-28048, (1994).
- 222 Kosinski, J. *et al.* Xlink Analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *Journal of structural biology* **189**, 177-183, (2015).
- 223 Constantinescu-Aruxandei, D., Petrovic-Stojanovska, B., Penedo, J. C., White, M. F. & Naismith, J. H. Mechanism of DNA loading by the DNA repair helicase XPD. *Nucleic acids research* **44**, 2806-2815, (2016).
- 224 Evans, E., Fellows, J., Coffey, A. & Wood, R. D. Open complex formation around a lesion during nucleotide excision repair provides a structure for cleavage by human XPG protein. *The EMBO journal* **16**, 625-638, (1997).
- 225 Sinha, R. P. & Hader, D. P. UV-induced DNA damage and repair: a review. *Photochemical & photobiological sciences : Official journal of the European Photochemistry Association and the European Society for Photobiology* **1**, 225-236, (2002).
- 226 de Lima-Bessa, K. M. *et al.* CPDs and 6-4PPs play different roles in UV-induced cell death in normal and NER-deficient human cells. *DNA Repair (Amst)* **7**, 303-312, (2008).
- 227 You, Y. H., Szabo, P. E. & Pfeifer, G. P. Cyclobutane pyrimidine dimers form preferentially at the major p53 mutational hotspot in UVB-induced mouse skin tumors. *Carcinogenesis* **21**, 2113-2117, (2000).

- 228 Cadet, J., Sage, E. & Douki, T. Ultraviolet radiation-mediated damage to cellular DNA. *Mutation research* **571**, 3-17, (2005).
- 229 Ravanat, J. L., Douki, T. & Cadet, J. Direct and indirect effects of UV radiation on DNA and its components. *Journal of photochemistry and photobiology. B, Biology* **63**, 88-102, (2001).
- 230 Mouret, S. *et al.* Cyclobutane pyrimidine dimers are predominant DNA lesions in whole human skin exposed to UVA radiation. *Proceedings of the National Academy of Sciences* **103**, 13765, (2006).
- 231 Blackburn, G. M. & Davies, R. J. H. The structure of DNA-derived thymine dimer. *Biochemical and Biophysical Research Communications* **22**, 704-706, (1966).
- 232 Kneuttinger, A. C. *et al.* Formation and Direct Repair of UV-induced Dimeric DNA Pyrimidine Lesions. *Photochemistry and photobiology* **90**, 1-14, (2014).
- 233 Selby, C. P., Drapkin, R., Reinberg, D. & Sancar, A. RNA polymerase II stalled at a thymine dimer: footprint and effect on excision repair. *Nucleic acids research* **25**, 787-793, (1997).
- 234 Tornaletti, S., Reines, D. & Hanawalt, P. C. Structural characterization of RNA polymerase II complexes arrested by a cyclobutane pyrimidine dimer in the transcribed strand of template DNA. *The Journal of biological chemistry* **274**, 24124-24130, (1999).
- 235 Shaewitz, J. W., Abbondanzieri, E. A., Landick, R. & Block, S. M. Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature* **426**, 684-687, (2003).
- 236 Galburt, E. A. *et al.* Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* **446**, 820-823, (2007).
- 237 Fei, J. & Chen, J. KIAA1530 protein is recruited by Cockayne syndrome complementation group protein A (CSA) to participate in transcription-coupled repair (TCR). *The Journal of biological chemistry* **287**, 35118-35126, (2012).
- 238 Groisman, R. *et al.* CSA-dependent degradation of CSB by the ubiquitin-proteasome pathway establishes a link between complementation factors of the Cockayne syndrome. *Genes & development* **20**, 1429-1434, (2006).
- 239 Wang, L. *et al.* Regulation of the Rhp26/ERCC6/CSB chromatin remodeler by a novel conserved leucine latch motif. *Proceedings of the National Academy of Sciences* **111**, 18566, (2014).
- 240 Kettenberger, H., Armache, K.-J. & Cramer, P. Architecture of the RNA Polymerase II-TFIIS Complex and Implications for mRNA Cleavage. *Cell* **114**, 347-357, (2003).
- 241 Conaway, R. C., Kong, S. E. & Conaway, J. W. TFIIS and GreB: Two Like-Minded Transcription Elongation Factors with Sticky Fingers. *Cell* **114**, 272-274, (2003).
- 242 Harper, J. W. & Elledge, S. J. The DNA damage response: ten years after. *Molecular cell* **28**, 739-745, (2007).
- 243 Jiang, G. & Sancar, A. Recruitment of DNA damage checkpoint proteins to damage in transcribed and nontranscribed sequences. *Mol Cell Biol* **26**, 39-49, (2006).
- 244 Ljungman, M. & Zhang, F. Blockage of RNA polymerase as a possible trigger for u.v. light-induced apoptosis. *Oncogene* **13**, 823-831, (1996).
- 245 Ljungman, M. Activation of DNA damage signaling. *Mutation research* **577**, 203-216, (2005).
- 246 Shin, J. H., Xu, L. & Wang, D. Mechanism of transcription-coupled DNA modification recognition. *Cell & bioscience* **7**, 9-9, (2017).
- 247 Hawryluk, P. J., Ujvári, A. & Luse, D. S. Characterization of a novel RNA polymerase II arrest site which lacks a weak 3' RNA-DNA hybrid. *Nucleic acids research* **32**, 1904-1916, (2004).

