Structural and functional characterization of TFIIH from *Chaetomium thermophilum*

Strukturelle und funktionale Charakterisierung von TFIIH aus *Chaetomium thermophilum*

Doctoral thesis

for a doctoral degree at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg,
Section Biomedicine

submitted by

Wolfgang Kölmel

from Bad Mergentheim

Würzburg 2017

Submitted on: .....

Office stamp

Members of the *Promotionskomitee*:

Chairperson:	**Prof. Dr. Utz Fischer** University of Würzburg
Primary Supervisor:	**Prof. Dr. Caroline Kisker** University of Würzburg
Supervisor (Second):	**Prof. Dr. Thomas Müller** University of Würzburg
Supervisor (Third):	**Dr. Katrin Paeschke** University of Groningen

Date of Public Defence: .....

Date of Receipt of Certificates: .....

Willst du dich am Ganzen erquicken,
so mußt du das Ganze im Kleinsten erblicken.

Johann Wolfgang von Goethe

Abstract

Gene expression and transfer of the genetic information to the next generation forms the basis of cellular life. These processes crucially rely on DNA, thus the preservation, transcription and translation of DNA is of fundamental importance for any living being. The general transcription factor TFIIH is a ten subunit protein complex, which consists of two subcomplexes: XPB, p62, p52, p44, p34, and p8 constitute the TFIIH core, CDK7, CyclinH, and MAT1 constitute the CAK. These two subcomplexes are connected via XPD. TFIIH is a crucial factor involved in both, DNA repair and transcription. The central role of TFIIH is underlined by three severe disorders linked to failure of TFIIH in these processes: xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy. Only limited structural and functional data of TFIIH are available so far. Here, the model organism *Chaetomium thermophilum* was utilized with the aim to structurally and functionally characterize TFIIH. By combining the expression and purification of single TFIIH subunits with the co-expression and co-purification of dual complexes, a unique and powerful modular system of the TFIIH core subunits could be established, encompassing all proteins in high quality and fully functional. This system permits the step-wise assembly of TFIIH core, thereby making it possible to assess the influence of the intricate interaction network within TFIIH core on the overall enzymatic activities of TFIIH, which has not been possible so far. Utilizing the single subunits and dual complexes, a detailed interaction network of TFIIH core was established, revealing the crucial role of the p34 subunit as a central scaffold of TFIIH by linking the two proteins p44 and p52. Our studies also suggest that p62 constitutes the central interface of TFIIH to the environment rather than acting as a scaffold. TFIIH core complexes were assembled and investigated via electron microscopy. Preliminary data indicate that TFIIH adopts different conformational states, which are important to fulfill its functions in transcription and DNA repair. Additionally, a shortened construct of p62 was used to develop an easy-to-use, low cost strategy to overcome the crystallographic phase problem via cesium derivatization.

Zusammenfassung

Die Expression von Genen und die Weitergabe des Erbguts an die nächste Generation bilden die Grundlage jeden Lebens. Bei diesen Vorgängen spielt die DNA eine entscheidende Rolle. Deshalb sind der Erhalt, die Transkription und die Translation der DNA von fundamentaler Bedeutung für alle Lebewesen. Der generelle Transkriptionsfaktor TFIIH ist ein Multi-Proteinkomplex und umfasst insgesamt zehn Untereinheiten. TFIIH kann in zwei Teilkomplexe unterteilt werden: XPB, p62, p52, p44, p34 und p8 bilden den TFIIH Core Komplex, CDK7, CyclinH und MAT1 bilden den CAK Komplex. Diese beiden Teilkomplexe werden durch XPD verbunden. TFIIH spielt eine entscheidende Rolle sowohl in der DNA Reparatur, als auch in der Transkription. Diese zentrale Rolle wird durch das Auftreten dreier schwerer Krankheiten deutlich, die mit dem Ausfall von TFIIH bei diesen Aufgaben in Verbindung stehen: Xeroderma pigmentosum, Cockayne-Syndrom und Trichothiodystrophie. Daten bezüglich der Struktur und Funktion von TFIIH stehen bisher nur in begrenztem Umfang zur Verfügung. In dieser Arbeit kam der Modellorganismus *Chaetomium thermophilum* zum Einsatz, mit dem Ziel die Struktur und Funktion von TFIIH näher zu beleuchten. Durch die Kombination der Expression und Aufreinigung einzelner TFIIH Untereinheiten mit der Koexpression und Koaufreinigung von dualen Komplexen konnte ein einmaliges und leistungsfähiges modulares System entwickelt werden, das die Darstellung aller Untereinheiten in hoher Qualität und voller Funktionalität erlaubt. Basierend auf diesen Ergebnissen wurde die schrittweise modulare Zusammensetzung von TFIIH Core ermöglicht, was es nun erlaubt den Einfluss der komplexen Wechselwirkungen innerhalb von TFIIH Core auf die enzymatischen Aktivitäten im Ganzen zu untersuchen, was bisher nicht möglich war. Mit Hilfe der Einzelproteine und dualen Komplexe wurde ein detailliertes Netzwerk aus Wechselwirkungen innerhalb TFIIH Core etabliert, welches die entscheidende Rolle der p34 Untereinheit als zentrales Gerüst für TFIIH offenbarte, da sie die Verbindung zwischen p44 und p52 herstellt. Unsere Untersuchungen deuten zudem darauf hin, dass p62 die zentrale Schnittstelle zur Umgebung von TFIIH darstellt, anstatt als Gerüst zu fungieren. Des Weiteren gelang die Assemblierung von TFIIH Core Komplexen, die mittels Elektronenmikroskopie untersucht wurden. Die Strukturen, die daraus hervorgingen, legen das Vorhandensein verschiedener TFIIH Konformationen nahe, welche vermutlich bei den verschiedenen Aufgaben von TFIIH in der Transkription und DNA Reparatur zum Tragen kommen. Außerdem wurde mit Hilfe eines gekürzten p62 Konstrukts eine einfach zu handhabende, kostengünstige Strategie zur Lösung des kristallografischen Phasenproblems mittels Cäsiumderivatisierung entwickelt.

Contents

1. Introduction

Two fundamental processes form the basis of cellular life: Gene expression and transfer of the genetic information to daughter cells. Both processes have in common that they crucially rely on DNA. The preservation, transcription and translation of DNA is thus highly important for the survival of any living being. A crucial factor that is involved in the repair of DNA and its transcription is the general transcription factor IIH (TFIIH) [1]. Consequently, mutations in TFIIH lead to severe disorders [2]. TFIIH is an essential multi protein complex of which only limited structural and functional data are available. Therefore, TFIIH was the subject of this doctoral thesis. The aim was the structural and functional characterization of TFIIH, utilizing the model organism *Chaetomium thermophilum*.

1.1. The TFIIH complex

1.1.1. Composition of TFIIH

TFIIH is a ten subunit protein complex consisting of XPB, XPD, p62, p52, p44, p34, p8, CDK7, CyclinH, and MAT1 [1, 3, 4]. TFIIH is subdivided into two subcomplexes: XPB, p62, p52, p44, p34, and p8 form the core complex (Figure 1-2 b, c); CDK7, CyclinH, and MAT1 form the CAK (CDK activating kinase) complex [1, 4-6]. These two subcomplexes are connected via XPD [1, 4, 6, 7].

XPB (xeroderma pigmentosum group B) and XPD (xeroderma pigmentosum group D) belong to the superfamily 2 helicase family [8-11]. They contain two RecA-like domains, termed helicase domain 1 (HD1) and helicase domain 2 (HD2), bearing the seven conserved helicase motifs [12, 13]. Both helicases have additional unique motifs. XPB contains a RED and a Thumb motif (ThM), which are involved in the ATPase dependent DNA anchoring [14-16]. XPB also harbors an N-terminal extension with a damage recognition domain (DRD), and it has been suggested that this domain can bind to some types of lesions [14]. However, the exact function is unclear, as the DRD seems to be not crucial for DNA repair [15, 17]. XPD contains two additional domains, the iron sulfur cluster domain and the arch domain [18-20]. XPB has been classified as a 3'->5' helicase [8, 9], whereas XPD is a 5'->3' helicase [10, 11]. However, it is under debate if XPB is actually functioning as a true helicase [21]. Nevertheless, its ATPase activity is indispensable for transcription and DNA repair [7, 22]. XPDs helicase activity is indispensable for DNA repair, but dispensable for transcription, where it acts solely as a scaffold [22-24].

The activities of XPB and XPD are regulated through other TFIIH subunits. XPBs ATPase activity is stimulated through p52 and p8, and the N-terminal extension of XPB directly interacts with p52 through two distinct binding regions within p52 [22, 25, 26]. P8 assumes an α/β split fold with a C-terminal helix [1, 27, 28] and it interacts with the very C-terminus of

p52 [28, 29]. XPDs ATPase/helicase activity is stimulated through p44 [30]. P44 contains an N-terminal von Willebrand factor A (vWA) like domain, a central zinc finger domain, and a C-terminal ring finger domain [31-35]. SSL1, the yeast homolog of p44, has been shown to display E3 ubiquitin ligase activity [36]. The vWA domain interacts with HD2 of XPD [30]. This interaction not only stimulates XPD, but also anchors it to the TFIIH core [30]. Furthermore, the arch domain of XPD interacts with MAT1, thereby inhibiting XPDs helicase activity [37, 38].

P62 constitutes an N-terminal pleckstrin homology (PH) domain and a BSD (**B**TF2-like transcription factors, **s**ynapse-associated and **D**OS2-like) tandem with unknown function [39-41]. No enzymatic or regulatory function has been assigned to p62 so far. However, the PH domain competitively interacts with factors involved in DNA repair as well as transcription, suggesting a crucial role in both processes [42-45]. Additionally, the C-terminus of p62 interacts with p44 [46].

P34 contains an N-terminal vWA like domain and a C-terminal zinc finger domain [31, 47]. The vWA domain interacts with the ring finger domain of p44 [48].

CDK7 (cyclin dependent kinase 7) is a kinase involved in transcription and cell cycle control [1, 49-51]. It contains a typical kinase domain, consisting of an N-terminal lobe and a C-terminal lobe [52, 53]. The kinase domain contains twelve conserved subdomains [54]. CDK7 is stimulated via interaction with CyclinH [55, 56].

CyclinH consists of an N-terminal helical motif, a central cyclin box with two α-helical repeats, and a C-terminal helical motif [57, 58]. CyclinH interacts with CDK7 via the cyclin box [55, 56].

MAT1 (ménage à trois 1) consists of an N-terminal Ring finger motif, a central coiled-coil domain, and a C-terminal hydrophobic domain [37, 59, 60]. MAT1 associates with CDK7 and CyclinH via the hydrophobic domain, thereby stabilizing their interaction [5, 59]. The middle part, including the coiled-coil domain, interacts with XPD, leading to the inhibition of its helicase activity [37, 38, 59].

The domain architecture of the TFIIH subunits is depicted in Figure 1-1.

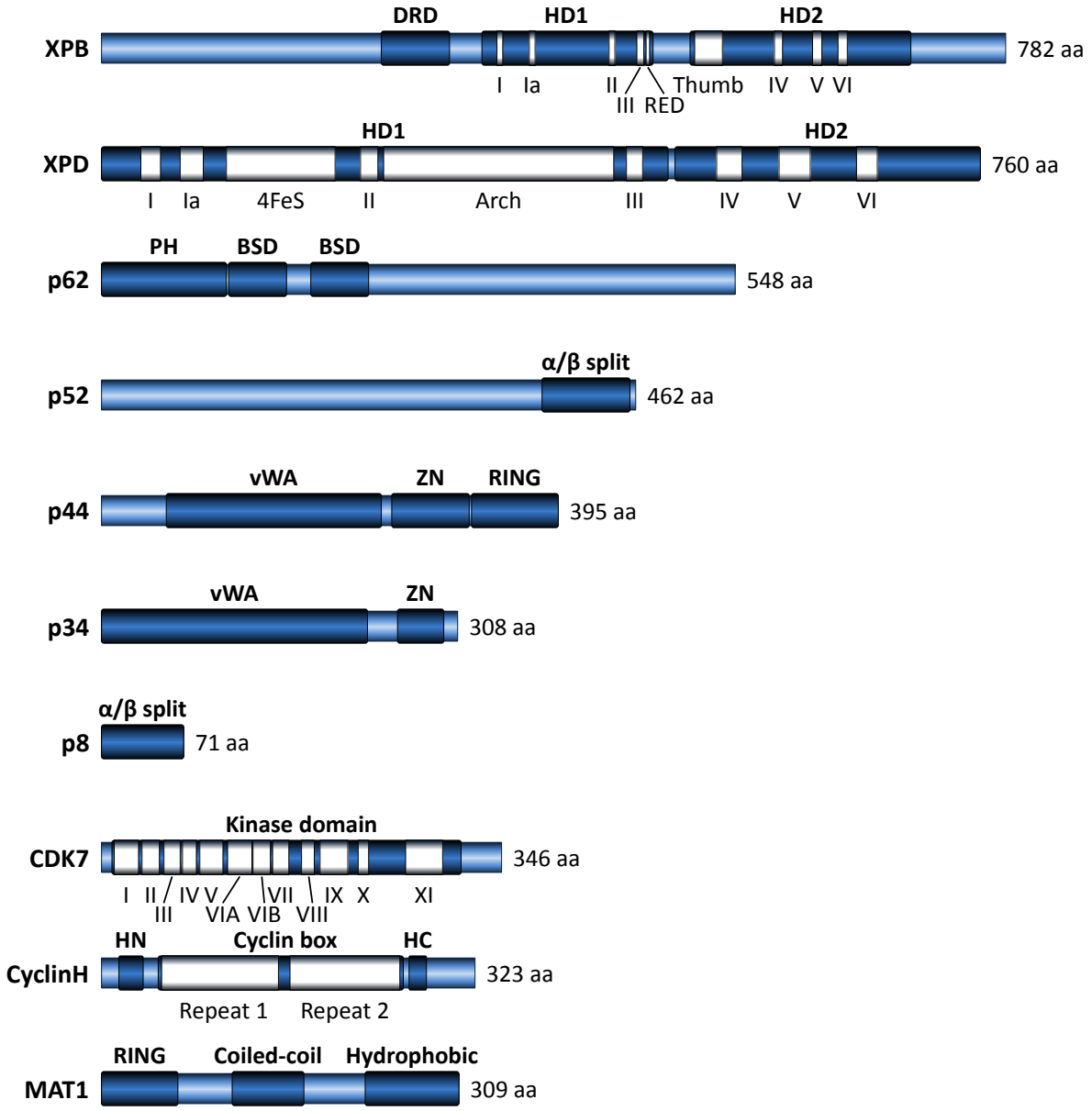


Figure 1-1. Domain architecture of the TFIIH subunits. DRD: damage recognition domain; HD1: helicase domain 1; HD2: helicase domain 2; 4FeS: iron sulfur cluster domain; Arch: arch domain; PH: pleckstrin homology domain; BSD: BTF2-like transcription factors, synapse-associated and DOS2-like domain; vWA: von Willebrand factor A like domain; ZN: zinc finger; RING: ring finger domain; HN: N-terminal helix; HC: C-terminal helix.

1.1.2. Structure of TFIIH

TFIIH has been the subject of numerous structural studies [61]. Regarding the individual subunits, structural information is available for each, even though to a quite varying extend and from different organisms. Full-length human structures are only available for CDK7 [52], CyclinH [57, 58], and p8 [27]. Partial human structures have been solved for XPB [62], p62 [41, 45, 63-65], p44 [34], and MAT1 [60]. Considering other organisms, full-length archaeal structures are available for orthologues of XPB [14] and XPD [18-20, 66, 67]. A full-length yeast structure for Tfb5 (p8) [28], and partial yeast structures have been solved for Tfb1 (p62) [43, 44, 68-73], Tfb2 (p52) [28] and SSL1 (p44) [35]. Additionally, there is a partial

structure of p34 from *Chaetomium thermophilum* [47]. These findings are depicted and summarized in Figure 1-2 a.

To date, structural data about the entire TFIIH complex is still quite limited. According to electron microscopy (EM) studies, TFIIH forms a ring like structure build by the core subunits and XPD (Figure 1-2 b, c), with the CAK attached [74, 75]. The resolution regarding core TFIIH is limited to about 10 Å (Figure 1-2 b, c) [76], and with respect to the entire TFIIH to about 20 Å [77]. Owing to the limited resolution, fitting of the TFIIH subunits to EM volumes is ambiguous, although chemical crosslinking/mass spectrometry has been utilized [76-78]. Especially the intricate interaction network between the TFIIH subunits remains elusive. It is known that, among the core subunits, XPB, p52, and p8 form one functionally associated subcomplex. XPD, p44, and p34 form another functionally associated subcomplex [79]. How the interfaces and precise interactions between the subunits of these subcomplexes are formed is poorly understood. Furthermore, how these two subcomplexes interact with each other within TFIIH is also poorly understood. A recent chemical crosslinking/mass spectrometry study suggested a connection of these subcomplexes via interaction of p52 with p44, and p52 with p34, as well as via bridging of XPB and XPD through the CAK subunit MAT1 [80]. How this intricate interaction network among the TFIIH subunits influences the overall enzymatic activities of TFIIH remains a key question to be answered.

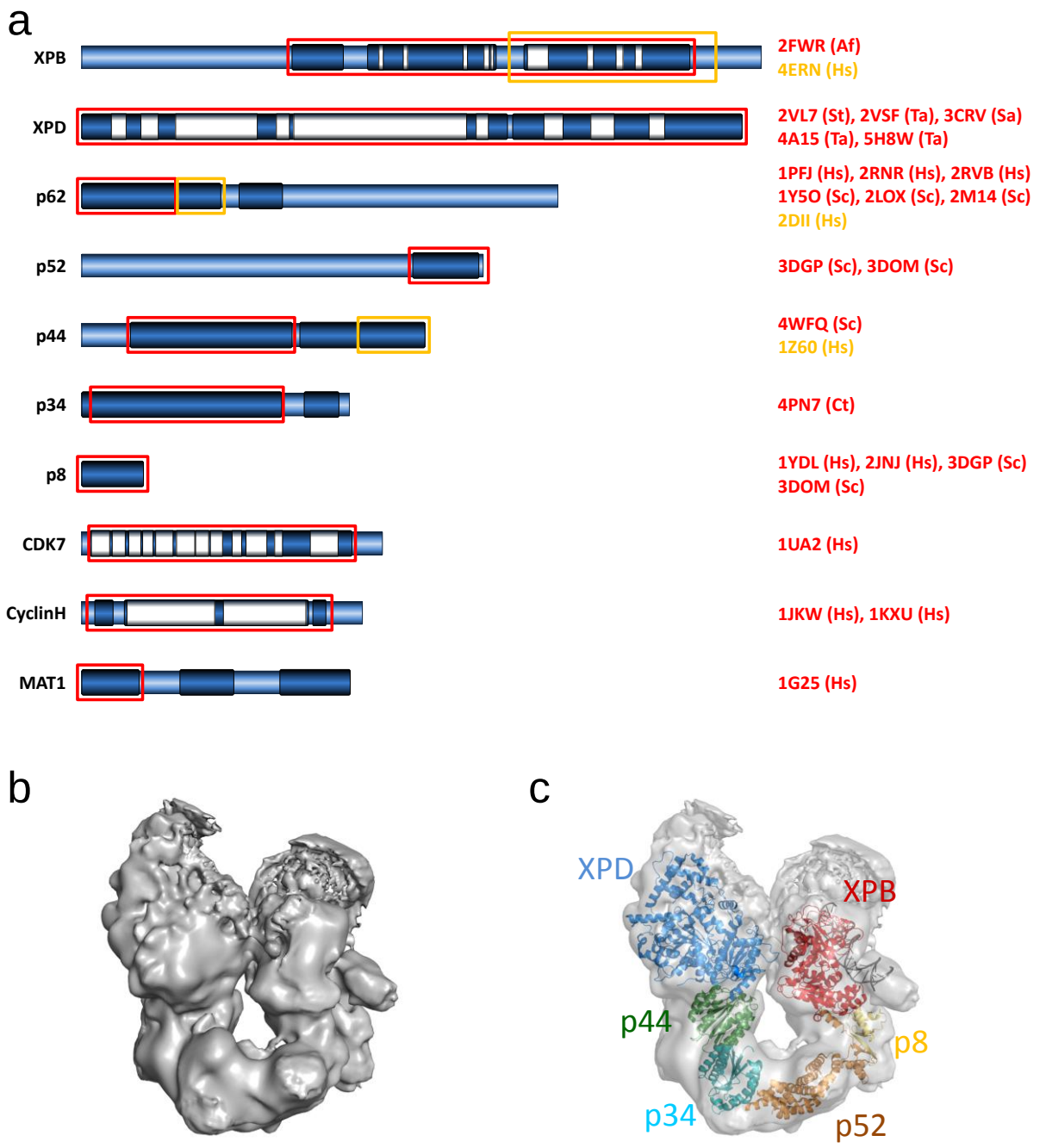


Figure 1-2. Structural characterization of TFIIH. (a) Structurally characterized regions of the different TFIIH subunits are indicated by colored boxes. Corresponding PDB codes are colored accordingly. The domain scheme corresponds to Figure 1-1. Af: *Archaeoglobus fulgidus*; Hs: *Homo sapiens*; St: *Sulfolobus tokodaii*; Ta: *Thermoplasma acidophilum*; Sa: *Sulfolobus acidocaldarius*; Sc: *Saccharomyces cerevisiae*; Ct: *Chaetomium thermophilum*. **(b)** Surface representation of an EM volume of TFIIH core (EMD-8131). **(c)** Fitted TFIIH subunits to the EM volume (5IVW).

1.1.3. The role of TFIIH

TFIIH is involved in several cellular processes of fundamental importance like transcription, DNA repair and cell cycle control [1, 79] and will be discussed in the following three sections 1.3, 1.4, and 1.5.

Several mutations in TFIIH subunits have been identified which give rise to different severe diseases with quite diverse phenotypes, namely xeroderma pigmentosum, Cockayne syndrome, and trichothiodystropy, reflecting the different cellular processes TFIIH is involved in [2, 81], which will be discussed in section 1.6.

1.2. History of TFIIH

The first available mention of a subject related to TFIIH is the description of the disorder xeroderma pigmentosum (XP; see 1.6.1) in 1874 by Ferdinand Hebra and Moriz Kaposi in their book “On diseases of the skin, including the exanthemata” [82, 83]. As apparent from the book title, XP was regarded as a disease of the skin at that time.

The history of TFIIH itself starts more than a century later. In the meantime, it has been recognized that XP actually originates from a defect in the ability of removing UV lesions from the genome [84].

In 1986 a factor involved in RNA polymerase II (RNAPII) transcription (see 1.4.2) was identified, termed BTF2 (which now corresponds to TFIIH) [85]. In 1988 Weber and coworkers reported that they cloned ERCC2, the gene encoding XPD, and showed that it could correct a nucleotide excision repair (NER; see 1.3.2) defect in a Chinese hamster ovary (CHO) cell line [86]. One and a half years later in 1989 Conaway and Conaway reported the purification of an essential RNAPII transcription factor from rat liver (termed δ factor) containing seven subunits, and demonstrated that it constitutes ATPase activity [87]. In 1990 it was shown that XPD displays high homology to the yeast 5’->3’ DNA helicase Rad3 (the helicase function of Rad3 was known at that time) [88, 89], and suggested that XPD is also a helicase [90]. In the same year Weeda and coworkers reported the cloning of ERCC3, the gene encoding XPB, and showed that it could also correct an NER defect in a CHO cell line [91]. Also in 1990, it was reported that the XPB amino acid sequence exhibits seven conserved helicase motifs, and it was suggested that XPB is a 3’->5’ helicase [8]. In 1991 Feaver and coworkers reported the purification of yeast transcription factor b (Tfb) containing at least three subunits, and demonstrated that it is crucial for RNAPII transcription [92]. One month later Gerard and coworkers reported the purification of BTF2 containing five subunits, and demonstrated that it is crucial for RNAPII transcription [93]. Shortly after, it was demonstrated that Tfb possesses kinase activity towards the C-terminal domain (CTD) of the large RNAPII subunit Rpb1 [94]. In 1992, Flores and coworkers reported the identification and purification of a novel human transcription factor, TFIIH, containing at least two subunits, and showed that it is essential for RNAPII transcription [95]. At that time it was not realized that BTF2 and TFIIH are identical and were thus regarded as two distinct transcription factors. In the same year, a kinase activity as shown for Tfb was also demonstrated for δ factor [96] as well as TFIIH [97], finally revealing that Tfb, δ factor, and TFIIH are homologs

from yeast, rat, and human respectively. Also in 1992, the cloning of a 62 kDa subunit from BTF2 was reported [98]. This subunit corresponds to p62 from TFIIH and it was finally realized that BTF2 and TFIIH are identical. Until that time the findings regarding XPD and XPB (NER) and TFIIH (transcription) were unrelated. This drastically changed in 1993 with the milestone observation that XPB is actually a part of TFIIH, linking TFIIH to both processes [99]. This finding was further supported in 1994 when it was shown that XPD is also part of TFIIH [100]. In parallel, two more subunits of TFIIH were cloned and characterized, p44 and p34 [31]. Also in 1994, Roy and coworkers reported that CDK7 is part of TFIIH, and demonstrated that it is responsible for its kinase activity [101]. In 1996 it was discovered that CyclinH and MAT1 are also part of TFIIH [5]. In 1997, p52 was cloned and characterized and was assumed to be the last subunit [102]. It took more than seven years more until the discovery of p8 in 2004, which was overlooked for a long time due to its small size (Figure 1-1) [3, 103].

1.3. DNA maintenance and repair

DNA carries the genetic information thereby possessing fundamental importance for life. It stores the blueprint for all proteins, which form the foundation of all cellular processes. Furthermore, DNA is the basis for the passage of the genetic information to the next generation of cells [104]. In order to fulfil its functions, the integrity of the DNA is crucial. As DNA is a chemical molecule it can undergo chemical reactions, leading to damages or mutations of the DNA. There are various sources for DNA lesions, both exo- and endogenous. Besides exogenous compounds [105, 106], most notably everyday sunlight poses a major threat, as the UV light component leads to DNA lesions [107]. Furthermore, the obligatory cell metabolism itself poses a threat to DNA as it produces reactive oxygen species like hydroxyl radicals or hydrogen peroxide [108, 109]. Also, spontaneous decomposition of chemical bonds in DNA, like hydrolysis of nucleotides or deamination, leads to damaged DNA [108, 110]. It is estimated that every cell encounters 10^4 to 10^5 lesions per day [111, 112]. These remarkable numbers immediately suggest a need for DNA maintenance and repair. Indeed, several pathways have evolved, ensuring the integrity of DNA and efficient repair of lesions. Five different repair pathways can be distinguished, each one dealing with a specific subset of lesions [81, 108, 113]. Apart from removing lesions from the genome, the cell has to deal with situations, where a yet unrepaired lesion actually interferes with DNA metabolism.

1.3.1. General DNA repair pathways, lesion bypass, and apoptosis

Base excision repair is responsible for the removal of oxidized or deaminated bases, abasic sites, and single strand breaks. Damaging agents include ionizing radiation, oxygen radicals, or alkylating agents [114, 115].

Crosslink repair refers to the removal of interstrand crosslinks. Damaging agents include lipid peroxidation products or chemical compounds (e.g. mitomycin C) [116-118].

Double strand break repair is responsible for the removal of double strand breaks. Damaging agents include ionizing radiation, free radicals, or chemical compounds (e.g. cis-platinum) [119, 120].

Mismatch repair recognizes and removes A-G/T-C mismatches, insertions, and deletions, which are caused by replication errors [121, 122].

Nucleotide excision repair removes UV light induced lesions as well as bulky DNA adducts [123] and is the focus of this thesis due to the involvement of TFIIH. It will therefore be discussed in greater detail in the following section.

If a lesion is encountered during DNA replication, the replicative DNA polymerase (δ/ε) [124] becomes stalled, leading to replicational stress. To bypass the lesion, the polymerase is temporarily replaced by a specialized polymerase capable of translesion synthesis [125-127]. Several translesion polymerases exist, each one being responsible for a certain subset of lesions [126-132]. This process enables to continue DNA replication, though it is error prone and may lead to persistent mutations [126, 133].

Another pathway which lesions interferes with is transcription. If a lesion is encountered on the transcribed stand, RNA polymerase II is stalled. This is specifically addressed by transcription coupled repair (TCR). As a subpathway of NER, TCR is described in more detail in section 1.3.2.2. Alternatively, lesion bypass may be a pathway to deal with transcriptional blocks [134-138]. This however, may result in incorrectly transcribed mRNA, leading to transcriptional mutagenesis [134, 139, 140]. Consequently, mutant proteins are synthesized, which may impair cellular functions [136, 139].

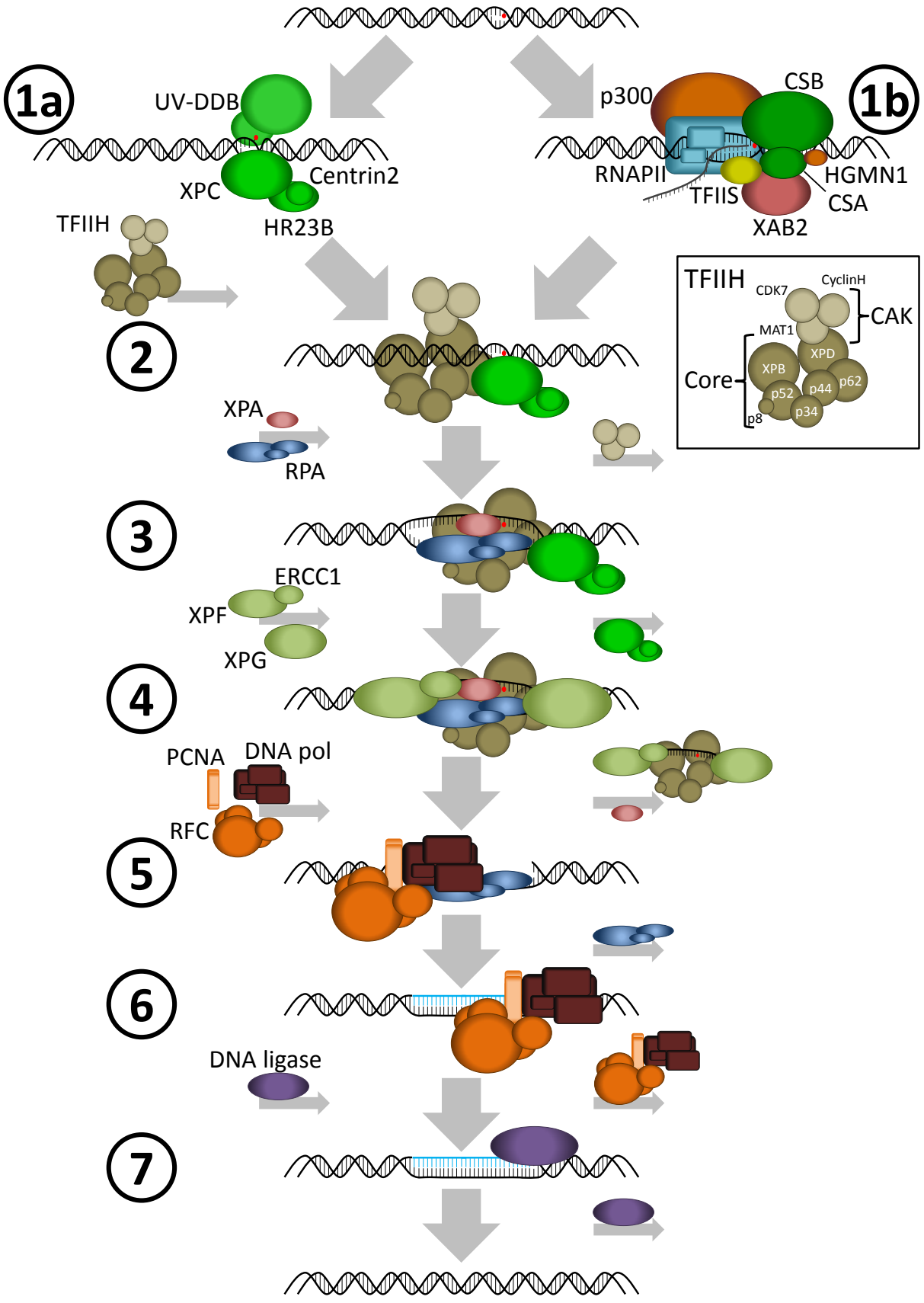
If the replicational or transcriptional stress is too high, the cell may induce apoptosis as a last resort mechanism [108, 141-147].

1.3.2. Nucleotide excision repair

Nucleotide excision repair (NER) is a major DNA repair pathway which repairs a broad range of quite diverse DNA lesions [123]. It addresses various bulky DNA adducts such as polycyclic aromates or intrastrand crosslinks. Most importantly, it is the only pathway in humans which repairs UV-light induced damages, like cyclobutane pyrimidine dimers (CPD) and 6,4-photoproducts [123]. There are two distinct pathways of NER, namely global NER and transcription coupled repair (TCR). They have different entry points but converge into a

common repair process [1, 144, 148]. Global NER constantly scans the entire genome in order to remove lesions [149]. TCR removes DNA damages in actively transcribed strands that are stalling RNA polymerase II (RNAPII) during transcription [150]. These two NER pathways will be discussed in more detail in the following two sections and are depicted in Figure 1-3.

Figure 1-3. Mechanism of nucleotide excision repair (NER). Damaged DNA is either recognized by XPC/HR23B/Centrin2 with or without the help of UV-DDB (global NER) (1a), or a stalled RNAPII (TCR) (1b). Both entry points converge into a common pathway upon recruitment of TFIIH (2). XPA and RPA are recruited, thereby CAK is released. TFIIH then unwinds the DNA around the lesion through the helicase activity of XPD (3). The endonucleases XPG and XPF-ERCC1 are recruited to the repair bubble. XPF-ERCC1 makes an incision 5' to the lesion and XPG makes an incision 3' to the lesion in the damaged strand (4). The excised strand is released with TFIIH and the endonucleases bound. DNA polymerase δ, ε, or κ is recruited via RFC and PCNA (5). The DNA polymerase synthesizes DNA, filling the single stranded gap (6). The DNA is finally ligated by DNA ligase I or XRCC1-DNA ligase IIIα (7).


