Towards an understanding of the intricate interaction network of TFIIH

Auf dem Weg zum Verständnis des komplexen TFIIH Interaktionsnetzwerkes

Doctoral Thesis

for a doctoral degree at the Graduate School of Life Sciences,

Julius-Maximilians-Universität Würzburg,

Section Biomedicine

submitted by

Elisabeth Sofie Schönwetter

from Munich

Würzburg 2018

Submitted on: …….

Office stamp

Members of the *Promotionskomitee*:

Chairperson:	Prof. Dr. Manfred Gessler
Primary Supervisor:	Prof. Dr. Caroline Kisker
Supervisor (Second):	Prof. Dr. Bennett Van Houten
Supervisor (Third):	Prof. Dr. Thomas Müller

Date of Public Defence: ……

Date of Receipt of Certificates: …….

“Dans la vie, rien n’est à craindre, tout est à comprendre.“

Marie Curie (1867-1934)

Abstract

The integrity of its DNA is fundamental for every living cell. However, DNA is constantly threatened by exogenous and endogenous damaging agents that can cause a variety of different DNA lesions. The severe consequences of an accumulation of DNA lesions are reflected in cancerogenesis and aging. Several DNA repair mechanisms ensure the repair of DNA lesions and thus maintain DNA integrity. One of these DNA repair mechanisms is nucleotide excision repair (NER), which is famous for its ability to address a large variety of structurally unrelated DNA lesions. A key component of eukaryotic NER is the transcription factor II H (TFIIH) complex, which is not only essential for DNA repair but also for transcription. The TFIIH complex is composed of ten subunits. How these subunits work together during NER to unwind the DNA around the lesion is, however, not yet fully understood. High-resolution structural data and biochemical insights into the function of every subunit are thus indispensable to understand the functional networks within TFIIH. The importance of an intact TFIIH complex is reflected in the severe consequences of patient mutations in the TFIIH subunits XPB, XPD or p8 leading to the hallmark diseases xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy. Defects in the NER pathway are further associated with several types of cancer including skin cancer.

The herein described work focused on five TFIIH subunits derived from the thermophilic fungus *Chaetomium thermophilum*, the p34/p44 pair and the ternary XPB/p52/p8 complex. The interaction between p34 and p44 was characterized based on a high-resolution structure of the p34_vWA/p44_RING minimal complex. Biochemical studies of the p34/p44 interaction led to the disclosure of an additional interaction between the p34 and p44 subunits, which had not been characterized so far. The p34/p44 interaction was shown to be central to TFIIH, which justifies the presence of several redundant interfaces to safeguard the interaction between the two proteins and might explain why so far, no patient mutations in these subunits have been identified. The p52 subunit of TFIIH was known to be crucial to stimulate the ATPase activity of XPB, which is required during NER. This work presents the first entire atomic resolution structural characterization of p52, which was derived of several crystal structures of p52 variants and a p52/p8 variant thereby demonstrating the interaction between p52 and p8. The precise structural model of p52 offered the possibility to investigate interactions with other TFIIH subunits in more detail. The middle domain 2 of p52 and the N-terminal domain of XPB were shown to mediate the main interaction between the two subunits. An analysis of the p52 crystal structures within

recently published cryo-electron microscopy structures of TFIIH provides a model of how p52 and p8 stimulate the ATPase activity of XPB, which is essential for NER and transcription. The structural and biochemical findings of this work provide an additional building block towards the uncovering of the architecture and function of this essential transcription factor.

Zusammenfassung

Die Unversehrtheit ihrer DNA ist für jede lebende Zelle elementar. Die DNA ist jedoch fortwährend exogenen und endogenen Toxinen ausgeliefert, die eine Vielfalt unterschiedlicher DNA-Schäden verursachen. Die sehr ernsthaften Konsequenzen einer Anhäufung von DNA-Schäden spiegeln sich in der Entstehung von Tumorerkrankungen und Alterung wider. Verschiedene DNA-Reparaturmechanismen sorgen für die Reparatur von DNA-Schäden und erhalten so die Unversehrtheit der DNA. Einer dieser DNA-Reparaturmechanismen ist die Nukleotid-Exzisions-Reparatur (NER), die bekannt dafür ist, eine Vielfalt an strukturell unterschiedlichen DNA-Schäden zu adressieren. Eine Schlüsselkomponente der eukaryotischen NER ist der Transkriptionsfaktor II H (TFIIH), welcher nicht nur für die DNA-Reparatur, sondern auch für die Transkription essentiell ist. Der TFIIH Komplex besteht aus zehn Untereinheiten. Wie diese Untereinheiten zusammenarbeiten, um die DNA um den Schaden herum zu entwinden, ist jedoch noch nicht hinreichend bekannt. Hochaufgelöste Strukturdaten und biochemische Einblicke in die Funktion jeder Untereinheit sind daher unabkömmlich, um das funktionelle Netzwerk innerhalb dieses Transkriptionsfaktors zu verstehen. Die Bedeutung eines intakten TFIIH Komplexes spiegelt sich in den verheerenden Folgen von Patientenmutationen in den TFIIH Untereinheiten XPB, XPD oder p8 wider, die zu den kennzeichnenden Krankheitsbildern von Xeroderma Pigmentosum, Cockayne Syndrom und Trichothiodystrophie führen. Ein fehlerhafter NER Reparaturweg ist ferner mit einigen Krebsarten wie Hautkrebs assoziiert.

Die hier beschriebene Arbeit hat sich auf fünf TFIIH Untereinheiten konzentriert, die aus dem thermophilen Pilz *Chaetomium thermophilum* stammen, das p34/p44 Heterodimer und der ternäre XPB/p52/p8 Komplex. Die Interaktion zwischen p34 und p44 wurde basierend auf einer hochaufgelösten Kristallstruktur des p34_vWA/p44_RING Minimalkomplexes charakterisiert. Biochemische Studien der p34/p44 Interaktion haben zur Aufdeckung einer weiteren Interaktion zwischen p34 und p44 geführt, die bisher noch nicht charakterisiert wurde. Die p34/p44 Interaktion ist von zentraler Bedeutung für TFIIH, was die Gegenwart mehrerer redundanter Schnittstellen zwischen p34 und p44, um die p34/p44 Interaktion abzusichern, rechtfertigt und erklären könnte, warum bislang keine Patientenmutationen in diesen Untereinheiten identifiziert wurden. Die p52 Untereinheit von TFIIH ist bekannt dafür, die ATPase-Aktivität von XPB zu stimulieren, die während der NER benötigt wird. Diese Arbeit zeigt die erste vollständige atomare strukturelle Charakterisierung von p52, die aus verschiedenen Kristallstrukturen von p52

Varianten und einer p52/p8 Variante, welche die Interaktion zwischen p52 und p8 darstellt, stammt. Das Strukturmodel von p52 bietet die Möglichkeit Interaktionen mit anderen TFIIH Untereinheiten zu analysieren. Es wurde gezeigt, dass die mittlere Domäne 2 von p52 und die N-terminale Domäne von XPB die hauptsächliche Interaktion zwischen den beiden Untereinheiten vermitteln. Eine Analyse der p52 Kristallstrukturen in neuesten publizierten cryo-Elektronenmikroskopie TFIIH-Strukturen ermöglichte die Erstellung eines Models, das zeigt, wie p52 und p8 die ATPase-Aktivität von XPB stimulieren, welche essentiell für die NER und die Transkription ist. Die strukturellen und biochemischen Erkenntnisse dieser Arbeit bieten einen wichtigen Beitrag zur Enthüllung der Architektur und Funktion von TFIIH, einem essentiellen zellulären Komplex.

Table of Contents

I Introduction

I.1 Genome Integrity

"I wanted to investigate if DNA, like tRNA, was susceptible to slow decomposition. This was a rather far-fetched idea, because DNA, as the carrier of genetic information in our cells, was believed to be very stable in the intracellular environment."

(Tomas Lindahl, Nobel lecture "The Intrinsic Fragility of DNA", December 8th, 2015)

The paradigm of DNA as 'the carrier of genetic information in our cells' to be very robust and stable was assumed for a long time, but Tomas Lindahl has been proven correct, as all macromolecules also DNA is subjected to multitudinous chemical alterations depicting the 'intrinsic fragility of DNA'. Thus, how do these chemical alterations arise and what are their consequences?

Independent of the source of the damage, chemical alterations in DNA bear a high risk for the entire organism as genetic instability is directly linked to cancer development. During our lifetime, diverse DNA lesions when left unrepaired, can accumulate and drive tumorigenesis. Understanding the causes of DNA damage and their effect on the genome is therefore indispensable for cancer prevention as well as cancer therapy.

I.1.1 Threats to genome integrity

Chemical alterations in DNA can arise from diverse sources and thus lead to a broad spectrum of different DNA lesions. Endogenous sources like by-products of the normal cellular metabolism and natural errors depict a daily cause of DNA lesions. More than 25 DNA modifications derived from reactive oxygen species or nitrogen oxide species have been identified in DNA (1). These free radical species are generated as by-products of the normal cellular metabolism derived, for example, from the respiratory chain cascade or inflammatory response. A well-known oxidative DNA lesion is 8-oxo-2'-deoxyguanosine (8-oxo-dG) (1). The replication machinery, to name another source of endogenous DNA damage, is highly accurate but not error-free. High-fidelity DNA polymerases like the replicative Pol ε have a low error rate of 1 in 10^{-7} for every newly

synthesized nucleotide due to their intrinsic proofreading capacity. Nevertheless, the replication machinery can be a source of mutagenesis, especially when combined with mutations in the replicative DNA polymerases or a misbalance of the cellular deoxynucleoside triphosphate pool (2). Low-fidelity polymerases like Pol η, Pol ι and Pol κ, that lack proofreading capacity, inherently bear a much higher risk of mutagenesis as they display a much higher error rate. In a process called translesion synthesis (TLS) these low-fidelity polymerases are able to continue synthesis past a DNA damage especially under cellular stress conditions. However, TLS DNA polymerases have a preference to incorporate an adenine opposite to an abasic site, which can lead to different mutational signatures depending on the original base at the site (3). Abasic sites are an example for spontaneous events in our cells, in this case the spontaneous hydrolysis of the nucleotide. Another event, which is occurring spontaneously in all four DNA bases, is deamination. One of the most abundant mutational signatures is the C-G -> T-A transition caused by the hydrolytic deamination of 5-methylcytosines at methylated CpG dinucleotides (2), which is correlated with age and associated with more than 25 different types of cancer (4).

Next to endogenous sources, a large number of exogenous physical and chemical agents can generate DNA lesions. An important environmental factor causing different modifications in DNA is the ultraviolet (UV) component of sunlight. UV exposure can lead to the generation of pyrimidine dimers like (6-4) pyrimidine photoproducts ((6-4)PPs) and cyclobutane pyrimidine dimers (CPDs) and is linked to different types of skin cancer, such as basal-cell carcinoma, squamous-cell carcinoma and melanoma (5, 6). Ionizing radiation can cause fatal single- or double-strand breaks in DNA. Unfortunately, humans are not only exposed to radiation, the exposure to several genotoxic compounds is likewise problematic. Tobacco smoke, for example, consists of more than 60 carcinogens causing a diverse spectrum of different DNA damages (7). Among them are N-nitrosamines, aromatic amines and polycyclic aromatic hydrocarbons. In our cells N-nitrosamines are converted to metabolites that can alkylate DNA bases at diverse positions. It is known that cigarette smoking is an important cause not only for lung cancer, but also for esophageal, oral or laryngeal cancers and many more (7). Other genotoxic compounds that can cause diverse DNA damages are industrial chemicals like formaldehyde (8) and chemotherapeutic agents like *cis*-Diaminodichloroplatinum (cisplatin) and temozolomide. Temozolomide is metabolized to a compound that preferably methylates guanines at their N7 or O6 positions (9).

Taken together, there is a plethora of different exogenous and endogenous damaging agents that can lead to diverse DNA lesions. It is estimated that 10^4 – 10^5 damages arise per mammalian cell per day (8, 10). The fatal consequence of unrepaired DNA lesions can be observed in immediate

effects such as the blockage of transcription or replication, which can cause cell cycle delay or arrest or cell death and hence, accelerate premature aging. Furthermore, long-term effects are at least as severe as short-term effect as DNA lesions can be converted into mutations if misinterpreted by the replication machinery and thus, drive tumorigenesis. Next to single mutations, also DNA rearrangements, insertions, deletions, loss of heterozygosity and chromosomal anomalies can evolve and contribute to the formation of cancer (11). DNA damage is thus a major cause for the development of cancer and the diseases of premature aging. The main difficulty, in comparison to RNA and proteins, might be that DNA cannot just be re-established. As the carrier of all genetic information it is crucial for all cellular processes, but accumulates lesions over a lifetime and relies exclusively on cellular repair mechanisms that are able to locate and remove lesions in the genome (12).

I.1.2 DNA repair mechanisms

To avoid the disastrous consequences of tumorigenesis and premature aging eukaryotic cells are equipped with an intricate network of DNA repair and checkpoint pathways that are tightly connected with routine cellular processes like replication, transcription or recombination. Due to the broad variety of different DNA lesions many different DNA repair mechanisms have evolved that ensure genome maintenance.

In mammals, there are two sets of enzymes that can reverse alkylated bases in a single-protein direct damage reversal mechanism. The O^6-alkylguanine-DNA alkyltransferase (AGT) enzyme, also known as O^6-methylguanine-DNA methyltransferase (MGM), removes alkyl groups from O-alkylated lesions in a reaction that leads to the irreversible inactivation of the enzyme itself (13). A second class of enzymes that can reverse alkylated bases consists of homologs of the *E. coli* AlkB enzyme, which are named ALKBH1 - 8 (alkylation repair homologs 1 to 8) and FTO (fat mass and obesity associated) in human cells (13). In addition to these rather simple repair enzymes several multistep repair mechanisms exist, which are dedicated to a specific category of DNA lesions. An overview of different DNA lesions, their sources and repair mechanisms is shown in Figure I.1.

Base excision repair (BER) is dedicated to oxidized DNA bases like 8-oxo-dG, alkylated or deaminated DNA bases or the incorporation of uracil instead of thymidine. Furthermore, it recognizes abasic sites and single-strand breaks as these represent intermediate stages of the BER mechanism. A set of at least eleven DNA glycosylases, each equipped with different substrate specificities, is responsible for damage recognition (14). On the one hand, this damage

specialization allows the removal of a large variety of lesions. On the other hand, the inactivation of individual glycosylases has no severe consequences as their functions are also partly overlapping.

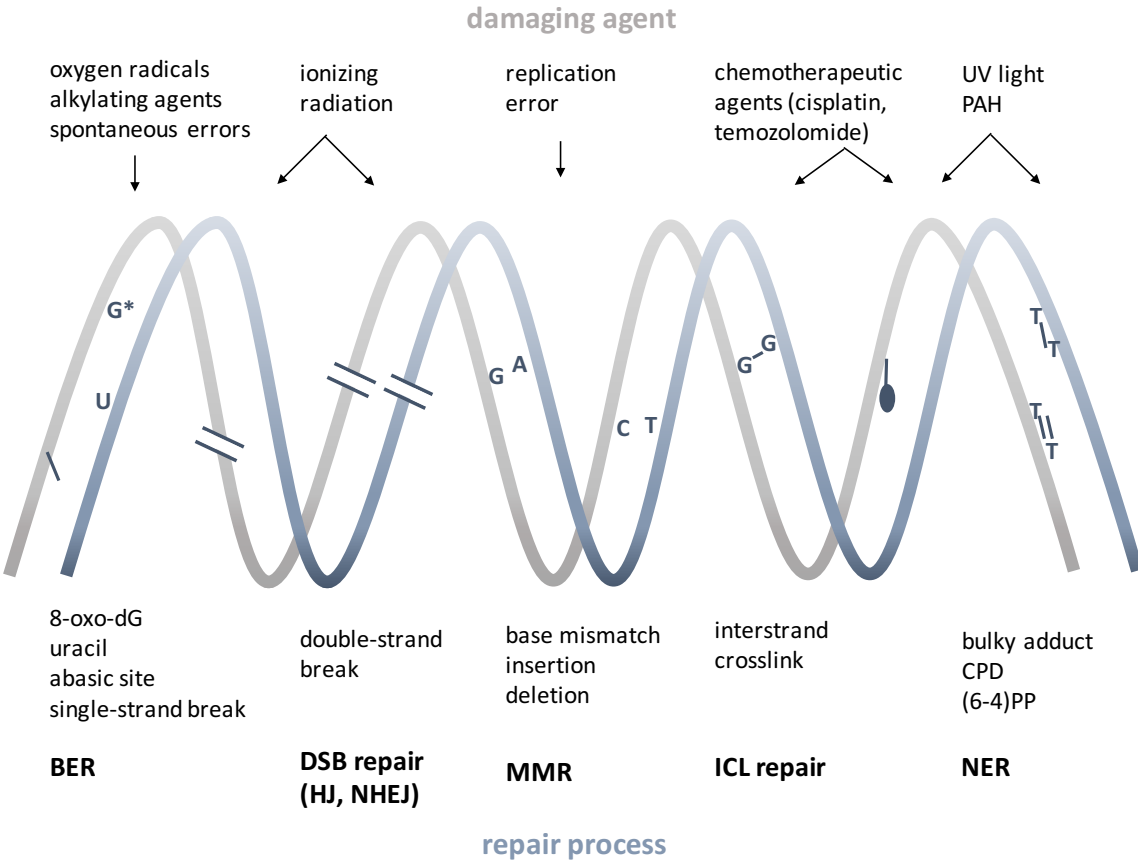


Figure I.1: Overview of different damaging agents, the lesions they cause and the respective repair processes. Various endogenous and exogenous damaging agents are depicted that lead to a variety of different lesions. Due to this variety, different DNA repair mechanisms are required to take care of these lesions, which are base excision repair (BER), double-strand break (DSB) repair (HJ = homologous end-joining; NHEJ = non-homologous end-joining), mismatch repair (MMR), interstrand crosslink (ICL) repair and nucleotide excision repair (NER). Abbreviations: UV = ultraviolet; PAH = polycyclic aromatic hydrocarbons; 8-oxo-dG = 8-oxo-2'-deoxyguanosine; CPD = cyclobutane pyrimidine dimers; (6-4)PP = (6-4) pyrimidine photoproducts.

Base mismatches generated, for example, during replication can be repaired by the mismatch repair (MMR) system. Besides the close connection to replication, MMR is also closely connected to other cellular processes as microsatellite stability, meiotic and mitotic recombination, class-switch recombination, apoptosis and somatic hypermutation. The importance of MMR for genome maintenance is shown by the severe effects of germline mutations in MMR genes that cause the Lynch syndrome, also known as hereditary non-polyposis colorectal cancer (13).

Interstrand crosslinks (ICLs) induced, for example, by chemotherapeutic agents are highly toxic for the cell as they can block replication and transcription. To overcome these blockages TLS polymerases are involved in resolving ICLs (13). Depending on the stage of the cell cycle, ICL

lesions are repaired differently. In the G0/G1-phase of the cell cycle, repair of ICL lesions was shown to involve nucleotide excision repair proteins. In the S/G2-phase of the cell cycle the replication-dependent ICL repair is dependent on the Fanconi Anemia (FA) pathway and homologous recombination-mediated replication and fork stabilization. FA is known to be a cancer predisposition syndrome (15).

Double-strand breaks (DSBs) are also a disastrous fate for cells as unresolved DSBs can lead to cell death or give rise to cancer. Two major mechanisms, homologous recombination (HR) and non-homologous end-joining (NHEJ), are responsible for resolving DSBs. NHEJ is more error-prone and takes place during the G0/G1-phase, whereas HR is preferred after replication in the late S- and G2-phase when the identical sister chromatid can be taken as a template (16). Deficiencies in DSB repair genes are associated with several human disorders like the Werner or Rothmund-Thomson syndrome, both caused by defects in RecQ-like helicases. A strong deposition for breast cancer is provoked by inherited defects in the HR-related genes BRCA1 and BRCA2 (breast cancer 1 and 2) (17).

Another famous DNA repair pathway is nucleotide excision repair, which will be explained in greater detail in the following chapter.

I.2 Nucleotide Excision Repair

Nucleotide excision repair (NER) is a crucially important DNA repair mechanism, which is particularly reflected in its association with different human disorders such as xeroderma pigmentosum (XP), Cockayne syndrome (CS), trichothiodystrophy (TTD) or combined phenotypes of these. Defective NER is additionally associated with several types of cancer such as skin cancer. The mammalian NER mechanism is a complex pathway, which involves various components and is divided into two sub-pathways that differ in the damage recognition step, the transcription-coupled (TC) and global genome (GG) - NER pathways. Damage recognition is a decisive step to activate the NER pathway. Sources and types of DNA lesions that can be recognized by NER will be explained in more detail in the following.

I.2.1 Lesions addressed by NER

One of the outstanding characteristics of NER is certainly its ability to address a large range of structurally unrelated DNA lesions. Among these lesions are DNA crosslinks of two adjacent

pyrimidine bases induced by UV light like (6-4)PPs and CPDs. CPDs are preferably formed between two adjacent thymine residues (18). The crystal structure of a CPD containing nucleosome indicated that CPD lesions are stably accommodated within the DNA helix and do not lead to remarkable DNA distortions (19), which, however, causes a more difficult detection of the lesion. However, the hydrogen bonds with the opposite adenine residues are weakened. (6-4)PP lesions cause, in contrast to CPD lesions, a more pronounced helix distortion by creating a flexible DNA backbone at the lesion site (20). Hence, the base pairing to the complementary adenine residues is lost. (6-4)PP and CPD lesions display clearly serious threats for cells, rendering the accurate recognition and repair by NER indispensable. Interestingly, prokaryotes, archaea and some eukaryotes are equipped with DNA photolyases that are able to successfully repair CPDs and (6-4)PPs by an electron-transfer mechanism (21), whereas placental mammals rely solely on the repair action of the NER machinery.

Another source of DNA lesions addressed by NER are electrophilic compounds that can form adducts with the nucleophilic sites in DNA. These compounds can either directly or after metabolic transformation react with the DNA and form bulky DNA adducts. Examples are polycyclic aromatic hydrocarbons (PAHs) and cisplatin (18). PAHs are hydrophobic compounds ubiquitously present in our environment, for example, in tobacco smoke, automobile exhausts or well-cooked food. PAHs can undergo different metabolic transformations before reacting with the DNA. An example are dihydrodiol epoxide groups that are generated in PAHs like benzo[a]pyrene (B[a]P) and primarily form DNA adducts with the exocyclic amino groups of guanines (18, 22). Cisplatin is a famous chemotherapeutic drug and by itself a neutral compound, which is transformed to an electrophilic molecule by substitution of the chloride ligands through water or hydroxide ions. The most abundant DNA adducts that are formed by this activated cisplatin are intrastrand crosslinks between two adjacent guanines (18). NER is equipped with an intricate network of proteins, which ensure that all of these lesions are detected and repaired.

I.2.2 Eukaryotic global genome NER

The intricate network of about 30 proteins that are required for NER follow a coordinated pathway, which involves initial damage recognition, damage verification, incision and removal of the damaged strand and re-synthesis. At the damage verification step, the GG-NER and TC-NER pathways converge into a common repair pathway. In the past decades, many studies have illuminated the steps and components that are responsible for the NER mechanism. Despite these efforts, many unanswered questions remain and the NER mechanism is not yet fully understood.

In the following, the NER pathway and the interplay of its components as it is known so far will be discussed. A schematical overview of the NER mechanism is provided in Figure I.2.

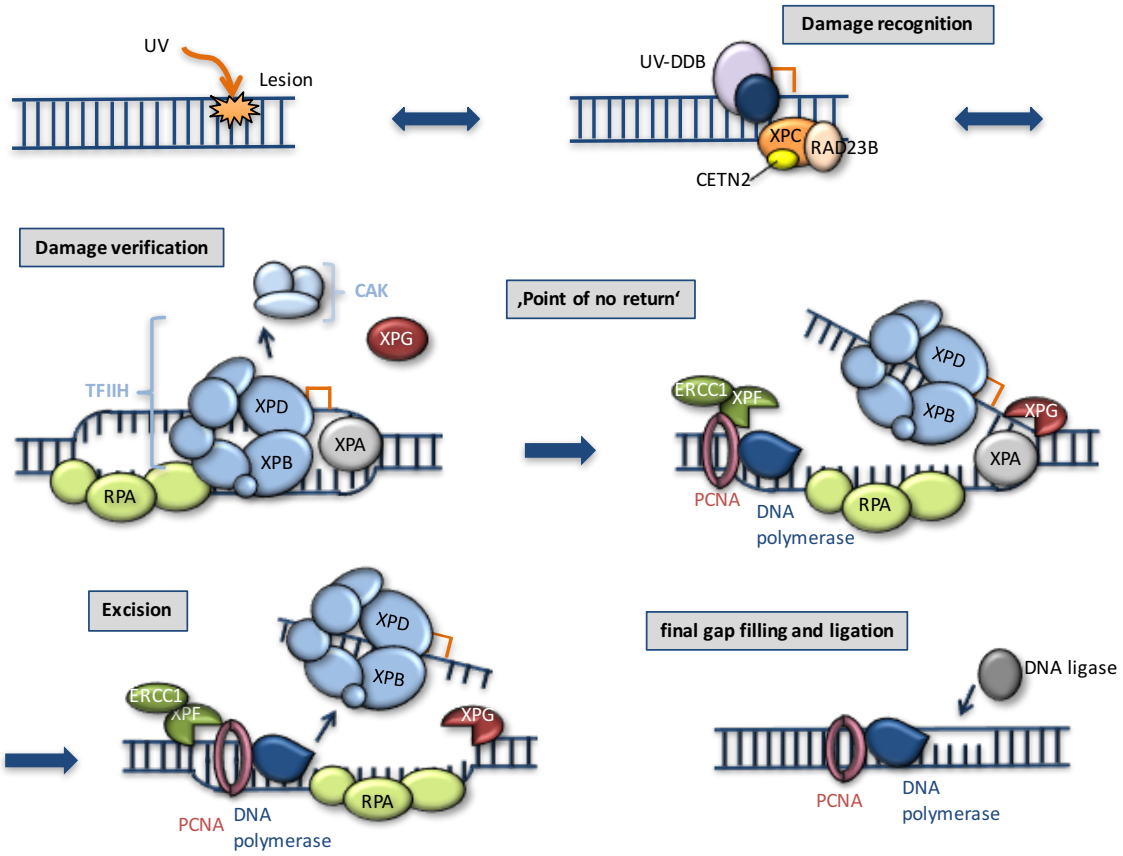


Figure I.2: Overview of the eukaryotic global genome NER pathway. After the DNA damage detection is performed by XPC-RAD23B-CENT2 and UV-DDB the TFIIH complex and further repair factors like XPA and RPA are recruited to the site of the damage. DNA damage verification and strand opening is performed by the TFIIH complex before the endonucleases ERCC1-XPF and XPG are mediating the 5' and 3' incision, respectively. Concomitant to the dual incision DNA synthesis takes places, then the damaged ssDNA strand is released and the DNA nicks are sealed by a ligase.

In GG-NER, initial damage detection is mainly mediated by two protein complexes, the XPC (xeroderma pigmentosum complementation group C) and the UV-DDB (UV-DNA damage binding protein) complex. The heterotrimeric XPC complex consists of XPC, RAD23B and CENT2 (centrin 2). A crystal structure of the yeast homologs of the XPC-RAD23B complex (Rad4-Rad23) in complex with a CPD-containing DNA double strand revealed closer insights into the damage detection of the XPC complex (23). Interestingly, the XPC-RAD23B complex does not bind to the DNA lesion itself, but rather to the single-stranded DNA (ssDNA) strand opposite to the lesion resulting in additional DNA bending of ~42°. Thereby, a β-hairpin of XPC is inserted through the DNA duplex and the two damaged nucleotides that contain the CPD lesion are flipped out. The fact that XPC does not bind to the lesion itself explains, why XPC is able to recognize such a broad

variety of structurally unrelated lesions (23). A more recent crystal structure illustrating that the XPC-RAD23B complex can also bind to undamaged DNA in a similar manner as to damaged DNA (24) and single-molecule studies on the XPC-RAD23B complex (25) propose a dynamic multi-step damage recognition model for the XPC complex. Furthermore, another recent study on XPC binding to B[a]P-induced DNA lesions supports this model and the theory that XPC uses different binding pathways for different lesions (25, 26).

In line with the observation that there are differences in lesion recognition and thus there are lesions that are better recognized and those that are poorly recognized by XPC, an assistant complex, the UV-DDB complex, is present in eukaryotic cells to facilitate lesion recognition. The UV-DDB complex consists of DDB1 and DDB2 (also named XPE) (27). In contrast to (6-4)PP lesions, CPDs lesions can also occur in highly condensed heterochromatin, but their detection by XPC is inefficient and requires the support of UV-DDB (28). The DDB2 subunit of the UV-DDB complex is suggested to facilitate the repair of CPD lesions in heterochromatin, most likely through chromatin remodeling, which makes the lesion site accessible for XPC (28-31). Because of their only mild effect on helix distortion CPD lesions are thought to be poor substrates for XPC in general requiring the UV-DDB complex to recruit XPC to the lesion site (19, 32). Crystal structures of the UV-DDB complex in complex with damaged DNA have been solved and show that the DDB2 subunit directly binds to the lesion within its binding pocket, bends the DNA by ~40° around the lesion and flips the lesion out of the DNA helix (33, 34). A short stretch of ssDNA strand is thus created, which facilitates XPC binding. Interestingly, DDB2 also contains a β-hairpin that is inserted into the duplex DNA as observed for XPC. In addition, the CRL (cullin 4A – regulator of cullins 1 (Roc1) E3 ubiquitin ligase) complex is recruited to the site of the damaged chromatin by DDB2. The subsequently formed UV-DDB-CRL complex was shown to mono-ubiquitinate histones, auto-ubiquitinate DDB2 and to additionally ubiquitinate XPC (35, 36). Auto-ubiquitination of DDB2 leads to its degradation, which might be crucial for handing over the lesion from the UV-DDB complex to the XPC complex (36, 37).

Damage detection is followed by damage verification, which requires the recruitment of the transcription factor IIH (TFIIH), a transcription initiation and repair factor, to the site of the damage. TFIIH is recruited to the lesion site immediately after XPC has bound to the damaged DNA strand, which functions as an ideal substrate to anchor TFIIH (38, 39). The chromatin remodeler CHD1 has recently been suggested to play a role in assisting XPC to recruit TFIIH to the damage site (40).

TFIIH is a complex of ten subunits that can be subdivided into the CAK (cyclin-dependent kinase (CDK)-activating kinase) complex composed of three subunits and the core TFIIH complex composed of seven subunits, which also includes the two helicases XPB and XPD. The ATPase activity of XPB is required for anchoring TFIIH at the site of the lesion (41), whereas the helicase activity of XPD is indispensable for unwinding the DNA around the lesion and creating the NER bubble (42, 43). As the XPC and UV-DDB complexes bind to several different lesions and not all of them are actual NER substrates, a lesion verification step after initial recognition is essential before NER can proceed (44, 45). TFIIH scans the DNA strand in a 5'-3' direction and it is assumed that XPD verifies the lesion during translocation (46, 47). Crystal structures of XPD homologs from archaea showed that the helicase domain 1 and the Arch and FeS cluster domains of XPD together form a pore for ssDNA (48-51). Undamaged ssDNA can pass through this pore, whereas bulky lesions may block XPD and thus allow the verification of the lesion by XPD (52). In a further study by Li *et al.* a tripartite lesion recognition model involving XPC, TFIIH and XPA was proposed (53). Upon recruitment of TFIIH to the lesion site by XPC the two helicases XPB and XPD of TFIIH are suggested to translocate towards the lesion on one strand each. XPA, a further component of the NER pathway, was shown to stimulate the translocation and simultaneous DNA unwinding by TFIIH. Upon encounter of a bulky lesion on the strand scanned by XPD TFIIH is suggested to be blocked and the stalling of TFIIH is enhanced by XPA (53). Interestingly, XPA was shown to also promote the release of the CAK from core TFIIH, which seems to be essential for the helicase activity of XPD to progress (54). A more detailed picture of TFIIH and its role in NER and transcription is given in chapter I.3.

XPA was shown to function not only in damage verification, but also in the assembly of several NER components, which becomes apparent when taking a look at the numerous interactions that XPA is fulfilling with further NER components (55). So far, it is uncertain how XPA is engaged on the NER bubble created by TFIIH. The preference for XPA to bind to ssDNA-dsDNA junctions suggests that it binds to either end of the NER bubble (56). Two crystal structures of the DNA binding domain of the yeast homolog of XPA (Rad14) in complex with two different damaged DNA duplexes revealed that a dimer of XPA binds to the damaged duplex DNA without any direct contact to the lesion itself and that the DNA in this complex is kinked by 70° (57). Noteworthy, XPA contains a β-hairpin that is inserted into the duplex DNA and crucial for lesion recognition as observed for XPC and DDB2 suggesting that it is a common feature for damage detection and verification within NER (58). It is therefore disputable if XPA is already involved in damage detection and thus recruited to the lesion site before the arrival of TFIIH. Some studies suggest

that an arrival of XPA prior to the recruitment of TFIIH can take place (59) and other studies argue for an arrival together with or after the recruitment of TFIIH (54, 60).

One of the direct interaction partners of XPA is the heterotrimeric replication protein A (RPA). RPA is recruited to the NER bubble to protect the undamaged ssDNA strand from nucleolytic damage (61). It is a common ssDNA binding protein that is known to protect ssDNA strands generated during various cellular processes from endonucleases, DNA reannealing and the formation of ssDNA hairpins (62) and consists of the three subunits RPA70, RPA32 and RPA14. Nuclear magnetic resonance (NMR) data indicate that XPA interacts with the RPA70 subunit of RPA (63).

In a next step, a 22 to 30 nucleotides long ssDNA patch containing the lesion is excised after 5' and 3' incision into the damaged DNA strand and the resulting gap is filled by DNA synthesis. The two endonucleases ERCC1(excision repair cross complementing 1)-XPF and XPG that are responsible for the 5' and 3' incision into the damaged DNA strand, respectively, are recruited to the NER bubble independent of each other (39). The recruitment of the heterodimer ERCC1-XPF to the NER bubble by XPA is mediated through an interaction between ERCC1 and XPA (64, 65). TFIIH seems to be important for the recruitment of XPG to the NER bubble (66, 67). All previous steps are reversible, but upon incision into the damaged DNA strand a 'point of no return' is reached. The coordination and defined temporal order of this action was surveyed in a detailed study by Staresincic *et al.* (68). According to this analysis a 'cut-patch-cut-patch' mechanism takes place, in which dual incision and repair synthesis are coordinated with each other. First, 5' incision by ERCC1-XPF occurs, but only in the presence of XPG proposing a verification step to ensure that all necessary proteins are recruited before the incision reaction starts. 5' incision was shown to be immediately followed by gap filling. About 18 to 20 nucleotides are filled in by the repair synthesis machinery, then XPG performs the 3' incision into the ssDNA strand triggered maybe by a stalling of the polymerase.

The presence of XPG is required for the recruitment of the classical DNA repair synthesis machinery consisting of the clamp loader RFC (replication factor C), the processivity factor PCNA (proliferating cell nuclear antigen) and polymerase δ (Pol δ) (68, 69). Dependent mostly on the state of the cell cycle either Pol δ, Pol ε or Pol κ, which are recruited by different factors, perform the DNA repair synthesis during NER (70, 71). Sealing of the nicks is performed by either DNA ligase I mostly in proliferating cells after DNA synthesis by Pol ε or DNA ligase III in quiescent cells after Pol δ and Pol κ have performed the DNA synthesis (71, 72). After dual incision, the excised

oligonucleotide was shown to be released in complex with TFIIH and XPG, sometimes including also XPF, before it is bound by RPA or degraded by nucleases (73, 74).
These main steps involving damage verification, incision and excision of the damaged strand and re-synthesis are the same in GG-NER and TC-NER. However, initial damage recognition differs in GG-NER and TC-NER and the latter is described below.

I.2.3 Eukaryotic transcription-coupled NER

Lesions occurring on transcribed strands of active genes can block transcription and lead to a stalling of RNA polymerase II (RNA Pol II). A stalling of RNA Pol II is the signal for TC-NER to be initiated. An advantage of transcription-coupled repair is certainly the fact that the nucleosome structure is already loosened by RNA Pol II (75). Human TC-NER is taking place in the absence of the GG-NER damage recognition factors DDB2, XPC and RAD23B. Interestingly, the yeast XPC homolog Rad4 was found to be important for both GG-NER and TC-NER (76).
CSB (Cockayne syndrome – B), a DNA-dependent ATPase, is the key player of TC-NER and translocates along with RNA Pol II as a transcription elongation factor. Upon stalling of RNA Pol II CSB binds to it and leads to a change in the interface between RNA Pol II and the DNA by wrapping the DNA around itself (77). CSB then recruits the CSA complex, core-NER factors and chromatin factors such as p300 to the lesion site (76). The CSA complex is a E3 ubiquitin ligase complex, which includes CSA, RBX1 and cullin 4A. CSA is required for further recruitment of XAB2 (XPA binding protein 2), HMGN1 (high mobility group nucleosome binding protein domain 1), UVSSA (UV-stimulated scaffold protein A) and TFIIS. The partner of UVSSA, Usp7 (ubiquitin specific peptidase 7), is important for deubiquitylation and stabilization of CSB and is thereby counteracting the CSA-induced degradation of CSB (78). Upon recruitment of TFIIH to the site of the lesion TC-NER and GG-NER converge into a common mechanism.
The fate of RNA Pol II still remains elusive (76). The chromatin factors p300 and HMGN1 and the elongation factor TFIIS could facilitate a backtracking of RNA Pol II and thus allow the NER machinery to obtain sufficient access to the lesion site. According to another model, RNA Pol II is remodeled directly at the lesion site to provide enough space for the NER machinery. Other options like bypassing of the lesion would result in transcriptional mutagenesis and persistent stalling of the RNA Pol II could initiate cell death, which should be avoided, if at all, possible (76).
It is currently discussed if transcription-coupled repair is also part of a crossover between NER and BER and necessary to resolve transcription-blocking oxidative damage (78). The importance

of both accurate TC-NER and GG-NER is demonstrated by the severe consequences of defects in any genes involved in these two sub-pathways.

I.2.4 NER-related clinical disorders

Defects in any genes involved in the NER pathway can lead to severe diseases like the rare, autosomal recessive disorders XP, CS or TTD. The phenotypes of patients suffering from XP, CS, TTD or combinations of XP/CS and XP/TTD are highly heterogeneous (79).

XP has already been described by Moritz Kaposi in 1874. To date, eight different complementation groups are known – *XPA* through *XPG*, which are required for GG-NER and the XP variant *XPV*, which is associated with defects in DNA Pol η and thereby impaired TLS (79). All XP patients show an extreme sensitivity to sunlight reflected in dry scaly skin and abnormal skin pigmentation and a high incidence of malignant skin cancers such as basal-cell carcinoma, squamous-cell carcinoma and melanoma. About 20 - 30% of the XP patients develop additional neurological and cognitive dysfunctions like microcephaly, mental retardation, ataxia and spasticity and at least 40% of them show ophthalmologic pathologies (80). The hallmark of XP is the increased risk in developing skin cancer, which depicts a clear link between impairments in the GG-NER pathway and an increase of mutations driving cancerogenesis (12).

CS is characterized by mutations in the *CSA* or *CSB* genes. Patients suffering from CS display clinically extreme heterogeneous phenotypes with mutations in the *CSB* gene leading to a much more severe outcome than mutations in the *CSA* gene. Characteristic features of CS are premature aging, growth failure and neurological impairment (80). CS is caused by an impairment in TC-NER, which blocks transcription when a damage is encountered and leads to an increase in cell death rather than an increase of mutations. This might explain why patients do not show an elevated cancer risk, but signs of premature aging (12). Mutations in the *XPB*, *XPD* and *XPG* genes can cause combined features of XP and CS and mutations in the *XPF* gene may cause a combined phenotype of XP, CS and FA (79).

Only 100 cases of TTD have been reported so far, indicating that it is a very rare disease. TTD is a severe disorder caused by mutations in the TFIIH genes *XPB*, *XPD* or *TTDA* (trichothiodystrophy group A – coding for the TFIIH subunit p8). A non-photosensitivity version of TTD is caused by mutations in the *TTDN1* (TTD nonphotosensitive 1) gene (80). The main indicator for TTD is the brittle and fragile hair of the patients caused by a lack of sulfur. Apart from the brittle hair structure most patients show additional photosensitivity, ichthyosis, intellectual impairment, neurological abnormalities, decreased fertility and growth retardation (11). TTD patients do not

show an increased cancer risk as it is observed for XP patients, which might be explained by impairments in transcription rather than in DNA repair due to the dual role of TFIIH in both NER and transcription. A combined XP/TTD phenotype was observed for some patients with mutations in the *XPD* gene (79, 80).

XPB, XPD and p8 are the only subunits of TFIIH that are associated with these three disorders and no patient mutations in any of the other TFIIH subunits are known so far. Patient mutations in the *XPB* gene include a point mutation resulting in the missense mutation F99S leading to XP or combined forms of XP/CS (81) and a point mutation resulting in the missense mutation T119P leads to TTD (82). Another known mutation displays a C to A transversion in the last intron resulting in a frameshift at the protein level, which changes the last 41 amino acids of XPB and leads to XP/CS (83). There are multiple known patient mutations in the *XPD* gene associated with XP, CS and TTD and most of them are leading to an impairment of the interaction between XPD and its partner p44 (80). Known patient mutations in the *TTDA* gene are only associated with TTD and comprise the nonsense mutation R56X creating a premature stop codon, the missense mutation L21P and a mutation in the start codon (M1T) leading to a shortened p8 protein (84).
To date, patients with XP, CS, TTD or a combination of these cannot be cured, but there are treatments available that can improve the quality of life of these patients (80).

I.3 The Transcription Factor II H

The dual role of TFIIH in transcription and NER and its implication in diseases like XP, CS and TTD clearly underline its importance. In 1989, TFIIH was first discovered as a new general transcription factor that is essential during transcription (85). A few years later it became apparent that the helicases XPB and XPD that were known to participate in DNA repair (83, 86, 87) are subunits of the transcription factor IIH (88, 89). This discovery led to the awareness that TFIIH is involved in two basic cellular processes, transcription and DNA repair. The complete composition of TFIIH was finally revealed in 2004, when the smallest TFIIH subunit, p8, was found to be an essential part of TFIIH and responsible for the repair syndrome TTD-A (84, 90). Besides p8, the core TFIIH complex consists of XPB, XPD, p62, p52, p44 and p34. XPD bridges the core complex with the CAK complex consisting of CDK7, cyclin H and MAT1 (ménage à trois 1).

I.3.1 The TFIIH complex

Early electron microscopy (EM) studies of the human and yeast TFIIH complexes revealed the structural composition of TFIIH and identified a donut-shaped structure for core TFIIH with an attached bulge harboring the bound CAK (91-93). The central hole of the donut was suggested to provide a passage for DNA. More detailed knowledge about the subunit composition of TFIIH was gained from an integrative approach combining previous EM maps with crosslinked assisted mass spectrometry data and further cryo-EM studies on the human and yeast pre-initiation complex (PIC), the assembly formed by RNA Pol II and the general transcription factors (94-96). Two recent cryo-EM studies on human TFIIH and yeast PIC including TFIIH provide closer insights into the subunit composition of TFIIH and reveal new details about interactions within TFIIH (97, 98). A model of the composition of TFIIH according to these results is shown in Figure I.3.

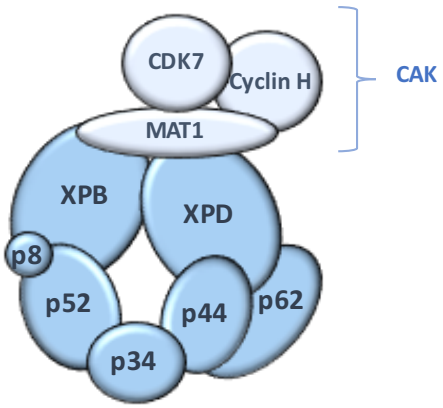


Figure I.3: Schematic composition of the TFIIH core complex and CAK. The TFIIH complex is composed of the two helicases XPB and XPD that are regulated by their partners p52/p8 and p44, respectively. Further core subunits are p62 and p34. The CAK complex consists of MAT1, CDK7 and Cyclin H.

The CAK complex is not required during NER, it even has an inhibitory effect on the NER activity and thus has to dissociate from core TFIIH before incision can take place (54, 99). The dissociation of the CAK complex from core TFIIH is dependent on the arrival of XPA to the damaged site (54). During transcription, the CAK complex stays incorporated into TFIIH and is necessary for the phosphorylation of the C-terminal domain of RNA Pol II. Apart from its role within TFIIH, the CAK complex also functions as a single complex during cell cycle control (100). XPD also assumes additional roles in the cell including roles in the regulation of the CAK activity, mitosis and chromosome segregation (101).

Core TFIIH comprises the two DNA-dependent helicases XPB and XPD. Both of them are superfamily 2 (SF2) helicases and contain two RecA like helicase domains (HD1 and HD2) that comprise seven conserved helicase motifs (102). The helicase activity of XPD is required for NER,

but is dispensable for transcription (43, 103, 104), whereas the ATPase activity of XPB is required for both NER and transcription (41, 104, 105). Interestingly, the helicase activity of XPB seems to be dispensable for both cellular processes and XPB is suggested to act rather as a translocase than as a helicase (106, 107). The activities of XPB and XPD are fine-tuned by their partners within TFIIH. The helicase activity of XPD is stimulated by the interaction with p44 and the ATPase activity of XPB is stimulated by the interaction with p52/p8 and thus dual incision during NER is only possible in the presence of an intact TFIIH complex (103, 105).

The natural product of a traditional Chinese medicinal plant, triptolide, was shown to display anti-proliferative and immunosuppressive activity and to target TFIIH (108). There is a strong indication that it targets XPB at residue C342 (109) and thus inhibits its DNA-dependent ATPase activity leading to TFIIH-dependent defects in transcription and DNA repair (108). In contrast to this finding, other studies showed that triptolide targets the CDK7 subunit of CAK, which leads to hyperphosphorylation and hence, degradation of the Rbp1 subunit of RNA Pol II during transcription (110, 111). Whether there is a connection between these two findings has not yet been investigated. However, it is clear that triptolide exhibits anticancer properties on a cellular level offering many therapeutic options that need to be investigated further.

I.3.2 The unconventional helicase XPB

XPB is a 3'-5' helicase that contains all seven conserved helicase motifs (Walker motifs I, Ia, II, III, IV, V and VI) and belongs to the SF2 helicases (83, 112). A crystal structure of an archaeal homolog of XPB from *Archaeoglobus fulgidus* depicted important structural features of XPB (113). Besides the two RecA like domains (HD1 and HD2), which contain the seven helicase motifs, a damage recognition domain (DRD), a unique R-E-D (arginine - glutamic acid - aspartic acid) motif and a flexible thumb-like motif (ThM) were identified (Figure I.4 A+B). Human XPB comprises additional N- and C-terminal extension domains (NTD and CTD) that were not present in this archaeal structure. The only crystal structure of human XPB that could be solved so far depicts the C-terminal half of XPB comprising HD2, ThM and approximately the first half of the CTD (114) (Figure I.4 C). A comparison of both structures shows that the orientation of HD1 and HD2 has to be different in human XPB to avoid clashes (114).

TFIIH displays only a 5'-3' helicase activity, which is based on the action of XPD, whereas the helicase activity of XPB seems to be dispensable for TFIIH actions (103). However, its ATPase activity is required to pursue NER, which is demonstrated by the severe effects of a mutation in the Walker A motif (K346R in human XPB) (41). Furthermore, it was shown that a reduced DNA-

dependent ATPase activity of XPB resulting in deficient NER was also obtained by mutations in either the R-E-D or the ThM motif, which clearly demonstrates the importance of these motifs for the function of XPB (41). In their model Oksenych *et al.* propose that binding of XPB to DNA stimulates its ATPase activity, which is used to bring the R-E-D and ThM motifs in close proximity of each other. This might be accomplished by a flip of HD2 and the formation of a closed complex, which seems to be required for anchoring TFIIH to the sites of the damage in NER (41).

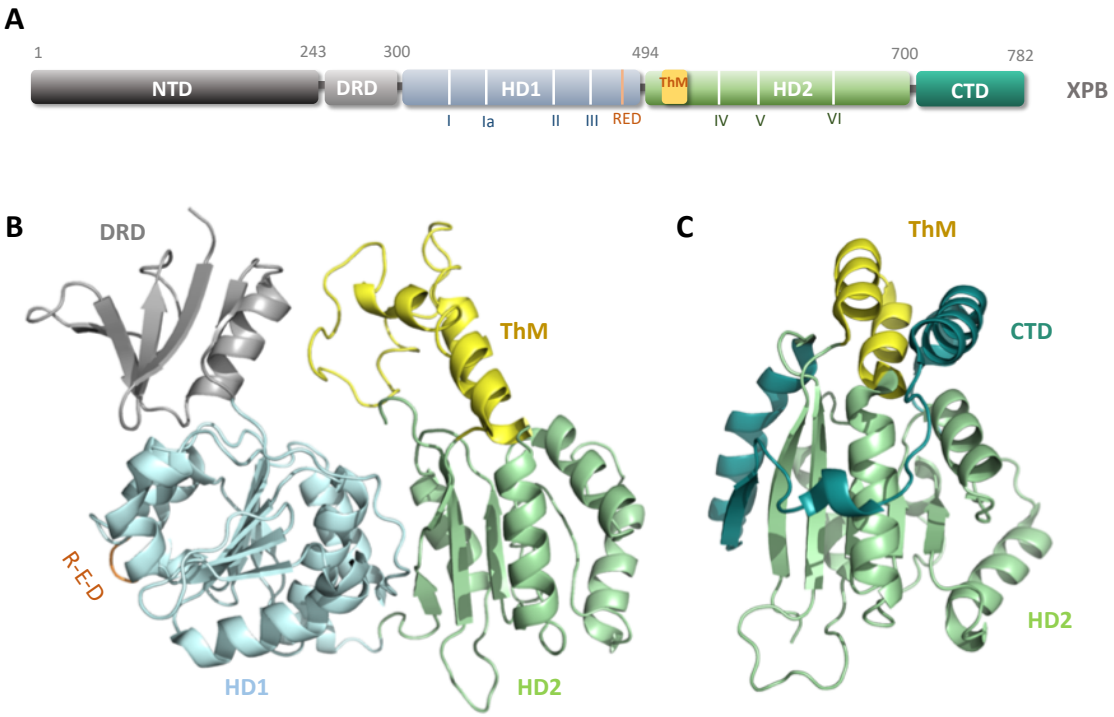


Figure I.4: Domain architecture and crystal structures of XPB. (A) Illustration of the domain architecture of human XPB, which is consisting of an N-terminal domain (NTD), a damage recognition domain (DRD), helicase domains 1 and 2 (HD1 and HD2) including the seven helicase motifs (I, Ia, II-VI), a R-E-D motif, a thumb-like motif (ThM) and a C-terminal domain (CTD). The grey numbers indicate the amino acids that approximately mark the borders of the individual domains. **(B)** Crystal structure of full-length XPB from *Archaeoglobus fulgidus* (PDB code: 2FWR) containing only the DRD (grey), HD1 (light blue), R-E-D (orange), ThM (yellow) and HD2 (light green). **(C)** Crystal structure of the C-terminal part of human XPB (PDB code: 4ERN) depicting HD2 (light green), ThM (yellow) and the N-terminal part of the CTD (dark green).

It is not yet clear how the functions of the two helicases XPB and XPD are coordinated during NER, but additional models have been suggested of how XPB might pursue its duty. Fan and DuPrez proposed a working hypothesis, in which XPB initiates unwinding of the dsDNA at the lesion and passes the duty on to XPD, which is a more productive helicase than XPB and fully opens the NER bubble (115). Recent data suggest that XPB acts as a translocase in a 5'-3' direction on the non-template strand (107). This model is in line with the previous one implying that XPB performs an initial unwinding of the DNA duplex by generating torsional stress in the constrained damaged

DNA. This initial unwinding would generate an optimal substrate for XPD thus facilitating binding of the helicase and subsequent damage verification through its unwinding activity (107).
A frameshift mutation found in XP/CS patients that alters the last 42 amino acids of XPB (83) was shown to reduce transcription and NER activity (112). A phosphorylation site at residue S751 is located in this altered C-terminal region of XPB, which requires dephoshorylation prior to 5' incision by ERCC1-XPF (116). The frameshift mutation might impair this regulatory step of NER and putative interactions between XPB and ERCC1-XPF.

The ATPase activity of XPB is not only required for NER, but also for transcription. The K346R mutation in the Walker A motif of human XPB prevents transcription, whereas a mutation in the ATP-binding site of XPD (K48R in human XPD) still displays transcriptional activity (41, 104). In another study, it could be shown that RNA Pol II promoter opening relies on the ATPase activity of XPB, which is likely to induce a conformational change in XPB (117). Interestingly, the R-E-D and ThM motifs of XPB are as important for transcription as for NER suggesting a similar mode of action for XPB in both processes (41). Several studies investigated the structural incorporation of TFIIH within the PIC (95-98, 118-120). In eukaryotes, the transcription of DNA genes into mRNA requires the assembly of RNA Pol II and a set of general transcription factors near the transcription start site (TSS) to form the PIC, which is stabilized through a co-activator complex, the Mediator. These general transcription factors include TFIIA, TFIIB, TFIID, TFIIF, TFIIE and TFIIH (121). Based on a cryo-EM structure of the PIC-core Mediator complex including TFIIH bound to a DNA duplex Schilbach *et al.* propose a model for the translocase activity of XPB, in which XPB actively pushes the dsDNA into the RNA Pol II cleft thereby leading to an opening of the DNA (98).

I.3.3 The interaction network between XPB, p52 and p8

Within core TFIIH, XPB is interacting with XPD, p52 and p8 (122-124). Furthermore, p52 and p8 are also interacting with each other (125). This interaction was captured in a crystal structure comprising the yeast homologs of the C-terminal domain of p52 (p52_CTD) and p8 (126) (Figure I.5 A+B). Human p8 contains only 71 amino acids and is by far the smallest subunit of TFIIH. Interestingly, p8 and p52_CTD adopt the same fold and form a pseudosymmetric heterodimer (126) (Figure I.5 B). Both p8 and p52_CTD as well as full-length p52 and an N-terminal variant of p52 form homodimers when expressed on their own (126, 127). The p52_CTD/p8 heterodimer resembles the p8 homodimer that was solved by NMR (Figure I.5 C) as both interactions are mediated by a β-strand addition involving mainly β-strand 1 of both subunits (126, 128) (Figure

I.5 C). Mutations that alter the interaction between p52_CTD and p8 were shown to have severe effects on NER (126). The lack of an interaction between p52_CTD and p8 results in the solvent exposure of the hydrophobic surface of p52_CTD and most likely causes a reduced stability of p52. This could explain the defects observed in patients suffering from TTD due to a mutation in the *TTDA* gene (coding for p8) that alters the start codon and thus, leads to either no p8 protein at all or a shortened p8 protein lacking β-strand 1 (125, 126, 128).