- 248 Sidorenkov, I., Komissarova, N. & Kashlev, M. Crucial Role of the RNA:DNA Hybrid in the Processivity of Transcription. *Molecular cell* **2**, 55-64, (1998).
- 249 Nudler, E., Mustaev, A., Goldfarb, A. & Lukhtanov, E. The RNA–DNA Hybrid Maintains the Register of Transcription by Preventing Backtracking of RNA Polymerase. *Cell* **89**, 33-41, (1997).
- 250 Theil, A. F., Hoeijmakers, J. H. & Vermeulen, W. TTDA: big impact of a small protein. *Experimental cell research* **329**, 61-68, (2014).
- 251 Theil, A. F. *et al.* Disruption of TTDA results in complete nucleotide excision repair deficiency and embryonic lethality. *PLoS genetics* **9**, e1003431, (2013).
- 252 Coin, F. *et al.* p8/TTD-A as a Repair-Specific TFIIH Subunit. *Molecular cell* **21**, 215-226, (2006).
- 253 Naegeli, H., Bardwell, L. & Friedberg, E. C. The DNA helicase and adenosine triphosphatase activities of yeast Rad3 protein are inhibited by DNA damage. A potential mechanism for damage-specific recognition. *The Journal of biological chemistry* **267**, 392-398, (1992).
- 254 Naegeli, H., Bardwell, L. & Friedberg, E. C. Inhibition of Rad3 DNA helicase activity by DNA adducts and abasic sites: implications for the role of a DNA helicase in damage-specific incision of DNA. *Biochemistry* **32**, 613-621, (1993).
- 255 Mathieu, N., Kaczmarek, N. & Naegeli, H. Strand- and site-specific DNA lesion demarcation by the xeroderma pigmentosum group D helicase. *Proceedings of the National Academy of Sciences* **107**, 17545, (2010).
- 256 Dubaele, S. *et al.* Basal transcription defect discriminates between xeroderma pigmentosum and trichothiodystrophy in XPD patients. *Molecular cell* **11**, 1635-1646, (2003).
- 257 Cleaver, J. E., Thompson, L. H., Richardson, A. S. & States, J. C. A summary of mutations in the UV-sensitive disorders: xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy. *Human mutation* **14**, 9-22, (1999).
- 258 Kralund, H. H. *et al.* Xeroderma Pigmentosum-Trichothiodystrophy overlap patient with novel XPD/ERCC2 mutation. *Rare diseases (Austin, Tex.)* **1**, e24932-e24932, (2013).
- 259 Coin, F. *et al.* Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH. *Nature genetics* **20**, 184-188, (1998).
- 260 Seroz, T., Perez, C., Bergmann, E., Bradsher, J. & Egly, J. M. p44/SSL1, the regulatory subunit of the XPD/RAD3 helicase, plays a crucial role in the transcriptional activity of TFIIH. *The Journal of biological chemistry* **275**, 33260-33266, (2000).
- 261 Botta, E. *et al.* Analysis of Mutations in the XPD Gene in Italian Patients with Trichothiodystrophy: Site of Mutation Correlates with Repair Deficiency, but Gene Dosage Appears to Determine Clinical Severity. *The American Journal of Human Genetics* **63**, 1036-1048, (1998).
- 262 Rudolf, J., Makrantonis, V., Ingledew, W. J., Stark, M. J. R. & White, M. F. The DNA Repair Helicases XPD and FancJ Have Essential Iron-Sulfur Domains. *Molecular cell* **23**, 801-808, (2006).
- 263 Wakasugi, M. & Sancar, A. Order of assembly of human DNA repair excision nuclease. *The Journal of biological chemistry* **274**, 18759-18768, (1999).
- 264 Jones, C. J. & Wood, R. D. Preferential binding of the xeroderma pigmentosum group A complementing protein to damaged DNA. *Biochemistry* **32**, 12096-12104, (1993).
- 265 Asahina, H. *et al.* The XPA protein is a zinc metalloprotein with an ability to recognize various kinds of DNA damage. *Mutation research* **315**, 229-237, (1994).