1a
UV-DDB
XPC
Centrin2
HR23B
TFIIH
1b
p300
CSB
RNAPII
TFIIS
HGMN1
CSA
XAB2
TFIIH
CDK7
CyclinH
CAK
MAT1
XPD
Core
XPB
p52
p44
p62
p34
p8
2
XPA
RPA
3
ERCC1
XPF
XPG
4
PCNA
DNA pol
RFC
5
6
DNA ligase
7

1.3.2.1. Global nucleotide excision repair

The major player in the initial damage detection process during global NER is the XPC/HR23B/Centrin2 complex [151-154]. Thereof, XPC is the crucial factor, responsible for damage detection [153, 155]. However, HR23B and Centrin2 play important roles in the stabilization and stimulation of XPC [153, 156-158]. XPC scans the DNA for lesions and binds to them [154, 159-162]. The specific mode of binding is the basis for the broad substrate range of NER. It has been shown that XPC does not bind to the lesion itself, but to the altered DNA structure, particularly to the unpaired DNA bases on the strand opposite to the lesion [154, 163, 164]. This mode of action immediately implies a certain distortion of the DNA in order to be recognized [165]. However, some lesions, especially CPDs, hardly introduce distortion into the DNA [123, 166]. In line with this, it has been shown that CPD lesions are only poorly recognized by the XPC complex [154, 159, 167]. Here UV-DDB comes into play, which is a heterodimer consisting of DDB1 and DDB2 [168, 169]. UV-DDB can bind to (but not only) CPDs and is thereby able to detect these less distorting lesions [170-173]. Upon binding, DDB2 flips the lesion out of the DNA duplex into its active site pocket [174, 175], which facilitates binding of XPC/HR23B/Centrin2 to the damaged DNA [176-179]. Subsequently, TFIIH is recruited to the lesion site [180-183]. Recruitment most likely occurs via interactions of XPC with p62 and XPB [42, 44, 45]. It has also been shown that XPC stimulates the ATPase activity of XPB [42, 184].

The anchoring of TFIIH to the repair site is achieved by XPB [15]. It has been suggested that XPB therefore undergoes a conformational change upon hydrolysis of ATP, bringing the RED motif and ThM domain (Figure 1-1) close together. The RED motif intrudes into the DNA, forming a wedge, which is clamped by the ThM domain [9, 14, 15]. Subsequent recruitment of XPA to the site of repair poses an important step, as this leads to the dissociation of CAK from core TFIIH [185]. As CAK inhibits the helicase activity of XPD, its release leads to the activation of XPD [37, 38, 184, 186]. XPD unwinds the DNA around the lesion thus creating a DNA bubble [22, 187]. Also, the single stranded DNA binding protein RPA (replication protein A) is recruited, which protects the undamaged single strand [188-194]. At this stage an additional damage verification step takes place. It is under debate how this verification is achieved. A likely candidate is XPD. It has been suggested that XPD threads the damaged single strand through a pore formed by HD1, the iron sulfur cluster domain, and the arch domain [18-20, 66, 67, 195]. Thereafter, the diameter of the pore would only allow undamaged single stranded DNA to pass through, whereas a lesion would lead to a physical block of the helicase movement [66, 186]. Indeed, it has been shown that TFIIH scans the lesion site with 5'->3' directionality [196], and that bulky lesions stall XPDs movement and inhibit its helicase activity [186, 197-199]. It has also been suggested that XPA is involved in damage verification as it is able to bind to damaged DNA [200-202].

Reminiscent of XPC, it recognizes kinked DNA duplexes rather than the lesion itself [203-205]. Furthermore, XPA seems to increase the inhibition that DNA lesions exert on XPDs helicase activity [186].

After verification of the damage, the endonucleases XPG and XPF-ERCC1 are recruited. In this process the XPC complex is released [183]. It has been suggested that the PH domain of p62 plays a crucial role in this process, as it competitively interacts with both XPC and XPG, thereby organizing the “hand-over” from XPC to XPG [43-45]. XPF-ERCC1 on the other hand is recruited via XPA [206-210]. In addition, RPA is important for the correct positioning of XPG and XPF-ERCC1 [190, 211, 212]. Dual incision is likely to be initiated by XPF-ERCC1 [213], making an incision 5’ to the lesion into the damaged strand [214-216]. Subsequently, XPG makes an incision 3’ to the lesion into the damaged strand [215, 217-219]. However, studies about the order of these incisions are contradictory [1, 187, 192, 213, 220]. Dual incision results in a 24-32 nucleotides long single stranded DNA stretch containing the damage asymmetrically located closer to the 3’ end [215, 221-225]. TFIIH, XPG, and presumably also XPF-ERCC1 remain bound to the excised strand and are released together with it, resulting in a single stranded gap in the DNA [225-227]. The clamp loader RFC (replication factor C) and the sliding clamp PCNA (proliferating cell nuclear antigen) are recruited, and the RFC/PCNA complex subsequently recruits DNA polymerase δ, ε, or κ [183, 228-232]. It has been shown that RPA and XPG are involved in proper recruitment of RFC/PCNA [233, 234]. The DNA polymerase then synthesizes DNA, filling the single stranded gap, and final ligation is carried out by DNA ligase I or XRCC1-DNA ligase IIIα [229, 230, 235].

1.3.2.2. Transcription coupled repair

The trigger of transcription coupled repair (TCR, also referred to as TC-NER) is the stalling of RNAPII due to the presence of a lesion in the actively transcribed strand [236-240]. During transcription, RNAPII is accompanied by the transcription elongation factor CSB (Cockayne syndrome group B) [138, 150, 241, 242]. Upon stalling, CSB binds tightly to RNAPII and DNA [138, 150, 242-245]. In this process CSB changes the DNA conformation, presumably by wrapping it [138, 246, 247]. It has been shown that XPG binds to stalled RNAPII and stimulates the ATPase activity of CSB, suggesting a role in recognition of stalled RNAPII [248]. Subsequently, CSB recruits the CSA E3 ubiquitin-ligase complex, the histone acetyltransferase p300, and the NER proteins, including TFIIH [249, 250]. CSA recruits further factors, like the nucleosome binding protein HMGN1, elongation factor TFIIS, and the XPA binding protein 2 (XAB2) [250]. The fate of RNAPII in the subsequent steps is under debate. As the stalled RNAPII occupies ~10 nucleotides in front and ~25 nucleotides behind the lesion it has to be removed in order to grant access to the NER machinery to the lesion

site [138, 251, 252]. One proposed mechanism is the degradation of RNAPII initiated by the ubiquitination through NEDD4. This would lead to the abortion of transcription in favor of DNA repair [253-259]. Another proposed mechanism is the backtracking of RNAPII with the help of p300, HMGN1, and TFIIS. The chromatin remodelers p300 and HMGN1 would disassemble the nucleosomes behind RNAPII to permit the backtracking, and TFIIS activates the nascent RNA cleavage by RNAPII [138, 150, 238, 252, 260]. Yet another proposed mechanism is the remodeling of RNAPII. Hereby, TFIIH would induce conformational changes in RNAPII, granting access to the damaged site with RNAPII still in place [248]. In either case TFIIH is recruited, triggering the same repair cascade as described in the section about global NER (see 1.3.2.1) after TFIIH has been recruited by the XPC complex [148, 150, 249, 250].

1.4. Transcription

The second major cellular pathway TFIIH is involved in is transcription. Transcription is the transmission of the genetic information from DNA into RNA [104]. This is carried out by RNA polymerases accompanied by various transcription factors and regulators [76, 77, 261-266]. In eukaryotes, three RNA polymerases can be distinguished: RNA polymerase I (RNAPI), II (RNAPII), and III (RNAPIII). Each polymerase is responsible for the synthesis of a different subset of RNAs [267-271]. TFIIH is a crucial general transcription factor in RNAPII dependent transcription [76, 77, 87, 92, 93, 97, 272]. Additionally, TFIIH is verifiably involved in RNAPI dependent transcription [272-275]. A few hints exist that TFIIH might also be involved in RNAPIII dependent transcription [276, 277]. In the following two sections transcription by the three polymerases will be briefly discussed as well as the role of TFIIH therein.

1.4.1. RNA polymerase I and III transcription

RNAPI synthesizes the large ribosomal RNA (rRNA) precursor, which encodes for the 28S, 18S, and 5.8S ribosome subunits [278-283]. The preinitiation complex is initiated by the promotor selectivity factor TIF-IB/SL1 together with the upstream binding factor UBF [284-290]. TIF-IB/SL1 consists of the TATA binding protein (TBP) and three TBP associated factors, which are specific for RNAPI [291-293]. TIF-IB/SL1 and UBF recruit RNAPI, alongside with additional factors [288, 289, 294-297]. It was found that TFIIH binds to the rDNA promotor and associates with TIF-IB/SL1 and RNAPI [272, 273]. Furthermore, TFIIH leaves the rDNA promotor together with RNAPI, and it has been suggested that TFIIH is involved in transcription elongation [274, 275]. The exact function of TFIIH in this process, however, is unclear, as neither its helicase, nor its kinase activity is required *in vitro* [273,

298]. *In vivo* however, it has been shown that XPBs ATPase activity enhances RNAPI transcription [275].

RNAPIII synthesizes transfer RNA (tRNA), the 5S ribosome subunit, and some noncoding RNAs [299]. A unique feature of RNAPIII transcription is that the majority of the transcribed genes do not require upstream promotors, but possess internal promotors [300-302]. These promotors are recognized by TFIIIC which subsequently recruits TFIIIB [300, 302-304]. However, a small subset of RNAPIII transcribed genes constitute upstream promotors containing a TATA box. These TATA boxes are recognized by TFIIIB, as TFIIIB contains TBP [300, 302, 305-307]. In both cases TFIIIB then recruits RNAPIII [300, 302, 308, 309]. To date, no direct evidence exists that TFIIH is involved in RNAPIII transcription. However, it could be shown that TFIIH, together with TFIIB, TFIIE, and RNAPII, associates with a range of RNAPIII transcribed genes [276, 277, 310].

1.4.2. RNA polymerase II transcription

RNAPII dependent transcription is the first step of protein expression and responsible for the synthesis of pre-messenger RNA (pre-mRNA) as well as noncoding RNAs [104, 311, 312]. Pre-mRNA is processed to mRNA [313, 314] and is then translated into the primary amino acid sequence in the second step of protein expression [104, 315-317].

The formation of the preinitiation complex (PIC) is initiated by TFIID [318, 319] which contains TBP and thereby binds to the promotor region [318, 320, 321]. Subsequently, TFIIA and TFIIB are recruited [318, 322]. Next, RNAPII together with TFIIF are recruited [323]. This complex is then joined by TFIIE. Finally, TFIIE recruits TFIIH [63, 322, 324-327]. Further factors are involved in PIC formation, like nuclear receptors (NR) and the multi subunit Mediator complex [1, 328-330]. The Meditator complex stimulates and controls the PIC formation [331-334] and it extensively interacts with RNAPII, transcription factors like TFIIB and TFIIH, as well as nuclear receptors [77, 266, 335, 336].

TFIIH, in particular its XPB subunit, is believed to fulfill a key role in promotor opening [7, 337]. The exact mechanism of this process is under debate. It could be shown that XPB binds downstream from the actual opening region, arguing against a common helicase mechanism [76-78, 338, 339]. It has been suggested that XPB functions as a DNA translocase. In the PIC, on the one hand, the DNA is fixed upstream of the opening region by TBP. On the other hand, XPB is fixed downstream of the opening region by TFIIE. Thus, the DNA translocase activity of XPB would lead to a rotation and insertion of DNA into the RNAPII active site cleft, resulting in DNA opening [21, 340]. Another hypothesis suggests that XPB functions as a wrench, which rotates the DNA in order to generate torque, resulting in DNA opening [339]. Furthermore, it has been suggested that XPB induces an ATP dependent conformational change into the PIC leading to DNA opening [341].

In stark contrast, a recent study suggested that XPB might not be involved in promotor opening at all, as transcription seemed to be possible even in the absence of XPB. The authors suggested that XPB might act as a regulatory transcriptional block, which is released upon ATP hydrolysis through XPB [342].

XPDs enzymatic activities are dispensable for transcription, yet it plays an important scaffolding role [23, 24]. XPD anchors CAK to TFIIH via the interaction with MAT1 [38, 59].

P8 seems to play an important role in transcription, as all known disease related mutations lead to the transcription related disorder trichothiodystrophy (see 1.6.3) [3, 343, 344]. P8 stabilizes TFIIH [28] and mutations in p8 result in reduced TFIIH levels [345].

MAT1, in particular the N-terminal ring finger domain (Figure 1-1), also plays an important role, as mutation or deletion of this domain impairs transcription [59].

CDK7 is important for promotor escape as it phosphorylates serine 5 and 7 of the heptad repeats of the CTD of the RNAPII Rbp1 subunit [50, 94, 96, 97, 101, 346]. This leads to the dissociation of Mediator from RNAPII enabling the movement of RNAPII from the initiation site [334, 347, 348]. Additionally, CTD phosphorylation is required for pre-mRNA processing [349, 350]. Furthermore, CDK7 is involved in phosphorylation of transcription factors [327] and transactivation of NRs [351-354]. The activity of CDK7 is in turn regulated by the CDK8 subunit of Mediator. CDK8 phosphorylates CyclinH, thereby reducing the activity of CDK7 [355].

TFIIH also seems to play a role in early elongation and suppression of promoter-proximal arrest of RNAPII [356-360].

The process of transcription initiation is depicted in Figure 1-4.

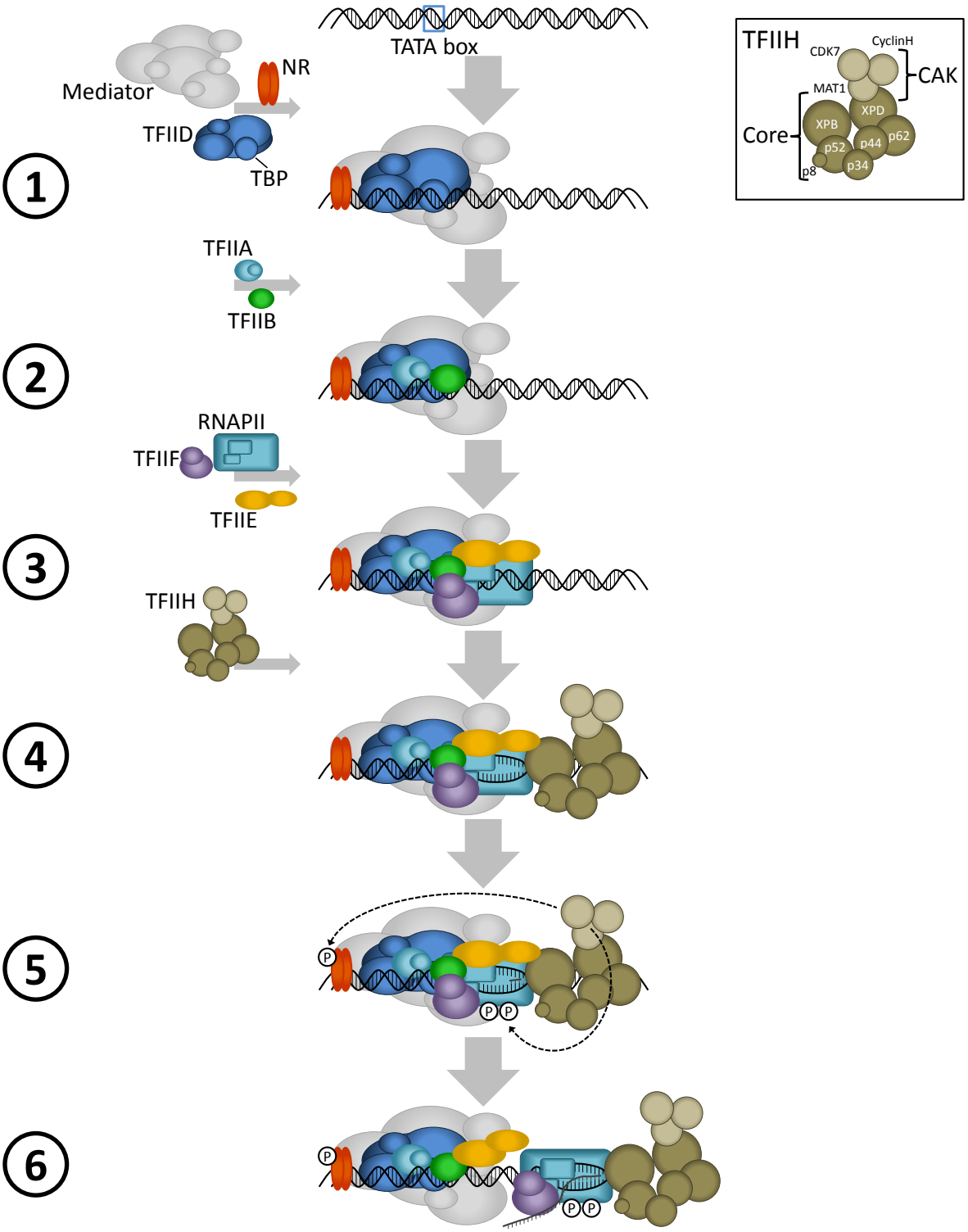


Figure 1-4. Mechanism of RNA polymerase II transcription initiation. Preinitiation complex (PIC) formation is initiated by TFIID, at which TBP recognizes the TATA box. Mediator and nuclear receptors (NR) stimulate and control PIC formation (1). Subsequently, TFIIA and TFIIB are recruited (2). This complex recruits RNA polymerase II (RNAPII) alongside with TFIIF. Next, TFIIE joins the complex (3). TFIIE recruits TFIIH, which then opens the promoter via its XPB subunit (4). CDK7 phosphorylates the C-terminal domain of the RNAPII subunit Rbp1 and nuclear receptors (5). RNAPII escapes the promoter and synthesizes pre-mRNA (6). Ⓟ: phosphorylation.

1.5. Cell cycle control and chromosome segregation

In addition to its roles in transcription and DNA repair, TFIIH is also involved in other important cellular processes. These functions are not carried out by the complete TFIIH complex, but rather by certain subparts or subunits, also in the context with other protein complexes [1, 79].

The CAK complex plays a crucial role in cell cycle control [1, 49, 51]. Interestingly, CAK is the only known CDK activating kinase in metazoans [1, 361]. It is able to stimulate a certain set of CDKs important for cell cycle control via T-loop phosphorylation through CDK7 [51]. This set comprises CDK1, CDK2, CDK4, and CDK6 [49, 55, 361-364]. T-loop phosphorylation promotes the association of these CDKs with different cyclin subunits or activates the preformed CDK/cyclin complexes [49, 365]. These CDK/cyclin complexes are the actual functional species, able to trigger cell cycle events [365-367]. It has been shown that CAK levels and activity remain constant during the cell cycle, suggesting that CAK is a constitutive activator of the CDKs [5, 368]. This observation immediately poses the question about possible regulators of CAK activity. One possible regulator is XPD [1, 79, 369]. It could be shown in *Drosophila* that XPD inhibits the kinase activity of CAK [370]. Indeed, XPD levels fluctuate during cell cycle in *Drosophila*, suggesting a mechanism how CAK might be regulated during cell cycle progression [370]. In addition, XPD controls the subcellular localization of CDK7 in *Drosophila* [371].

XPD is also involved in chromosome segregation. It was found that XPD is part of the MMXD complex [1, 79, 372, 373]. Besides XPD, MMXD consists of MMS19, MIP18, Ciao1, and ANT2. MMXD seems to localize to mitotic spindles during metaphase, and knock-down of XPD, MMS19, or MIP18 leads to a disturbed spindle pole formation [373]. Furthermore, MMS19 and MIP18 are required for the correct positioning of Aurora B at the central spindle and midbody [373]. However, it has been found recently that MMS19, MIP18, and Ciao1 are part of the cytoplasmic Fe-S assembly pathway [374-376]. Hence, XPD most likely acquires its iron sulfur cluster via MMXD [376]. Therefore, these observed detrimental effects on chromosome segregation might be indirect through impaired XPD synthesis. Furthermore, lack of XPD leads to nuclear division defects and chromosomal instability in *Drosophila* [371]. XPDs involvement in chromosome segregation is also indicated by its participation in the CGX complex in *Drosophila* [1, 377]. Besides XPD, CGX consists of Crumps, Galla-1, and Galla-2. Here, Galla-1 shows sequence homology to the MMXD subunit MIP18 [377]. CGX localizes to mitotic spindles and is required for correct chromosome segregation in early embryogenesis [377].

1.6. TFIIH and disease

The fundamental importance of TFIIH for transcription and DNA repair becomes evident through the severe diseases caused by mutations in TFIIH. Three quite diverse autosomal recessive disorders, xeroderma pigmentosum, Cockayne syndrome, and trichothiodystropy, are linked with malfunctioning TFIIH [2, 81]. The diversity of these syndromes reflects the multiple and diverse roles of TFIIH in the cell. The diseases and the corresponding mutations in TFIIH will be discussed in the following four sections.

1.6.1. Xeroderma pigmentosum

Xeroderma pigmentosum (Greek ξηρός (xerós) for dry, δερμα (derma) for skin; New Latin pigmentosum for pigmentation) (XP) is characterized by a highly elevated photosensitivity, with an up to 10,000 fold increased risk to develop skin cancer [83]. Sun exposed areas of the skin develop several abnormalities like numerous hyperpigmented macules. Some patients suffer from severe sun burns, even after minimal sun exposure [378]. Also, some patients display ocular abnormalities [379]. Sometimes XP is accompanied by progressive neurological degeneration, like mental degeneration or hearing loss [83, 380]. The photosensitivity and skin cancer phenotype of XP can directly be linked to TFIIHs function in NER, as it is the only pathway in humans capable of repairing UV-light induced lesions [123]. It has been suggested that the progressive neurological degeneration is caused by a decay of non-dividing nerve cells due to insufficient DNA repair [380, 381] or subtle transcriptional defects [382, 383].

1.6.2. Cockayne syndrome

Cockayne syndrome (named after Edward Alfred Cockayne, who first described this syndrome [384]) (CS) is a disorder typically attributed to a defective TCR pathway [385]. Specifically, this syndrome is caused by mutations in CSA or CSB [2, 386, 387]. No mutations in TFIIH are known that cause solely CS. Here, only combined XP/CS phenotypes have been observed so far [83, 388, 389]. CS is characterized by severe mental retardation, microcephaly, a bird-like face, retinal degeneration, long limps, and gait defects [2, 81, 379, 390]. Some patients display photosensitivity, but without abnormal skin pigmentation or increased skin cancer risk. CS is clinically heterogeneous, with severe and moderate phenotypes, which lack a clear genotype-phenotype relationship [2, 390, 391]. The observed photosensitivity is mainly attributed to impaired repair of UV lesions through defective TCR [4, 385]. Presumably, the neurological symptoms are caused by impaired resolving of transcriptional blocks, thereby interfering with proper transcription [387], or shifting the balance between cell survival and apoptosis [136, 141, 146]. Also TCR unrelated impairment

of RNAPI and RNAPII transcription through defects in CSA and CSB or TFIIH has been proposed as being causative for the neurological disorders [4, 136, 274, 392]. Regarding TFIIH, the occurrence of only combined XP/CS may well reflect its roles in global NER, TCR, and transcription.

1.6.3. Trichothiodystrophy

Trichothiodystrophy (Greek τριχο (tricho) for hair, θειον (theion) for sulfur, δυσ (dys) for bad, τροφεῖν (trophein) for nutrition) (TTD) is characterized by brittle hair and nails [393, 394]. It is often accompanied by mental retardation, short stature, ichthyotic skin, and photosensitivity, but without abnormal skin pigmentation or increased skin cancer risk [2, 394]. In contrast to XP, TTD is attributed to the impaired function of TFIIH during transcription [275, 395, 396]. TFIIH levels are significantly reduced in all known TTD cases [397] and it has been suggested that in specific differentiated cells (like keratinocytes) most of the genes are transcriptionally repressed, including those encoding for TFIIH. These cells therefore have to rely on residual TFIIH levels, which might be too low in case of TTD cells, resulting in the TTD phenotype [398]. However, this cannot explain all symptoms of TTD, especially regarding the fact that reduced TFIIH levels were also found in some XP cells [397]. Hence, impairment of other functions, like transactivation of NRs might also contribute to TTD [4, 344, 399]. One enigma however remains: TTD is often accompanied by photosensitivity and defects in NER, but without display of the XP phenotype of abnormal pigmentation and skin cancer. It has thus been suggested that the transcriptional defect of TTD prevents the occurrence of an XP phenotype. According to that hypothesis, the expression of crucial genes would be greatly reduced, the overexpression of which would be needed to cause abnormal skin pigmentation and cancer [400].

1.6.4. Disease related mutations in TFIIH

Regarding TFIIH, disease related mutations in only three subunits have been observed so far in humans: XPD, XPB, and remarkably, the smallest subunit p8 [1, 3, 344, 388, 389, 401]. A likely explanation of lacking mutations in other TFIIH subunits might be an incompatibility with life due to their essential role in maintaining the integrity of TFIIH [1]. A particular case might be p44. In humans the gene encoding p44 is duplicated, and the gene products differ in three amino acids [402]. It has been shown that both versions are functional within TFIIH [7]. Therefore, the lack of p44 disease mutations might be due to the little likelihood of mutations in both alleles of both genes [30]. The mutations found in XPD, XPB, and p8 will be discussed below.

1.6.4.1. XPD

Mutations in XPD can result in XP, XP/CS, TTD, or XP/TTD [81, 388]. Several mutations have been identified, most of them cluster in HD2 of XPD (Figure 1-5) [18, 388, 403]. Most of these mutations are missense mutations, and surprisingly, some mutations are adjacent to each other but cause different disorders. It has been suggested that the different disorders are related to the impairment of specific properties of XPD. According to this theory, mutations impairing ATP hydrolysis or DNA binding lead to XP, mutations impairing the conformational flexibility between HD1 and HD2 lead to XP/CS, and mutations impairing the structural integrity of XPD or protein-protein interactions within TFIIH lead to TTD [18, 369]. This hypothesis is supported by several mutations.

A frequent XP mutation is R683W/Q, occurring in ~80 % of the patients [388, 403]. R683 is located in the DNA binding channel in HD2, contacting the DNA backbone and the mutation therefore impedes DNA binding. Similarly, several other XP mutations are located within the DNA binding channel (T76A, S541R, Y542C, and R601L/W) or the ATP binding site (D234N). A mutation leading to XP/CS is G675R. It is located at the bottom of HD2, and substitution of a flexible G with a more rigid R reduces the conformational flexibility of HD2 [18]. Another XP/CS mutation is G602D, which similar to G675R, also substitutes a flexible G with a more rigid amino acid. Remarkably, this mutation is directly adjacent to the XP mutant position R601. A frequent TTD mutation is R722W, which disrupts the XPD-p44 interaction, thereby impairing the recruitment of XPD to TFIIH [24, 396]. As the disruption of the XPD-p44 interaction also abolishes the helicase activity of XPD, it is puzzling why this mutation displays no XP phenotype. An attempt to explain is given in the section about TTD (see 1.6.3). Another TTD mutation is C259Y, which is located in the arch domain of XPD. This mutation affects the interaction with CAK, impairing the transcriptional activity of TFIIH [38].

Also mutations causing a combined XP/TTD phenotype have been reported, which might well reflect the concurrent impairment of different XPD properties [404].

A remarkable observation regarding XP/CS mutations of XPD are aberrant DNA incisions distant from repair sites which were shown to be transcription dependent [405, 406]. In accordance with reduced conformational flexibility of XPD, one might speculate that TFIIH is locked in a repair conformation, leading to erroneous TFIIH mediated repair events at transcription sites [18].

1.6.4.2. XPB

Mutations in XPB can result in XP, XP/CS, or TTD [389]. In contrast to XPD, very few mutations of XPB have been identified so far. Most likely, this is due to XPBs essential role in basal RNAPII transcription, making most XPB alterations incompatible with life [1]. Only three

causative mutations are known: The missense mutations F99S and T119P, and a frameshift mutation alternating the sequence of the last 42 amino acids (fs740) (Figure 1-5) [389, 401].
F99S causes either a mild form of XP or a mild form of XP/CS [389]. It is unclear, why one mutation causes two different phenotypes. However, these patients differ in the second allele. For the XP phenotype, the second allele bears a point mutation, leading to a stop codon at amino acid 425 (full-length: 782 amino acids). For the XP/CS phenotype, the second allele bears a splice donor mutation in intron 3, resulting in a frameshift after amino acid 157 and a stop codon at amino acid 162 [389]. It is tempting to speculate that for the XP phenotype the less truncated protein can at least partially compensate for some of the detrimental effects of the F99S mutation on the first allele, leading to a less severe phenotype. It has been shown that F99S impairs the interaction with p52, thereby reducing the ATPase activity of XPB [22]. Furthermore, it has been shown that NER activity is strongly reduced [22, 337, 389]. This is in line with the need of XPBs ATPase activity for NER [15] and with the XP symptoms. Transcriptional activity, however, seems to be much less affected [22, 337].
T119P causes TTD without mental defects. NER and transcription seem to be only slightly affected [22, 389], whereas the cellular level of TFIIH is significantly reduced [397].
The frameshift mutation fs740 causes a severe XP/CS phenotype. This mutation affects a phosphorylation site (S751), which was shown to be of importance for regulation of 5' DNA incision by XPF-ERCC1 [407]. Accordingly, fs740 leads to impaired 5' DNA incision during NER [187]. Also, it has been shown that the ATPase activity of XPB is impaired, as well as the transcriptional activity of TFIIH [408]. Additionally, it has been shown that fs740 decreases the cellular level of TFIIH [62].

1.6.4.3. p8

Three causative mutations in p8 are known, all leading to TTD (Figure 1-5) [3, 343, 344]. One mutation is a point mutation of the start codon [3], thus no p8 is expressed. The missense mutation L21P [3] interferes with the secondary structure of p8 [28]. The nonsense mutation R56stop (full-length: 71 amino acids) [3] significantly reduces the binding affinity to p52 [28]. It has been suggested that these mutations impair the interaction with p52, thereby destabilizing p52 by exposing its hydrophobic C-terminus [28]. Combined, this leads to a decreased cellular concentration of TFIIH [3, 345, 397] and might also impede XPBs function [26]. However, the transcriptional activity seems to be not affected, at least *in vitro* [26, 345]. In contrast, a recent study suggested a direct role of p8 in transcription as p8 depletion in *Drosophila* testes did not decrease TFIIH stability but showed transcriptional defects [409]. This might well explain why p8 mutations lead to TTD.

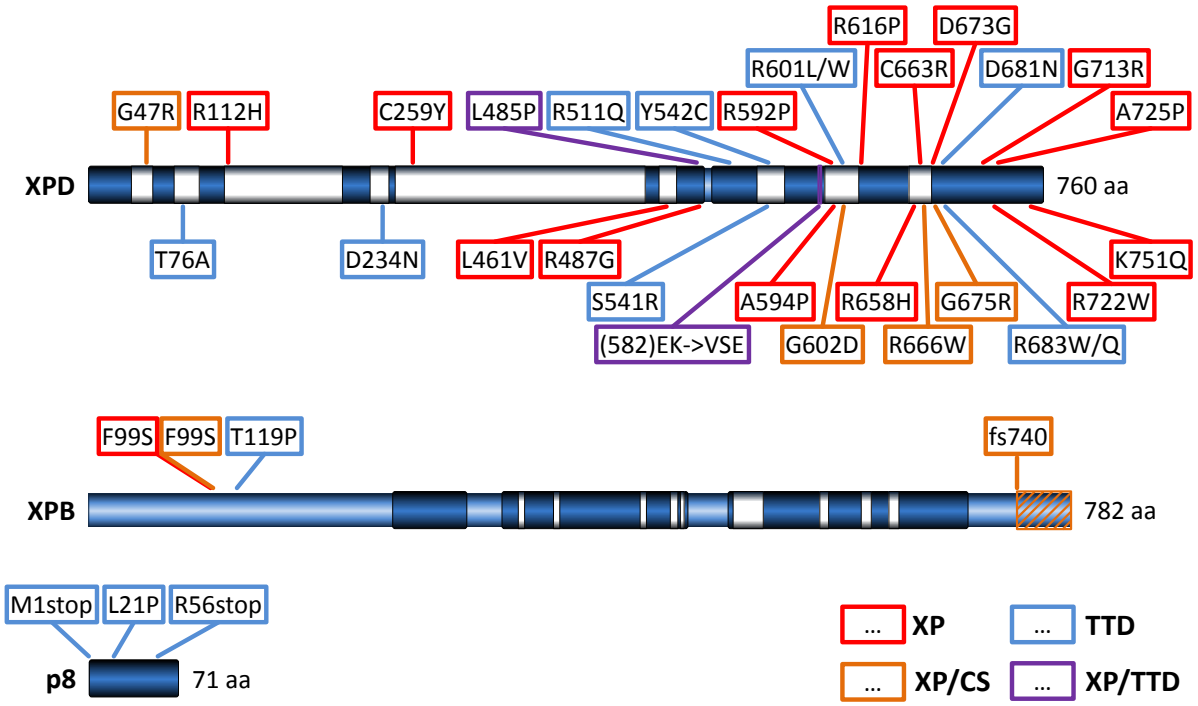


Figure 1-5. Disease related mutations of the TFIIH subunits XPD, XPB, and p8. Locations of mutation sites are indicated by lines. Colored boxes correspond to disorder as indicated. fs: frameshift; XP: xeroderma pigmentosum; TTD: trichothiodystropy; CS: Cockayne syndrome; XP/CS and XP/TTD: combined syndroms. The domain scheme corresponds to Figure 1-1.

2. Materials and Methods

2.1. Materials

2.1.1. Chemicals, dyes, and ladders

Table 2-1. Chemicals, dyes, and ladders.