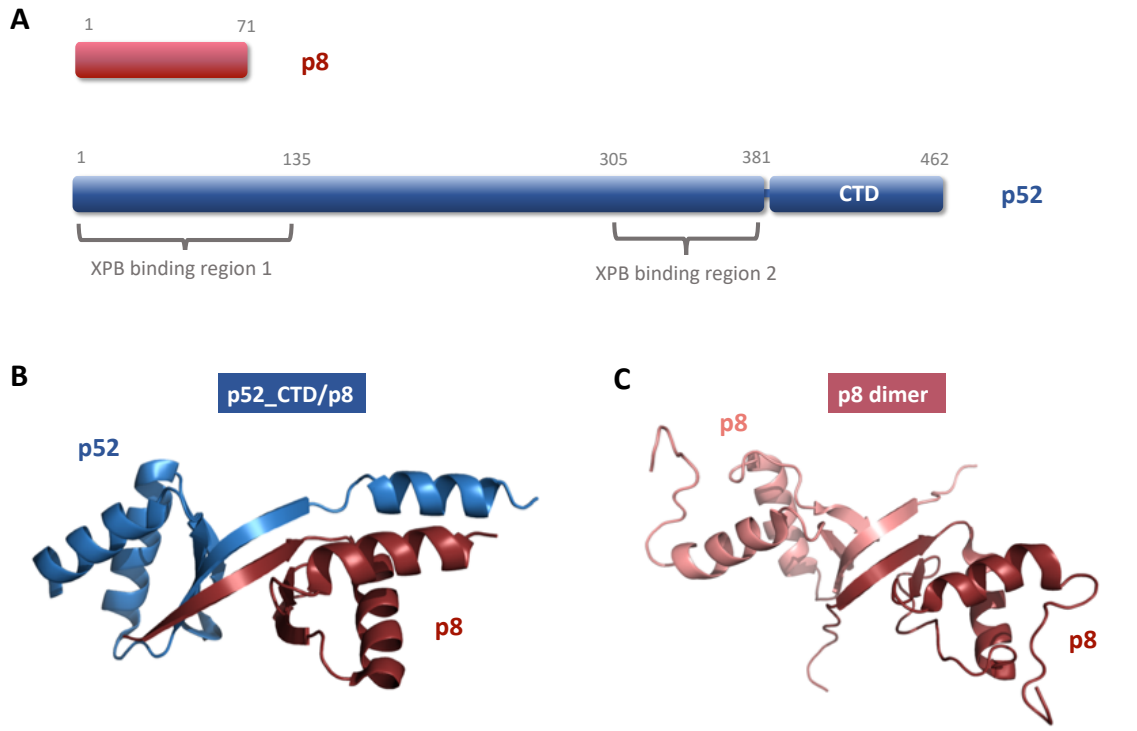


Figure I.5: Domain architecture of p8 and p52 and structures of p52_CTD/p8 and a p8 dimer. (A) Illustration of the domain architecture of human p8 and p52. P52 is known to contain a C-terminal domain (CTD), which is interacting with p8. In addition, there are two proposed XPB binding sites in p52, which are annotated. The grey numbers indicate the amino acids that approximately classify the individual domains. **(B)** Crystal structure of the heterodimeric complex formed by the yeast homologs of p8 (red) and p52_CTD (blue) (PDB code: 3DOM). **(C)** NMR structure of a p8 dimer (PDB code: 2JNJ).

The absence of p8 causes an overall reduced concentration of TFIIH and a reduced NER activity of TFIIH but has no effect on its transcription activity (124, 125, 129). The reduced NER activity can be explained by the fact that p8 contributes to the stimulation of the ATPase activity of XPB, which is crucial for NER (124). Furthermore, p8 seems to be required for the recruitment of XPA to the site of the damage (124). Interestingly, p8 is not only found in the nucleus of cells in association with TFIIH, but also as an unbound fraction that shuttles between the cytoplasm and the nucleus (130). Upon DNA damage the free dimerized p8 shuttling between the cytoplasm and

the nucleus increasingly associates with TFIIH to form a heterodimer with p52_CTD and to pursue NER (128, 130).

Except for the crystal structure covering the C-terminal domain of yeast p52 no precise structural knowledge of p52 is available. As p52 was shown to be important for anchoring XPB within TFIIH more detailed studies were pursued to investigate the domains that underlie this interaction (123). Co-infected baculovirus extracts expressing different truncated variants of human p52 were analyzed for their ability to interact with XPB by co-immunopurification and led to the identification of two XPB binding regions in p52 (123). The first XPB binding region in p52 comprises residues 1 - 135 and the second one residues 305 - 381 (Figure I.5 A). A p52 variant lacking parts of the second XPB binding region (p52_1-358 in human p52) was shown to be inactive in NER activity and only slightly active in transcription. Further analyses revealed that the N-terminus of XPB (residues 44 - 208 in human XPB) is required for the interaction with p52 (123). P52 was shown to stimulate the DNA-dependent ATPase activity of XPB, which is indispensable for both NER and transcription (41, 105). The mutation F99S found in the *XPB* gene of XP/CS patients weakens the interaction between XPB and p52 and results in a highly reduced ATPase activity of XPB within TFIIH comparable to a mutation in its Walker A motif (105). Further investigations by Coin *et al.* showed that the ATPase activity of XPB in the absence of p52 is not affected by this mutation highlighting the crucial stimulating role p52 assumes towards XPB's ATPase activity within TFIIH. Only the F99S mutation, but not the T119P mutation found in the *XPB* gene of TTD patients, leads to defects in damaged DNA opening. Interestingly, transcription was not much affected by either of the two mutations. An investigation of the two proposed XPB binding sites in p52 revealed that only the second proposed binding site in p52 was required for stimulating the ATPase activity of XPB (105).

As there are no known patients with mutations in any TFIIH subunits aside from XPB, XPD or p8, mutations in p52 that effect the interaction between XPB and p52 are only known from further genetic or biochemical studies. One of these studies is based on the analysis of the *Drosophila melanogaster* homolog of p52 (131). Interestingly, some of the defects generated by mutations in the p52 homolog in *Drosophila melanogaster* resembled phenotypes of patients with XP, CS or TTD. The drosophila mutations were also introduced into the human *p52* gene and the single mutation p52_E310K (human p52) as wells as the double mutation p52_E310K/R314E (human p52) were shown to impair binding to XPB and to reduce its ATPase activity. The double mutant further led to reduced NER and transcription activities (131). Smurnyy *et al.* investigated the identification of drug-resistant variants utilizing triptolide as a drug and XPB as its target (132). The resulting resistant clones did not only carry mutations in XPB, but also in p52 and hence two

mutations in p52 (R337L and L440P in human p52) have been identified that were proposed to inhibit an interaction with XPB.

I.3.4 The p34/p44 pair

The p44 subunit of TFIIH is known for its pivotal role in stimulating XPD's helicase activity, which is required for NER (103, 133). The domain architectures of p44 and p34 (Figure I.6 A) resemble each other indicating the possibility that they have been derived from gene duplication (94). Both proteins contain an N-terminal *von Willebrand factor A* like (vWA) domain, followed by a C4 zinc-binding domain. P44 contains an additional C-terminal C4C4 RING finger domain, which binds two zinc ions (134, 135). Each of the three domains of p44 is known to interact with a neighboring subunit within TFIIH.

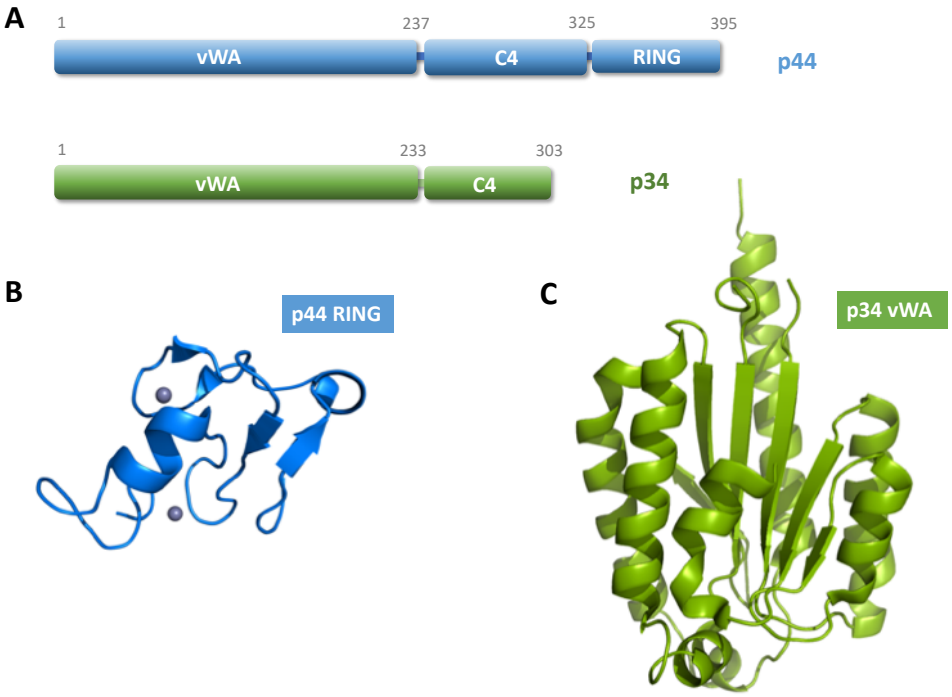


Figure I.6: Domain architecture of p44 and p34 and structures of p44 RING and p34 vWA domains. (A) Illustration of the domain architecture of human p44 and p34 both containing a *von Willebrand factor A* like (vWA) domain and a C4 zinc binding domain. P44 additionally contains a C4C4 RING domain. The grey numbers indicate the amino acids that approximately define the borders of the individual domains. **(B)** NMR structure of the human p44 RING domain (PDB code: 1Z60). The two bound zinc ions are depicted as spheres. **(C)** Crystal structure of the p34 vWA domain from *Chaetomium thermophilum* (PDB code: 4PN7).

The vWA domain of p44 is interacting with the HD2 and CTD domains of XPD and stimulates its helicase activity (103, 133, 136, 137). The importance of this interaction is reflected in various mutations that were found in the *XPD* gene of XP, XP/CS and TTD patients and shown to impair

the interaction between XPD and p44 (136). The crystal structure of the vWA domain of the yeast homolog of p44 was solved, which identified that the β4-α5 loop of p44 is essential for the interaction with XPD (137). The C4 domain of p44 interacts with the TFIIH subunit p62, an interaction that was shown to be important for the stability of TFIIH (122, 138). The p62 subunit was shown to interact with many additional partners among them XPC, XPG and TFIIE (139-141). In addition to the interactions with XPD and p62, the third domain of p44, the C4C4 RING domain, interacts with p34 (135, 142, 143). The RING domain was shown to be important for promoter escape during transcription (138) and a yeast homolog of this domain was further suggested to possess an E3 ubiquitin ligase activity (144).

Based on previous studies that depicted an interaction between p34 and p44 (122), co-expression attempts by Fribourg *et al.* were pursued to verify this interaction and to identify the underlying domains of p34 and p44 (142). In this study, the p34 vWA domain and the p44 RING domain were identified as the required domains within p34 and p44 to permit complex formation. The minimal subunits were co-expressed and were sufficient to form a stable interaction. An NMR structure of the human RING domain of p44 was solved (Figure I.6 B) and provided first insights into the overall architecture of this domain. The structure revealed that the first zinc binding site is required for a proper fold of this domain and its interaction with the p34 vWA domain (135). A crystal structure of the vWA domain of p34 from *Cheatomium thermophilum* (Figure I.6 C) showed that it assumes the typical vWA like architecture and resembles the structure of the von Willebrand factor A1 domain (143, 145). Based on these two structures putative binding sites for the interaction between p34 and p44 could be modelled. Schmitt *et al.* suggested that the first zinc binding site and helix α2 of p44 form contacts with the hydrophobic region around α-helices 5 to 7 and β-strand 3 of p34 (143). The function of the C4 domain of p34 still remains unclear as well as the role of p34 for NER and transcription.

I.4 The Model Organism *Chaetomium thermophilum*

We decided to investigate the TFIIH subunits from the model organism *Chaetomium thermophilum* (*C. thermophilum*; ct). The thermophilic, filamentous ascomycete *C. thermophilum* is frequently found in soil, dung, rotting plants and composting piles. Because of its ability to thrive at temperatures up to 60°C, the proteins of this thermophilic fungus are thought to be very stable and thus ideal for structural investigations (146). The genome of *C. thermophilum* was

sequenced in 2011 and the proteome of *C. thermophilum* was shown to have 73% homology to the one of *Saccharomyces cerevisiae* (146). A further study confirmed that the proteome of *C. thermophilum* is similarly composed to the ones of mesophilic organisms (147). *C. thermophilum* contains all subunits of core TFIIH and the CAK. The following Table displays the homology and identity of the core TFIIH subunits from *C. thermophilum* to those of their human counterparts.

Table I.1: Identity and homology of core TFIIH subunits from *C. thermophilum* and human.

TFIIH	**Identical***		**Homolog***	
subunit	Number of residues	Percentage	Number of residues	Percentage
XPD	392/737	53%	538/737	72%
XPB	371/720	52%	519/720	72%
p62	88/369	24%	154/369	41%
p52	172/497	35%	261/497	52%
p44	146/361	40%	201/361	55%
p34	108/380	28%	158/380	41%
p8	23/57	40%	41/57	71%

*All values were derived from a Basic Local Alignment Search Tool (BLAST) search from the National Center for Biotechnology Information (NCBI).

I.5 Aim of the Work

Although numerous studies have investigated TFIIH on a structural and biochemical basis, still many structural details and further information on interactions and functions of individual TFIIH subunits are missing. Examples are the structural characterization of full-length p52 and the characterization of its interaction with XPB. Although structural knowledge on p34 and p44 is partially available, the functional role of this interaction for TFIIH remains unknown. The aim of this work was to elucidate the structure of individual subunits and subunit complexes of TFIIH and to analyze their interaction and function. Of particular interest were the characterization of the p34/p44 interaction and the characterization of the XPB/p52/p8 network and its implication on the function of XPB.

The structure of a minimal complex between the p34 vWA domain and the p44 RING domain could be solved in previous works, which are depicted in the doctoral thesis of Dominik Schmitt and the Master thesis of Elisabeth Schönwetter (148). However, the characterization of this

interaction and its implication on the TFIIH complex were not known. Thus, one aim of this work was to analyze the interaction between p34 and p44 based on previous findings.

The second aim of this work was to provide the first complete structural characterization of the p52 subunit. Furthermore, the interaction of p52 with XPB was intended to be analyzed to identify the domains of p52 and XPB that are essential for this interaction. The implication of the interactions between XPB, p52 and p8 on the function of XPB were intended to be investigated in detail.

II Materials and Methods

II.1 Materials

II.1.1 Chemicals

All chemicals used during this work were purchased at analytical grade or better. Reagent-grade water provided by a TKA GenPure system was used for the preparation of all buffers and solutions. The following Table enlists all chemicals used during this work including abbreviations or molecular formulas if substances will be named according to these subsequently. Chemicals used for the in-house production of crystallization screens by the LISSY 2002 liquid handing robot are not included in the Table.

Substance	**Abbreviation/ molecular formula**	**Supplier**
2-(*N*-morpholino)-ethanesulfonic acid	MES	Sigma-Aldrich
2-Propanol		Carl Roth
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	HEPES	Carl Roth
Acetic Acid		Carl Roth
Ampicillin sodium salt		Carl Roth
Ammonium acetate		Sigma-Aldrich
Ammonium chloride	NH_4Cl	Sigma-Aldrich
Ammonium persulfate	APS	Carl Roth
Brillant Blue R-250		Carl Roth
Bromophenol blue sodium salt		Carl Roth
Calcium acetate		Sigma-Aldrich
Calcium chloride dihydrate	$CaCl_2$	Carl Roth
Chloramphenicol		Carl Roth
Citric acid monohydrate		Sigma-Aldrich
D-(+)-Biotin		Carl Roth
Dimethyl sulfoxide	DMSO	Carl Roth
Disodium hydrogen phosphate	Na_2HP0_4	Carl Roth
Dithiothreitol	DTT	Carl Roth
Ethylenediaminetetraacetic acid	EDTA	Carl Roth
Ethanol		Carl Roth
Glucose		Carl Roth
Glycerol		Carl Roth
Glycine		Carl Roth

Hydrochloric acid	HCl	Carl Roth
Imidazole		Carl Roth
Iodine	I_2	Fluka/Sigma-Aldrich
Iodoacetamide		Sigma-Aldrich
Iron(II) sulfate heptahydrate	$Fe(II)SO_4$	Carl Roth
Isopropyl-β-D-thiogalactopyranoside	IPTG	Carl Roth
Kanamycin sulfate		Carl Roth
L-(+)-Arabinose		Carl Roth
Lithium citrate tetrahydrate		Fluka/Sigma-Aldrich
L-Isoleucine	Iso/I	Fluka/Sigma-Aldrich
L-Leucine	Leu/L	Carl Roth
L-Lysine	Lys/K	Fluka/Sigma-Aldrich
L-Phenylalanine	Phe/F	Fluka/Sigma-Aldrich
L-Threonine	Thr/T	Fluka/Sigma-Aldrich
L-Valine	Val/V	Fluka/Sigma-Aldrich
Magnesium chloride hexahydrate	$MgCl_2$	Carl Roth
Magnesium acetate tetrahydrate		Sigma-Aldrich
Magnesium sulfate	$MgSO_4$	Sigma-Aldrich
Monopotassium phosphate	KH_2PO_4	Carl Roth
N-Cyclohexyl-2-aminoethanesulfonic acid	CHES	Carl Roth
Phenylmethylsulfonyl fluoride	PMSF	Carl Roth
Polyethylene glycol 400	PEG 400	Sigma-Aldrich
Polyethylene glycol 3350	PEG 3350	Sigma-Aldrich
Polyethylene glycol 4000	PEG 4000	Sigma-Aldrich
Polyethylene glycol 8000	PEG 8000	Fluka/Sigma-Aldrich
Ponceau S		Carl Roth
Potassium chloride	KCl	Carl Roth
Potassium dihydrogen phosphate	KH_2PO_4	Carl Roth
Potassium iodide	KI	Sigma-Aldrich
Seleno methionine	Se-Met	Acros Organics
Sodium bromide	NaBr	Fluka/Sigma-Aldrich
Sodium cacodylate trihydrate		Sigma-Aldrich
Sodium chloride	NaCl	Carl Roth
Sodium dodecyl sulfate	SDS	Carl Roth
Sodium dihydrogen phosphate	NaH_2PO_4	Carl Roth
Sodium hydroxide	NaOH	Carl Roth
SYPRO® Orange		Thermo Fisher Scientific
Tetramethylethylenediamin	TEMED	Carl Roth
Trimethylamine *N*-oxide	TMAO	Sigma-Aldrich
Tris[2-carboxyethyl]-phosphine	TCEP	Carl Roth
Tris[hydroxymethyl]-aminomethane	Tris	Carl Roth
Urea		Carl Roth

II.1.2 General consumables

Manufactured composites

Enzyme	Type	Supplier
DNase I	Deoxyrionuclease	AppliChem
Trypsin	Restriction endonuclease	Carl Roth or in-house production
Tobacco Etch Virus (TEV) protease	TEV cysteine protease	In-house production

Name	Type	Supplier
Bayer silicon grease medium viscosity	Silicon grease	Jena Bioscience
PageRuler™ Plus Prestained Protein Ladder	Protein molecular-weight size marker	Thermo Fisher Scientific
Rotiphorese® Gel 30 (37.5:1)	30% acrylamide/bisacrylamide solution	Carl Roth

Disposable labware

Type	Model	Supplier
24-well crystallization plate	Crystalgen SuperClear™ Plate	Jena Bioscience
96-well crystallization plate	Crystalquick™ 1 square well, flat bottom, low profile	Greiner Bio-One
96-well thin-wall PCR plate	Microplate 96 well, for PCR	Greiner Bio-One
Cannula	Sterican® single-use cannula – various sizes	B. Braun
Centrifugal concentrator	Amicon® Ultra – 0.5, 4 and 15 ml	Millipore
Centrifuge tube	Cellstar® centrifuge tube – 15 and 50 ml	Greiner Bio-One
Cover slides	Circular cover slides – siliconized, Ø 22 mm	Jena Bioscience
Cuvette	Rotilabo®-single-use cells – 1.6 ml	Carl Roth
Dialysis tube	D-Tube™ Dialyzer – Mini, Midi	Novagen
Filter paper	Filter paper	Sartorius
Gloves	Nitril gloves	Star Lab
Optical quality sealing foil	VIEWseal™	Greiner Bio-One
Parafilm	Parafilm®	Bemis
PCR tubes	PCR reaction tubes – 0.2 ml	Sarstedt
Petri dishes	Petri dishes for agar plates	
Pipette tips	Pipette tips – 10, 200, 1000 and 5000 µl	Mettler-Toledo
Reaction tubes	Reaction tubes – 0.5, 1.5 and 2 ml	Sarstedt
Serological pipette	Serological pipette – 5, 10 and 25 ml	Sarstedt
Sterile filter	Acrodisc® sterile filter for syringe – 0.22 and 0.45 µm	Pall
Syringes	Omnifix® syringes – 1, 5, 10 and 20 ml	B. Braun

II.1.3 Molecular biology consumables

Bacterial strains

Name	Genotype	Supplier
E. coli DH5α™	F^- ϕ80 *lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(r_K^- m_K^+) *pho*A *sup*E44 λ^- *thi*-1 *gyr*A96 *rel*A1	Invitrogen
E. coli BL21-CodonPlus® (DE3)-RIL	B F^- *omp*T *hsd*S(r_B^- m_B^-) *dcm*$^+$ Tet^r *gal* λ(DE3) *end*A Hte[*arg*U *ile*Y *leu*W Cam^r]	Stratagene

Enzymes, buffer and others reagents

Enzyme	Type	Supplier
Dpn I	Restriction endonuclease	New England BioLabs
Phusion®	High-Fidelity DNA polymerase	Thermo Fisher Scientific
T4 DNA Polymerase	DNA polymerase	New England BioLabs

Name	Type	Supplier
2'-Deoxyadenosine 5'-triphosphate (dATP), sodium salt solution	Deoxynucleotide	Thermo Fisher Scientific
2'-Deoxycytidine 5'-triphosphate (dCTP), sodium salt solution	Deoxynucleotide	Jena Biosciences
2'-Deoxyguanosine 5'-triphosphate (dGTP), sodium salt solution	Deoxynucleotide	Thermo Fisher Scientific
2'-Deoxythymidine 5'-triphosphate (dTTP), sodium salt solution	Deoxynucleotide	Thermo Fisher Scientific
Agar	Agar-agar	Carl Roth
Bovine Serum Albumin (BSA)	Carrier protein	New England BioLabs
GC buffer	DNA polymerase reaction buffer	Thermo Fisher Scientific
HF buffer	DNA polymerase reaction buffer	Thermo Fisher Scientific
Lysogeny broth (LB) medium (Lennox)	Standard *E. coli* culture medium	Carl Roth
NEBuffer™ 2, NEBuffer™ 2.1	Restriction endonuclease reaction buffer	New England BioLabs
RecA	Single-stranded DNA binding protein	New England BioLabs
T4 ligation buffer	DNA ligase reaction buffer	New England BioLabs

Plasmids

Plasmid	Resistance	Description	Induction by	Supplier
pBADM-11	Ampicillin	*E. coli* expression vector: araBAD promotor; N-terminal His_6-Tag sequence; N-terminal TEV cleavage site	L-(+)-Arabinose	A. Geerlof; European Molecular Biology Laboratory (EMBL)
pETM-11	Kanamycin	*E. coli* expression vector: T7 promotor; N- and C-terminal His_6-Tag sequence; N-terminal TEV cleavage site	IPTG	G. Stier; EMBL

Constructs

The following Table enlists all constructs that were already generated prior to this work, but were used in the course of this thesis. This comprises constructs that were produced by the Kisker group members Agnes Elias (AE), Dominik Schmitt (DRS), Wolfgang Kölmel (WK) or constructs that were produced during the Master thesis of Elisabeth Schönwetter (ES-MT). A list of all newly cloned gene products produced in the course of this thesis is provided in VI.2.

Construct	Synonym	Plasmid	Reference
p8		pBADM-11	AE
p8_noTag		pBADM-11	ES-MT
p34		pETM-11	AE
p34		pBADM-11	DRS
p34_noTag		pBADM-11	DRS
p34_1-277	p34_vWA	pBADM-11	DRS
p34_1-277_K155E	p34_vWA_K155E	pBADM-11	DRS
p44		pETM-11	AE
p44		pBADM-11	WK
p44_368-534	p44_RING	pETM-11	DRS
p44_368-534_F490E	p44_RING_F490E	pETM-11	DRS
p44_368-534_noTag	p44_RING_noTag	pETM-11	DRS
p44_368-534_Δloop_noTag		pETM-11	ES-MT
p52		pETM-11	AE
p52_noTag		pETM-11	ES-MT
p52_1-120		pETM-11	ES-MT
p52_1-120_noTag		pETM-11	ES-MT
p52_121-End_dLinker	p52_121-EdL	pETM-11	WK
p52_350-435		pETM-11	ES-MT
p52_350-435_noTag		pETM-11	ES-MT

p52_350-End		pETM-11	ES-MT
p62		pETM-11	WK
p62_noTag		pETM-11	WK
XPB_noTag		pBADM-11	WK
XPB_1-345		pBADM-11	WK
XPB_60-345		pBADM-11	WK
XPB_116-245		pBADM-11	WK
XPB_116-345		pBADM-11	WK

Primers

All primers were ordered from Sigma-Aldrich.

Primer Name*	**Sequence (5`-3`)**
T7	TAATACGACTCACTATAGGG
T7-term	CTAGTTATTGCTCAGCGGT
vector_lin_F	GTCGACAAGCTTGCGGCCG
vector_lin_R	GGCGCCCTGAAAATAAAGATTC
p52_inside_primer_F	GACGGTGGTAGATCAGATC
primer_vector_end_F	GCTCGAGAGATCCGGCTGC
p8_lin_F	GAATGAGCGGATGCGGACGAGCGTCAGGGCTATTCGCGG
p8_lin_R	CGGCCGCAAGCTTGTCGACCTAGTCACTGTCCGCAAGGG
p34_noTag_F	CTAACAGGAGGAATTAACCATGAGCGCCCAGGACGCCGTCG
p34_noTag_R	CGACGGCGTCCTGGGCGCTCATGGTTAATTCCTCCTGTTAG
p34_K155E(GAG)_F	CGCCCTCGCTCACATTAACGAGACTGCCCTCTCCCTGACCG
p34_K155E(GAG)_R	CGGTCAGGGAGAGGGCAGTCTCGTTAATGTGAGCGAGGGCG
p34_A151E_F	GGCCTTGACTCTCGCCCTCGAACACATTAACAAGACTGCCC
p34_A151E_R	GGGCAGTCTTGTTAATGTGTTCGAGGGCGAGAGTCAAGGCC
p34_18-x_F	CTTTATTTTCAGGGCGCCACGGACGACATCCCGTCGCTGC
p34_18-x_R	GCAGCGACGGGATGTCGTCCGTGGCGCCCTGAAAATAAAG
p34_18-x_noTag_F	CTAACAGGAGGAATTAACCATGACGGACGACATCCCGTCGCTGC
p34_18-x_noTag_R	GCAGCGACGGGATGTCGTCCGTCATGGTTAATTCCTCCTGTTAG
p34_278-x_F	CTTTATTTTCAGGGCGCCAGTGCGCCATCGAACCCTGCTGC
p34_278-x_R	GCAGCAGGGTTCGATGGCGCACTGGCGCCCTGAAAATAAAG
p34_x-274_F	GTACCTCATGTTCGGCTTTTAACTCGAGAGCTTGGCTGTTTTGG
p34_x-274_R	CCAAAACAGCCAAGCTCTCGAGTTAAAAGCCGAACATGAGGTAC
p44_F490E_pETM11_F	CTTTATTTTCAGGGCGCCATGGCTGATTCCGATGGCG
p44_F490E_pETM11_R	CGGCCGCAAGCTTGTCGACTTAATCCAGGACCATTGCG
p52_noTag_F	GAAGGAGATATACCATGTCGATCCCCGCCG
p52_noTag_R	CGGCGGGGATCGACATGGTATATCTCCTTC
p52_CtermHis_pETM11_F	GAGTCGGAAGAAGGAGGGTGTCGACAAGCTTGCGGCCG
p52_CtermHis_pETM11_R	CGGCCGCAAGCTTGTCGACACCCTCCTTCTTCCGACTC
p52_dLinker2_F	CCAAGCAAAAGCCGGCTCGAACGGCAACGGCTCGAAGCCAGATAGC
p52_dLinker2_R	GCTATCTGGCTTCGAGCCGTTGCCGTTCGAGCCGGCTTTTGCTTGG
p52_235-x_F	GCCGCCGACCAAGCAAAATAAGTCGACAAGCTTGCGGCCG

p52_235-x_R	CGGCCGCAAGCTTGTCGACTTATTTTGCTTGGTCGGCGGC
p52_x-321_SdM_F	CGCTAACCTCGTCCGCATCCTAACTGCGTTCCGTTTCCTCCG
p52_x-321_SdM_R	CGGAGGAAACGGAACGCAGTTAGGATGCGGACGAGGTTAGCG
p52_x-349_F	CCCTGGGGGGAGCCGACCCCTAAGTCGACAAGCTTGCGGCCG
p52_x-349_R	CGGCCGCAAGCTTGTCGACTTAGGGGTCGGCTCCCCCCAGGG
p52_x-454_F	GGCAGTTGGAGAATGAGCGGTAAGTCGACAAGCTTGCGGCC
p52_x-454_R	GGCCGCAAGCTTGTCGACTTACCGCTCATTCTCCAACTGCC
p52_19-x_F	CTTTATTTTCAGGGCGCCCTGCCGGGCACCACCTTTAG
p52_19-x_R	CTAAAGGTGGTGCCCGGCAGGGCGCCCTGAAAATAAAG
p52_noTag_19-x_F	CTTTAAGAAGGAGATATACCATG CTGCCGGGCACCACCTTTAG
p52_noTag_19-x_R	CTAAAGGTGGTGCCCGGCAGCATGGTATATCTCCTTCTTAAAG
p52_121-x_F	CTTTATTTTCAGGGCGCCGCCGCCCACAACAGCTTCGG
p52_121-x_R	CCGAAGCTGTTGTGGGCGGCGGCGCCCTGAAAATAAAG
p52_noTag_121-x_F	GAAGGAGATATACCATGGCCGCCCACAACAGCTTCGG
p52_noTag_121-x_R	CCGAAGCTGTTGTGGGCGGCCATGGTATATCTCCTTC
p52_121-EdL-TEV_nT_F	GAAGGAGATATACCATGATCCCCACTACTGAGAATC
p52_121-EdL-TEV_nT_R	GATTCTCAGTAGTGGGGATCATGGTATATCTCCTTC
p52_noTag_248-x_F	CTTTAAGAAGGAGATATACCATG AATGCCCCCCCGACCAAG
p52_noTag_248-x_R	CTTGGTCGGGGGGGCATTCATGGTATATCTCCTTCTTAAAG
p52_248-x_F	CTTTATTTTCAGGGCGCC AATGCCCCCCCGACCAAG
p52_248-x_R	CTTGGTCGGGGGGGCATTGGCGCCCTGAAAATAAAG
p52_x-465_F	GCCCGGGGTTCTTGTTCAAG TAAGTCGACAAGCTTGCGGCCG
p52_x-465_R	CGGCCGCAAGCTTGTCGACTTACTTGAACAAGAACCCCGGGC
p52_x-421_F	CCTACCTCGCCAGCCACGCACATTAAGTCGACAAGCTTGCGGC
p52_x-421_R	GCCGCAAGCTTGTCGACTTAATGTGCGTGGCTGGCGAGGTAGG
p52_x-458_F	GAGCGGATGCGGACGAGCTAAGTCGACAAGCTTGCGGCCG
p52_x-458_R	CGGCCGCAAGCTTGTCGACTTAGCTCGTCCGCATCCGCTC
p52_458_CtermHis_F	GAGCGGATGCGGACGAGCGTCGACAAGCTTGCGGCCG
p52_458_CtermHis_R	CGGCCGCAAGCTTGTCGACGCTCGTCCGCATCCGCTC
p52_458_lin_F	CCCTTGCGGACAGTGACTAGGTCGACAAGCTTGCGGCCG
p52_458_lin_R	CCGCGAATAGCCCTGACGCTCGTCCGCATCCGCTCATTC
p52_345-x_F	CTTTATTTTCAGGGCGCCGGGGGAGCCGACCCCTCCGC
p52_345-x_R	GCGGAGGGGTCGGCTCCCCCGGCGCCCTGAAAATAAAG
p52_noTag_345-x_F	GAAGGAGATATACCATGGGGGGAGCCGACCCCTCCGCGC
p52_noTag_345-x_R	GCGCGGAGGGGTCGGCTCCCCCCATGGTATATCTCCTTC
p52_x-422_F	CCTACCTCGCCAGCCACGCACATTAAGTCGACAAGCTTGCGGC
p52_x-422_R	GCCGCAAGCTTGTCGACTTAATGTGCGTGGCTGGCGAGGTAGG
p52_x-453_F	GCTGTGGCAGTTGGAGAATGAGTAAGTCGACAAGCTTGCGGCC
p52_x-453_R	GGCCGCAAGCTTGTCGACTTACTCATTCTCCAACTGCCACAGC
p52_E359K_QC_F	CAAGGGCTCCATCATCATCAAGACGAACTACCGCCTCTAC
p52_E359K_QC_R	GTAGAGGCGGTAGTTCGTCTTGATGATGATGGAGCCCTTG
p52_R363E_QC_F	CATCATCGAGACGAACTACGAACTCTACGCCTACACCTCCTC
p52_R363E_QC_R	GAGGAGGTGTAGGCGTAGAGTTCGTAGTTCGTCTCGATGATG
p52_R386L_QC-F	CTTTACTCACCTCAACATGCTCTTTGCCGGCATGGTGACAG
p52_R386L_QC-R	CTGTCACCATGCCGGCAAAGAGCATGTTGAGGTGAGTAAAG
XPB_x-249_F	GGGAACTGACACGACCTAAGTCGACAAGCTTGCGGCCG

XPB_x-249_R	CGGCCGCAAGCTTGTCGACTTAGGTCGTGTCAGTTCCC
XPB_x-270_F	CGCCGCTGGCGTCCGGTAAGTCGACAAGCTTGCGGCCG
XPB_x-270_R	CGGCCGCAAGCTTGTCGACTTACCGGACGCCAGCGGCG
XPB_x-284_F	CGGTTGCCGAAAAGAAGCCCTAAGTCGACAAGCTTGCGGCCG
XPB_x-284_R	CGGCCGCAAGCTTGTCGACTTAGGGCTTCTTTTCGGCAACCG

*F = forward primer; R = reverse primer

II.1.4 Kits and screens

Kits

Kit name	Purpose	Supplier
JBS Floppy Choppy	Limited proteolysis or *in situ* proteolysis; contains the following proteases: α-chymotrypsin, papain, subtilisin A, trypsin	Jena Biosciences
JBScreen Heavy	Heavy metal soaking; used substances: lead(II) acetate trihydrate, mercury(II) acetate, potassium dicyanoaurate(I), potassium tetranitroplatinate(II)	Jena Biosciences
NucleoSpin® Gel and PCR Clean-up	Purification of PCR products	MACHEREY-NAGEL
NulceoSpin® Plasmid	Isolation of plasmid DNA from *E. coli*	MACHEREY-NAGEL

ThermoFluor screens

Screen	Reference
ThermoFluor Standard*	Ericsson *et al*., 2006 (149)
ThermoFluor Advanced*	Ericsson *et al*., 2006 (149)

*Compositions of ThermoFluor screens are given in VI.3.

Crystallization screens

Screen	Supplier
$AmSO_4$ Suite	Qiagen
Crystal Screen™ 1 + 2	Hampton Research
Index	Hampton Research
JCSG+	Molecular Dimensions

MbClass II Suite	Qiagen
Nextal -PEG Suite	Qiagen
Nextal - PEG Suite II	Qiagen
Nextal pH Clear	Qiagen
Nextal pH Clear II	Qiagen
Nucleix Suite	Qiagen
Protein Complex Suite	Qiagen
Topaz™ OptiMix 3	Fluidigm
Topaz™ OptiMix PEG	Fluidigm
Wizard 1 + 2	Emerald
Wizard 3 + 4	Emerald
96-well fine screen 10 - 34	self-designed
24-well fine screen 1 - 10	self-designed

II.1.5 Peptides

Name	Sequence	Supplier
XPB-18	DAERLRHYDAATPYF	GenScript
XPB-67	LDIGYPMLEEYDFRN	GenScript

II.1.6 Equipment

Instruments

Type	Model	Supplier
Autoclave	Systec V-150	Systec
Balance, analytical	XS 105 Dual Range	Mettler Toledo
Balance	XS 6002S Dual Range	Mettler Toledo
Block thermstat	Rotilabo® block thermostat H 250	Carl Roth
Circular dichroism (CD) spectropolarimeter	J-715	JASCO
CD cuvette	Cylindrical absorption cuvette, path length 1 mm	Hellma
Cell disruption	M-110P	Microfluidics
Centrifuge	5417 R	Eppendorf
Centrifuge	5424	Eppendorf
Centrifuge	5804 R	Eppendorf
Centrifuge	Avanti J-26 XP	Beckmann Coulter
Centrifuge	Avanti J-HC	Beckmann Coulter
Crystallographic – cryo loop	CryoLoop	Hampton Research
Crystallographic – sample holder	Crystal Cap™ HT (Spine)	Hampton Research
Crystallographic – sample vial	CryoVial	Hampton Research

Crystallographic – handling tool	CrystalWand™ Magnetic	Hampton Research
Crystallographic – sample basket	ESRF/EMBL Sample Changer Basket	Hampton Research
Fast protein liquid chromatography (FPLC) system	ÄKTA™ avant 25	GE Healthcare
FPLC system	ÄKTA™ pure 25	GE Healthcare
FPLC system	ÄKTA™ purifier 10	GE Healthcare
FPLC system	ÄKTAxpress™	GE Healthcare
Electrophoresis	Mini-PROTEAN Tetra Cell	Bio-Rad Laboratories
Electrophoresis power supply	PowerPac™ Basic	Bio-Rad Laboratories
Gel-drying device	GelAir Gel Dryer	Bio-Rad Laboratories
Imager	Odyssey	LI-COR Biosciences
Incubator	B15 Compact Incubator	Heraeus
Isothermal titration calorimeter	VP-ITC	MicroCal, GE Healthcare
Isothermal titration calorimeter	MicroCal ITC_{200}	MicroCal, GE Healthcare
Liquid handling robot	HoneyBee 963	Digilab
Liquid handling robot	LISSY 2002	Zinsser Analytic
Magnetic stirrer	MR 3002	Heidolph Instruments
MALS detector	DAWN® 8 + HELEOS® II	Wyatt Technology
Microscope	STEMI 2000-C	ZEISS
Microscope	SteREO Discovery.V12	ZEISS
Microscope camera	AxioCam MRc	ZEISS
Microscope light source	KL 2500 LCD	ZEISS
Microscope light source	CL 1500 Eco	ZEISS
PCR cycler	Mastercycler® EPgradient S	Eppendorf
PCR cycler	Mastercycler pro S	Eppendorf
Real-time PCR cycler	Thermocycler Stratagene Mx 3005P	Agilent Technologies
pH meter	BlueLine 14pH	SCHOTT
Pipette	Pipet-LTS,	Mettler-Toledo
Pipette	Pipet-XLS	Mettler-Toledo
Pipette	Pipetus	Hirschmann Laborgeräte
Refractometer	Optilab T-rEX	Wyatt Technology
Rotor	JLA-25.50	Beckmann Coulter
Rotor	JLA-8.100	Beckmann Coulter
Rotor	JS 5.0	Beckmann Coulter
Sealing robot	RoboSeal	HJ-BIOANALYTIC
Shaking incubator	ISF-1-W	Kühner
Shaking incubator	ISF-1- X	Kühner
Shaking incubator	LT-X	Kühner
Spectrophotometer	BioPhotometer	Eppendorf
Spectrophotometer	NanoDrop ND 1000	Peqlab
Thermomixer	Thermomix comfort	Eppendorf
Ultra pure water system	TKA GenPure	Thermo Fisher Scientific
Ultrasonic bath sonicator	Sonorex RK 255 H	BANDELIN electronic
Vortex mixer	Vortex-Genie 2	Scientific Industries

X-ray cryosystem	X-Stream™ 2000	Rigaku
X-ray detector	R-Axis HTC	Rigaku
X-ray generator	MicroMax-007 HF	Rigaku
X-ray optics	VariMax™	Rigaku

Chromatography columns and media

Type	Model	Supplier
Column body for affinity chromatography	Econo-Column® 2.5 x 20 cm	Bio-Rad Laboratories
Preparative size exclusion chromatography (SEC) FPLC column	HiLoad™ 16/60 Superdex™ 200	GE Healthcare
Analytical SEC FPLC column	Superdex™ 200 5/150 GL	GE Healthcare
Analytical SEC FPLC column	Superdex™ 200 10/300 GL	GE Healthcare
Analytical SEC FPLC column	Superdex™ 200 Increase 10/300 GL	GE Healthcare
Analytical SEC FPLC column	Superose™ 6 10/300 GL	GE Healthcare
Cation exchange column	MonoS® 5/50 GL	GE Healthcare
Immobilized metal ion affinity chromatography resin	Protino® Ni-TED	MACHEREY-NAGEL

II.1.7 Software

Computer programs

Software	Description	Supplier/Reference
AIMLESS	Scaling and merging of diffraction data	Evans and Murshudov, 2013 (150)
Astra VI	Control for Multi-angle light scattering (MALS) detector and refractometer; MALS data collection and analysis	Wyatt
AxioVision	Recording of microscopy images	ZEISS
BLAST	Database search; sequence analysis	Altschul *et al.*, 1990 (151)
Buccaneer	X-ray crystallography; statistical protein chain tracing	Cowtan, 2006 (152)
BUSTER	X-ray crystallography; structure refinement	Bricogne *et al.*, 2017; Global Phasing Ltd.
CCP4	X-ray crystallography; software suite for structure determination	Winn *et al.*, 2011 (153)
CCP4i	X-ray crystallography; graphical interface to *CCP4*	Potterton *et al.*, 2003 (154)
CLC sequence viewer	Design of constructs and analysis of sequencing data	QIAGEN bioinformatics
Clustal Omega	Multiple sequence alignment	Sievers *et al.*, 2011 (155)

ClustalW/W2	Multiple sequence alignment	Larkin *et al*., 2007 (156)
COBALT	Multiple sequence alignment	Papadopoulos and Agarwala, 2007 (157)
Coot	X-ray crystallography; model-building software	Emsley *et al*., 2010 (158)
CrystalClear	X-ray data collection and basic processing	Rigaku
ESPript 3.0	Analysis of primary to quaternary protein structure information	Robert and Gouet, 2014 (159)
ExPASy ProtParam tool	Computation of physical and chemical parameters of proteins	Artimo *et al*., 2012 (160)
DALI	Structural homology search	Holm and Laasko, 2016 (161)
iTASSER	Protein 3D structure prediction	Zhang, 2008 (162)
ITC_{200}	Experimental design of ITC measurements and ITC instrument control	MicroCal
Microsoft Excel	Spreadsheet software	Microsoft Corporation
MolProbity	X-ray crystallography; geometric evaluation of macromolecular structure models	Chen *et al*., 2010 (163)
MxCuBE	Synchrotron beamline control	Gabadinho *et al.*, 2010 (164)
MxPro	RT PCR cycler control used for thermofluor measurements (recording and basic processing of data)	Agilent Technologies
Origin	Data analysis and graphing	OriginLab Corporation
PDBeFold	Protein structure comparison service	Krissinel and Henrick, 2004 (165)
PDBsum	Summaries and analyses of PDB structures	Laskowski, 2001 (166)
PHASER	X-ray crystallography; phasing software	McCoy *et al*., 2007 (167)
Phenix	X-ray crystallography; software suite for X-ray structure determination and refinement	Adams *et al*., 2010 (168)
Phyre2	Protein 3D-structure prediction	Kelley *et al*., 2015 (169)
POINTLESS	Space-group determination	Evans, 2006 (170)
PyMOL	3-dimensional visualization and graphical illustration of structures	Schrödinger
RaptorX	Protein 3D-structure prediction	Kallberg *et al*., 2012 (171)
REFMAC	X-ray crystallography; structure refinement	Murshudov *et al*., 1997 (172)
Reverse Complement	Determination of reverse complement structure	Stothard, 2000 (173)
SHARP	X-ray crystallography; automated structure solution pipeline; heavy atom refinement and phasing	Bricogne *et al*., 2003 (174)
Spectra Manager	CD spectropolarimeter control, CD data recording and analysis	JASCO
STARANISO	Density improvement of anisotropic data sets	Global Phasing Ltd.
Thermofluor Script	*Excel* script for thermofluor data analysis	SGC, Oxford

UNICORN	FPLC instrument control; recording, analysis and management of chromatograms	GE Healthcare
XDS	X-ray crystallography; processing, indexing and integrating of diffraction images	Kabsch, 2010 (175)

Databases

Database	Description	Web address	Reference
PDB	Structural data of biological macromolecules	www.rcsb.org	Berman *et al.*, 2000 (176)
PubMed	Biomedical literature online library	www.ncbi.nlm.nih.gov/pubmed	National Center for Biotechnology Information (NCBI)
UniProt	Protein sequences	www.uniprot.org	UniProt Consortium, 2015

II.2 Methods

II.2.1 Molecular biology

II.2.1.1 Transformation

LB agar plates: 30 g/l LB medium, 16 g/l agar

Antibiotics: Ampicillin (stock concentration: 100 mg/ml; final concentration: 100 µg/ml)
Chloramphenicol (stock concentration: 34 mg/ml; final concentration: 34 µg/ml)
Kanamycin (stock concentration: 100 mg/ml; final concentration: 100 µg/ml)

All bacterial cells (*E. coli* DH5α™ or *E. coli* BL21-CodonPlus® (DE3)-RIL) were transformed by heat shock transformation. An aliquot of 50 µl of competent *E. coli* cells was thawed on ice for 30 min and supplemented with 100 ng of the respective plasmid DNA. For the co-expression of two proteins, whose genes were cloned into two different plasmids, 100 ng of each plasmid DNA was used. The cell mixture was incubated for 30 min at 4 °C, before a heat shock treatment for 90 s at 42 °C was performed. Then, 300 µl of LB medium was added and the cells were incubated for 45 - 60 min at 37 °C and 600 rpm. Depending on the purpose of the transformation, 100 - 250 µl of the transformed cells were plated on LB agar plates including the appropriate antibiotics and incubated overnight at 37 °C.

II.2.1.2 DNA amplification and isolation

For DNA amplification and subsequent DNA isolation, *E. coli* DH5α™ cells were transformed with the respective DNA (see II.2.1.1), plated on LB agar plates including the appropriate antibiotics and incubated overnight at 37 °C. One to three of the colonies occurring on the agar plate were selected for DNA isolation. A centrifuge tube containing 10 ml of LB medium supplemented with the appropriate antibiotics was inoculated with a single colony and incubated overnight at 37 °C and 200 rpm. On the next day, the DNA was isolated using the NucleoSpin® Plasmid kit and following the manufacturer's instructions. After determining the DNA concentration (see II.2.3.1) 500 ng of DNA were sent for sequencing to Eurofins Genomics or Microsynth Seqlab to verify the result.

II.2.1.3 Sequence and ligation independent cloning

The generation of recombinant DNA of choice was performed by sequence and ligation independent cloning (SLIC) (177). Hereby, either a classic approach was followed, where an insert and a vector with complementary ends were amplified separately and then annealed or an alternative approach was pursued, where the insert remained in the respective vector, but was shortened. For the classical approach, primers that were designed for the insert contained overhangs that were homologous to the vector ends and vice versa. Thereby, complementary ends were generated that allowed facile annealing of vector and insert. For the alternative approach, the primers were designed in a way that they were excluding an undesired part of the insert or vector. Forward and reverse primers were usually 35 - 43 bases long and contained complementary sequences. The polymerase chain reaction (PCR) setup and program was the same for both approaches (Figure II.1).

Final concentration	component
100 ng	template DNA
1x	HF buffer or
(1x	GC buffer +
2 %	DMSO)
200 µM	dNTPs
200 nM	forward primer
200 nM	reverse primer
1 U	Phusion® polymerase
add to 50 µl	ddH_2O

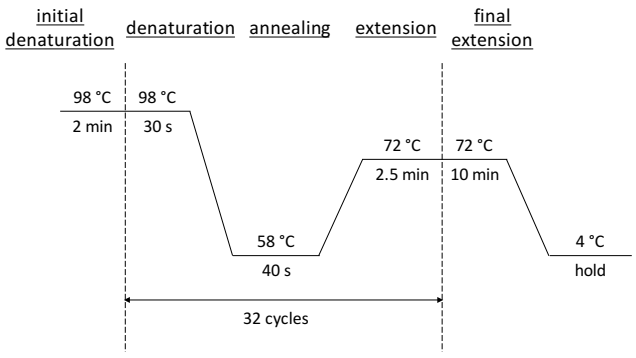


Figure II.1: Setup and program for a PCR during SLIC.

PCR reactions were usually performed in the presence of HF buffer. In some cases, an additional attempt was performed in the presence of GC buffer and DMSO. The linearized PCR product was treated with 20 U Dpn1 overnight at 37 °C to digest the remaining template DNA. On the next day, the PCR product was purified using the NucleoSpin® Gel and PCR Clean-up kit and following the manufacturer's instructions. During the subsequent T4 DNA Polymerase treatment, which was carried out at room temperature, single stranded overhangs were generated. The reaction mix contained 500 ng DNA, 1x BSA and 1.2 U T4 DNA polymerase in the presence of NEBuffer™ 2 or NEBuffer™ 2.1. The reaction was stopped after 30 min by adding dCTP to a final concentration of 1 mM and transferring the sample on ice. Next, the annealing reaction was set up containing specific amounts of DNA, 20 ng RecA and 1x T4 ligation buffer and incubated for 30 min at room temperature. For the classical approach, two different ratios of vector to insert (1:1 and 1:2) resulting in 100 ng of vector DNA and 100 ng or 200 ng of insert DNA, were tested in the annealing reaction. For the alternative approach, 100 ng of the shortened DNA product was taken. After successful transformation (see II.2.1.1) of 15 µl DNA product into *E. coli* DH5α™ cells, the cells were plated on LB agar plated including the appropriate antibiotics and incubated overnight at 37 °C. DNA amplification, isolation and sequencing was performed as described in II.2.1.2.

II.2.1.4 Site-directed mutagenesis

In order to introduce point mutations or premature stop codons, site-directed mutagenesis (SdM) was performed (178). The designed primers contained the altered nucleotides coding for the mutation or stop codon and approximately 20 nucleotides of the target sequence on either side of the mutation. Forward and reverse primers were complementary to each other. Depending on the construct two different strategies were pursued. Following strategy 1, first two single reactions were performed for eight cycles containing either the forward or the reverse primer (Figure II.2). Then, the reactions were combined for another 25 cycles. Strategy 2 was performed as outlined in Figure II.3.

The PCR product was digested with 10 U Dpn1 overnight at 37 °C. The next day, *E. coli* DH5α™ cells were transformed with 10 µl of the DNA product (see II.2.1.1), distributed on LB agar plates including the appropriate antibiotics and incubated overnight at 37 °C. The DNA was isolated from one to three colonies and sent for sequencing as described above (see II.2.1.2) to verify the introduced mutation.

Final concentration	component
100 ng	template DNA
1x (1x 2 %	HF buffer or GC buffer + DMSO)
400 µM	dNTPs
400 nM	forward or reverse primer
1 U	Phusion® polymerase
add to 25 µl	ddH_2O

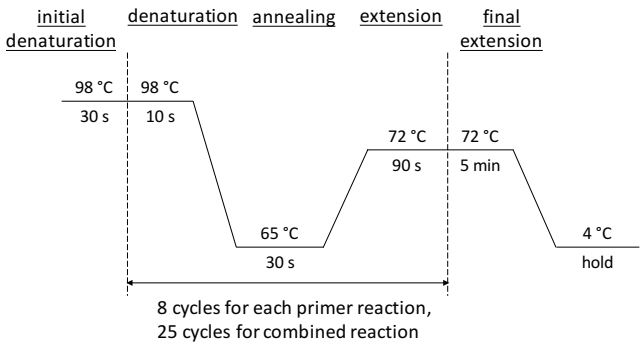


Figure II.2: Setup and program for strategy 1.

Final concentration	component
100 ng	template DNA
1x (1x 1 %	HF buffer or GC buffer + DMSO)
400 µM	dNTPs
300 nM	forward primer
300 nM	reverse primer
1 U	Phusion® polymerase
add to 25 µl	ddH_2O

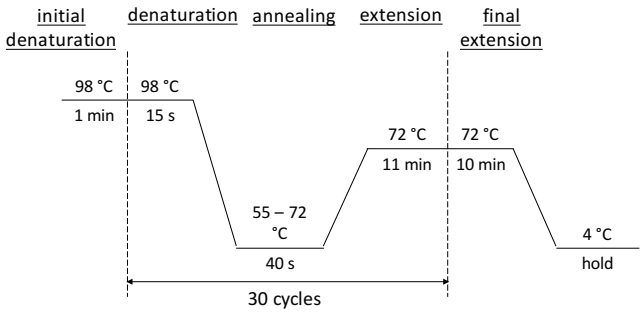


Figure II.3: Setup and program for strategy 2.

II.2.1.5 Expression and harvest

For recombinant protein expression*, E. coli* BL21-CodonPlus® (DE3)-RIL cells were transformed with the DNA of interest at least two days prior to expression using the standard heat shock protocol (see II.2.1.1). One day prior to expression, 100 - 200 ml of LB medium supplemented with the appropriate antibiotics was inoculated with a single colony and the pre-culture was incubated overnight at 37 °C and 200 rpm. On the next day, 20 ml of the overgrown culture were taken to inoculate 2 l of LB medium containing the appropriate antibiotics. The inoculated cell culture was incubated at 37 °C and 200 rpm until an OD_{600} of 0.6 - 0.7 was reached. Expression of the target protein was induced by the addition of either IPTG to a final concentration of 0.5 mM or L-(+)-arabinose to a final concentration of 0.05% or both if two proteins, whose DNA was located on two different plasmids, were co-expressed. The incubation of the expression culture was continued in the same shaker incubator for 19 - 21 h at 200 rpm but at a reduced temperature of 15 °C. The cells were harvested by centrifugation for 12 min at 4 °C and 6056 x g. The cell pellet was either purified subsequently or stored at -20 °C (most p34, p44 and XPB variants) or -80 °C (most p52 variants).

II.2.1.6 Expression and harvest of seleno-methionine derivatives

M9 medium:	0.1 mM $CaCl_2$, 2 mM $MgSO_4$, 4.2 mg/l $Fe_{II}SO_4$, 0.4% glucose, 10 µg/50 ml biotin, 1x M9 mix
5x M9 mix:	33.9 g/l Na_2HPO_4, 15 g/l KH_2PO_4, 5 g/l NH_4Cl, 2.5 g/l NaCl, pH 7.0 (autoclaved)
Amino acid mix:	10 mg/100 ml Lys/Phe/Thr, 5 mg/100 ml Val/Leu/Iso, 3 mg/100 ml Se-Met

In order to obtain seleno-methionine (Se-Met) containing crystals, which were used for structure determination via phasing with single-wavelength anomalous diffraction data (see II.2.5.4), Se-Met derivatives were produced. The transformation of the plasmid DNA into *E. coli* BL21-CodonPlus® (DE3)-RIL cells and the inoculation and incubation of the pre-culture was performed with standard LB agar plates and in standard LB medium as described before (see II.2.1.1 and II.2.1.5).

The expression was carried out in 4 x 2 l M9 medium. Prior to inoculation of the expression culture, the pre-culture was centrifuged for 15 min at 20 °C and 3220 x g and the pellet was resuspended in 5 ml M9 medium. The resuspended pellet was used to inoculate 2 l of M9 medium containing the appropriate antibiotics and the cell culture was incubated at 37 °C and 200 rpm until an OD_{600} of 0.5 was reached. In contrast to a standard expression in LB medium, where cells need 3 - 4 h to reach an OD_{600} of 0.6, it takes up to 25 h for the cells to reach an OD_{600} of 0.5 in M9 medium. At an OD_{600} of 0.5 a special amino acid mix was added to facilitate feedback inhibition of the methionine biosynthesis pathway in the cells and they were further incubated for 15 min before the induction by the addition of IPTG to a final concentration of 0.5 mM was performed. The shaker incubator was cooled down to 15 °C and expression was conducted for 22 - 23 h. The cells were harvested by centrifugation for 12 min at 4 °C and 6056 x g and the cell pellet was purified subsequently.

II.2.2 Protein purification

II.2.2.1 Cell lysis

Lysis buffer I:	50 mM Tris, 500 mM KCl, 5 mM imidazole, 0.5 mM TCEP, pH 8.0
Lysis buffer II:	20 mM Tris, 500 mM KCl, 0 – 5 mM imidazole, 0.5 – 1 mM TCEP, pH 8.0
Lysis buffer III:	20 mM HEPES, 375 mM NaCl, 5 mM imidazole, pH 8.0
Lysis buffer IV:	20 mM Tris, 300 mM NaCl, 5 mM imidazole, pH 7.5
Lysis buffer V:	20 mM Tris, 300 mM NaCl, 5 mM imidazole, 1 mM TCEP, pH 7.5
Lysis buffer VI:	20 mM Tris, 150 mM KCl, 0.5 mM TCEP, pH 8.0

For cell lysis, cell pellets stored at -20 °C or -80 °C were thawed either on ice or at room temperature. The thawed or freshly harvested cell pellet was resuspended in 5 ml of the respective lysis buffer per 1 g of wet pellet, supplemented with 150 U of self-made DNase I per 1 g of wet pellet and stirred for about 30 min at 4 °C and 250 rpm until complete resuspension was reached. The cells were lysed in three cycles at a pressure of 1.5 kbar using a mechanical cell disrupter and the cell suspension was centrifuged at 38,000 - 48,000 x g and 4 °C for 1 h. After centrifugation, samples of the pellet and the supernatant were taken for SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) analysis before the supernatant was used for subsequent purification via affinity chromatography.