- 266 Robins, P., Jones, C. J., Biggerstaff, M., Lindahl, T. & Wood, R. D. Complementation of DNA repair in xeroderma pigmentosum group A cell extracts by a protein with affinity for damaged DNA. *The EMBO journal* **10**, 3913-3921, (1991).
- 267 He, Z., Henriksen, L. A., Wold, M. S. & Ingles, C. J. RPA involvement in the damage-recognition and incision steps of nucleotide excision repair. *Nature* **374**, 566-569, (1995).
- 268 Batty, D., Ropic'-Otrin, V., Levine, A. S. & Wood, R. D. Stable binding of human XPC complex to irradiated DNA confers strong discrimination for damaged sites. *Journal of molecular biology* **300**, 275-290, (2000).
- 269 Dienemann, C., Schwalb, B., Schilbach, S. & Cramer, P. Promoter Distortion and Opening in the RNA Polymerase II Cleft. *Molecular cell* **73**, 97-106.e104, (2019).
- 270 Lisica, A. *et al.* Mechanisms of backtrack recovery by RNA polymerases I and II. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 2946-2951, (2016).
- 271 Schutz, P. *et al.* Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 9564-9569, (2008).
- 272 Montpetit, B. *et al.* A conserved mechanism of DEAD-box ATPase activation by nucleoporins and InsP6 in mRNA export. *Nature* **472**, 238-242, (2011).
- 273 Selby, C. P. & Sancar, A. Structure and function of transcription-repair coupling factor. II. Catalytic properties. *The Journal of biological chemistry* **270**, 4890-4895, (1995).
- 274 Selby, C. P. & Sancar, A. Structure and function of transcription-repair coupling factor. I. Structural domains and binding properties. *The Journal of biological chemistry* **270**, 4882-4889, (1995).
- 275 Adebali, O., Chiou, Y.-Y., Hu, J., Sancar, A. & Selby, C. P. Genome-wide transcription-coupled repair in Escherichia coli is mediated by the Mfd translocase. *Proceedings of the National Academy of Sciences*, 201700230, (2017).
- 276 Grishkevich, V. & Yanai, I. Gene length and expression level shape genomic novelties. *Genome research* **24**, 1497-1503, (2014).
- 277 Lee, S.-K., Yu, S.-L., Prakash, L. & Prakash, S. Requirement of Yeast RAD2, a Homolog of Human XPG Gene, for Efficient RNA Polymerase II Transcription: Implications for Cockayne Syndrome. *Cell* **109**, 823-834, (2002).
- 278 Sarker, A. H. *et al.* Recognition of RNA polymerase II and transcription bubbles by XPG, CSB, and TFIIH: insights for transcription-coupled repair and Cockayne Syndrome. *Molecular cell* **20**, 187-198, (2005).
- 279 Bakshi, S., Choi, H. & Weisshaar, J. C. The spatial biology of transcription and translation in rapidly growing Escherichia coli. *Frontiers in microbiology* **6**, 636, (2015).
- 280 Kohler, R., Mooney, R. A., Mills, D. J., Landick, R. & Cramer, P. Architecture of a transcribing-translating expressome. *Science (New York, N.Y.)* **356**, 194-197, (2017).
- 281 Tennyson, C. N., Klamut, H. J. & Worton, R. G. The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. *Nature genetics* **9**, 184-190, (1995).
- 282 Murata, K. & Wolf, M. Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1862**, 324-334, (2018).
- 283 Yang, M. Q. *et al.* The emerging genomics and systems biology research lead to systems genomics studies. *BMC genomics* **15 Suppl 11**, I1-I1, (2014).