Compound	Abbreviation	Supplier
1,4-piperazinediethanesulfonic acid	Pipes	Carl Roth
2-(N-morpholino)ethanesulfonic acid	MES	Sigma-Aldrich
2-propanol		Carl Roth
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	Hepes	Carl Roth
Acetic acid	HAc	Carl Roth
Acrylamide/Bis-acrylamide	AA/BAA	Carl Roth
Adenosine triphosphate	ATP	Carl Roth
Agarose		Carl Roth
Ammonium persulfate	APS	Carl Roth
Ampicillin (sodium salt)	Amp	Carl Roth
Bromophenol blue		Sigma-Aldrich
Cesium chloride	CsCl	Sigma-Aldrich
Chloramphenicol	Cam	Carl Roth
cOmplete™ Protease inhibitor cocktail		Roche Diagnostics
Coomassie Brilliant Blue G250	Coomassie G250	Carl Roth
Coomassie Brilliant Blue R250	Coomassie R250	Carl Roth
Disodium hydrogen phosphate	Na_2HPO_4	Fluka
Dithiothreitol	DTT	Carl Roth
Ethanol	EtOH	Carl Roth
Ethylenediaminetetraacetic acid	EDTA	Carl Roth
GeneRuler™ 1 kb		Thermo Scientific
Glycerol		Carl Roth
Glycine		Carl Roth
Hen egg white lysozyme	HEWL	Carl Roth
Hydrochloric acid	HCl	Carl Roth
Imidazole		Carl Roth
Isopropyl β-D-1-thiogalactopyranoside	IPTG	Carl Roth
Kanamycin sulfate	Kan	Carl Roth
L-(+)-Arabinose	Arabinose	Carl Roth
LB Broth (Lennox)	LB-medium	Carl Roth
Magnesium chloride	$MgCl_2$	Carl Roth
Midori Green Advance		NIPPON Genetics
N-Cyclohexyl-2-aminoethanesulfonic acid	Ches	Sigma-Aldrich
Nickel sulfate	$NiSO_4$	Carl Roth
Nicotinamide adenine dinucleotide disodium salt	NADH	Sigma-Aldrich
Orange G		Carl Roth
PageRuler™ Prestained		Thermo Scientific
Phenylmethylsulfonyl fluoride	PMSF	AppliChem
Phosphoenolpyruvate monosodium salt		Sigma-Aldrich

Polyethylene glycol 4000	PEG 4000	Sigma-Aldrich
Ponceau S		Carl Roth
Potassium chloride	KCl	Carl Roth
Potassium dihydrogen phosphate	KH_2PO_4	Carl Roth
Potassium iodide	KI	Sigma-Aldrich
Protino® Ni-IDA Resin	Ni-IDA	Machery-Nagel
Protino® Ni-TED Resin	Ni-TED	Machery-Nagel
Sodium acetate		Fluka
Sodium bromide	NaBr	Fluka
Sodium chloride	NaCl	Carl Roth
Sodium dodecyl sulfate	SDS	Carl Roth
Sodium hydroxide	NaOH	Carl Roth
SYPRO® Orange Protein Gel Stain	SYPRO orange	Invitrogen
Terrific Broth	TB-medium	Carl Roth
Tetramethylethylenediamine	TEMED	Carl Roth
Triptolide		Cayman Chemical
Tris(2-carboxyethyl)phosphine	TCEP	Carl Roth
Tris(hydroxymethyl)aminomethane	Tris	Carl Roth
Uranyl acetate		SERVA
Xylenecyanol FF		Fluka

Table 2-1 continued

2.1.2. Enzymes, enzyme buffers, and kits

Table 2-2. Enzymes and buffers.

Enzyme/Buffer	Abbreviation	Company
Bovine serum albumin	BSA	New England BioLabs
Deoxyadenosine triphosphate	dATP	Thermo Scientific
Deoxycytidine triphosphate	dCTP	Thermo Scientific
Deoxyguanosine triphosphate	dGTP	Thermo Scientific
Deoxythymidine triphosphate	dTTP	Thermo Scientific
Deoxyribonuclease I	DNase I	AppliChem
DpnI		New England BioLabs
HF buffer		Thermo Scientific
NEB2 buffer		New England BioLabs
Phusion DNA polymerase		Thermo Scientific
Pyruvate kinase/Lactate dehydrogenase		Sigma-Aldrich
RecA		New England BioLabs
T4 DNA polymerase		New England BioLabs
T4 ligation buffer		New England BioLabs
Taq DNA polymerase		New England BioLabs

Table 2-3. Kits.

Kit	Company
NucleoSpin® Gel and PCR Clean-Up	Machery-Nagel
NucleoSpin® Plasmid	Machery-Nagel

2.1.3. Vectors, constructs, and primer

Full-length XPB, XPD, p62, p52, p44, p34, and p8 from *Chaetomium thermophilum* were cloned individually into a pBADM-11 vector by Agnes Elias. These constructs constituted the starting point and templates for this work.

Table 2-4. Vectors.

Vector	Promoter	Tags	Cleavage site	Marker	Source
pBADM-11	araBAD	N-term[1] His_6[2]	N-term TEV[4]	Amp	EMBL[5]
pETM-11	T7-lac	N-term His_6, C-term[3] His_6	N-term TEV	Kan	EMBL
pGH	T7-lac			Amp	GeneArt

1) N-terminal
2) Hexa-histidine
3) C-terminal
4) Tobacco etch virus
5) European Molecular Biology Laboratory

The constructs cloned for this work are listed in Table 2-5, together with the respective primers. In the course of this work further constructs were attempted and their state concerning cloning, purification, and crystallization is given in the appendix (see 6.5, Table 6-2). All primers were purchased from Sigma-Aldrich.

Table 2-5. Constructs and primers.

Construct[1]	Primer sequences[2]	Template[3]	Vector[4]	Method[5]
XPBnT [pBADM-11]	5'-CTAACAGGAGGAATTAACCATGCCTCCCAAGCGAAAGGC-3' 5'-GCCTTTCGCTTGGGAGGCATGGTTAATTCCTCCTGTTAG-3'	XPB [pBADM-11]		SLIC
XPB_1-345 [pBADM-11]	5'-GAGTATGATTTCAGAAATTAAGTCGACAAGCTTGCGGCCG-3' 5'-CGGCCGCAAGCTTGTCGACTTAATTTCTGAAATCATACTC-3'	XPB [pBADM-11]		SLIC
XPB_1-345_F143S [pBADM-11]	5'-GCAGGCAACTGATTCTCTCATCACGATTGC-3' 5'-GCAATCGTGATGAGAGAATCAGTTGCCTGC-3'	XPB_1-345 [pBADM-11]		Site dir mut
XPB_116-E [pBADM-11]	5'-CTTTATTTTCAGGGCGCCAATCGCCCTTTGTGGATTGAC-3' 5'-GTCAATCCACAAAGGGCGATTGGCGCCCTGAAAATAAAG-3'	XPB [pBADM-11]		SLIC
XPB_116-768 [pBADM-11]	5'-CTTTATTTTCAGGGCGCCAATCGCCCTTTGTGGATTGAC-3' 5'-CGGCCGCAAGCTTGTCGACTTACTCGTTGTCGACCAGCGTGCG-3'	XPB [pBADM-11]	pBADM-11	SLIC
XPB_116-345 [pBADM-11]	5'-GAGTATGATTTCAGAAATTAAGTCGACAAGCTTGCGGCCG-3' 5'-CGGCCGCAAGCTTGTCGACTTAATTTCTGAAATCATACTC-3'	XPB_116-E [pBADM-11]		SLIC
XPB_116-245 [pBADM-11]	5'-GGCAAGCTGCGCGTTCAGGGATAAGTCGACAAGCTTGCGGCCG-3' 5'-CGGCCGCAAGCTTGTCGACTTATCCCTGAACGCGCAGCTTGCC-3'	XPB_116-E [pBADM-11]		SLIC
XPB_541-E [pBADM-11]	5'-CTTTATTTTCAGGGCGCCATGGAGCTCTCACAGCAGGGCC-3' 5'-GGCCCTGCTGTGAGAGCTCCATGGCGCCCTGAAAATAAAG-3'	XPB [pBADM-11]		SLIC
XPBopti[6] [pBADM-11]	5'-CTTTATTTTCAGGGCGCCATGCCGCCGAAACGCAAAGCACC-3' 5'-CGGCCGCAAGCTTGTCGACTTACTGACGCGCTGCACGAG-3'	XPBopti [pGH]	pBADM-11	SLIC
p62 [pETM-11]	5'-CTTTATTTTCAGGGCGCCATGAGCATCCCCCGCAGCCAGACCACC-3' 5'-CGGCCGCAAGCTTGTCGACTTAAACCGTGCCCTTTCCAGACACAGC-3'	p62 [pBADM-11]	pETM-11	SLIC
p62nT [pETM-11]	5'-CTTTAAGAAGGAGATATACCATGAGCATCCCCCGCAGCCAG-3' 5'-CTGGCTGCGGGGGATGCTCATGGTATATCTCCTTCTTAAAG-3'	p62 [pETM-11]		SLIC
p62_1-285 [pETM-11]	5'-CCGCGGCGAGCGCTAAACGGAGAACGACAAC-3' 5'-GTTGTCGTTCTCCGTTTAGCGCTCGCCGCGG-3'	p62 [pETM-11]		Site dir mut
p62_1-109 [pETM-11]	5'-CTCGCCGCGGCAAGGTAAAATGACCCGAATG-3' 5'-CATTCGGGTCATTTTACCTTGCCGCGGCGAG-3'	p62 [pETM-11]		Site dir mut
p62_145-290 [pETM-11]	5'-CTTTATTTTCAGGGCGCCCAAGCCCCCCGATGGTTCGACGATGCTG-3' 5'-CGGCCGCAAGCTTGTCGACTTAGTCGTTCTCCGTGTAGCGCTCGCC-3'	p62 [pETM-11]	pETM-11	SLIC
p62_290-E [pETM-11]	5'-CTTTATTTTCAGGGCGCCGACAACCCCGATCCGCTGTTC-3' 5'-GAACAGCGGATCGGGGTTGTCGGCGCCCTGAAAATAAAG-3'	p62 [pETM-11]		SLIC
p62_435-E [pETM-11]	5'-CTTTATTTTCAGGGCGCCGGGCAGGCGTCGGAAGATGC-3' 5'-GCATCTTCCGACGCCTGCCCGGCGCCCTGAAAATAAAG-3'	p62 [pETM-11]		SLIC
p62nT_435-E [pETM-11]	5'-GAAGGAGATATACCATGGGGCAGGCGTCGGAAGATGC-3' 5'-GCATCTTCCGACGCCTGCCCCATGGTATATCTCCTTC-3'	p62_435-E [pETM-11]		SLIC

p52 [pETM-11]	5'-CTTTATTTTCAGGGCGCCATGTCGATCCCCGCCGTGCC-3' 5'-CGGCCGCAAGCTTGTCGACTTAACCCTCCTTCTTCCGAC-3'	p52 [pBADM-11]	pETM-11	SLIC
p52nT [pETM-11]	5'-CTTTAAGAAGGAGATATACCATGTCGATCCCCGCCGTGCC-3' 5'-CGGCCGCAAGCTTGTCGACTTAACCCTCCTTCTTCCGAC-3'	p52 [pETM-11]		SLIC
p52nT_dL [pETM-11]	5'-CCGCATCCTCTAACGGCAACGGCGGGGGAGCCGACCCCTCCGCG-3' 5'-GCTCCCCCGCCGTTGCCGTTAGAGGATGCGGACGAGGTTAGCGTC-3'	p52nT [pETM-11]		SLIC
p52_121-E_dL [pETM-11]	5'-CCGCATCCTCTAACGGCAACGGCGGGGGAGCCGACCCCTCCGCG-3' 5'-GCTCCCCCGCCGTTGCCGTTAGAGGATGCGGACGAGGTTAGCGTC-3'	p52_121-E [pETM-11]		SLIC
p44_1-367 [pBADM-11]	5'-GGTCTCACCCTGATCTAAGCACCCACTTGGCGC-3' 5'-GCGCCAAGTGGGTGCTTAGATCAGGGTGAGACC-3'	p44 [pBADM-11]		Site dir mut
p44_1-326 [pBADM-11]	5'-CGCACACTCGCATCTGCCTAAGTCGACAAGCTTGCGGCCG-3' 5'-CGGCCGCAAGCTTGTCGACTTAGGCAGATGCGAGTGTGCG-3'	p44 [pBADM-11]		SLIC
p44_286-E [pBADM-11]	5'-CTTTATTTTCAGGGCGCCCCGCCGCCAGCCACAGCATC-3' 5'-GATGCTGTGGCTGGCGGCGGGGCGCCCTGAAAATAAAG-3'	p44 [pBADM-11]		SLIC
p44_286-367 [pBADM-11]	5'-CTGTGGTCTCACCCTGATCTAAGTCGACAAGCTTGCGGCCG-3' 5'-CGGCCGCAAGCTTGTCGACTTAGATCAGGGTGAGACCACAG-3'	p44_286-E [pBADM-11]		SLIC
p44_327-E [pBADM-11]	5'-CTTTATTTTCAGGGCGCCAGCCACGTCTCCCTCTGTGC-3' 5'-GCACAGAGGGAGACGTGGCTGGCGCCCTGAAAATAAAG-3'	p44 [pBADM-11]		SLIC

1) Residue range is indicated. E: C-terminal end; nT: no tag; dL: delta linker. Vector is given in squared brackets.
2) Upper sequence corresponds to forward, lower sequence to reverse primer.
3) Template for insert. Vector is given in squared brackets.
4) Linearized empty vector for SLIC, if applicable.
5) SLIC: sequence and ligation independent cloning; Site dir mut: site directed mutagenesis.
6) Optimized XPB gene.

Table 2-5 continued

2.1.4. Bacterial strains

Table 2-6. Bacterial strains.

Strain	Antibiotic resistance	Purpose	Company
ArcticExpress (DE3) RIL	Gentamycin	Expression	Agilent
BL21 (DE3)	-	Expression	New England BioLabs
BL21 CodonPlus (DE3) RIL	Chloramphenicol	Expression	Agilent
DH5α™	-	Cloning	NEB
Rosetta™ 2 (DE3)	Chloramphenicol	Expression	Merck Millipore
SoluBL21™	-	Expression	Amsbio

2.1.5. Culture media, LB-Agar plates, and antibiotics

LB-medium and TB-medium were used for bacterial cultures. Both media were purchased as premixed powders. 20 g of LB-medium powder was dissolved per 1 l H_2O. 50.8 g of TB-medium powder was dissolved per 1 l H_2O and supplemented with 4 ml glycerol. The media were then autoclaved at 121 °C at 2.1 bar for 15 min and antibiotics were added afterwards. The composition of both media is given below.

LB-medium

- 10 g/l Trypton
- 5 g/l yeast extract
- 5g/l NaCl

TB-medium

- 12 g/l Casein (digested enzymatically)
- 24 g/l yeast extract
- 12.54 g/l K_2HPO_4
- 2.31 g/l KH_2PO_4

LB-Agar plates were casted with the respective antibiotics directly added to the LB-Agar medium after autoclaving at 121 °C at 2.1 bar for 15 min. The composition of the LB-Agar is given below.

LB-Agar

- 30 g/l LB-medium powder
- 16 g/l Agar

The antibiotics used and the respective concentrations are listed in Table 2-7.

Table 2-7. Antibiotics.

Antibiotic	Abbreviation	Concentration (mg/l)
Ampicillin	Amp	100
Kanamycin	Kan	50
Chloramphenicol	Cam	34

2.1.6. Solutions for gel electrophoresis

All solutions listed are aqueous solutions and were prepared with ultrapure water (GenPure/xCAD, TKA). Rotiphorese® Gel 30 (37.5:1) ready-to-use AA/BAA solution was used for gel casting.

Agarose gel

- 1 % (w/v) Agarose
- 40 mM Tris
- 20 mM HAc
- 1 mM EDTA pH 8.0

6x Agarose gel loading dye

- 0.25 % (w/v) Bromophenol blue
- 0.25 % (w/v) Xylenecyanol FF
- 30 % (v/v) Glycerol

TAE buffer

- 40 mM Tris
- 20 mM HAc
- 1 mM EDTA pH 8.0

SDS-PAGE stacking gel

- 5 % (w/v) AA/BAA (37.5:1)
- 0.125 M Tris-HCl pH 6.8
- 0.1 % (w/v) SDS
- 0.1 % (w/v) APS
- 0.1 % (v/v) TEMED

SDS-PAGE separating gel

- 15 % (w/v) AA/BAA (37.5:1)
- 0.375 M Tris-HCl pH 8.8
- 0.1 % (w/v) SDS
- 0.1 % (w/v) APS
- 0.04 % (v/v) TEMED

5x SDS-PAGE loading buffer

- 0.25 M Tris-HCl pH 6.8
- 0.5 M DTT
- 10 % (w/v) SDS
- 0.5 % (w/v) Bromophenol Blue
- 50 % (v/v) Glycerol

SDS-PAGE running buffer

- 25 mM Tris
- 0.192 M Glycine
- 0.1 % (w/v) SDS

Native PAGE gel

- 3.5 % (w/v) AA/BAA (37.5:1)
- 12.5 mM Tris
- 96 mM Glycine
- 0.07 % (w/v) APS
- 0.1 % (v/v) TEMED

5x Native PAGE loading dye

- 0.1 % (w/v) Ponceau S
- 50 % (v/v) Glycerol

Tris/Glycine buffer

- 25 mM Tris
- 0.192 M Glycine

Coomassie R250 staining solution

- 0.05 % (w/v) Coomassie R250
- 50 % (v/v) EtOH
- 10 % (v/v) HAc

Coomassie R250 destaining solution

- 10 % (v/v) EtOH
- 5 % (v/v) HAc

Coomassie G250 staining solution

- 0.008 % (w/v) Coomassie G250
- 36 mM HCl

2.1.7. Standard purification buffers

All buffer components were mixed together in ultrapure water at a temperature of 4 °C. The pH value was then adjusted with hydrochloric acid. The gel filtration buffer was degassed and filtered through a 0.2 µm filter.

Lysis buffer

- 20 mM Tris-HCl pH 7.5
- 0.3 M NaCl
- 5 mM Imidazole

Elution buffer

- 20 mM Tris-HCl pH 7.5
- 0.3 M NaCl
- 0.25 M Imidazole

Gel filtration buffer

- 20 mM Tris-HCl pH 7.5
- 0.25 M NaCl

2.1.8. Crystallization, additive and Thermofluor screens

The screens used for crystallization were based on commercially available screens, but were prepared in-house with the liquid handling system LISSY 2002. These screens are listed in Table 2-8.

Table 2-8. Crystallization screens.

Screen	Company
AmSO4 Suite	Quiagen
Crystal Screen HT™	Hampton Research
Index HT™	Hampton Research
JCSG+ Suite	Quiagen
Mb Class Suite	Quiagen
Mb Class II Suite	Quiagen
Nucleix Suite	Quiagen
OptiMix™ 3	Fluidigm
OptiMix™ PEG	Fluidigm
Opti-Salts Suite	Quiagen
PEGs Suite	Quiagen
PEGs II Suite	Quiagen
pHClear Suite	Quiagen
pHClear II Suite	Quiagen
Protein Complex Suite	Quiagen
Wizard™ Classic 1 & 2 HT96	Rigaku
Wizard™ Classic 3 & 4 HT96	Rigaku

For the supplement of additives to the crystallization conditions commercially available additive screens in 96 well format were used. These screens are listed in Table 2-9.

Table 2-9. Additive screens.

Screen	Company
Additive Screen HT™	Hampton Research
Silver Bullets™	Hampton Research
Silver Bullets Bio™	Hampton Research

For the Thermofluor assays a standard screen containing 40 buffer conditions ranging from pH 4.5 to 9.8 at a concentration of 0.1 M each was used. This screen was prepared in-house with the liquid handling system LISSY 2002. The exact composition of the standard Thermofluor screen is given in the appendix (see 6.2).

2.1.9. Consumables

Table 2-10. Consumables.

Material	Specification	Company
24 well plate	SuperClear	Crystalgen
96 well plate	Crystalquick™ LP Plate, square	Greiner Bio-One
Centrifugal filter units	Amicon® Ultra - 0.5mL, Ultracel® - 3/10/30 Amicon® Ultra - 15, Ultracel® - 3/10/30/50 Vivaspin® 20, 3k/10k/30k/50k/100k MWCO	Merck Millipore Merck Millipore Sartorius Stedim Biotech
Cover slide	22 mm circular, siliconized	Jena Bioscience
EM grid	R2/4, R3/3; 2 nm carbon layer	Quantifoil Micro Tools
Filter for buffers	0.2 µm Sartolon Polyamid	Sartorius Stedim Biotech
Silicone paste	KORASILON®, mittelviskos	Kurt Obermeier
Syringe filter unit	0.2 µm Minisart®	Sartorius Stedim Biotech

2.1.10. Instruments and columns

Table 2-11. Instruments and columns.

Instrument/Column	Model	Company
Anion exchange chromatography column	Mono Q 10/100 GL	GE Healthcare
Crystallization robot	Honeybee 961/963	Genomic Solutions
Electron microscope	Tecnai™ G^2 Spirit Twin	FEI
FPLC systems	ÄKTA pure ÄKTA purifier ÄKTA xpress	GE Healthcare GE Healthcare GE Healthcare
Gel scanner	Odyssey® Scanjet G2710	LI-COR Hewlett-Packard
Glass chromatography column	Econo-Column® 2.5 x 20 cm	Bio-Rad
Light microscope camera	AxioCam MRc	Zeiss
Light microscopes	Stemi 2000 SteREO Discovery.V12	Zeiss Zeiss
Liquid handling system	LISSY 2002	Zinsser Analytic
Microplate reader	CLARIOstar®	BMG LABTECH
PCR cyclers	GeneAmp® PCR System 9700 Mastercycler epgradient S Mastercycler pro S	Applied Biosystems Eppendorf Eppendorf
pH meter	BlueLine 14 pH	Schott
Plasma cleaner	PDC-002	Harrick Plasma
RT-PCR cycler	Stratagene Mx3005P	Agilent Technologies
Size exclusion chromatography columns	HiLoad 26/60 Superdex 200 pg HiLoad 16/600 Superdex 200 pg Superdex 200 10/300 GL Superose 6 10/300 GL Superose 6 Increase 3.2/300	GE Healthcare GE Healthcare GE Healthcare GE Healthcare GE Healthcare
Spectrophotometers	BioPhotometer NanoDrop® ND 1000	Eppendorf Peqlab
Vitrification device	Vitrobot™ Mark IV	FEI
X-ray generator	MicroMax™-007HF	Rigaku

2.1.11. Software

Table 2-12. Software.

Software	Version	Reference
ADXV	1.9.8	Andrew Arvai
Aimless	0.5.28	[410]
ApE	2.0.37	M. Wayne Davis
Buccaneer	1.6.3	[411]
CCP4 suite	7.0.021	[412]
UCSF Chimera	1.11.2	[413]
Coot	0.8.8	[414]
CTFFIND3		[415]
EMAN2		[416]
HySS		[417]
Leginon		[418]
Phaser	2.6.1	[419]
Phenix	1.9-1692	[420]
Pointless	1.10.26	[421]
PyMOL	1.7.6.0	[422]
Refmac5	5.8.0155	[423]
RESOLVE		[424, 425]
ShelxC/D/E	C: 2006/1 D: 2013/2 E: 2016/2	[426]
SPARX		[427]
Spider	14	[428]
XDS	2016/11	[429]

2.1.12. Web sources and databases

Table 2-13. Web sources and databases.

Web source	URL	Reference
ClustalW2	http://www.ebi.ac.uk/Tools/msa/clustalw2/	[430]
COBALT	https://www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi?CMD=Web	[431]
EMBOSS Needle	http://www.ebi.ac.uk/Tools/psa/emboss_needle/	[432]
ExPASy ProtParam	http://web.expasy.org/protparam/	[433]
I-TASSER	http://zhanglab.ccmb.med.umich.edu/I-TASSER/	[434]
Molprobity	http://molprobity.biochem.duke.edu	[435]
PDBe EMDB	http://www.ebi.ac.uk/pdbe/emdb/index.html/	
Phyre2	http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index	[436]
PubMed	https://www.ncbi.nlm.nih.gov/pubmed	
RaptorX	http://raptorx.uchicago.edu	[437]
RCSB Protein Data Bank	http://www.rcsb.org	[438]
UniProtKB	http://www.uniprot.org	[439]

2.2. Methods

2.2.1. Cloning

2.2.1.1. Sequence and ligation independent cloning

Sequence and ligation independent cloning (SLIC) utilizes single stranded DNA (ssDNA) overhangs which are generated by an exonuclease in the linearized vector and insert, and the recombination via these ssDNA overhangs [440]. Two slightly different approaches were applied.

For the cloning of an insert from one vector into another, or deletion of parts at both sides of the insert within the same vector, the following procedure was applied. Vector and insert where linearized and amplified separately via polymerase chain reaction (PCR) with an appropriate primer pair. Here, the vectors (pETM-11, pBADM-11) where linearized with the primer pair 5'-GTCGACAAGCTTGCGGCCG-3' (forward primer) and 5'-GGCGCCCTGAAAATAAAG-3' (reverse primer). The forward primer for insert amplification consisted of the sequence with reverse complementarity to the vector reverse primer (5'-CTTTATTTTCAGGGCGCC) and an approximately 20 nt long sequence corresponding to the 5'-end of the insert. The reverse primer for insert amplification consisted of the sequence with reverse complementarity to the vector forward primer (5'-CGGCCGCAAGCTTGTCGAC) and an approximately 20 nt long sequence corresponding to the reverse complement of the 3'-end of the insert.

For deletion of a sequence within a gene or the removal of the sequence encoding the His-tag and TEV cleavage site, the following procedure was applied. The vector containing the insert was linearized via PCR with an appropriate primer pair. The 5'-end of the forward primer consisted of about 20 nt and corresponded to the vector/insert sequence directly 5' to the deletion. The 3'-end of the forward primer consisted of about 20 nt and corresponded to the vector/insert sequence directly 3' to the deletion. The reverse primer corresponded to the reverse complement of the forward primer. The reaction mixture and PCR protocol are given below.

Reaction mixture (linearization)

- 100 ng vector/template plasmid
- 0.5 µl Phusion polymerase (2 U/µl)
- 1 µl forward primer (10 µM)
- 1 µl reverse primer (10 µM)
- 1 µl dNTP (10 µM each)
- 10 µl HF buffer (5x)
- H_2O (fill to 50 µl)

PCR protocol

98 °C	2 min	
98 °C	30 s	32 cycles
58 °C	40 s	
72 °C	2 min 30 s	
72 °C	10 min	

The linearized vector was incubated with 20 U DpnI over night at 37 °C to digest the original template. Linearized vector and insert were then purified with the NucleoSpin® Gel and PCR Clean-Up kit and subjected to the exonuclease activity of T4 polymerase for 30 min at room temperature to generate 5' overhangs. The reaction was stopped by addition of 1 mM dCTP. The reaction mixture is given below.

Reaction mixture (T4 treatment)

- 500 ng vector/insert
- 0.4 µl T4 polymerase (3 U/µl)
- 2 µl NEB2 buffer (10x)
- 2 µl BSA (10x)
- H_2O (fill to 20 µl)

For annealing, the T4 polymerase treated vector and (if applicable) insert were mixed together and incubated for 30 min at room temperature. Afterwards, 15 µl of the mixture was transformed into DH5α cells (see 2.2.2). The reaction mixture is given below.

Reaction mixture (annealing)

- 100 ng vector
- 200 ng insert (if applicable)
- 1 µl RecA (20 ng/µl)
- 2 µl T4 ligation buffer (10x)
- H_2O (fill to 20 µl)

Single colonies were picked and checked via colony PCR. For the pBADM-11 vector the primers 5'-GACGCTTTTTATCGCAACTCTC-3' (forward) and 5'-CTCATCCGCCAAAACAGCCAAG-3' (reverse) were used. For the pETM-11 vector T7 promoter (5'-TAATACGACTCACTATAGGG-3') and T7 terminator (5'-GCTAGTTATTGCTCAGCGG-3') primers were used. The reaction mixture and PCR protocol are given below.

Reaction mixture (Colony PCR)

- 0.5 µl Taq polymerase (5 U/µl)
- 1 µl forward primer (10 µM)
- 1 µl reverse primer (10 µM)
- 1 µl dNTP (10 µM each)
- 2 µl NEB2 buffer (10x)
- 14.5 µl H_2O
- 1 suspended colony

PCR protocol

95 °C	2 min	
95 °C	20 s	30 cycles
55 °C	30 s	
72 °C	2 min	
72 °C	10 min	

The size of the amplified PCR product was analyzed via agarose gel electrophoresis. Positive single colonies were further grown in 10 ml LB-medium over night at 37 °C and 200 rpm, supplemented with the appropriate antibiotics. The plasmid DNA was purified from the cell culture with the NucleoSpin® Plasmid kit, and the product was checked by sequencing (Eurofins Genomics).

2.2.1.2. Site directed mutagenesis

For introduction of single point mutations site directed mutagenesis was applied [441, 442]. Primers with a length of about 30 nt were used with the desired base changes in the center but otherwise completely corresponding to the insert sequence. Forward and reverse primers were reverse complementary to each other. To avoid the formation of primer dimers, two successive rounds of PCR were applied [441]. First, two reactions in separate tubes were performed, one with the forward primer, and one with the reverse primer. Then, the PCR products were combined and the second reaction was performed. The reaction mixture and PCR protocol are given below.

Reaction mixture (forward and reverse primer separately)

- 100 ng template plasmid
- 0.5 µl Phusion polymerase (2 U/µl)
- 1 µl forward/reverse primer (10 µM)
- 1 µl dNTP (10 µM each)
- 5 µl HF buffer (5x)
- H_2O (fill to 25 µl)

PCR protocol

98 °C	30 s	
98 °C	10 s	8 cycles (forward/reverse primer separately)
65 °C	30 s	25 cycles (combined reaction)
72 °C	1 min 30 s	
72 °C	5 min	

The PCR product from the combined reaction was treated with 20 U of DpnI over night at 37 °C to digest the original template. 10 µl of the reaction mixture was transformed into DH5α cells. Single colonies were grown in 10 ml LB-medium over night at 37 °C and 200 rpm, supplemented with appropriate antibiotics. The plasmid DNA was purified from the cell culture with the NucleoSpin® Plasmid kit, and the product was analyzed by sequencing (Eurofins Genomics).

2.2.2. Transformation

Chemically competent cells were stored at -80 °C. For transformation, cells were thawed for 30 min on ice. 100 ng of plasmid DNA was added and incubated for 30 min on ice. Afterwards, cells were subjected to 42 °C for 90 s. Then, 300 µl of LB-medium was added and the cells were shaken at 500 rpm at 37 °C for 1h. 150 µl of the cell suspension was plated on a LB-Agar plate with the appropriate antibiotics. The plate was incubated over night at 37 °C and stored at 4 °C afterwards.

For co-expression studies, two plasmids were co-transformed. The steps were similar as described above, but after thawing of the cells 100 ng of each plasmid DNA was added. Also, 250 µl of the cell suspension was plated on a LB-Agar plate with the appropriate antibiotics.

2.2.3. Gel electrophoresis

2.2.3.1. Agarose gel electrophoresis

Agarose gel electrophoresis was used to check the size of the amplified PCR products. To stain the DNA, 3 µl of Midori Green Advance was added per 50 ml of agarose gel solution during gel casting. The samples were mixed with 6x agarose gel loading dye in a 5:1 ratio and the mixture was loaded on the gel. As size marker GeneRuler™ 1 kb was used. Gel electrophoresis was performed for 40 min at 110 V. The DNA fragments were visualized via fluorescence of the Midori Green Advance with a Molecular Imager® Gel Doc™ XR system (Bio-Rad).

2.2.3.2. SDS-PAGE

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) separates proteins by their molecular weight, independent of their amino acid sequence [443]. The proteins are denatured by addition of SDS and heating. The detergent SDS is negatively charged and binds to the denatured protein chain. The resulting negative charge of the SDS-protein complex correlates with the molecular weight of the protein. The mobility through the gel then depends on the size of the protein, smaller proteins move faster than bigger ones. Protein samples were mixed with 5x SDS-PAGE loading buffer in a 4:1 ratio and incubated at 95 °C for 5 min. The sample mixture was then loaded on the gel. As size marker PageRuler™ Prestained was used. Gel electrophoresis was performed for 50 min at 200 V. Gels were either stained with Coomassie R250 staining solution and destained with Coomassie R250 destaining solution, or stained with Coomassie G250 staining solution and destained with water.

2.2.3.3. Native PAGE

Native polyacrylamide gel electrophoresis (Native PAGE) is performed under none denaturing conditions and the running behavior of a sample depends on its charge, size, conformation and oligomeric state. It can thus be used to detect protein-protein or protein-nucleic acids interactions. For the analysis of protein-protein interactions 50 pmol of protein was used. To determine if an interaction can be observed, the respective proteins were mixed together and incubated for 30 min on ice prior to gel electrophoresis. The respective single proteins were used as control. To compensate for differences in the buffer conditions, equal amounts of buffer from the lacking protein components were added to the respective samples. The samples were mixed with 5x Native PAGE loading dye in a 5:1 ratio and loaded on the gel. Gel electrophoresis was performed at 4 °C for 90 min at 90 V. As running buffer Tris/Glycine buffer, cooled to 4 °C, was used. Gels were stained with Coomassie G250 staining solution and destained with water. Pictures of the gels were taken with a cell phone camera.

2.2.4. Chromatography

2.2.4.1. Immobilized metal affinity chromatography

Immobilized metal affinity chromatography (IMAC) exploits the affinity of histidine side chains for divalent cations like Ni^{2+} or Co^{2+} [444, 445]. To separate the proteins of interest from other host proteins during purification, the desired proteins were expressed with an N-terminal hexa-histidine tag (His-tag). These proteins were then immobilized on a solid phase via the His-tag. As solid phase either Ni-TED, or Ni-IDA was used. Ni-TED consists of silica beads with tris-carboxymethyl ethylene diamine (TED) attached. TED is a chelator of Ni^{2+} ions,

occupying five of the six available ligand binding sites [446]. Ni-IDA consists of silica beads with iminodiacetic acid (IDA) attached. IDA is also a chelator of Ni^{2+} ions, yet occupying only three of the six available ligand binding sites [446]. Therefore, Ni-IDA has a higher binding capacity than Ni-TED, whereas the latter has a higher specificity. The histidine sidechains of the His-tag occupy the remaining ligand binding sites of the Ni^{2+} ion and the His-tagged protein is thereby adsorbed to the solid phase. The host proteins remain in the flow through and can thereby be separated from the target protein. The bound proteins were then eluted with a buffer containing high imidazole concentrations, competing with the His-tag for the Ni^{2+} binding sites.

2.2.4.2. Size exclusion chromatography

Size exclusion chromatography (SEC), also referred to as gel filtration, separates particles by their hydrodynamic radius [447, 448]. As a particle solution passes through a porous column material, particles with a smaller hydrodynamic radius have a higher likelihood of entering porous channels, hence they have to pass a larger volume than bigger particles. As the molecular weight correlates with the hydrodynamic radius, larger proteins usually elute earlier than smaller proteins and can thus be separated from each other. As SEC usually does not disrupt protein-protein interactions, it can also be used to determine oligomeric states or monitor complex formation through the analysis of the elution profile. The protein solution was applied to a column packed with a composition of dextran and cross-linked agarose under constant flow rate using an FPLC system. The elution profile was monitored via the absorption at 280 nm (A280). Depending on the FPLC system used, absorption at 260 nm (A260) was monitored as well. The eluate was collected in fixed sized fractions.

2.2.4.3. Ion exchange chromatography

Ion exchange chromatography (IEX) separates molecules by their surface charge [449]. A highly charged stationary phase is used and proteins with an opposite net charge bind to the stationary phase. Depending on the charge of the stationary phase, anion exchange (positively charged) (AIEX) and cation exchange (negatively charged) chromatography can be distinguished. According to the net surface charge, the binding affinity varies between different proteins. The bound proteins can then be eluted by increasing amounts of counter ions. More weakly bound proteins elute at lower counter ion concentrations and can thus be separated from tighter bound proteins. Anion exchange chromatography was applied using a FPLC system. The charged group of the column material was a quaternary ammonium cation. To elute the bound proteins, a linear gradient of an increasing NaCl concentration was used. The elution profile was monitored via the A280 and, if applicable, A260. The eluate was collected in fixed sized fractions.