II.2.2.2 Affinity chromatography

Elution buffer I:	50 mM Tris, 500 mM KCl, 150 mM imidazole, 0.5 mM TCEP, pH 8.0
Elution buffer II:	20 mM Tris, 300 mM KCl, 250 mM imidazole, 1 mM TCEP, pH 8.0
Elution buffer III:	20 mM HEPES, 375 mM NaCl, 250 mM imidazole, pH 8.0
Elution buffer IV:	20 mM Tris, 300 mM NaCl, 250 mM imidazole, pH 7.5
Elution buffer V:	20 mM Tris, 300 mM NaCl, 250 mM imidazole, 1 mM TCEP, pH 7.5
Elution buffer VI:	20 mM Tris, 150 mM KCl, 150 mM imidazole, 1 mM TCEP, pH 8.0

All proteins purified in this work either contained an N- or C-terminal Hexahistidine-tag (His_6-tag) or were co-purified together with a His_6-tagged protein. An immobilized metal ion affinity chromatography (IMAC) procedure was thus applied as an initial purification step, which during the herein described work was based on the nonbiospecific interaction between the His_6-tags and Ni^{2+} ions immobilized on silica beads via the chelating ligand tris(carboxymethyl)-ethylene diamine (TED). IMAC was performed by a batch procedure, which started in a glass beaker and was further pursued in a column. All steps of the affinity chromatography were performed at 4° C and the protein samples were kept on ice during the entire purification process for all described purifications except for the special purifications of p52 and p52/p8 at room temperature.

A volume of 100 - 300 ml of the supernatant obtained after cell lysis was incubated with 1.0 - 2.0 g Protino® Ni-TED beads and 50 - 200 mg PMSF for 1 h under constant stirring at 150 rpm in the cold room. Then, the supernatant (flow through) was removed and the beads were embedded in the respective lysis buffer (see II.2.2.1) and transferred into a column body for further washing of the beads with 200 - 300 ml of the respective lysis buffer. The His_6-tagged proteins were eluted from the Ni^{2+}-coated beads with a buffer containing high concentrations of imidazole, which has a strong affinity to Ni^{2+} and competes with the target protein for the same binding surface on the

beads. The protein was eluted with the respective elution buffer in 8 - 10 fractions of 7.5 - 9.0 ml each. Samples of the flow through, wash and all 8 - 10 elution fractions were taken and analyzed by SDS-PAGE as described in II.2.3.2. According to the result of the SDS-PAGE, elution fractions containing the target protein were pooled and, if not subjected to TEV cleavage or ion exchange chromatography (see II.2.2.3 or II.2.2.4), concentrated to 2.5 - 5.0 ml using Amicon® ultra centrifugal filters with a molecular weight cut-off of either 3,000 Da, 10,000 Da or 30,000 Da corresponding to the size of the target protein.

II.2.2.3 His_6-tag cleavage by the TEV protease

TEV dialysis buffer: 20 mM Tris, 300 mM NaCl, 0.5 mM EDTA, 1 mM DTT, pH 7.5

During some of the purifications, it was necessary to remove the His_6-tag of the respective protein by cleavage with the TEV protease. The absorbance at 280 nm (A_{280}) of the pooled target protein was determined by UV/Vis spectrophotometry (see II.2.3.1) and the TEV protease was added to the protein in a A_{280} - ratio of 1:50. The mixture was dialyzed in TEV dialysis buffer overnight at 4 °C. On the next day, 1 mM $MgCl_2$ was added to the sample before it was applied to a reverse IMAC step to clear the cleaved sample from the His_6-tagged TEV protease and the non-cleaved sample. The protein sample was incubated with 0.3 g Ni-TED beads per 1 ml of protein solution for 75 min at 4 °C. The flow through containing the cleaved protein was collected and the beads were washed with the respective SEC buffer and the His_6-tagged protein was eluted with the respective elution buffer. Samples of the protein before and after TEV cleavage were collected and analyzed by SDS-PAGE. The cleaved protein sample was concentrated as described in II.2.2.2 and subjected to a preparative size exclusion chromatography step.

II.2.2.4 Ion exchange chromatography

Dilution IEC buffer: 20 mM Tris, pH 7.5
Low salt IEC buffer: 20 mM Tris, 50 mM NaCl, pH 7.5
High salt IEC buffer: 20 mM Tris, 1 M NaCl, pH 7.5

For ion exchange chromatography (IEC) the respective protein sample was diluted in dilution buffer to obtain an overall salt concentration of 50 mM NaCl. The sample was applied onto a MonoS® 5/50 GL cation exchange chromatography column, which was attached to an FPLC Äkta™ system. Prior to sample application, the IEC column was first equilibrated in high salt IEC buffer to remove potential proteins and then in low salt IEC buffer. After sample application, the IEC

column was washed over 10 CV using the low salt IEC buffer. The protein was eluted in a linear gradient rising to 50% of the high salt IEC buffer in 80 CV and elution fractions of 1 ml were collected. According to the chromatogram obtained during the elution, appropriate samples were chosen for SDS-PAGE analysis. Elution fractions containing the target protein were pooled and the protein sample was subsequently concentrated as described in II.2.2.2.

II.2.2.5 Preparative size exclusion chromatography

SEC buffer I:	20 mM Tris, 200 mM KCl, 1 mM TCEP, pH 8.0
SEC buffer II:	20 mM Tris, 150 mM KCl, 1 mM TCEP, pH 8.0
SEC buffer III:	20 mM HEPES, 375 mM NaCl, 250 mM imidazole, pH 8.0
SEC buffer IV:	20 mM Tris, 250 mM NaCl, pH 7.5
SEC buffer V:	20 mM Tris, 250 mM NaCl, 1 mM TCEP, pH 7.5

The concentrated sample (2.5 - 5.0 ml) was usually centrifuged for 20 - 30 min at 11,000 x g and 4 °C prior to the size exclusion chromatography (SEC) step. Then, the sample was applied onto a HiLoad™ 16/60 Superdex™ 200 (SD 200 16/60) column, which was attached to a Fast Protein Liquid Chromatography (FPLC) system (here: an Äkta™ system) and was equilibrated in the respective SEC buffer in advance. The protein was eluted in the respective SEC buffer over 1.2 column volumes (CV) and in fractions of 2 ml. According to the chromatogram obtained during elution, appropriate samples were taken for SDS-PAGE analysis (see II.2.3.2). Selective fractions were pooled and concentrated to 2 - 20 mg/ml using amicon® ultra centrifugal filters with a suitable molecular weight cut-off. The final concentration of the protein was determined as described in II.2.3.1. The portion of the freshly purified protein that was not directly taken for crystallization trials or biochemical studies was split into aliquots, flash frozen in liquid nitrogen and stored at -80 °C.

II.2.3 Protein characterization

II.2.3.1 UV/Vis spectrophotometry

DNA as well as protein concentrations were determined by measuring their UV absorbance using a spectrophotometer. All measurements were corrected against the spectrum of the respective reference buffer. The DNA concentration was measured at 260 nm based on the absorbance of purines and pyrimidines in DNA. The protein concentration (c) was determined based on the

Beer-Lambert law (Eq.) taking its absorbance at 280 nm (A_{280}), its extinction coefficient (ε) and the path length (d) of the light through the sample into account.

$$A_{280} = \varepsilon \times c \times d \qquad \text{(Eq.)}$$

The extinction coefficient, characterized by the number of tryptophan, tyrosine and cysteine residues, and the molecular weight of the respective protein was determined for each protein utilizing the online tool ExPASy ProtParam.
The purity of the DNA or protein sample was verified by the absorption ratio of 260 nm/280 nm.

II.2.3.2 Separation of proteins by SDS-polyacrylamide gel electrophoresis

15% separation gel:	15% (v/v) acrylamide/bisacrylamide mix (37.5:1), 167 mM Tris-HCl pH 8.8, 0.1% (w/v) SDS, freshly added 0.1% (w/v) APS and 0.04% (v/v) TEMED
5% stacking gel:	5% (v/v) acrylamide/bisacrylamide mix (37.5:1), 125 mM Tris-HCl pH 6.8, 0.1% (w/v) SDS, freshly added 0.1% (w/v) APS and 0.1% (v/v) TEMED
5x SDS sample buffer:	250 mM Tris-HCl pH 6.8, 500 mM DTT, 0.5% (w/v) bromphenol blue, 10% (w/v) SDS, 50% glycerol
SDS running buffer:	192 mM glycine, 0.1% SDS, 25 mM Tris-HCl
Coomassie® staining:	50% ethanol, 10% acetic acid, 0.1% brilliant blue R-250
Coomassie® destaining:	10% ethanol, 5% acetic acid

Proteins were separated by SDS-PAGE analysis using self-prepared gels that consisted of a short 5% stacking gel and an adjacent 10 - 15% separation gel in order to separate the proteins according to their molecular weight. If not stated differently the SDS-PAGE consisted of a 15% separation gel. Samples for SDS-PAGE analysis consisted of 20 µl sample and 5 µl 5x SDS sample buffer. SDS-PAGE samples of the pellet after cell lysis were in particular resuspended in 1 ml of 6 M urea and samples of the supernatant and flow through of affinity chromatography were diluted in the respective lysis buffer in a 1:10 ratio. All samples, with the exception of p34 containing samples, were heated to 95 °C for 2 - 5 min prior to loading 10 µl of them and 2.5 µl of the PageRuler™ Plus Prestained Protein Ladder onto the gel. The gels were run at 250 V for 30 - 40 min at room temperature, stained with Coomassie® staining solution for 5 - 15 min and subsequently destained until the bands appeared clearly on the gel and then stored in water.

II.2.3.3 Limited proteolysis

One aim of performing limited proteolysis was to determine degradation variants that can be cloned and expressed and might facilitate subsequent crystallization. Another purpose of limited proteolysis was to perform a pre-screening to identify the best protease and the best conditions to perform *in situ* proteolysis (see II.2.5.1). Limited proteolysis was performed using either the protease papain, subtilisin A, α-chymotrypsin (JBS Floppy-Choppy kit) or trypsin (JBS Floppy-Choppy kit, Carl Roth or in-house production). 10 µl of protein at a concentration of 2.5 mg/ml was incubated with 2.5 µl of the respective protease at room temperature. In order to determine the best degradation result, the concentration of the protease (0.1 mg/ml, 0.01 mg/ml and 0.001 mg/ml corresponding to ratios of protease-to-protein = 1:100, 1:1000 and 1:10,000) and the incubation time (5 min, 15 min, 25 min, 30 min, 35 min, 45 min) were varied. After the respective incubation time, 3.2 µl of 5 x SDS sample buffer were added and the sample was heated for 5 min at 95 °C to stop the protease reaction.

The different reaction mixtures were compared by SDS-PAGE analysis (see II.2.3.2). For this purpose, 10 µl of each reaction was loaded onto a SDS-PAGE gel. For pre-screening purposes, the results were compared and the most promising protease-to-protein ratio was taken for *in situ* proteolysis. In order to determine a degradation variant that can subsequently be cloned, the most promising degradation product was chosen and reproduced for mass spectrometry analysis. Prior to loading the sample for mass spectrometry analysis onto the SDS-PAGE gel Iodoacetamide was added to a final concentration of 240 mM to alkylate cysteines and the reaction was incubated in the dark for 20 min. Respective bands were thoroughly cut out of the SDS-PAGE gel, transferred into a tube and sent to the mass spectrometry department (Schlosser group, Rudolf Virchow Center, University of Würzburg).

II.2.3.4 Circular dichroism spectroscopy

CD buffer: 20 mM K_2H/H_2K-$P0_4$, pH 8.0

Circular dichroism (CD) spectroscopy measurements were performed to confirm the structural integrity of a protein and conducted in a 1 mm quartz cuvette employing a spectropolarimeter. For all measurements, protein samples were equilibrated in the same CD buffer conditions and their CD spectra were recorded at a wavelength of 260 to 190 nm and at room temperature. A spectrum of the reference buffer was recorded and subtracted from the protein sample spectra. A total of 10 spectra were accumulated to optimize the signal-to-noise ratio.

II.2.3.5 Thermal shift assay – ThermoFluor

ThermoFluor analysis was performed to determine the stability of a protein or protein complex in different buffer compositions. SYPRO® Orange served as fluorescent dye, which, with increasing temperatures, interacts with hydrophobic amino acids side chains in the unfolding process of the protein. A mixture of 0.1 - 0.5 mg/ml protein, 5x SYPRO® Orange and 92 mM of the respective buffer in a total volume of 25 µl was filled in one tube of a 96-well plate, which was analyzed in a real time PCR cycler. The plate was sealed and heated from 25 °C to 95 °C in 1 °C steps while the fluorescence was recorded. At least three controls containing the protein in the respective reference buffer were measured to compare the results with a standard. A summary of the utilized screens can be found in II.1.4.

II.2.3.6 Size exclusion chromatography coupled to multi-angle light scattering

MALS buffer A: 20 mM Tris, 250 mM NaCl, pH 7.5

MALS buffer B: 20 mM HEPES, 200 mM NaCl, pH 8.0

The molecular mass distribution of a protein sample was determined by size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS). The protein sample was first fractionated according to size by a Superdex™ 200 10/300 GL (SD 200 10/300) column attached to an ÄKTA™ purifier 10 system and then continued processing further by a MALS detector and a refractive index monitor. SEC-MALS analyses were conducted in MALS buffer A or B at room temperature. All samples were centrifuged for 20 - 30 min at 4 °C and 11,228 x g prior to analysis. SEC-MALS was also used for interaction analyses of two proteins. Here, the two centrifuged protein samples were mixed in a 1:1 ratio and incubated for 1 h on ice prior to the chromatography step. For the measurement, 100 µl of sample at a concentration of at least 2.5 mg/ml were injected onto the column. The result was processed and analyzed using the Astra VI software.

II.2.4 Interaction analysis

II.2.4.1 Analytical size exclusion chromatography

A-SEC buffer A: 20 mM Tris, 250 mM NaCl, 1 mM TCEP, pH 7.5

A-SEC buffer B: 20 mM Tris, 150 mM KCl, 1 mM TCEP, pH 8.0

A-SEC buffer C: 20 mM HEPES, 200 mM NaCl, pH 8.0

One of the methods used to identify protein interactions was analytical size exclusion chromatography (A-SEC). A-SEC was performed using a small-scale analytical column, which was attached to an FPLC Äkta™ system working at 4 °C. For sample preparation, the proteins were thawed on ice, centrifuged for 30 min at 11,228 x g and 4 °C and then diluted to the desired concentration. To verify the interaction between two proteins, the samples of each single protein and a sample containing both proteins in a 1:1 ratio were prepared and incubated for 1 h on ice prior to the experiment. Depending on the size of the column either 40 µl of 50 µM protein (5/150 column) or 400 µl of 10 - 12 µM protein (10/300 columns) were applied onto the respective column, which was pre-equilibrated with the respective A-SEC buffer. Sample separation was performed over 1.2 CV and fractions were taken throughout the entire elution. To verify the results observed on the chromatogram, fractions from a defined elution range were chosen and analyzed via SDS-PAGE (see II.2.3.2).

II.2.4.2 Native polyacrylamide gel electrophoresis

3.5% Native gel:	3.5% (v/v) acrylamide/bisacrylamide mix (37.5:1), 0.5x Tris/glycine, freshly added 0.07% (w/v) APS and 0.1% (v/v) TEMED
10x Tris/glycine:	250 mM Tris, 1.92 M glycine
Ponceau S sample buffer:	50% glycerol, 0.1% Ponceau S

Native polyacrylamide gel electrophoresis (native PAGE) was used to determine protein interactions while keeping the protein in its native conformation. For this purpose, 3.5% native polyacrylamide gels consisting of a separation gel only were used. A native gel was prepared as such that all proteins analyzed on that gel were also present as single protein samples. Since many protein samples purified in different buffer compositions were analyzed on one native gel, the samples had to be prepared as such that the buffer composition was precisely equal for each lane of the gel. All samples were prepared simultaneously and contained an end volume of 5 µl and a final protein concentration of 10 µM. The samples were incubated for 1 h on ice, then 1 µl of Ponceau S sample buffer was added and the complete sample was loaded onto the gel. Native gels were run in 0.5x Tris/glycine buffer for 1 - 2 h at 4 °C and 90 V. Coomassie® staining as described in II.2.3.2 was used to visualize the bands on the native gel.

II.2.4.3 Isothermal titration calorimetry

ITC buffer A:	20 mM CHES, 150 mMKCl, 1 mM TCEP, pH 9.5
ITC buffer B:	20 mM HEPES, 150 mM NaCl, pH 8.0

ITC buffer C: 20 mM Tris, 150 mM NaCl, pH 7.5

The thermodynamic analysis of molecular interactions was performed by isothermal titration calorimetry (ITC) using either a VP-ITC (for p34/p44 samples) or an ITC_{200} (for p52/XPB samples) instrument. Protein as well as peptide samples were dialyzed in the respective ITC buffer conditions overnight at 4 °C to ensure that the buffer conditions for each experiment were identical. After dialysis, the samples were centrifuged for 30 min at 4 °C and 11,228 x g and then diluted in the respective dialysis buffer to reach the desired concentration, which was determined by UV/Vis spectrophotometry (see II.2.3.1). The respective beforehand filtered ITC buffer served as a reference buffer for the ITC experiment. For measurements employing the VP-ITC instrument, the respective ITC buffer and all samples were degassed for 6 min at 23 °C prior to the experiment. All ITC measurements were carried out at either at 25 °C or 37 °C. The following Table lists the settings that were used for the individual measurements. Control experiments were performed, in which the ligand was titrated into the respective ITC buffer. For data analysis, the Origin software was employed, a one site-binding model was assumed and the first injection was discarded.

Setting	**p34/p44 samples**	**p52/XPB samples**
Number of injections	30	15 - 16
Initial delay [min]	5	4
Injection speed [µl/s]	0.5	0.5
Injection volume [µl]	10	2.5
Injection spacing [min]	4	4
Stirring speed [rpm]	260	500
Reference power [µcal/s]	12	11

II.2.5 X-ray crystallography

II.2.5.1 Protein crystallization

Crystal growth was achieved by the vapor diffusion method. Freshly purified protein was typically used to set up initial crystallization trials. For developing further crystallization trials or fine screens, the protein was thawed on ice and centrifuged for 30 min at 4 °C and 11,228 x g prior to setting up the crystal plates. Initial screens and most of the fine screens were set up by the sitting drop method using the HoneyBee 963 crystallization robot at room temperature. The crystallization robot prepared 96-well crystallization plates in a way that each well contained

40 µl of reservoir solution and a sitting drop consisting of 0.3 µl of the same reservoir solution mixed with 0.3 µl of the protein solution. During the plate setup, the protein sample was kept on ice if not being used by the crystallization robot. Once the 96-well crystallization plate setup was completed, the plate was sealed airtight with an adhesive film and usually stored at 20 °C or in some rare cases also at 4 °C, 30 °C or 37 °C. A list of the different initial screens that were used during this work and all fine screens designed for individual crystallization purposes can be found in II.1.4. All screens used for 96-well crystallization plate setups were generated by the LISSY 2002 liquid handing robot.

In some cases, fine screen conditions were additionally tested in a 24-well hanging drop setup. For this purpose, a reservoir solution of 1 ml was pipetted manually into a 24-well plate and mixed thoroughly before a drop of 0.5 - 2 µl of this reservoir solution was mixed with 0.5 - 2 µl of the protein solution on a cover slip. The cover slip was placed headfirst onto the respective well and sealed with silicon grease. The 24-well setup was performed at room temperature and plates were stored at 20 °C.

For co-crystallization attempts of p52 and p8, the proteins were thawed on ice and centrifuged for 30 min at 4 °C and 11,228 x g. Then, p52 was mixed with p8 in a 1:2 ratio and the mixture was incubated on ice for 60 - 90 min before 96-well crystallization trials were setup using the crystallization robot.

For *in situ* proteolysis, the protein sample and the protease were prepared by thawing and centrifugation as described above. *In situ* proteolysis was performed based on the results derived from the limited proteolysis experiments (see II.2.3.3). According to these results, the protein was mixed with the protease in a defined ratio, incubated for a defined time at room temperature and kept on ice during the subsequent crystal plate setup, which was performed in 96-well crystallization plates using the crystallization robot.

To improve the crystal quality, micro-seeding with pre-exiting crystals was performed. The protein was prepared as described above and mixed with a seeding solution in a 2:1 ratio prior to setting up the crystal plates using the crystallization robot. The seeding solution was produced by collecting different drops containing pre-existing crystals and diluting these with mother liquor from the reservoir until the collected drops form only 3 - 5% of the final solution. Next, the seeding solution was crushed with a glass pipette, vortexed for 2 - 3 min at maximum speed and subsequently sonicated for 2 min.

II.2.5.2 Crystal preparation

Crystals that were taken for data collection were fished out of the drop, in which they grew, cryo-protected and flash frozen in liquid nitrogen. For cryo-protection, crystals were fished out of the drop with a cryo-loop of proper size and placed into a new drop containing the mother liquor and 20 - 25% (v/v) glycerol. In some rare cases, a drop containing 3 M - 4 M TMAO without mother liquor was used for cryo-protection. Subsequently, the crystals were fished out of the cryo-drop, excrescent cryo-solution was knocked off and cryo-loops harboring the crystals were flash frozen in liquid nitrogen and stored in a precooled sample basket.

Besides co-crystallization attempts of p52 and p8, p52 crystals were also soaked in a drop containing purified p8 to obtain p52/p8 crystals. For this approach, p52 crystals were fished out of the drop with a mounted cryo-loop of proper size and placed into a new drop containing a 10-fold excess of p8 compared to the concentration at which p52 was crystallized. In some of these attempts, 15% (v/v) glycerol was added to the p8 drop. The p52 crystal was incubated in the p8 drop for 1 - 30 min.

Besides the production of Se-Met derivatives, structure determination via phasing with single-wavelength anomalous diffraction data from crystals that were soaked in heavy-atom containing solutions was attempted. Crystals were soaked in different solutions for varying incubation times at room temperature. The following Table enlists the various combinations tested in this thesis. The preparation of iodine derivatives by vaporizing iodine labelling was performed according to the protocol described in Miyatake *et al*., 2006 (179).

cryo-solution	**additive**	**incubation time**
cryo-solution + 25 % glycerol	0.1 M NaBr 0.3 M NaBr 0.6 M NaBr 0.1 M KI 0.3 M KI 0.6 M KI	15 s, 30 s, 45 s, 60 s
no cryo-protection or addition of 25 % glycerol only	I_2 from KI/I_2 (0.67 M/0.47 M) droplet	5 min, 10 min, 20 min, 30 min, 60 min, 120 min
cryo-solution + 25 % glycerol	5 mM potassium dicyanoaurate(I) 5 mM potassium tetranitroplatinate(II) 5 mM mercury(II) acetate 5 mM lead(II) acetate trihydrate	10 min, 20 min, 30 min, 60 min, 75 min, 90 min, 300 min

II.2.5.3 Data collection and processing

Data sets of adequately diffracting crystals were collected based on the crystal quality and parameters of initial test images obtained at one of the synchrotron radiation facilities, the European Synchrotron Radiation Facility (ESRF, Grenoble, France), the Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung (BESSY, Berlin) or the Deutsches Elektronen-Synchrotron (DESY, EMBL, Hamburg). Structures solved in the course of the herein described thesis were based on the following data sets.

Collection parameter	**p52_121-EdL SAD**	**p52_121-End**	**p52_121-EdL/p8**	**p52_121-EdL/p8 SAD**	**p52_1-321 SAD**
Beamline	BESSY 14.1	ESRF ID30A-3	ESRF ID30A-3	ESRF ID30A-3	BESSY 14.1
Wavelength [Å]	0.979770	0.96500	0.96770	0.96770	0.97973
Detector	Dectris PILATUS 6M	Rayonix MarCCD 225	Dectris EIGER X 4M	Dectris EIGER X 4M	Dectris PILATUS 6M
Detector distance [nm]	500.0	271.363	227.31	251.29	574.71
Number of images	7200	360	2500	3600	3600
Oscillation [°]	0.1	1.0	0.1	0.1	0.1

All diffraction data were processed, integrated and scaled using XDS, Pointless and Aimless (both CCP4 suite).

II.2.5.4 Structure solution and model refinement

First, the structure of p52_121-EdL (p52_121-End_ΔLinker) was solved using a data set of a Se-Met derivative, which was pivotal for phasing by single-wavelength anomalous diffraction (SAD). The structure was solved by applying the automated structure solution pipeline SHARP and the automated model building program buccaneer. The initial model was first improved by combining density modifications with the phases from the model. Then, following model building and refinement cycles using the buccaneer/REFMAC pipeline the model was further improved by manual model building. For later stages of the refinement process the data were improved by utilizing the STARANISO server and the p52_121-EdL structure was refined using Phenix refine. This structure served as a model for solving the structure of p52_121-End by molecular replacement applying PHASER.

The structure of p52_1-321 was solved by molecular replacement using PHASER and domain MD1 (residues 121 - 321) of p52_121-EdL as a search model. The additional electron density that

appeared at the N-terminus of the model allowed to manually build some α-helices of the N-terminal domain (NTD) of p52 in Coot. Seven α-helices per molecule of the asymmetric unit could be built after intensive model building and multiple refinement cycles by REFMAC and Phenix refine. The anomalous density signal of the SAD data allowed the localization of three out of four Se-Met residues per NTD and served as a guideline for subsequent model building and sequence assignment. The structure was refined by REFMAC using an amplitude-based twin refinement strategy.

The structure of p52_121-EdL in complex with p8 (p52_121-EdL/p8) was solved by molecular replacement with PHASER using the structure of p52_121-EdL as a model. The additional electron density, which corresponded to p52_CTD and p8, was used to build the structure of the yeast homologs of p52_CTD/p8 (PDB code: 3DOM) and the sequences were adjusted to the *C. thermophilum* equivalents. The data of p52_121-EdL/p8 were improved by the utilizing the STARANISO server and the native p52_121-EdL/p8 structure was refined using BUSTER. A SAD data set from Se-Met containing crystals of p52_121-EdL/p8 was solved by molecular replacement applying PHASER EP and using the structure of p52_121-EdL/p8 as a search model to unambiguously define the position of p52_CTD and p8.

More details about the structure determination of individual p52 structures and their purposes are provided in the respective result parts. Data collection, structure solution and refinement of the p52 structures was kindly assisted by Jochen Kuper and Wolfgang Kölmel.

III Results

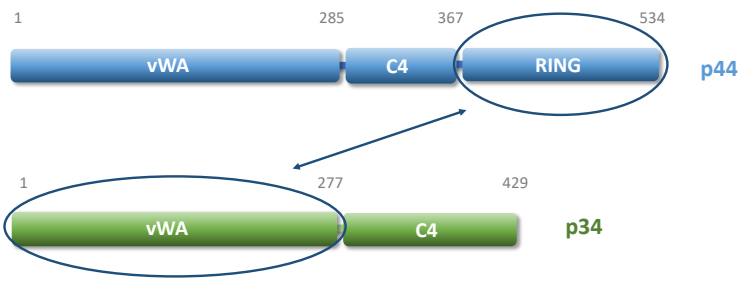
III.1 The p34/p44 Complex

This section presents the structural and functional analysis of the interaction between p34 and p44. The model organism *Chaetomium thermophilum* (*C. thermophilum*; ct) was chosen for all protein studies presented here. The results are based on the structure of a minimal complex between the p34 vWA domain (p34_vWA) and the p44 RING domain (p44_RING), which was the objective of previous works described in the doctoral thesis of Dominik Schmitt (180) and the Master thesis of Elisabeth Schönwetter (148). The crystal structure of the p34_vWA/p44_RING minimal complex and further biochemical studies on the interaction between p34 and p44, which are presented in more detail in the following, have been originally presented and discussed in the publication Radu; Schoenwetter *et al.*; 2017 (181). Some of the results presented below have been adapted from this publication.

III.1.1 Structure of the p34/p44 minimal complex

The interaction between the vWA domain of p34 and the RING domain of p44 has been suggested in several previous studies (135, 142, 143). A crystal structure of a minimal complex consisting of the p34_vWA and p44_RING domains of the *C. thermophilum* proteins (p34/p44 MC I) (Figure III.1) could be obtained at a resolution of 3.7 Å by a former PhD student of the Kisker group, Dominik Schmitt (180).

A



B

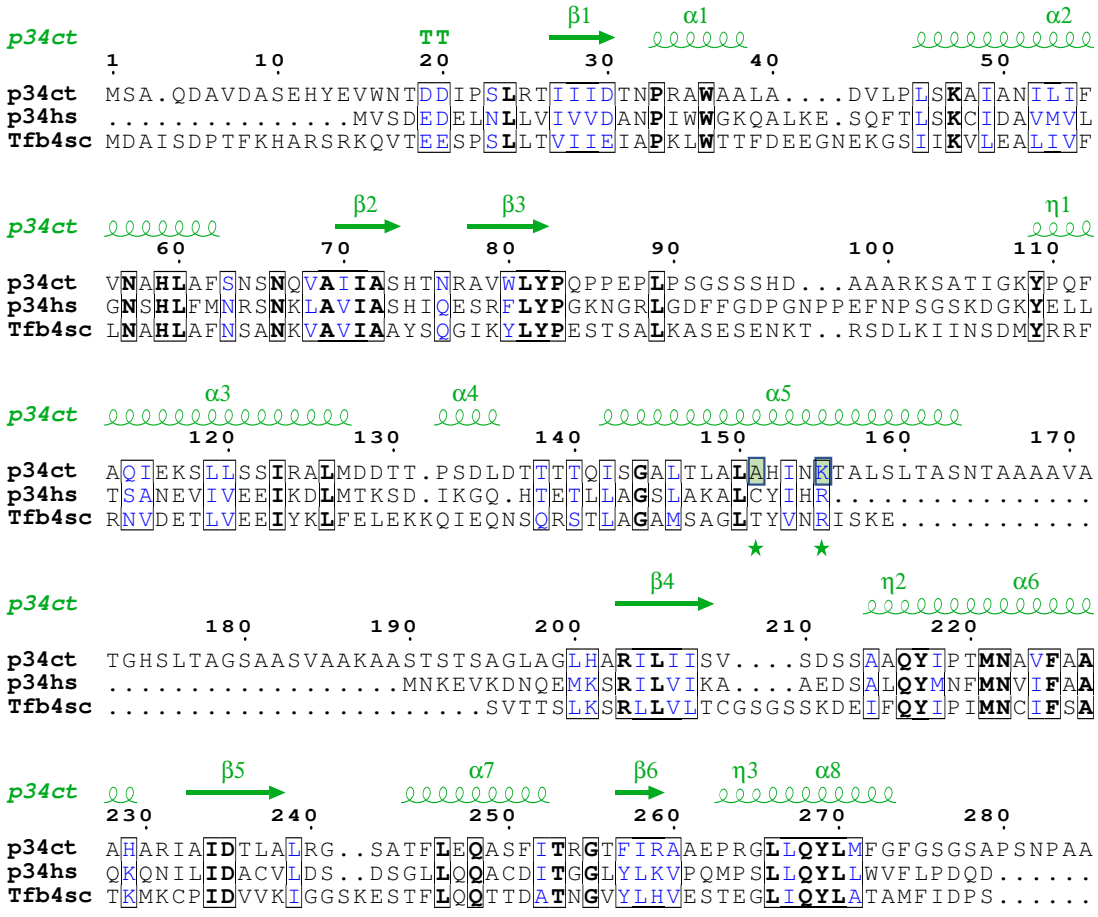
Name according to residues	Name according to domain
p44_368-534	p44_RING
p34_1-277	p34_vWA

Figure III.1: Domain architecture and nomenclature of p44 and p34 from *C. thermophilum* and representation of their interaction. (A) Illustration of the domain architecture of p44 and p34 from *C. thermophilum* both containing a *von Willebrand factor A* like (vWA) domain and a C4 zinc binding domain. P44 additionally contains a C4C4 RING domain. The grey numbers indicate the amino acids that approximately mark the borders of the

individual domains. The interaction between p44_RING and p34_vWA as present in the crystal structure is represented by an arrow. **(B)** Nomenclature of p44_RING and p34_vWA domains.

The amino acid sequences of p34 and p44 from *C. thermophilum* contain flexible linker insertions that are missing in their human homologs (Figure III.2) and have already been described for the p34_vWA structure from *C. thermophilum* (143). The p44_RING domain contains a prominent insertion reaching from residue 410 to 468, which is missing in human p44 and not visible in the electron density map of the p34/p44 MC I structure. In the Master thesis of Elisabeth Schönwetter the resolution of the p34_vWA/p44_RING minimal complex could be improved to 2.2 Å by replacing this flexible linker insertion in p44_RING by an artificial linker consisting of five amino acids, S-N-G-N-G (148). The improved p34_vWA/p44_RING minimal complex structure (p34/p44 MC II) permitted a more detailed view of the interface. The p34/p44 MC I and II structures are described in more detail in the previous works of Dominik Schmitt and Elisabeth Schönwetter (148, 180).

A



B

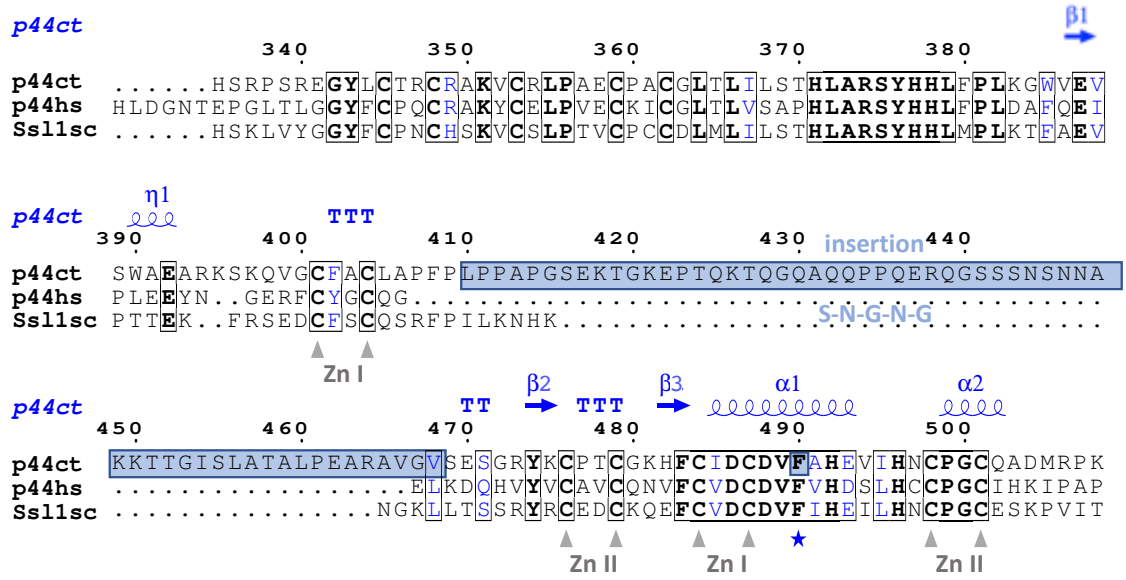


Figure III.2: Sequence alignment of p34 and p44. (A) Sequence alignments of p34_vWA from *C. thermophilum* (ct), human (hs) and *Saccharomyces cerevisiae* (*S. cerevisiae*; sc; Tfb4). **(B)** Sequence alignment of p44_RING from ct, hs and sc (Ssl1). Both sequence alignments were generated with the Clustal Omega (1.2.4) multiple sequence alignment tool. Secondary structure elements (p34/p44 MC II for p34 and p34/p44 MC I for p44) are indicated by arrows, spirals, η or TT for β-strands, α-helices, 3_{10} helices and β-turns, respectively (generated by ESPript). Conserved residues are highlighted in black boxes with the most conserved residues depicted in bold. The flexible linker region reaching from residue 410 to 468 (insertion) of p44ct, which was replaced by S-N-G-N-G, is marked in blue. Residues, which were mutated and analyzed, are marked by stars. The nomenclature is according to the sequences of ct.

One aim of the present work was to improve the final model of the p34/p44 MC I structure by the use of the higher resolution model of p34/p44 MC II. The fully refined model of the p34/p44 MC II structure was used to obtain the final model of the p34/p44 MC I structure by molecular replacement utilizing the program PHASER. The structure of p34/p44 MC I was refined to a resolution of 3.7 Å with an R_{work} of 19.6% and R_{free} of 24.6% using REFMAC (Table VI.1), which depicts a significant improvement as compared to the previous model of p34/p44 MC I (R_{work} of 27.8% and R_{free} of 32.4%) (180). The new model of p34/p44 MC I contains 200 out of 277 residues of p34_vWA and 74 out of 167 residues of p44_RING. The p34/p44 MC I and p34/p44 MC II structures revealed no significant differences and can be superimposed with a root mean square deviation (rmsd) of 0.7 Å for p34_vWA and 0.8 Å for p44_RING (Figure III.3).

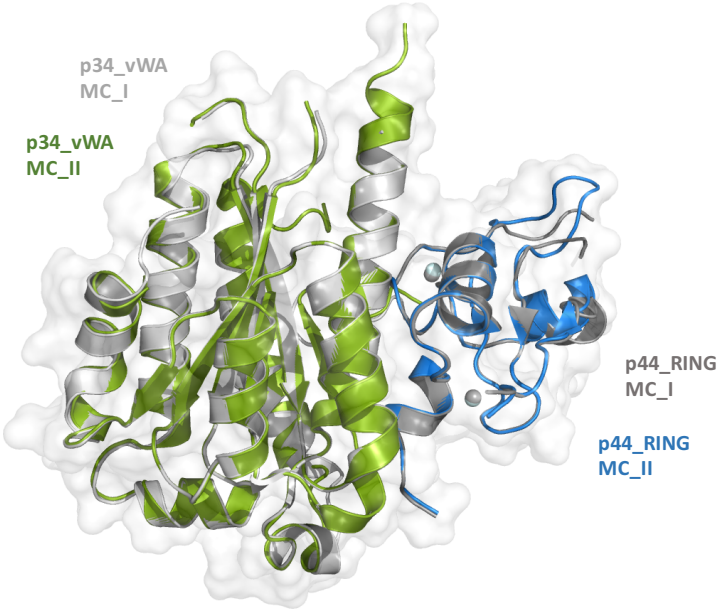
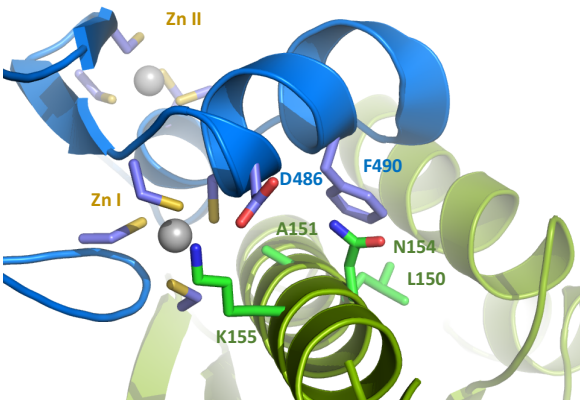


Figure III.3: The p34/p44 MC I and MC II structures. Superposition of the p34/p44 MC I (light grey/dark grey) with the p34/p44 MC II complex (green/blue) resulting in an rmsd of 0.7 Å for p34 and 0.8 Å for p44. The two bound zinc ions are depicted as cyan (MC I) and grey (MC II) spheres. The aligned structures show that the overall fold is the same in both structures.

The p34_vWA/p44_RING structure reveals a highly conserved interface composed of interlocking hydrophobic and polar sections that are responsible for the tight interaction indicated by a K_D of ~ 11 nM (180). A mutational analysis of the p34_vWA/p44_RING interface identified two residues that are crucial for their interaction, K155 of p34 and F490 of p44 (135, 180). The p34_vWA_K155E variant disrupts the electrostatic interaction with D486 of p44_RING and the p44_RING_F490E variant seems to weaken van der Waals contacts in close proximity to the first zinc binding site (Zn I), which was shown to be involved in the interaction to p34_vWA (Figure III.4). In the course of the herein described work a new interface variant, p34_vWA_A151E, which most likely also weakens van der Waals contacts in close proximity to the Zn I binding site, was analyzed (Figure III.4).



◀ **Figure III.4: Interface between p34_vWA and p44_RING.** The part of the interface between p34_vWA (green) and p44_RING (blue) shows the two zinc binding sites (Zn I and Zn II) as well as selected residues. Three of these residues (p34_vWA_A151, p34_vWA_K155 and p44_RING_F490) were mutated to analyze their importance to stabilize the interface. P34_vWA_K155 makes an electrostatic interaction with D486 of p44_RING, the p34_vWA_A151 and p44_RING_F490 residues are involved in van der Waals contacts in close proximity to Zn I. The p34/p44 MC II structure was chosen for the representation of the interface.

III.1.2 Interface variant p34_vWA_A151E

The generation of all novel p34 variants including the respective cloning strategy and their molecular weights is depicted in the appendix in VI.2. All proteins purified for the p34/p44 project were expressed in *E. coli* BL21-CodonPlus® (DE3)-RIL cells (see II.2.1.5) and the cell pellet was either stored at -20 °C (or -80 °C for p44/p62_noTag) or directly used for subsequent purification of the respective protein. After cell lysis and centrifugation (see II.2.2.1) the supernatant was applied to immobilized metal ion affinity chromatography (IMAC) using Ni-TED beads (see II.2.2.2). Fractions containing the respective protein were pooled, concentrated and applied to preparative size exclusion chromatography (SEC) via a HiLoad™ 16/60 Superdex™ 200 (SD 200 16/60) column attached to an Äkta™ system (see II.2.2.4). The final product was concentrated, split into aliquots and stored at -80 °C. Table III.1 summarizes all protein variants of the p34/p44 project including the buffers that were used for their purifications. Several proteins that were used for biochemical analyses were purified according to protocols that were already defined in previous works described either by Dominik Schmitt (DRS) or by Wolfgang Kölmel (WK) (180, 182). An SDS-PAGE gel and chromatogram of representative purifications of novel protein variants are depicted in the appendix as indicated.

Table III.1: Summary of protein variants purified for the p34/p44 project.

Protein variant	Buffer set*	Reference or representative purification
p34	I	DRS
p34_vWA	II	DRS
p44_RING	II	DRS
p44/p62_noTag	V	WK
p44_RING_F490E	II	DRS
p34_vWA_A151E	II	Figure VI.1
p34_18-277	II	Figure VI.2
p34_K155E	I	Figure VI.3
p34_A151E	I	Figure VI.4

*Buffer sets are composed of lysis, elution and SEC buffers and are listed in II.2.2.

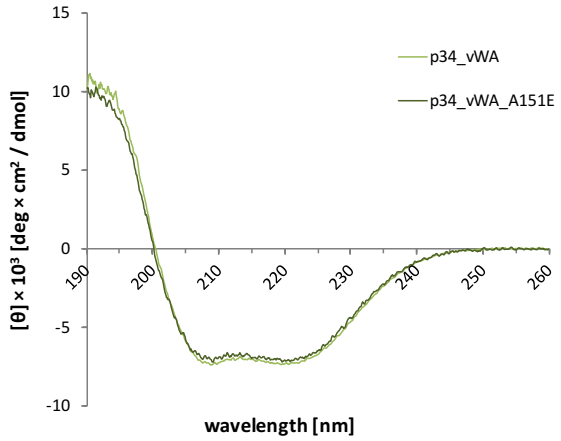


Figure III.5: CD spectra of wild-type p34_vWA and p34_vWA_A151E. CD spectra of p34_vWA and p34_vWA_A151E were recorded at a wavelength of 260 to 190 nm at room temperature. A total of 10 spectra was accumulated for each sample and confirms that the A151E variant displays the same fold as wild-type p34_vWA.

The p34_vWA_A151E variant was generated as indicated above and its impact on the p34_vWA/p44_RING interaction analyzed. The folding integrity of the p34_vWA_A151E variant was verified by circular dichroism (CD) spectroscopy (see II.2.3.4). The CD spectrum of the p34_vWA_A151E variant was highly comparable to that of wild-type p34_vWA confirming that the overall fold of the vWA domain had been maintained (Figure III.5).

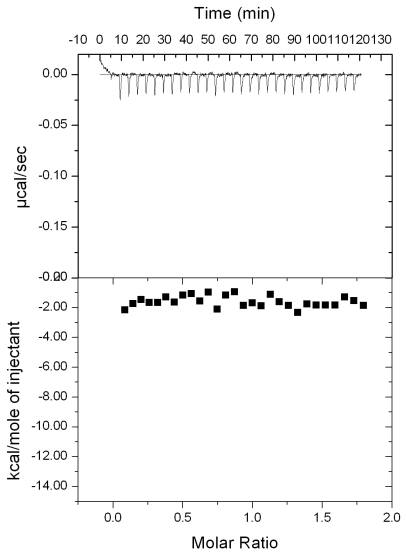


Figure III.6: ITC analysis of p34_vWA_A151E and p44_RING. The ITC analysis of p34_vWA_A151E and p44_RING was conducted according to the same parameters as the ITC analysis of the wild-type proteins using the VP-ITC device and ITC buffer A. However, no heat release could be detected, when 40 µM of p44_RING was titrated into 5 µM of p34_vWA_A151E in 30 injections at 37 °C.

The ability of the p34_vWA_A151E variant to interact with p44_RING was investigated by isothermal titration calorimetry (ITC) and analytical size exclusion chromatography (A-SEC). ITC experiments were performed according to II.2.4.3 using a VP-ITC instrument at 37 °C and ITC buffer A. Previous ITC analyses had already shown that wild-type p34_vWA and p44_RING form a tight interaction indicated by a K_D of ~ 11 nM (180). In contrast, when an equal concentration of p44_RING was titrated to p34_vWA_A151E the interaction was clearly lost (Figure III.6). This

confirmed the hypothesis that the A151E mutation weakens van der Waals contacts in close proximity to the Zn I binding site and thus leads to a loss of interaction between p34_vWA and p44_RING.

Additional analyses by A-SEC experiments using a Superdex^{TM} 200 10/300 GL (SD 200 10/300) column attached to an Äkta^{TM} system and A-SEC buffer B were performed as described in II.2.4.1. The p34_vWA, p34_vWA_A151E and p44_RING samples were applied individually and in complex with each other onto the same column. For these A-SEC experiments an optimized wild-type p34_vWA variant was used, which lacks the first 17 flexible amino acids (p34_18-277).

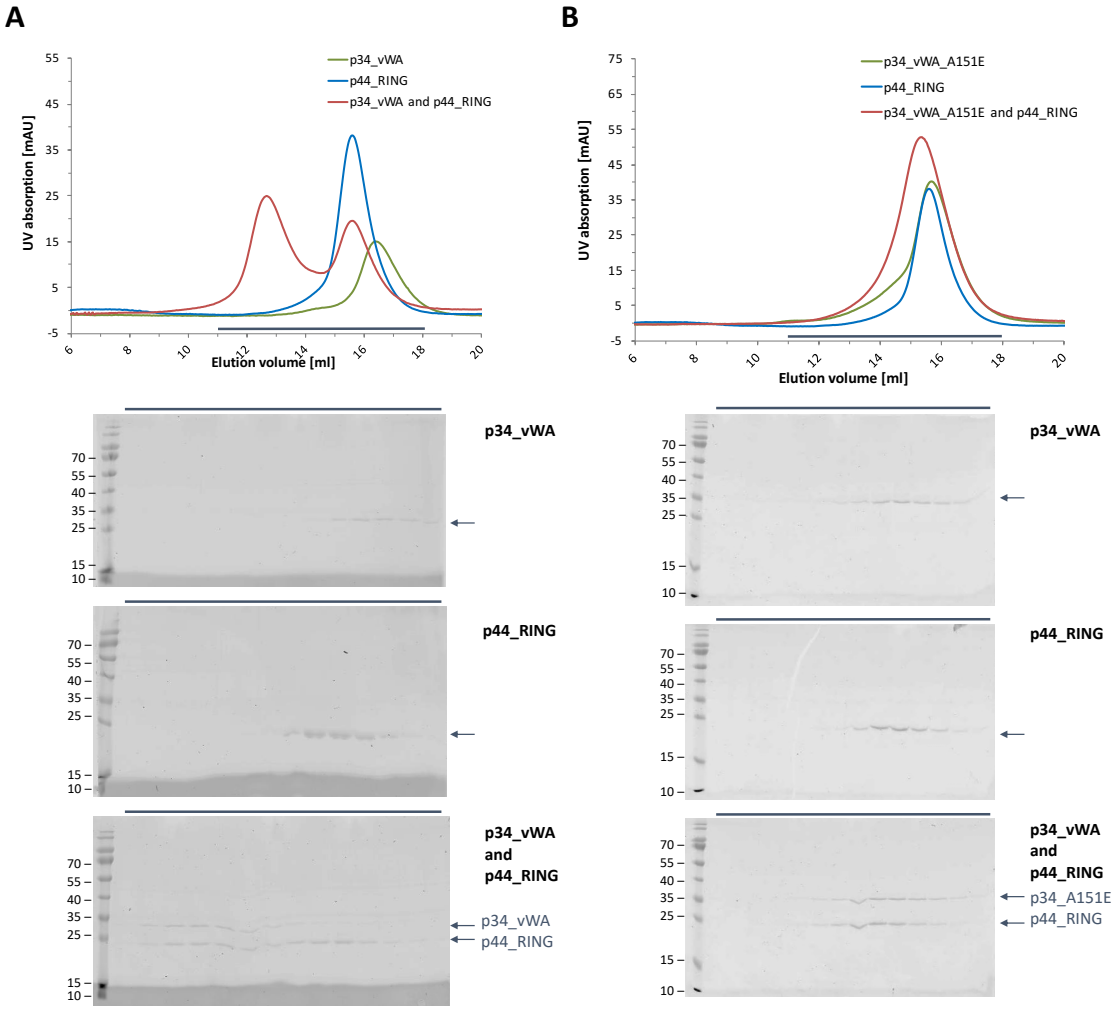


Figure III.7: A-SEC of p34_vWA WT or A151E and p44_RING. (A) A-SEC (SD 200 10/300) of the individual p34_vWA (green) and p44_RING (blue) proteins and both proteins together (red) shows complex formation indicated by a peak shifted to earlier elution volumes (12.5 ml). For this A-SEC experiment an optimized p34_vWA variant (p34_18-277) lacking the first 17 flexible amino acids was used. **(B)** A-SEC of the individual p34_vWA_A151E (green) and p44_RING (blue) proteins and both proteins together (red) does not show any complex formation. To verify the results seen on the chromatogram, fractions from a defined elution range (indicated by the bar) were taken and analyzed via SDS-PAGE (arrows indicate the respective proteins; the first lane always depicts the marker (kDa)).

In line with results described in the doctoral thesis of Dominik Schmitt (180) p34_vWA and p44_RING form a tight complex with each other indicated by a peak shifted to earlier elution volumes (12.5 ml) (Figure III.7 A). Complex formation was verified by SDS-PAGE of samples that were taken within the same elution range of each of the wild-type and the mutant chromatography steps (11 - 18 ml) (Figure III.7 A). Furthermore, all A-SEC analyses were performed consecutively using the same column and without changing the setup at the Äkta™ system. The chromatogram and SDS-PAGE analysis clearly show that the p34_vWA_A151E variant abolished complex formation with p44_RING (Figure III.7 B). This loss of interaction obtained for p34_vWA_A151E/p44_RING resembles the disrupted interaction observed for p34_vWA_K155E/p44_RING and p34_vWA/p44_RING_F490E (180). Thus, a third residue identified in the interface of the crystal structure, A151 of p34, was verified to be crucial for the interaction between p34_vWA and p44_RING.

III.1.3 Impact of interface mutations on TFIIH

The identification of three prominent mutations (A151E and K155E of p34, F490E of p44) that clearly disrupt the interface between p34_vWA and p44_RING led to the analysis of a disrupted p34/p44 interaction on the integrity and function of TFIIH. For this purpose, the mutations that disrupted the interaction between the p34_vWA/p44_RING domains were introduced and characterized in the full-length proteins and the interaction of full-length p34_A151E or p34_K155E with full-length p44 was investigated first.

Unfortunately, full-length p44 is forming higher oligomers, when purified on its own (180). Thus, a complex of p44 and its interaction partner p62 (p44/p62_noTag), which stabilizes p44, was used for this approach. Beforehand, it was ensured that p62 is not interacting with p34 and that p62 is just a stabilization factor, which is not influencing the interaction between p44 and p34. As native PAGE analyses were difficult to interpret due to the fact that p34, p44/p62 and p62 appear at similar heights in the gel an A-SEC analysis of p34 and p62 was performed. The full-length p62 protein used for A-SEC was kindly provided by Wolfgang Kölmel. A-SEC of p34 and p62 was performed using a SD 200 5/150 column and A-SEC buffer A. The p34 and p62 samples were applied individually and in complex with each other onto the same column. The chromatogram and SDS-PAGE indicate that p34 and p62 do not interact with each other (Figure III.8). Thus, the p44/p62 complex was used for testing the interaction between p44 and p34 with p62 being only a stabilization factor.

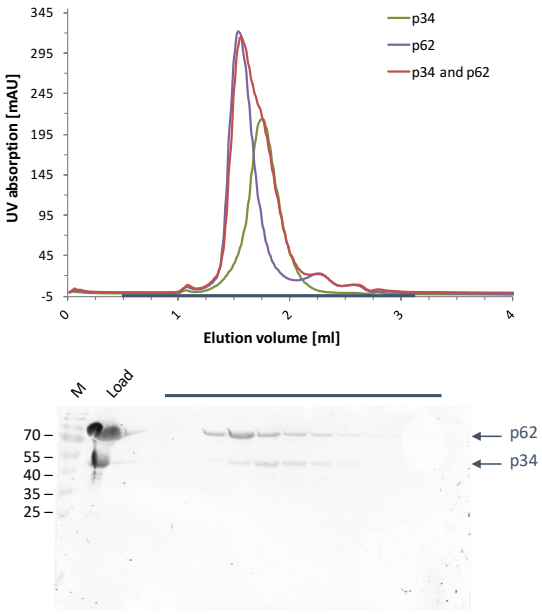
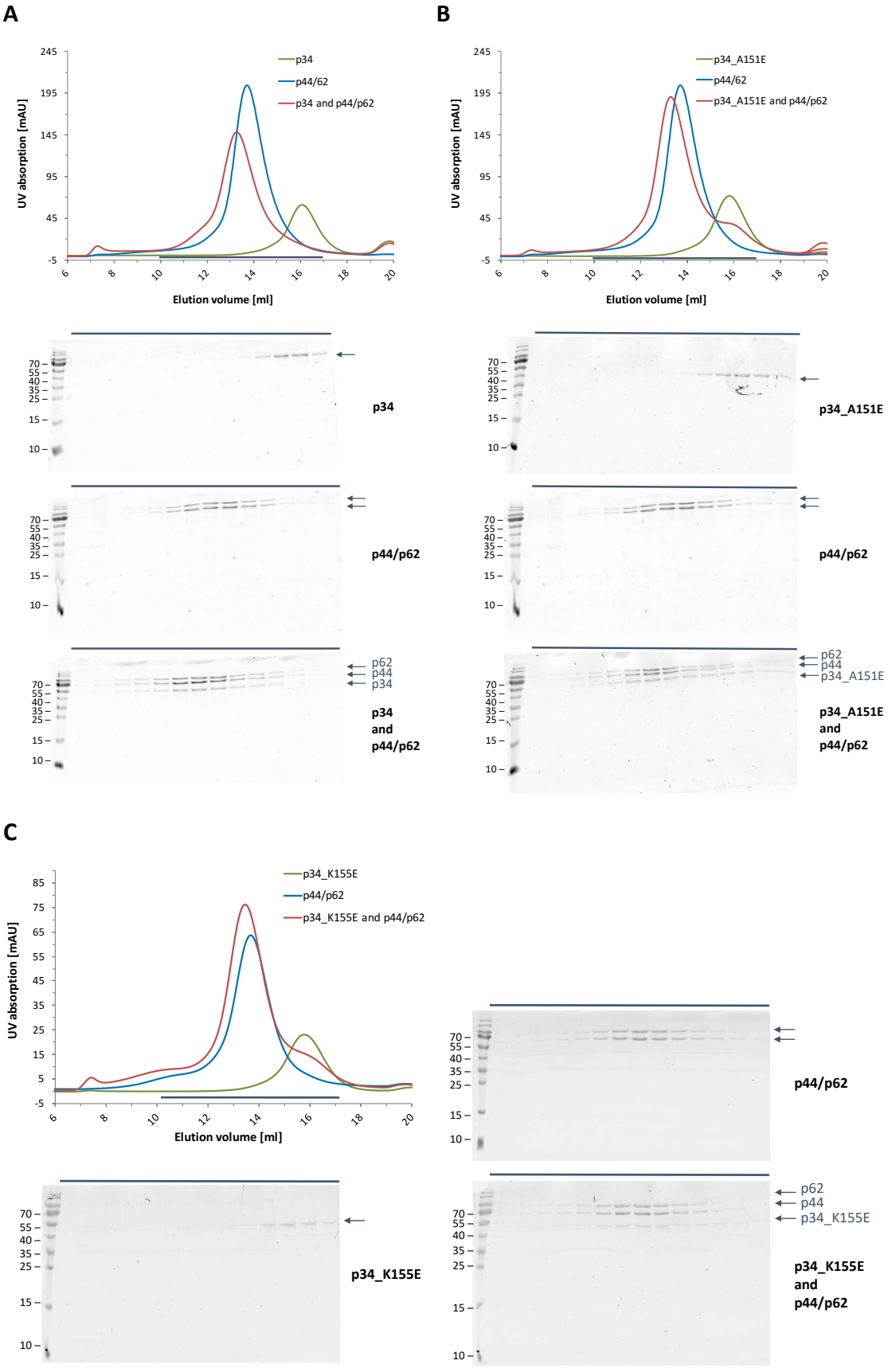


Figure III.8: A-SEC of p34 and p62. A-SEC (SD 200 5/150) of the individual full-length p34 (green) and p62 (purple) proteins and both proteins together (red) shows no complex formation. The result was verified by SDS-PAGE of samples from a defined elution range (indicated by the bar). Abbreviation: M = marker (kDa).