- 284 Dasari, S. & Tchounwou, P. B. Cisplatin in cancer therapy: molecular mechanisms  
of action. *European journal of pharmacology* **740**, 364-378, (2014).
- 285 Fennell, D. A. *et al.* Cisplatin in the modern era: The backbone of first-line  
chemotherapy for non-small cell lung cancer. *Cancer Treatment Reviews* **44**, 42-50,  
(2016).
- 286 Szymkowski, D. E., Yamada, M., Moggs, J. G., Karran, P. & Wood, R. D.  
Differential Human Nucleotide Excision Repair of Paired and Mismatched Cisplatin-  
DNA Adducts. *Nucleic acids research* **25**, 480-490, (1997).
- 287 Rosell, R. *et al.* Nucleotide excision repair pathways involved in Cisplatin resistance  
in non-small-cell lung cancer. *Cancer control : journal of the Moffitt Cancer Center*  
**10**, 297-305, (2003).
- 288 Gentile, F., Tuszyński, J. A. & Barakat, K. H. New design of nucleotide excision  
repair (NER) inhibitors for combination cancer therapy. *Journal of molecular  
graphics & modelling* **65**, 71-82, (2016).
- 289 Burgers, P. M. J. & Kunkel, T. A. Eukaryotic DNA Replication Fork. *Annual review  
of biochemistry* **86**, 417-438, (2017).
- 290 Bernecky, C., Pletzko, J. M. & Cramer, P. Structure of a transcribing RNA  
polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nature  
structural & molecular biology* **24**, 809-815, (2017).
- 291 Brosey, C. A., Chagot, M.-E. & Chazin, W. J. Preparation of the modular multi-  
domain protein RPA for study by NMR spectroscopy. *Methods in molecular biology  
(Clifton, N.J.)* **831**, 181-195, (2012).
- 292 Berger, I., Fitzgerald, D. J. & Richmond, T. J. Baculovirus expression system for  
heterologous multiprotein complexes. *Nature Biotechnology* **22**, 1583, (2004).
- 293 Vos, S. M. *et al.* Architecture and RNA binding of the human negative elongation  
factor. *eLife* **5**, e14981, (2016).
- 294 Vos, S. M., Stewart, N. K., Oakley, M. G. & Berger, J. M. Structural basis for the  
MukB-topoisomerase IV interaction and its functional implications in vivo. *The  
EMBO journal* **32**, 2950-2962, (2013).
- 295 Vos, S. M. *et al.* Structure of activated transcription complex Pol II-DSIF-PAF-  
SPT6. *Nature* **560**, 607-612, (2018).
- 296 Sydow, J. F. *et al.* Structural basis of transcription: mismatch-specific fidelity  
mechanisms and paused RNA polymerase II with frayed RNA. *Molecular cell* **34**,  
710-721, (2009).
- 297 Boehning, M. *et al.* RNA polymerase II clustering through carboxy-terminal domain  
phase separation. *Nature structural & molecular biology* **25**, 833-840, (2018).
- 298 Butenandt, J. P. M. E., André Carell, Thomas. *Synthesis, Crystal Structure, and  
Enzymatic Evaluation of a DNA-Photolesion Isostere*. Vol. 4 642-654 (1998).
- 299 Combe, C. W., Fischer, L. & Rappsilber, J. xiNET: cross-link network maps with  
residue resolution. *Molecular & cellular proteomics : MCP* **14**, 1137-1147, (2015).
- 300 Yang, B. *et al.* Identification of cross-linked peptides from complex samples. *Nature  
methods* **9**, 904-906, (2012).
- 301 Tang, G. *et al.* EMAN2: an extensible image processing suite for electron  
microscopy. *Journal of structural biology* **157**, 38-46, (2007).
- 302 Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM  
structure determination. *Journal of structural biology* **180**, 519-530, (2012).
- 303 Kimanius, D., Forsberg, B. O., Scheres, S. H. W. & Lindahl, E. Accelerated cryo-  
EM structure determination with parallelisation using GPUs in RELION-2. *eLife* **5**,  
e18722, (2016).



- 304 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* **66**, 486-501, (2010).
- 305 Buchan, D. W. A., Minneci, F., Nugent, T. C. O., Bryson, K. & Jones, D. T. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic acids research* **41**, W349-W357, (2013).
- 306 Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research* **42**, W252-258, (2014).
- 307 Bordoli, L. *et al.* Protein structure homology modeling using SWISS-MODEL workspace. *Nature protocols* **4**, 1-13, (2009).
- 308 Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Current protocols in protein science* **86**, 2.9.1-2.9.37, (2016).
- 309 Kim, J. L. *et al.* Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. *Structure (London, England : 1993)* **6**, 89-100, (1998).
- 310 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography* **66**, 12-21, (2010).