2.2.5. Standard expression procedure

After transformation, a single colony was picked from the LB-Agar plate and transferred to 200 ml of LB-medium supplemented with the appropriate antibiotics. This pre-culture was incubated over night at 37 °C and 200 rpm. The LB-medium main culture was supplemented with the appropriate antibiotics and inoculated with 10 ml pre-culture per 1 l main culture. The following steps are similar for all bacterial strains, except ArcticExpress (DE3) RIL. Therefore expression in ArcticExpress (DE3) RIL and in other bacterial strains will be described separately. Overexpression was monitored by taking samples from the culture at different time points before and after induction. The optical density at 600 nm (OD_{600}) of the samples was measured and the cells were pelleted by centrifugation at 25000 x g for 1 min. The supernatant was discarded and the concentration was normalized to 1 OD_{600}/ml by resuspending the pellet in an appropriate amount of water. The samples were then checked via SDS-PAGE.

2.2.5.1. ArcticExpress (DE3) RIL

The inoculated main culture was incubated at 30 °C and 200 rpm. Culture growth was monitored by measuring OD_{600}. At an OD_{600} of 0.8 the main culture was cooled to 11 °C and the expression of the cold-adapted chaperonins Cpn10 and Cpn60 were induced with 0.5 mM IPTG. Depending on the expression vector(s) used, the main culture was additionally induced with 0.03 % (w/v) Arabinose. Expression was performed at 11 °C at 200 rpm for 24 h.

2.2.5.2. All other bacterial expression strains

The inoculated main culture was incubated at 37 °C and 200 rpm. Culture growth was monitored by measuring OD_{600}. At an OD_{600} of 0.6 the main culture was cooled to 15 °C and induced with 0.5 mM IPTG and/or 0.03 % (w/v) Arabinose, according to the expression vector(s) used. Expression was performed at 15 °C at 200 rpm for 21 h.

2.2.6. Standard purification procedure

In most of the cases the following procedure was applied for protein purification. Deviations from this procedure are explicitly mentioned in the respective sections. For the standard purification a two-step purification protocol was applied. The first step was IMAC, followed by SEC. All steps were performed at 4 °C with buffers cooled to 4 °C. The buffers used for the standard purification procedure are listed in section 2.1.7.

After expression, the bacteria were harvested by centrifugation at 6056 x g for 12 min at 4 °C. The supernatant was decanted and discarded. The obtained cell pellet was either stored at -80 °C or directly subjected to further purification steps.

The cell pellet was resuspended in about 5 ml lysis buffer per 1 g of cell pellet. The suspension was supplemented with 4 U/ml of DNase I. Cells were lysed by two rounds with a Microfluidizer® (Microfluidics). The lysate was centrifuged at 33768 x g for 1 h at 4 °C. The supernatant was decanted and supplemented with a spatula tip of PMSF. About 2-3 g Ni-TED beads were added to ~300-500 ml of supernatant and the mixture was incubated for 1.5 h on a magnetic stirrer at 150 rpm. The supernatant was removed and the beads were washed 2-3 times with lysis buffer. The beads were then transferred to a gravity flow glass chromatography column. After washing with 1 column volume of lysis buffer, the His-tagged protein was eluted with elution buffer. 10 ml elution fractions were collected and inspected via SDS-PAGE. Elution fractions containing the desired protein(s) were pooled and concentrated to 5 ml using a centrifugal filter unit and subjected to SEC, usually using a HiLoad 16/600 Superdex 200 pg column. Elution fractions of 2 ml were collected and inspected via SDS-PAGE. Fractions containing the desired protein(s) were pooled and concentrated using a centrifugal filter unit. The resulting protein solution was flash frozen in liquid nitrogen and stored at -80 °C.

2.2.7. Thermofluor

Thermofluor is a fluorescence-based thermal shift assay, which monitors the unfolding of a protein with increasing temperature. It can be used to screen for buffer conditions that stabilize a protein of interest, by comparing the shift of the melting temperature utilizing different conditions [450, 451]. The method makes use of a dye with a low fluorescence quantum yield in aqueous solution, but increased fluorescence in a hydrophobic environment and can therefore be used to analyze the stability of soluble proteins. By increasing the temperature, a protein begins to unfold leading to an exposure of its hydrophobic amino acids. These hydrophobic amino acids interact with the dye, resulting in an increase in fluorescence [450]. The Thermofluor assay was performed in a 96 well PCR plate (Greiner Bio-One) with SYPRO orange as fluorescence dye. In each well 1 µl protein at a concentration of 5-12.5 mg/ml was mixed with 23 µl of a certain buffer condition and 1 µl of 5x SYPRO orange. The temperature of these mixtures was increased step-wise by 1 °C in a RT-PCR cycler, starting at 25 °C and ending at 95 °C. At every temperature step the fluorescence for each well was monitored. The fluorescence was plotted against the temperature, usually resulting in a curve with a steep increase reaching a maximum. The inflection point of this increase defined the melting temperature. For direct comparison of different buffer conditions the curves were normalized.

2.2.8. ATPase assay

For the biochemical characterization of XPB, its ATPase activity was investigated utilizing a coupled enzyme assay. The ATPase activity of an enzyme results in the conversion of ATP to ADP. ADP forms the cofactor for pyruvate kinase, which then converts phosphoenolpyruvate to pyruvate. Pyruvate in turn is the substrate of lactate dehydrogenase, catalyzing the reaction to lactate, thereby converting the cofactor NADH to NAD+. Accordingly, the decrease of the NADH concentration is correlated to the hydrolysis of ATP. As NADH absorbs at 340 nm, the ATPase activity can be determined by monitoring the decrease in absorption at that wavelength.

ATPase assays were performed at 37 °C in a microplate reader. All components of the reaction mixture except ATP were pre-incubated at 37 °C until a stable baseline was reached. The reaction was then started by the addition of ATP. The composition of the reaction mixture is given below. To test the influence of DNA on the ATPase activity, 2 µM of different DNA substrates were used in the assay.

Reaction mixture

- 0.25-1 µM protein
- 2 U pyruvate kinase/lactate dehydrogenase
- 2 mM phosphoenolpyruvate
- 0.3 mM NADH
- 2 mM ATP
- 20 mM Tris-HCl pH 8.0
- 10 mM KCl
- 1 mM $MgCl_2$
- 1 mM TCEP
- 2 µM DNA (if applicable)

2.2.9. X-ray crystallography

X-ray crystallography is a technique to determine the three dimensional structure of molecules at atomic resolution. It is in principle suitable for molecules of any size and can therefore be used to investigate proteins. A prerequisite is the crystallization of the respective proteins. Single protein crystals are then exposed to x-rays, giving rise to a protein diffraction pattern. The diffraction pattern is recorded with a detector and processed afterwards. X-ray crystallography yields an electron density map, which represents the distribution of electrons within a protein crystal. This electron density map is interpreted by fitting amino acids into it according to their shape, resulting in a structure model. This model is then iteratively improved and refined.

2.2.9.1. Protein crystallization

In order to investigate proteins via x-ray crystallography, they have to be crystallized first. To retain the native conformation of the protein, it has to be crystallized in its aqueous environment. The vapor diffusion method was applied for crystallization [452, 453]. A protein solution was usually mixed in a 1:1 ratio with a precipitant solution (protein drop). This mixture was equilibrated against an excess of the same precipitant solution (reservoir). The process of vapor diffusion requires that the equilibration process takes place in a closed system and therefore each individual set-up was sealed. As the precipitant was diluted in the protein drop compared to the reservoir, vapor diffusion of water from the protein solution to the reservoir takes place, until an equilibrium is reached. Thus, water is removed from the protein drop, increasing the concentration of protein and precipitant, resulting in a metastable solution. Eventually, the protein forms nuclei, resulting in crystal formation [452-456]. This process is depicted in Figure 2-1.

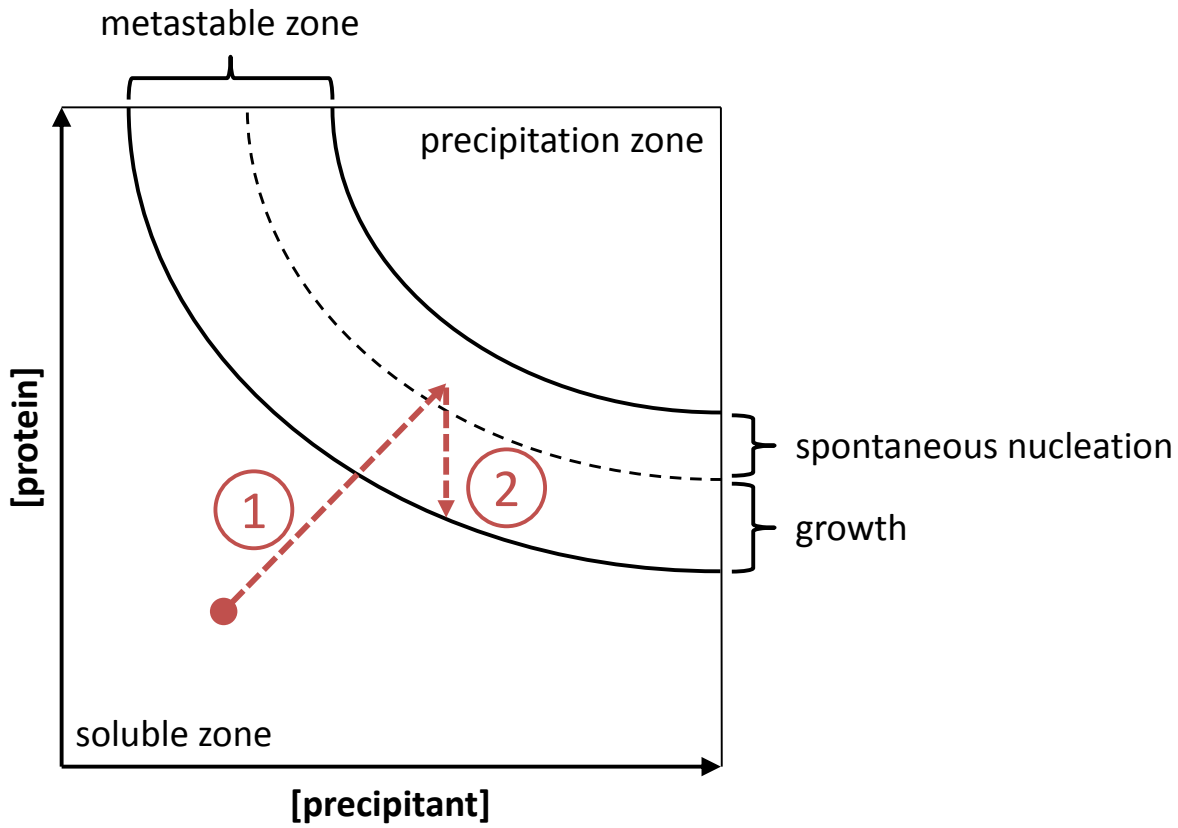


Figure 2-1. Phase diagram of a protein/precipitant solution. The process of crystal formation during a vapor diffusion experiment is indicated by dashed arrows. Upon vapor diffusion of water from the protein drop to the reservoir, the protein and precipitant concentration increases, eventually reaching the zone of spontaneous nucleation (1). Nucleation leads to the decrease of the protein concentration, shifting into the growth zone, resulting in the formation of crystals from the nuclei (2).

All solutions used for crystallization were filtered through 0.2 µm filters and crystallization was generally performed in two ways. The initial screening for suitable crystallization conditions

was performed via sitting drop vapor diffusion in 96 well plates with the Honeybee 961/963 crystallization robot using commercial crystallization screens (Table 2-8). Here, 0.3 µl of protein solution was mixed with 0.3 µl of precipitant solution and equilibrated against 40 µl of the precipitant solution. Conditions leading to crystals were then fine screened by variation of different parameters, like the pH value, precipitant concentration, drop size, temperature, or supplement of additives. On the one hand this optimization process was performed in 96 well plates with the crystallization robot. In this case, custom crystallization screens were prepared with the LISSY 2002 liquid handling system. On the other hand manual crystallization was performed utilizing the hanging drop vapor diffusion method in 24 well plates. Here, 1 µl of protein solution was mixed with 1 µl of precipitant solution and equilibrated against 1 ml of the precipitant solution. The protein drop was pipetted on a cover slide, which inverted, thus facing the well containing the reservoir. The well was sealed by applying silicone paste. Crystallization plates were usually kept at 20 °C.

To improve crystal growth and quality, additive screens in the 96 well format were pursued (Table 2-9). One promising prior crystallization condition was chosen and 45 µl of the corresponding reservoir solution was applied to each reservoir of the 96 well plate. Next, 5 µl of a 10x additive stock solution was added to each reservoir and mixed. 0.3 µl from this mixture were transferred to the sitting shelf and subsequently supplemented with 0.3 µl of the protein solution.

In order to control and improve nucleation of crystals, seeding was applied. Seeding circumvents the necessity of reaching the zone of spontaneous nucleation (Figure 2-1), thereby enabling to shift the crystallization process into the more favorable growth zone [453, 457]. This approach might then result in fewer and larger crystals. In this work, microseeding was applied. Crystals obtained from initial crystallization trials were transferred to ~1 µl of the reservoir solution, crushed by vortexing and sonication, and diluted with reservoir solution according to the amount of seeds in the solution. 1 µl of this solution was added to 40 µl of the protein solution directly prior to crystallization.

Crystals were harvested with cryo loops (CryoLoop™, Hampton Research) and transferred to a cryo protectant solution corresponding to the reservoir composition, usually supplemented with 25 % (v/v) glycerol. Crystals were flash frozen in liquid nitrogen and exposed to synchrotron radiation at the ESRF (Grenoble, France), PETRAIII (Hamburg, Germany) or BESSYII (Berlin, Germany) for data acquisition.

2.2.9.2. Diffraction, phase problem and SAD phasing

Exposure of a protein crystal with x-rays yields a characteristic diffraction pattern. Each diffraction spot can be assigned to a triplet of integer values (hkl). These indices correspond to a specific family of imaginary parallel planes within the crystal lattice, at which the x-rays

have been reflected [452]. The diffraction experiment yields an intensity I(hkl) for every recorded reflection.

The electron density at a point xyz within the unit cell of a crystal can be calculated via the following equation:

$$\rho(xyz) = \frac{1}{V}\sum_{hkl} F(hkl)e^{-2\pi i(hx+ky+lz)} \quad (1)$$

V denotes the unit cell volume, F(hkl) is the structure factor of reflection hkl. By plotting ρ(xyz) against x, y, and z, an electron density map is obtained. The structure factor F(hkl) is a complex number and can be represented in the plane of complex numbers (Figure 2-2).

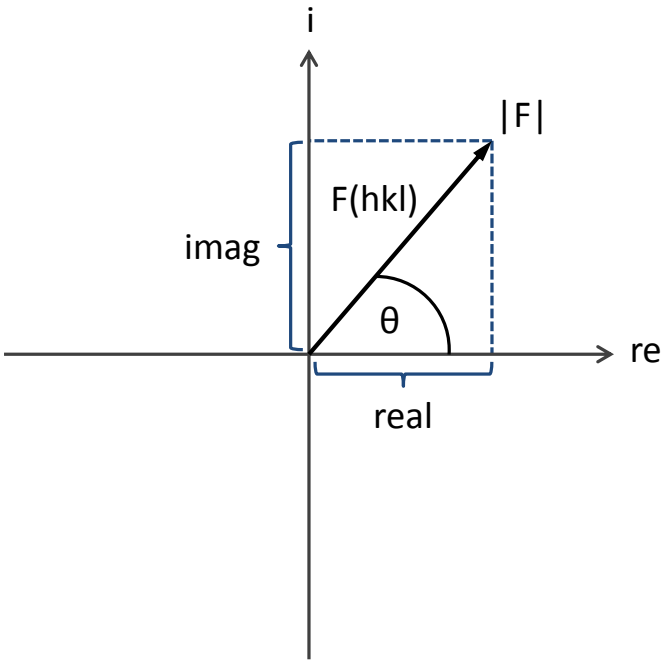


Figure 2-2. Structure factor representation in the plane of complex numbers. A structure factor F(hkl) is a complex number and can be represented as a vector in the plane of complex numbers. The real and imaginary parts correspond to the projections on the axes as indicated. The amplitude |F| of the structure factor corresponds to the length of the vector, the phase θ to the angle with the real axis. i: imaginary axis; re: real axis.

Thereafter, F(hkl) consists of an amplitude |F| and a phase angle θ. The measured intensity I(hkl) relates to F(hkl) via the complex conjugate:

$$I(hkl) = FF^*(hkl) \quad (2)$$

F* is the complex conjugate of F. From equation (2) follows that |F| is proportional to the square root of I. Thus, |F| can be retrieved from the diffraction experiment. However, for

solving equation (1) the phase angle θ from F(hkl) is also needed. This information cannot be retrieved and is lost during the diffraction experiment. This constitutes the phase problem in x-ray crystallography [453, 458].

Different methods exist to obtain the missing phase information. One such method is single wavelength anomalous diffraction (SAD) [459]. As this method is of relevance for this thesis, it will be briefly discussed.

SAD phasing is an experimental approach for de novo phase determination. The popularity of this method has increased more and more over the last 20 years, and nowadays it is the most commonly used experimental phasing approach [460]. SAD is based on the fact that atoms are capable of absorbing photons at certain energy levels. If an atom is exposed to x-rays with an energy at, or slightly above, such an absorption edge, the scattering obtains a certain anomalous contribution. The amount of this contribution increases with the weight of the anomalous scattering atom. The anomalous scattering consists of a real component f' and an imaginary component f'' [452, 453, 456]. These scattering contributions are shown exemplarily for selenium in Figure 2-3.

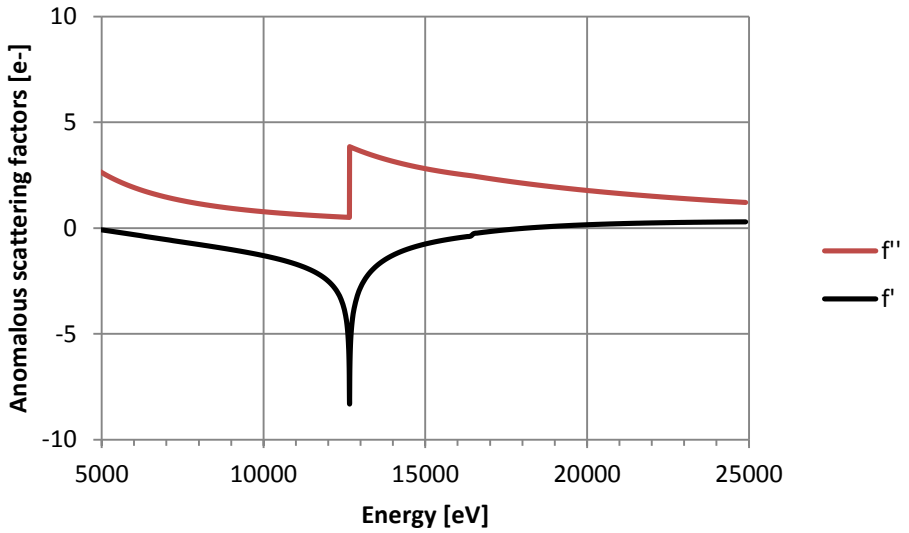


Figure 2-3. Anomalous scattering of selenium. The components f' and f'' of the anomalous scattering of selenium are plotted against the x-ray energy. The peak of f'' at around 12.66 keV corresponds to the K-edge of selenium and constitutes a suitable energy for SAD data collection with an anomalous contribution f'' of about 4 e^-. The plot was generated with data from [461].

SAD exploits the fact that the occurrence of anomalous scattering leads to the breakdown of Friedel's law. This breakdown is caused by f''. As f'' is an imaginary part, it is delayed compared to the real scattering part. This delay can be visualized in the complex plane by a counterclockwise rotation by 90°. This is depicted in Figure 2-4, demonstrating how this results in different amplitudes for the Friedel pairs [453, 458].

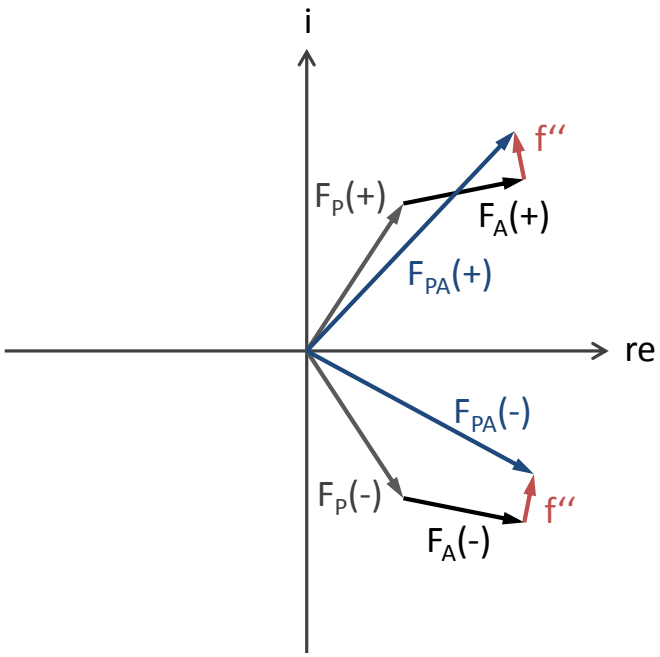


Figure 2-4. Breakdown of Friedel's law via anomalous scattering. The imaginary component f'' of the anomalous scattering is delayed relative to the real scattering by a counterclockwise rotation by 90°. This results in different amplitudes for $F_{PA}(+)$ and $F_{PA}(-)$. Thus, Friedel's law no longer holds true. F_P: scattering contribution of the protein without anomalous scatterers; F_A: non-anomalous scattering of the anomalous scatterer, combined with the real part of the anomalous scattering; F_{PA}: scattering from protein and anomalous scatterer; (+) and (-) correspond to Friedel pairs (hkl) and (-h-k-l), respectively.

The Friedel pair differences can be used to determine the positions of the anomalous scatterers [452, 453]. For this, a Patterson map is calculated with these differences as Fourier coefficients:

$$P(uvw) = \frac{1}{V}\sum_{hkl}\left||F(hkl)| - \left|F(\overline{hkl})\right|\right|^2 \cos\left(2\pi(hu + kv + lw)\right) \qquad (3)$$

V denotes the unit cell volume; u, v, and w are relative unit cell coordinates. From the resulting peaks in the Patterson map the positions of the anomalous scatterers can be determined. Knowing the positions of these scatterers a Harker construction can be performed, yielding two possible solutions for the phase angle θ (Figure 2-5) [453, 458].

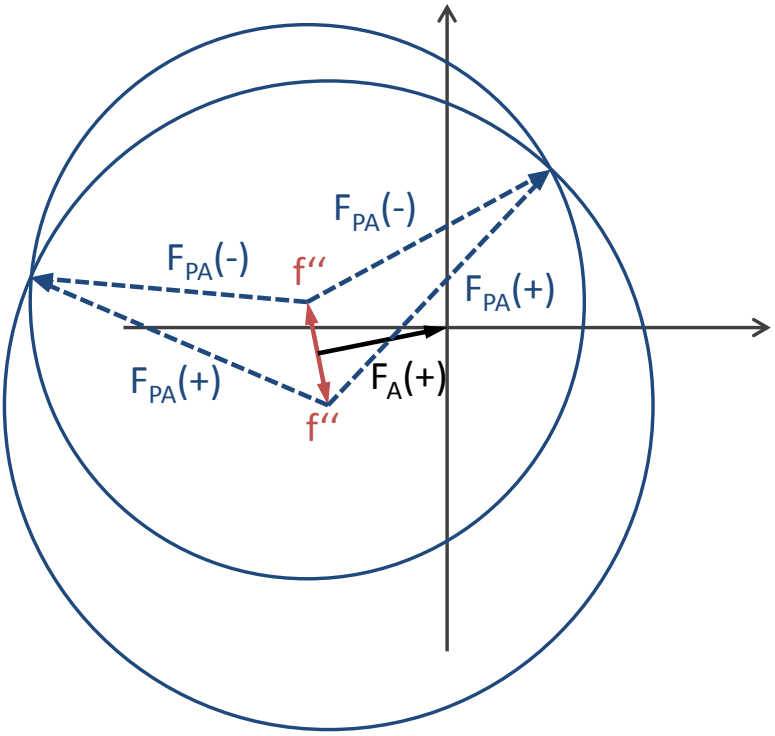


Figure 2-5. Harker construction for single wavelength anomalous diffraction (SAD). Demonstrated is the construction for the $F_{PA}(+)$ mate. The intersection points of the two circles constitute two possible solutions for the phase θ of $F_{PA}(+)$. The labels correspond to Figure 2-4.

To break this phase ambiguity two main steps are applied. First, an initial set of phases is derived for the structure factors, taking into account the phases of the anomalous scatterers. Second, these initial phases are used as input for density modification approaches. In an iterative approach the phases are improved until an interpretable electron density map is obtained [462].

As already stated, SAD depends on significant anomalous scattering. However, anomalous scattering from proteins is usually very weak, as they mainly consist of the light elements C, N, and H. S yields significant anomalous scattering, albeit it is still relatively weak. However, the corresponding x-ray energy is not as easily accessible as for example for Se. Furthermore, it depends on the presence of cysteine or methionine in the protein. To obtain strong anomalous scattering, heavy atoms can be introduced into the crystal. In this work, cesium was used as anomalous scatterer. The focus was on the efficient introduction of cesium into crystals in order to derive a general, easy-to-use strategy to overcome the phase problem via SAD.

2.2.10. Electron microscopy

Electron microscopy (EM) is another technique to obtain structural information and especially suited for the structural investigation of large protein complexes. Recent advances in EM, particularly the invention of direct electron detectors [463, 464], made it possible to obtain

near atomic resolution maps with EM [465, 466]. EM is especially powerful in combination with x-ray crystallography. Fitting crystal structures of atomic resolution into near atomic resolution EM maps of a protein complex can yield a pseudo-atomic model of this protein complex [467, 468]. In this work transmission electron microscopy has been applied. In general, a protein sample is deposited on a grid, so that a single layer of particles, well separated from each other, is obtained. These particles are then exposed to an electron beam. The electrons pass through the sample and are scattered [467]. The scattered electrons are focused by electromagnetic lenses and recorded with a detector [467, 469]. This results in projection images of the particles on the detector. As the electrons pass through a particle on the grid, they can interact with each atom on their path, therefore also transmitting information from the interior of the particle [468]. The resulting projection images have a very poor signal to noise ratio. However, when a large amount of particle projections are recorded, they can be averaged to drastically improve the signal to noise ratio and to recover the structural details [467, 469]. For this purpose, the particles are grouped according to their orientation on the grid. The members of a group have to be properly aligned and will then be averaged. This procedure yields two dimensional class averages. The class averages can then be used to reconstruct a three dimensional map [468-470].
Two techniques for sample preparation have been pursued, negative stain and vitrification. Using vitrified samples is referred to as cryo-electron microscopy. Both approaches will be described in the following two sections.

2.2.10.1. Negative stain

Negative staining is a simple method to image particles and is usually used for screening the quality of a sample. A heavy atom salt solution, such as uranyl acetate or ammonium molybdate, is applied to the sample on the grid and dried afterwards [471, 472]. The heavy atoms strongly interact with the electrons, and therefore the distribution of the heavy atom stain is imaged. Since the particles are embedded in the stain, the exclusion volume is monitored, yielding a high contrast at the edges of the particle [472]. However, the information from the interior of the particle is not recorded. Furthermore, if the particles are not fully embedded in the staining solution, parts of the particle are missing in the projection, impeding 3D reconstructions [472]. The resolution of negative stain imaging is limited due to the grain size of the heavy atom [468, 473].
In this study, uranyl acetate was used as stain. All steps were performed at room temperature. 3 µl of a protein sample at a concentration of 0.01-0.05 mg/ml was applied to the grid, incubated for 30 s, and blotted with filter paper. Subsequently, 20 µl of a 3 % (w/v) uranyl acetate solution was applied to the grid, incubated for 30 s, and blotted with filter

paper. The grid was then dried at room temperature. Data acquisition was performed at room temperature.

2.2.10.2. Cryo-electron microscopy

To preserve the high resolution information of a sample, cryo-electron microscopy (cryo-EM) is applied. In cryo-EM vitrified protein samples are used. The protein sample is applied to a grid and flash frozen in liquid ethane [467]. This leads to the formation of amorphous ice, keeping the protein in its native, hydrated state [467, 474].

Vitrification was performed with a Vitrobot™ at 4 °C and 100 % humidity. 3.5 µl of the sample at a concentration of 0.03-0.2 mg/ml was applied to a grid and blotted for 4 s with zero blot force. For blotting, two filter papers on each side of the grid were used. The sample was then flash frozen in liquid ethane and kept in liquid nitrogen for storage. Data acquisition was performed at cryogenic temperature.

3. Results

The aim of this thesis was the structural and functional characterization of TFIIH. This task encompassed different levels of complexity. First, the expression and purification of the single core subunits were addressed. During this task also shortened constructs were designed for structural studies. Hereof, a p62 construct was used for crystallographic studies, investigating the usage of cesium as a common element for experimental phasing approaches. Enzymatic activities of the XPB subunit were analyzed. Second, dual subcomplexes between these TFIIH subunits and shortened constructs were established to stabilize these subunits and to elucidate the interaction network within TFIIH. Third, the subcomplexes and single subunits were combined to assemble higher order TFIIH core complexes for structural studies via EM. The resulting EM maps were fitted with available models of the TFIIH subunits. Finally, the TFIIH core was extended by a shortened construct of the CAK complex subunit MAT1 and investigated via EM.

3.1. Sequence conservation of TFIIH in *Chaetomium thermophilum*

In this thesis *Chaetomium thermophilum* was used as model organism. *C. thermophilum* is a thermophilic eukaryote [475]. Thermophiles usually offer proteins of increased stability compared to their mesophilic orthologues, which can be highly beneficial for structural studies. Most importantly, *C. thermophilum* possesses all TFIIH subunits. If not explicitly stated otherwise, all TFIIH subunits mentioned here refer to the homologs from *C. thermophilum*. Table 3-1 provides an overview of all subunits and their sequence conservation.

Table 3-1. Comparison of the TFIIH subunits from *Homo sapiens* and *Chaetomium thermophilum.*

	Homo sapiens			*Chaetomium thermophilum*				
Protein	**mol. w.[1] (kDa)[2]**	**length (aa)[3]**	**UNP[4]**	**mol. w. (kDa)**	**length (aa)**	**UNP**	**Identity[5] (%)**	**Similarity[5] (%)**
XPB	89.3	782	P19447	97.5	867	G0SHF8	44.2	61.9
XPD	86.9	760	P18074	89.7	780	G0RZH0	49.1	67.5
p62	62.0	548	P32780	74.4	677	G0RZ08	19.3	33.7
p52	52.2	462	Q92759	56.8	514	G0S965	33.0	50.5
p44	44.4	395	Q13888	58.0	534	G0RZE6	30.7	42.4
p34	34.4	308	Q13889	43.6	429	G0RXV8	25.1	38.0
p8	8.1	71	Q6ZYL4	8.4	72	G0SDQ1	31.7	53.7
CDK7	39.0	346	P50613	49.1	437	G0SFC6	32.7	49.2
CyclinH	37.6	323	P51946	46.9	425	G0SH78	24.7	36.5
MAT1	35.8	309	P51948	40.8	363	G0SF48	24.8	44.1

1) Molecular weight.
2) Kilo Dalton.
3) Amino acids.
4) UniProtKB identifier.
5) Identity/Similarity to the human homolog. Determined with EMBOSS Needle; Matrix: EBLOSUM62; gap penalty: 10.0; extend penalty: 0.5.

3.2. XPB

3.2.1. Secondary structure prediction and construct design

Different constructs of XPB were designed in order to improve its expression, purification and crystallization propensities, or to test different regions of the protein for interactions with other TFIIH subunits. In order to detect presumably disordered regions within XPB, secondary structure predictions (Figure 3-1, see also 6.3.1) and sequence alignments with the human XPB amino acid sequence were performed (see 6.3.1). Regions within the protein, which are predicted as coil or as helix/sheet with low confidence score were assumed to be disordered. Furthermore, insertions and not conserved regions according to sequence alignments were considered as disordered as well. To improve the stability of the protein, these disordered regions were deleted. Different XPB constructs used in this work alongside a secondary structure prediction are depicted in Figure 3-1. Further XPB constructs were attempted in the course of this thesis, which are given in the appendix (see 6.5, Table 6-2).

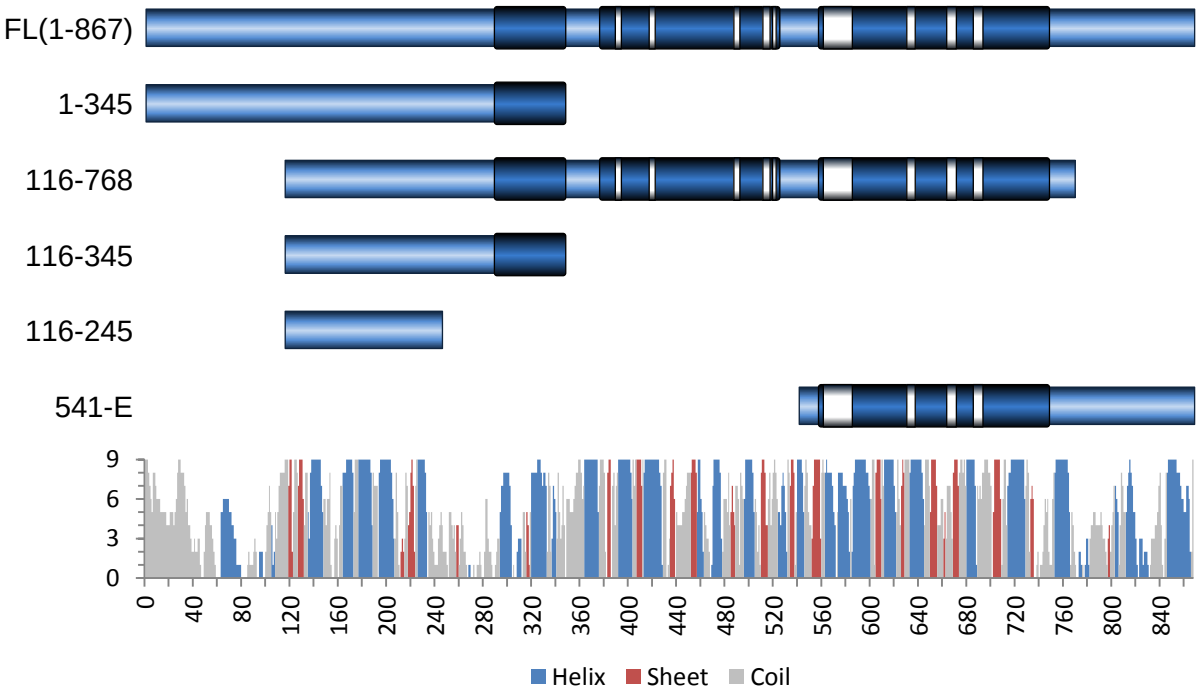


Figure 3-1. XPB constructs and secondary structure prediction. The construct schemes are aligned to the secondary structure prediction at the bottom. The domain scheme corresponds to Figure 1-1. Secondary structure prediction was performed with I-TASSER.