A-SEC experiments of full-length p34_A151E or p34_K155E with p44/p62 were performed using A-SEC buffer B. Due to the large molecular weight of a p34/p44/p62 complex (182 kDa) a Superose^{TM} 6 10/300 GL (SR 6 10/300) column was used, which optimized the separation of the sample in comparison to a SD 200 10/300 column. The p34 variants and p44/p62 were applied individually or in complex with each other onto the same SR 6 10/300 column. Wild-type p34 forms a complex with p44/p62 indicated by a peak shifted to earlier elution volumes and visualized via SDS-PAGE (Figure III.9 A). Surprisingly, the A151E variant of p34 was still able to interact with p44/p62, although it was shown to disrupt the interaction between the p34_vWA and p44_RING domains (Figure III.7 B). The SDS-PAGE analysis clearly showed that p34_A151E was shifted to earlier elution volumes illustrating successful complex formation of p34_A151E and p44/p62. However, the chromatogram of p34_A151E and p44/p62 showed a little shoulder that could indicate a weakened interaction of p34_A151E and p44/p62. The samples taken for the SDS-PAGE analysis of the p34_A151E A-SEC experiments covered the same elution range as the wild-type samples (10 - 17 ml). A-SEC analyses of the K155E variant of p34 revealed an interaction with p44/p62 as well (Figure III.9 C). However, the complex formed between p34_K155E and p44/p62 appeared to be weaker than the complex formed between p34_A151E and p44/p62.


A
B
C
UV absorption [mAU]
Elution volume [ml]
p34
p44/62
p34 and p44/62
p34_A151E
p34_A151E and p44/62
p34_K155E
p34_K155E and p44/62
p34
p44/62
p34 and p44/62
p34_A151E
p34_A151E and p44/62
p34_K155E
p34_K155E and p44/62
p62
p44
p34
p34_A151E
p34_K155E

◀ **Figure III.9: A-SEC of p34 WT or A151E or K155E and p44/p62. (A)** A-SEC (SR 6 10/300) of individual wild-type p34 (green) and p44/p62 (blue) samples and combined (red) shows complex formation indicated by a peak shifted to earlier elution volumes. The complex is also visible on SDS-PAGE. **(B)** A-SEC of p34_A151E (green) and p44/p62 (blue) alone and combined (red) and **(C)** A-SEC of p34_K155E (green) and p44/p62 (blue) alone and combined (red) also show complex formation for both variants. The SDS-PAGE gels (arrows indicate the respective protein variants; the first lane always depicts the marker (kDa)) show fractions that were taken from a defined elution range (indicated by the bar). The corresponding wild-type run of (C) is not shown.

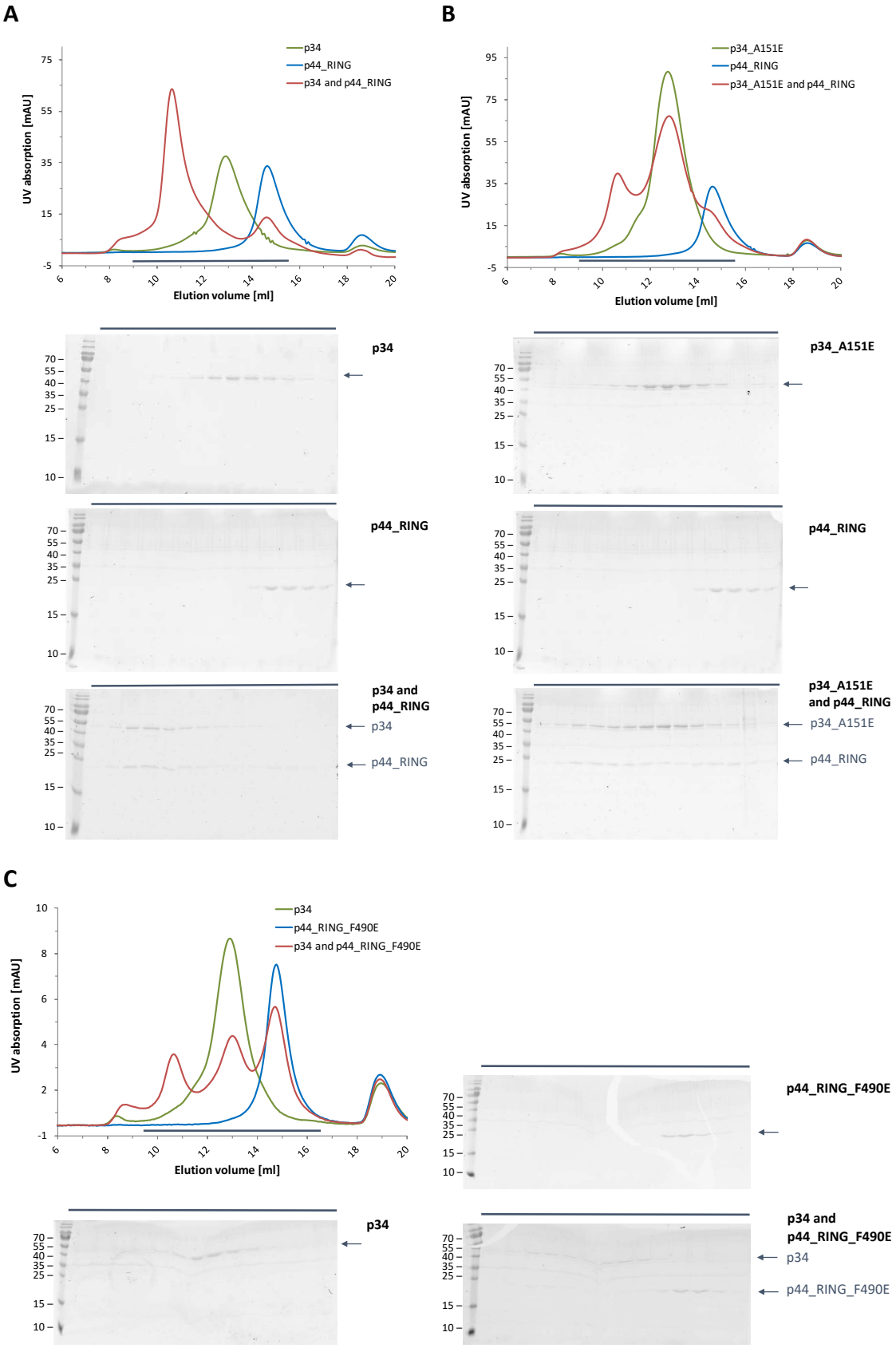
Interestingly, mutations that abolished an interaction between the p34_vWA and p44_RING domains had, if at all, only minor effects on the full-length proteins. An interaction between p34 and p62 could be excluded due to previous findings that showed that p34 and p62 were not interacting with each other. Furthermore, no interaction between p34 and p44 other than via its p34_vWA and p44_RING domains as present in the crystal structure had been reported in the literature. However, the A-SEC results of the A151E and K155E variants of p34 suggested the presence of an additional interface between p34 and p44 that stabilizes their interaction.

Further experiments performed by our collaboration partners (groups of Arnaud Poterszman and Jean-Marc Egly) strengthened our hypothesis. A western blot analysis of recombinant human core TFIIH subunits did not show an effect on the composition of TFIIH, when individual p34_vWA/p44_RING interface variants had been used instead of the wild-type proteins (181). Furthermore, the activity of TFIIH was still intact as these variants were able to accurately perform RNA synthesis in *in vitro* transcription assays and dual incision in *in vitro* NER assays.

III.1.4 Characterization of the additional interface

Since our experiments pointed towards an additional interface between p34 and p44 that was so far uncharacterized A-SEC experiments were performed to investigate the possible presence of an additional interface between p34 and p44 in greater detail. In addition, the new interface was intended to be characterized structurally.

To distinguish whether the additional interface is constituted by p34, p44 or a combination of both A-SEC experiments were conducted. We initiated our studies, in which the interface between full-length p34 and p44_RING was characterized. To this end we utilized the variants A151E (p34) or F490E (p44_RING) to “disable” the known interface thereby ensuring that only an additional interaction permits complex formation. This approach permitted us to determine the influence of the C4 domain of p34 (Figure III.1) towards complex formation. A-SEC experiments were performed either between p34 and p44_RING, the full-length p34_A151E variant and p44_RING or between p34 and the p44_RING_F490E variant.


A
p34
p44_RING
p34 and p44_RING
UV absorption [mAU]
Elution volume [ml]
p34
p44_RING
p34 and p44_RING
p44
B
p34_A151E
p44_RING
p34_A151E and p44_RING
UV absorption [mAU]
Elution volume [ml]
p34_A151E
p44_RING
p34_A151E and p44_RING
p34_A151E
p44
C
p34
p44_RING_F490E
p34 and p44_RING_F490E
UV absorption [mAU]
Elution volume [ml]
p44_RING_F490E
p34
p34 and p44_RING_F490E
p34
p44_RING_F490E

◀ **Figure III.10: A-SEC of p34 WT or A151E and p44_RING WT or F490E. (A)** A-SEC (SD 200 Increase 10/300) of individual wild-type p34 (green) and p44_RING (blue) samples and combined (red) shows complex formation indicated by a peak shifted to earlier elution volumes (10.5 ml) and visible on the SDS-PAGE. **(B)** A-SEC of p34_A151E (green) and p44_RING (blue) alone and combined (red) and **(C)** A-SEC of p34 (green) and p44_RING_F490E (blue) alone and combined (red) also show complex formation for both variants. The SDS-PAGE gels (arrows indicate the respective protein variant; the first lane always depicts the marker (kDa)) show fractions that were taken from a defined elution range (indicated by the bar). The corresponding wild-type run of (C) is not shown.

All A-SEC experiments were performed using a SD 200 Increase 10/300 column and A-SEC buffer B. All samples were applied individually and in complex with each other onto the same column. Wild-type p34 formed a complex with p44_RING as indicated by a peak shifted to earlier elution volumes (10.5 ml) and visualized via SDS-PAGE (Figure III.10 A). The A151E variant of p34, which disrupted the interaction between p34_vWA and p44_RING, was still able to interact with p44_RING although this interaction was reduced as compared to the interaction of the wild-type proteins (Figure III.10 B). The samples taken for the SDS-PAGE analysis of the p34_A151E A-SEC experiments covered the same elution range as the wild-type samples (9 - 15.5 ml) and strikingly showed that p34_A151E was shifted to earlier elution volumes (10.5 ml) as it was also indicated by the chromatogram. Furthermore, an interaction between full-length p34 and the p44_RING_F490E variant, which disrupted the interaction to p34_vWA, was observed (Figure III.10 C). Although this interaction could hardly be seen on the SDS-PAGE gel, the peak at 10.5 ml clearly indicated complex formation between p34 and p44_RING_F490E. These findings verified the hypothesis that there is an additional interface between p34 and p44 and moreover showed that this additional interface involves the C4 domain of p34 (p34_C4). The A-SEC results revealed that the interaction between the p34_C4 and p44_RING domains permits complex formation even when the interaction between the p34_vWA and p44_RING domains is lost.

Our results were further supported by experiments performed by our collaboration partners (groups of Arnaud Poterszman and Jean-Marc Egly) that underline the importance of the p34_C4 domain for the p34/p44 interaction and TFIIH. Automated Hydrogen-Deuterium eXchange coupled to Mass Spectrometry (HDX-MS) experiments of human p34 in the presence and absence of p44 did not only validate the p34_vWA/p44_RING interface present in the crystal structure, but moreover, identified a region in the C4 domain of p34 that showed a significant reduction in exchange (181). Pull-down experiments of recombinant human core TFIIH subunits showed that the incorporation of p62, p52 and XPB was affected in complexes containing a truncated p34 lacking the C4 domain (181). Furthermore, the activity of a core TFIIH complex containing a truncated p34 that lacks the C4 domain was impaired in RNA synthesis and in dual incision as observed in *in vitro* transcription assays and in *in vitro* NER assays, respectively (181).

In order to visualize the additional interface between the p34_C4 and p44_RING domains we co-expressed and co-purified different p34 and p44 variants. Crystallization trials of a complex of co-purified full-length p34 and p44_RING yielded tiny needle-like crystals that resembled those of the p34_vWA/p44_RING minimal complex. Unfortunately, data sets of these crystals did not show any additional density for the p34_C4 domain. Analysis of the crystal content by a silver stained SDS-PAGE gel of the crystals further indicated that full-length p34 had been degraded during the crystallization procedure and that crystallization of the p34_vWA/p44_RING minimal complex was preferred. Different attempts to stabilize the interaction between p34_C4 and p44_RING as well as co-purifications of further p34/p44 complexes were tested but remained unsuccessful. These efforts are described in more detail in the Master thesis of Stefan Peißert (183).

III.1.5 The p34/p44 interaction within TFIIH

After completion of the p34/p44 analysis, two cryo-EM structures of human and yeast TFIIH at resolutions of 4.4 Å and 4.7 Å, respectively, were published (97, 98). These cryo-EM structures allowed us to validate the findings on the p34/p44 interaction in the context of TFIIH. In comparison to the human TFIIH structure (97), which depicts only parts of p34 and p44, both proteins could be modelled entirely in the cryo-EM structure of the PIC-TFIIH complex from *S. cerevisiae* (98). Thus, the cryo-EM structure of yeast TFIIH was taken as a template to analyze the high-resolution structure of p34_vWA/p44_RING from *C. thermophilum* in the environment of TFIIH. An overview of the donut-shaped TFIIH cryo-EM structure is depicted in Figure III.11 A.
Despite the resolution of 4.7 Å and the fact that all domains of p34 and p44 could be built into the electron density of the cryo-EM structure of yeast TFIIH, many parts were built as poly-alanine stretches and many stretches of the sequence could not be assigned. The yeast and *C. thermophilum* p34_vWA/p44_RING complexes are structurally similar and can be superimposed with an rmsd of 1.6 Å for p34 and 1.9 Å for p44 (Figure III.11 B+C) thus further supporting our identified p34_vWA/p44_RING interaction as obtained for the isolated domains. Interestingly, the p34_C4 domain, which could be modelled in the yeast TFIIH structure, shows an interaction with p44_RING via an additional α-helix of p44_RING. This α-helix of p44_RING is located at the very N-terminus of this domain (amino acids 368 - 379 in ct) and could not be depicted structurally in the crystal structure of p34_vWA/p44_RING from *C. thermophilum* (Figure III.11 D). Nevertheless, this finding verified the presence of an additional interface between the p44_RING and p34_C4 domains.

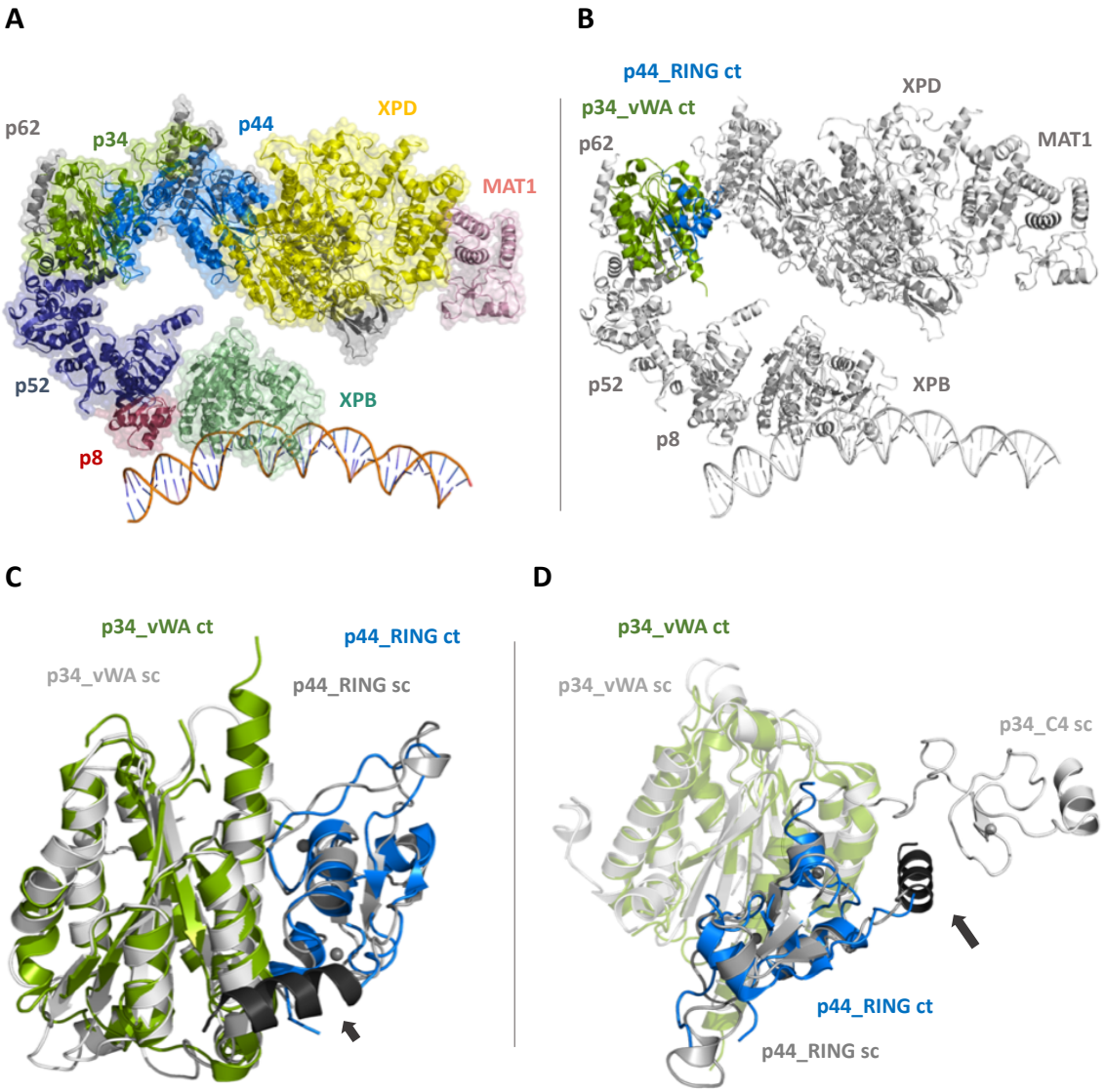


Figure III.11: Comparison of the p34/p44 interaction from *C. thermophilum* without and within TFIIH from *S. cerevisiae*. (A) The TFIIH structure is shown as a cutout of the cryo-EM structure of the PIC-TFIIH complex from *S. cerevisiae* (PDB code: 5OQJ). The structure comprises XPB (HD1 and HD2), XPD, p62, p52, p44, p34, p8 and the N-terminal part of MAT1. The DNA duplex is bound by XPB. **(B)** The p34_vWA/p44_RING MC II crystal structure from *C. thermophilum* (ct; green/blue) was built into the TFIIH cryo-EM structure of the PIC-TFIIH complex from *S. cerevisiae* (sc; grey). **(C)** The p34_vWA and p44_RING domains from *C. thermophilum* (green/blue) and *S. cerevisiae* (grey) can be superimposed with an rmsd of 1.6 Å for p34 and 1.9 Å for p44. **(D)** Depiction of the interface between p34_C4 and an additional α-helix in p44_RING as it could be modelled in the cryo EM structure. The additional α-helix of p44_RING is depicted in black and marked by an arrow in (C) and (D). The two zinc ions bound by p44_RING and the zinc ion bound by p34_C4 are depicted as grey spheres in (C) and (D).

III.2 The Interaction Network of XPB, p52 and p8

This section describes the structural and functional analysis on the interaction between XPB, p52 and p8 from *C. thermophilum*. Apart from the structural characterization of the C-terminal domain of p52 and its interaction with p8 (126) no structural information on p52 was so far available. Hence, this work focused on the structural characterization of the entire p52 protein, which in turn permitted the analysis of the interaction between p52 and its binding partners. The interaction between p52 and XPB had been partially investigated in previous studies and two binding sites for XPB within p52 had been proposed, the second of which was shown to be crucial to stimulate XPB's ATPase activity, which is indispensable for NER and transcription (105, 123). However, the lack of a structure of p52 or of p52 and its binding partners made it so far impossible to elucidate how this stimulation leads to the activation of XPB's ATPase activity.

III.2.1 Construct design of p52 variants

Different strategies were pursued to identify p52 variants that were suitable for crystallization and subsequent structure determination. Unfortunately, full-length p52 was not stable and prone to precipitation when purified on its own. By coincidence, one purification of full-length p52 yielded a stable protein sample, which also crystallized but the purification could not be reproduced and thus crystal improvement was limited. One feasible alternative of producing stable full-length p52 was its co-purification with p8 (p52_noTag/p8) as described in the Master Thesis of Elisabeth Schönwetter (148). It was observed that protein samples containing full-length p52 were prone to precipitate when subjected to temperature changes from 4 °C to room temperature (RT) or vice versa, thus purifications of p52 and the p52/p8 complex were also conducted at RT instead of the standard purification at 4 °C, which seemed to improve their stability.

In addition to purification and crystallization attempts of full-length p52, various shortened p52 variants were designed and analyzed. Construct design of shortened p52 variants was assisted by secondary structure predictions as defined by the bioinformatic tools iTASSER, RaptorX and Phyre2. Table III.2 summarizes different p52 variants that were tested for expression, purification and crystallization. A Table summarizing the generation of these p52 variants can be found in the appendix in VI.2. The crystal structures of two of these shortened p52 variants, p52_121-EdL and p52_1-321, could be solved and their purification, crystallization and structure solution will be described in more detail in the following.

Table III.2: Summary of p52 variants tested for crystallization.

p52 variant	Soluble expression*	Purification	Initial crystallization screens**	Crystals
p52	yes	yes, but not stable	9x standard screens***	no
p52 (not reproducible)	yes	yes, stable	9x standard screens	yes
p52 @RT	yes	@RT: yes	Crystal Screen™ 1 + 2, JCSG+, Nextal - PEG Suite, Nextal pH Clear, Nucleix Suite and Protein Complex Suite	no
p52_noTag/p8	yes	yes	9x standard screens	no
p52_noTag/p8 @RT	yes	@RT: yes	$AmSO_4$ Suite, Crystal Screen™ 1 + 2, Index, JCSG+, Nextal - PEG Suite, Nextal pH Clear, Nucleix Suite and Protein Complex Suite	no
p52_1-321	yes	yes	9x standard screens and 96-well fine screen 21	yes
p52_1-321_dL2	yes	yes	Index, JCSG+, Nextal - PEG Suite II, Nextal pH Clear, Nextal pH Clear II, Protein Complex Suite, Topaz™ OptiMix PEG, Wizard 3 + 4 and 96-well fine screens 23 - 25	yes
p52_1-349	yes	yes	9x standard screens	yes
p52_1-454	yes	yes	9x standard screens and JCSG+	yes
p52_121-349	yes	yes	9x standard screens	no
p52_121-454_dL	yes	yes	8x standard screens (without Topaz™ OptiMix PEG) and MbClass II Suite, Nextal - PEG Suite II, Wizard 3 + 4, 96-well fine screens 22, 23 and 29	yes
p52_121-End	yes	yes	Crystal Screen™ 1 + 2, Nextal - PEG Suite, Nextal pH Clear and 96-well fine screen 14	yes
p52_121-EdL	yes	yes	9x standard screens	yes
p52_121-EdL_CtermHis	yes	yes, but very low yield	Crystal Screen™ 1 + 2, JCSG+, Protein Complex Suite and Wizard 1 + 2	no

*soluble expression was tested in *E. coli* BL21-CodonPlus® (DE3)-RIL cells

**initial crystallization screening was performed according to II.2.5.1

***the nine standard screens comprise the Crystal Screen™ 1 + 2, Index, Nextal - PEG Suite, Nextal pH Clear, Nucleix Suite, Protein Complex Suite, Topaz™ OptiMix 3, Topaz™ OptiMix PEG and Wizard 1 + 2 screen

III.2.2 Structural characterization of p52_121-EdL

The p52_121-EdL variant lacks the N-terminal 120 amino acids but includes the second proposed XPB binding site (amino acids 354 - 435) and the C-terminal domain (amino acids 436 - 514), which is known to interact with p8. Sequence alignments of human and *C. thermophilum* p52 (p52hs and p52ct) performed by ClustalW and COBALT identified linker regions that are only present in p52ct but are missing in p52hs as it has already been described for p34 and p44 (III.1.1). One of those insertion regions of p52ct is reaching from amino acid 322 to 344. Based on the successful results of replacing an insertion region in p44_RING by an artificial linker, the insertion region reaching from amino acid 322 to 344 was replaced by an artificial linker consisting of the amino acids S-N-G-N-G in the p52_121-End variant resulting in the optimized variant p52_121-EdL (p52_121-End_ΔLinker).

III.2.2.1 Production of p52_121-EdL

The p52_121-EdL variant was inserted into the pETM-11 vector containing an N-terminal His_6-tag and a TEV cleavage site and its expression was conducted in *E. coli* BL21-CodonPlus® (DE3)-RIL cells. The cell pellet was either stored at -80 °C or directly used for subsequent purification. The purification was performed by IMAC using Ni-TED beads, followed by SEC via a SD 200 16/60 column (Figure III.12 A+B) and conducted with buffer set IV. Fractions containing the desired protein were selected, pooled and the final product was concentrated to 11 - 22 mg/ml. The final sample was either used for subsequent crystallization trials or split into aliquots and stored at - 80 °C. Attempts to cleave the N-terminal His_6-tag of the p52_121-EdL variant by the TEV protease during the purification as described in II.2.2.3 remained unsuccessful (Figure VI.5). Interestingly, a SEC-MALS analysis of p52_121-EdL (44.4 kDa) showed that the protein forms a dimer in solution indicated by an average mass of 86.36 ± 0.52 kDa (Figure III.12 C).

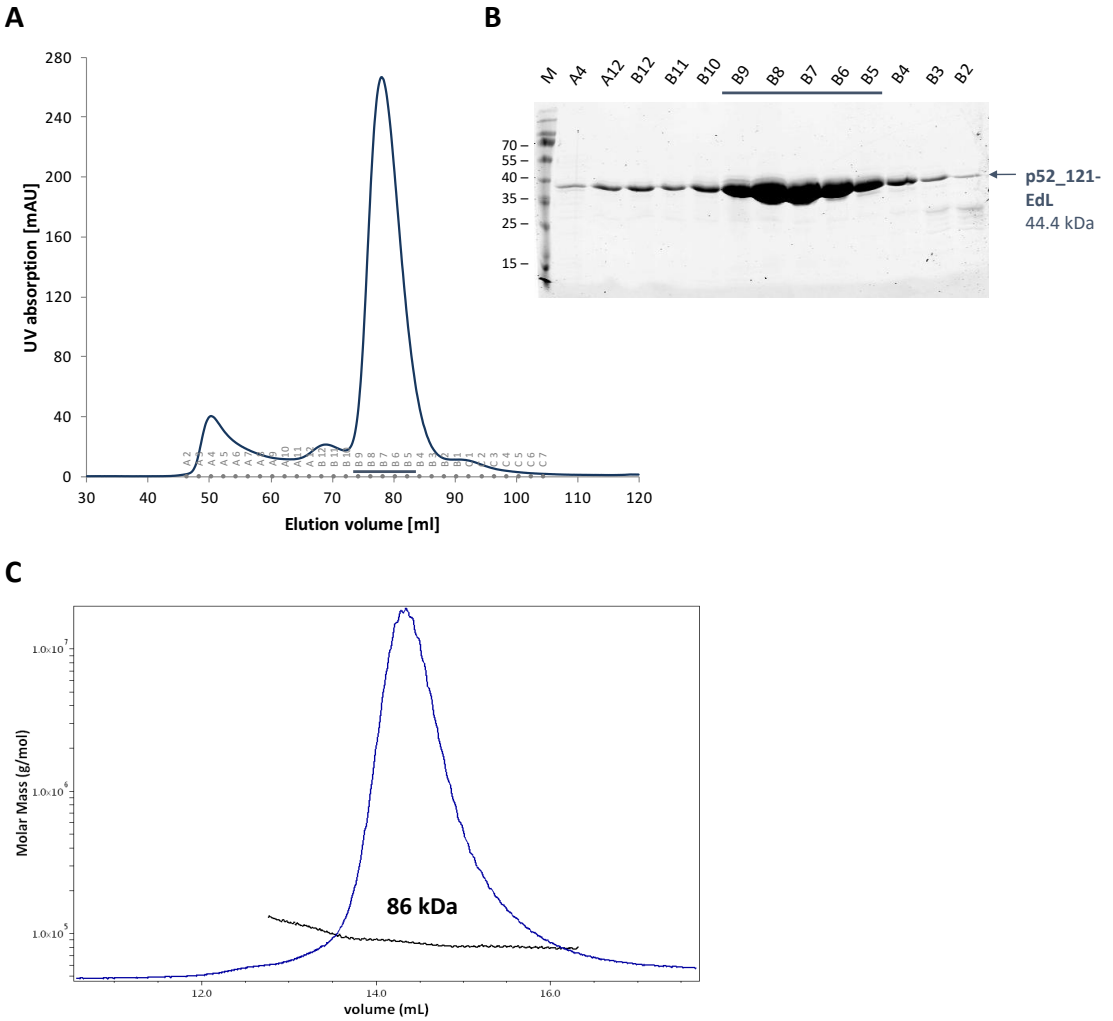


Figure III.12: Size exclusion chromatography and SEC-MALS of p52_121-EdL. (A) Chromatogram and **(B)** SDS-PAGE of the size exclusion chromatography (SD 200 16/60) of p52_121-EdL (44.4 kDa). The chromatogram shows the UV absorption at 280 nm. The bar indicates fractions that were finally pooled. Abbreviation: M = marker (kDa). **(C)** The SEC-MALS analysis (SD 200 16/60) of p52_121-EdL shows a relatively homogenous peak with an average mass of 86.36 ± 0.52 kDa indicating dimer formation of p52_121-EdL.

III.2.2.2 Crystallization of p52_121-EdL

Initial crystallization trials of the p52_121-EdL variant were performed by the sitting drop method using the crystallization robot HoneyBee 963 (see II.2.5.1) and initial crystals were improved in several 24-well and 96-well fine screens through optimization of the most promising crystallization conditions. Crystals of p52_121-EdL grew at concentrations of 5 - 6 mg/ml as very thin plates and in more than 20 different crystallization conditions (Figure III.13). Most of the crystals appeared already after 1 - 5 days, but some larger plates took 10 - 14 days to appear. The crystals grew preferably in the sitting drop method and displayed a very anisotropic diffraction

pattern. Many plates grew on top of each other, which additionally complicated obtaining single plates for data collection. Crystallization conditions varied, but most often contained either PEG 400, PEG 3350 or PEG 4000. The buffer of the crystallization conditions varied from MES pH 6.0 to HEPES pH 7.0 or Tris pH 8.5 and often salts like lithium citrate tribasic, $CaCl_2$ or $MgCl_2$ were additionally present.

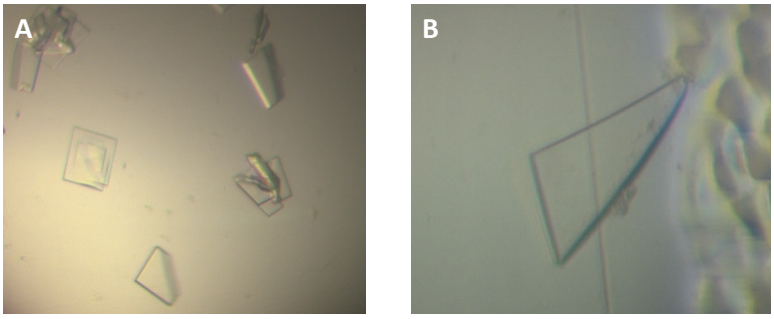


Figure III.13: Protein crystals of p52_121-EdL. Crystals obtained of p52_121-EdL in **(A)** with the precipitant solution containing 80 mM $CaCl_2$, 50 mM sodium cacodylate pH 6.5, 1 mM cobalt (III) hexamine chloride and 2 mM spermine and **(B)** a fine screen condition containing 100 mM $MgCl_2$, 100 mM HEPES pH 7.5 and 5% PEG 4000.

Diffraction of most of the p52_121-EdL crystals was limited to 4 - 3 Å. In order to solve the phase problem for the data of p52_121-EdL a Se-Met derivative of p52_121-EdL was expressed as described in II.2.1.6 and purified and crystallized in the same way as native p52_121-EdL. The p52_121-EdL Se-Met derivative crystals grew also as thin plates and one of them diffracted up to 2.6 Å and showed a comparably high anomalous signal, which was a great improvement compared to the moderate diffraction and moderate anomalous signal of all other Se-Met crystals. This crystal grew in a 96-well fine screen condition containing 100 mM HEPES pH 7.0 and 7.5% PEG 4000.

III.2.2.3 Structure of p52_121-EdL

The p52_121-EdL variant contains ten methionine residues, thus structure solution was attempted by using the anomalous scattering of the selenium atoms in a p52_121-EdL Se-Met derivative for phasing by the SAD method. Structure solution was complicated by the fact that all plate like crystals grew in space group P1 and displayed a very anisotropic diffraction pattern. Nevertheless, the structure of p52_121-EdL could be solved by SAD phasing using one data set of a p52_121-EdL Se-Met derivative crystal, which diffracted to a maximum resolution of 2.6 Å. The data set was recorded at the selenium peak at 0.979770 Å (12,656 eV) to enhance the anomalous scattering. The structure was solved in space group P1 and contained four molecules in the

asymmetric unit. Structure solution was performed by applying the automated structure solution pipeline SHARP and the model building program buccaneer. SHARP identified initially 7 selenium atoms of the 10 total sites per p52 molecule in the asymmetric unit. The initial model was improved by combining density modification with the phases from the model. After subsequent model building and refinement cycles using the buccaneer/REFMAC pipeline the model was further improved by manual model building. For later stages of the refinement process data were improved by utilization of the STARANISO server due to the anisotropy of the initial data. The final model was refined to an R_{work} of 23.6% and R_{free} of 26.0% using Phenix refine (Table III.3).

Table III.3: Data collection and refinement statistics of p52_121-EdL and p52_121-End.

Data collection	**p52_121-EdL SAD**	**p52_121-End**
Space Group	P 1	$P2_1$
Cell Dimensions		
a, b, c [Å]	60.5, 83.2, 86.0	86.0, 107.7, 91.4
α, β, γ [°]	82.7, 80.0, 77.5	90.0, 90.0, 90.0
Resolution [Å]	49.16 – 2.60 (2.69 – 2.60)	46.41 – 3.00 (3.15 – 3.00)
Wavelength [Å]	0.979770	0.96500
Unique Reflections	48,667 (4,455)	33,483 (4,429)
I/σI	9.4 (1.5)	3.3 (0.6)
CC (1/2)	0.995 (0.494)	0.947 (0.353)
Rmerge [%]	20.1 (134.2)	93.9 (466.3)
Completeness [%]	98.7 (98.0)	99.8 (99.9)
Redundancy	7.0 (7.1)	7.6 (7.7)
Anomalous Completeness	95.9 (95.2)	-
Anomalous Multiplicity	3.5 (3.6)	-
Refinement		
Resolution [Å]	49.16 – 2.60	
Unique Reflections	41,609	
Number of Atoms	8,767	
R_{work} (R_{free}) [%]	23.6 (26.0)	
Mean B-Factor [Å^2]	31.89	
Bond Lengths [Å]	0.002	
Bond Angles [°]	0.467	
Ramachandran Statistics (favored/allowed/outliers)* [%]	96.09/3.72/0.19	

Values in parentheses refer to the highest resolution shell. *Ramachandran statistics were performed using MolProbity.

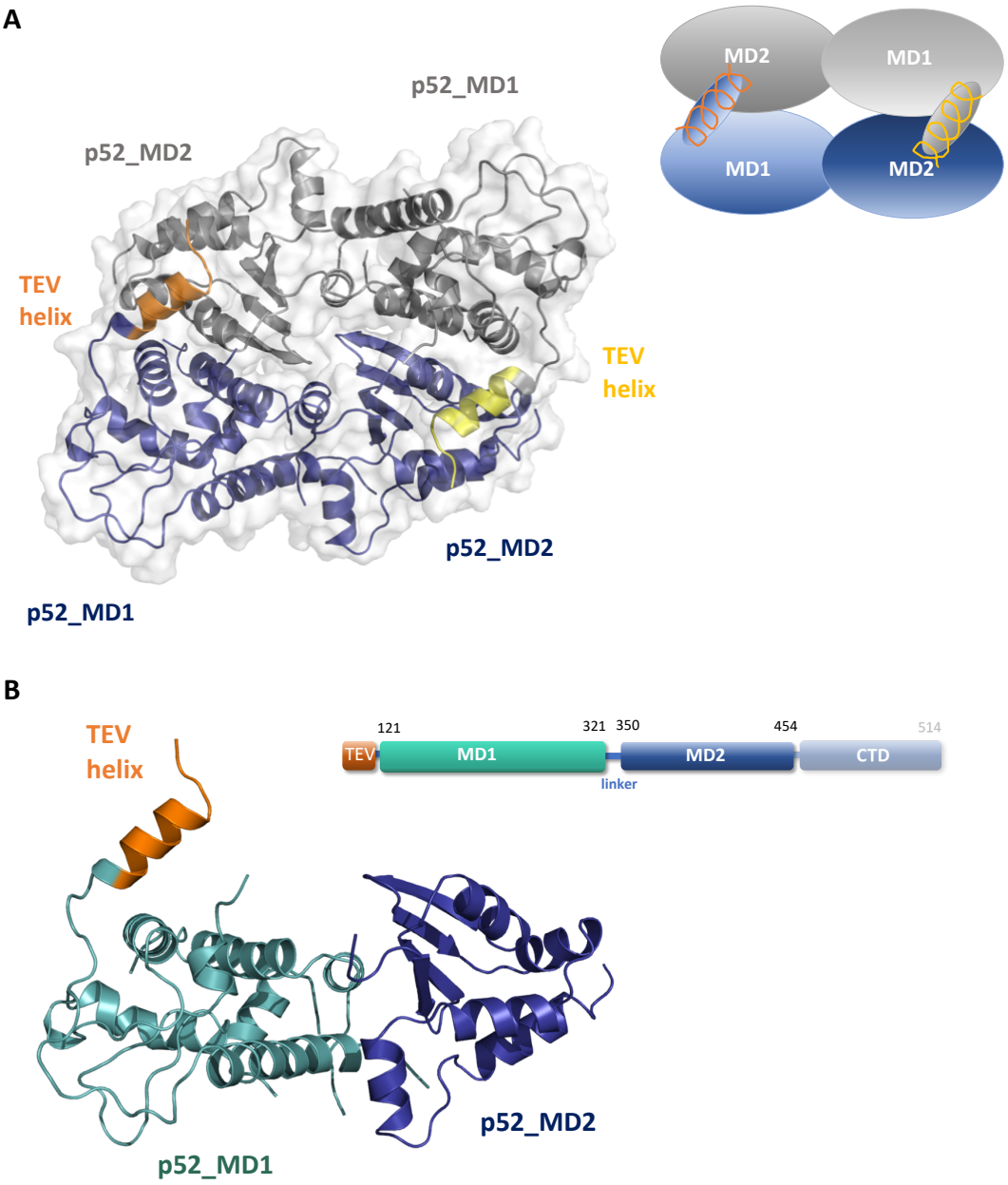


Figure III.14: Crystal structure of p52_121-EdL. (A) Structure of a p52_121-EdL dimer. Molecule A is colored in dark blue, whereas molecule B is colored in grey. Each molecule of the dimer consists of two domains, middle domains 1 and 2 (MD1 and MD2). Interestingly, the dimer is highly favored by two α-helices that protrude from each molecule and were shown to be the crystallized TEV cleavage site (TEV helix). A schematic representation of the dimer structure is shown in the upper right corner. **(B)** Structure of one monomer showing MD1 in teal and MD2 in dark blue. The domain architecture of p52_121-EdL is shown in the upper right corner. MD1 and MD2 as well as a part of the TEV site are modelled, whereas the CTD was missing in the p52_121-EdL structure.

The p52_121-EdL structure contains four molecules in the asymmetric unit that form two tight dimers. Each molecule consists of two domains, middle domains 1 and 2 (MD1 and MD2), that dimerized with the opposite domain of the adjacent molecule (Figure III.14 A). The dimerization of p52_121-EdL as seen in the crystal structure revealed a broad interface area of approximately

2,700 Å^2 (according to PDBsum), which was unexpected as p52 was not thought to be a dimer when incorporated into TFIIH. However, p52_121-EdL was shown to be a dimer in solution as depicted by the SEC-MALS analysis (Figure III.12 C). Interestingly, dimer formation is further supported by two α-helices that are not part of p52 but are formed by residues from the N-terminal TEV cleavage site, which was included in the p52 variant for purification purposes and unfortunately not possible to be removed as described above (Figure VI.5). The strong dimerization of p52_121-EdL blocked the access to the TEV cleavage site. The two α-helices that are not part of p52 protrude from each molecule to the other and might promote dimerization. Analysis of the dimer interface revealed that 52 residues of molecule A and 54 residues of molecule B participate in dimer formation. Four salt bridges, eleven hydrogen bonds and 252 van der Waals contacts are formed.

The final model of the best resolved molecule in the asymmetric unit contains 283 out of 394 residues of p52 with residues 165 to 176, 234 to 253, 320 to 349 and 454 to 514 presumably being disordered. The final model further contains twelve additional residues at the N-terminus that belong to the TEV cleavage site. An overview of the residues that could be built for all other molecules in the asymmetric unit can be found in Table VI.2. A linker region reaching from amino acid 322 to 349 connects MD1 comprising amino acids 121 to 321 with MD2 comprising amino acids 350 to 454 (Figure III.14 B). Although the linker region originally reaching from amino acid 322 to 344 was artificially shortened to five amino acids in p52_121-EdL it could not be visualized in the crystal structure due to its flexibility. Surprisingly, the last 61 amino acids of p52_121-EdL that form the CTD of p52, which interacts with p8, were not visible in the crystal structure. To test whether the CTD was degraded during the crystallization process, crystals of p52_121-EdL as well as a protein sample of full-length p52 were sent to the in-house mass spectrometry department for further analysis. A comparison of the peptide intensities of the crystal sample and full-length p52 revealed that the CTD of p52 was still present in the p52_121-EdL crystals. Most likely the CTD was too flexible to be visualized in the crystal structure.

To investigate if the artificial linker had any influence on the orientation of MD1 and MD2 towards each other, the structure was compared with the structure of native p52_121-End. The crystal structure of p52_121-End was solved in space group $P2_1$ utilizing a native data set of p52_121-End, which diffracted to a maximum resolution of 3.0 Å (Table III.3). Phasing was accomplished by molecular replacement applying PHASER and using the structure of p52_121-EdL as a search model. The p52_121-EdL structure contained eight domains per asymmetric unit that formed four molecules in total. The p52_121-EdL and p52_121-End structures can be superimposed with an rmsd of 0.5 Å (Figure III.15). The superposition showed that the overall orientation of the two

domains was not influenced by the artificial linker. The structures highly resembled each other as p52_121-End dimerized as well including the crystallized α-helices of the TEV site. Unfortunately, the p52_121-End structure also lacked additional electron density for the last 60 amino acids of p52.

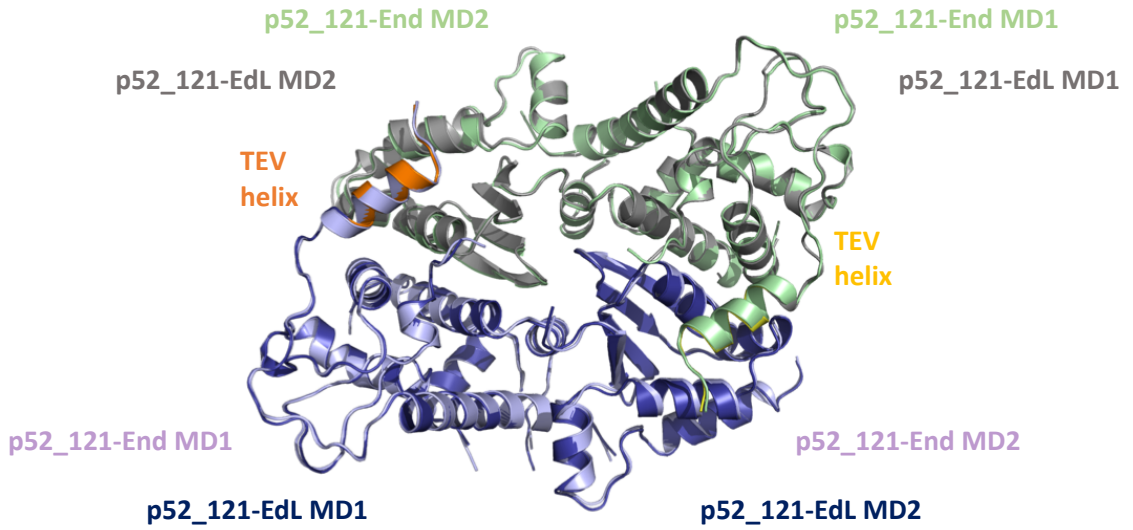


Figure III.15: Superposition of the p52_121-EdL and p52_121-End dimer. The p52_121-EdL and p52_121-End structures were superimposed with an rmsd of 0.5 Å. The structures are highly similar and indicate that the orientation of MD1 towards MD2 is not influenced by the artificial linker.

III.2.3 Structural characterization of p52_1-321

To structurally characterize the N-terminal 120 amino acids, which were not part of the p52_121-EdL variant, shortened p52 variants including this N-terminal region of the protein were investigated. The p52_1-321 variant comprised the N-terminal 321 amino acids of p52 and was terminated prior to the insertion region that was replaced by an artificial linker in the p52_121-EdL variant.

III.2.3.1 Production of p52_1-321

The p52_1-321 variant was inserted into the pETM-11 vector containing an N-terminal His_6-tag and a TEV cleavage site and its expression was pursued in *E. coli* BL21-CodonPlus® (DE3)-RIL cells. The cell pellet was either stored at -80 °C or directly used for subsequent purification utilizing buffer set IV. Purification of p52_1-321 was performed by IMAC using Ni-TED beads followed by SEC using a SD 200 16/60 column (Figure III.16). The final fractions were pooled, concentrated to

5 - 15 mg/ml and either used for subsequent crystallization trials or split into aliquots and stored at -80 °C. During some of the purifications of the p52_1-321 variant an additional His_6-tag cleavage step was performed after the IMAC to cleave the N-terminal His_6-tag of p52_1-321 (Figure VI.5).

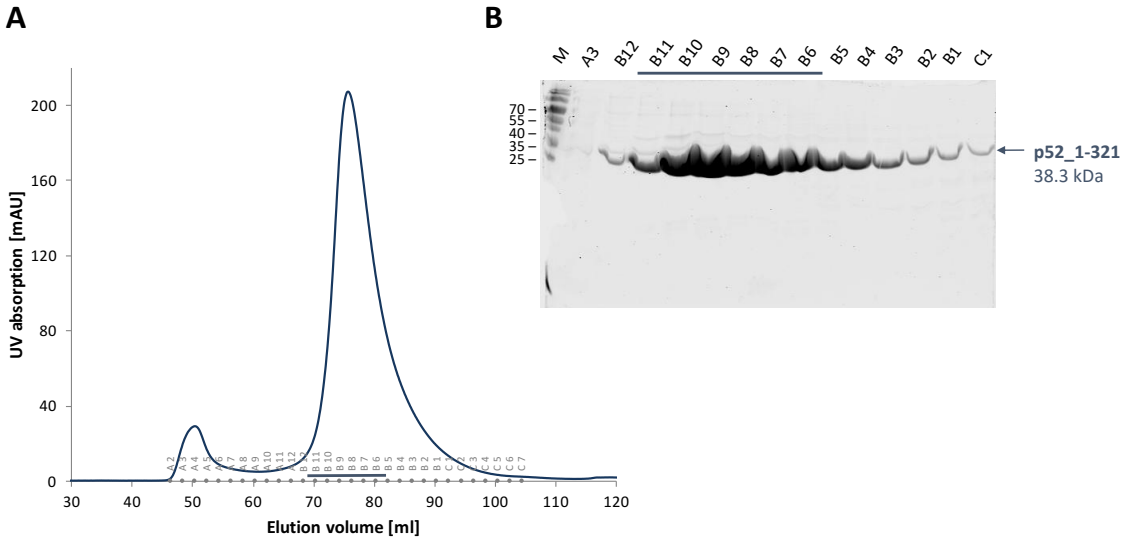


Figure III.16: Size exclusion chromatography of p52_1-321. (A) Chromatogram and **(B)** SDS-PAGE of the size exclusion chromatography (SD 200 16/60) of p52_1-321 (38.3 kDa). The chromatogram shows the UV absorption at 280 nm. The bar indicates fractions that were finally pooled. Abbreviation: M = marker (kDa).

III.2.3.2 Crystallization of p52_1-321

Initial crystallization trials of the p52_1-321 variant were performed by the sitting drop method using the crystallization robot HoneyBee 963 and initial crystals were improved in several 24-well and 96-well fine screens of the most promising crystallization conditions. The p52_1-321 variant crystallized at concentrations of 4 - 7 mg/ml in more than 20 different conditions (Figure III.17). The crystals usually took 1 - 3 days to grow, only a few of them took up to 10 days to grow. Interestingly, although the crystallization conditions were relatively diverse they resembled the ones of p52_121-EdL. They most often contained either PEG 3350, PEG 4000 or PEG 8000, buffers like HEPES pH 7.0 - 7.5 or Tris pH 8.0 - 8.5 and salts like $MgCl_2$ or calcium acetate. All p52_1-321 crystals assumed a hexagonal shape forming either hexagonal needles or hexagonal plates. The needle like crystals grew in conditions containing 100 mM citric acid pH 4.5 - 5.0 and 1.5 - 2 M NaCl.

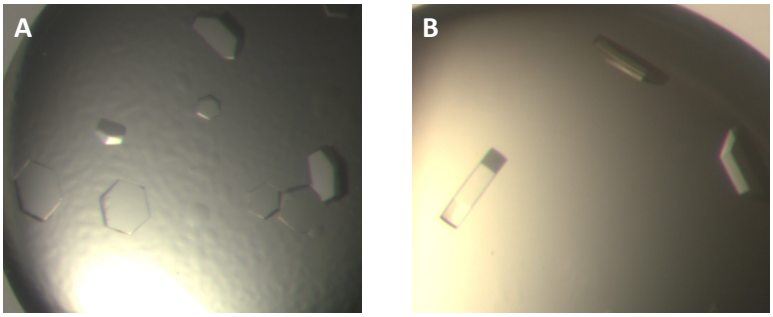


Figure III.17: Protein crystals of p52_1-321. Hexagonal plate like crystals of p52_1-321 obtained in **(A)** a fine screen condition with a precipitant solution containing 80 mM calcium acetate, 100 mM HEPES pH 7.5 and 12.75% PEG 8000 and **(B)** a fine screen condition with a precipitant solution containing 100 mM calcium acetate, 100 mM HEPES pH 7.5 and 7.5% PEG 8000.

The diffraction of the p52_1-321 crystals was in most cases limited to 5 - 4 Å and only a few of them diffracted up to 3.5 - 3.0 Å. Unfortunately, all p52_1-321 crystals were twinned, which hindered structure solution of this p52 variant. An attempt to obtain better diffracting crystals or crystals grown in a different space group through micro-seeding with pre-existing crystals of the p52_1-349 and p52_1-321 variants as described in II.2.5.1 was unsuccessful. In a further attempt the His_6-tag of p52_1-321 was cleaved by the TEV protease during the purification process as depicted above, which as well did not improve crystal quality. In order to solve the phase problem for the p52_1-321 data, the crystals were soaked in heavy-atom containing solutions as depicted in II.2.5.2. Unfortunately, the soaked crystals were either not diffracting any more or the anomalous signal was very weak. The same was observed for most of the crystals of a p52_1-321 Se-Met derivative. However, one p52_1-321 Se-Met derivative crystal that was grown in a precipitant condition containing 100 mM calcium acetate, 100 mM HEPES pH 7.5 and 12.5% PEG 8000 diffracted up to 2.8 Å.

III.2.3.3 Structure of p52_1-321

Twinning of the p52_1-321 crystals severely complicated solving their structure. Processing of the data was difficult as the space group could not be unambiguously determined and different space groups such as $P6_222$, P3 and $P3_121$ were taken under consideration. The best data set of p52_1-321 was obtained of a Se-Met derivative crystal that diffracted up to 2.8 Å. A SAD data set was recorded at a wavelength of 0.97973 Å to enhance the anomalous scattering. The statistics of pointless unambiguously identified $P3_1$ as the correct space group and indicated tetartohedral twinning of the data with four twinning fractions of 53%, 14%, 8% and 25%. Structure solution was attempted using the anomalous scattering of the selenium atoms in the p52_1-321 Se-Met

derivative for phasing by the SAD method but remained unsuccessful due to the tetartohedral twinning.

Table III.4: Data collection and refinement statistics of p52_1-321.

Data collection	**p52_1-321 SAD**
Space Group	P 3_1
Cell Dimensions	
a, b, c [Å]	104.3, 104.3, 165.0
α, β, γ [°]	90.0, 90.0, 120.0
Resolution [Å]	46.97 – 2.80 (2.89 – 2.80)
Wavelength [Å]	0.97973
Unique Reflections	49,518 (4,628)
I/σI	10.7 (1.0)
CC (1/2)	0.997 (0.325)
Rmerge [%]	16.2 (282.9)
Completeness [%]	100.0 (100.0)
Redundancy	10.4 (10.8)
Anomalous Completeness	100.0 (100.0)
Anomalous Multiplicity	5.2 (5.4)
Refinement	
Resolution [Å]	46.97 – 2.80
Unique Reflections	46,873
Number of Atoms	8,528
R_{work} (R_{free}) [%]	17.7 (20.55)
Mean B-Factor [Å^2]	90.695
Bond Lengths [Å]	0.011
Bond Angles [°]	1.618
Ramachandran Statistics* (Favored/Allowed/Outliers)[%]	96.55/2.69/0.77

Values in parentheses refer to the highest resolution shell. *Ramachandran statistics were performed according to the statistics provided by Coot.

However, molecular replacement with PHASER using MD1 of the p52_121-EdL structure (residues 121 - 321) as a search model revealed all four molecules in the asymmetric unit and sufficient additional space to fit an N-terminal domain (NTD) for each molecule into the model. After refinement by REFMAC additional electron density appeared at the N-terminus of p52_121-321, which permitted to manually build some α-helices of the NTD in Coot. Intensive model building

and multiple refinement cycles by REFMAC and Phenix refine made it possible to finally build seven α-helices per molecule. The anomalous density signal of the SAD data permitted the localization of three out of four Se-Met residues per NTD in the asymmetric unit, which served as a guideline for subsequent model building and sequence assignment. The final model of the best resolved molecule in the asymmetric unit contains 280 out of 321 residues with residues 1 to 9, 72 to 78, 166 to 175, 241 to 253 and 320 to 321 presumably being disordered. An overview of the residues that could be built for all other molecules in the asymmetric unit can be found in Table VI.2. The p52_1-321 structure was refined to an R_{work} of 17.7% and an R_{free} of 20.6% by REFMAC using an amplitude-based twin refinement strategy (Table III.4).