## ACKNOWLEDGMENTS

I am thankful to Patrick Cramer for providing me with an opportunity to do my PhD research in his lab. It is uncommon to be given such a freedom and amazing resources I have enjoyed during my PhD work. I am thankful for his mentorship and fruitful scientific discussions. I am also thankful to members of my thesis committee: Prof. Holger Stark for finding the time to advise me even outside of my TAC meetings and for allowing me to use the ProteoPlex pipeline his lab developed, Prof. Claudia Höbartner for providing advice on DNA chemistry and allowing me to use her lab equipment whenever I needed to, and finally Dr. Vladimir Pena for nice discussions and supportive words. I am also thankful to Steffen Burkhard and Kerstin Grüniger for providing such an amazing support to students in my graduate program and for helping me out on many occasions.

I was lucky to be surrounded with amazing colleagues, many of which became dear friends. I am in debt to Seychelle Vos for all of her help in the beginning of my PhD project, she was always happy to advise and share protocols. Thanks to Franz Fischer for teaching me about the insect cell culture and always keeping it cool. I am thankful to Carrie Bernecky for teaching me a lot about EM and also for many discussions and a lot of support when thing in the lab just did not work (and let me just say, there were quite some of those days), and for laughing at my silly puns. Huge thanks to Svetlana Dodonova, for many scientific discussions, for feeding me chocolate and for all the fun chats and coffee breaks. Dimitry Tegunov, you Netflix junky, thanks for all the help with the science stuff and for making sure I wouldn't starve during the late nights in the lab! Thanks to Christian for teaching me a lot about EM and model building. Marco, I have nothing to thank you for. I want to thank Haibo, Simon, Hauke, Felix (the best flatmate!), Sara, Paulina, Ying, Anna, Kristina, Katharina and Saskia for scientific discussion, fun chats and support. Servus to Kerstin, you are the best, thank you for proofreading my thesis! Thanks to Manuela and Geli for good chats about life and gardens and cats – it was a welcomed distraction when I needed a little break.

I am thankful for great scientific collaborations throughout my PhD. Biggest thanks goes to Aleksandar Chernev from the lab of Prof. Henning Urlaub, who did countless mass-spectrometry experiments for me and who always went along with my crazy ideas. I am thankful to Jan Seikowski, Dr. Vladimir Belov and Prof. Claudia Höbartner for synthesizing the CPD-containing oligo I used for TCR reconstitutions. I also briefly collaborated with Jana Schmitzova from the lab of Dr. Vladimir Pena. Her love for science is truly inspiring.

Maybe the most important acknowledgements go to the people who unconditionally supported me through my PhD journey, my mum, my dad and my amazing sister. *Veliko hvala na ogromnoj potpori tijekom mog doktorata, ne bih to mogao odgurati bez vas!* Thanks to Petra Jug for always making me laugh. Also, I am grateful to have formed a family of my own in Göttingen, where I have met some amazing people. Siv, I loved our walks and long hours spent Geocaching, thanks for being there for me when I needed support. Irena, thanks for the tough love, P10 forever! Shiru, I will never forget those delicious green tea egg rolls from Taiwan. Silvia, thanks for all the cozy blanket moments and for all the joy and laughter you are always bringing. Barratt, thank you for cooking amazing vegan food and for your support. Tahere, thanks for listening when I needed to talk.

## List of abbreviations

ATP	adenosine triphosphate
APS	ammonium peroxydisulfate
CTD	C-terminal domain of RNA-polymerase II
EM	electron microscopy
FRET	fluorescence resonance energy transfer
GTF	general transcription factor
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
His	histidine
IPTG	isopropyl- $\beta$ -D-1-thiogalactopyranoside
NTP	nucleotide triphosphate
PAGE	polyacrylamide gel electrophoresis
RNAP	RNA-polymerase II
s.d.	standard deviation
TEMED	N,N,N',N'-Tetramethylethylenediamine
UV	ultra violet
v/v	volume/volume
w/v	weight/volume