3.2.2. Expression and Purification

Different expression strains were tested for overexpression of full-length XPB, and the protein amount over time was monitored using SDS-PAGE analysis of crude extract samples (Figure 3-2).

Expression of XPB could be observed in different cell lines, but degradation was observed already after a few hours after induction of protein expression as seen for BL21 (DE3) and BL21 CodonPlus (DE3) RIL cells (Figure 3-2 a, b). Purifications were attempted from ArcticExpress (DE3) RIL, BL21 CodonPlus (DE3) RIL, and Rosetta™ 2 (DE3) according to the standard protocol (see 2.2.6).

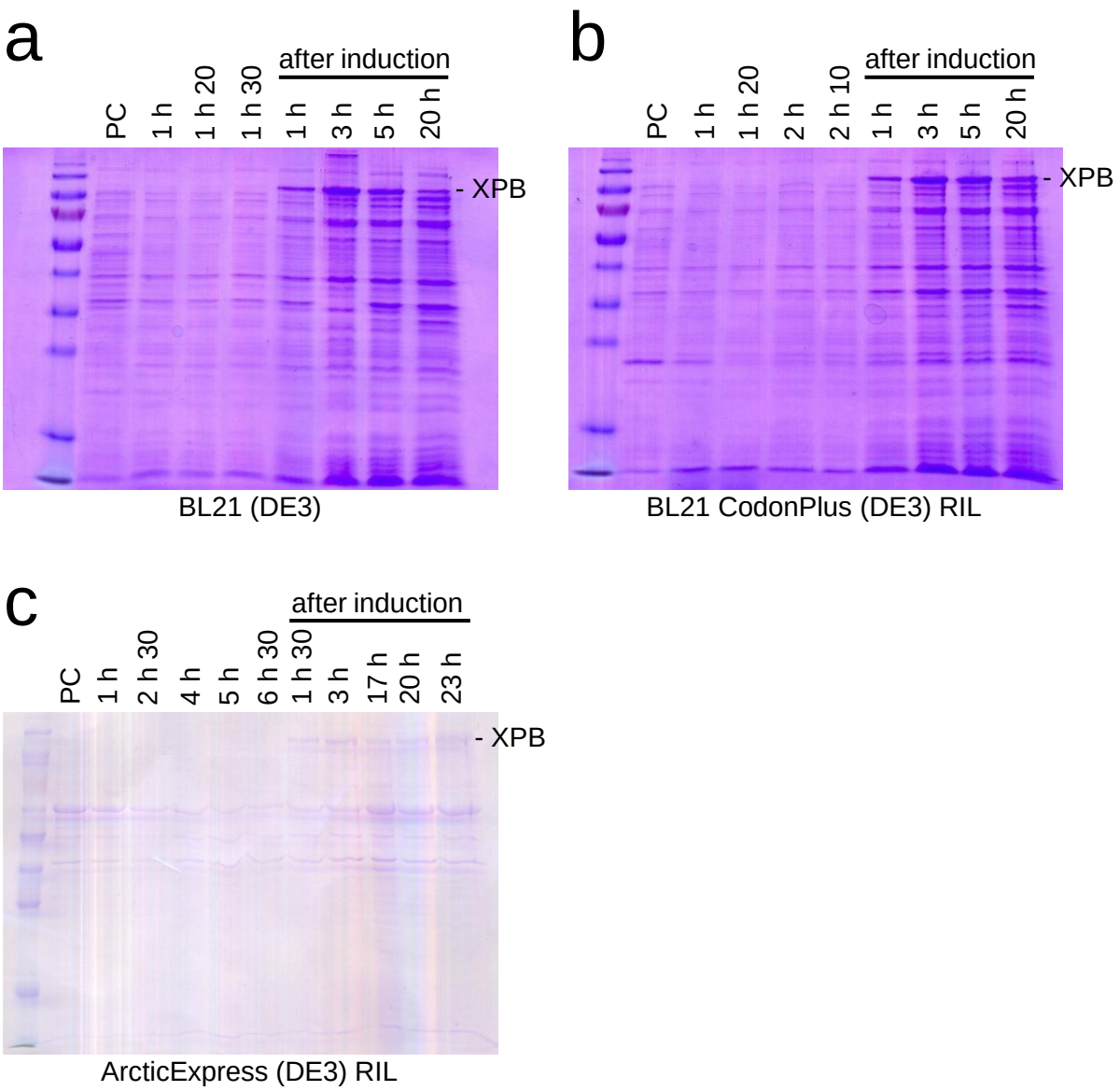


Figure 3-2. Overexpression of XPB in different bacterial expression strains. The amount of XPB before and after induction was monitored over time via SDS-PAGE. **(a)** BL21 (DE3). **(b)** BL21 CodonPlus (DE3) RIL. **(c)** ArcticExpress (DE3) RIL. PC: pre-culture.

Purification from ArcticExpress (DE3) RIL yielded the highest amount of XPB, yet tightly bound to a chaperone (Figure 3-3). The purification from BL21 CodonPlus (DE3) RIL seemed to be promising, however the yield was very low and the reproducibility of the purification was poor (Figure 3-4, Figure 3-7). Purification from Rosetta™ 2 (DE3) led to less pure protein compared to BL21 CodonPlus (DE3) RIL, yet the results were reproducible (Figure 3-5). Based on these initial results extensive attempts to optimize expression and purification from ArcticExpress (DE3) RIL and BL21 CodonPlus (DE3) RIL cells were conducted.

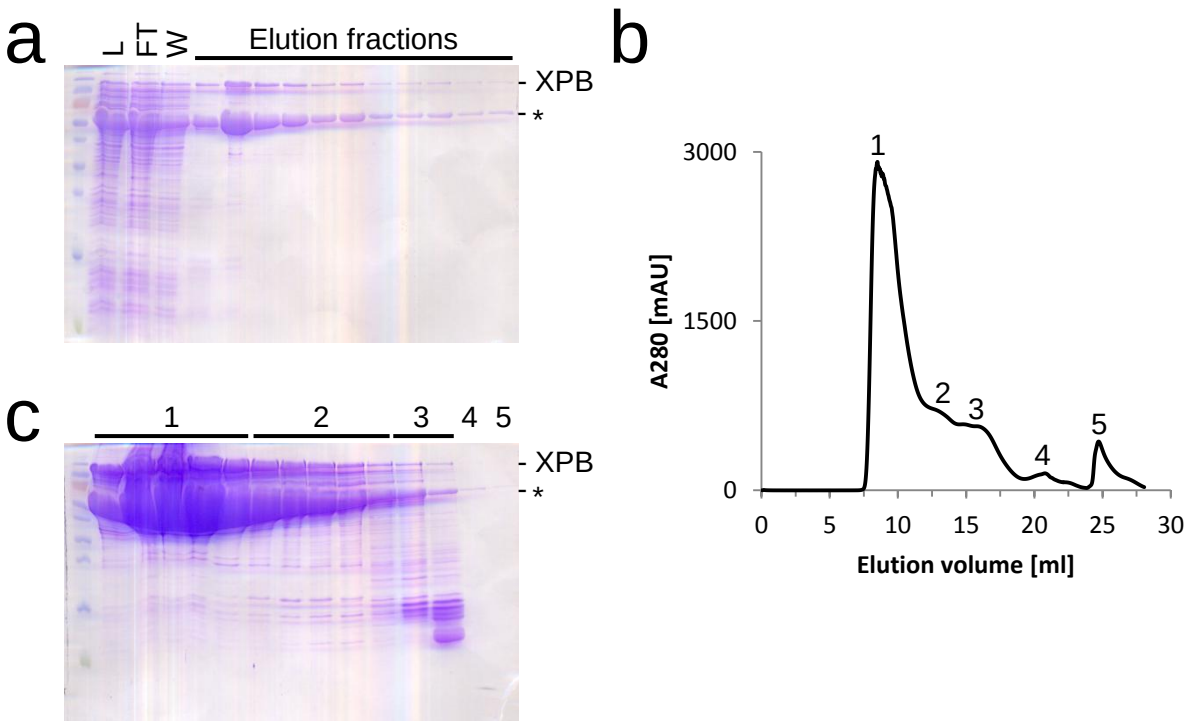


Figure 3-3. Purification of XPB from ArcticExpress (DE3) RIL cells. (a) SDS-PAGE analysis of IMAC samples. L: lysate; FT: flow through; W: wash. **(b)** SEC elution profile from a Superdex 200 10/300 GL column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b). *: chaperonin.

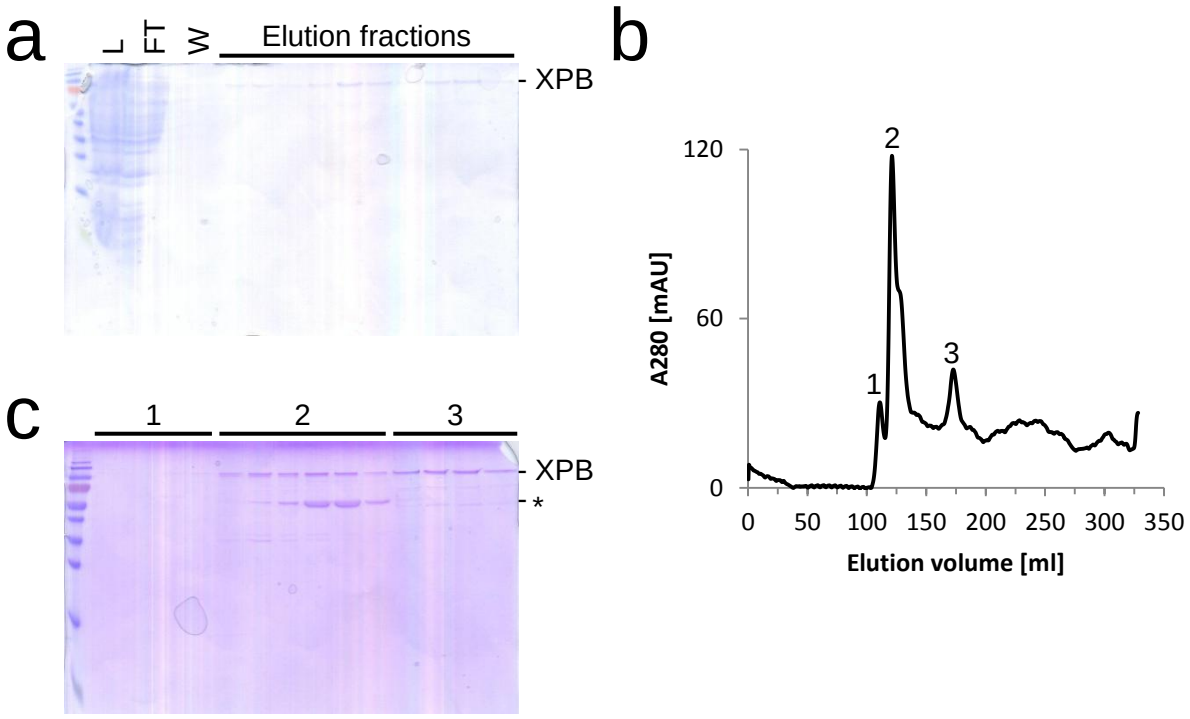


Figure 3-4. Purification of XPB from BL21 CodonPlus (DE3) RIL cells. (a) SDS-PAGE analysis of IMAC samples. L: lysate; FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 26/60 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b). *: chaperonin.

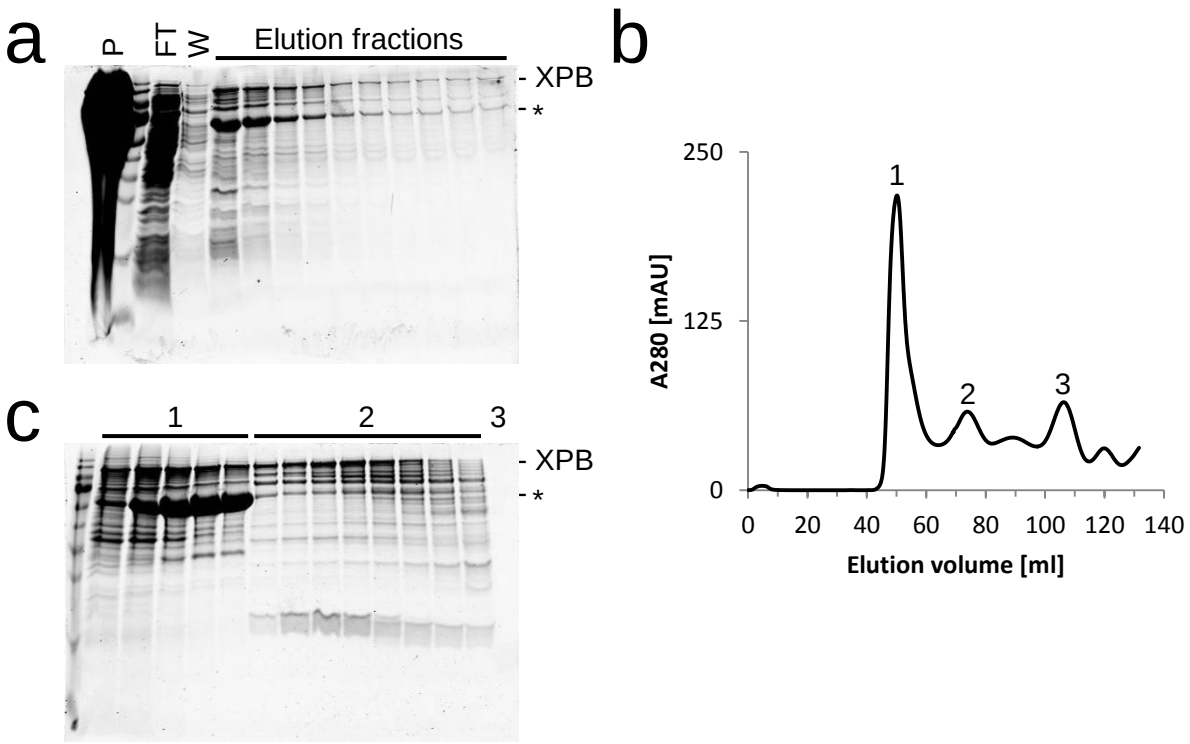


Figure 3-5. Purification of XPB from Rosetta™ 2 (DE3) cells. (a) SDS-PAGE analysis of IMAC samples. P: pellet; FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b). *: chaperonin.

In order to separate the chaperone from the high amounts of XPB expressed by ArcticExpress (DE3) RIL cells different approaches were applied, like IEX, addition of ATP, or purification under denaturing conditions. All approaches failed so far. The result from AIEX is exemplarily shown in Figure 3-6. Here, the XPB sample was diluted with dilution buffer to a final NaCl concentration of 0.1 M and loaded on the column. Bound proteins were eluted via a linear increasing ratio of high salt to low salt buffer (0-100 % over 20 column volumes). Buffer compositions are given below.

Dilution buffer

- 20 mM Tris-HCl pH 7.5

Low salt buffer

- 20 mM Tris-HCl pH 7.5
- 0.1 M NaCl

High salt buffer

- 20 mM Tris-HCl pH 7.5
- 1 M NaCl

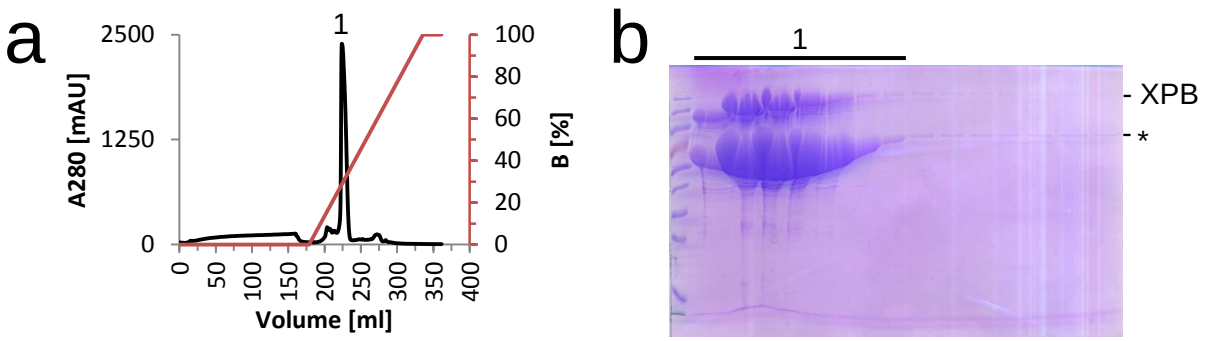


Figure 3-6. Anion exchange chromatography of XPB bound to a chaperone. (a) AIEX elution profile from a Mono Q 10/100 GL column. B: percentage ratio of high salt to low salt buffer. **(b)** SDS-PAGE analysis of elution fractions from AIEX. The number corresponds to the peak in (a). *: chaperonin.

Even though the results seemed to be promising initially, optimization of the purification from BL21 CodonPlus (DE3) RIL cells failed and was poorly reproducible. Figure 3-7 shows an example of such a purification where XPB was purified in the presence of a Pipes buffer system. The composition of all buffers is given below.

Lysis buffer

- 20 mM Pipes-NaOH pH 7.5
- 0.3 M NaCl
- 10 mM $MgCl_2$
- 5 mM Imidazole

Elution buffer

- 20 mM Pipes-NaOH pH 7.5
- 0.3 M NaCl
- 10 mM $MgCl_2$
- 0.25 M Imidazole

Gel filtration buffer

- 20 mM Pipes-NaOH pH 7.5
- 0.25 M NaCl
- 10 mM $MgCl_2$

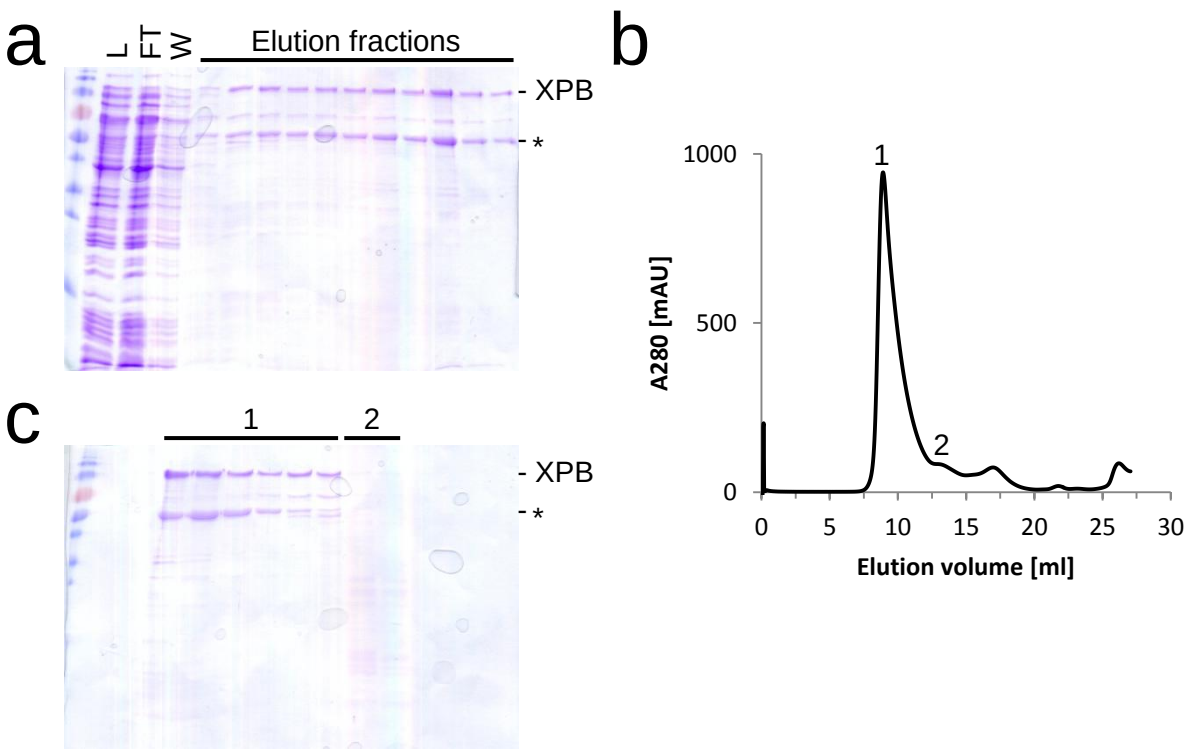


Figure 3-7. Purification of XPB from BL21 CodonPlus (DE3) RIL cells with a Pipes buffer system. (a) SDS-PAGE analysis of IMAC samples. L: lysate; FT: flow through; W: wash. **(b)** SEC elution profile from a Superdex 200 10/300 GL column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b). *: chaperonin.

From the observations made during the expression and purification attempts of full-length XPB, a new purification strategy was derived, implementing three major considerations:

1. During expression, degradation of XPB is already observed after a few hours.
2. The usage of Rosetta™ 2 (DE3) cells yields protein separated from the chaperonin in a reproducible manner.
3. A codon optimization of the XPB gene might increase the expression levels.

According to the assumptions stated above, Rosetta™ 2 (DE3) cells were used in future attempts, but the expression time was reduced from 21 h to 15 h. The cell pellet was freshly utilized for purification, without prior freezing, and all purification steps were performed within one day. Additionally, the XPB construct was codon optimized to make it more suitable for the bacterial expression system.

The optimized gene was ordered at ATG:biosynthetics. Different parameters were considered, like codon usage, GC-content, and secondary structure at the RNA level. The optimized gene was delivered in a pGH vector. The gene was then cloned into a pBADM-11 vector for further usage. An alignment between the wild-type and optimized cDNA sequence is given in the appendix (see 6.4).

The above described strategy yielded pure XPB in a reproducible manner, using the standard buffer system (Figure 3-8).

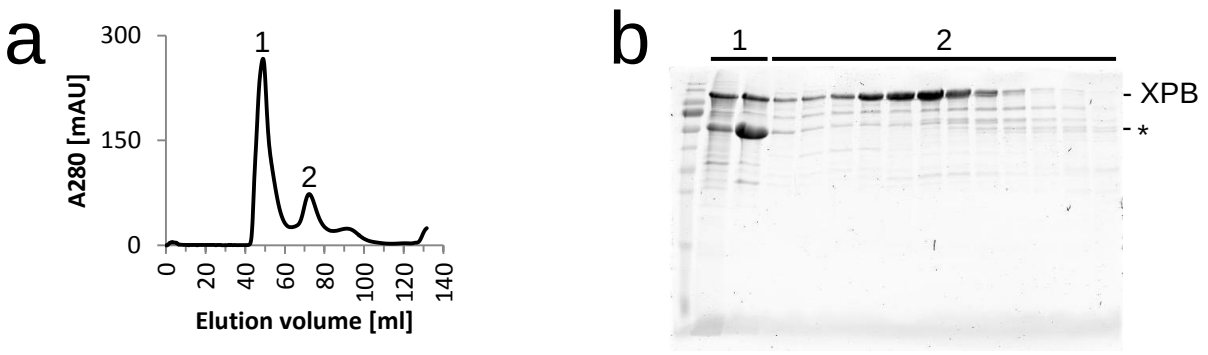


Figure 3-8. Purification of XPB after the revised strategy with the standard buffer system. (a) SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(b)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (a). *: chaperonin.

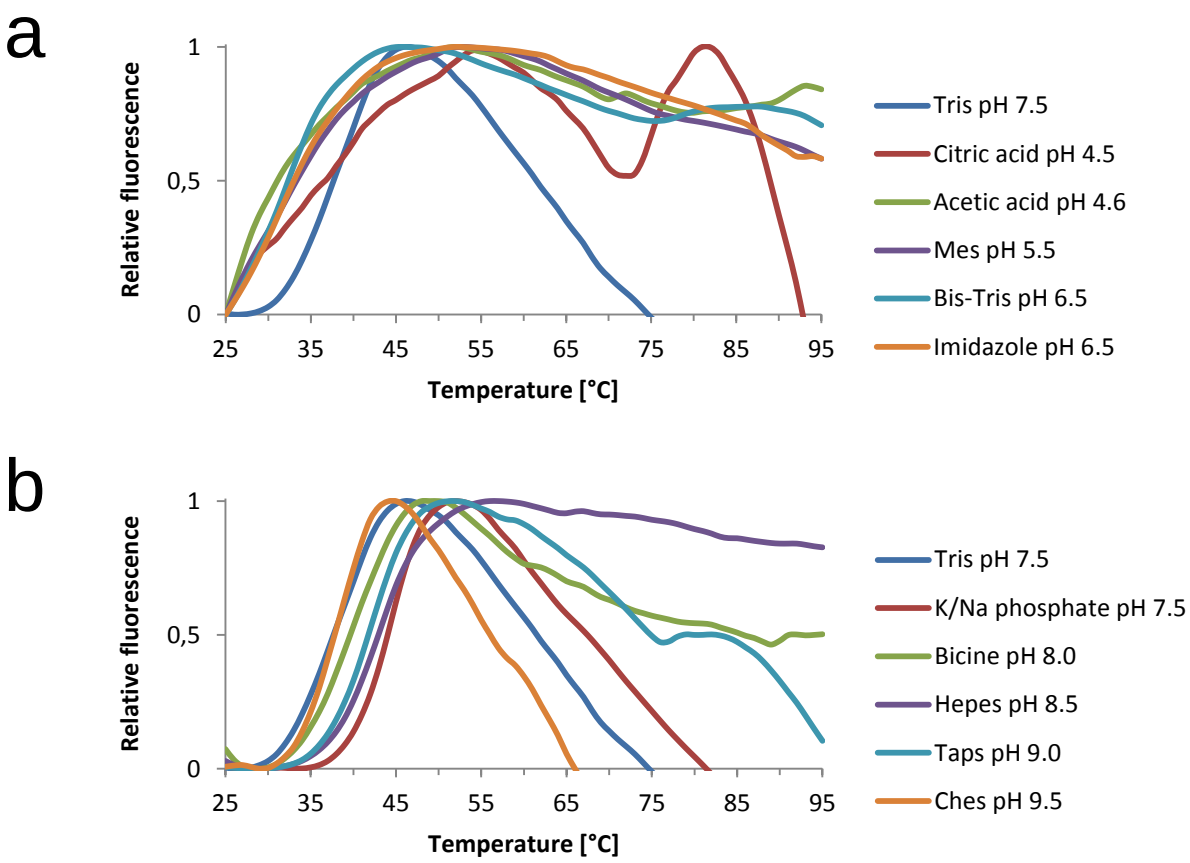


Figure 3-9. Thermofluor analysis of XPB using the Thermofluor standard screen. (a) Comparison of acidic buffers to the reference buffer (Tris-HCl pH 7.5). **(b)** Comparison of basic buffers to the reference buffer.

A Thermofluor assay was performed with the protein obtained after completion of the revised purification strategy. The standard Thermofluor screen was used and the results are depicted in Figure 3-9. The reference buffer (100 mM Tris-HCl pH 7.5) displayed a typical unfolding

curve, indicating proper folding of the protein and a melting temperature of ~37 °C. Acidic buffer conditions were not suitable, as no typical unfolding curves could be observed (Figure 3-9 a). Basic buffer conditions however mostly led to proper unfolding curves, with some buffers shifting the melting temperature to higher temperatures. The most pronounced shift of approximately 6 K to higher melting temperature could be observed with a K/Na phosphate pH 7.5 buffer (Figure 3-9 b).

As K/Na phosphate buffer at pH 7.5 led to a distinct shift to higher melting temperature, purification with a K/Na phosphate buffer system following the revised strategy was attempted. The results were comparable to the purification with the standard buffer system and are depicted in Figure 3-10. From the SDS-PAGE it appeared that impurities and degradation were reduced. However, this might also be due to the lower protein amounts in the gel and the weaker staining. The K/Na phosphate buffer was prepared by mixing 2 M KH_2PO_4 with 2 M Na_2HPO_4 in an approximately 1:4 ratio, so that pH 7.5 was reached. The buffer compositions are given below.

Lysis buffer

- 20 mM K/Na phosphate pH 7.5
- 0.3 M NaCl
- 5 mM Imidazole

Elution buffer

- 20 mM K/Na phosphate pH 7.5
- 0.3 M NaCl
- 0.25 M Imidazole

Gel filtration buffer

- 20 mM K/Na phosphate pH 7.5
- 0.25 M NaCl

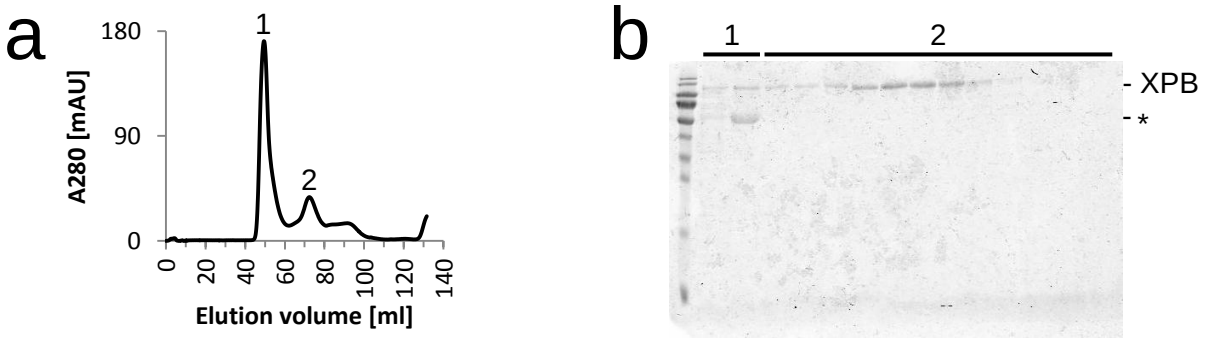


Figure 3-10. Purification of XPB following the revised strategy with a phosphate buffer system. (a) SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(b)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (a). *: chaperonin.

3.2.3. ATPase assays

The ATPase activity of XPB was tested via a coupled enzyme assay (see 2.2.8). These assays were performed by Gudrun Sander and Jochen Kuper. Different XPB concentrations (0.25, 0.5, and 1 µM) were tested for ATPase activity in dependence of different DNA substrates and the presence of p52/p8. Here, single stranded DNA (ssDNA), double stranded DNA (dsDNA), and dsDNA with a 3' overhang (3'OH) were tested. The sequences of the DNA substrates are given in Table 3-2. The results for the ATPase activity with increasing XPB concentrations are depicted in Figure 3-11 a. In the presence of ssDNA alone no ATPase could be observed. In contrast, dsDNA supported induction of ATPase activity. Using a 3'OH DNA substrate led to an even more pronounced activation of XPBs ATPase activity, well in line with the notion that XPB possesses a 3'->5' directionality [9]. Remarkably, addition of p52/p8 (added in equimolar ratio) also stimulated ATPase activity, even in the absence of DNA. This behavior has also been reported before [15, 186]. When adding p52/p8 in addition to dsDNA, the strongest activation of the ATPase activity was observed.

Table 3-2. Sequences of the DNA substrates used for the ATPase assays of XPB.

Substrate	Sequence
ssDNA	5'-GCTACCATGCCTGCACGAATTAAGCAATTCGTAATCATGGTCATAGC-3'
dsDNA	5'-GCTACCATGCCTGCACGAATTAAGCAATTCGTAATCATGGTCATAGC-3' 3'-CGATGGTACGGACGTGCTTAATTCGTTAAGCATTAGTACCAGTATCG-5'
3'OH	5'-AAGCAATTCGTAATCATGGTCATAGC-3' 3'-GATGTCAAGCAGTCCTAAGGTTCGTTAAGCATTAGTACCAGTATCG-5'

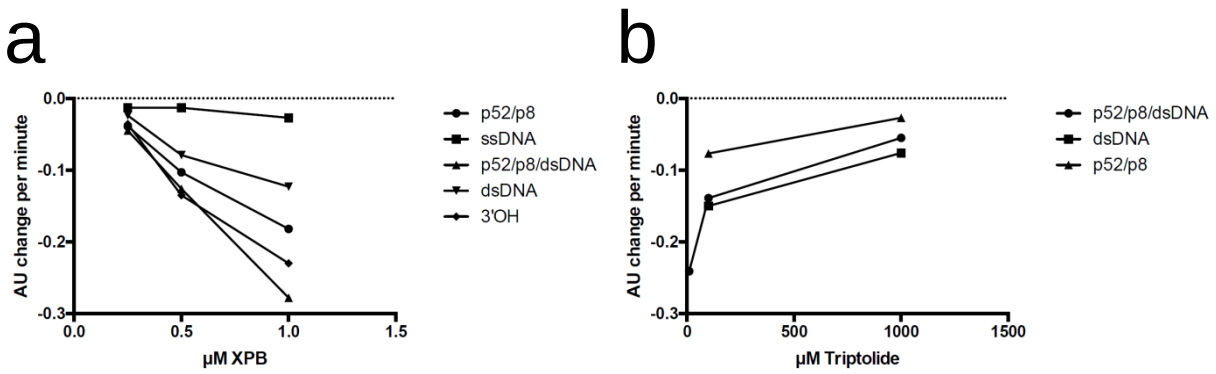


Figure 3-11. ATPase activity of XPB. (a) ATPase activity in dependence of different DNA substrates and p52/p8. **(b)** Inhibitory effect of Triptolide on the ATPase activity of XPB. ssDNA: single stranded DNA; dsDNA: double stranded DNA; 3'OH: 3' overhang. The data was provided by courtesy of Jochen Kuper.

Triptolide is a known natural inhibitor of ATPase activity of human XPB [298]. It covalently attaches to cysteine residue C342 in human XPB, which is located near the first helicase motif [476]. According to sequence alignments (see 6.3.1), C342 is conserved in *C. thermophilum* XPB and corresponds to C388. Therefore, the effect of triptolide on the ATPase activity of XPB from *C. thermophilum* was investigated, and the results are depicted in Figure 3-11 b. Here, 1 µM of XPB were used, and the effects of increasing amounts of triptolide were monitored. The ATPase activity in the presence of dsDNA, p52/p8, or both was tested. In all cases, addition of triptolide strongly inhibited the ATPase activity. These findings underlined the validity and usefulness of the *C. thermophilum* model system.