The final model of p52_1-321 consists of the NTD and MD1 domains and is shown in Figure III.18 A. The MD1 of p52_1-321 can be superimposed with the MD1 of p52_121-EdL with an rmsd of 0.87 Å (Figure III.18 B). Interestingly, the connection between the NTD and MD1 in p52_1-321 is different than the connection between the TEV helix and MD1 in p52_121-EdL. However, the TEV helix is located at the same position as helix α3 in p52_1-321. Intriguingly, the fold of the four molecules present in the asymmetric unit of p52_1-321 is not completely identical, which might have provoked the tetartohedral twinning and could be favored by the two intricate dimers that are formed between them (Figure III.18 C+D). A PDBsum analysis showed that the dimer interface between molecules A and B covers an area of approximately 1,200 $Å^2$. The dimer interface expands over both domains, the NTD and MD1, and involves 18 residues of molecule A and 20 of molecule B. Altogether, ten charged and 15 hydrophobic residues are located in the interface and one salt bridge, six hydrogen bonds and 96 non-bonded contacts are formed.

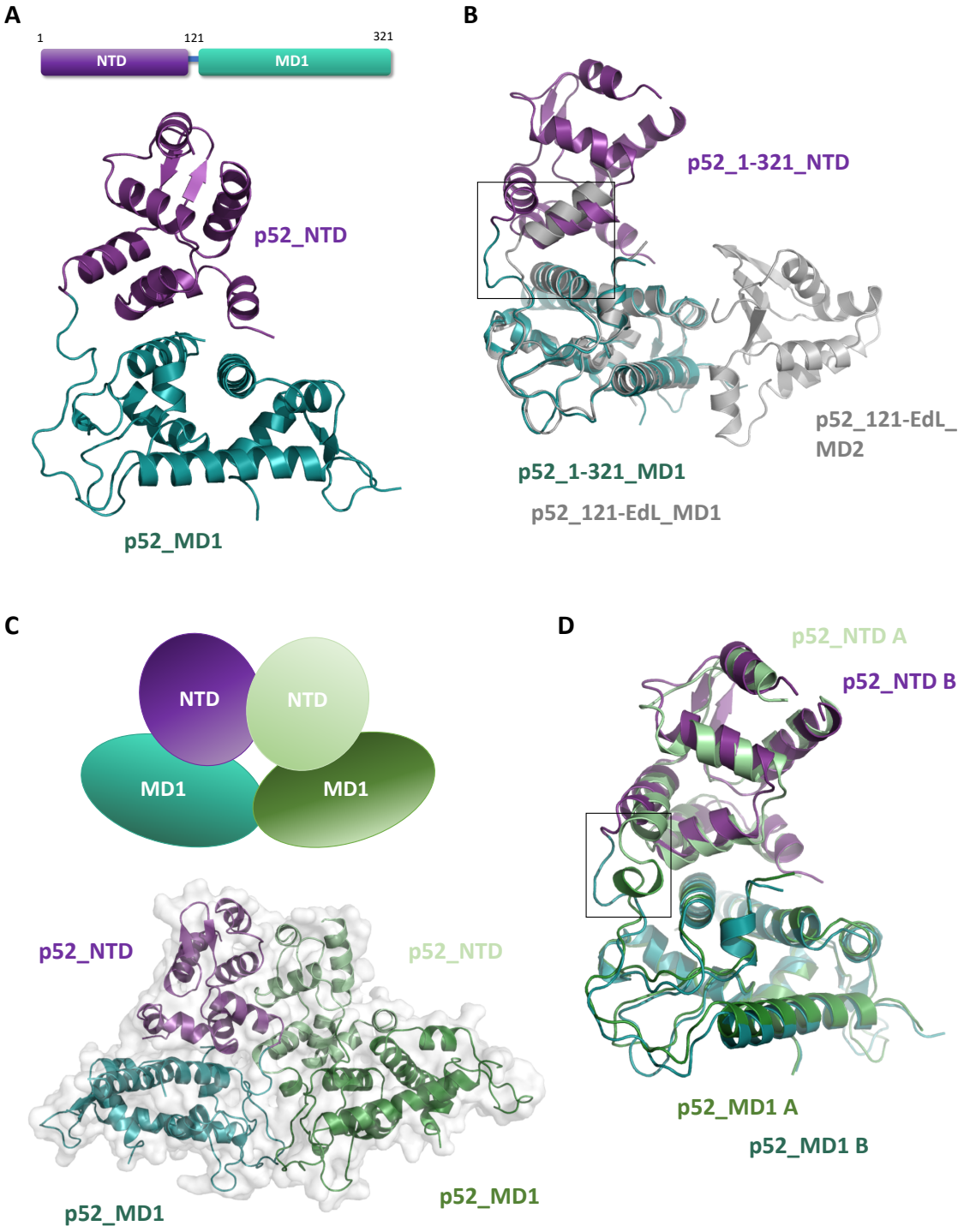


Figure III.18: Crystal structure of p52_1-321. (A) Structure of p52_1-321 consisting of the NTD (purple) and MD1 (teal) domains. The domain architecture of p52_1-321 is indicated above the structure. **(B)** The MD1 domains of p52_1-321 and p52_121-EdL can be superimposed with an rmsd of 0.87 Å. The connection between the NTD and MD1 in p52_1-321 is different than the connection between the TEV helix and MD1 in p52_121-EdL as indicated by the frame. **(C)** Structure of a p52_1-321 dimer showing the NTD and MD1 domains of molecule A in light green and dark green and of molecule B in purple and teal, respectively. A schematic representation of the dimer structure is shown above. **(D)** A superposition of molecules A and B of p52_1-321 shows that their fold is slightly different with the main difference being located in the linker region between the NTD and MD1 domains (marked by a frame).

III.2.4 The interaction between p52 and p8

The interaction between the p52_CTD domain and p8 was revealed by a crystal structure of the p52_CTD /p8 complex from yeast (126). However, this interaction was only investigated with the isolated protein domains and the arrangement of this interaction in the global environment of p52 was unknown at the beginning of the herein described work. Thus, the interaction between p52 and p8 was analyzed on the basis of the crystal structure of p52_121-EdL.

III.2.4.1 The p52_121-EdL/p8 complex

Apart from purifications of the full-length p52/p8 complex several minimal complexes between p8 and shortened p52 variants were analyzed. These included co-purifications of the p52_121-EdL variant with p8 in different variations altering, for example, the combination of un-tagged and tagged protein samples. However, these purifications resulted either in only one of the two components or in a non-homogenous sample that formed oligomers.

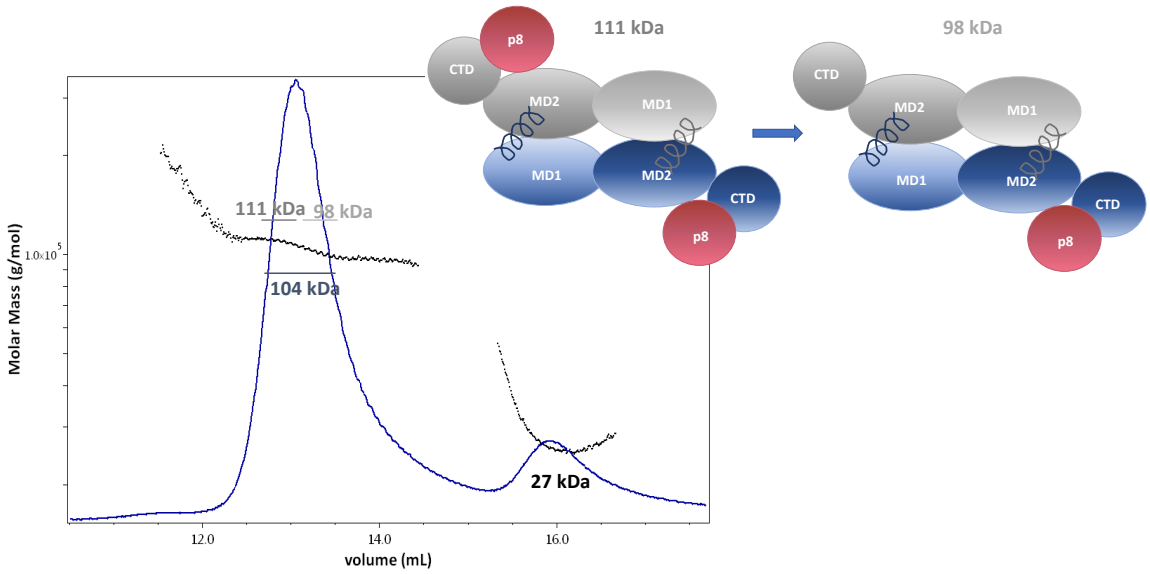


Figure III.19: SEC-MALS analysis of the p52_121-EdL dimer in complex with p8. The SEC-MALS analysis (SD 200 16/60) indicated an average mass of 104.2 ± 0.71 kDa for the first peak that can be divided into two halves of distinct masses. The first half of the peak most likely represents a p52_121-EdL dimer (86.3 kDa) with two p8 (11.4 kDa per p8) proteins bound to it (110.9 ± 0.84 kDa) and the second half of the sample indicates a p52_121-EdL dimer that has only one p8 protein bound (97.86 ± 0.87 kDa). This is depicted graphically in the upper right corner. The unbound p8 is represented in the SEC-MALS analysis by the smaller peak with an average mass of 27.09 ± 0.38 kDa, corresponding to dimerized p8 (22.8 kDa).

The ability of p52_121-EdL to bind p8 was verified in a SEC-MALS experiment. The SEC-MALS analysis of single p52_121-EdL (44.4 kDa) showed that it forms a dimer in solution with an average

mass of 86 kDa (see Figure III.12 C). Incubation of p52_121-EdL with p8 in a 1:1 ratio for 1 h at 4 °C prior to the SEC-MALS experiment leads to a shift of the peak to earlier elution volumes indicating successful complex formation. The SEC-MALS analysis indicated an average mass of 104.2 ± 0.71 kDa for the p52_121-EdL/p8 sample distributed over the entire peak. However, the peak showed a prominent decrease of the mass, splitting the peak into two areas with different masses. Analysis of these two halves of the peak resulted in an average mass of 110.9 ± 0.84 kDa for the first half and 97.86 ± 0.87 kDa for the second half after the decrease in mass. A comparison with the mass of a previous experiment, in which p52_121-EdL (86.3 kDa) was analyzed, indicated that the first half of the sample contained a p52_121-EdL dimer with two p8 (11.4 kDa per p8) molecules bound. In the second half of the peak presumably one of the p8 molecules was lost indicated by a decrease in mass of around 13 kDa, thus containing the p52_121-EdL dimer with one p8 molecule. The unbound p8 was giving rise to the smaller peak with an average mass of 27.09 ± 0.38 kDa indicating dimerized p8. The SEC-MALS analysis thus revealed that p8 can bind to the p52_121-EdL dimer and that the two proteins constitute an optimal substrate for crystallization attempts.

III.2.4.2 Crystallization of p52_121-EdL/p8

In order to obtain crystals of the p52_121-EdL/p8 complex two different strategies were pursued. The p8 subunit is a very small protein of 11.4 kDa and the crystal structure of p52_121-EdL showed additional space to fit p8 into the same crystal packing (diameter of 30 - 40 Å). Thus, pre-existing p52_121-EdL crystals were soaked with a 10-fold excess of p8 for different incubation times before they were flash frozen in liquid nitrogen. Only crystals, which were soaked for 60 sec showed a diffraction pattern, but unfortunately no incorporation of p8. Soaking for 17 - 30 min resulted in crystals that did not diffract and after 30 min of soaking, the crystal began to crack. In a second strategy, p52_121-EdL was co-crystallized with p8 in a 1:2 ratio. The proteins were incubated for 1 h at 4 °C prior to setting up the crystallization trials utilizing the sitting drop method with the crystallization robot HoneyBee 963. Co-crystallization of p52_121-EdL and p8 resulted in plate like crystals that grew after 1 - 4 days and resembled those of single p52_121-EdL crystals (Figure III.20). The crystals grew in several different crystallization conditions and most of them were similar to the conditions of the p52_121-EdL crystals. The best crystals diffracted up to 2.7 - 3.0 Å. In order to unambiguously identify the location of the p52_CTD (see III.2.4.3 for more details) co-crystallization of a Se-Met derivative of p52_121-EdL, which was originally purified for the structure solution of this protein variant in the absence of p8, was

performed together with native p8. The p52_121-EdL Se-Met/p8 crystals had a similar shape and diffraction quality as their native equivalents.

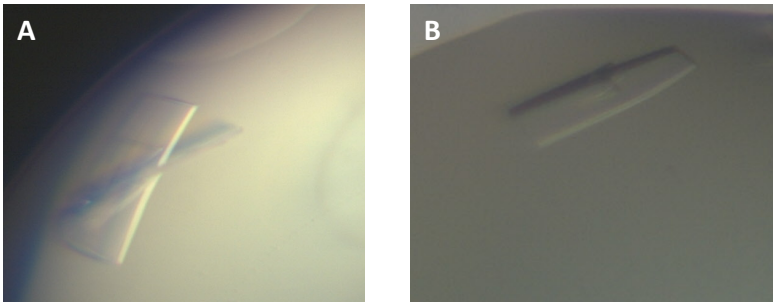


Figure III.20: Protein crystals of co-crystallized p52_121-EdL/p8 complexes. Crystals obtained of p52_121-EdL co-crystallized with p8 in a 1:2 ratio in **(A)** an initial 96-well screen condition with a precipitant solution containing 2% 1,4-dioxane, 100 mM sodium citrate pH 5.5 and 15% PEG 1000 and **(B)** a 96-well fine screen condition with a precipitant solution containing 150 mM ammonium chloride, 100 mM Tris pH 8.5 and 15% PEG 4000.

III.2.4.3 Structure of the p52_121-EdL/p8 complex

The structure of the p52_121-EdL/p8 complex was solved in space group $P2_1$ by molecular replacement applying PHASER and a data set of the co-crystallized p52_121-EdL and p8 proteins. Molecular replacement was performed utilizing the MD1 and MD2 domains of p52 (p52_121-EdL structure) as a search model. Structure solution revealed that one p52_121-EdL dimer was present in the asymmetric unit. The additional electron density, which was present at the C-terminal end of one of the monomers, was used to build the structure of the p52_CTD/p8 complex from yeast (PDB code: 3DOM) into the electron density. The sequence was adapted to the *C. thermophilum* homolog and the structure was initially refined with REFMAC. Unfortunately, it was difficult to locate any residues in the p52_CTD and p8 domains due to the flexibility of this part of the structure and the rather weak overall data quality. In addition, the overall fold of the p52_CTD and p8 is nearly the same and it was therefore not possible to unambiguously define, which part of the electron density belonged to p52_CTD and which to p8. Two possibilities had to be investigated in more detail (Figure III.21 A+B). Interestingly, the other monomer of p52 did not seem to carry p8 as no additional electron density could be obtained for neither p52_CTD nor p8.

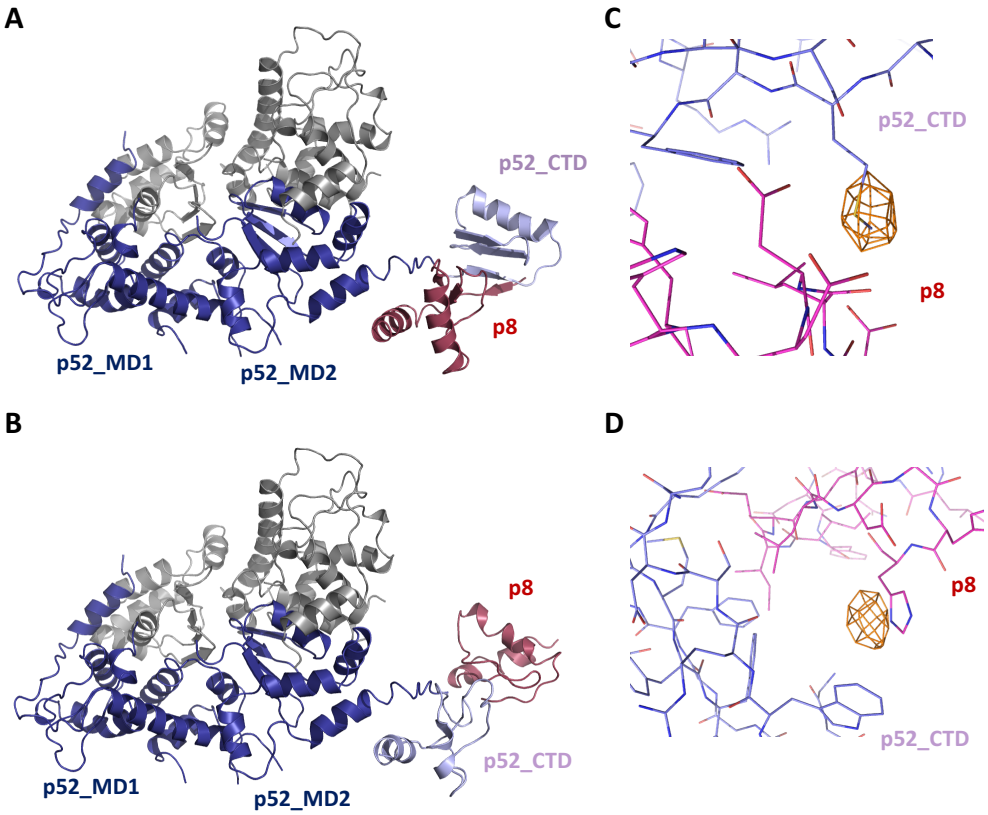


Figure III.21: Initial structure solutions of the p52_121-EdL/p8 complex. (A) Structure solution for possibility A, in which p8 would be bound beneath p52_MD2 and p52_CTD. **(B)** Structure solution for possibility B, in which p8 would be bound above p52_CTD. The anomalous density map of **(C)** for the first possibility, i.e. A, clearly indicates the presence of a Se-Met residue at the corresponding electron density peak, **(D)** whereas possibility B does not offer a fitting solution. The contour level of the anomalous density map was set to 4.0 σ.

As the p52_CTD from *C. thermophilum* contains three methionines a Se-Met derivative of p52_121-EdL was co-crystallized with native p8 to obtain an anomalous density signal for p52_CTD and to define its position. The structure was solved by molecular replacement applying PHASER EP and using the p52_121-EdL structure as a search model (Table III.5).

An anomalous density signal was obtained for one out of the three Se-Met residues within p52_CTD. Both possibilities were tested but only in one of them the anomalous density signal could be assigned to a Se-Met residue in p52_CTD (Figure III.21 C+D). Thus, the p52_CTD and p8 were positioned correctly into the native structure of the p52_121-EdL/p8 complex and the structure was refined to a resolution of 2.7 Å. During refinement the map was improved by the STARANISO server due to the anisotropy of the data and the model was refined to an R_{work} of 19.3% and R_{free} of 23.7% using BUSTER (Table III.5).

Table III.5: Data collection and refinement statistics of p52_121-EdL/p8.

Data collection	**p52_121-EdL/p8**	**p52_121-EdL/p8 SAD**
Space Group	P 2_1	P 2_1
Cell Dimensions		
a, b, c [Å]	74.2, 85.9, 92.1	74.2, 85.5, 92.0
α, β, γ [°]	90.0, 94.9, 90.0	90.0, 95.2, 90.0
Resolution [Å]	49.28 – 2.68 (2.81 – 2.68)	49.24 – 3.00 (3.18 – 3.00)
Wavelength [Å]	0.96770	0.96770
Unique Reflections	31,937 (3,898)	22,807 (3,528)
I/σI	2.8 (0.2)	7.2 (1.4)
CC (1/2)	0.982 (0.087)	0.994 (0.750)
Rmerge [%]	38.3 (476.6)	24.4 (142.1)
Completeness [%]	98.1 (91.4)	98.8 (95.2)
Redundancy	4.4 (3.3)	7.1 (6.3)
Anomalous Completeness	-	98.1 (93.4)
Anomalous Multiplicity	-	3.6 (3.2)
Refinement		
Resolution [Å]	19.92 – 2.68	
Unique Reflections	17,099	
Number of Atoms	5,621	
R_{work} (R_{free}) [%]	19.3 (23.7)	
Mean B-Factor [Å^2]	51.46	
Bond Lengths [Å]	0.010	
Bond Angles [°]	1.200	
Ramachandran Statistics* (Favored/Allowed/Outliers) [%]	92.30/4.94/2.76	

Values in parentheses refer to the highest resolution shell. *Ramachandran statistics were performed according to the statistics provided by Coot.

The structure of p52_121-EdL/p8 depicts the previously characterized interface of p52/p8 within the environment of the two middle domains of p52 (Figure III.22). The final model contains 334 out of 394 residues of p52_121-EdL with residues 166 to 175, 195 to 199, 236 to 252, 322 to 349 being presumably disordered as well as 16 additional residues of the TEV cleavage site of p52 and residues 4 to 69 of p8. Interestingly, the p52_CTD is relatively distant from p52_MD2 and almost forms a distinct entity together with p8. This might explain, why the p52_CTD was too flexible to be visualized without p8. The interaction between p52 and p8 is mediated by β-strands 12 to 15 of p52 and β-strands 1 to 3 of p8 covering an interface area of approximately 1050 Å^2 (PDBsum).

Analysis of the interface shows that 16 residues of p52_CTD and 19 residues of p8 participate in the interaction. The rather electrostatic interface is formed by five salt bridge, ten hydrogen bonds and 111 non-bonded contacts. The p52_CTD/p8 structure can be superimposed with the p52_CTD/p8 structure from yeast (PDB code: 3DOM) with an rmsd of 1.10 Å for p52_CTD and an rmsd of 0.97 Å for p8.

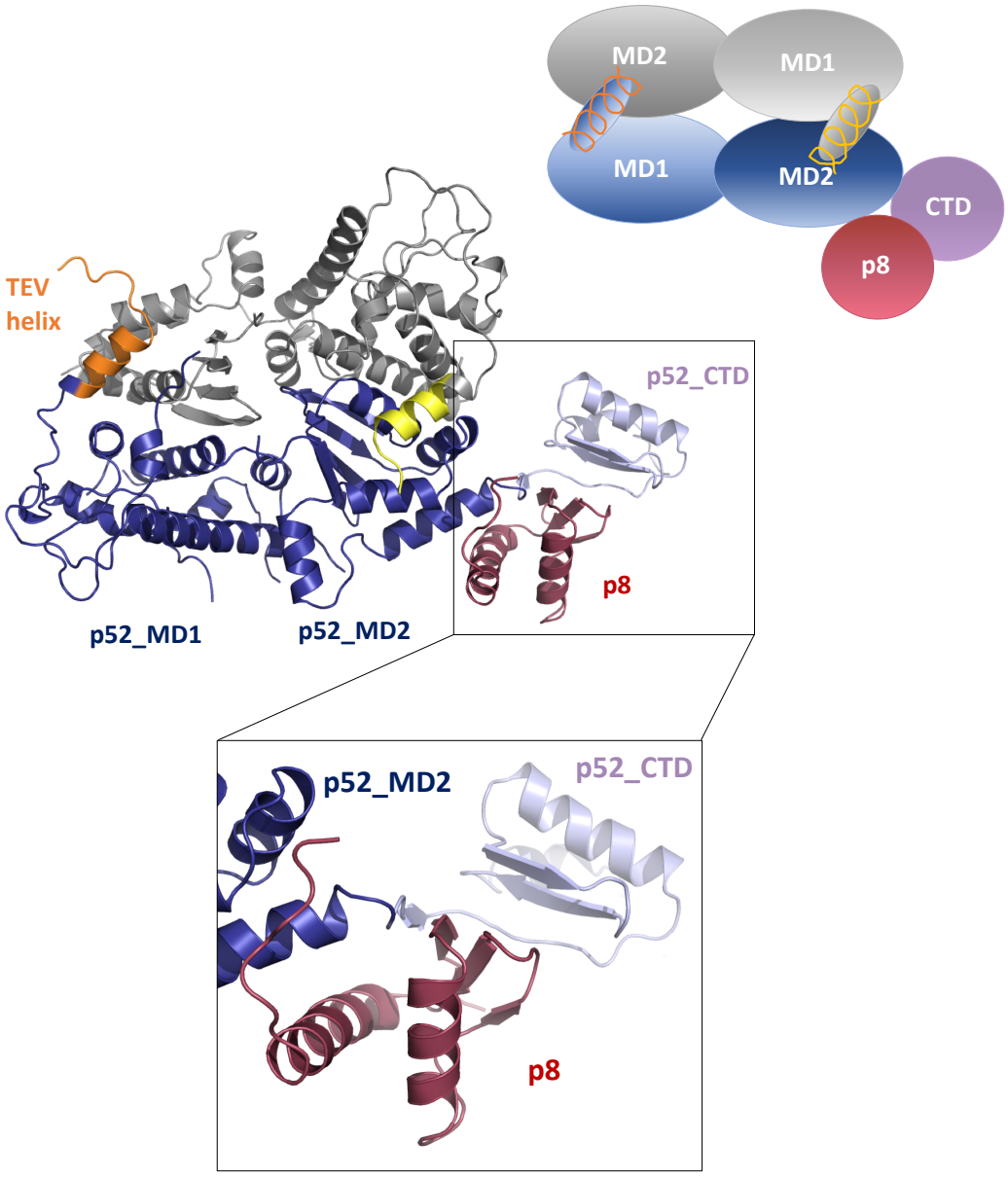


Figure III.22: Crystal structure of the p52_121-EdL/p8 complex. Structure of the p52_121-EdL/p8 complex showing the two molecules of the dimer in grey and dark blue and the corresponding TEV helices in yellow orange, respectively. The p52_CTD (light purple) and p8 (red) can only be observed in one of the two monomers. A schematic representation of the structure is shown in the upper right corner and a closer view of the interaction between p52 and p8 is shown in the frame.

III.2.5 Model of full-length p52

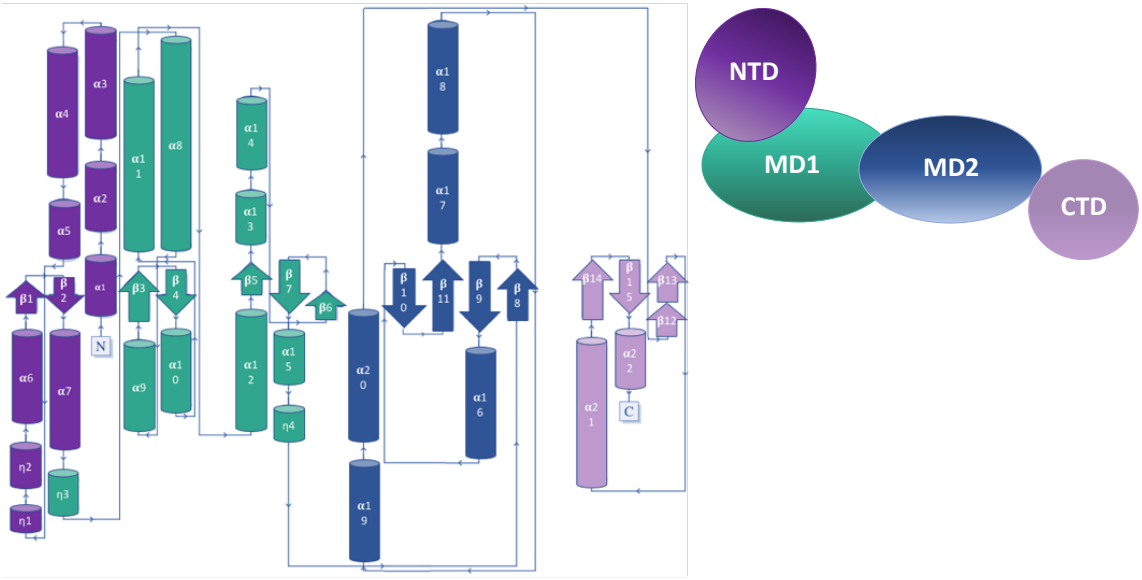
The p52_1-321 and p52_121-EdL/p8 crystal structures could be combined to obtain a structural model of full-length p52. The MD1 domains of p52_1-321 and p52_121-EdL/p8 were superimposed to generate the final model of full-length p52 (Figure III.23). Full-length p52 consists of four domains, the NTD, two middle domains (MD1 and MD2) and the CTD, which binds p8. Figure III.23 A depicts a 2D topology of the p52 full-length model, Figure III.23 B shows the p52 full-length model in different orientations and Figure III.23 C shows a sequence alignment of p52 depicting the secondary structure elements of the p52 full-length model. The p52 full-length model shows that the NTD, MD1 and MD2 domains form a rather compact entity, whereas the CTD is located at a larger distance to the other domains and is interconnected and stabilized through p8. The structure of full-length p52 is predominantly formed by α-helices and consists of 22 α-helices and 15 β-strands in total.

The NTD reaches from amino acid 1 to 120 and consists of seven α-helices and a β-hairpin. MD1 is the largest domain reaching from amino acid 121 to 319. It consists of several α-helices and β-strands including two long helices α8 and α11, which consist of 20 and 17 amino acids, respectively. A 28 amino acids long linker region, which is presumably highly disordered, bridges MD1 with MD2. MD2 reaches from amino acid 250 to 454 and contains five α-helices and an antiparallel β-sheet consisting of β-strands 8 to 11. The CTD reaches from amino acid 455 to 514 and consists of two α-helices and four β-strands, which are mediating the interaction to p8.

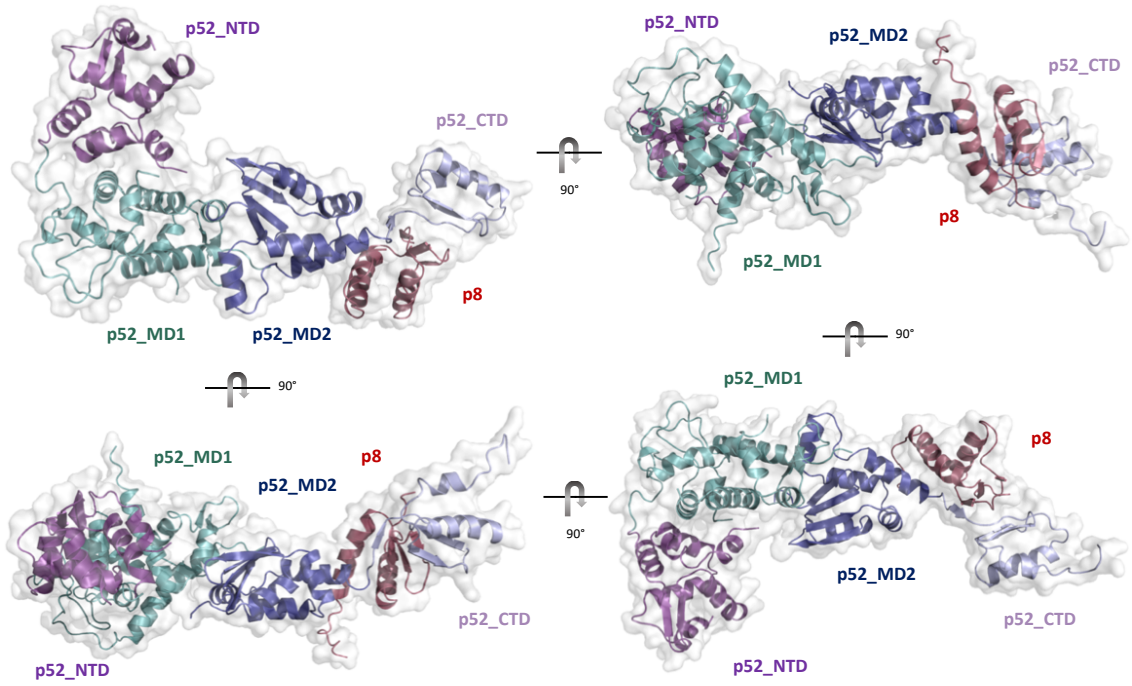
The p52 full-length model was subjected to a DALI search against the PDB-90 database to compare its fold with homologous structures deposited in the PDB, which resulted in 14 hits. However, except for the closest homolog, the yeast TFIIH cryo-EM structure (PDB code: 5OQJ), all other matches covered, if at all, only individual domains of p52. Thus, the three novel identified domains of p52 (NTD, MD1 and MD2) were subjected to individual DALI searches against the PDB-25 database and their results were analyzed.

Table III.6 shows the top three homologs of the individual p52 domains ranked according to their Z-values. The homologs found for the NTD and MD1 domains showed reasonable matches, whereas the homologs found for MD2 showed comparably low matches. Out of the best ten homologs found for the NTD nine homologs were DNA binding proteins or associated with DNA binding processes including three homologs of the MarR family of transcription factors and out of the best ten homologs found for MD1 seven homologs were DNA binding proteins or associated with DNA binding processes.

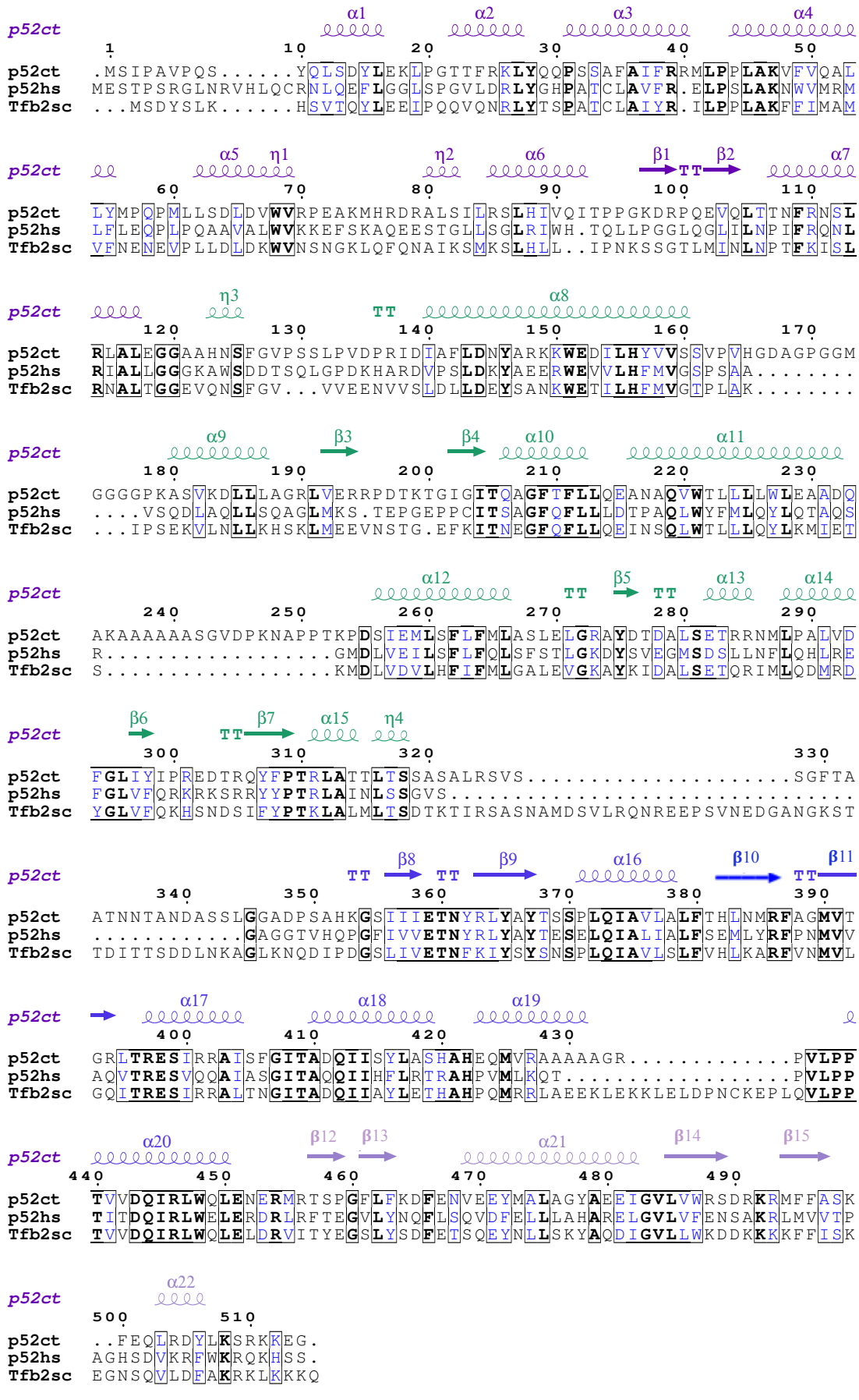
A



B



C


p52ct
α1 α2 α3 α4
1 10 20 30 40 50
p52ct .MSIPAVPQS.....YQLSDYLEKLPGTTFRKLYQQPSSAFAIFRRMLPPLAKVFVQAL
p52hs MESTPSRGLNRVHLQCRNLQEFLGGLSPGVLDRLYGHPATCLAVFR.ELPSLAKNWVMRM
Tfb2sc ...MSDYSLK.....HSVTQYLEEIPQQVQNRLYTSPATCLAIYR.ILPPLAKFFIMAM
p52ct α5 η1 η2 α6 β1 TT β2 α7
60 70 80 90 100 110
p52ct LYMPQPMLLSDLDVWVRPEAKMHRDRALSILRSLHIVQITPPGKDRPQEVQLTTNFRNSL
p52hs LFLEQPLPQAAVALWVKKEFSKAQEESTGLLSGLRIWH.TQLLPGGLQGLILNPIFRQNL
Tfb2sc VFNENEVPLLDLDKWVNSNGKLQFQNAIKSMKSLHLL..IPNKSSGTLMINLNPTFKISL
p52ct η3 TT α8
120 130 140 150 160 170
p52ct RLALEGGAAHNSFGVPSSLPVDPRIDIAFLDNYARKKWEDILHYVVSSVPVHGDAGPGGM
p52hs RIALLGGGKAWSDDTSQLGPDKHARDVPSLDKYAEERWEVVLHFMVGSPSAA.......
Tfb2sc RNALTGGEVQNSFGV...VVEENVVSLDLLDEYSANKWETILHFMVGTPLAK.......
p52ct α9 β3 β4 α10 α11
180 190 200 210 220 230
p52ct GGGGPKASVKDLLLAGRLVERRPDTKTGIGITQAGFTFLLQEANAQVWTLLLLWLEAADQ
p52hs ....VSQDLAQLLSQAGLMKS.TEPGEPPCITSAGFQFLLLDTPAQLWYFMLQYLQTAQS
Tfb2sc ...IPSEKVLNLLKHSKLMEEVNSTG.EFKITNEGFQFLLQEINSQLWTLLLQYLKMIET
p52ct α12 TT β5 TT α13 α14
240 250 260 270 280 290
p52ct AKAAAAAASGVDPKNAPPTKPDSIEMLSFLFMLASLELGRAYDTDALSETRRNMLPALVD
p52hs R.....................GMDLVEILSFLFQLSFSTLGKDYSVEGMSDSLLNFLQHLRE
Tfb2sc S.....................KMDLVDVLHFIFMLGALEVGKAYKIDALSETQRIMLQDMRD
p52ct β6 TT β7 α15 η4
300 310 320 330
p52ct FGLIYIPREDTRQYFPTRLATTLTSSASALRSVS.....SGFTA
p52hs FGLVFQRKRKSRRYYPTRLAINLSSGVS.....
Tfb2sc YGLVFQKHSNDSIFYPTKLALMLTSDTKTIRSASNAMDSVLRQNREEPSVNEDGANGKST
p52ct TT β8 TT β9 α16 β10 TT β11
340 350 360 370 380 390
p52ct ATNNTANDASSLGGADPSAHKGSIIIETNYRLYAYTSSPLQIAVLALFTHLNMRFAGMVT
p52hs .....GAGGTVHQPGFIVVETNYRLYAYTESELQIALIALFSEMLYRFPNMVV
Tfb2sc TDITTSDDLNKAGLKNQDIPDGSLIVETNFKIYSYSNSPLQIAVLSLFVHLKARFVNMVL
p52ct α17 α18 α19
400 410 420 430
p52ct GRLTRESIRRAISFGITADQIISYLASHAHEQMVRAAAAAGR.....PVLPP
p52hs AQVTRESVQQAIASGITAQQIIHFLRTRAHPVMLKQT.....PVLPP
Tfb2sc GQITRESIRRALTNGITADQIIAYLETHAHPQMRRLAEEKLEKKLELDPNCKEPLQVLPP
p52ct α20 β12 β13 α21 β14 β15
440 450 460 470 480 490
p52ct TVVDQIRLWQLENERMRTSPGFLFKDFENVEEYMALAGYAEEIGVLVWRSDRKRMFFASK
p52hs TITDQIRLWELERDRLRFTEGVLYNQFLSQVDFELLLAHARELGVLVFENSAKRLMVVTP
Tfb2sc TVVDQIRLWQLELDRVITYEGSLYSDFETSQEYNLLSKYAQDIGVLLWKDDKKKKFFISK
p52ct α22
500 510
p52ct ..FEQLRDYLKSRKKEG.
p52hs AGHSDVKRFWKRQKHSS.
Tfb2sc EGNSQVLDFAKRKLKKKQ

◀ **Figure III.23: Model of full-length p52. (A)** The MD1 domains of the p52_1-321 and p52_121-EdL/p8 structures were superimposed and a model of full-length p52 was generated. Schematic representation of the p52 full-length 2D topology and a schematic representation of the structure. **(B)** Different orientations of the full-length p52 model with the NTD in purple, MD1 in teal, MD2 in dark blue and the CTD in light purple. P8 (red) is bound to p52. **(C)** Sequence alignment of the *C. thermophilum* (ct), human (hs) and *S. cerevisiae* (sc; TfB2) sequences of p52 performed by the ClustalW2 multiple sequence alignment tool. Conserved residues are highlighted in black boxes with the most conserved residues depicted in bold. Secondary structure elements of the p52 full-length model are indicated by arrows, spirals, η or TT for β-strands, α-helices, 3_{10} helices and β-turns, respectively (generated by ESPript).

Table III.6: Structural homologs of the DALI search against individual domains of p52.

Domain - rank of the hit out of all found homologs*	PDB code (chain)	Z-value	Rmsd	Identical (%)	Protein class**
NTD - 1/252	5OQJ (2)	9.8	2.3	24	Transcription
NTD - 2/252	4ASN (A)	6.6	2.6	10	Transcription
NTD - 3/252	1W5S (B)	6.6	2.6	8	Replication
MD1 - 1/220	5OQJ (2)	16.4	2.4	18	Transcription
MD1 - 2/220	4M9S (B)	6.0	3.0	13	Apoptosis
MD1 - 3/220	5JUY (B)	5.6	3.3	14	Apoptosis
MD2 - 1/75	5JLT (A)	3.4	3.2	17	Viral protein/DNA
MD2 - 2/75	5N7L (A)	3.3	2.0	9	Protein transport
MD2 - 3/75	5WJ7 (B)	3.3	4.0	10	Actin/DNA binding protein

*ranking was performed according to the Z-value **according to the classification of the Protein data bank

In summary, it can be noted that the full-length p52 structure seems to represent a novel composition of differently folded domains that partly resemble the fold of other proteins involved in DNA associated processes.

III.2.6 The interaction between XPB and p52

The interaction between XPB and p52 had already been pursued in a few studies prior to our analysis but could not be determined accurately. Two XPB binding regions had been proposed within p52 and the N-terminal part of XPB was found to be likewise important for the interaction with p52 (105, 123). The precise structural knowledge of p52, as depicted above, constituted an ideal starting point to investigate the interaction between XPB and p52 further.

III.2.6.1 Production of XPB

Eukaryotic XPB is composed of several domains, an N-terminal extension domain (NTD), a damage-recognition domain (DRD), two RecA-like helicase domains (HD1 and HD2) and a C-terminal extension domain (CTD) (Figure I.4). The region of XPB that was identified to be sufficient for an interaction with p52 in the human proteins (amino acids 44 - 208) (123) corresponds approximately to amino acids 81 to 253 of XPB from *C. thermophilum* as defined by the sequence alignment tool Clustal_Omega.

Since previous studies showed that the N-terminus of XPB was sufficient to form an interaction with p52 and due to the fact that the expression of full-length XPB from *C. thermophilum* was difficult in *E. coli* cells, interaction studies between XPB and p52 concentrated on N-terminal variants of XPB. XPB variants that comprised the NTD only proved to be not soluble in *E. coli* cells, thus an XPB variant that additionally comprised the DRD (approximately amino acids 290 - 348 in XPBct), XPB_60-345, was used for the interaction analyses with p52. A Table summarizing the generation of XPB and p52 variants used for the XPB/p52 project can be found in the appendix in VI.2. All XPB and p52 variants were expressed in *E. coli* BL21-CodonPlus® (DE3)-RIL cells and subsequently purified via IMAC (Ni-TED column) followed by SEC (SD 200 16/60 column). Purifications including full-length p52 were pursued using buffer set III, whereas all other proteins or protein combinations were purified using buffer set IV. Due to the fragile behavior of the XPB variants, expressions of single XPB variants and co-expressions of XPB and p52 variants were immediately followed by protein purification after cell harvest. A representative purification of XPB_60-345 is shown in Figure VI.6.

III.2.6.2 Interaction between XPB and p52

The interaction between XPB_60-345 and the p52/p8 complex was analyzed in detail. Analytical SEC analyses were performed using a SD 200 Increase 10/300 column and A-SEC buffer C. The XPB_60-345 sample and the p52/p8 sample were applied individually and in a 1:1 complex onto the same column. The peak shifted to earlier elution volumes clearly indicating complex formation between the two samples, which was as well visualized via SDS-PAGE (Figure III.24 A). Due to its rather small size p8 could hardly be visualized on the SDS-PAGE gel. The analytical SEC data confirmed that the N-terminus of XPB is indeed mediating the interaction to p52. Furthermore, the optimized N-terminal XPB_60-345 variant, which is lacking the first 60 amino acids in comparison to the XPB_1-345 construct, seemed to be sufficient to form a stable complex with p52/p8.

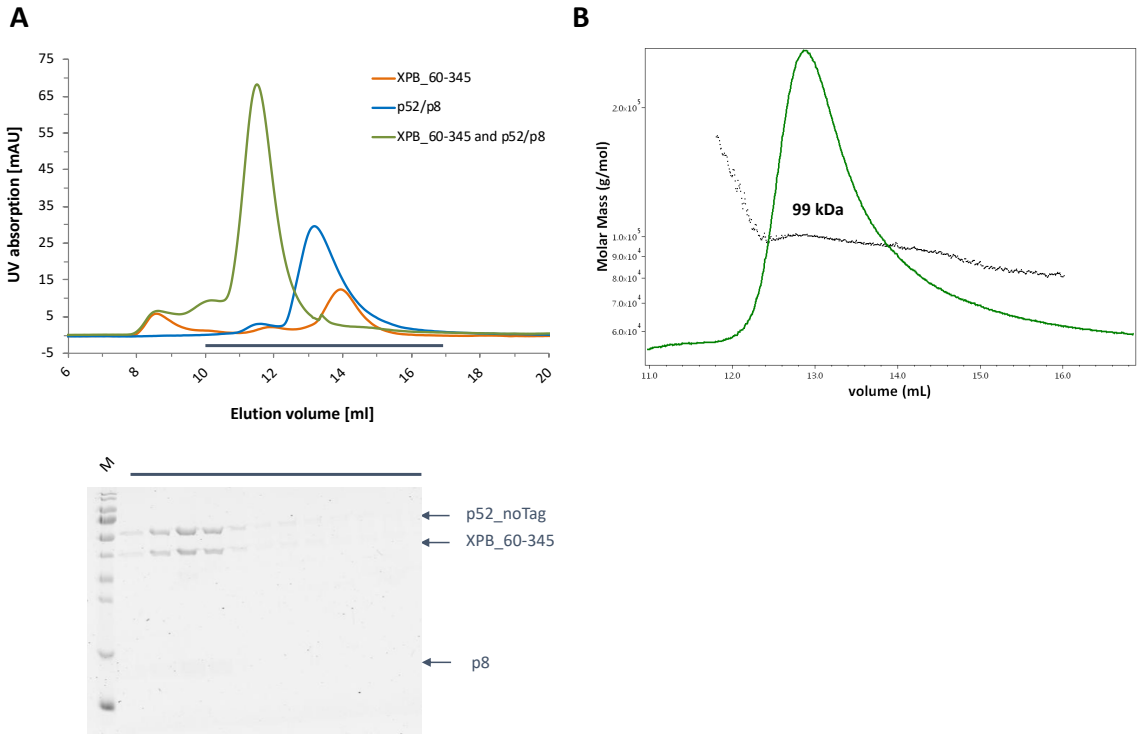


Figure III.24: A-SEC and SEC-MALS analysis of the complex between XPB_60-345 and p52/p8. (A) The interaction between XPB_60-345 and p52_noTag/p8 was analyzed on a SD 200 Increase 10/300 column and verified by SDS-PAGE analysis. The SDS-PAGE gel of the combined run is shown below. **(B)** SEC-MALS analysis (SD 200 16/60) of the XPB_60-345/p52_noTag/p8 ternary complex indicates a mass of 98.80 ± 0.99 kDa that corresponds to an equal monomeric distribution of the proteins in the ternary complex.

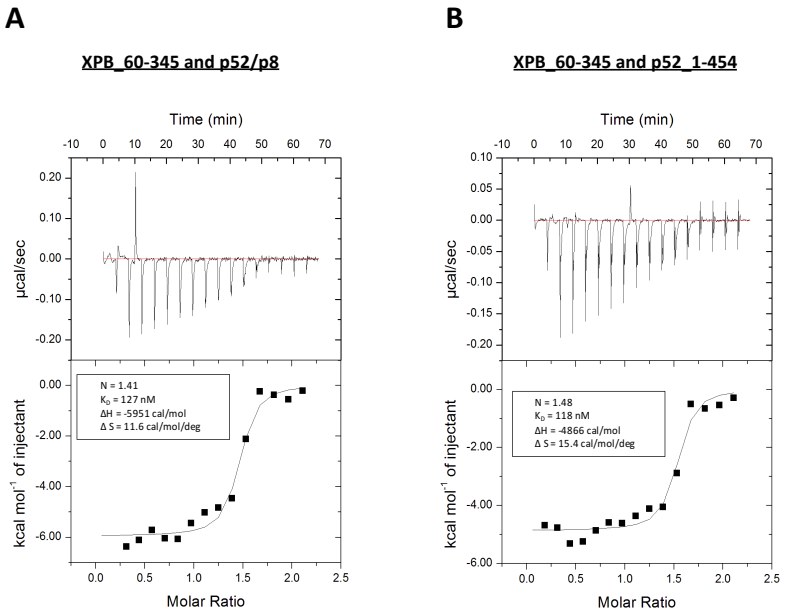


Figure III.25: ITC analysis of XPB_60-345 and p52/p8 or p52_1-454. ITC analyses were conducted using the ITC_{200} instrument and ITC buffer B. **(A)** 250 µM of XPB_60-345 were titrated into 25 µM of p52_noTag/p8 in 16 injections at 25 °C and showed an interaction with a K_D of 127 nM. **(B)** 250 µM of XPB_60-345 were titrated into 25 µM of p52_1-454 in 16 injections at 25 °C and showed an interaction with a similar K_D of 118 nM.

The XPB_60-345/p52/p8 ternary complex was further analyzed by SEC-MALS, which showed a nearly homogenous peak with an average mass of 98.80 ± 0.99 kDa that corresponds to a 1:1:1 ratio in the ternary complex (Figure III.24 B). Analysis of the mass at the beginning of the peak (101.6 ± 1.118 kDa) and its end (96.64 ± 0.934 kDa) revealed that the difference in mass is rather minor and most likely does not correspond to a loss of the p8 protein.

Titration of XPB_60-345 to p52/p8 in an ITC analysis revealed a K_D of 127 nM (Figure III.25 A). The stoichiometry of N = 1.4 was surprising as the ternary complex was shown to be of equal monomeric distribution of the proteins during SEC-MALS but might indicate that a fraction of the XPB variant formed dimers. Interestingly, when XPB_60-345 was titrated to a p52 variant that was lacking the CTD (p52_1-454) the binding was as efficient as observed for p52/p8 resulting in a K_D of 118 nM (Figure III.25 B). It could thus be concluded that the CTD of p52 as well as p8 are not required for the interaction to the N-terminus of XPB.

III.2.6.3 Interaction between XPB_NTD and p52_MD2

The N-terminal part of XPB was found to be sufficient to bind p52 as illustrated above. Two regions in p52 had been proposed to be important for the binding to XPB and correspond to the NTD and MD2 domains (123). However, only MD2 was found to be required to stimulate the ATPase activity of XPB (105). To accurately determine the domains of p52 that are important for this interaction various p52 variants were analyzed against XPB_1-345 and XPB_60-345 via native PAGE. Figure III.26 depicts native PAGE gel A, which comprises the results of XPB_1-345 and XPB_60-345, and native PAGE gel B, which shows the results of further p52 variants and XPB_60-345. The same p52 variants as used in native PAGE B were also tested against XPB_1-345 leading to the same results for all analyzed p52 variants. In comparison to the XPB_60-345 variant, the XPB_1-345 variant comprised additionally the N-terminal 59 amino acids of XPB.

The p52/p8 complex showed successful complex formation with the N-terminal XPB variants as indicated by a shift of both bands to a unique height in the native PAGE. This further verified the results obtained by ITC and A-SEC. However, the p52_1-321 variant, which comprises the NTD and MD1 domains and thus the first proposed XPB binding site, did not show an interaction with XPB. In contrast, p52_121-EdL, which comprised the MD1, MD2 and CTD domains and thus the second proposed binding site, showed an interaction with XPB. This result suggests that the second proposed XPB binding site but not the first proposed XPB binding site is sufficient for the interaction to XPB_60-345.

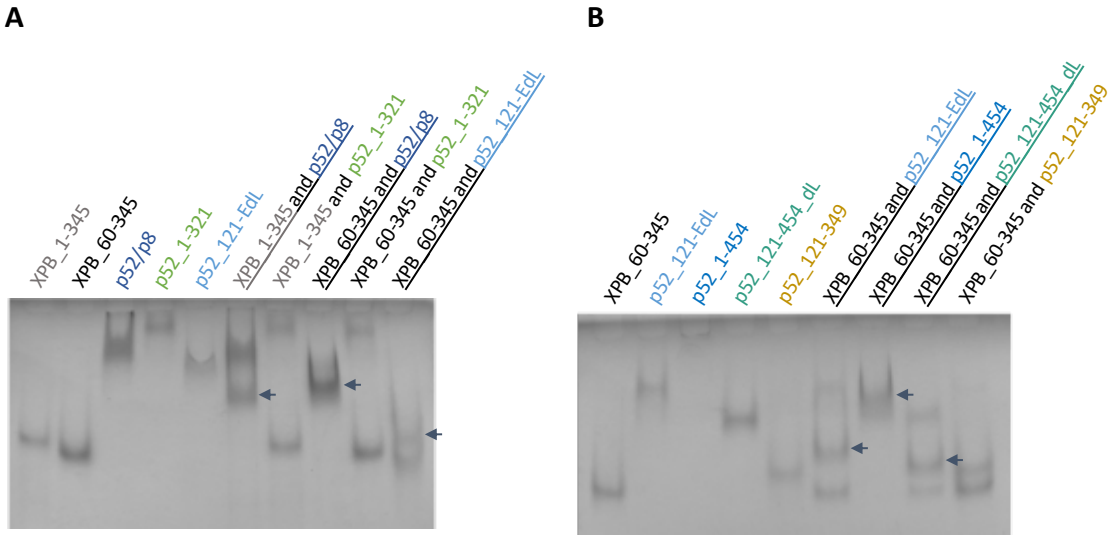


Figure III.26: Native PAGE analyses of XPB and p52 variants. (A) The interaction between XPB_1-345 or XPB_60-345 was analyzed by native PAGE. Bands at unique heights indicate complex formation of XPB_1-345 or XPB_60-345 and p52/p8 as well as of XPB_60-345 and p52_121-EdL. No interaction could be observed for XPB_1-345 or XPB_60-345 and p52_1-321 (NTD and MD1). **(B)** Native PAGE analysis of further p52 variants and XPB_60-345 showed an interaction for all three p52 variants containing MD2, but not for p52_121-349 (MD1). Bands that indicate complex formation are marked by an arrow and the respective constructs are underlined.