## List of figures

Figure 1   The main steps of nucleotide excision DNA repair (NER).....	4
Figure 2   Schematic representation of human global-genome nucleotide excision repair (GGR).....	8
Figure 3   The fate of lesion arrested RNAP.....	12
Figure 4   Cryo-EM structure of human TFIIH.....	15
Figure 5   Cryo-EM structure of Rad26 bound to RNA-polymerase II.....	16
Figure 6   Purification of the human TFIIH core.....	18
Figure 7   In vitro reconstitution of the TFIIH complex.....	19
Figure 8   Binding of XPC-Rad23b to DNA as measured by fluorescence anisotropy.....	20
Figure 9   Reconstitution and the crosslinking mass-spectrometry network of the core TFIIH-XPA-XPC-Rad23b-DNA complex.....	21
Figure 10   Negative stain reconstruction of the core TFIIH-XPA-XPC-Rad23b-DNA complex.....	22
Figure 11   Biochemical characterization of the core TFIIH ATPases.....	23
Figure 12   Regulation of ATPases in core TFIIH.....	25
Figure 13   Cryo-EM data processing.....	27
Figure 14   Structure of human TFIIH-XPA-DNA complex.....	29
Figure 15   Cryo-EM map quality.....	30
Figure 16   Detailed analysis of XPB and XPD bound to the DNA.....	31
Figure 17   Large-scale structural changes in TFIIH upon activation for repair.....	32
Figure 18   XPA-DNA interactions.....	34
Figure 19   XPD-DNA interactions.....	36
Figure 20   XPD regulation by the kinase module.....	37
Figure 21   Structural basis for XPD inhibition by the kinase module.....	38
Figure 22   Structural basis for XPD activation by XPA.....	39
Figure 23   Crosslinking mass-spectrometry network of the core TFIIH-XPA-XPG-DNA complex.....	40
Figure 24   Organic synthesis of a chemical CPD analogue.....	43
Figure 25   Testing the CPD-containing DNA scaffold.....	44
Figure 26   Complex formation between the RNAP and TCR-specific factors.....	45
Figure 27   The effect of CsB and UVSSA on transcription in vitro.....	46
Figure 28   TCR-specific factors do not stimulate transcription past the CPD lesion nor the dissociation of the lesion-arrested RNAP from the DNA template.....	48
Figure 29   Core TFIIH inhibits transcription in the presence of XPG and XPA.....	50
Figure 30   Core TFIIH induces RNAP backtracking in the presence of XPG and XPA... ..	51
Figure 31   XPB translocase pushes RNAP backwards.....	52
Figure 32   Active RNAP backtracking by XPB exposes CPD lesions to repair.....	53
Figure 33   Crosslinking mass-spectrometry network of the RNAP-core TFIIH-XPA-XPG complex and the effect of DNA lesions on XPB translocase activity.....	54
Figure 34   NER model which incorporates novel insights obtained in this study.....	57
Figure 35   Extension of the previous model for the human TFIIH core.....	58
Figure 36   Different conformations of the TFIIH core in transcription and DNA repair... ..	59
Figure 37   DNA-bound TFIIH core at 3.6Å reveals structural etiology of Xeroderma Pigmentosum (XP) and Trichothiodystrophy (TTD).....	61
Figure 38   Comparison between Rad14 bound to the damaged DNA duplex and XPA bound to the duplex-single strand junction.....	64
Figure 39   A possible role for XPB in DNA opening.....	68
Figure 40   New insights into the mechanism of human TCR.....	72

## List of tables

Table 1   List of NER factors and their role in DNA repair .....	6
Table 2   List of chemicals and consumables. ....	75
Table 3   List of buffers. ....	77
Table 4   List of growth media for E. coli and insect cells. ....	77
Table 5   List of antibiotics and supplements for E. coli cultures.....	78
Table 6   Bacterial strains used in this study.....	78
Table 7   Insect cell strains used in this study.....	78
Table 8   Vectors produced for protein expression in this study. ....	79
Table 9   Overhang sequences for the ligation-independent cloning.....	82
Table 10   PCR program used for gene amplification. ....	82
Table 11   LIC reaction for purified inserts and target vectors.....	83
Table 12   List of DNA constructs used in this study.....	93
Table 13   List of RNA constructs used in this study. ....	94
Supplemental Table 1   Core TFIIH-XPA-XPC-Rad23b crosslinking data.....	107
Supplemental Table 2   Core TFIIH-XPA-XPG crosslinking data. ....	108
Supplemental Table 3   RNAP-CsB-CsA-DDB1-UVSSA crosslinking data. ....	111
Supplemental Table 4   RNAP-core TFIIH-XPA-XPG crosslinking data. ....	115
Supplemental Table 5   Cryo-EM data collection, refinement and validation statistics....	121