3.2.4. XPB_116-768

Based on sequence alignments and secondary structure predictions (Figure 3-1, see also 6.3.1) numerous expression constructs of XPB were engineered, in order to improve its protein expression and purification behavior. Several constructs did not lead to an improvement compared to full-length XPB, and only the results for the XPB_116-768 construct are exemplarily described here. Further constructs are listed in Table 6-2 (see 6.5).
Secondary structure predictions suggested extensive disorder for roughly the first 120 amino acids and the last 110 amino acids of the full-length protein (see 6.3.1, Figure 6-1 a, b). Accordingly, these regions were predicted to be predominantly unstructured (Figure 3-1, see also 6.3.1). Sequence alignments with human XPB (see 6.3.1) furthermore indicated extensions and insertions at the N- and C-terminus. Hence, the construct XPB_116-768 was designed, lacking the first 115 and the last 99 amino acids (Figure 3-1). XPB_116-768 was expressed in ArcticExpress (DE3) RIL cells and purified according to the standard protocol. Similar to the full-length protein, the construct was bound by a chaperonin and aggregated during the SEC (Figure 3-12).

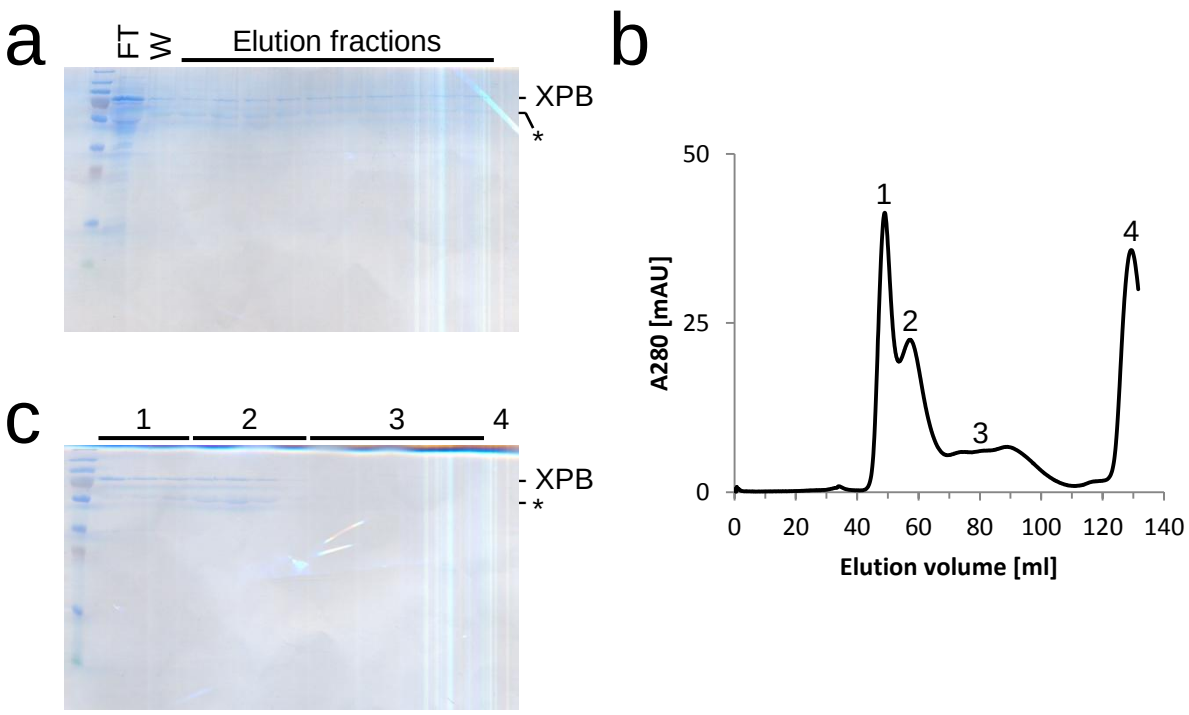


Figure 3-12. Purification of XPB_116-768 from ArcticExpress (DE3) RIL cells. (a) SDS-PAGE analysis of IMAC samples. FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b). *: chaperonin.

3.2.5. XPB_1-345

To investigate the N-terminal extension of XPB, a construct consisting of the N-terminus, as well as the so called DRD domain was designed (Figure 3-1). The crystal structure of an archaeal XPB [14] suggests that the N-terminal extension, including the DRD domain, encompasses roughly the first 300 amino acids of human XPB. According to sequence alignments (see 6.3.1) this range corresponds to the first 345 amino acids of XPB from *C. thermophilum* and therefore the construct XPB_1-345 was cloned, expressed, and purified. Purification from BL21 CodonPlus (DE3) RIL cells with the standard buffer system is depicted in Figure 3-13. As apparent from the SEC elution profile, XPB_1-345 eluted as two distinct species (Figure 3-13 b). Fractions from both peaks were pooled and concentrated separately. Both protein pools were subjected to crystallization trials. The protein pool from the peak at higher elution volume was used for interaction studies. Crystallization trials were conducted at 7.8 mg/ml and 9.8 mg/ml utilizing the crystallization robot (see 2.2.9.1) and the commercial crystallization screens (see 2.1.8, Table 2-8) yielding no crystals.

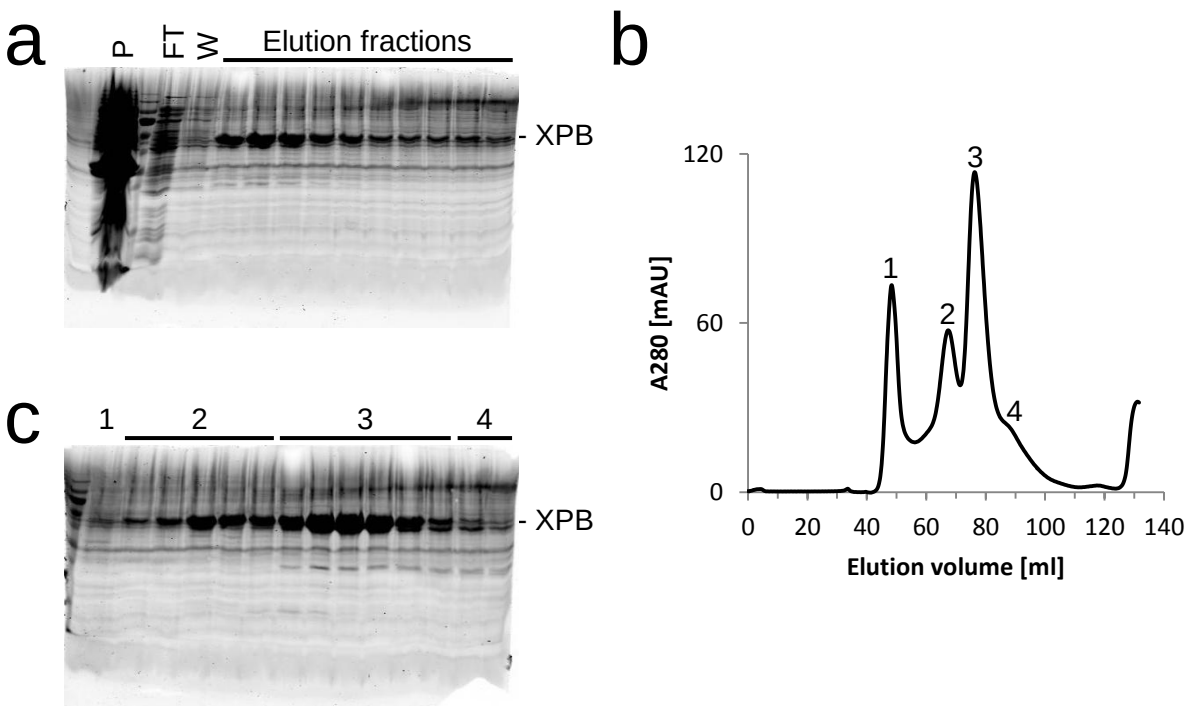


Figure 3-13. Purification of XPB_1-345 from BL21 CodonPlus (DE3) RIL cells. (a) SDS-PAGE analysis of IMAC samples. P: pellet; FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

3.2.6. XPB_1-345_F143S

To investigate the disease related mutation F99S in human XPB (see 1.6.4, Figure 1-5) [22, 389], this missense mutation was introduced into the XPB_1-345 fragment. According to sequence alignments (see 6.3.1), residue F99 corresponds to F143 in XPB from *C. thermophilum*. XPB_1-345_F143S was generated via site directed mutagenesis, and the purification from BL21 CodonPlus (DE3) RIL cells with the standard buffer system is shown in Figure 3-14. Compared to XPB_1-345, the protein yield and purity was reduced for the mutant construct.

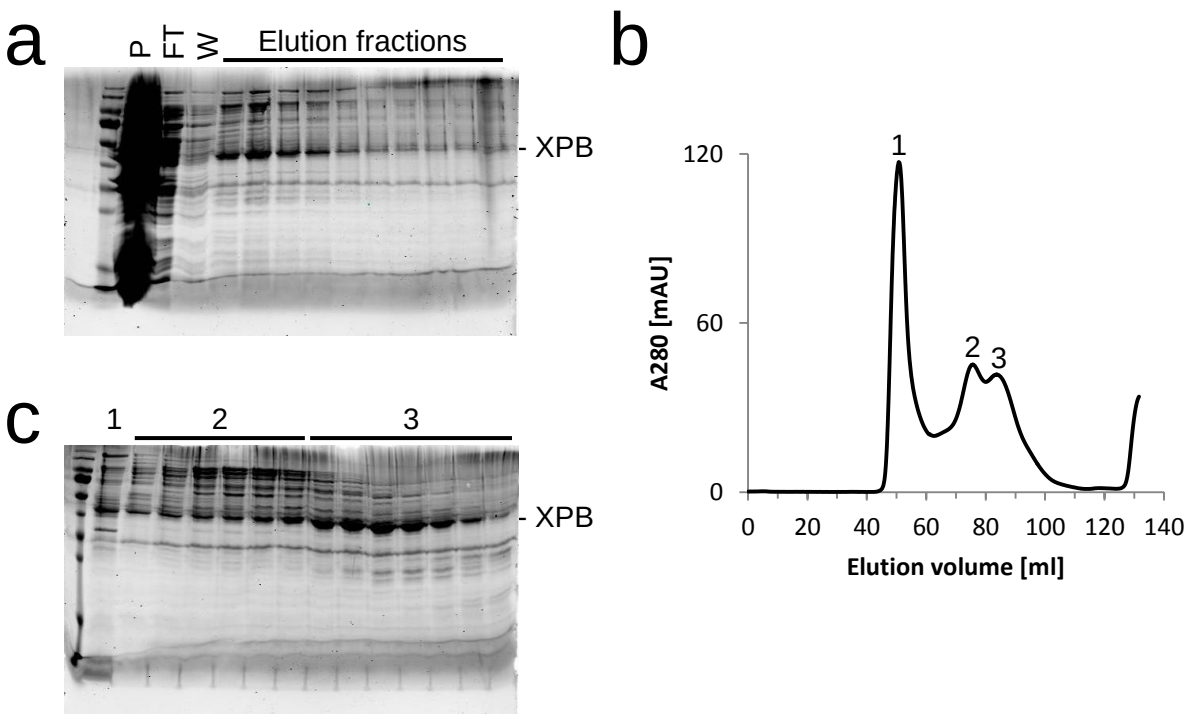


Figure 3-14. Purification of XPB_1-345_F143S from BL21 CodonPlus (DE3) RIL cells. (a) SDS-PAGE analysis of IMAC samples. P: pellet; FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

3.2.7. XPB_116-345

As seen from secondary structure prediction (Figure 3-1, see also 6.3.1) and described for construct XPB_116-768 (see 3.2.4), roughly the first 120 amino acids of XPB are predicted to be disordered. Therefore, a construct based on XPB_1-345, but without the first 115 amino acids was designed. Accordingly, construct XPB_116-345 was cloned (Figure 3-1). Expression and purification from BL21 CodonPlus (DE3) RIL cells with the standard buffer system was attempted, and the result after IMAC purification is shown in Figure 3-15. As the protein was insoluble and remained in the pellet fraction, purification was not continued.

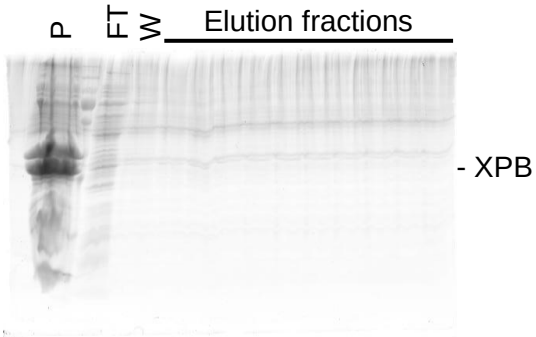


Figure 3-15. SDS-PAGE analysis of IMAC samples of XPB_116-345 purified from BL21 CodonPlus (DE3) RIL cells. P: pellet; FT: flow through; W: wash.

3.2.8. XPB_541-E

Hilario et al. reported the crystallization of a human XPB construct, encompassing the complete C-terminal half [62]. This construct starts at residue 494 and contains HD2, the ThM motif, and the C-terminal extension. In order to obtain a soluble construct of the C-terminal half for further studies, the corresponding construct for *C. thermophilum* was cloned. According to sequence alignments (see 6.3.1), residue 494 from human XPB corresponds to residue 541 in *C. thermophilum*. Thus, construct XPB_541-E was cloned, starting at residue 541 and ending at the very C-terminus with residue 867 (Figure 3-1). The purification utilizing ArcticExpress (DE3) RIL cells with the standard buffer system is shown in Figure 3-16. Besides eluting between 80 and 90 ml, which is the elution volume expected for this fragment if it is a monomer, a significant amount of XPB_541-E eluted also after 120 ml (Figure 3-16 b). Fractions from both species were pooled and concentrated separately. The protein pool from elution fractions between 80 and 90 ml was used for further studies. The late elution after 120 ml was quite unusual and might indicate partial unfolding of the protein and/or sticking to the column. This fraction was therefore not analyzed further.

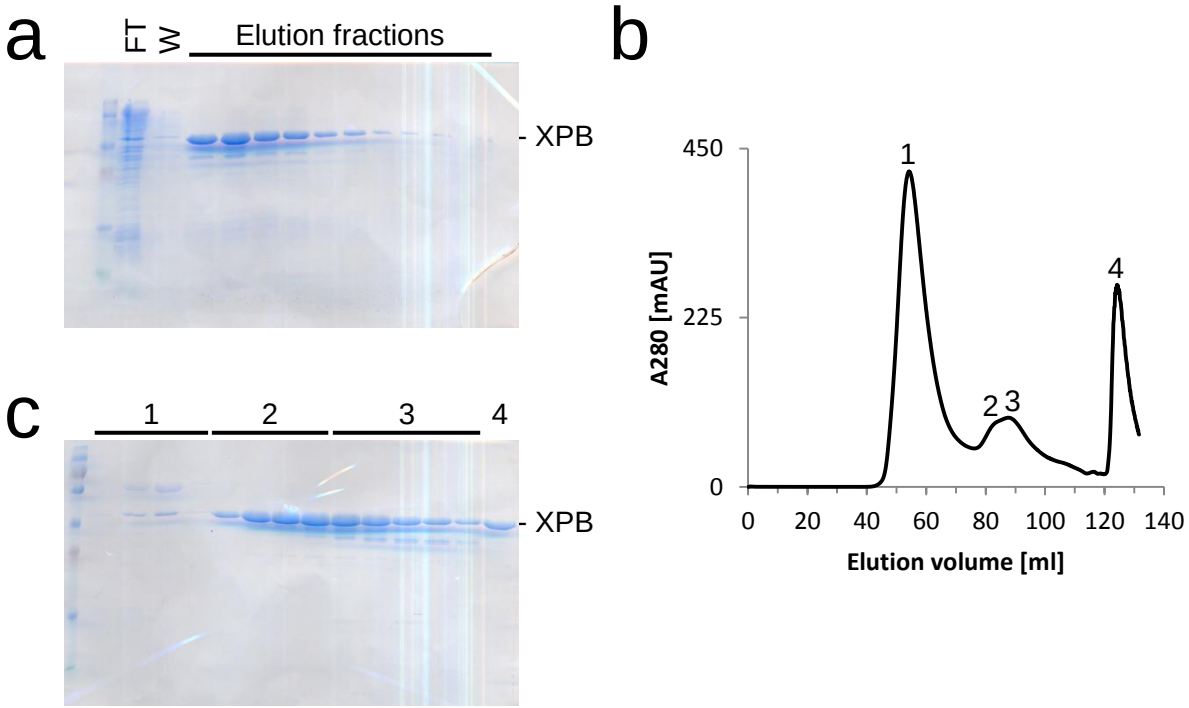


Figure 3-16. Purification of XPB_541-E from ArcticExpress (DE3) RIL cells. (a) SDS-PAGE analysis of IMAC samples. FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

3.3. XPD

The XPD sequence used in this work differs from the sequence given in the UniProtKB reference G0RZH0, as in the latter one exon seems to be missing. An alignment between the sequence used in this work and the sequence from the UniProtKB entry is given in Figure 3-17 to illustrate this difference.

```
ct     1 MEFMIDDLPVLFPYPRIYPEQYAYMCDLKKTLDAGGNCVLEMPSGTGKTITLLSLIVAYQQHYAEHRKLIYCSRT   75
UNP    1 MEFMID-----------------YMCDLKKTLDAGGNCVLEMPSGTGKTITLLSLIVAYQQHYAEHRKLIYCSRT   58

ct    76 MSEIEKALVELKALMKFRAERLGYVEEFRGLGLTSRKNLCLHPSVKREKSGTIVDARCRSLTAGFVKEKKQRGED  150
UNP   59 MSEIEKALVELKALMKFRAERLGYVEEFRGLGLTSRKNLCLHPSVKREKSGTIVDARCRSLTAGFVKEKKQRGED  133

ct   151 VDVCIYHDNLDLLEPHNLIPNGIWTLDNLLKYGEEHKQCPYFTARRMLQYCNVVIYSYHYLLDPKIAERVSRDLS  225
UNP  134 VDVCIYHDNLDLLEPHNLIPNGIWTLDNLLKYGEEHKQCPYFTARRMLQYCNVVIYSYHYLLDPKIAERVSRDLS  208

ct   226 SDSIVVFDEAHNIDNVCIEALSTDITEESLRRATRGAQNLENRINEMKETDQQKLQDEYEKLVEGLRGNDDGTRE  300
UNP  209 SDSIVVFDEAHNIDNVCIEALSTDITEESLRRATRGAQNLENRINEMKETDQQKLQDEYEKLVEGLRGNDDGTRE  283

ct   301 DSFMTSPVLPQDLLKEAVPGNIRRAEHFVAFLRRFIEYLKTRMKVRQVISETPPSFLAHLKEYTFIEKKPLRWCA  375
UNP  284 DSFMTSPVLPQDLLKEAVPGNIRRAEHFVAFLRRFIEYLKTRMKVRQVISETPPSFLAHLKEYTFIEKKPLRWCA  358

ct   376 ERLTSLVRTLELTNIEDYHALQEVATFATLVATYEKGFLLILEPYESDTAEVPNPVLHFCCLDAAIAIKPVFDKF  450
UNP  359 ERLTSLVRTLELTNIEDYHALQEVATFATLVATYEKGFLLILEPYESDTAEVPNPVLHFCCLDAAIAIKPVFDKF  433

ct   451 RNVIITSGTISPLEMYPKMLNFTTVVQESYSMTLARRSFLPLIVTRGSDQASISTGFQVRNEPSVVRNYGNLLTE  525
UNP  434 RNVIITSGTISPLEMYPKMLNFTTVVQESYSMTLARRSFLPLIVTRGSDQASISTGFQVRNEPSVVRNYGNLLTE  508

ct   526 FAKITPDGMVVFFPSYLYMESIISMWQGMGILDEVWKYKLILVETPDAQETSLALETYRTACCNGRGAVLLCVAR  600
UNP  509 FAKITPDGMVVFFPSYLYMESIISMWQGMGILDEVWKYKLILVETPDAQETSLALETYRTACCNGRGAVLLCVAR  583

ct   601 GKVSEGIDFDHQYGRTVLCIGVPFQYTESRILKARLEFLRETYRIRENDFLSFDAMRHAAQCLGRVLRGKDDYGL  675
UNP  584 GKVSEGIDFDHQYGRTVLCIGVPFQYTESRILKARLEFLRETYRIRENDFLSFDAMRHAAQCLGRVLRGKDDYGL  658

ct   676 MVLADRRFQKKRNQLPKWIAQALLDADTNLSTDMAVSSARRFLKTMAQPFKAKDQEGISTWSLEDLKRHQQKMDE  750
UNP  659 MVLADRRFQKKRNQLPKWIAQALLDADTNLSTDMAVSSARRFLKTMAQPFKAKDQEGISTWSLEDLKRHQQKMDE  733

ct   751 ERMKELEAQREKESAPAVERPADQGRESDYEVDEEEEREMMALDVGN  797
UNP  734 ERMKELEAQREKESAPAVERPADQGRESDYEVDEEEEREMMALDVGN  780
```

Figure 3-17. Alignment between the XPD amino acid sequence used in this work (ct) and the sequence from the UniProtKB entry G0RZH0 (UNP).

3.3.1. Expression and purification

Expression and purification of XPD was performed following a protocol established by Gudrun Sander und Jochen Kuper. XPD was overexpressed in ArcticExpress (DE3) RIL cells cultured in TB-medium. The cell culture was grown to an OD_{600} of 1.9 and then induced with 0.03 % (w/v) Arabinose and 0.5 mM IPTG. Expression and cell harvest was performed according to the standard procedure (see 2.2.5.1 and 2.2.6). The cell pellet was resuspended in lysis buffer supplemented with DNase I, three protease inhibitor cocktail tablets, and 1 mM TCEP. After cell lysis, Ni-IDA beads were used for IMAC. Elution fractions were concentrated and subjected to SEC. Elution fractions from the peak containing XPD were pooled and concentrated with a centrifugal filter unit. SDS-PAGE analysis of the purification and the SEC elution profile are depicted in Figure 3-18. The composition of the purification buffers is given below.

Lysis buffer

- 20 mM Hepes-NaOH pH 7.0
- 0.3 M NaCl
- 5 mM $MgCl_2$
- 5 mM Imidazole

Elution buffer

- 20 mM Hepes-NaOH pH 7.0
- 0.3 M NaCl
- 5 mM $MgCl_2$
- 1 mM TCEP
- 0.25 M Imidazole

Gel filtration buffer

- 20 mM Hepes-NaOH pH 7.0
- 0.15 M NaCl
- 5 mM $MgCl_2$
- 1 mM TCEP

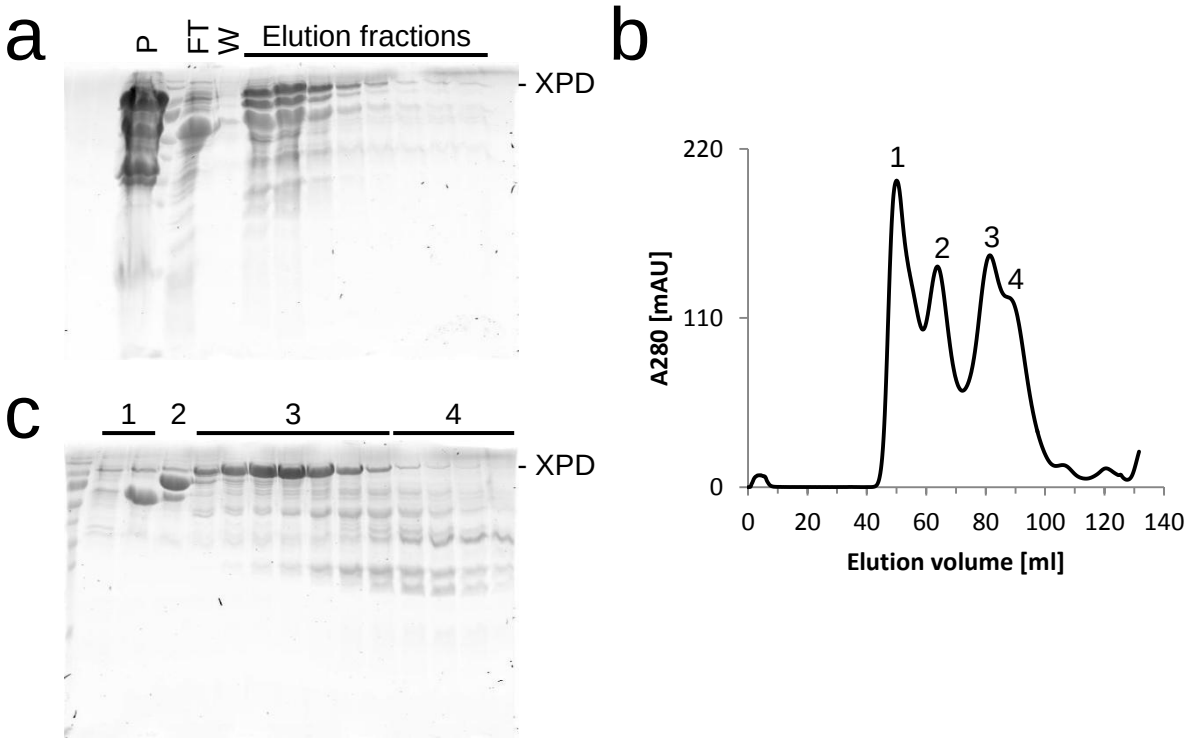


Figure 3-18. Purification of XPD. (a) SDS-PAGE analysis of IMAC samples. P: pellet; FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

3.4. p62

3.4.1. Secondary structure prediction and construct design

To improve crystallization propensities and to investigate interactions with other TFIIH subunits, different constructs of p62 were designed (Figure 3-19). Secondary structure predictions and sequence alignments with the human p62 sequence were performed, which formed the basis for construct design (Figure 3-19, see also 6.3.2).

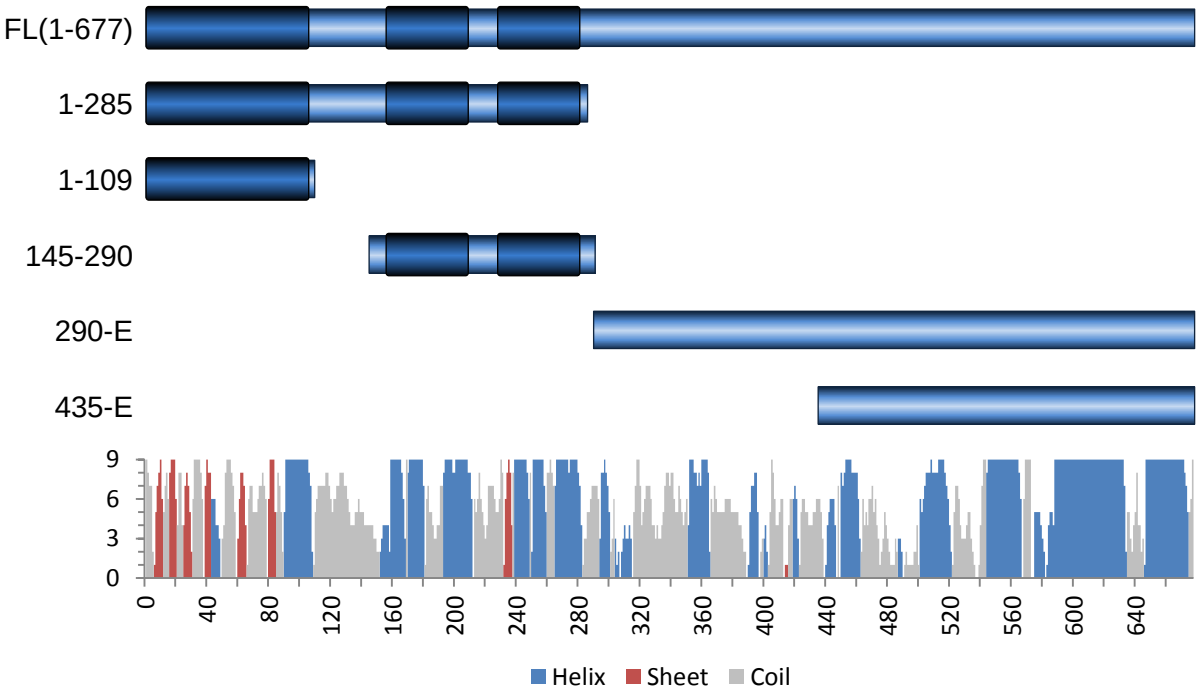


Figure 3-19. p62 constructs and secondary structure prediction. The construct schemes are aligned to the secondary structure prediction at the bottom. The domain scheme corresponds to Figure 1-1. Secondary structure prediction was performed with I-TASSER.

3.4.2. Expression and purification

Initial attempts to express and purify p62 from BL21 CodonPlus (DE3) RIL cells implementing the standard protocol failed, and the result is shown in Figure 3-20.

Expression from ArcticExpress (DE3) RIL cells, however, was successful and yielded protein of high purity (Figure 3-21). The resulting protein sample was subjected to crystallization (see 3.4.3).

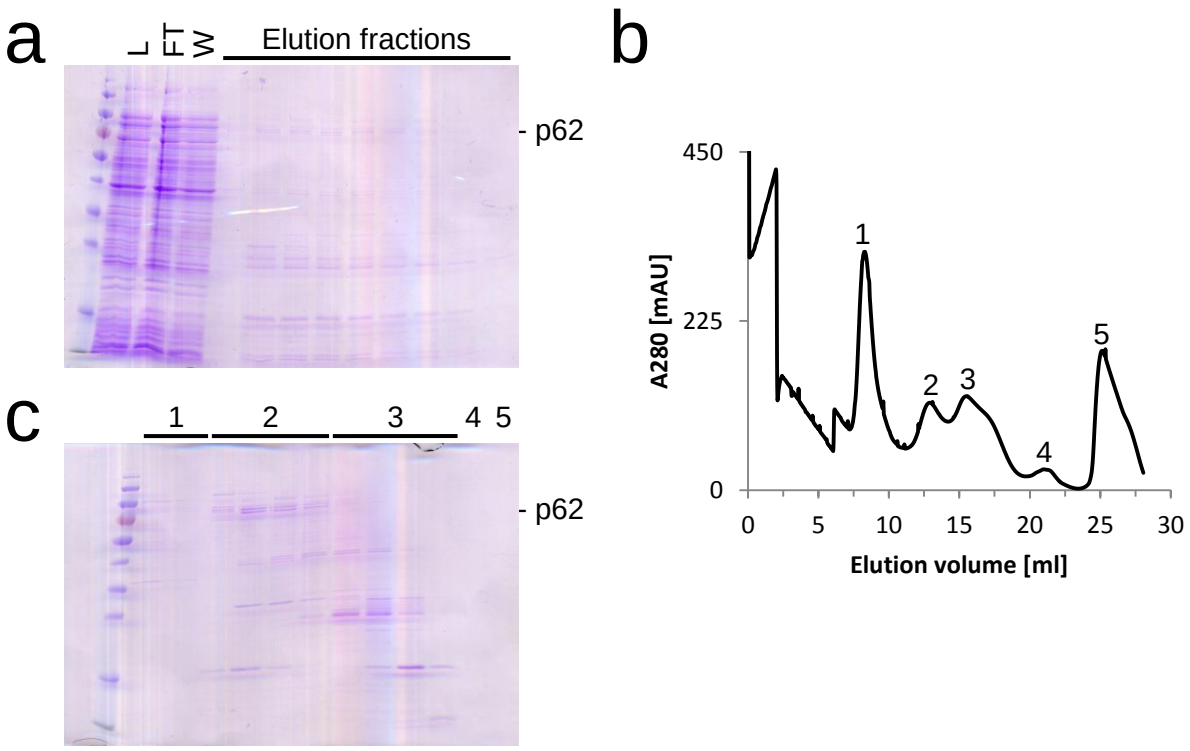


Figure 3-20. Purification of p62 from BL21 CodonPlus (DE3) RIL cells. (a) SDS-PAGE analysis of IMAC samples. L: lysate; FT: flow through; W: wash. **(b)** SEC elution profile from a Superdex 200 10/300 GL column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

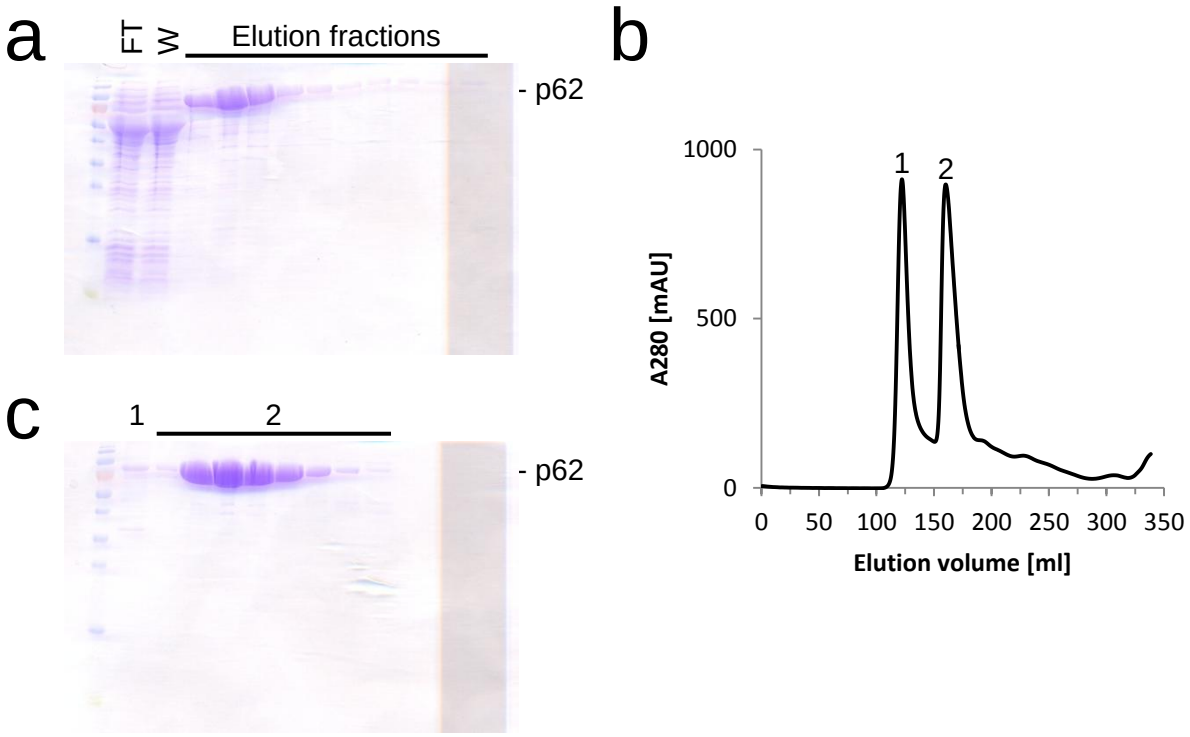


Figure 3-21. Purification of p62 from ArcticExpress (DE3) RIL cells. (a) SDS-PAGE analysis of IMAC samples. FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 26/60 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

To stabilize p62 and improve crystallization propensities, a Thermofluor assay was performed, using the standard Thermofluor screen (Figure 3-22). The assay revealed a preference of p62 for basic buffers, increasing the melting temperature compared to the reference buffer at a pH of 7.5 (Figure 3-22 b). Acidic buffers on the other hand, led to decreased melting temperatures (Figure 3-22 a). Ches buffer at pH 9 displayed an increase of the melting temperature of about 3-4 K (Figure 3-22 b).