A p52 variant, which comprised the NTD, MD1 and MD2 (p52_1-454) did show an interaction indicated by a shifted band compared to the sample of single p52_1-454, which was stacked in the pocket of the gel. To further verify if MD2 was sufficient for binding XPB two p52 variants were tested that comprised either MD1 and MD2 (p52_121-454_dL) or only MD1 (p52_121-349). MD1 alone did not interact with XPB_60-345, but the additional presence of MD2 led to an interaction with XPB. The results clearly showed that MD2 of p52 is responsible for the interaction with XPB_60-345 and is sufficient to accomplish this interaction. An additional interaction between the NTD of p52 and XPB could not be excluded but p52_NTD was not sufficient to form a complex with XPB_60-345 in a native gel and most likely does not participate in the interaction between the two proteins at all. This hypothesis was supported by co-expression attempts of XPB_60-345 in complex with a p52_NTD variant (p52_1-120_noTag) that led to the expression of XPB only.

The determination of MD2 as the domain, which is decisive for the interaction with the N-terminal part of XPB, was the basis to generate several minimal complexes of XPB and p52 variants. Thus, co-expression and -purification of an XPB/p52 minimal complex that would be suitable for crystallization was pursued. All co-expression attempts were conducted in *E. coli* BL21-CodonPlus® (DE3)-RIL cells as these were shown to yield the highest expression of XPB_60-345 and most of the single p52 variants. For co-expression attempts XPB variants were inserted into the pBADM-11 vector and p52 variants into the pETM-11 vector.

Table III.7: Summary of XPB/p52 complexes that were tested for expression, purification and crystallization.

XPB variant	p52 variant	Co - expression	Co - purification	Initial crystallization screens
60-345	full-length_noTag	yes	homogenous complex	Index, JCSG+, Crystal Screen™ 1 + 2, Nextal - PEG Suite, Nextal pH Clear, Nucleix Suite, Protein Complex Suite, Wizard 1 + 2 and Wizard 3 + 4
60-345	full-length_noTag/p8	purified separately	homogenous complex	Index, JCSG+, Crystal Screen™ 1 + 2, Nextal - PEG Suite, Nextal pH Clear, Nucleix Suite, Protein Complex Suite, Wizard 1 + 2 and Wizard 3 + 4
60-345	121-EdL	no	–	–
60-345	345-End/p8*	no	–	–
60-345	350-End*	no	–	–
60-345	1-454	yes	non-homogenous complex	JCSG+, Crystal Screen™ 1 + 2, Nextal pH Clear, Nucleix Suite, Protein Complex Suite, Wizard 1 + 2 and Wizard 3 + 4
60-345	121-454_dL	yes	non-homog. complex	–
60-345	121-454_dL_noTag**	yes	homogenous complex	Index, JCSG+, Crystal Screen™ 1 + 2, Nextal - PEG Suite, Nextal pH Clear, Nucleix Suite, Protein Complex Suite, Wizard 1 + 2 and Wizard 3 + 4
60-345	345-458*	yes	homogenous complex	Index, JCSG+, Crystal Screen™ 1 + 2, Protein Complex Suite and Wizard 1 + 2
60-345	345-458_noTag**	no	–	–
60-345	345-421*	no	–	–
60-345	350-435*	no	–	–
60-345	350-435_noTag**	yes	weak complex formation	–
1-345	345-458*	yes	weak complex formation	–
116-345*	345-458*	no	–	–
116-245**	345-458*	no	–	–
60-249*	345-458*	no	–	–
60-270*	345-458*	no	–	–
60-249*	121-454_dL	no	–	–
60-270*	121-454_dL	no	–	–
60-249*	full-length_noTag	yes	homogenous complex	Index, JCSG+ and Crystal Screen™ 1 + 2
60-270*	full-length_noTag	yes	homogenous complex	Index, JCSG+ and Crystal Screen™ 1 + 2

*protein could not be expressed as single entity in a soluble form **expression as single entity has not been tested

Table III.7 summarizes all XPB/p52 minimal complexes that have been pursued. Unfortunately, only a few of these combinations could be co-expressed in a soluble form, which allowed the subsequent co-purification of the complex.

A co-expression and subsequent co-purification of XPB_60-345 and p52_noTag yielded a stable complex of stoichiometric amounts. Unfortunately, initial crystallization trials of this complex did not yield any crystals. Similarly, initial co-crystallization trials of separately purified XPB_60-345 and p52/p8 did not yield any crystals either. The non-tagged variant of p52_121-454_dL, which comprised both middle domains of p52 and the artificial linker that was also introduced in the p52_121-EdL variant, formed a homogenous complex with XPB_60-345 during a co-purification attempt (Figure VI.9 A+B). Although the complex appeared to be very homogenous initial crystallization trials did not yield any crystals.

The smallest minimal complex entity was obtained through co-expression and -purification of XPB_60-345 and the p52_345-458 variant, which comprises only MD2. Purification of this XPB/p52 minimal complex (Figure VI.9 C+D) yielded very low amounts of protein that impeded crystallization. In order to improve the yield and quality of this XPB/p52 minimal complex several optimization strategies were pursued. A thermofluor analysis of XPB_60-345/p52_345-458 did not indicate different buffer conditions than the ones already used to improve the purification. Furthermore, p52 variants with slightly different domain boundaries, but comprising only MD2, i.e. p52_345-421 and p52_350-435(_noTag), were tested. However, none of these led to an improvement of the XPB/p52 minimal complex. Likewise, several XPB variants such as XPB_1-345, XPB_116-345 and XPB_116-245 did not lead to minimal complex improvement. The construct design of XPB was further complicated due to the lack of structural knowledge of the N-terminal part of XPB. In addition, limited proteolysis (see II.2.3.3) of the XPB/p52 minimal complex was performed to identify constructs that would be more suitable for crystallization (Figure VI.10). Mass spectrometry analysis of a degradation product with a molecular weight of 25 kDa on the SDS-PAGE indicated degradation of XPB_60-345 solely from the C-terminus. The combination of the results obtained by limited proteolysis with those of secondary structure predication tools led to the design of two new XPB variants, XPB_60-249 and XPB_60-270. In comparison to the XPB variants reaching until amino acid 345 the two new XPB variants comprised only the NTD and lacked the DRD. However, single expressions of these XPB variants as well as co-expressions with p52_345-458 or p52_121-454_dL led to insoluble protein production. Interestingly, co-purifications with p52_noTag were possible for both new XPB variants, but initial crystallization trials were unsuccessful. Based on these results it could be reasoned that the NTD of XPB is

sufficient for complex formation with p52, but the additional presence of the DRD most likely favors soluble expression.

In summary, the interaction between XPB and p52 that could be characterized in this work involves the NTD of XPB and MD2 of p52. These two domains are probably the main actors for the XPB/p52 interaction and are sufficient for stable complex formation. Several XPB/p52 minimal complexes could be produced, but unfortunately did not yield any crystals.

III.2.6.4 Towards the interface characterization of XPB/p52

As attempts to structurally characterize the interaction between XPB and p52 were not successful alternative approaches were pursued in parallel. In a first approach, a peptide array of XPB against p52/p8 was developed and performed by our collaboration partner Hans Maric (at the time of the experiments present at the University of Kopenhagen, currently at the University of Würzburg). The peptide array was intended to identify very short fragments within XPB that are mediating the interaction towards p52 and could then be used for co-crystallization studies with p52 variants. The peptide array was composed of 15 amino acid long peptides of the N-terminal region of XPB (XPB_1-345). The peptides were designed as such that they were partially overlapping with the sixth amino acid of the previous peptide being the first amino acid of the following peptide. The peptide array was incubated with p52/p8 and interactions between XPB peptides and p52/p8 were detected by an anti-His peroxidase antibody. The signal strength was normalized against a His_9 peptide. Two peptides, XPB-18 and XPB-67, with significant albeit not extremely high intensities were identified (Figure VI.7). The peptide intensities surrounding these two hits were slightly enriched as compared to other peptide intensities, which underpinned the results. The XPB-18 peptide was located in the NTD (amino acids 86 - 100), whereas XPB-67 was located in the DRD region of XPB (amino acids 331 - 345).

Table III.8: Summary of ITC results of p52/p8 and XPB-18 or XPB-67.

Concentration p52/p8	**Concentration XPB-18**	**Concentration XPB-67**	**Temperature**	**Interaction**
25 μM	250 μM		25 °C	no
25 μM	250 μM		37 °C	no
25 μM	500 μM		37 °C	no
15 μM		100 μM	37 °C	no
22 μM		230 μM	37 °C	no
22 μM		230 μM	25 °C	no
25 μM		465 μM	37 °C	no

To verify the results, the XPB-18 and XPB-67 peptides were commercially synthesized and their interaction with p52/p8 was analyzed in ITC experiments. Due to the presence of two tyrosine residues in each of the peptides their concentrations could be determined spectrophotometrically. Several ITC analyses were performed at different concentrations and temperatures to verify the results (Table III.8). However, no interaction could be observed when XPB-18 or XPB-67 was titrated to p52/p8. Since the initially identified interactions could not be verified by ITC the XPB-18 and XPB-67 peptides were not suitable for crystallization trials with p52 variants. The suitability of the peptide array to dissect the XPB/p52 interaction remains thus questionable.

The F99S mutation found in the *XPB* gene of XP/CS patients was shown to weaken the interaction between XPB and p52 (105). However, no peptide containing the F99 homolog F143 of *C. thermophilum* was identified in the peptide array. The peptides comprised a length of 15 amino acids, which might have been too short to form secondary structure elements like the β-strands that underlie the p52_MD2/XPB_NTD interaction (see III.2.7).

In a further approach, a collaboration with the Borchers/Sickmann group of the ISAS (Leibniz institute for analytical science) in Dortmund was initiated to perform crosslinking studies combined with mass spectrometry analysis. Samples of XPB_60-345, p52/p8 and the complex of all three of them were analyzed by our collaboration partners. The ternary complex was generated by a separate purification of p8 and the XPB_60-345/p52_noTag complex. The purified proteins were combined in a 1:1 ratio and applied to a final A-SEC to generate the ternary complex (Figure VI.8). The sample obtained after the A-SEC was used for the crosslinking studies. Bis(sulfosuccinimidyl)suberate (BS3) was used to crosslink lysine residues and successfully crosslinked peptides were analyzed by mass spectrometry.

Table III.9: Summary of inter-crosslinks between p52 and p8 or XPB.

Protein	Residue	Protein	Residue
p52	S2	p8	K51
p52	S2	p8	K53
p52	S2	XPB	K73
p52	S2	XPB	K83
p52	K183	XPB	K83
p52	K508	XPB	K83

Four crosslinked peptide pairs between XPB and p52, two peptide pairs between p52 and p8 (Table III.9) as well as several intra-crosslinks in p52 and XPB were identified. The crosslinks were mediated between several lysine residues as well as serine residue S2 of p52. Figure III.27 visualizes the position of the crosslinked residues in the full-length p52/p8 structure model. S2 of p52 is located in the NTD, K183 in MD1 and K508 in the CTD. The K26 and K499 residues mediated only intra-crosslinks. Unfortunately, no crosslink between the MD2 domain of p52 and XPB could be identified. This might be explained by the fact that only one lysine residue is located in p52_MD2. Interestingly, all p52 and XPB residues that performed inter-crosslinks with each other also performed intra-crosslinks.

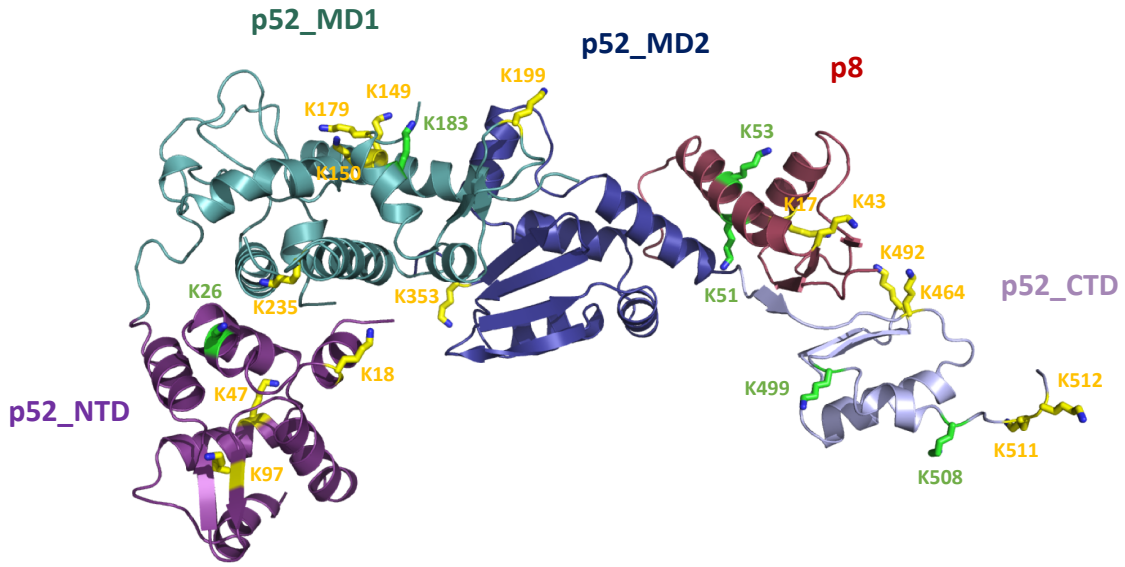


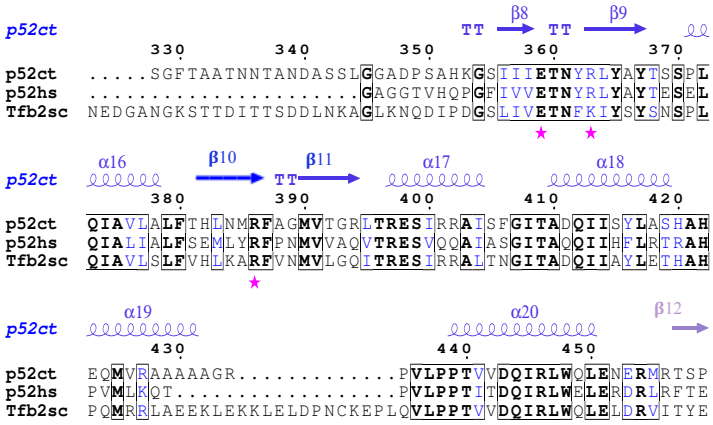
Figure III.27: Results from crosslinking studies visualized in the p52/p8 structure model. Lysine residues are shown as sticks in the full-length p52/p8 model. All lysine residues that did not crosslink are shown in yellow, whereas lysine residues that crosslinked are shown in green.

III.2.6.5 Mutational analysis of the XPB/p52 interface

Two studies investigated p52 mutations that disrupted the interaction with XPB. In a first study, the single mutation p52_E310K as wells as the double mutation p52_E310K/R314E were shown to impair the binding to XPB and to reduce its ATPase activity in the human proteins (131). The double mutant further led to reduced NER and transcriptional activities. In a further study, the human p52_ R337L and p52_L440P mutations had been proposed to inhibit the interaction to XPB due to their resistance against triptolide (132).

Interestingly, L440 (M494 in *C. thermophilum*) is located in the CTD and in the interface to p8. Thus, the L440P mutation was not investigated further. A sequence alignment of human, *C. thermophilum* and *S. cerevisiae* p52 showed that the other three residues E310, R314 and R337 are conserved (Figure III.28 A). In the structural model of full-length p52, the E310, R314 and R337 homologs of *C. thermophilum* (E359, R363 and R386, respectively) are all located in the antiparallel β-sheet of MD2 (Figure III.28 B). Moreover, the entire MD2 domain is highly conserved (see Figure III.28 A), which might reflect its importance and the location of the three residues suggests that the interaction towards XPB is mediated by the antiparallel β-sheet of MD2.

A



B

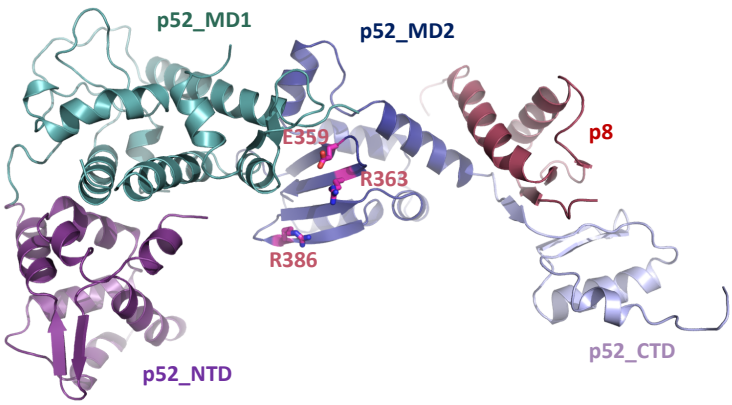


Figure III.28: Localization of p52 residues that have been identified in previous studies to be critical for the interaction with XPB. (A) Sequence alignment of the *C. thermophilum* (ct), human (hs) and *S. cerevisiae* (sc; TfB2) sequences of p52_MD2. The sequence alignment was performed by the ClustalW2 multiple sequence alignment tool. Secondary structure elements of MD2 are indicated by arrows, spirals or TT for β-strands, α-helices and β-turns, respectively (generated by ESPript). Conserved residues are highlighted in black boxes with the most conserved residues depicted in bold. The three p52ct residues E359, R363 and R386 are marked by a star. The nomenclature is according to the sequence of p52ct. **(B)** The three p52ct residues E359, R363 and R386 are shown as sticks in the full-length p52/p8 model and were found to be all located in the antiparallel β-sheet of MD2.

The three mutations E359K, R363E and R386L were introduced into p52_noTag, co-expressed with p8 and purified via IMAC (Ni-TED column) and SEC (SD 200 16/60 column) according to the protocol of the wild-type p52/p8 complex (the purifications are depicted in Figure VI.11, Figure VI.12, Figure VI.13). It should be noted that the p52_E359K/p8 complex behaved differently during the purification procedure as the peak of the p52_E359K/p8 complex eluted at higher elution volumes compared to the p52_R363E/p8 complex, which was eluting from the same column (compare Figure VI.13 and Figure VI.12). In addition, the SDS-PAGE of p52_E359K/p8 indicated at least a partial loss of p8. The interaction of the three p52 variants with XPB_60-345 was analyzed by native PAGE (Figure III.29).

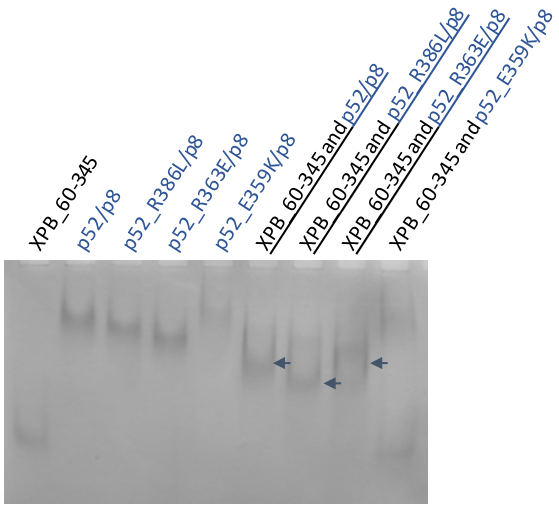


Figure III.29: Native PAGE analysis of potential XPB/p52 interface variants. The interaction between XPB_60-345 and p52 interface variants was analyzed by native PAGE. Bands at unique heights indicate complex formation of XPB_60-345 and wild-type p52/p8, p52_R386L/p8 and p52_R363E/p8. No interaction could be observed for XPB_60-345 and p52_E359K/p8 indicating that the p52_E359 residue is crucial for the interaction with XPB. Bands that indicate complex formation are marked by an arrow and the respective constructs are underlined.

The p52_R386L and p52_R363E variants did not disrupt the interaction towards XPB. The E359K variant, however, had a strong effect on the XPB/p52 interface and was shown to completely disrupt the interaction to XPB. This variant depicts a promising first candidate to investigate the XPB/p52 interface further and to perform functional studies on XPB's ATPase activity.

III.2.7 The XPB/p52/p8 network within TFIIH

As already introduced in chapter III.1.5 two recent cryo-EM structures of the human and yeast TFIIH complexes at resolutions of 4.4 Å and 4.7 Å, respectively, permitted the analysis of our structural findings of the *C. thermophilum* proteins within TFIIH. In both cryo-EM structures,

secondary structures elements of p52 could be built as poly-alanine stretches that roughly positioned the individual domains. However, both complexes display many gaps in the p52 structure and its sequence could not be adequately assigned except for the CTD. In the yeast TFIIH structure, the sequence was assigned to some additional parts of p52, but these parts were nevertheless only built as poly-alanine chains. Thus, both EM structures lacked a coherent model of the entire p52 structure.

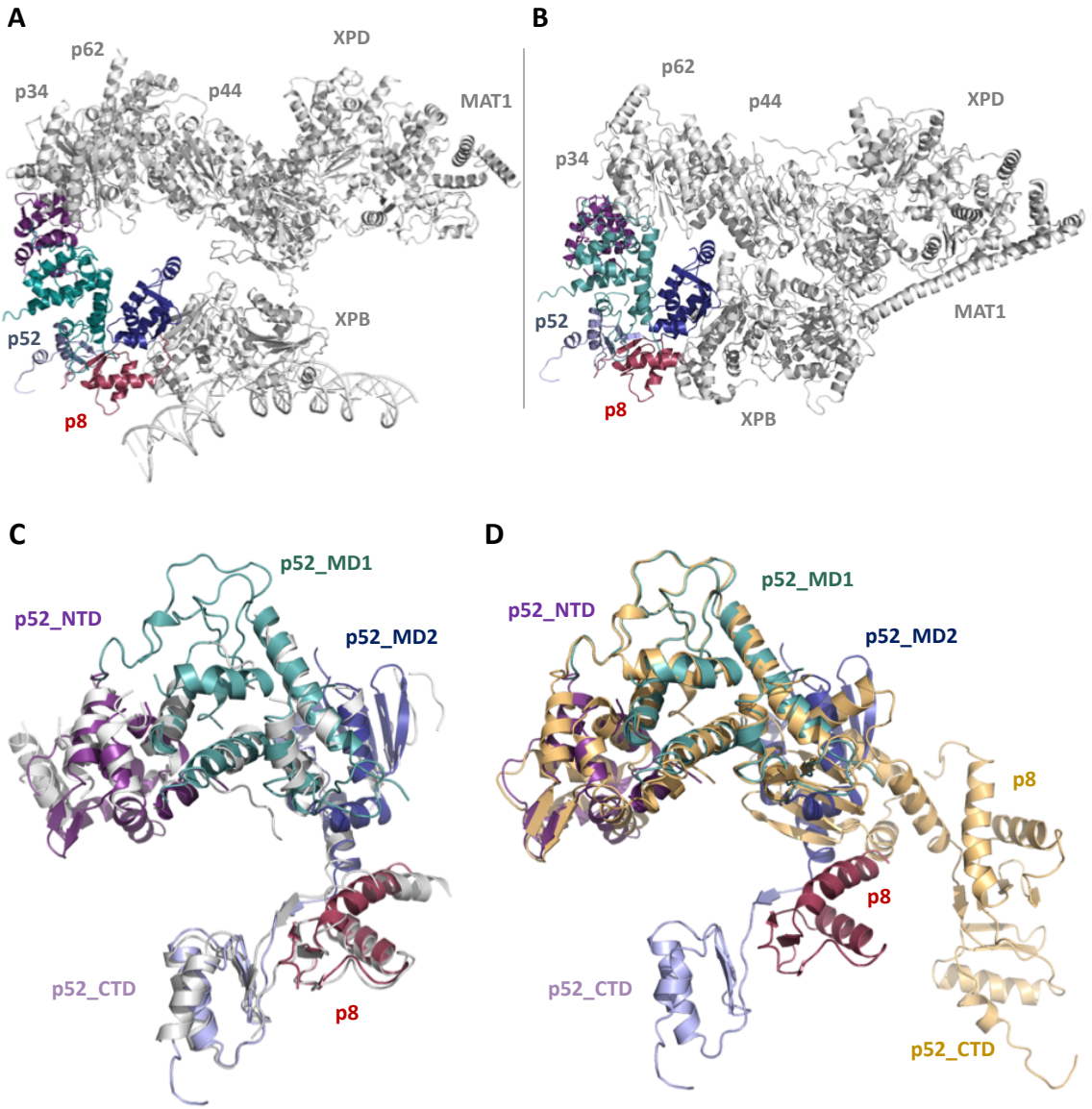


Figure III.30: Comparison of the p52/p8 structure from *C. thermophilum* without and within TFIIH. (A) Individual domains of the p52/p8 crystal structures from *C. thermophilum* (colored) built into the TFIIH cryo-EM structure from *S. cerevisiae* (grey). The TFIIH structure is shown as a cutout of the cryo-EM structure of the PIC-TFIIH complex from *S. cerevisiae* (PDB code: 5OQJ). The DNA duplex is bound by XPB. **(B)** Individual domains of the p52/p8 crystal structures from *C. thermophilum* (colored) built into the human TFIIH cryo-EM structure (PDB code: 5OF4). **(C)** Superposition of the p52 domains and p8 from *C. thermophilum* modelled into the human TFIIH cryo-EM structure (colored) and the corresponding parts as shown in the human TFIIH cryo-EM structure (grey). **(D)** Superposition of the p52 domains and p8 from *C. thermophilum* modelled into the human TFIIH cryo-EM structure (colored) and the p52/p8 crystal structure model from *C. thermophilum* without TFIIH (light orange).

The structure of p52/p8 from *C. thermophilum* as obtained from the individual crystal structures was modelled into both cryo-EM maps (TFIIH/p52ct/p8ct structure model) (Figure III.30 A+B). Interestingly, MD1 and MD2 had to be treated as individual entities to fit them into the EM maps and could not be built as continuous domains. The individual domains, however, conclusively fitted into the electron densities. Overall, it was possible to insert the p52 crystal structure from *C. thermophilum* into the poly-alanine stretches that could be built in the human and yeast EM structures and thereby several new insights were obtained that could not have been analyzed in either of the EM structures due to their only very rough secondary structures elements that had not been connected with each other (Figure III.30 C). Interestingly, a comparison of the domain arrangement of p52 of *C. thermophilum* as modelled into TFIIH with the one observed in the individual crystal structures clearly indicated differences in the overall arrangement. When the NTD and MD1 domains of p52 as built into TFIIH were superimposed with those of the isolated p52 structure, the orientation of MD2 varied significantly and hence the location of p52_CTD/p8 in relation to the rest of p52 was quite different (Figure III.30 D).

The more detailed analysis of the interaction between p52 and XPB within TFIIH focused on the human model (Figure III.31 A). The individual domains of p52 from *C. thermophilum* accurately fitted into the EM densities and not only revealed the localization of the individual domains within TFIIH but also uncovered interactions of p52 that could not have been analyzed so far. Within TFIIH, the p52_MD2 domain interacts with another domain that was not assigned to any particular subunit in the human cryo-EM structure but assumed to depict the NTD of XPB and was built as incoherent poly-alanine stretches (97). However, combining our structural and biochemical knowledge the domain could be unambiguously assigned to the NTD of XPB and the human TFIIH/p52ct/p8ct model depicted its interaction with p52_MD2. The different domain arrangement of p52 within TFIIH as compared to the isolated p52 brings the antiparallel β-sheet of p52_MD2 in close proximity to XPB_NTD, which as well consists of an antiparallel β-sheet that points directly towards p52 (Figure III.31 B). Their interaction seems to be mediated by an orthogonal packing of their β-sheets. The p52_E359K variant, which showed a loss of interaction in the native PAGE analysis, most likely disturbed the electrostatic interaction with a positively charged residue within the opposite β-strand of XPB_NTD. Unfortunately, the poly-alanine stretches that were modelled for the human XPB_NTD could not be connected, which prevented a sequence assignment for this domain. In the yeast TFIIH structure, the XPB_NTD was not modelled at all but there is additional density at the same position that highly resembles the fold of XPB_NTD and further supported our findings.

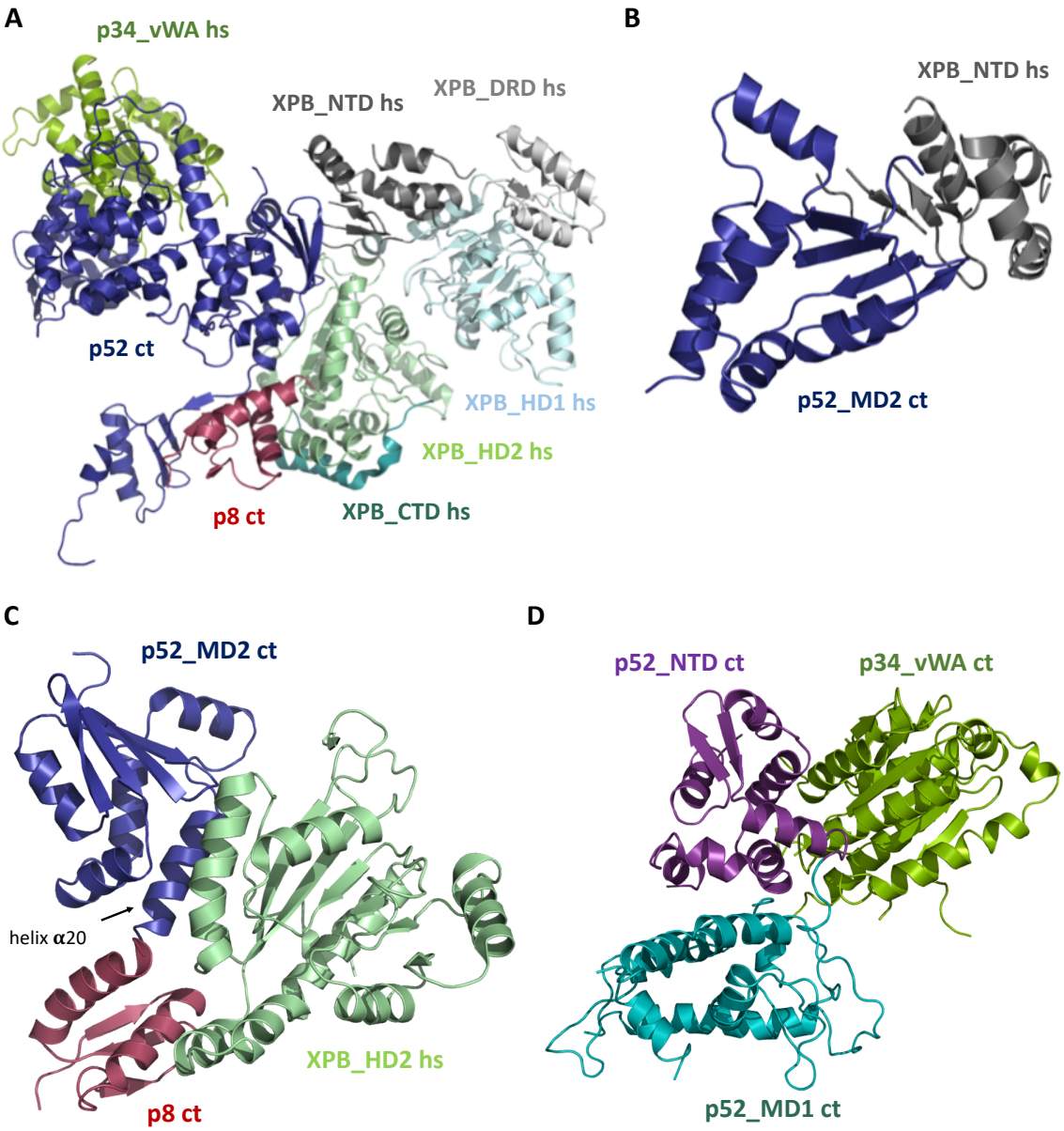


Figure III.31: Depiction of p52 interactions within TFIIH. (A) Individual domains of the p52/p8 crystal structures from *C. thermophilum* (ct: dark blue/red) were built into the human TFIIH cryo-EM structure (PDB code: 5OF4) and are shown with their interaction partners p34 and XPB of human TFIIH. The human p34_vWA (green), XPB_NTD (dark grey), XPB_DRD (light grey), XPB_HD1 (light blue), XPB_HD2 (light green) domains and part of the XPB_CTD (dark teal) domain are indicated. **(B)** The interaction between p52_MD2 from *C. thermophilum* (ct: dark blue) and XPB_NTD (hs: dark grey) as present in the human TFIIH cryo-EM model is shown, which is mediated by the two antiparallel β-sheets of p52_MD2 and XPB_NTD. **(C)** The interaction between p52_MD2 and p8 from *C. thermophilum* (ct: dark blue/red; helix α20 of p52_MD2 is indicated) and XPB_HD2 (hs: light green) is shown. **(D)** The interaction between the NTD and MD1 domains of p52 (ct: purple and teal, respectively) and p34_vWA (ct: green) as built in the yeast TFIIH cryo-EM structure (PDB code: 5OQJ) is shown.

According to both cryo-EM structures and in addition to previous knowledge p52_MD2 is also interacting with the HD2 of XPB and the latter interacts with p8 as well (Figure III.31 C). The interaction between p52_MD2/p8 from *C. thermophilum* as built into the human TFIIH structure and human XPB_HD2 was analyzed by PDBsum. The interaction to XPB_HD2 involves mainly

helix α20 of p52 (Figure III.31 C) as well as the β-turns that are connecting the β-strands of p52_MD2. Eleven residues of XPB_HD2 and twelve residues of p52_MD2 participate in the formation of the interface, which covers an area of approximately 530 Å^2. The interface between XPB_HD2 and p52_MD2 is rather electrostatic involving nine charged residues. Three salt bridges, one hydrogen bond and 81 non-bonded contacts are formed. The interface between XPB_HD2 and p8 covers a surface area of approximately 330 Å^2 involving only nine residues altogether, which form two hydrogen bonds and 55 non-bonded contacts.

Interestingly, the p52_NTD domain interacts with p34_vWA but not with XPB (Figure III.31 A+D), which is in line with our native PAGE analyses. The interaction between p52 and p34 was observed in the human and yeast TFIIH structures and analyzed in greater detail in the yeast TFIIH/p52ct/p8ct model due to its better resolution of this part of the model. The p34_vWA structure from *C. thermophilum* (PDB code: 5NUS) that was used for this analysis was built into the yeast TFIIH EM structure from amino acid 18 to 285 as depicted in III.1.5 and the interaction to p52_NTD_MD1 from *C. thermophilum* as built into yeast TFIIH (Figure III.31 D) was analyzed by PDBsum. The interface between p52 and p34_vWA involves mainly residues of p52_NTD and the linker, which is connecting the NTD and MD1 domains of p52 (residues 116 - 119), as well as one residue of p52_MD1 (T317). The interface is mainly hydrophobic and engages 17 hydrophobic or neutral residues out of the 26 residues of both proteins that are participating in the interface. 119 van der Waals contacts and two salt bridges are formed in an interface area that covers approximately 800 Å^2. According to the yeast TFIIH/p52ct/p8ct model three α-helices of p62 might also participate in the interface of p34_vWA and p52_NTD.

IV Discussion

Prior to this work, very little structural knowledge on the TFIIH complex was available. Furthermore, detailed biochemical data on the intricate interaction network within TFIIH was lacking. XPD's helicase activity and XPB's ATPase/translocase activity were known to be indispensable for NER (41, 43, 105). However, detailed knowledge of how these two enzymes work together during NER was still missing. In this work, we investigated the structural and functional details that underlie the two sub-complexes p34/p44 and XPB/p52/p8 of TFIIH and propose a model of how the ATPase activity of XPB is stimulated by p52/p8 and DNA, which allows the enzyme to fulfill its translocase activity during NER.

IV.1 The p34/p44 Complex

Many interactions had been proposed that underline the composition of the ten-subunit complex TFIIH. The p44 subunit was described to interact with XPD, p62 and p34 within the TFIIH network (103, 122, 142). Yet, structural knowledge of these interactions was lacking and the p34_vWA/p44_RING crystal structure from *C. thermophilum* was the first high resolution structural information on any interaction of p44. High-resolution structural data on TFIIH interactions are, however, indispensable to understand the functional networks within TFIIH. It is still puzzling how the individual subunits of TFIIH work together to execute its function in NER or in transcription initiation. P44 was known to stimulate the helicase activity of XPD, which is indispensable for NER (43, 103, 104), but its interaction with p34 was not as well characterized, thus, we aimed to investigate the interaction between the p34/p44 pair in more detail.

IV.1.1 The interaction between p34 and p44

The high-resolution structure of the p34/p44 minimal complex permitted the identification of several residues that were essential for the interaction between p34_vWA and p44_RING. Two essential residues, K155 of p34 and F490 of p44, had already been described (135, 180). In the work herein an additional residue, A151 of p34, was found to be likewise important for the p34_vWA/p44_RING interaction. The A151 residue was mutated to the negatively charged residue glutamic acid, which most likely weakened van der Waals contacts in close proximity to

the Zn I binding site and therefore disrupted the interaction between p34_vWA and p44_RING (Figure III.6 and Figure III.7). Despite the rather drastic mutation, the overall fold of p34_vWA_A151E had been maintained (Figure III.5) and hence, the lost interaction was not based on an incorrect folding of the p34_vWA_A151E variant.

Having identified three residues that were crucial for the p34_vWA/p44_RING interaction, their effect on the integrity of TFIIH was analyzed. Therefore, the A151E and K155E mutations were introduced into full-length p34 and their interaction with full-length p44 was investigated. Interestingly, the full-length p34 and p44 variants formed a complex even though the p34_vWA/p44_RING interface was disrupted by the A151E or K155E mutation, respectively. So far, no further interactions between p34 and p44 had been identified that could have compensated for the lost interaction between the p34_vWA and p44_RING domains.

Thus, the presence of an additional interface between p34 and p44 was investigated and a possible interaction between the p44_RING and p34_C4 domains was analyzed. A-SEC experiments of p44_RING and full-length p34 variants indicated complex formation and verified that the C4 domain of p34 was able to partially rescue a disrupted interaction between p34_vWA and p44_RING. Having identified a previously unknown interaction between p34 and p44 this additional interface was attempted to be characterized structurally. Unfortunately, the crystallization of the p34_vWA/p44_RING minimal complex was preferred and the interface between p44_RING and p34_C4 could not be visualized most probably because the p34_C4 domain was relatively unstable and tends to degrade (183). In comparison to the sequence of the human and yeast p34 the vWA and C4 domains of p34 from *C. thermophilum* are connected by a relatively long insertion region (Figure IV.1 A), which might have further complicated the structural characterization of this domain. So far, the role of p34_C4 had been unknown and no particular interactions had been proposed for this domain. The additional interface between p44_RING and p34_C4 defines its function and highlights its role within TFIIH.

The effect of a disrupted p34_vWA/p44_RING interaction on the activity of TFIIH was analyzed by our collaboration partners (groups of Jean-Marc Egly and Arnaud Potzersman) utilizing the human proteins. Interestingly, mutations that disrupted an interaction between the p34_vWA and p44_RING domains had no influence on the NER or transcriptional activity of TFIIH, whereas a TFIIH complex containing a p34 variant that lacked the C4 domain was not able to pursue NER or transcription any more (181). These findings emphasized the importance of p34_C4 for TFIIH and for the integrity of the p34/p44 protein pair, which is most likely central to TFIIH stability and function. The strong and redundant interaction network between the p34/p44 pair proposes the presence of a failsafe mechanism, which ensures an intact protein pair. The multipoint interaction

between p34 and p44 indicates why so far no patient mutations have been identified in these two TFIIH subunits.

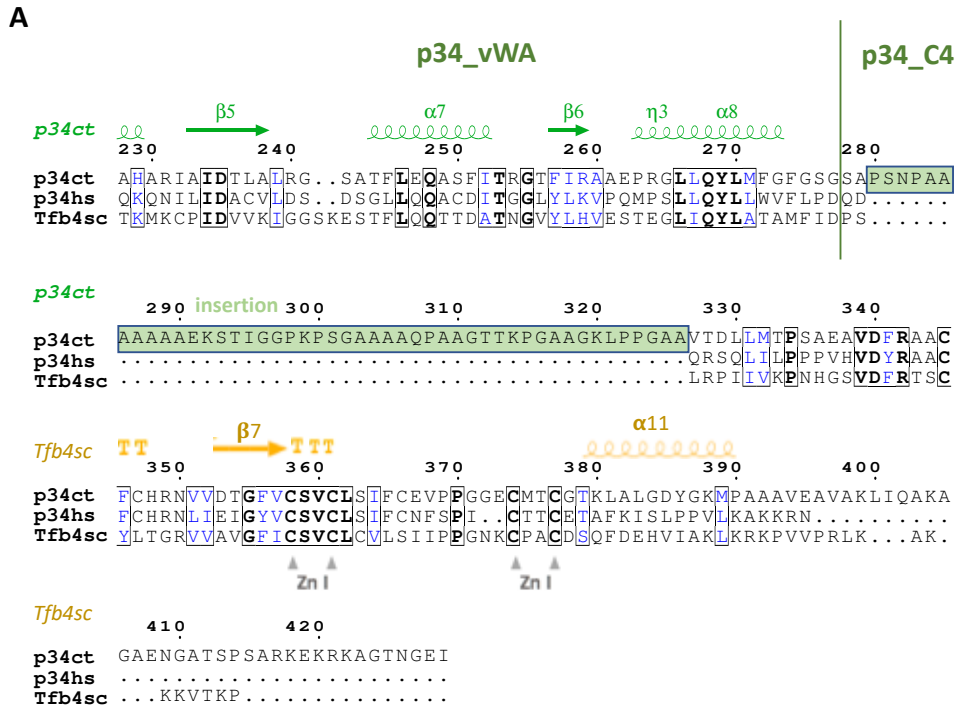
IV.1.2 The p34/p44 interaction within the TFIIH structure

To gain more information on the p34/p44 pair, the p34_vWA/p44_RING crystal structure from *C. thermophilum* was analyzed within the yeast TFIIH structure (98) as the full-length structures of both proteins were modelled into this cryo-EM structure, which was an advantage in comparison to the human TFIIH cryo-EM structure.

The high-resolution structure of the p34_vWA/p44_RING complex from *C. thermophilum* was built into the electron density map of the cryo-EM structure and the *C. thermophilum* and yeast structures were found to be very similar as indicated by an rmsd of 1.6 Å for p34 and 1.9 Å for p44.

The interaction of p34 with further TFIIH subunits was analyzed in the EM structure, which showed that p34_vWA interacts with three rather separated α-helices that were built as a poly-alanine chain. Although this α-helical stretch was not connected to any subunit, it was attributed to p62 in the cryo-EM structure, a subunit, which is spread over several distinct areas within TFIIH. Even if these three α-helices represent an interaction between p34 and p62, the interaction would most likely not be strong enough to hold both proteins together if they were isolated. In the course of the A-SEC experiments of the *C. thermophilum* proteins, an interaction between these two proteins was excluded because the chromatogram and SDS-PAGE of the A-SEC experiment unambiguously indicated that the p34 and p62 did not interact with each other (Figure III.8). Hence, p62 was thought to only be a stabilizing factor for p44 when performing A-SEC experiments between p44/p62 and p34 variants. The reason for using the p44/p62 complex in this set of experiments was the fact that p44 could not be purified on its own in sound quality as it tends to form higher oligomers that cannot be separated from a monomeric fraction of p44 (180). Due to this fact, the p44/p62 complex was used for interaction analysis with full-length p34. When expressed and purified together with its binding partner p62 full-length p44 was stable and of sound quality (182). However, an interaction between the three α-helices, which are relatively unaffiliated with respect to the remaining parts of p62 might not be sufficient to show an interaction between p34 and p62 during A-SEC. The interaction as seen in the A-SEC experiments of p44/p62 with p34_A151E and p34_K155E, respectively, can thus be attributed to the additional interface between p44_RING and p34_C4. A-SEC experiments of p44_RING and full-length p34 variants clearly validated these findings.

Despite many efforts a crystal structure of p34/p44 variants from *C. thermophilum* that would allow a structural characterization of the additional interface could not be obtained (183). The yeast cryo-EM structure, however, offers a structural model of full-length p34 and p44, which illustrates the additional interface. An additional α-helix at the N-terminus of p44_RING (helix α8) could be identified that is pivotal for the interaction between p44_RING and p34_C4. Helix α8 of p44 is connecting the p44_RING and p44_C4 domains and was included in the p44_RING variant of *C. thermophilum* but was not visible in the minimal complex crystal structure (Figure IV.1 B+C). Interestingly, a sequence alignment of the human, yeast and *C. thermphilum* sequences of p44 showed that helix α8 is highly conserved (Figure IV.1 B). The high sequence conservation emphasizes the importance of this α-helix for the p34/p44 interaction. According to the EM structure helix α8 protrudes from the p44_RING domain and intercalates in between the vWA and C4 domains of p34 (Figure IV.1 C). Helix α8 seems to slightly interact with p34_vWA but the main interaction is executed towards p34_C4 involving two β-turns of p34_C4, one of which is located N-terminal to β-strand 7 and one is involved in the Zn binding site of p34_C4. Helix α8 of p44 seems to anchor the C4 domain of p34, which tends to be very unstable at least in p34 from *C. thermophilum*, with the vWA domain of p34.



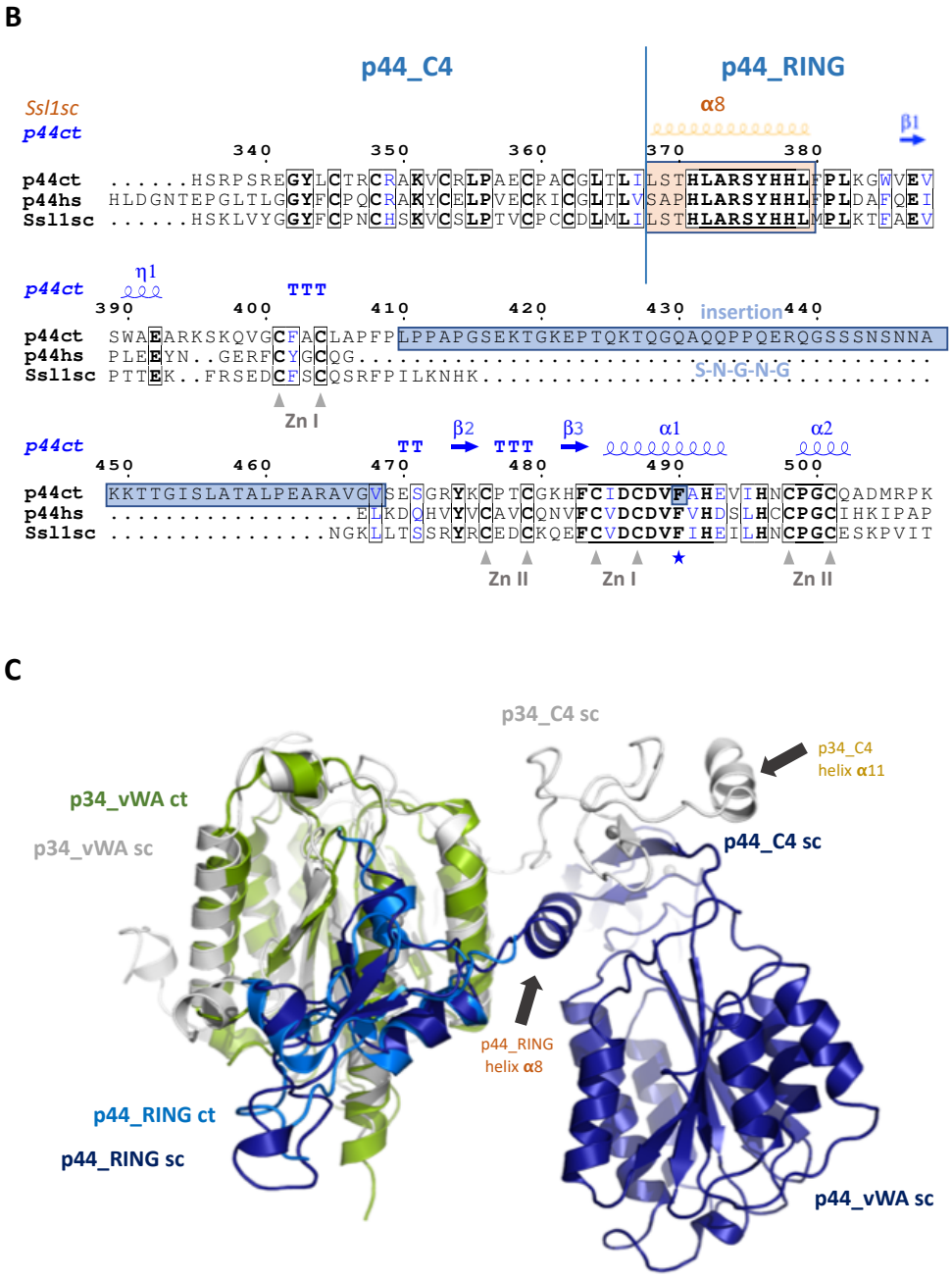


Figure IV.1: The p34/p44 interaction. (A) Sequence alignment of the C-terminal end of p34_vWA and p34_C4 from *C. thermophilum* (ct), human (hs) and *S. cerevisiae* (sc; Tfb4). **(B)** Sequence alignment of the C-terminal end of p44_C4 and N-terminal part of p44_RING from *C. thermophilum* (ct), human (hs) and *S. cerevisiae* (sc; Ssl1). Both sequence alignments were performed by Clustal Omega. Secondary structure elements (p34/p44 MC II for p34ct and p34/p44 MC I for p44ct) are indicated by arrows, spirals, η or TT for β-strands, α-helices, 3_{10} helices and β-turns, respectively (generated by ESPript). Additional secondary structure elements of p34 (Tfb4) or p44 (Ssl1) as defined in the yeast cryo-EM structure (PDB code: 5OQJ) are indicated according to the same symbols. The approximate domain boundaries are marked by lines. Conserved residues are highlighted in black boxes with the most conserved residues depicted in bold. Residues, which were mutated and analyzed, are marked by stars. The nomenclature is according to the sequences of ct. The flexible linker region reaching from residue 410 – 468 (insertion) of p44ct, which was replaced by S-N-G-N-G, is marked in blue. **(C)** The p34_vWA and p44_RING domains from ct (green/blue) and sc (grey/dark blue) were superimposed and the full-length structures of p34 and p44 from sc are shown. The interface between p34_C4 and p44 is depicted and helix α8 of p44_RING and helix α11 of p34_C4 are marked by an arrow. The bound zinc ions are depicted as grey spheres.

However, the interaction between p34_C4 and p44 is not only mediated by helix α8 of p44_RING. On its opposite site β-strand 7 and helix α11 of the p34_C4 domain interact with β-strand 7 of p44, which is located at the N-terminal end of the p44_C4 domain (Figure IV.1 C). Thus, the interaction between p34_C4 and p44 is even more intricate. The fact that p44_RING does only partially represent the interaction region towards p34_C4 might account for the rather weak interaction that was obtained for the full-length p34 and p44_RING variants compared to the interaction between full-length p44 and p34. The HDX-MS experiments of our collaboration partners identified a region within p34_C4 (amino acids 288 - 297 in human p34) that showed a significant reduction in exchange when interacting with p44 (181). Interestingly, this region corresponds to helix α11 of p34_C4, which is interacting with β-strand 7 of p44_C4, highly validating the findings of the HDX-MS experiments.

The interactions with helix α8 of p44_RING on the one hand and β-strand 7 of p44_C4 on the other hand seem to stabilize p34_C4 and anchor it to the p34/p44 pair. Furthermore, the zinc binding site of p34_C4 seems to have a stabilizing effect on its architecture and although it is not directly interacting with p44 the secondary structure elements that stabilize the Zn binding site are involved in interactions with helix α8 of p44_RING. Interestingly, the p34 and p44 proteins bind four zinc ions altogether. The Zn^{2+}-cysteine complexes might structurally stabilize the proteins, especially the C4 domains, which are small domains that are relatively unstable and seem to require their interaction partners for stability. As stated above, the p34_C4 was rather unstable and degraded during the crystallization process of several p34/p44 variants. Furthermore, the p34_C4 domain could not be located in the electron density maps of full-length p34 from *C. thermophilum* (143). The p44_C4 domain displayed the same instability. Purifications of full-length p44 led to the formation of higher oligomers if not purified together with its interaction partner p62, a phenomenon that was attributed to the C4 domain of p44 (180). P62 was shown to interact with the p44_C4 domain and thus seems to have a stabilizing effect on this domain (182).

The additional interface that was observed between the p34_C4 and p44_RING domains of *C. thermophilum* was thus confirmed by modelling the p34_vWA/p44_RING crystal structure into the cryo-EM structure of yeast TFIIH. The cryo-EM structure of yeast TFIIH did not only visualize the interface between p34_C4 and p44_RING but revealed an even more intertwined interaction network between p34 and p44 involving an additional interface between p34_C4 and p44_C4. These findings strengthen the conclusion that the p34/p44 pair forms a central entity to anchor further TFIIH subunits as XPD and p52 within TFIIH.

IV.2 The Interaction Network of XPB, p52 and p8

The p44 subunit of TFIIH is known to stimulate the helicase activity of XPD, whereas p52 stimulates the ATPase activity of XPB (103, 105). During NER, the helicase activity of XPD and the ATPase activity of XPB are both required (41, 43, 105). Having investigated the interaction between the p34/p44 pair on the one hand, we aimed to have a closer look at the interaction network between XPB and its partners p52 and p8 on the other hand. Structural knowledge about the p52 subunit was lacking with the exception of its CTD (126). Thus, a structural characterization of the entire p52 subunit was pursued to obtain a basis for its interaction with XPB.

IV.2.1 The p52 structure

In this work, the first complete atomic resolution structure of p52 from *C. thermophilum* was obtained. Despite the lack of a crystal structure of full-length p52 the entire structure of p52 could be obtained by solving the crystal structures of the p52_1-321, p52_121-EdL and p52_121-EdL/p8 variants. As the MD1 domains are present in the p52_1-321 and p52_121-EdL/p8 crystal structures and are highly similar as indicated by an rmsd of 0.87 Å, they can be superimposed to obtain a structural model of full-length p52 and its interaction with p8. The structural model of p52/p8 illustrates the architecture of p52 and revealed that p52 is composed of four distinct domains. It further visualized the interaction between p52 and p8 in the environment of full-length p52 and offered a precise structural basis to analyze the interactions of p52 with additional TFIIH subunits.

Interestingly, both p52 variants, p52_1-321 and p52_121-EdL, formed dimers in the respective crystal structures. A SEC-MALS analysis of the p52_121-EdL variant also verified dimer formation in solution (Figure III.12 C). However, the p52_121-EdL dimer seemed to be essentially favored by two α-helices that were not part of the protein but represented the crystallized TEV cleavage site (Figure III.14). The TEV helices protrude from each monomer and embrace the opposite monomer. The intricate dimerization hindered access to the TEV cleavage site in the p52_121-EdL variant and prevented a His_6-tag cleavage (Figure VI.5). Attempts to purify and crystallize p52_121-EdL variants that lacked the N-terminal His_6-tag and TEV cleavage site sequences remained unsuccessful. P52_121-EdL variants lacking the protective TEV helix behaved differently and were not as stable as the original p52_121-EdL variant. The dimerization of the p52_121-EdL variant was thus presumably favored by the N-terminal His_6-tag and TEV cleavage site sequences, which seemed to have a highly stabilizing effect on p52. Out of the 106 residues that participate

in the formation of the dimer interface of p52_121-EdL 52 residues are hydrophobic and 30 residues are neutral. Altogether 252 van der Waals contacts are formed rendering the largest part of the dimer interface hydrophobic. However, 17 charged residues are present as well and four salt bridges and eleven hydrogen bonds are formed. The dimerization protects most likely the mainly hydrophobic surface of the MD1 and MD2 domains from solvent exposure and thus stabilizes the p52 variant.

Interestingly, the artificial linker (amino acids S-N-G-N-G replacing amino acids 322 - 344) present in the p52_121-EdL variant could not be visualized in the crystal structure due to its flexibility, which renders the MD1 and MD2 domains of p52 unconnected with each other. After structure solution of p52_121-EdL, the structure of p52_121-End was determined to analyze the influence of the artificial linker. Both structures could be superimposed with an rmsd of 0.5 Å and the orientation of the domains with respect to each other was found to be the same, which excluded an influence of the artificial linker on the domain orientation (Figure III.15). The fact that the MD1 and MD2 domains are not connected with each other in either crystal structure offers the possibility that the MD1 and MD2 domains are connected differently in the dimer as demonstrated in Figure IV.2 B. However, a structure of alternatively composed MD1 and MD2 domains would not explain the stabilizing effect of the TEV helix on the p52_121-EdL dimer, which renders the composition as presented in the result part and Figure IV.2 A the most probable one.