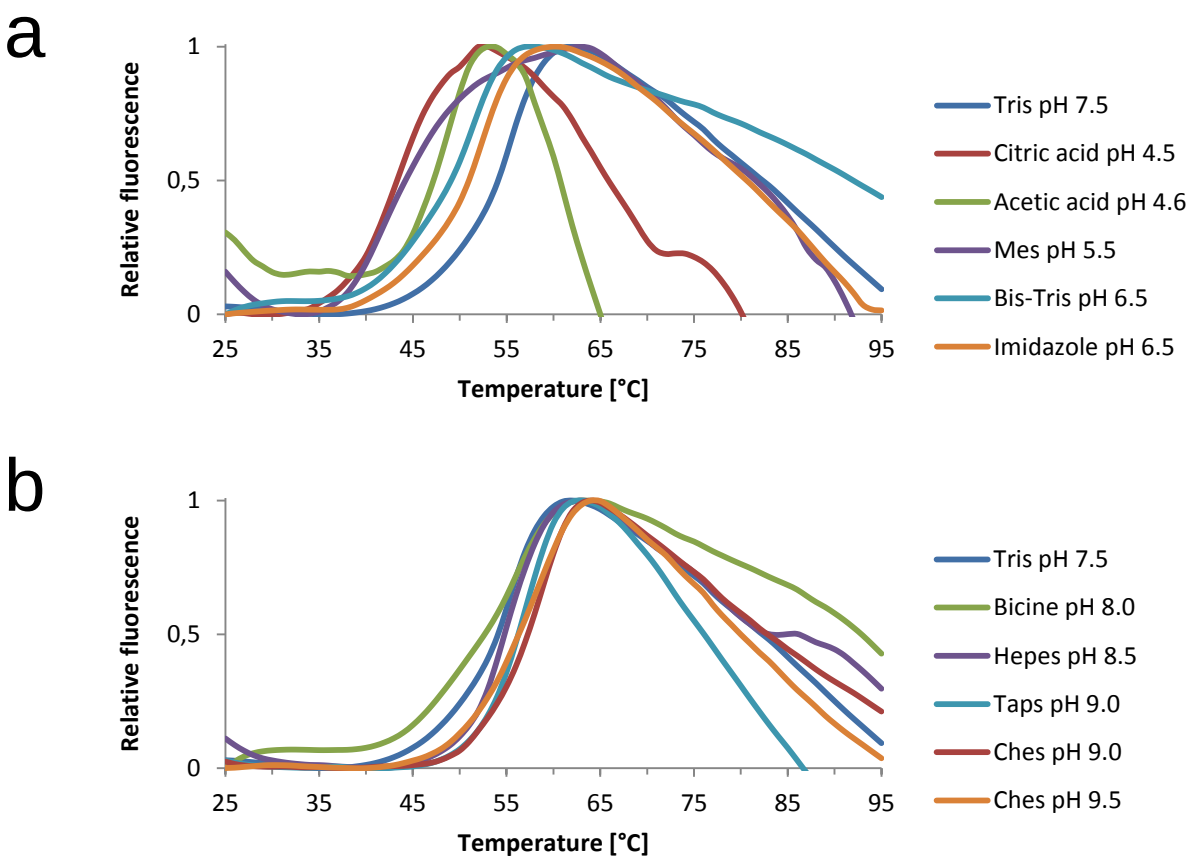


Figure 3-22. Thermofluor analysis of p62 using the Thermofluor standard screen. (a) Comparison of acidic buffers to the reference buffer (Tris-HCl pH 7.5). **(b)** Comparison of basic buffers to the reference buffer.

From these results a Ches based buffer system at pH 9 was derived. Purification with this buffer system led to about three times higher relative yields and the result from the SEC is depicted in Figure 3-23. Protein purified in the Ches buffer system was subjected to crystallization as well (see 3.4.3). The composition of the Ches buffer system is given below.

Lysis buffer

- 50 mM Ches-NaOH pH 9.0
- 0.3 M NaCl
- 5 mM Imidazole

Elution buffer

- 50 mM Ches-NaOH pH 9.0
- 0.3 M NaCl
- 0.25 M Imidazole

Gel filtration buffer

- 50 mM Ches-NaOH pH 9.0
- 0.25 M NaCl

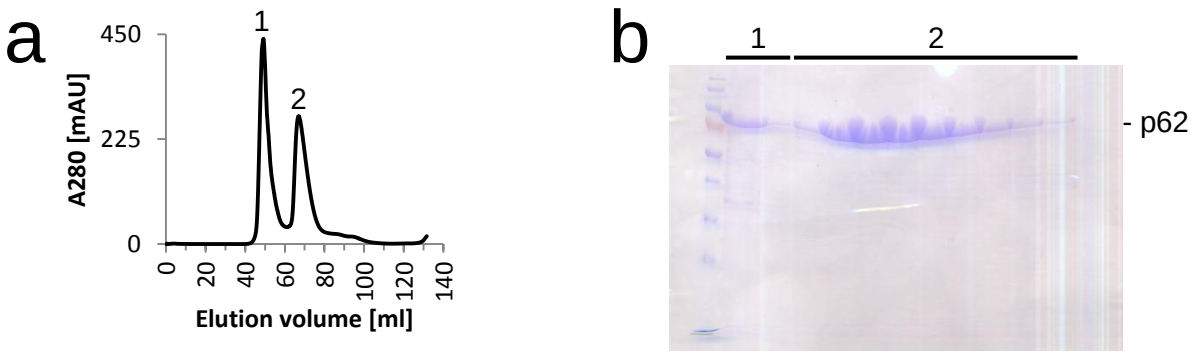


Figure 3-23. Purification of p62 from ArcticExpress (DE3) RIL cells with a Ches buffer system. (a) SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(b)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (a).

3.4.3. Crystallization

Protein purified with the standard buffer system and the Ches buffer system was subjected to extensive crystallization trials. Initially, crystallization with the crystallization robot, using the commercial crystallization screens was attempted. With the standard buffer system, crystals could be obtained at a concentration of 6.7 mg/ml with Crystal Screen HT™, PEGs Suite, and OptiMix™ PEG (Figure 3-24).

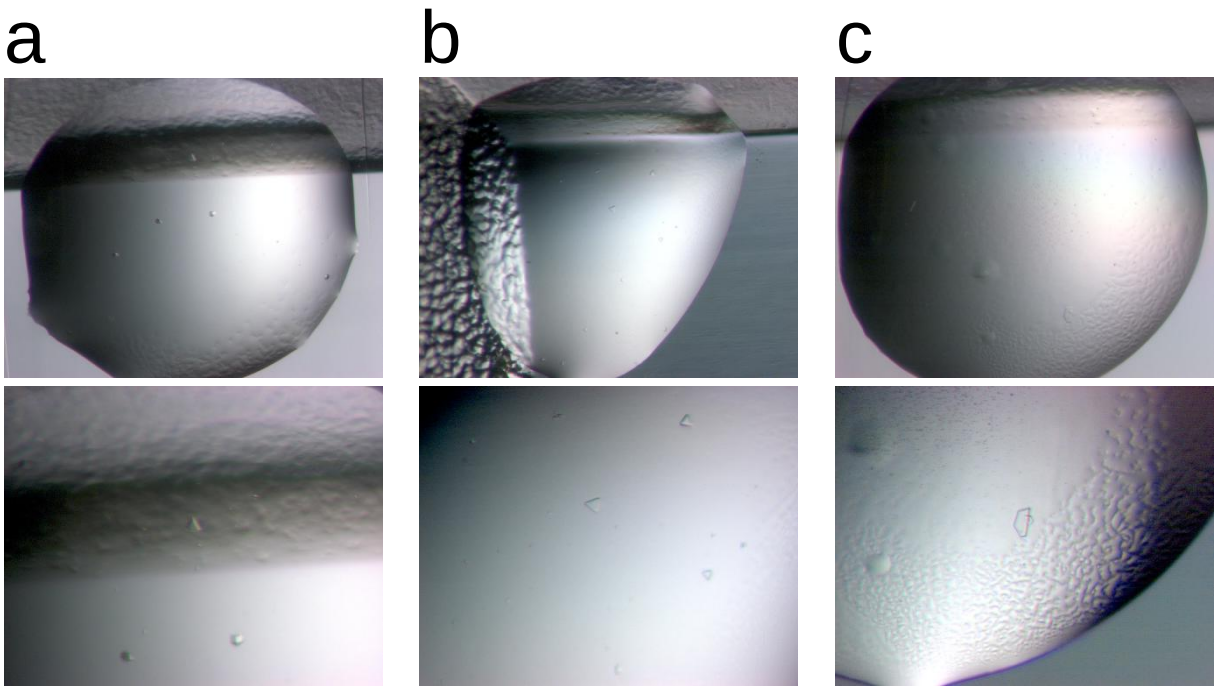


Figure 3-24. Initial crystallization trials with full-length p62. (a) Crystal Screen HT™. Reservoir solution: 0.1 M Tris-HCl pH 8.5; 0.2 M $MgCl_2$; 30 % (w/v) PEG 4000. **(b)** PEGs Suite. Reservoir solution: 0.2 M lithium citrate; 20 % (w/v) PEG 3350. **(c)** OptiMix™ PEG. Reservoir solution: 0.8 M potassium nitrate; 27 % (w/v) PEG 3350. Top row: overview; bottom row: zoom in.

Fine screens in 96 well format based on the initial crystallization conditions were designed to obtain bigger crystals. Different protein concentrations were tested, ranging from 5 to 30 mg/ml, but to no avail. Crystallization was also pursued in the 24 well format using the hanging drop method. However, the crystals could not be improved by this approach (Figure 3-25).

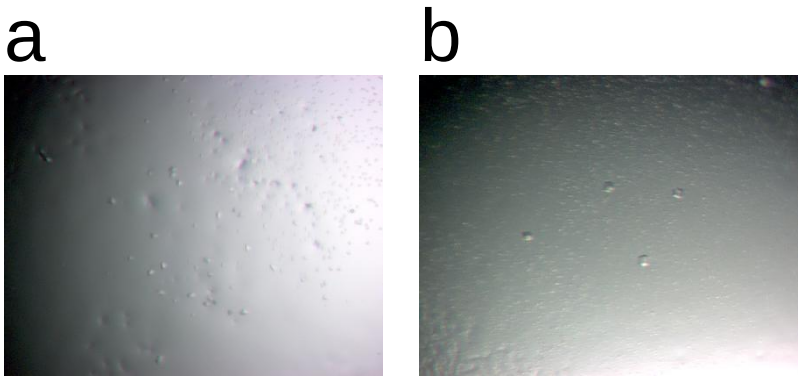


Figure 3-25. Crystallization of p62 with fine screens derived from the initial crystallization conditions. (a) Reservoir solution: 0.1 M Tris-HCl pH 8.5; 0.2 M $MgCl_2$; 20 % (w/v) PEG 4000. **(b)** Reservoir solution: 0.4 M $CaCl_2$; 30 % (w/v) PEG 3350.

Crystallization trays in the 24 well format at different temperatures (4 °C, 12 °C, 20 °C, and 37 °C) were set up as well. Crystals grew at 4 °C and 20 °C, but crystal the quality could not be improved. Additionally, additive screens were tested, as well as microseeding with

crystals from the initial or fine screen conditions. These approaches did not improve crystal quality, and the crystals obtained from seeding are shown in Figure 3-26. Further parameters varied were the salt concentration in the protein buffer (from 0 to 0.25 M NaCl) and the drop sizes (0.5 to 2 µl).

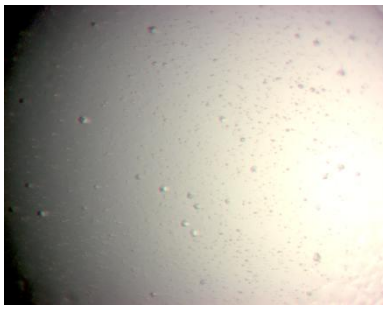


Figure 3-26. Crystals of p62, obtained by microseeding. Reservoir solution: 0.4 M $CaCl_2$; 20 % (w/v) PEG 3350.

Crystals were harvested and subjected to synchrotron radiation at beamline ID23-2 at the ESRF. Protein crystal diffraction could be obtained, but the highly mosaic pattern rendered all processing attempts unsuccessful (Figure 3-27).

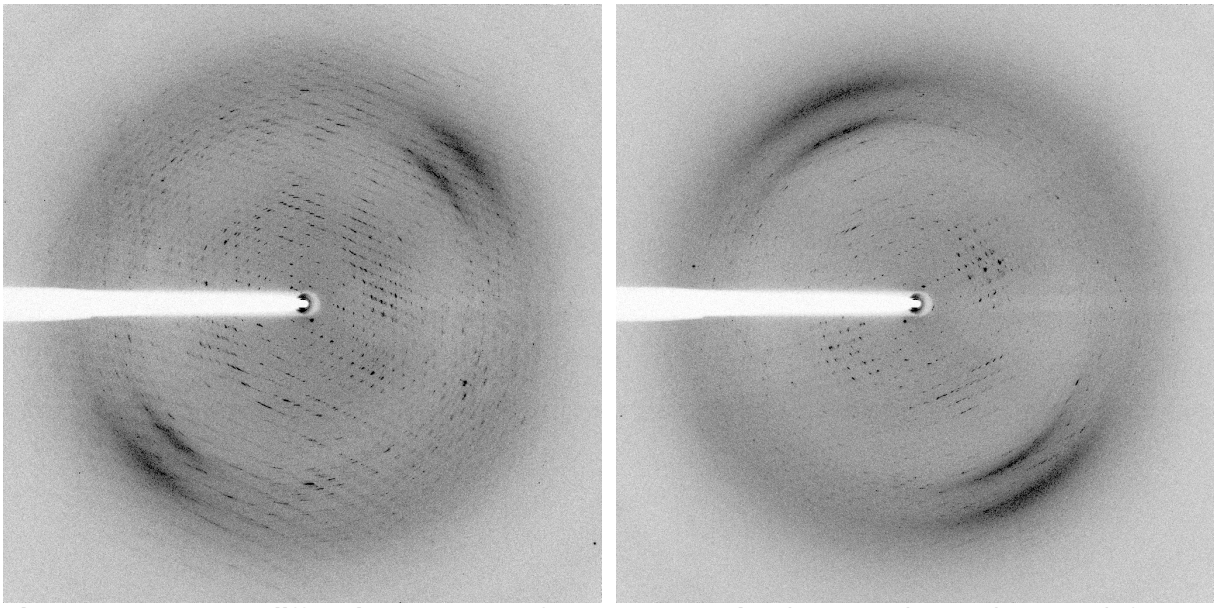


Figure 3-27. X-ray diffraction pattern of a p62 crystal. The crystal was harvested from a crystallization condition consisting of 0.2 M $CaCl_2$ and 15 % (w/v) PEG 3350 in the reservoir. Two orientations of the crystal in the beam, 90° apart from each other, are shown.

3.4.4. p62_1-109

Human p62 contains a PH domain comprising residues 1-108 [41]. A construct of the PH domain of p62 from *C. thermophilum* was designed. According to sequence alignments, residue 108 would correspond to residue 155 in *C. thermophilum* (see 6.3.2). However, this residue range would also contain an insertion ranging from 106-149, which is unstructured

according to secondary structure predictions (Figure 3-19, see also 6.3.2). It was therefore assumed that the C-terminal boundary of the *C. thermophilum* PH domain is at residue 105, and a construct p62_1-109, encompassing the first 109 amino acids, was cloned. P62_1-109 was expressed in ArcticExpress (DE3) RIL cells and purified with a Ches buffer system (Figure 3-28). Buffer compositions are given below.

Lysis buffer

- 50 mM Ches-NaOH pH 9.0
- 0.3 M NaCl
- 5 mM Imidazole

Elution buffer

- 50 mM Ches-NaOH pH 9.0
- 0.3 M NaCl
- 0.25 M Imidazole

Gel filtration buffer

- 20 mM Ches-NaOH pH 9.0
- 0.25 M NaCl

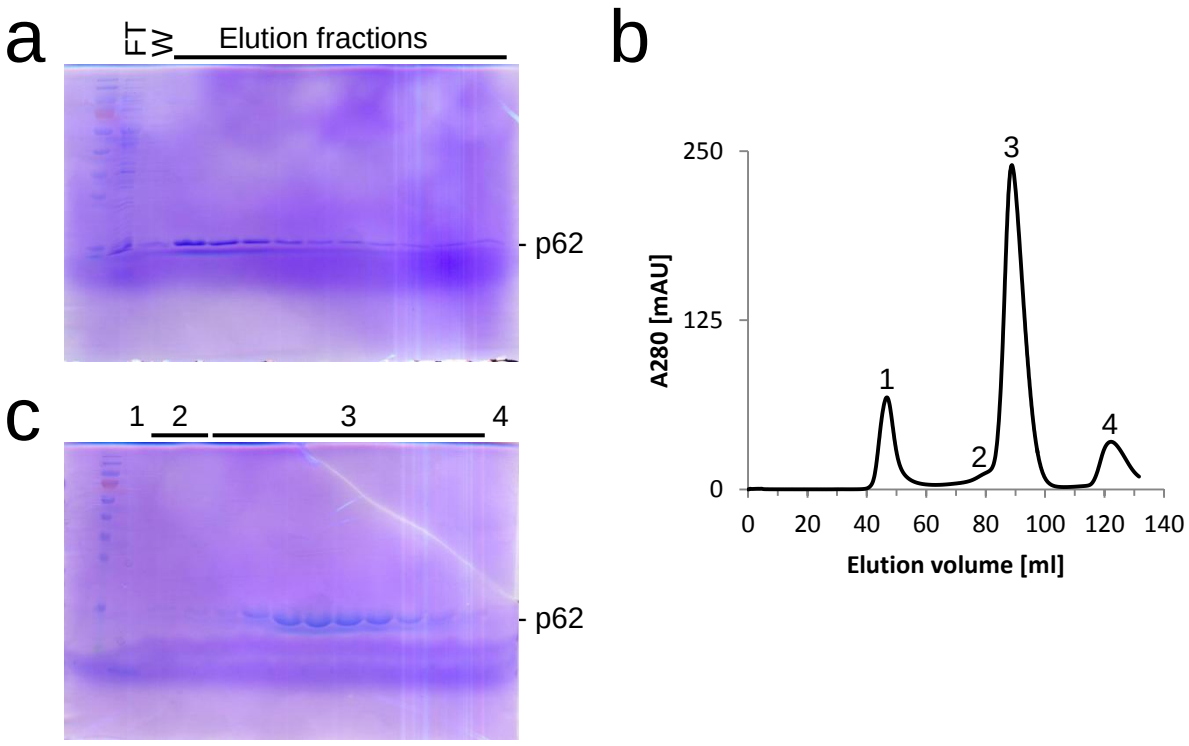


Figure 3-28. Purification of p62_1-109 from ArcticExpress (DE3) RIL cells with a Ches buffer system. (a) SDS-PAGE analysis of IMAC samples. FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

The construct was used for a crystallographic study, exploring cesium for experimental phasing. This study will be described in detail in section 3.10.

3.4.5. p62_1-285

In addition to the PH domain, human p62 contains a BSD tandem comprising residues 109-232 [1]. A construct of p62 from *C. thermophilum*, encompassing the PH domain and the BSD tandem was designed. According to sequence alignments, this region includes the first 281 amino acids of *C. thermophilum* p62 (see 6.3.2). The construct p62_1-285, encompassing the first 285 amino acids, was cloned, expressed in ArcticExpress (DE3) RIL cells, and purified according to the standard protocol (Figure 3-29). In the SDS-PAGE a double band was observed at the expected molecular weight (Figure 3-29 a, c), the cause of which was unclear. The protein was subjected to crystallization trials at a concentration of 9.7 mg/ml with the crystallization robot using the commercial crystallization screens, but yielded no crystals.

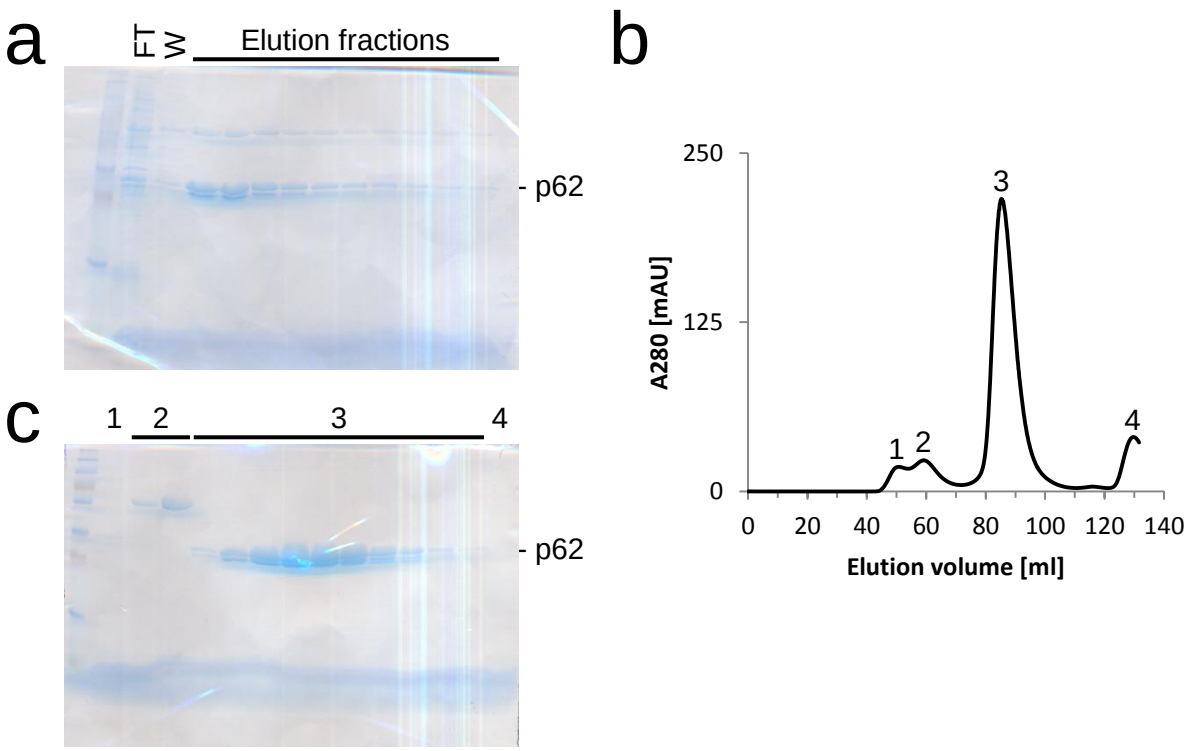


Figure 3-29. Purification of p62_1-285 from ArcticExpress (DE3) RIL cells. (a) SDS-PAGE analysis of IMAC samples. FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

3.4.6. p62_145-290

A construct encompassing only the BSD tandem was designed, ranging from residue 145 to 290. P62_145-290 was expressed in ArcticExpress (DE3) RIL cells and purified according to the standard procedure (Figure 3-30). Similar to the p62_1-285 construct (Figure 3-29 a, c), the protein appeared as a double band in an SDS-PAGE (Figure 3-30 a, c). Crystallization

trials at a concentration of 11.3 mg/ml with the crystallization robot were performed using the commercial crystallization screens, but yielded no crystals.

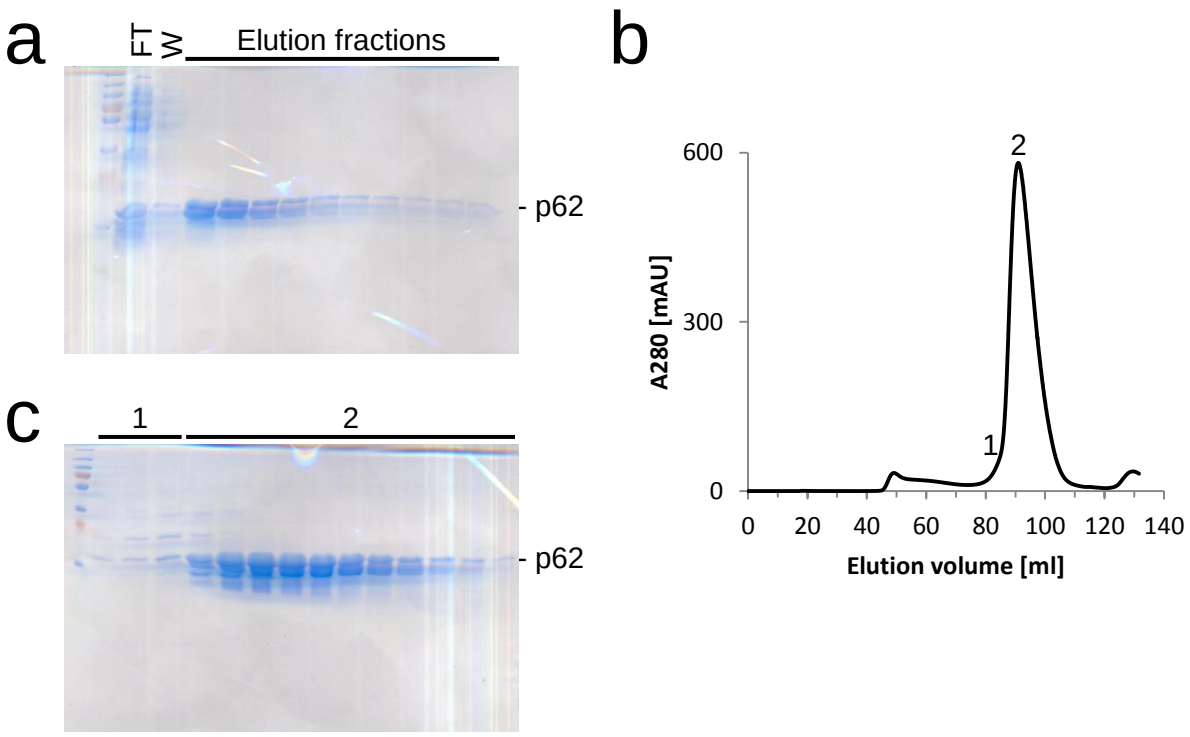


Figure 3-30. Purification of p62_145-290 from ArcticExpress (DE3) RIL cells. (a) SDS-PAGE analysis of IMAC samples. FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

3.4.7. p62_290-E

To investigate the C-terminal half of p62, especially the interaction of the C-terminus with p44, as shown for the *Saccharomyces cerevisiae* proteins [46], the construct p62_290-E, starting after the BSD tandem and encompassing the last 388 amino acids, was designed (Figure 3-19). The construct was expressed in ArcticExpress (DE3) RIL cells and purification according to the standard protocol with a Ches buffer system was attempted. The result from the IMAC purification is depicted in Figure 3-31, showing that no p62_290-E could be obtained. The Ches buffer system used is given below.

Lysis buffer

- 20 mM Ches-NaOH pH 9.0
- 0.3 M NaCl
- 5 mM Imidazole

Elution buffer

- 20 mM Ches-NaOH pH 9.0
- 0.3 M NaCl
- 0.25 M Imidazole

Gel filtration buffer

- 20 mM Ches-NaOH pH 9.0
- 0.25 M NaCl

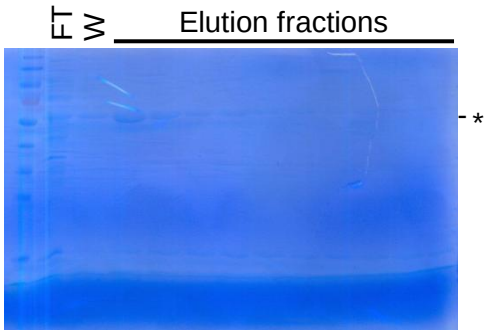


Figure 3-31. SDS-PAGE analysis of IMAC samples of p62_290-E purified from ArcticExpress (DE3) RIL cells. FT: flow through; W: wash; *: chaperonin.

3.4.8. p62_435-E

As the purification of the C-terminal construct p62_290-E was not successful (see 3.4.7), a shorter C-terminal construct was designed, leaving out residues 290-434, which may be disordered (Figure 3-19). Construct p62_435-E, encompassing the last 435 amino acids, was expressed in ArcticExpress (DE3) RIL cells and purified according to the standard protocol (Figure 3-32). P62_435-E eluted from the SEC column in two species, at ~70 ml and ~120 ml, the cause of which was unclear (Figure 3-32 b). Fractions from both species were pooled and concentrated separately. The pool from the fractions at ~70 ml was concentrated to 5.5 mg/ml, the pool from fractions at ~120 ml was concentrated to 7.9 mg/ml. Both pools were subjected to crystallization trials with the crystallization robot using the commercial crystallization screens, but yielded no crystals.

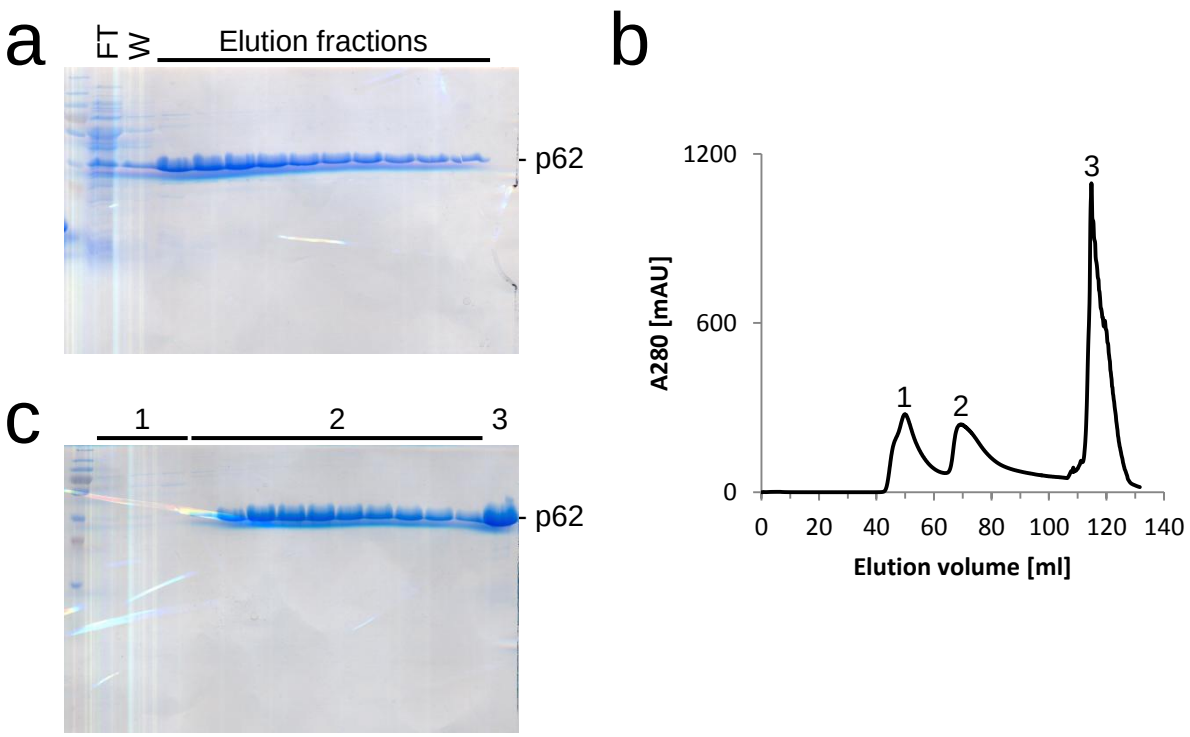


Figure 3-32. Purification of p62_435-E from ArcticExpress (DE3) RIL cells. (a) SDS-PAGE analysis of IMAC samples. FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

3.5. p52

3.5.1. Secondary structure prediction and construct design

To improve crystallization propensities and to investigate interactions of different protein regions with other TFIIH subunits, different constructs of p52 were designed (Figure 3-33). Secondary structure predictions and sequence alignments with the human p52 sequence were performed, which formed the basis for construct design (Figure 3-33, see also 6.3.3). Two XPB binding regions within human p52 are known (residues 1-135 and 304-381) [25]. Hence, constructs of p52 from *C. thermophilum* were designed lacking either one or the other binding site. According to sequence alignments the first XPB binding site would encompass roughly the first 130 amino acids (see 6.3.3). Furthermore, secondary structure prediction proposed a domain encompassing the first 120 amino acids (Figure 3-33). Thus, a construct lacking the first 120 amino acids (p52_121-E) was designed. Based on sequence alignments the second XPB binding site would begin roughly at amino acid 350. Secondary structure prediction proposed a domain at the C-terminal end of the protein starting at residue 360 (Figure 3-33). Thus, another construct lacking the last 165 amino acids (p52_1-349) was designed. The constructs p52_121-E and p52_1-349 have been cloned by Elisabeth Schönwetter. Sequence alignments revealed furthermore, that *C. thermophilum* p52 contains a 24 amino acids long insertion (residues 322-345) compared to human p52

(see 6.3.3). Correspondingly, this region is predicted as unstructured (Figure 3-33). Therefore, amino acids 322-344 were deleted in full-length p52 as well as p52_121-E and replaced by an SNGNG linker (termed delta linker, dL), resulting in constructs p52_dL and p52_121-E_dL, respectively (Figure 3-33).