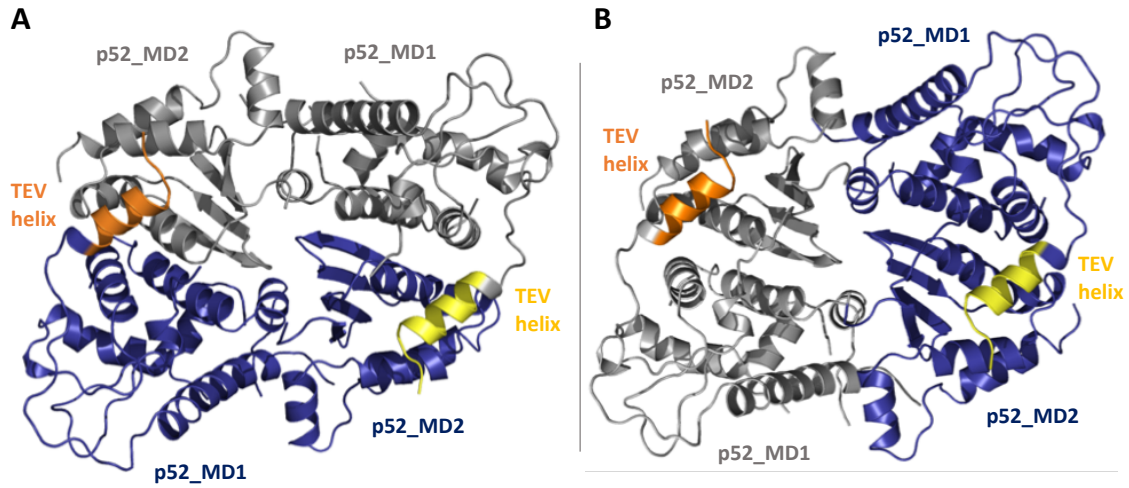


Figure IV.2: Alternative composition of the p52_121-EdL dimer. (A) Structure of a p52_121-EdL dimer. One molecule is colored in dark blue, whereas the other molecule is colored in grey. Each molecule of the dimer consists of MD1 and MD2 and the crystallized TEV cleavage site (TEV helix). **(B)** Alternative composition of the p52_121-EdL dimer. One molecule is colored in grey, whereas the other molecule is colored in dark blue. Each molecule of the dimer consists of MD1 and MD2 and the TEV helix, however, the strong dimerization is not favored by the TEV helices in the alternative composition.

Despite the pronounced dimerization of the p52_121-EdL variant an equivalent dimer formation in full-length p52 would not be possible due to the presence of the NTD. A superposition of the p52_121-EdL and p52_1-321 structures revealed that the orientation of the NTD would overlap with the formation of the p52_121-EdL dimer (Figure III.18 B). Interestingly, the position of the TEV helix overlaps with the position of helix α3 of the NTD in p52_1-321, yet, the connection of the NTD to MD1 differs significantly in both structures underlining the artificial character of the p52_121-EdL dimer, which is most likely not relevant *in vivo*.

The tendency of isolated p52 to form dimers was already reported in several previous studies. Full-length p52 was shown to elute as a dimer from the SEC and variants of yeast p52 that correspond to the CTD and the NTD-MD1 domains were also shown to dimerize (126, 127). However, when incorporated into TFIIH p52 was found to be monomeric (97, 98). Thus, the structures of p52 and p8 from *C. thermophilum* were analyzed within the cryo-EM structures of human and yeast TFIIH (TFIIH/p52ct/p8ct structures) (97, 98). Interestingly, the MD2 domain of p52 was found to be oriented differently when built into TFIIH as compared to its orientation in the crystal structure of p52_121-EdL (Figure III.30 D). The antiparallel β-sheet of MD2 consisting of β-strands 8 to 11 is pointing towards MD1 in the isolated p52_121-EdL structure, whereas it is pointing towards XPB_NTD when incorporated into TFIIH. It thus seems to be a prerequisite to shelter the antiparallel β-sheet by forming an interaction. The artificial p52_121-EdL dimer protects the antiparallel β-sheet from solvent exposure, which might contribute to the enhanced stability of the p52_121-EdL dimer. As stated above, the MD1 and MD2 domains could theoretically be composed differently in the p52_121-EdL dimer. Thus, the orientation of MD2 as present in TFIIH was compared with the alternative domain composition of the p52_121-EdL dimer. However, the alternative domain composition of the dimer varied as well from the orientation of MD2 as observed within TFIIH. This emphasizes that the antiparallel β-sheet of MD2 is only oriented towards XPB_NTD when incorporated into TFIIH but protected by the strong dimer interface within p52_121-EdL.

The human and yeast TFIIH/p52ct/p8ct models further revealed that p52_NTD interacts with p34_vWA (Figure III.31 D). Interestingly, the interface between p52_NTD and p34_vWA covers nearly the same area as the p52 dimer interface present in the p52_1-321 crystal structure. The p52_1-321 dimer is mainly mediated by interactions between the NTD domains of both molecules and only a few additional contacts are formed between residues of the MD1 domains of both molecules. The interactions between the NTD domains of the p52_1-321 dimer and the p52_NTD/p34_vWA domains are both mediated by helices α3, α4 and α7 of p52_NTD. Additional secondary structure elements of the NTD and MD1 domains of p52 that are involved in the

dimerization are giving rise to the more extended interface of the p52_1-321 dimer (1,200 Å^2) as compared to the p52/p34 interface (800 Å^2).

The ability of p52 to form diverse dimers reflects its ability to interact with several TFIIH subunits and supports its role as a scaffolding protein within TFIIH. The interactions of p52 with additional TFIIH subunits traps p52 into its position within TFIIH and alters the domain arrangement, which is distinct from the domain arrangement of isolated p52. The NTD of p52 forms a homodimer within the p52_1-321 variant but is interacting with p34_vWA within TFIIH via the same interface. MD2 of p52 participates in the formation of a strong dimer interface of the p52_121-EdL variant but interacts with XPB_NTD within TFIIH. The CTD of p52 forms a homodimer on its own but interacts with p8 via the same interface creating a pseudosymmetric heterodimer (126). The same applies for p8, which forms a homodimer, if not being incorporated into TFIIH (128, 130), illustrating the high similarity between p8 and p52_CTD. It can thus be summarized that the presence of its binding partners XPB, p34 and p8 anchors p52 within TFIIH and withdraws the freedom from p52 to build dimers or reorient its domains.

IV.2.2 The p52/p8 interaction

The interaction of p52_CTD with p8 has already been characterized for the yeast proteins (126) but was only described for the isolated p52_CTD/p8 heterodimer so far. The p52_121-EdL/p8 crystal structure illustrates the interaction between p52_CTD and p8 in a more global environment. The p52_CTD/p8 heterodimer forms a compact entity, which is, however, located at a distance to the remaining domains of p52 and only linked by a long bridge consisting of helix α20 and β-strand 12.

The p52/p8 structure raises the question why the CTD is located in such a distance to the remainder of the protein. Obviously, the distant location renders the CTD very flexible and makes it dependent on anchoring by its binding partner p8 to be stabilized. P8 appears to be a mediator between the MD2 and CTD domains of p52 and anchors the CTD to the remainder of p52, which explains why the CTD could only be visualized when it interacts with p8. Built into the cryo-EM structure of human TFIIH p8 was found to be enclosed between the p52_CTD, p52_MD2 and XPB_HD2 domains (Figure IV.4 B). The stabilizing role of p8 towards these interacting domains might explain why shortened variants of p8 lead to an overall reduced concentration of TFIIH in TTD-A cells (129). It further explains the reduced NER activity of TFIIH in the absence of p8 (124, 125). The integrity of p8 is crucial for the function of TFIIH, which is also reflected in the defects observed in patients suffering from TTD (84, 125, 126). P8 seems to be a highly important

contributor to stimulate XPB's ATPase activity and the p52_CTD domain might ensure that p8 is stably associated with TFIIH by positioning it in-between itself and XPB_HD2. Hence, p8 has a stabilizing role towards p52_CTD and vice versa. Without p52_CTD p8 might not be able to bind as tightly to XPB as necessary in order to stimulate its ATPase activity. This is in line with previous observations, which revealed that a TFIIH complex comprising a p52 variant that is lacking the CTD was impaired in NER (126). The transcriptional activity of this complex was, however, not affected. The absence of either p52_CTD or p8 has severe impacts on the activity of TFIIH during NER enlightening their role within TFIIH. As the architecture of TFIIH seems to be unaffected in TFIIH complexes lacking either p52_CTD or p8 (124, 126) the severe consequences of their absence during NER have to be attributed to a reduced stimulation of XPB's ATPase activity. XPB is not only interacting with p8 but also with p52, an interaction that has been investigated further in the herein described work.

IV.2.3 The XPB/p52 interaction

Prior to this work co-immunopurification assays had indicated an interaction between p52 and the N-terminal region of XPB (123). This interaction was analyzed with the isolated proteins from *C. thermophilum* and the XPB_60-345 variant was shown to form a stable complex of equimolar amounts with p52/p8 (Figure III.24) indicated by a K_D of around 130 nM (Figure III.25). The XPB_60-345 variant comprised not only the NTD domain, which was shown to mediate the interaction to p52, but also the DRD domain, which is not required for the interaction with p52/p8 (Figure III.31 A). In addition, the region of p52 that is important for the interaction with XPB was investigated and out of the two proposed XPB binding sites in p52 (123), which correspond to the NTD and MD2 domains of p52, only p52_MD2 was able to interact with XPB_60-345 in native PAGE analyses (Figure III.26 A). These findings are in line with ATPase activity assays that indicated that only the second proposed XPB binding site was required to stimulate XPB's ATPase activity (105) and the TFIIH/p52ct/p8ct models, which revealed that p52_NTD is not interacting with XPB but with p34_vWA. The structural characterization of the p52_MD2/XPB_NTD interaction was attempted by co-expression and -purification of several p52/XPB minimal complexes. Unfortunately, the soluble expression of shortened p52 and XPB variants proved to be difficult and none of the minimal complexes that could be expressed and purified crystallized so far (Table III.7).

According to the human TFIIH/p52ct/p8ct model the interaction between p52 and XPB is mainly executed by an orthogonal packing of their antiparallel β-sheets, which are located in the

p52_MD2 and XPB_NTD domains, respectively (Figure III.31 B). In addition, p52_MD2 and p8 are also interacting with XPB_HD2 (Figure III.31 C). First mutational analyses on the p52_MD2/XPB_NTD interface identified residue E359 of p52 to be crucial for their interaction (Figure III.29). The human TFIIH/p52ct/p8ct model permitted the identification of further residues of p52 that might be involved in the formation of the interface. A combined analysis of the p52_MD2/XPB_NTD interface as present in the human TFIIH/p52ct/p8ct model and the sequence conservation of p52_MD2 led to the identification of residues p52_N361, p52_Y365, p52_F387 and p52_M390 that might be crucial for the interaction and should be analyzed in future studies (Figure IV.3).

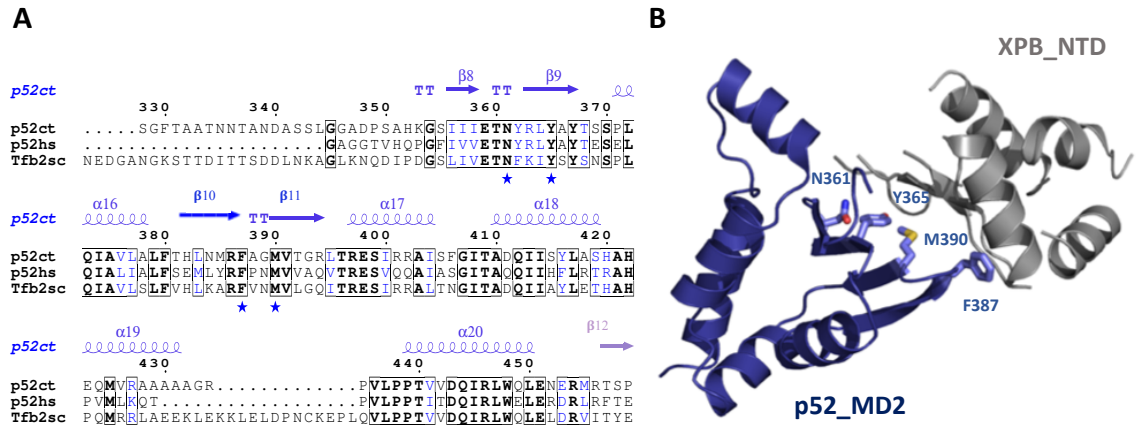


Figure IV.3: Analysis of the p52_MD2/XPB_NTD interface.(A) Sequence alignment of the *C. thermophilum* (ct), human (hs) and *S. cerevisiae* (sc; TfB2) sequences of p52_MD2. The sequence alignment was performed by ClustalW. Secondary structure elements of MD2 are indicated by arrows, spirals or TT for β-strands, α-helices and β-turns, respectively (generated by ESPript). Conserved residues are highlighted in black boxes with the most conserved residues depicted in bold. Four putative interface residues are marked by a star. The nomenclature is according to the sequence of p52ct. **(B)** The interaction between p52_MD2 from *C. thermophilum* (ct: dark blue) and XPB_NTD (dark grey) as present in the human TFIIH cryo-EM structure (PDB code: 5OF4) is shown. The four putative interface residues of p52_MD2 from *C. thermophilum* (N361, Y365, F387 and M390) are shown as sticks

The identification of corresponding XPB residues is, however, more difficult due to the lack of a precise structural model. The F99 residue, which was found to be mutated to a serine in the *XPB* gene of XP/CS patients and thus leads to a highly reduced ATPase activity of XPB (105), is conserved among the human, yeast and *C. thermophilum* proteins, which suggests that this residue is located in one of the β-strands of XPB_NTD. The human and yeast TFIIH/p52ct/p8ct models allowed us to have a closer look on the XPB/p52/p8 interaction network and to analyze the influence of the p52/p8 pair on the function of XPB.

IV.2.4 Model of XPB stimulation

The ATPase activity of XPB and the helicase activity of XPD are both required for NER (41, 43, 105). The importance of the interaction between p52 and XPB is reflected in the F99S mutation of XPB, which weakens the XPB/p52 interaction and leads to a highly reduced ATPase activity of XPB, an effect that is comparable to a mutation in the Walker A motif of XPB (105). The DNA-dependent ATPase activity of XPB is further influenced by its R-E-D and ThM motifs, which are located in the HD1 and HD2 domains, respectively (41). During transcription, the ATPase activity of XPB and the integrity of the R-E-D and ThM motifs seem to be likewise important for RNA Pol II promoter opening suggesting a similar mode of action for XPB in both processes (41). The ATPase activity of XPB seems to be dependent on the close proximity of the R-E-D and ThM motifs to each other, which leads to a closing of its two RecA like domains and is presumably provoked by DNA binding (41, 113). In the DNA bound structure of yeast TFIIH the R-E-D and ThM motifs of XPB come into close proximity (approximately 20 Å) to each other and both interact with the DNA duplex. The structure was captured with an empty ATPase active site in the pre-translocation state (98). The DNA seems to enhance a closing of the RecA like domains as the distance between the R-E-D and ThM motifs was larger in the non-DNA bound state of human TFIIH (approximately 30 Å; compare Figure IV.4 A and C) (97).

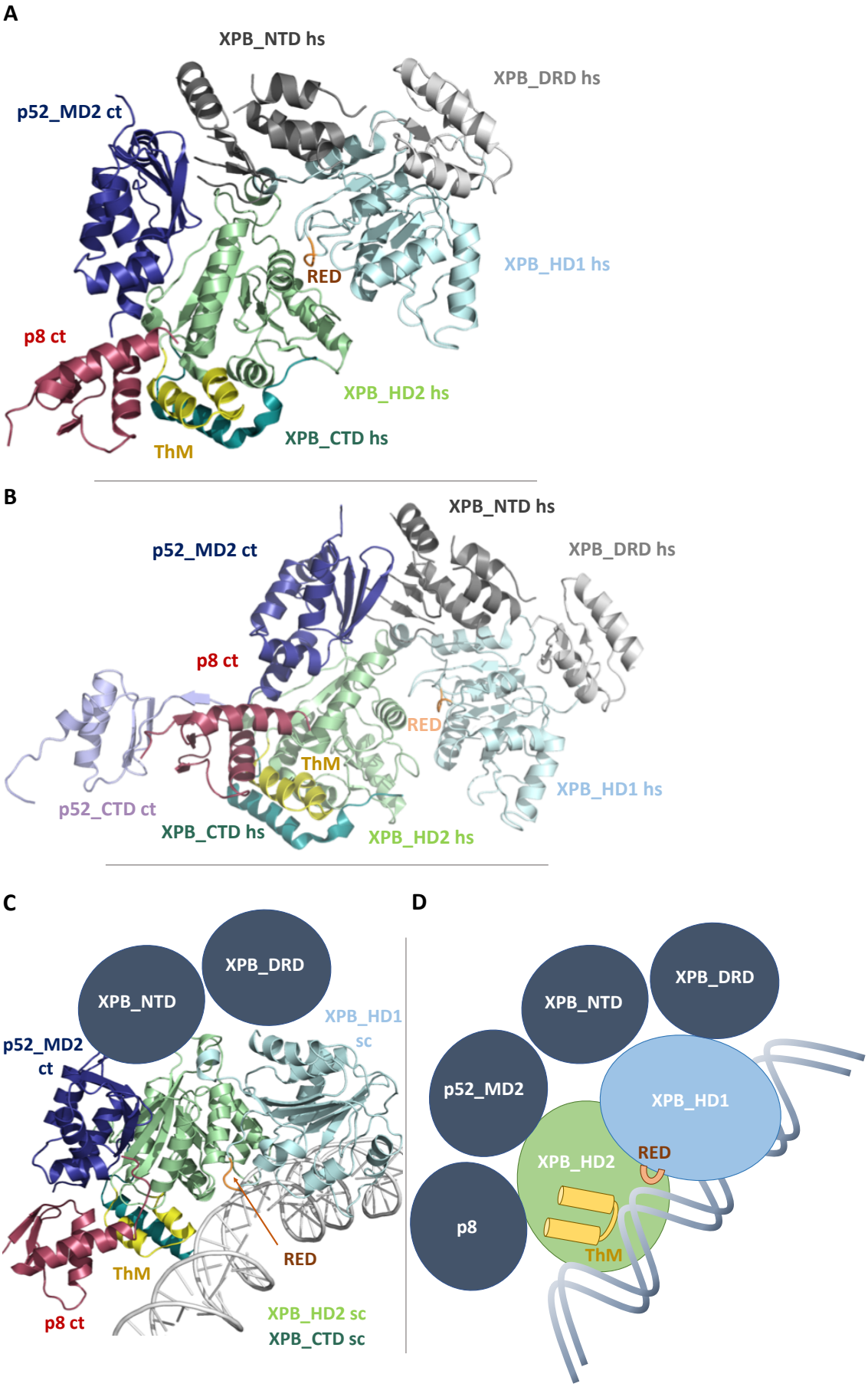
DNA binding, a functional Walker A motif and intact R-E-D and ThM motifs, which execute the closing of the RecA like domains, seem thus to be prerequisites for the activity of XPB during NER. The presence and interaction of p52/p8 with XPB is required to further stimulate its functional ATPase activity. The human TFIIH/p52ct/p8ct model illustrates that four entities comprising p8, p52_MD2, XPB_NTD and XPB_DRD encompass the RecA like domains of XPB and may stimulate its ATPase activity by bringing the R-E-D and ThM motifs close together (Figure IV.4). The four entities embrace the two RecA like domains from one side and may facilitate binding to the DNA, which enters from the other side. P8 directly binds to the ThM motif of XPB and may thereby ensure a close contact between the two RecA like domains and the DNA. The p52_CTD domain enhances this stimulation by interacting with p8 on its other side (Figure IV.4 B). The interaction between the p52_MD2 and XPB_NTD domains seems to stabilize the lunate ring in its middle and to ensure that the four entities fulfill their duty. The additional interaction between p52_MD2 and XPB_HD2 further stabilizes this intricate complex.

The TFIIH/p52ct/p8ct models highlight the importance of these interactions in order to stimulate the ATPase activity of XPB. These observations render a model as proposed based on the crystal structure of XPB from *Archaeoglobus fulgidus*, which suggests a flip of XPB_HD2 of around 170°

upon ATP hydrolysis, unlikely (113, 115). The orientation of the RecA like domains in the isolated archaeal XPB as observed in this crystal structure is different than the one of eukaryotic XPB observed within TFIIH. Within TFIIH, the stimulation of XPB's ATPase activity is affected by the entire lunate ring that seems to work as a corporate unit. This effect brings the R-E-D and ThM motifs and hence the two RecA like domains into close proximity of each other. It further explains the drastic effects of a loss of either p52_CTD, p8 or the p52_MD2/XPB_NTD interaction that were all observed to lead to an impairment in NER (84, 105, 125, 126).

The cryo-EM structures of TFIIH were taken as a template model to study the action of XPB during NER. Furthermore, the DNA binding of XPB can be investigated in the yeast TFIIH/p52ct/p8ct model. During transcription XPB was shown to bind the DNA downstream of RNA Pol II and to use ATP hydrolysis to translocate away from RNA Pol II. Due to the fact that the location of TFIIH is fixed by anchoring to the PIC, mostly by interactions to TFIIE, the translocation motion of XPB actively pushes the dsDNA into the RNA Pol II cleft, which leads to an opening of the DNA (98). During GG-NER TFIIH is recruited to the lesion site after damage recognition by XPC-RAD23B-CENT2 and UV-DDB (184). XPB was shown to be required for anchoring TFIIH to the lesion site (41). After DNA binding by XPB its R-E-D and ThM motifs are brought more closely together and the XPB ATPase is activated. This is enhanced by the presence of the lunate ring formed by p8, p52_MD2 and two additional domains of XPB. Just like during transcription the location of TFIIH during GG-NER seems to be fixed by interactions between the TFIIH subunit p62 and XPC (140, 185). Assuming that the mode of action of XPB is similar during transcription and NER a model could be envisioned, in which XPB would bind the dsDNA and use its translocase activity to actively push the dsDNA towards XPD. Assuming that the DNA binding of XPB is the same during transcription and NER a simultaneous binding of dsDNA by XPB and XPD is unlikely. A model, in which XPB pushes the DNA towards XPD would explain, how both enzymes are engaged on the DNA. The initial opening of the dsDNA provoked by XPB's translocase activity would generate an optimal substrate for XPD (107).

Figure IV.4: Model of XPB stimulation. (A) The p52_MD2 and p8 domains of the p52/p8 crystal structures from *C. thermophilum* (ct) were built into the cryo-EM structure of human TFIIH (hs; PDB code: 5OF4) and are shown towards their interaction with XPB. The four entities p8, p52_MD2, XPB_NTD and XPB_DRD form a lunate ring that seems to work as a corporate unit to bring the R-E-D and ThM motifs of XPB into close proximity to each other and to stimulate the ATPase activity of XPB. **(B)** In addition to the composition as in (A) the influence of p52_CTD on p8 is shown. **(C)** The p52_MD2 and p8 domains of the p52/p8 crystal structures from *C. thermophilum* (ct) were built into the cryo-EM structure of yeast TFIIH (sc; PDB code: 5OQJ). The DNA duplex is bound by XPB and brings the R-E-D and ThM motifs in close proximity to each other. The XPB_NTD and XPB_DRD domains had not been built in the yeast TFIIH structure and are thus only represented schematically. **(D)** Schematic overview of the lunate ring, which enhances the rapprochement of the R-E-D and ThM motifs within XPB, which bind dsDNA.


A
XPB_NTD hs
XPB_DRD hs
p52_MD2 ct
XPB_HD1 hs
RED
p8 ct
XPB_HD2 hs
XPB_CTD hs
ThM
B
XPB_NTD hs
p52_MD2 ct
XPB_DRD hs
p8 ct
RED
ThM
XPB_HD1 hs
p52_CTD ct
XPB_CTD hs
XPB_HD2 hs
C
XPB_DRD
XPB_NTD
XPB_HD1
sc
p52_MD2
ct
RED
ThM
p8 ct
XPB_HD2 sc
XPB_CTD sc
D
XPB_NTD
XPB_DRD
p52_MD2
XPB_HD1
XPB_HD2
RED
p8
ThM

IV.3 Final conclusions

The herein described work focused on several subunits of TFIIH, the p34/p44 pair and the XPB/p52/p8 ternary complex. In addition to the known p34_vWA/p44_RING interaction an interaction between the p34_C4 and p44_RING domains, which can compensate for a loss of interaction between the p34_vWA/p44_RING domains, was identified. This finding was further supported by the cryo-EM structure of yeast TFIIH, which identified the interaction of helix α8 of p44_RING with p34_C4. The EM structure visualized an even more intricate network between the p34/p44 pair strongly highlighting our previous assumptions that the interaction between p34 and p44 is central to TFIIH, which justifies safeguarding their interaction by the presence of multiple redundant interfaces. A stable interaction between p44 and p34 seems to be crucial to anchor all subunits within the core TFIIH complex. The p34 subunit of TFIIH further connects the XPD/p44/p62 and XPB/p52/p8 sub-complexes with each other via interactions with p44 and p52, respectively. The interaction to p52 is mediated by the p52_NTD and p34_vWA domains.

The herein described work offers the first high resolution structural model of full-length p52, which is composed of the crystal structures of the p52_1-321 and p52_121-EdL/p8 variants from *C. thermophilum*. The crystal structures highlight the ability of p52 to form dimers via interfaces that form interactions with additional subunits when incorporated into TFIIH. The presence of its binding partners XPB, p34 and p8 anchors p52 within TFIIH. The p52 crystal structures provide a detailed model to investigate the interactions between p52 and its binding partners within TFIIH. The interaction between the p52_MD2 and XPB_NTD domains was investigated in more detail and found to be mediated via an orthogonal packing of their antiparallel β-sheets. The human and yeast TFIIH structures further illustrated an interaction between p52_MD2 and XPB_HD2 as well as between p8 and XPB_HD2. The p52/p8 pair stimulates the ATPase activity of XPB, which is required during transcription and NER (41, 104, 105). The TFIIH/p52ct/p8ct models offered a basis to analyze the ATPase stimulation of XPB, which is mediated by a lunate ring consisting of XPB, p52 and p8 domains, which ensure the close proximity of the R-E-D and ThM motifs present in the RecA like domains of XPB.

So far, the NER mechanism is not completely understood and the modality of how TFIIH executes its function during NER and how XPB and XPD work together are not precisely known. The herein described work offers new observations on five of the seven core TFIIH subunits including the first high resolution description of the entire p52 structure and thus contributes to illuminating the function of TFIIH, a complex involved in two basic cellular processes, transcription and DNA repair.

V Literature

1. Evans MD, Dizdaroglu M, & Cooke MS (2004) Oxidative DNA damage and disease: induction, repair and significance. *Mutat Res* 567(1):1-61.
2. Helleday T, Eshtad S, & Nik-Zainal S (2014) Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet* 15(9):585-598.
3. Strauss BS (2002) The "A" rule revisited: polymerases as determinants of mutational specificity. *DNA Repair (Amst)* 1(2):125-135.
4. Alexandrov LB*, et al.* (2013) Signatures of mutational processes in human cancer. *Nature* 500(7463):415-421.
5. Narayanan DL, Saladi RN, & Fox JL (2010) Ultraviolet radiation and skin cancer. *Int J Dermatol* 49(9):978-986.
6. Pfeifer GP, You YH, & Besaratinia A (2005) Mutations induced by ultraviolet light. *Mutat Res* 571(1-2):19-31.
7. Pfeifer GP*, et al.* (2002) Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene* 21(48):7435-7451.
8. Swenberg JA*, et al.* (2011) Endogenous versus exogenous DNA adducts: their role in carcinogenesis, epidemiology, and risk assessment. *Toxicol Sci* 120 Suppl 1:S130-145.
9. Thomas A*, et al.* (2017) Temozolomide in the Era of Precision Medicine. *Cancer Res* 77(4):823-826.
10. Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature* 362(6422):709-715.
11. Diderich K, Alanazi M, & Hoeijmakers JH (2011) Premature aging and cancer in nucleotide excision repair-disorders. *DNA Repair (Amst)* 10(7):772-780.
12. Hoeijmakers JH (2009) DNA damage, aging, and cancer. *N Engl J Med* 361(15):1475-1485.
13. Chatterjee N & Walker GC (2017) Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mol Mutagen* 58(5):235-263.
14. Christmann M, Tomicic MT, Roos WP, & Kaina B (2003) Mechanisms of human DNA repair: an update. *Toxicology* 193(1-2):3-34.
15. Deans AJ & West SC (2011) DNA interstrand crosslink repair and cancer. *Nat Rev Cancer* 11(7):467-480.
16. Johnson RD & Jasin M (2000) Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. *EMBO J* 19(13):3398-3407.
17. Hoeijmakers JH (2001) Genome maintenance mechanisms for preventing cancer. *Nature* 411(6835):366-374.
18. Gillet LC & Scharer OD (2006) Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chem Rev* 106(2):253-276.
19. Horikoshi N*, et al.* (2016) Crystal structure of the nucleosome containing ultraviolet light-induced cyclobutane pyrimidine dimer. *Biochem Biophys Res Commun* 471(1):117-122.
20. Osakabe A*, et al.* (2015) Structural basis of pyrimidine-pyrimidone (6-4) photoproduct recognition by UV-DDB in the nucleosome. *Sci Rep* 5:16330.
21. Zhang M, Wang L, & Zhong D (2017) Photolyase: Dynamics and electron-transfer mechanisms of DNA repair. *Arch Biochem Biophys* 632:158-174.
22. Ewa B & Danuta MS (2017) Polycyclic aromatic hydrocarbons and PAH-related DNA adducts. *J Appl Genet* 58(3):321-330.

23. Min JH & Pavletich NP (2007) Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* 449(7162):570-575.
24. Chen X*, et al.* (2015) Kinetic gating mechanism of DNA damage recognition by Rad4/XPC. *Nat Commun* 6:5849.
25. Kong M*, et al.* (2016) Single-Molecule Imaging Reveals that Rad4 Employs a Dynamic DNA Damage Recognition Process. *Mol Cell* 64(2):376-387.
26. Mu H, Geacintov NE, Min JH, Zhang Y, & Broyde S (2017) Nucleotide Excision Repair Lesion-Recognition Protein Rad4 Captures a Pre-Flipped Partner Base in a Benzo[a]pyrene-Derived DNA Lesion: How Structure Impacts the Binding Pathway. *Chem Res Toxicol* 30(6):1344-1354.
27. Dualan R*, et al.* (1995) Chromosomal localization and cDNA cloning of the genes (DDB1 and DDB2) for the p127 and p48 subunits of a human damage-specific DNA binding protein. *Genomics* 29(1):62-69.
28. Sugasawa K (2016) Molecular mechanisms of DNA damage recognition for mammalian nucleotide excision repair. *DNA Repair (Amst)* 44:110-117.
29. Datta A*, et al.* (2001) The p48 subunit of the damaged-DNA binding protein DDB associates with the CBP/p300 family of histone acetyltransferase. *Mutat Res* 486(2):89-97.
30. Pines A*, et al.* (2012) PARP1 promotes nucleotide excision repair through DDB2 stabilization and recruitment of ALC1. *J Cell Biol* 199(2):235-249.
31. Han C, Srivastava AK, Cui T, Wang QE, & Wani AA (2016) Differential DNA lesion formation and repair in heterochromatin and euchromatin. *Carcinogenesis* 37(2):129-138.
32. Kusumoto R*, et al.* (2001) Diversity of the damage recognition step in the global genomic nucleotide excision repair in vitro. *Mutat Res* 485(3):219-227.
33. Scrima A*, et al.* (2008) Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex. *Cell* 135(7):1213-1223.
34. Yeh JI*, et al.* (2012) Damaged DNA induced UV-damaged DNA-binding protein (UV-DDB) dimerization and its roles in chromatinized DNA repair. *Proc Natl Acad Sci U S A* 109(41):E2737-2746.
35. Kapetanaki MG*, et al.* (2006) The DDB1-CUL4ADDB2 ubiquitin ligase is deficient in xeroderma pigmentosum group E and targets histone H2A at UV-damaged DNA sites. *Proc Natl Acad Sci U S A* 103(8):2588-2593.
36. Takedachi A, Saijo M, & Tanaka K (2010) DDB2 complex-mediated ubiquitylation around DNA damage is oppositely regulated by XPC and Ku and contributes to the recruitment of XPA. *Mol Cell Biol* 30(11):2708-2723.
37. Sugasawa K*, et al.* (2005) UV-induced ubiquitylation of XPC protein mediated by UV-DDB-ubiquitin ligase complex. *Cell* 121(3):387-400.
38. Yokoi M*, et al.* (2000) The xeroderma pigmentosum group C protein complex XPC-HR23B plays an important role in the recruitment of transcription factor IIH to damaged DNA. *J Biol Chem* 275(13):9870-9875.
39. Volker M*, et al.* (2001) Sequential assembly of the nucleotide excision repair factors in vivo. *Mol Cell* 8(1):213-224.
40. Ruthemann P, Balbo Pogliano C, Codilupi T, Garajova Z, & Naegeli H (2017) Chromatin remodeler CHD1 promotes XPC-to-TFIIH handover of nucleosomal UV lesions in nucleotide excision repair. *EMBO J* 36(22):3372-3386.
41. Oksenych V, Bernardes de Jesus B, Zhovmer A, Egly JM, & Coin F (2009) Molecular insights into the recruitment of TFIIH to sites of DNA damage. *EMBO J* 28(19):2971-2980.
42. Winkler GS*, et al.* (2000) TFIIH with inactive XPD helicase functions in transcription initiation but is defective in DNA repair. *J Biol Chem* 275(6):4258-4266.

43. Kuper J*, et al.* (2014) In TFIIH, XPD helicase is exclusively devoted to DNA repair. *PLoS Biol* 12(9):e1001954.
44. Sugasawa K*, et al.* (2001) A multistep damage recognition mechanism for global genomic nucleotide excision repair. *Genes Dev* 15(5):507-521.
45. Wittschieben BO, Iwai S, & Wood RD (2005) DDB1-DDB2 (xeroderma pigmentosum group E) protein complex recognizes a cyclobutane pyrimidine dimer, mismatches, apurinic/apyrimidinic sites, and compound lesions in DNA. *J Biol Chem* 280(48):39982-39989.
46. Sugasawa K, Akagi J, Nishi R, Iwai S, & Hanaoka F (2009) Two-step recognition of DNA damage for mammalian nucleotide excision repair: Directional binding of the XPC complex and DNA strand scanning. *Mol Cell* 36(4):642-653.
47. Mathieu N, Kaczmarek N, & Naegeli H (2010) Strand- and site-specific DNA lesion demarcation by the xeroderma pigmentosum group D helicase. *Proc Natl Acad Sci U S A* 107(41):17545-17550.
48. Wolski SC*, et al.* (2008) Crystal structure of the FeS cluster-containing nucleotide excision repair helicase XPD. *PLoS Biol* 6(6):e149.
49. Kuper J, Wolski SC, Michels G, & Kisker C (2012) Functional and structural studies of the nucleotide excision repair helicase XPD suggest a polarity for DNA translocation. *EMBO J* 31(2):494-502.
50. Liu H*, et al.* (2008) Structure of the DNA repair helicase XPD. *Cell* 133(5):801-812.
51. Fan L*, et al.* (2008) XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. *Cell* 133(5):789-800.
52. Mathieu N, Kaczmarek N, Ruthemann P, Luch A, & Naegeli H (2013) DNA quality control by a lesion sensor pocket of the xeroderma pigmentosum group D helicase subunit of TFIIH. *Curr Biol* 23(3):204-212.
53. Li CL*, et al.* (2015) Tripartite DNA Lesion Recognition and Verification by XPC, TFIIH, and XPA in Nucleotide Excision Repair. *Mol Cell* 59(6):1025-1034.
54. Coin F*, et al.* (2008) Nucleotide excision repair driven by the dissociation of CAK from TFIIH. *Mol Cell* 31(1):9-20.
55. Sugitani N, Sivley RM, Perry KE, Capra JA, & Chazin WJ (2016) XPA: A key scaffold for human nucleotide excision repair. *DNA Repair (Amst)* 44:123-135.
56. Yang Z*, et al.* (2006) 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(51):15921-15930.
57. Koch SC*, et al.* (2015) Structural insights into the recognition of cisplatin and AAF-dG lesion by Rad14 (XPA). *Proc Natl Acad Sci U S A* 112(27):8272-8277.
58. Kuper J & Kisker C (2012) Damage recognition in nucleotide excision DNA repair. *Curr Opin Struct Biol* 22(1):88-93.
59. Nocentini S, Coin F, Saijo M, Tanaka K, & Egly JM (1997) DNA damage recognition by XPA protein promotes efficient recruitment of transcription factor II H. *J Biol Chem* 272(37):22991-22994.
60. Ziani S*, et al.* (2014) Sequential and ordered assembly of a large DNA repair complex on undamaged chromatin. *J Cell Biol* 206(5):589-598.
61. Hermanson-Miller IL & Turchi JJ (2002) Strand-specific binding of RPA and XPA to damaged duplex DNA. *Biochemistry* 41(7):2402-2408.
62. Fanning E, Klimovich V, & Nager AR (2006) A dynamic model for replication protein A (RPA) function in DNA processing pathways. *Nucleic Acids Res* 34(15):4126-4137.
63. Daughdrill GW*, et al.* (2003) Chemical shift changes provide evidence for overlapping single-stranded DNA- and XPA-binding sites on the 70 kDa subunit of human replication protein A. *Nucleic Acids Res* 31(14):4176-4183.

64. Tsodikov OV*, et al.* (2007) Structural basis for the recruitment of ERCC1-XPF to nucleotide excision repair complexes by XPA. *EMBO J* 26(22):4768-4776.
65. Orelli B*, et al.* (2010) The XPA-binding domain of ERCC1 is required for nucleotide excision repair but not other DNA repair pathways. *J Biol Chem* 285(6):3705-3712.
66. Dunand-Sauthier I*, et al.* (2005) The spacer region of XPG mediates recruitment to nucleotide excision repair complexes and determines substrate specificity. *J Biol Chem* 280(8):7030-7037.
67. Zotter A*, et al.* (2006) Recruitment of the nucleotide excision repair endonuclease XPG to sites of UV-induced dna damage depends on functional TFIIH. *Mol Cell Biol* 26(23):8868-8879.
68. Staresincic L*, et al.* (2009) Coordination of dual incision and repair synthesis in human nucleotide excision repair. *EMBO J* 28(8):1111-1120.
69. Mocquet V*, et al.* (2008) Sequential recruitment of the repair factors during NER: the role of XPG in initiating the resynthesis step. *EMBO J* 27(1):155-167.
70. Ogi T*, et al.* (2010) Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells. *Mol Cell* 37(5):714-727.
71. Lehmann AR (2011) DNA polymerases and repair synthesis in NER in human cells. *DNA Repair (Amst)* 10(7):730-733.
72. Moser J*, et al.* (2007) Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase III alpha in a cell-cycle-specific manner. *Mol Cell* 27(2):311-323.
73. Kemp MG, Reardon JT, Lindsey-Boltz LA, & Sancar A (2012) Mechanism of release and fate of excised oligonucleotides during nucleotide excision repair. *J Biol Chem* 287(27):22889-22899.
74. Hu J*, et al.* (2013) Nucleotide excision repair in human cells: fate of the excised oligonucleotide carrying DNA damage in vivo. *J Biol Chem* 288(29):20918-20926.
75. Pani B & Nudler E (2017) Mechanistic insights into transcription coupled DNA repair. *DNA Repair (Amst)* 56:42-50.
76. Spivak G (2016) Transcription-coupled repair: an update. *Arch Toxicol* 90(11):2583-2594.
77. Beerens N, Hoeijmakers JH, Kanaar R, Vermeulen W, & Wyman C (2005) The CSB protein actively wraps DNA. *J Biol Chem* 280(6):4722-4729.
78. Spivak G & Ganesan AK (2014) The complex choreography of transcription-coupled repair. *DNA Repair (Amst)* 19:64-70.
79. Spivak G & Hanawalt PC (2015) Photosensitive human syndromes. *Mutat Res* 776:24-30.
80. Bukowska B & Karwowski BT (2018) Actual state of knowledge in the field of diseases related with defective nucleotide excision repair. *Life Sci*.
81. Oh KS*, et al.* (2006) Phenotypic heterogeneity in the XPB DNA helicase gene (ERCC3): xeroderma pigmentosum without and with Cockayne syndrome. *Hum Mutat* 27(11):1092-1103.
82. Weeda G*, et al.* (1997) A mutation in the XPB/ERCC3 DNA repair transcription gene, associated with trichothiodystrophy. *Am J Hum Genet* 60(2):320-329.
83. Weeda G*, et al.* (1990) A presumed DNA helicase encoded by ERCC-3 is involved in the human repair disorders xeroderma pigmentosum and Cockayne's syndrome. *Cell* 62(4):777-791.
84. Giglia-Mari G*, et al.* (2004) A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A. *Nat Genet* 36(7):714-719.
85. Conaway RC & Conaway JW (1989) An RNA polymerase II transcription factor has an associated DNA-dependent ATPase (dATPase) activity strongly stimulated by the TATA region of promoters. *Proc Natl Acad Sci U S A* 86(19):7356-7360.

86. Weber CA, Salazar EP, Stewart SA, & Thompson LH (1990) ERCC2: cDNA cloning and molecular characterization of a human nucleotide excision repair gene with high homology to yeast RAD3. *EMBO J* 9(5):1437-1447.
87. Weeda G*, et al.* (1990) Molecular cloning and biological characterization of the human excision repair gene ERCC-3. *Mol Cell Biol* 10(6):2570-2581.
88. Schaeffer L*, et al.* (1993) DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. *Science* 260(5104):58-63.
89. Schaeffer L*, et al.* (1994) The ERCC2/DNA repair protein is associated with the class II BTF2/TFIIH transcription factor. *EMBO J* 13(10):2388-2392.
90. Ranish JA*, et al.* (2004) Identification of TFB5, a new component of general transcription and DNA repair factor IIH. *Nat Genet* 36(7):707-713.
91. Chang WH & Kornberg RD (2000) Electron crystal structure of the transcription factor and DNA repair complex, core TFIIH. *Cell* 102(5):609-613.
92. Schultz P*, et al.* (2000) Molecular structure of human TFIIH. *Cell* 102(5):599-607.
93. Gibbons BJ*, et al.* (2012) Subunit architecture of general transcription factor TFIIH. *Proc Natl Acad Sci U S A* 109(6):1949-1954.
94. Luo J*, et al.* (2015) Architecture of the Human and Yeast General Transcription and DNA Repair Factor TFIIH. *Mol Cell* 59(5):794-806.
95. He Y*, et al.* (2016) Near-atomic resolution visualization of human transcription promoter opening. *Nature* 533(7603):359-365.
96. Robinson PJ*, et al.* (2016) Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. *Cell* 166(6):1411-1422 e1416.
97. Greber BJ*, et al.* (2017) The cryo-electron microscopy structure of human transcription factor IIH. *Nature* 549(7672):414-417.
98. Schilbach S*, et al.* (2017) Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature* 551(7679):204-209.
99. Araujo SJ*, et al.* (2000) 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 Dev* 14(3):349-359.
100. Fisher RP (2005) Secrets of a double agent: CDK7 in cell-cycle control and transcription. *J Cell Sci* 118(Pt 22):5171-5180.
101. Houten BV, Kuper J, & Kisker C (2016) Role of XPD in cellular functions: To TFIIH and beyond. *DNA Repair (Amst)* 44:136-142.
102. Fuss JO & Tainer JA (2011) XPB and XPD helicases in TFIIH orchestrate DNA duplex opening and damage verification to coordinate repair with transcription and cell cycle via CAK kinase. *DNA Repair (Amst)* 10(7):697-713.
103. Coin F*, et al.* (1998) Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH. *Nat Genet* 20(2):184-188.
104. Tirode F, Busso D, Coin F, & Egly JM (1999) Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Mol Cell* 3(1):87-95.
105. Coin F, Oksenych V, & Egly JM (2007) Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. *Mol Cell* 26(2):245-256.
106. Grunberg S, Warfield L, & Hahn S (2012) Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nat Struct Mol Biol* 19(8):788-796.

107. Fishburn J, Tomko E, Galburt E, & Hahn S (2015) Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation. *Proc Natl Acad Sci U S A* 112(13):3961-3966.
108. Titov DV*, et al.* (2011) XPB, a subunit of TFIIH, is a target of the natural product triptolide. *Nat Chem Biol* 7(3):182-188.
109. He QL*, et al.* (2015) 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. *Angew Chem Int Ed Engl* 54(6):1859-1863.
110. Manzo SG*, et al.* (2012) Natural product triptolide mediates cancer cell death by triggering CDK7-dependent degradation of RNA polymerase II. *Cancer Res* 72(20):5363-5373.
111. Yi JM*, et al.* (2016) Triptolide Induces Cell Killing in Multidrug-Resistant Tumor Cells via CDK7/RPB1 Rather than XPB or p44. *Mol Cancer Ther* 15(7):1495-1503.
112. Hwang JR*, et al.* (1996) A 3' --> 5' XPB helicase defect in repair/transcription factor TFIIH of xeroderma pigmentosum group B affects both DNA repair and transcription. *J Biol Chem* 271(27):15898-15904.
113. Fan L*, et al.* (2006) Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair. *Mol Cell* 22(1):27-37.
114. Hilario E, Li Y, Nobumori Y, Liu X, & Fan L (2013) Structure of the C-terminal half of human XPB helicase and the impact of the disease-causing mutation XP11BE. *Acta Crystallogr D Biol Crystallogr* 69(Pt 2):237-246.
115. Fan L & DuPrez KT (2015) XPB: An unconventional SF2 DNA helicase. *Prog Biophys Mol Biol* 117(2-3):174-181.
116. Coin F*, et al.* (2004) Phosphorylation of XPB helicase regulates TFIIH nucleotide excision repair activity. *EMBO J* 23(24):4835-4846.
117. Lin YC, Choi WS, & Gralla JD (2005) TFIIH XPB mutants suggest a unified bacterial-like mechanism for promoter opening but not escape. *Nat Struct Mol Biol* 12(7):603-607.
118. He Y, Fang J, Taatjes DJ, & Nogales E (2013) Structural visualization of key steps in human transcription initiation. *Nature* 495(7442):481-486.
119. Murakami K*, et al.* (2013) Architecture of an RNA polymerase II transcription pre-initiation complex. *Science* 342(6159):1238724.
120. Murakami K*, et al.* (2015) Structure of an RNA polymerase II preinitiation complex. *Proc Natl Acad Sci U S A* 112(44):13543-13548.
121. Nogales E, Louder RK, & He Y (2017) Structural Insights into the Eukaryotic Transcription Initiation Machinery. *Annu Rev Biophys* 46:59-83.
122. Iyer N, Reagan MS, Wu KJ, Canagarajah B, & Friedberg EC (1996) Interactions involving the human RNA polymerase II transcription/nucleotide excision repair complex TFIIH, the nucleotide excision repair protein XPG, and Cockayne syndrome group B (CSB) protein. *Biochemistry* 35(7):2157-2167.
123. Jawhari A*, et al.* (2002) p52 Mediates XPB function within the transcription/repair factor TFIIH. *J Biol Chem* 277(35):31761-31767.
124. Coin F*, et al.* (2006) p8/TTD-A as a repair-specific TFIIH subunit. *Mol Cell* 21(2):215-226.
125. Zhou Y, Kou H, & Wang Z (2007) Tfb5 interacts with Tfb2 and facilitates nucleotide excision repair in yeast. *Nucleic Acids Res* 35(3):861-871.
126. Kainov DE, Vitorino M, Cavarelli J, Poterszman A, & Egly JM (2008) Structural basis for group A trichothiodystrophy. *Nat Struct Mol Biol* 15(9):980-984.
127. Kainov DE, Selth LA, Svejstrup JQ, Egly JM, & Poterzsman A (2010) Interacting partners of the Tfb2 subunit from yeast TFIIH. *DNA Repair (Amst)* 9(1):33-39.

128. Vitorino M*, et al.* (2007) Solution structure and self-association properties of the p8 TFIIH subunit responsible for trichothiodystrophy. *J Mol Biol* 368(2):473-480.
129. Vermeulen W*, et al.* (2000) Sublimiting concentration of TFIIH transcription/DNA repair factor causes TTD-A trichothiodystrophy disorder. *Nat Genet* 26(3):307-313.
130. Giglia-Mari G*, et al.* (2006) Dynamic interaction of TTDA with TFIIH is stabilized by nucleotide excision repair in living cells. *PLoS Biol* 4(6):e156.
131. Fregoso M*, et al.* (2007) DNA repair and transcriptional deficiencies caused by mutations in the Drosophila p52 subunit of TFIIH generate developmental defects and chromosome fragility. *Mol Cell Biol* 27(10):3640-3650.
132. Smurnyy Y*, et al.* (2014) DNA sequencing and CRISPR-Cas9 gene editing for target validation in mammalian cells. *Nat Chem Biol* 10(8):623-625.
133. Seroz T, Perez C, Bergmann E, Bradsher J, & Egly JM (2000) p44/SSL1, the regulatory subunit of the XPD/RAD3 helicase, plays a crucial role in the transcriptional activity of TFIIH. *J Biol Chem* 275(43):33260-33266.
134. Fribourg S*, et al.* (2000) Structural characterization of the cysteine-rich domain of TFIIH p44 subunit. *J Biol Chem* 275(41):31963-31971.
135. Kellenberger E*, et al.* (2005) Solution structure of the C-terminal domain of TFIIH P44 subunit reveals a novel type of C4C4 ring domain involved in protein-protein interactions. *J Biol Chem* 280(21):20785-20792.
136. Dubaele S*, et al.* (2003) Basal transcription defect discriminates between xeroderma pigmentosum and trichothiodystrophy in XPD patients. *Mol Cell* 11(6):1635-1646.
137. Kim JS*, et al.* (2015) Crystal structure of the Rad3/XPD regulatory domain of Ssl1/p44. *J Biol Chem* 290(13):8321-8330.
138. Tremeau-Bravard A, Perez C, & Egly JM (2001) A role of the C-terminal part of p44 in the promoter escape activity of transcription factor IIH. *J Biol Chem* 276(29):27693-27697.
139. Gervais V*, et al.* (2004) TFIIH contains a PH domain involved in DNA nucleotide excision repair. *Nat Struct Mol Biol* 11(7):616-622.
140. Okuda M, Kinoshita M, Kakumu E, Sugasawa K, & Nishimura Y (2015) Structural Insight into the Mechanism of TFIIH Recognition by the Acidic String of the Nucleotide Excision Repair Factor XPC. *Structure* 23(10):1827-1837.
141. Okuda M*, et al.* (2016) Dynamics of the Extended String-Like Interaction of TFIIE with the p62 Subunit of TFIIH. *Biophys J* 111(5):950-962.
142. Fribourg S*, et al.* (2001) Dissecting the interaction network of multiprotein complexes by pairwise coexpression of subunits in E. coli. *J Mol Biol* 306(2):363-373.
143. Schmitt DR, Kuper J, Elias A, & Kisker C (2014) The structure of the TFIIH p34 subunit reveals a von Willebrand factor A like fold. *PLoS One* 9(7):e102389.
144. Takagi Y*, et al.* (2005) Ubiquitin ligase activity of TFIIH and the transcriptional response to DNA damage. *Mol Cell* 18(2):237-243.
145. Maita N*, et al.* (2003) Crystal structure of von Willebrand factor A1 domain complexed with snake venom, bitiscetin: insight into glycoprotein Ibalpha binding mechanism induced by snake venom proteins. *J Biol Chem* 278(39):37777-37781.
146. Amlacher S*, et al.* (2011) Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile. *Cell* 146(2):277-289.
147. Bock T*, et al.* (2014) An integrated approach for genome annotation of the eukaryotic thermophile Chaetomium thermophilum. *Nucleic Acids Res* 42(22):13525-13533.
148. Schoenwetter E (2014) Structural Characterization of the p34/p44 and p52/p8 Subcomplexes of TFIIH from Chaetomium thermophilum. Master Thesis (Julius-Maximilians-University of Wuerzburg, Wuerzburg).

149. Ericsson UB, Hallberg BM, Detitta GT, Dekker N, & Nordlund P (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem* 357(2):289-298.
150. Evans PR & Murshudov GN (2013) How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* 69(Pt 7):1204-1214.
151. Altschul SF, Gish W, Miller W, Myers EW, & Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403-410.
152. Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* 62(Pt 9):1002-1011.
153. Winn MD*, et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):235-242.
154. Potterton E, Briggs P, Turkenburg M, & Dodson E (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr* 59(Pt 7):1131-1137.
155. Sievers F*, et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539.
156. Larkin MA*, et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21):2947-2948.
157. Papadopoulos JS & Agarwala R (2007) COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* 23(9):1073-1079.
158. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486-501.
159. Robert X & Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42(Web Server issue):W320-324.
160. Artimo P*, et al.* (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res* 40(Web Server issue):W597-603.
161. Holm L & Laakso LM (2016) Dali server update. *Nucleic Acids Res* 44(W1):W351-355.
162. Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9:40.
163. Chen VB*, et al.* (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66(Pt 1):12-21.
164. Gabadinho J*, et al.* (2010) MxCuBE: a synchrotron beamline control environment customized for macromolecular crystallography experiments. *J Synchrotron Radiat* 17(5):700-707.
165. Krissinel E & Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2256-2268.
166. Laskowski RA (2001) PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res* 29(1):221-222.
167. McCoy AJ*, et al.* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4):658-674.
168. Adams PD*, et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213-221.
169. Kelley LA, Mezulis S, Yates CM, Wass MN, & Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10(6):845-858.
170. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62(Pt 1):72-82.
171. Kallberg M*, et al.* (2012) Template-based protein structure modeling using the RaptorX web server. *Nat Protoc* 7(8):1511-1522.
172. Murshudov GN, Vagin AA, & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53(Pt 3):240-255.

173. Stothard P (2000) The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28(6):1102, 1104.
174. Bricogne G, Vonrhein C, Flensburg C, Schiltz M, & Paciorek W (2003) Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr D Biol Crystallogr* 59(Pt 11):2023-2030.
175. Kabsch W (2010) Xds. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):125-132.
176. Berman HM*, et al.* (2000) The Protein Data Bank. *Nucleic Acids Res* 28(1):235-242.
177. Li MZ & Elledge SJ (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat Methods* 4(3):251-256.
178. Wang W & Malcolm BA (1999) Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange Site-Directed Mutagenesis. *Biotechniques* 26(4):680-682.
179. Miyatake H, Hasegawa T, & Yamano A (2006) New methods to prepare iodinated derivatives by vaporizing iodine labelling (VIL) and hydrogen peroxide VIL (HYPER-VIL). *Acta Crystallogr D Biol Crystallogr* 62(Pt 3):280-289.
180. Schmitt D (2014) Structural Characterization of TFIIH Subunits p34 and p44 from C. thermophilum. PhD Thesis (Julius-Maximilians-University of Wuerzburg, Wuerzburg).
181. Radu L*, et al.* (2017) The intricate network between the p34 and p44 subunits is central to the activity of the transcription/DNA repair factor TFIIH. *Nucleic Acids Res* 45(18):10872-10883.
182. Koelmel W (2017) Structural and Functional Characterization of TFIIH from Chaetomium thermophilum. PhD Thesis (Julius-Maximilians-Universityof Wuerzburg, Wuerzburg).
183. Peißert S (2016) Deciphering the interaction of the TFIIIH subunits p34 and p44 from Chaetomium thermophilum. Master Thesis (Julius-Maximilians_University of Wuerzburg, Wuerzburg).
184. Compe E & Egly JM (2016) Nucleotide Excision Repair and Transcriptional Regulation: TFIIH and Beyond. *Annu Rev Biochem* 85:265-290.
185. Okuda M & Nishimura Y (2015) Real-time and simultaneous monitoring of the phosphorylation and enhanced interaction of p53 and XPC acidic domains with the TFIIH p62 subunit. *Oncogenesis* 4:e150.

VI Appendix

VI.1 List of Abbreviations

(6-4)PP	(6-4) pyrimidine photoproducts
8-oxo-dG	8-oxo-2'-deoxyguanosine
A	Adenine
A_{280}	Absorbance at 280 nm
A. fulgidus	*Archaeoglobus fulgidus*
AGT	O^6-alkylguanine-DNA alkyltransferase
ALKBH	Alkylation repair homolog
APS	Ammonium persulfate
A-SEC	Analytical size exclusion chromatography
ATP	Adenosine triphosphate
B[a]P	Benzo[a]pyrene
BER	Base excision repair
BESSY	Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung
BLAST	Basic Local Alignment Search Tool
BRCA1/2	Breast cancer 1/2
BS3	Bis(sulfosuccinimidyl)suberate
BSA	Bovine Serum Albumin
C	Cytosine
CAK	CDK-activating kinase
CD	Circular dichroism
CDK	Cyclin-dependent kinase
CENT2	Centrin 2
CHES	*N*-Cyclohexyl-2-aminoethanesulfonic acid
Cisplatin	*cis*-Diaminodichloroplatinum
CPD	Cyclobutane pyrimidine dimer
CpG	5'-cytosine-phosphate-guanine-3'
CRL	cullin 4A – Roc1 E3 ubiquitin ligase
CS	Cockayne syndrome
ct	*Chaetomium thermophilum*
CTD	C-terminal domain
Cterm	C-terminal
C. thermophilum	*Chaetomium thermophilum*
CV	Column volume
dATP	2'-Deoxyadenosine 5'-triphosphate
dCTP	2'-Deoxycytidine 5'-triphosphate
DESY	Deutsches Elektronen-Synchrotron
dGTP	2'-Deoxyguanosine 5'-triphosphate

DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DRD	Damage recognition domain
DSB	Double-strand break
DTT	Dithiothreitol
dTTP	2'-Deoxythymidine 5'-triphosphate
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
EMBL	European Molecular Biology Laboratory
ERCC1	Excision repair cross complementing 1
ESRF	European Synchrotron Radiation Facility
FA	Fanconi Anemia
FeS	Iron-sulfur
FPLC	Fast protein liquid chromatography
FTO	Fat mass and obesity associated
G	Guanine
HD1/2	Helicase domain 1/2
HDX-MS	Hydrogen-Deuterium eXchange coupled to Mass Spectrometry
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His_6-tag	Hexahistidine-tag
HJ	Homologous end-joining
HMGN1	High mobility group nucleosome binding protein domain 1
hs	*Homo sapiens*
ICL	Interstrand crosslink
IEC	Ion exchange chromatography
IMAC	Immobilized metal ion affinity chromatography
IPTG	Isopropyl-β-D-thiogalactopyranoside
ITC	Isothermal titration calorimetry
K_D	Dissociation constant
LB	Lysogeny broth
MALS	Multi-angle light scattering
MAT1	Ménage à trois 1
MC	Minimal complex
MD1/2	Middle domains 1/2
MES	2-(*N*-morpholino)-ethanesulfonic acid
MMR	Mismatch repair
N	Nitrogen
NCBI	National Center for Biotechnology Information
NER	Nucleotide excision repair
NHEJ	Non-homologous end-joining
Ni-TED	Ni^{2+} ions immobilized on silica beads via the chelating ligand tris(carboxymethyl)-ethylene diamine
NMR	Nuclear magnetic resonance
NTD	N-terminal domain
Nterm	N-terminal

O	Oxygen
OD_{600}	Optical density at 600 nm
PAGE	Polyacrylamide gel electrophoresis
PAH	Polycyclic aromatic hydrocarbons
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PMSF	Phenylmethylsulfonyl fluoride
Pol	Polymerase
RFC	Replication factor C
RING	Really interesting new gene
RMSD	Root mean square deviation
Roc1	Regulator of cullins 1
RPA	Replication protein A
RT	Room temperature
SAD	Single-wavelength anomalous diffraction
sc	*Saccharomyces cerevisiae*
S. cerevisiae	*Saccharomyces cerevisiae*
SD	Superdex™
SdM	Site-directed mutagenesis
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
SF2	Superfamily 2
SLIC	Sequence and ligation independent cloning
SR	Superose™
ssDNA	Single-stranded DNA
T	Thymine
TCEP	Tris[2-carboxyethyl]-phosphine
TEMED	Tetramethylethylenediamin
TEV	Tobacco Etch Virus
TFIIH	Transcription factor II H
ThM	Thumb-like motif
TLS	Translesion synthesis
TMAO	Trimethylamine *N*-oxide
TRIS	Tris[hydroxymethyl]-aminomethane
tRNA	transfer Ribonucleic acid
TTD	Trichthiodystrophy
TTDN1	TTD nonphotosensitive 1
Usp7	Ubiquitin specific peptidase 7
UV	Ultraviolet
UV-DDB	UV-DNA damage binding protein
UVSSA	UV-stimulated scaffold protein A
vWA	*von Willebrand factor A* like
WT	Wild-type
XAB2	XPA binding protein 2
XP	Xeroderma pigmentosum

Abbreviations of amino acids

Alanine	**Ala**	**A**
Arginine	**Arg**	**R**
Asparagine	**Asn**	**N**
Aspartic acid	**Asp**	**D**
Cysteine	**Cys**	**C**
Glutamic acid	**Glu**	**E**
Glutamine	**Gln**	**Q**
Glycine	**Gly**	**G**
Histidine	**His**	**H**
Isoleucine	**Ile**	**I**
Leucine	**Leu**	**L**
Lysine	**Lys**	**K**
Methionine	**Met**	**M**
Seleno methionine	**Se-Met**	**-**
Phenylalanine	**Phe**	**F**
Proline	**Pro**	**P**
Serine	**Ser**	**S**
Threonine	**Thr**	**T**
Tryptophan	**Trp**	**W**
Tyrosine	**Tyr**	**Y**
Valine	**Val**	**V**

VI.2 Constructs

Construct	**Cloning strategy or reference**	**Plasmid**	**Affinity tag**	**Molecular weight (kDa)***
p34	see II.1.3	pBADM-11	Nterm-His_6	46.8
p34_K155E	SdM strategy 1	pBADM-11	Nterm-His_6	46.8
p34_A151E	SdM strategy 1	pBADM-11	Nterm-His_6	46.8
p34_vWA	see II.1.3	pBADM-11	Nterm-His_6	31.8
p34_vWA_A151E	SdM strategy 1	pBADM-11	Nterm-His_6	31.8
p34_18-277	SLIC alternative	pBADM-11	Nterm-His_6	30.0
p34_18-277_noTag	SLIC alternative	pBADM-11	–	27.0
p44	see II.1.3	pETM-11	Nterm-His_6	61.0
p44	see II.1.3	pBADM-11	Nterm-His_6	61.0
p44_RING	see II.1.3	pETM-11	Nterm-His_6	20.9
p44_RING_F490E	see II.1.3	pETM-11	Nterm-His_6	20.9
p44_RING_noTag	see II.1.3	pETM-11	–	17.9
p62	see II.1.3	pETM-11	Nterm-His_6	77.4
p62_noTag	see II.1.3	pETM-11	–	74.4

Construct	Cloning strategy or reference	Plasmid	Affinity tag	Molecular weight (kDa)*
p8	see II.1.3	pBADM-11	Nterm-His_6	11.4
p52	see II.1.3	pETM-11	Nterm-His_6	59.8
p52_noTag	see II.1.3	pETM-11	–	56.8
p52_E359K_noTag	SdM strategy 2	pETM-11	–	56.8
p52_R363E_noTag	SdM strategy 2	pETM-11	–	56.8
p52_R386L_noTag	SdM strategy 1	pETM-11	–	56.8
p52_1-321	SdM strategy 1	pETM-11	Nterm-His_6	38.3
p52_1-321_dL2	SLIC alternative	pETM-11	Nterm-His_6	37.5
p52_1-349	SLIC alternative	pETM-11	Nterm-His_6	40.9
p52_1-349_noTag	SLIC alternative	pETM-11	–	37.9
p52_1-454	SLIC alternative	pETM-11	Nterm-His_6	52.6
p52_121-End	SLIC alternative	pETM-11	Nterm-His_6	46.2
p52_121-End_noTag	SLIC alternative	pETM-11	–	43.2
p52_121-EdL	see II.1.3	pETM-11	Nterm-His_6	44.4
p52_121-EdL_noTag	SLIC alternative	pETM-11	–	41.4
p52_121-EdL_CtermHis	SLIC alternative	pETM-11	Cterm-His_6	44.2
p52_121-349	SLIC alternative	pETM-11	Nterm-His_6	27.3
p52_121-454_dL	SLIC alternative	pETM-11	Nterm-His_6	37.2
p52_121-454_dL_noTag	SLIC alternative	pETM-11	–	34.2
p52_345-421	SLIC alternative	pETM-11	Nterm-His_6	11.5
p52_345-458	SLIC alternative	pETM-11	Nterm-His_6	15.7
p52_345-458_noTag	SLIC alternative	pETM-11	–	12.7
p52_345-458_CtermHis	SLIC alternative	pETM-11	Cterm-His_6	15.5
p52_345-End	SLIC alternative	pETM-11	Nterm-His_6	22.4
p52_345-End_noTag	SLIC alternative	pETM-11	–	19.4
p52_350-435	see II.1.3	pETM-11	Nterm-His_6	12.6
p52_350-435_noTag	see II.1.3	pETM-11	–	9.6
p52_350-End	see II.1.3	pETM-11	Nterm-His_6	22.0
XPB_1-345	see II.1.3	pBADM-11	Nterm-His_6	41.5
XPB_60-249	SLIC alternative	pBADM-11	Nterm-His_6	25.1
XPB_60-270	SLIC alternative	pBADM-11	Nterm-His_6	27.1
XPB_60-345	see II.1.3	pBADM-11	Nterm-His_6	35.4
XPB_116-245	see II.1.3	pBADM-11	Nterm-His_6	18.0
XPB_116-345	see II.1.3	pBADM-11	Nterm-His_6	28.7

*according to the ExPASy ProtParam tool

VI.3 ThermoFluor Screens

ThermoFluor Standard

Tray Location	Composition
A1	0.1 M citric acid ph 4.5
A2	0.1 M bis-tris ph 7.0
A3	0.1 M imidazole ph 6.5
A4	0.1 M hepes ph 8.0
A5	0.1 M tris ph 8.5
B1	0.1 M acetat ph 4.6
B2	0.1 M ada ph 6.5
B3	0.1 M imidazole ph 8.0
B4	0.1 M hepes ph 8.5
B5	0.1 M tris ph 9.0
C1	0.1 M mes ph 5.5
C2	0.1 M ada ph 7.0
C3	0.1 M sodium potassium phosphate ph 6.8
C4	0.1 M bicine ph 8.0
C5	0.1 M taps ph 8.0
D1	0.1 M mes ph 6.5
D2	0.1 M bis-tris propane ph 6.0
D3	0.1 M sodium potassium phosphate ph 7.5
D4	0.1 M bicine ph 8.5
D5	0.1 M taps ph 9.0
E1	0.1 M sodium cacodylate ph 6.0
E2	0.1 M bis-tris propane ph 7.0
E3	0.1 M mops ph 7.0
E4	0.1 M bicine ph 9.0
E5	0.1 M glycyl-glycine ph 8.5
F1	0.1 M sodium cacodylate ph 6.5
F2	0.1 M pipes ph 6.5
F3	0.1 M mops ph 7.5
F4	0.1 M tris ph 7.0
F5	0.1 M ches ph 9.0
G1	0.1 M bis-tris ph 5.5
G2	0.1 M pipes ph 7.0
G3	0.1 M hepes ph 7.0
G4	0.1 M tris ph 7.5
G5	0.1 M ches ph 9.5
H1	0.1 M bis-tris ph 6.5
H2	0.1 M pipes ph 7.5
H3	0.1 M hepes ph 7.5
H4	0.1 M tris ph 8.0
H5	0.1 M caps ph 9.8

ThermoFluor Advanced

Tray Location	Composition			
A1	-			
A2	0.1 M	citric acid ph 4.0		
A3	0.1 M	sodium acetate ph 4.5		
A4	0.1 M	citric acid ph 5.0		
A5	0.1 M	mes ph 6.0		
A6	0.1 M	potassium phosphate ph 6.0		
A7	0.1 M	citric acid ph 6.0		
A8	0.1 M	bis-tris ph 6.5		
A9	0.1 M	sodium cacodylate ph 6.5		
A10	0.1 M	sodium phosphate ph 7.0		
A11	0.1 M	potassium phosphate ph 7.0		
A12	0.1 M	hepes ph 7.0		
B1	0.1 M	mops ph 7.0		
B2	0.1 M	ammonium acetate ph 7.3		
B3	0.1 M	tris ph 7.5		
B4	0.1 M	sodium phosphate ph 7.5		
B5	0.1 M	imidazole ph 8.0		
B6	0.1 M	hepes ph 8.0		
B7	0.1 M	tris ph 8.0		
B8	0.1 M	tricine ph 8.0		
B9	0.1 M	bicine ph 8.0		
B10	0.1 M	bicine ph 8.5		
B11	0.1 M	tris ph 8.5		
B12	0.1 M	bicine ph 9.0		
C1	0.15 M	sodium chloride		
C2	0.15 M	sodium chloride	0.1 M	citric acid ph 4.0
C3	0.15 M	sodium chloride	0.1 M	sodium acetate ph 4.5
C4	0.15 M	sodium chloride	0.1 M	citric acid ph 5.0
C5	0.15 M	sodium chloride	0.1 M	mes ph 6.0
C6	0.15 M	sodium chloride	0.1 M	potassium phosphate ph 6.0
C7	0.15 M	sodium chloride	0.1 M	citric acid ph 6.0
C8	0.15 M	sodium chloride	0.1 M	bis-tris ph 6.5
C9	0.15 M	sodium chloride	0.1 M	sodium cacodylate ph 6.5
C10	0.15 M	sodium chloride	0.1 M	sodium phosphate ph 7.0
C11	0.15 M	sodium chloride	0.1 M	potassium phosphate ph 7.0
C12	0.15 M	sodium chloride	0.1 M	hepes ph 7.0
D1	0.15 M	sodium chloride	0.1 M	mops ph 7.0
D2	0.15 M	sodium chloride	0.1 M	ammonium acetate ph 7.5
D3	0.15 M	sodium chloride	0.1 M	tris ph 7.5
D4	0.15 M	sodium chloride	0.1 M	sodium phosphate ph 7.5
D5	0.15 M	sodium chloride	0.1 M	imidazole ph 8.0
D6	0.15 M	sodium chloride	0.1 M	hepes ph 8.0
D7	0.15 M	sodium chloride	0.1 M	tris ph 8.0
D8	0.15 M	sodium chloride	0.1 M	tricine ph 8.0
D9	0.15 M	sodium chloride	0.1 M	bicine ph 8.0
D10	0.15 M	sodium chloride	0.1 M	bicine ph 8.5
D11	0.15 M	sodium chloride	0.1 M	tris ph 8.5

D12	0.15 M	sodium chloride	0.1 M	bicine ph 9.0
E1	0.25 M	sodium chloride		
E2	0.25 M	sodium chloride	0.1 M	citric acid ph 4.0
E3	0.25 M	sodium chloride	0.1 M	sodium acetate ph 4.5
E4	0.25 M	sodium chloride	0.1 M	citric acid ph 5.0
E5	0.25 M	sodium chloride	0.1 M	mes ph 6.0
E6	0.25 M	sodium chloride	0.1 M	potassium phosphate ph 6.0
E7	0.25 M	sodium chloride	0.1 M	citric acid ph 6.0
E8	0.25 M	sodium chloride	0.1 M	bis-tris ph 6.5
E9	0.25 M	sodium chloride	0.1 M	sodium cacodylate ph 6.5
E10	0.25 M	sodium chloride	0.1 M	sodium phosphate ph 7.0
E11	0.25 M	sodium chloride	0.1 M	potassium phosphate ph 7.0
E12	0.25 M	sodium chloride	0.1 M	hepes ph 7.0
F1	0.25 M	sodium chloride	0.1 M	mops ph 7.0
F2	0.25 M	sodium chloride	0.1 M	ammonium acetate ph 7.3
F3	0.25 M	sodium chloride	0.1 M	tris ph 7.5
F4	0.25 M	sodium chloride	0.1 M	sodium phosphate ph 7.5
F5	0.25 M	sodium chloride	0.1 M	imidazole ph 8.0
F6	0.25 M	sodium chloride	0.1 M	hepes ph 8.0
F7	0.25 M	sodium chloride	0.1 M	tris ph 8.0
F8	0.25 M	sodium chloride	0.1 M	tricine ph 8.0
F9	0.25 M	sodium chloride	0.1 M	bicine ph 8.0
F10	0.25 M	sodium chloride	0.1 M	bicine ph 8.5
F11	0.25 M	sodium chloride	0.1 M	tris ph 8.5
F12	0.25 M	sodium chloride	0.1 M	bicine ph 9.0
G1			0.02 M	hepes ph 7.5
G2			0.05 M	hepes ph 7.5
G3			0.1 M	hepes ph 7.5
G4			0.25 M	hepes ph 7.5
G5			0.02 M	sodium phosphate ph 7.5
G6			0.05 M	sodium phosphate ph 7.5
G7			0.1 M	sodium phosphate ph 7.5
G8			0.2 M	sodium phosphate ph 7.5
G9			0.02 M	tris ph 8.0
G10			0.05 M	tris ph 8.0
G11			0.1 M	tris ph 8.0
G12			0.25 M	tris ph 8.0
H1	0.05 M	sodium chloride	0.05 M	hepes ph 7.5
H2	0.13 M	sodium chloride	0.05 M	hepes ph 7.5
H3	0.25 M	sodium chloride	0.05 M	hepes ph 7.5
H4	0.5 M	sodium chloride	0.05 M	hepes ph 7.5
H5	0.75 M	sodium chloride	0.05 M	hepes ph 7.5
H6	1 M	sodium chloride	0.05 M	hepes ph 7.5
H7	0.05 M	sodium chloride	0.05 M	tris ph 8.
H8	0.13 M	sodium chloride	0.05 M	tris ph 8.0
H9	0.25 M	sodium chloride	0.05 M	tris ph 8.0
H10	0.5 M	sodium chloride	0.05 M	tris ph 8.0
H11	0.75 M	sodium chloride	0.05 M	tris ph 8.0
H12	1 M	sodium chloride	0.05 M	tris ph 8.0

VI.4 Supplementary Figures

Data collection and refinement statistics of p34/p44 MC I

Table VI.1: Data collection and refinement statistics of the new model of p34/p44 MC I.

Data collection	**New model p34/p44 MC I**
Space Group	P 6_3 2 2
Cell Dimensions	
a, b, c [Å]	138.21, 138.21, 95.24
α, β, γ [°]	90, 90, 120
Resolution [Å]	45.24 – 3.70 (4.05 – 3.70)
Wavelength [Å]	0.8726
Unique Reflections	6,120 (1,418)
I/σI	9.5 (1.4)
CC (1/2)	0.998 (0.601)
Rmerge [%]	20.0 (142.5)
Completeness [%]	99.8 (100.0)
Redundancy	8.1 (8.4)
SigAno	-
Refinement	
Resolution [Å]	45.24 – 3.70
Unique Reflections	5,828
Number of Atoms	2,088
R_{work} (R_{free}) [%]	19.6 (24.6)
Mean B-Factor [Å^2]	152.7
Bond Lengths [Å]	0.011
Bond Angles [°]	1.610
Ramachandran Statistics (favored/allowed/outliers)* [%]	93.56/4.92/1.52

Values in parentheses refer to the highest resolution shell. *Ramachandran statistics were performed using MolProbity.

Purification of p34_vWA_A151E

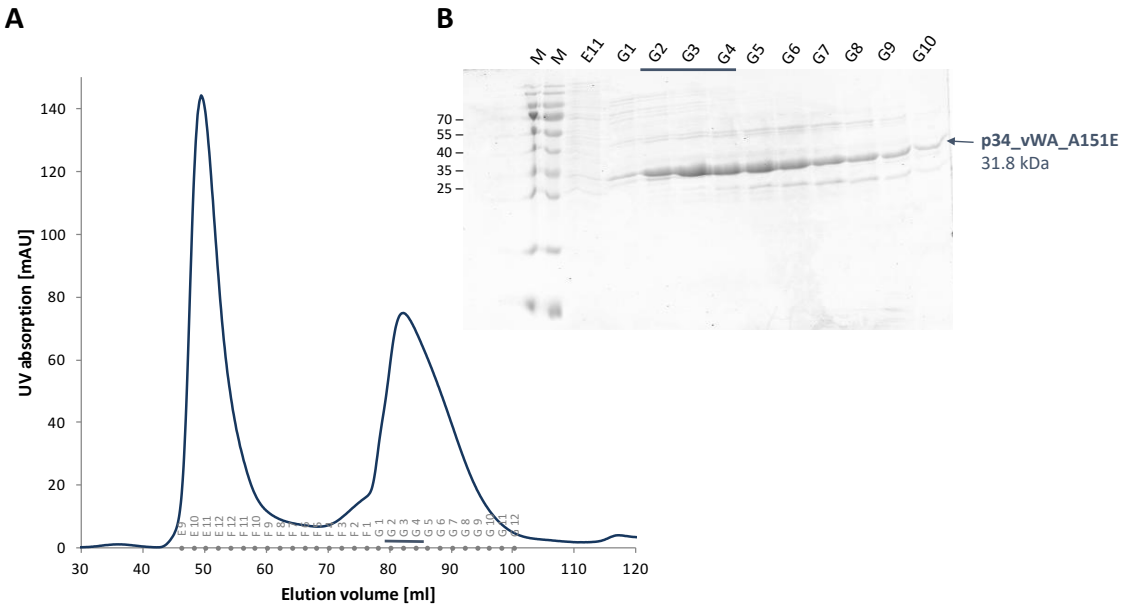


Figure VI.1: Size exclusion chromatography of p34_vWA_A151E. (A) Chromatogram and **(B)** SDS-PAGE of the size exclusion chromatography (SD 200 16/60 column) of p34_vWA_A151E (31.8 kDa). The chromatogram shows the UV absorption at 280 nm. The bar indicates fractions that were finally pooled. Abbreviations: M = marker (kDa); S = supernatant; P = pellet fraction; FT = flow through; E1 – E8 =elution fraction 1 – 8.

Purification of optimized p34_vWA (18-277)

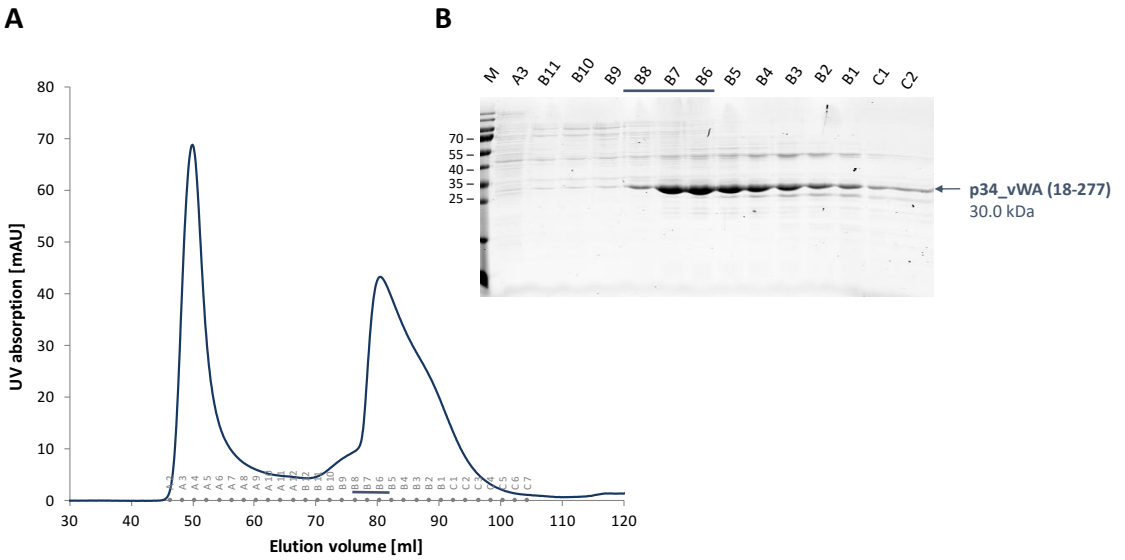


Figure VI.2: Size exclusion chromatography of optimized p34_vWA (18-277). (A) Chromatogram and **(B)** SDS-PAGE of the size exclusion chromatography (SD 200 16/60) of optimized p34_vWA (18-277). The chromatogram shows the UV absorption at 280 nm. The bar indicates fractions that were finally pooled. Abbreviation: M = marker (kDa).

Purification of p34_K155E

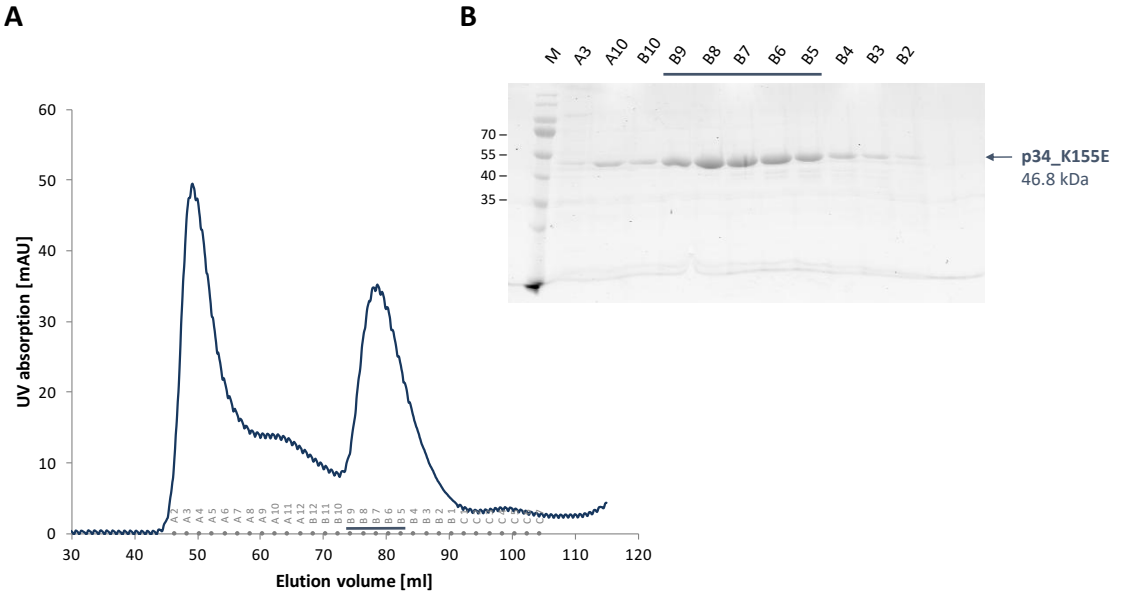


Figure VI.3: Size exclusion chromatography of p34_K155E. (A) Chromatogram and **(B)** SDS-PAGE of the size exclusion chromatography (SD 200 16/60 column) of p34_K155E (46.8 kDa). The chromatogram shows the UV absorption at 280 nm. The bar indicates fractions that were finally pooled. Abbreviation: M = marker (kDa).

Purification of p34_A151E

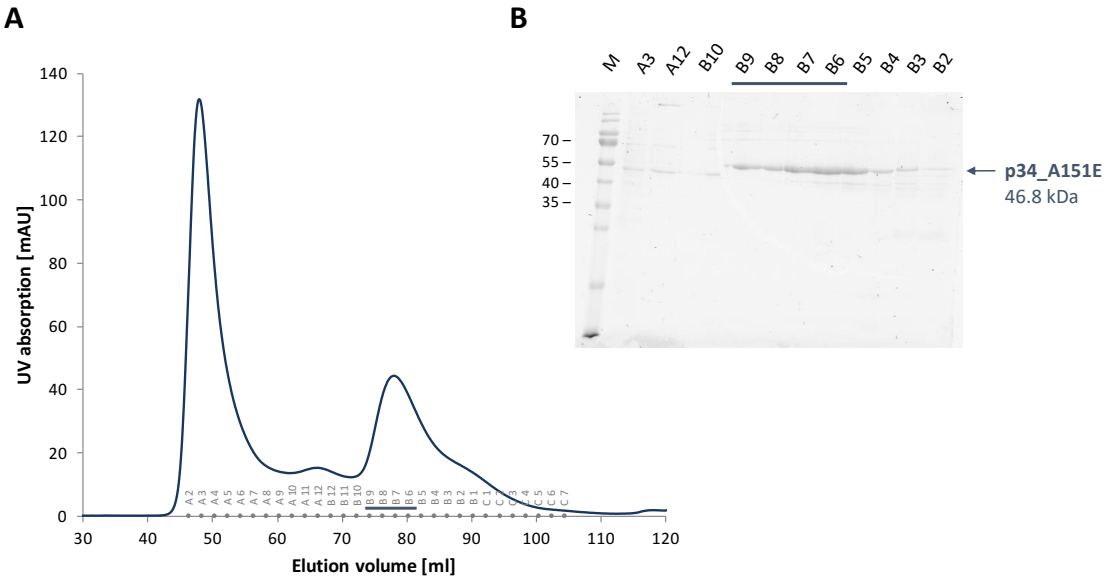


Figure VI.4: Size exclusion chromatography of p34_A151E. (A) Chromatogram and **(B)** SDS-PAGE of the size exclusion chromatography (SD 200 16/60 column) of p34_A151E (46. 8 kDa). The chromatogram shows the UV absorption at 280 nm. The bar indicates fractions that were finally pooled. Abbreviation: M = marker (kDa).

His_6-tag cleavage of p52 variants

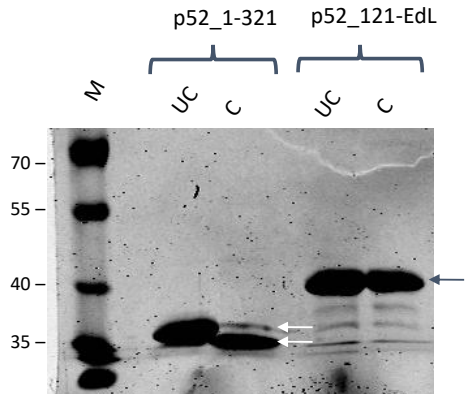


Figure VI.5: His_6-tag cleavage by the TEV protease of p52_1-321 and p52_121-EdL. SDS-PAGE analysis of samples before (UC = uncleaved) and after (C = cleaved) the His_6-tag cleavage by the TEV protease using a 10% SDS-PAGE gel. The His_6-tag was cleaved for most of the p52_1-321 sample as indicated by the band at lower molecular weight, but the His_6-tag of the p52_121-EdL sample could not be cleaved. Abbreviation: M = marker (kDa).

p52 structures

Table VI.2: Summary of residues that could be built per molecule of the asymmetric unit.

Structure	Residues built in molecule A	Residues built in molecule B	Residues built in molecule C	Residues built in molecule D
p52_121-EdL	109 - 120 (TEV helix) 121 - 165 177 - 194 199 - 234 254 - 319 349 - 453	109 - 120 (TEV helix) 121 - 164 177 - 233 254 - 319 350 - 453	109 - 120 (TEV helix) 121 - 161 177 - 232 254 - 319 350 - 453	109 - 120 (TEV helix) 121 - 164 178 - 194 201 - 234 254 - 319 350 - 451
p52_1-321	11 - 69 79 - 164 176 - 238 253 - 319	10 - 71 79 - 165 176 - 240 254 - 319	11 - 67 81 - 95 106 - 164 176 - 238 253 - 318	11 - 71 80 - 94 103 - 165 176 - 235 255 - 318
p52_121-EdL/p8	109 - 120 (TEV helix) 121 - 165 176 - 238 254 - 320 349 - 453	105 - 120 (TEV helix) 121 - 165 176 - 194 200 - 235 253 - 321 350 - 514	-	4 - 69 (p8)

Purification of XPB_60-345

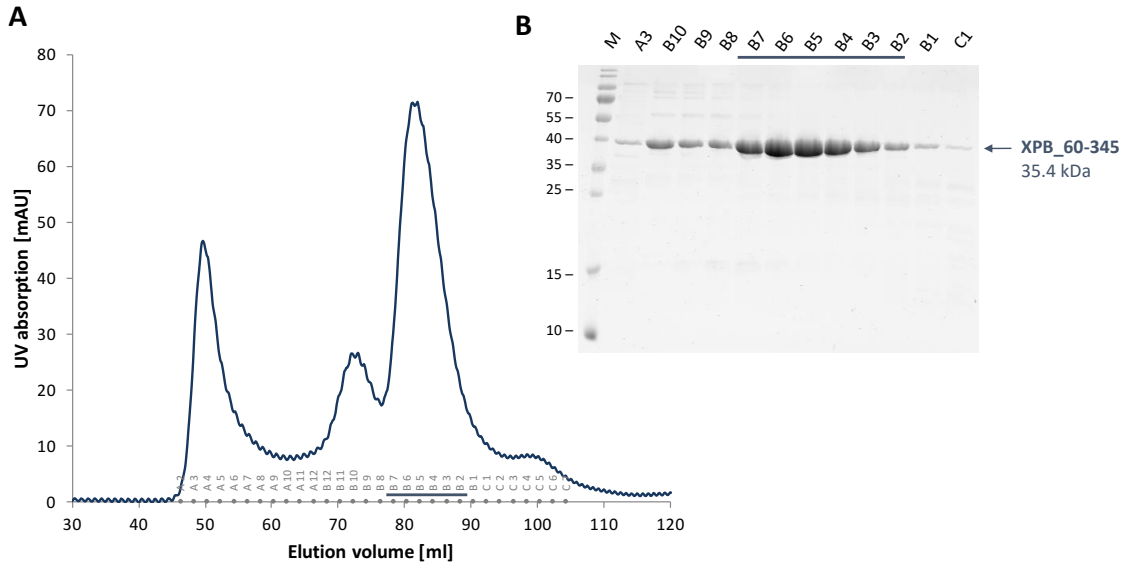


Figure VI.6: Size exclusion chromatography of XPB_60-345. (A) Chromatogram and **(B)** SDS-PAGE of the size exclusion chromatography (SD 200 16/60) of XPB_60-345 (35.4 kDa). The chromatogram shows the UV absorption at 280 nm. The bar indicates fractions that were finally pooled. Abbreviation: M = marker (kDa).

XPB peptide array against p52/p8

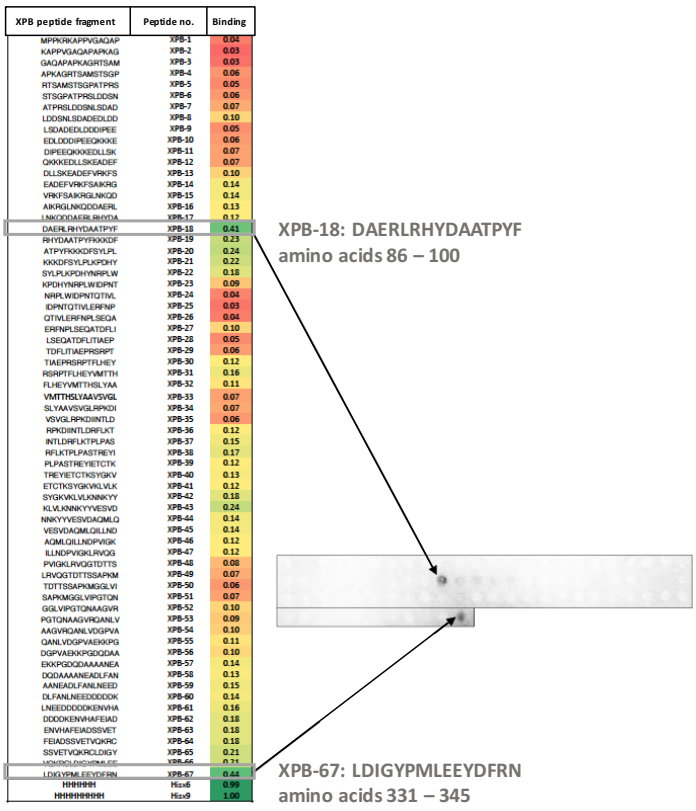


Figure VI.7: Summary of the XPB peptide array against p52/p8. Summary of the results of a XPB peptide array against p52_noTag/p8. All XPB peptides (XPB-1 – XPB-67) as well as the His_6 and His_9 peptides used for normalization of the results are listed. Their signal intensities are marked and color coded and the two peptides XPB-18 and XPB-67 that yielded the highest intensities are marked on the respective peptide array. This figure is based on a summary of the XPB peptide array results, which was kindly provided by Hans Maric.

Analytical SEC of XPB/p52/p8 triple complex

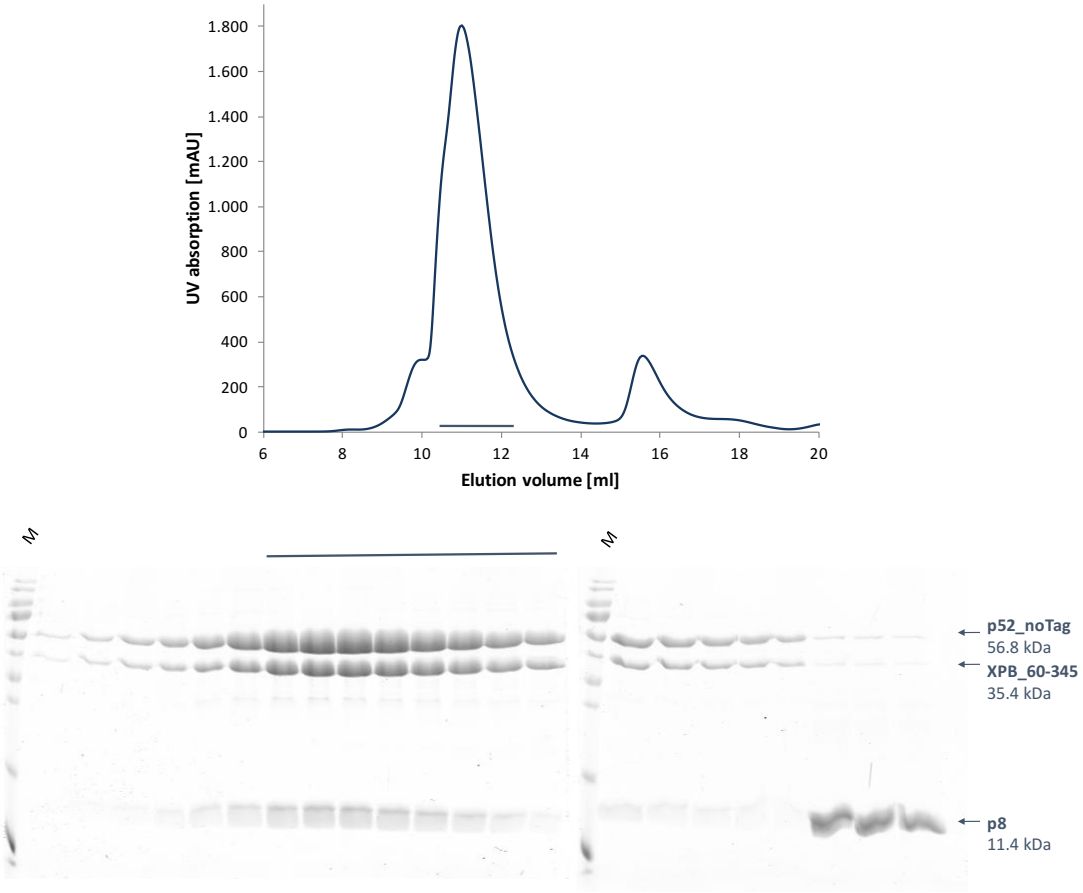


Figure VI.8: Analytical SEC of XPB_60-345/p52 and p8. The XPB_60-345/p52_noTag sample was applied to a SD 200 increase 10/300 column in a 1:1 ratio with p8. The triple complex eluted at around 11 ml, an excess of p8 was present in the second smaller peak at 15.5 ml elution volume. The chromatogram shows the UV absorption at 280 nm. The result was analyzed by SDS-PAGE. The bar indicates the final fractions that were pooled. Abbreviation: M = marker (kDa). The assembly of the ternary complex was performed together with an internship student (Max Meirow).

Purification of XPB/p52 minimal complexes

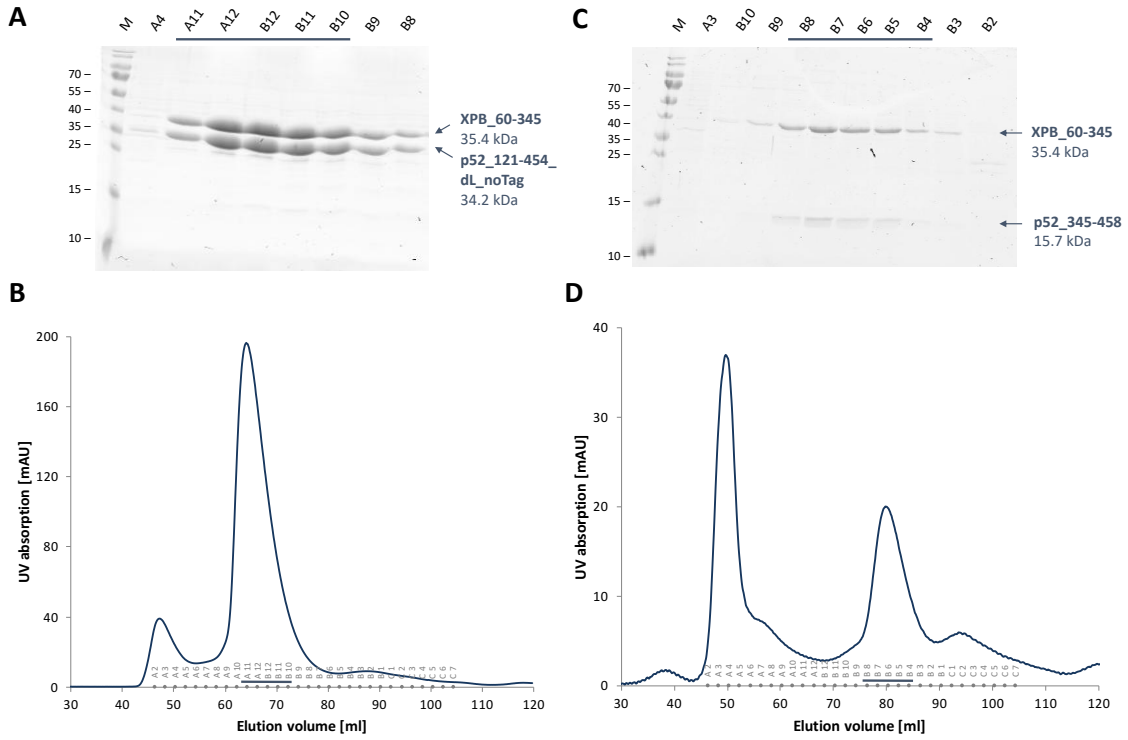


Figure VI.9: Size exclusion chromatography of two different XPB/p52 minimal complexes. (A) SDS-PAGE and **(B)** chromatogram of the size exclusion chromatography (SD 200 16/60) of XPB_60-345 (35.4 kDa) and p52_121-454_dL_noTag (34.5 kDa). **(C)** SDS-PAGE and **(D)** chromatogram of the size exclusion chromatography (SD 200 16/60) of XPB_60-345 (35.4 kDa) and p52_345-458 (15.7 kDa) (performed together with two short-term students, Leonie Müller and Yugendra Raghunadan). The chromatogram shows the UV absorption at 280 nm. The bar indicates fractions that were finally pooled. Abbreviation: M = marker (kDa).

Limited proteolysis of XPB/p52 minimal complex

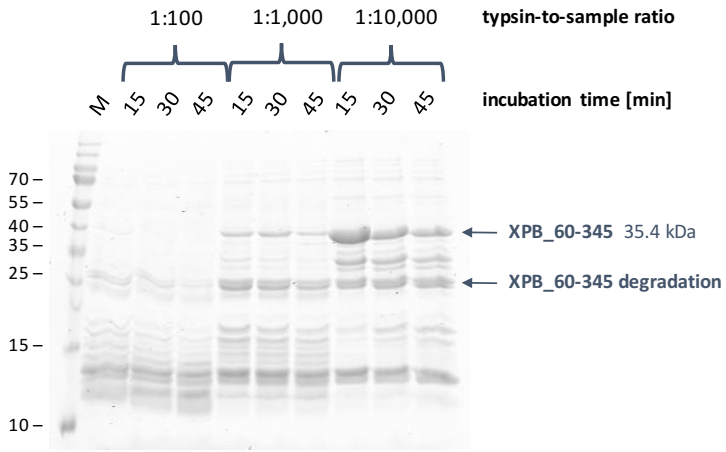


Figure VI.10: Limited proteolysis of XPB_60-345/p52_345-458. The XPB_60-345/p52_345-458 minimal complex was incubated with trypsin in different trypsin-to-sample ratios and for varying incubation times. All reactions resulted in a degradation product of XPB of around 25 kDa. The lower one of the double bands at 25 kDa was cut out and sent to mass spectrometry analysis.

Purification of p52_R386L/p8

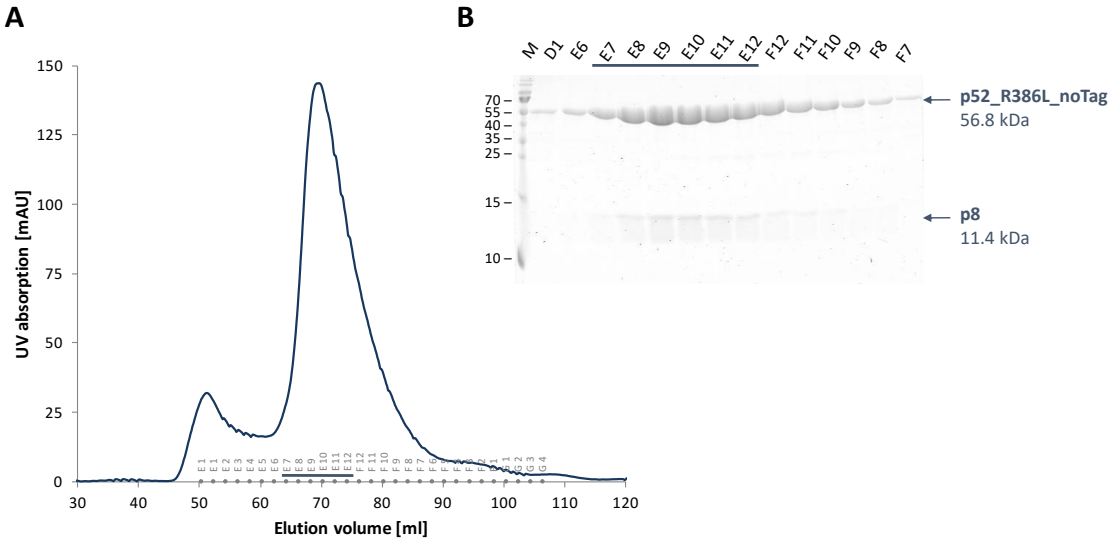


Figure VI.11: Size exclusion chromatography of p52_R386L/p8.

(A) Chromatogram and **(B)** SDS-PAGE of the size exclusion chromatography (SD 200 16/60 column) of p52_R386L_noTag/p8 (56.8 kDa and 11.4 kDa, respectively). The chromatogram shows the UV absorption at 280 nm. The bar indicates fractions that were pooled. Abbreviation: M = marker (kDa).

Purification of p52_R363E/p8

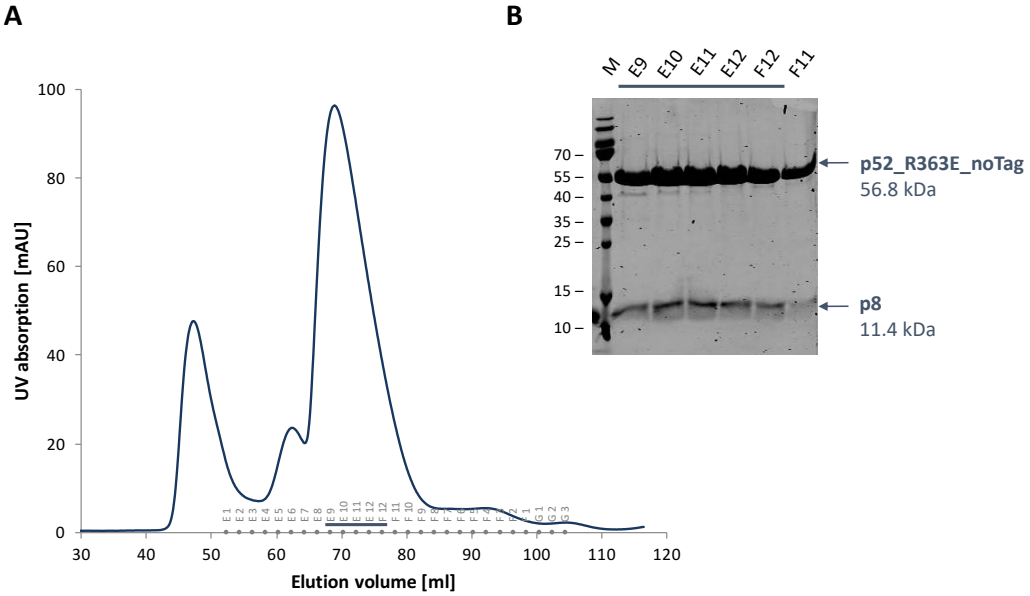


Figure VI.12: Size exclusion chromatography of p52_R363E/p8. (A) Chromatogram and **(B)** SDS-PAGE of the size exclusion chromatography (SD 200 16/60 column) of p52_R363E_noTag/p8 (56.8 kDa and 11.4 kDa, respectively). The chromatogram shows the UV absorption at 280 nm. The bar indicates fractions that were pooled. Abbreviation: M = marker (kDa). The purification was performed in collaboration with Jeannette Kappenberger.

Purification of p52_E359K/p8

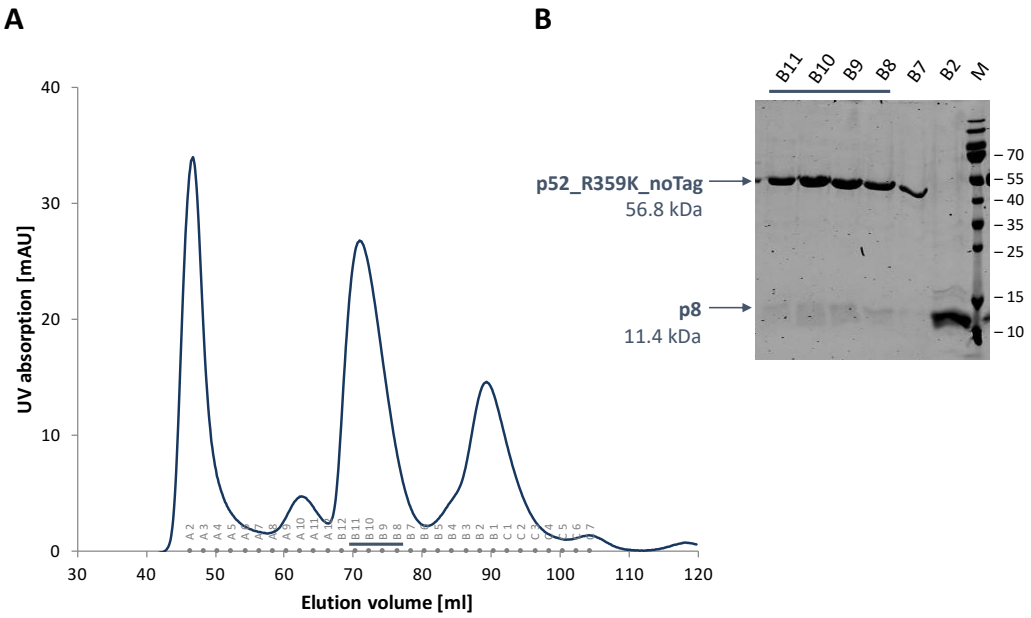


Figure VI.13: Size exclusion chromatography of p52_E359K/p8. (A) Chromatogram and **(B)** SDS-PAGE of the size exclusion chromatography (SD 200 16/60 column) of p52_E359K_noTag/p8 (56.8 kDa and 11.4 kDa, respectively). Interestingly, the elution volume of the p52/p8 peak and the SDS-PAGE indicate that the complex lost parts of p8. The chromatogram shows the UV absorption at 280 nm. The bar indicates fractions that were pooled. Abbreviation: M = marker (kDa). The purification was performed in collaboration with Jeannette Kappenberger.

VII List of Publications

Publication

Radu L*, Schoenwetter E*, Braun C, Marcoux J, Koelmel W, Schmitt DR, Kuper J, Cianférani S, Egly JM, Poterszman A, Kisker C; **The intricate network between the p34 and p44 subunits is central to the activity of the transcription/DNA repair factor TFIIH**, *Nucleic Acids Research,* Volume 45, Issue 18, 13 October 2017, Pages 10872-10883

*these authors contributed equally to the work

Congress Contributions

Poster presentation at the 10th International GSLS Student Symposium "EUREKA! 2015", October 14-15, 2015, Wuerzburg (Germany): *Towards Structural Characterization of the p52 Subunit of TFIIH from C. thermophilum.*

Poster presentation at the 11th International GSLS Student Symposium "EUREKA! 2016", October 12-13, 2016, Wuerzburg (Germany): *Structural Characterization of the p52 Subunit of TFIIH from C. thermophilum.*

Poster presentation at the Gordon Research Conference on Nucleic Acids "RNA and DNA Processes Throughout Biology: Mechanistic Insights and Applications to Human Health", June 4-9, 2017, Biddeford, Maine (USA): *Structural and Biochemical Insights into the Intricate Network between the p34 and p44 Subunits of TFIIH.*

Poster presentation at the 12th International GSLS Student Symposium “EUREKA! 2017”, October 11-12, 2017, Wuerzburg (Germany): *Structural and Biochemical Insights into the Intricate Network between the p34 and p44 Subunits of TFIIH.*

VIII Curriculum vitae

IX Acknowledgements

Many people accompanied and supported me during this work and I would like to express my sincere gratitude towards them. First of all, I would like to thank my supervisor Caroline Kisker for giving me the possibility to continue working on the two fascinating projects I already pursued during my Master thesis. I am thankful for her constant advice, her positive attitude and especially for taking me with her to the Gordon Research Conference on Nucleic Acids in Maine, where we spent a most wonderful and informative time.
I would like to thank my second and third supervisors Bennett Van Houten and Thomas Müller for accompanying my projects, for our fruitful discussions and for the useful advice they offered me throughout the time of my PhD.

Furthermore, I would like to thank Jochen Kuper, who always supported me and offered his constructive guidance to lead my projects into the right direction. I am especially thankful for his valuable advice during the preparation of our p34/p44 manuscript and this Thesis as well as for his constant support during the structure solution and refinement processes of all my structures.
Wolfgang Kölmel helped me a lot in the earlier stages of the project and I would like to thank him especially for his advice and support during construct design, structure solution and refinement processes of the p52 project and for all the scientific and non-scientific discussions we shared.
I would like to thank Dominik Schmitt, who initially worked on the p34/p44 project, for all his informative support and all the data he shared with me on this project.
Furthermore, I am grateful to Hermann Schindelin for his constant interest in my projects and his scientific support.

During the time of my PhD work I supervised a few students and I want to thank all of them for their contribution to my projects and for giving me the opportunity to practice my teaching skills.
I would like to thank Stefan Peißert for working on the p34/p44 project during his Master Thesis, Max Meirow for spending two summer internships investigating p52 interactions and Leonie Müller and Yugendra Raghunadan for working on the p52/XPB interaction with me. Furthermore, I would like to thank Jeannette Kappenberger, who joined the p52/XPB project with so much enthusiasm.

I would like to thank our collaboration partners the groups of Jean-Marc Egly and Arnaud Poterszman, Hans Maric and the Borchers/Sickmann group of the ISAS for sharing projects, data and many scientific discussions with us.
I am very grateful to Andrea Schott-Heinzmann and Teresa Frank for managing all administrative work and to Bernhard Fröhlich and Roland Markert for the technical help they offered whenever needed. In addition, I am grateful to the GSLS team for supporting me with all the administrative work concerning my funding and the doctoral graduation regulations.

Furthermore, I would like to thank all colleagues of the research groups Kisker, Schindelin, Lorenz and Tessmer for creating such a lovely and cheerful working atmosphere. Special thanks go to my office mates for their nice company and the various synchrotron teams I joined for spending such an instructive but also entertaining time at the beamlines and on the road.
Very special thanks go to Anna Liess, who was always there for me and supported me in every possible manner and Sandra Eltschkner, who always knew how to cheer me up.

I would like to warmly thank my family and friends, especially my husband Matthias, for their constant support and encouragement throughout the last years.

Affidavit

I hereby confirm that my thesis entitled

Towards an understanding of the intricate interaction network of TFIIH

is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Würzburg, …… …….

(Date) (Signature)

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation

Auf dem Weg zum Verständnis des komplexen TFIIH Interaktionsnetzwerkes

eigenständig, d. h. insbesondere selbstständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Würzburg, …… …….

(Datum) (Unterschrift)