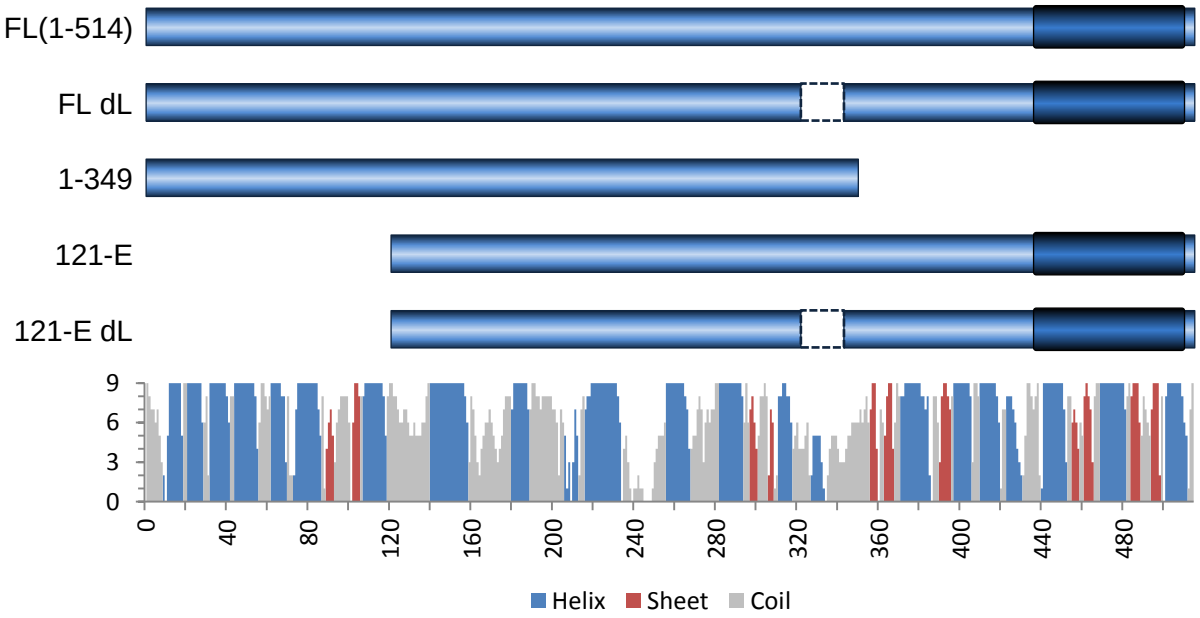


Figure 3-33. p52 constructs and secondary structure prediction. The construct schemes are aligned to the secondary structure prediction at the bottom. The domain scheme corresponds to Figure 1-1. Secondary structure prediction was performed with I-TASSER. dL: delta linker, denotes the deletion of residues 322-344 and the replacement with a short SNGNG linker.

3.5.2. Expression and purification

P52 was expressed in B21 CodonPlus (DE3) RIL cells and purified following the standard protocol. SDS-PAGE analysis of the purification and the SEC elution profile are depicted in Figure 3-34.

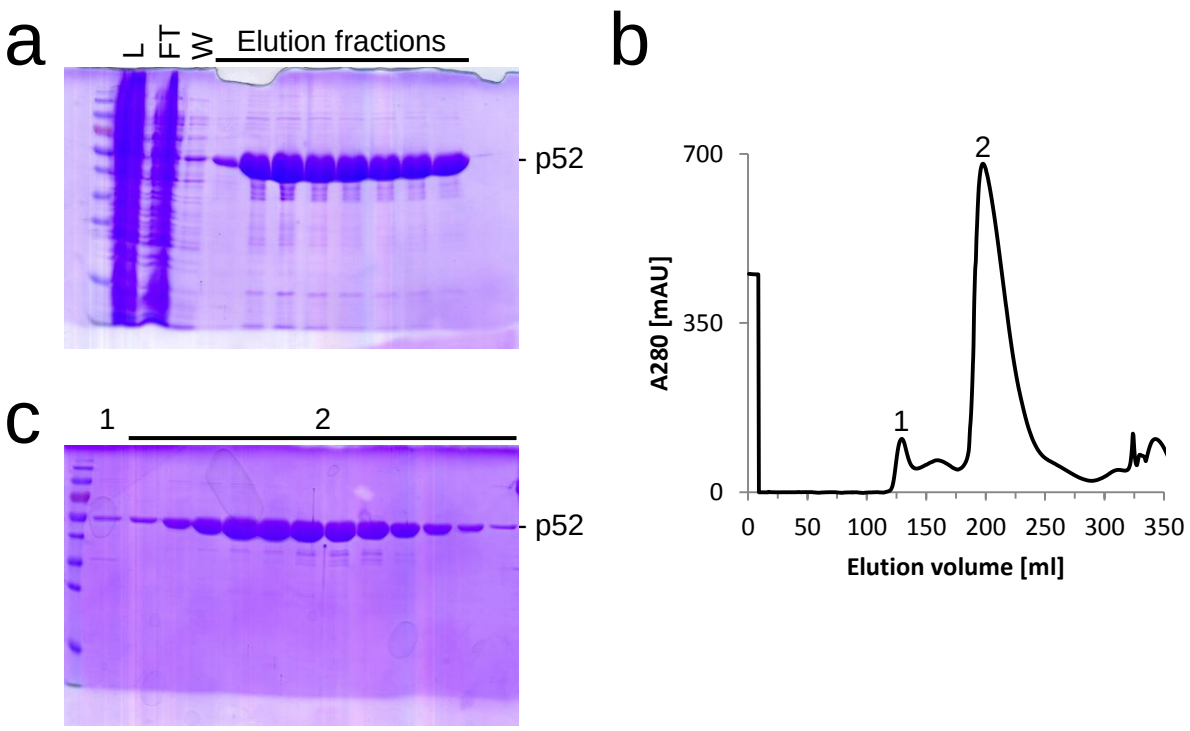


Figure 3-34. Purification of p52 from BL21 CodonPlus (DE3) RIL cells. (a) SDS-PAGE analysis of IMAC samples. L: lysate; FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 26/60 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

SEC was also performed in a Ches buffer, and the elution profile is shown in Figure 3-35. The composition of the Ches buffer is given below.

Ches buffer

- 50 mM Ches-NaOH pH 9.0
- 0.25 M NaCl

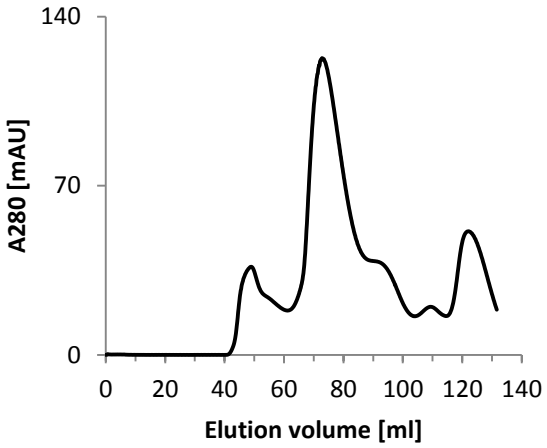


Figure 3-35. SEC elution profile of p52 in Ches buffer from a HiLoad 16/600 Superdex 200 pg column.

3.5.3. Crystallization

P52 in standard as well as Ches buffer was subjected to extensive crystallization trials. Initial plates were set up with the crystallization robot using the commercial crystallization screens. For p52 in standard buffer, small crystals could be obtained at 2.1 mg/ml; for p52 in Ches buffer, small crystals could be obtained at 2.9 mg/ml with Index HT™ and Crystal Screen HT™ (Figure 3-36).

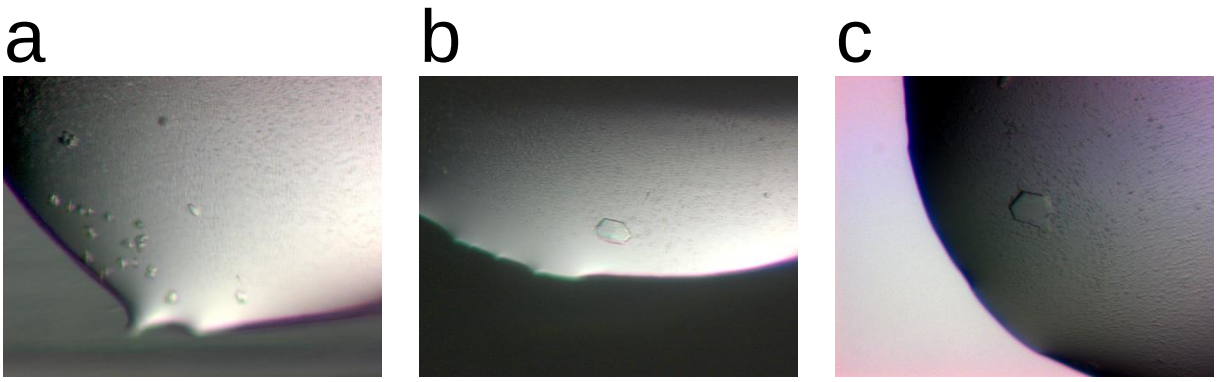


Figure 3-36. Initial crystallization trials with p52. (a) Index HT™. Reservoir solution: 0.1 M Bis-Tris pH 6.5; 0.2 M lithium sulfate; 25 % (w/v) PEG 3350. **(b)** Index HT™. Reservoir solution: 0.1 M Hepes pH 7.5; 0.2 M lithium sulfate; 25 % (w/v) PEG 3350. **(c)** Crystal Screen HT™. Reservoir solution: 0.1 M Mes pH 6.5; 12 % (w/v) PEG 20000.

Initial conditions were fine screened in the 96 well format using the crystallization robot, as well as in the 24 well format using hanging drop. Additionally, different protein concentrations (ranging from 1.5 to 7.5 mg/ml) and temperatures (4 °C, 20 °C, 37 °C) were tested. However, the size or diffraction properties of the crystals could not be significantly improved (Figure 3-37).

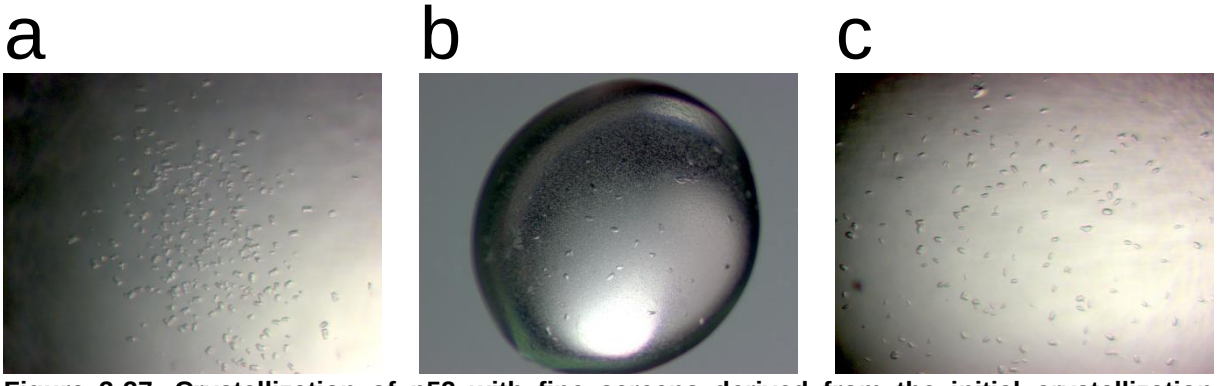


Figure 3-37. Crystallization of p52 with fine screens derived from the initial crystallization conditions. (a) Reservoir solution: 0.1 M Bis-Tris pH 5.5; 0.3 M ammonium sulfate; 20 % (w/v) PEG 3350. **(b)** Reservoir solution: 0.1 M Bis-Tris pH 5.5; 0.3 M lithium sulfate; 15 % (w/v) PEG 3350. **(c)** Reservoir solution: 0.1 M Bis-Tris pH 6.5; 0.3 M lithium sulfate; 15 % (w/v) PEG 3350.

P52 crystals were exposed to synchrotron radiation at beamline P14 of PETRAIII. Protein diffraction patterns could be observed, yet the resolution was limited to about 6-7 Å (Figure 3-38).

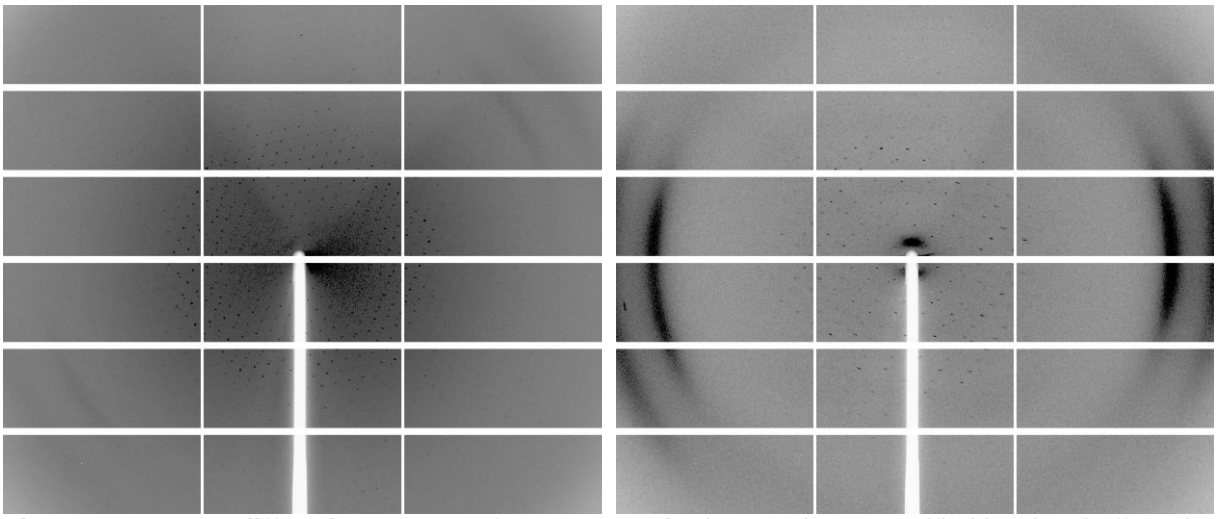


Figure 3-38. X-ray diffraction pattern of a p52 crystal. The protein was purified in Ches buffer. The crystal was harvested from a crystallization condition consisting of 0.1 M Hepes pH 7.5, 0.2 M lithium sulfate, and 25 % (w/v) PEG 3350 in the reservoir. Two orientations of the crystal in the beam, 90° apart from each other, are shown. The outmost areas of the detector are omitted.

3.6. p44

3.6.1. Secondary structure prediction and construct design

Different constructs of p44 were designed in order to improve crystallization propensities and to investigate interactions of different protein regions with other TFIIH subunits (Figure 3-39). P44 is composed of three domains, an N-terminal vWA like domain, a central zinc finger domain, and a C-terminal ring finger domain [34, 35]. Secondary structure predictions and sequence alignments with the human p44 amino acid sequence were performed to map the domain boundaries known from human p44 to *C. thermophilum* p44 (Figure 3-39, see also 6.3.4). According to the alignments, *C. thermophilum* p44 was subdivided into three regions: 1-285 containing the N-terminal vWA like domain, 286-367 containing the central zinc finger domain, and 368-534 containing the C-terminal ring finger domain. Constructs lacking one or the other domain were designed accordingly (Figure 3-39). Construct p44_1-285, containing the N-terminal vWA like domain has been designed and cloned by Dominik Schmitt. To further investigate the central zinc finger domain, the constructs p44_1-326 and p44_327-E were designed, containing the first half or second half of the central domain, respectively (Figure 3-39). To determine a suitable position for splitting of the central zinc finger domain sequence alignments with the human p44 were used (see 6.3.4). Fribourg et al. reported the presence of a highly conserved C4 zinc finger motif (C-X_2-C-X_{10}-C-X_2-C) within the central domain [32]. This motif is also conserved in *C. thermophilum* encompassing residues C345,

C348, C359, and C362. The presence of additional histidines and cysteines was also considered when the construct was generated, to avoid an accidental destruction of the zinc finger. The first potential zinc coordinating amino acid is H328. Hence, the central domain was splitted after residue 326, so that p44_327-E contains all histidines and cysteines (including the conserved C4 zinc finger motif) present in the central zinc finger domain.

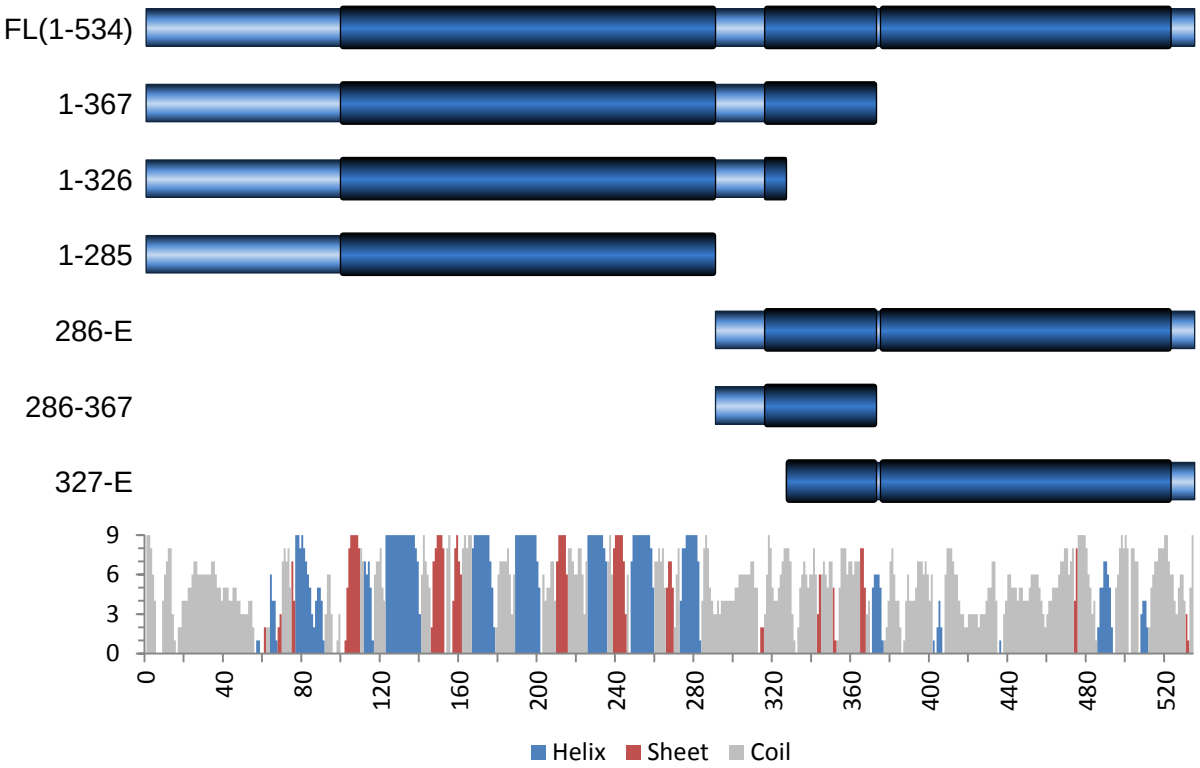


Figure 3-39. p44 constructs and secondary structure prediction. The construct schemes are aligned to the secondary structure prediction at the bottom. The domain scheme corresponds to Figure 1-1. Secondary structure prediction was performed with I-TASSER.

3.7. p34

3.7.1. Secondary structure prediction and construct design

P34 is composed of an N-terminal vWA like domain and a C-terminal zinc finger domain [31, 47]. Two constructs, containing either the vWA like domain or the C-terminal zinc finger domain, have been designed (Figure 3-40). Construct p34_1-277 has been designed and cloned by Dominik Schmitt, construct p34_278-E by Elisabeth Schönwetter.

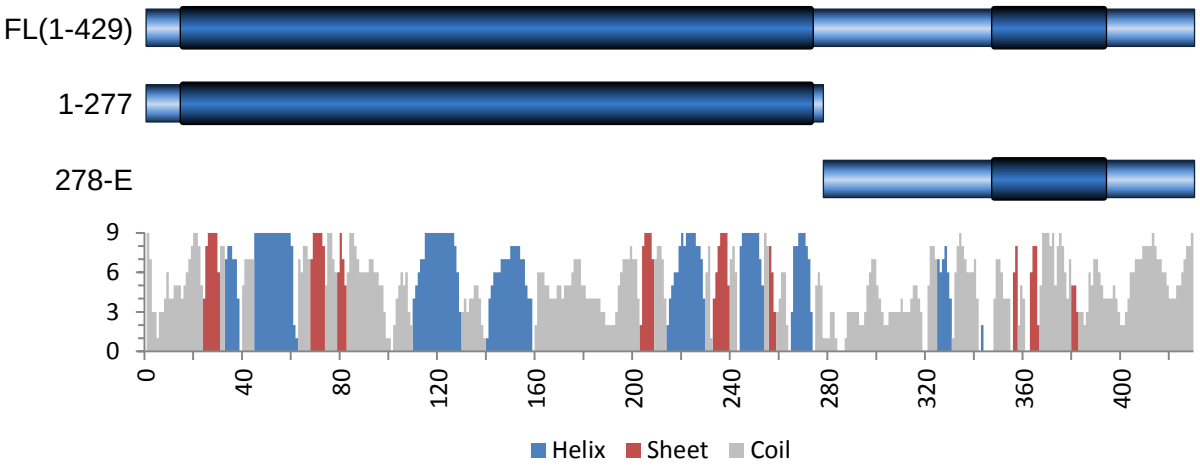


Figure 3-40. p34 constructs and secondary structure prediction. The construct schemes are aligned to the secondary structure prediction at the bottom. The domain scheme corresponds to Figure 1-1. Secondary structure prediction was performed with I-TASSER.

3.7.2. Expression and purification

Expression and purification was performed with a buffer system established by Dominik Schmitt. The protein was expressed in BL21 CodonPlus (DE3) RIL cells and purified according to the standard procedure. The result from the SEC is depicted in Figure 3-41, and the buffer compositions are given below.

Lysis buffer

- 50 mM Tris-HCl pH 8.0
- 0.15 mM KCl
- 0.5 mM TCEP

Elution buffer

- 50 mM Tris-HCl pH 8.0
- 0.15 M KCl
- 0.15 M Imidazole
- 1 mM TCEP

Gel filtration buffer

- 20 mM Tris-HCl pH 8.0
- 0.15 M KCl
- 1 mM TCEP

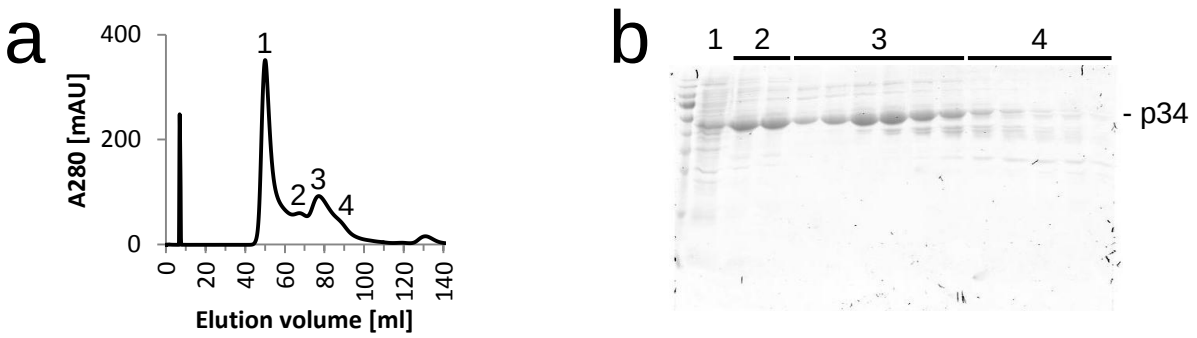


Figure 3-41. Purification of p34. (a) SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(b)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (a).

3.8. p8

3.8.1. Expression and purification

P8 was expressed from BL21 CodonPlus (DE3) RIL cells and purified according to the standard protocol. High yield and purity could be obtained, therefore no further optimization was necessary. The result from the SEC is depicted in Figure 3-42.

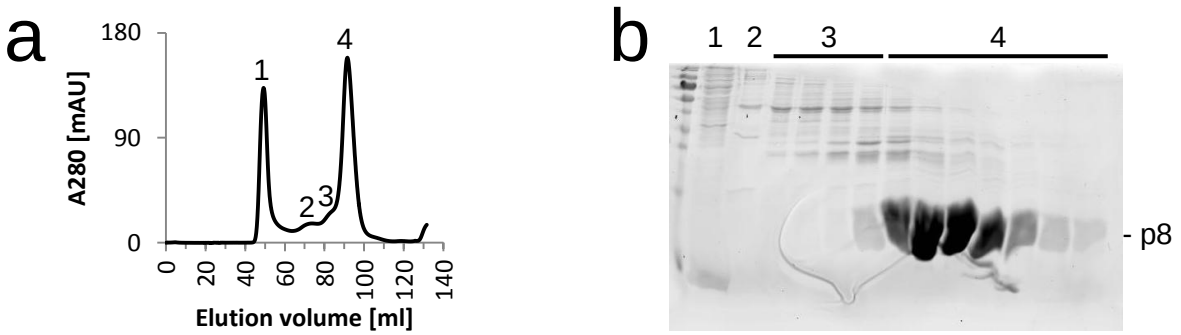


Figure 3-42. Purification of p8. (a) SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(b)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (a).

3.9. MAT1

3.9.1. Secondary structure prediction and construct design

To investigate the interaction between the CAK subunit MAT1 and core TFIIH, the construct MAT1_1-248 was used, lacking the C-terminal hydrophobic domain (Figure 3-43). MAT1_1-248 has been designed, cloned, and purified by Florian Sauer.

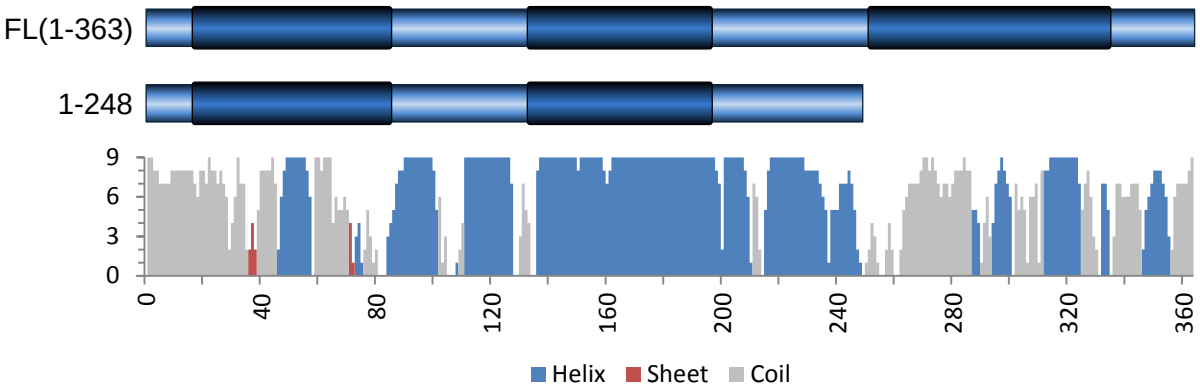


Figure 3-43. MAT1 constructs and secondary structure prediction. The construct schemes are aligned to the secondary structure prediction at the bottom. The domain scheme corresponds to Figure 1-1. Secondary structure prediction was performed with I-TASSER.

3.10. Cesium for experimental phasing - A crystallographic study

To overcome the phase problem by means of SAD, a potent anomalous scatterer needs to be present in the crystal. In the course of this study, different stages for introduction of an anomalous scatterer were explored. The anomalous scatterer was included during protein purification, crystallization, or cryo-protection. A suitable heavy atom compound was required for this approach, which was ideally compatible with each of these stages. Common components of protein buffers and crystallization solutions are chloride salts of alkali metals, like NaCl or KCl. These salts are proposed to facilitate crystallization either by stabilizing the protein in solution (salting-in) or acting as precipitant (salting-out) [452, 453]. Elements in the periodic table ordered in groups generally display similar chemical properties. The group of alkali metals consists of lithium (Li), sodium (Na), potassium (K), rubidium (Rb), cesium (Cs), and francium (Fr) (Figure 3-44). The chemical similarity within this group is apparent from the fact that monovalent chloride salts of all alkali metals exist. Within the group the number of electrons and anomalous scattering propensity increases downwards. The heaviest group member, Fr, possesses the strongest anomalous scattering. However, all known isotopes are radioactive, therefore, Fr was not considered further in this study. The second heaviest member of the alkali metals, Cs, is also a very potent anomalous scatterer. At an x-ray energy of 7 keV (1.7712 Å) the contribution of f'' is about 10 e^- (Figure 3-45). These anomalous properties, combined with the chemical similarity to NaCl and KCl, made CsCl a suitable candidate for introduction during purification, crystallization, and cryo-protection.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1 H																	2 He
3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne
11 Na	12 Mg											13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc*	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
55 Cs	56 Ba	57 La	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po*	85 At*	86 Rn*
87 Fr*	88 Ra*	89 Ac*	104 Rf*	105 Db*	106 Sg*	107 Bh*	108 Hs*	109 Mt*	110 Ds*	111 Rg*	112 Cn*	113 Nh*	114 Fl*	115 Mc*	116 Lv*	117 Ts*	118 Og*

Figure 3-44. Periodic table. Cesium is the heaviest not radioactive member of the alkali metals. *: No stable isotope known.

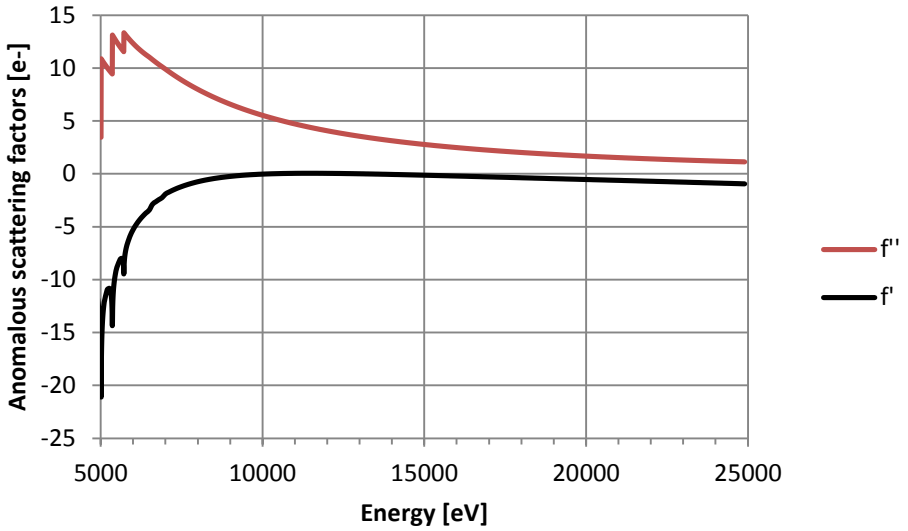


Figure 3-45. Anomalous scattering of cesium. The components f' and f'' of the anomalous scattering of cesium are plotted against the x-ray energy. The plot was generated with data from [461].

3.10.1. Proteins, purification and crystallization

To test the broader applicability of this approach, three proteins were considered in this study. First, p62_1-109 as a novel target, the structure not yet determined by other means. Second, hen egg white lysozyme (HEWL) as protein well characterized for methodical studies in crystallography. Third, p34_1-277 as a medium resolution target, as it yields crystals diffracting to about 3 Å [47].

To keep track of the datasets of the different proteins, numbers are introduced in Table 3-3 (p62_1-109), Table 3-4 (HEWL), and Table 3-5 (p34_1-277). These numbers are used to refer to the corresponding crystal/dataset in other tables and the text.

P62_1-109 was purified as described in section 3.4.4. For introduction of cesium during purification, CsCl buffer was used instead of NaCl buffer for the SEC. Usage of CsCl during the SEC seemed to have no influence on the purification compared to NaCl buffer and is depicted in Figure 3-46. The composition of both buffers is given below.

NaCl buffer

- 20 mM Ches-NaOH pH 9.0
- 0.25 M NaCl

CsCl buffer

- 20 mM Ches-NaOH pH 9.0
- 0.25 M CsCl

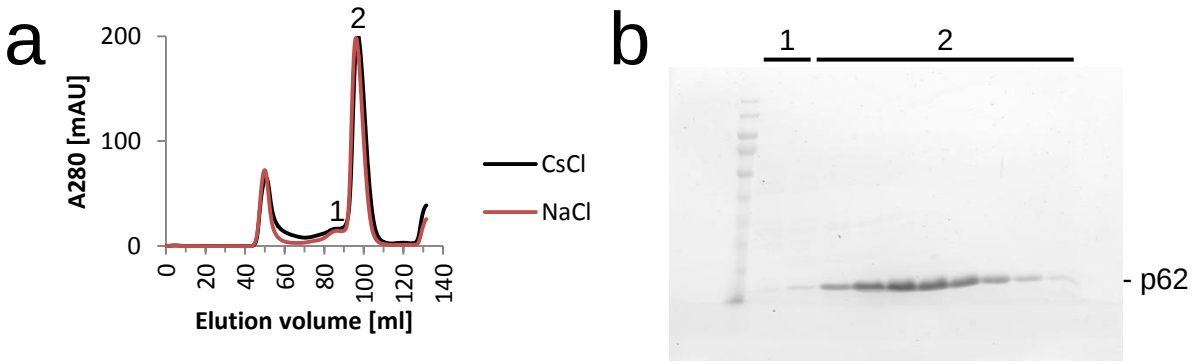


Figure 3-46. Purification of p62_1-109 in CsCl buffer. (a) SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. The curve is overlaid with an elution profile from a purification with NaCl buffer. **(b)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (a).

P62_1-109 was concentrated to 11-13 mg/ml and subjected to crystallization in 24 well plates utilizing the hanging drop vapor diffusion method. 1 µl of protein solution was mixed with 1 µl of mother liquor. Crystallization was attempted in presence of KCl or CsCl. As cryo-protectant glycerol was used. The precise crystallization and cryo-protectant conditions of crystals used in this study are given in Table 3-3.

Similar crystals of p62_1-109 grew in the presence of KCl or CsCl. The crystals had a plate like appearance with the long edges mostly ranging from 200 to 600 µm (Figure 3-47). The thickness of the plates was between 20 and 30 µm.

Table 3-3. Crystallization and cryo-protectant conditions of p62_1-109 crystals.

#	Protein buffer	Crystallization	Cryo-protectant
1	20 mM Ches pH 9.0 0.25 M CsCl	0.8 M KCl 17 % (w/v) PEG 4000	0.75 M KCl 17.5 % (w/v) PEG 4000 20 % (v/v) glycerol
2	20 mM Ches pH 9.0 0.25 M CsCl	0.6 M KCl 15 % (w/v) PEG 4000	0.25 M CsCl 0.5 M KCl 17.5 % (w/v) PEG 4000 20 % (v/v) glycerol
3	20 mM Ches pH 9.0 0.25 M NaCl	0.9 M KCl 17 % (w/v) PEG 4000	0.75 M CsCl 17.5 % (w/v) PEG 4000 20 % (v/v) glycerol
4	20 mM Ches pH 9.0 0.25 M CsCl	0.9 M KCl 17 % (w/v) PEG 4000	0.75 M CsCl 17.5 % (w/v) PEG 4000 20 % (v/v) glycerol
5	20 mM Ches pH 9.0 0.25 M NaCl	0.7 M CsCl 18 % (w/v) PEG 4000	0.7 M CsCl 18 % (w/v) PEG 4000 20 % (v/v) glycerol
6	20 mM Ches pH 9.0 0.25 M NaCl	0.9 M KCl 19 % (w/v) PEG 4000	0.9 M KCl 19 % (w/v) PEG 4000 20 % (v/v) glycerol

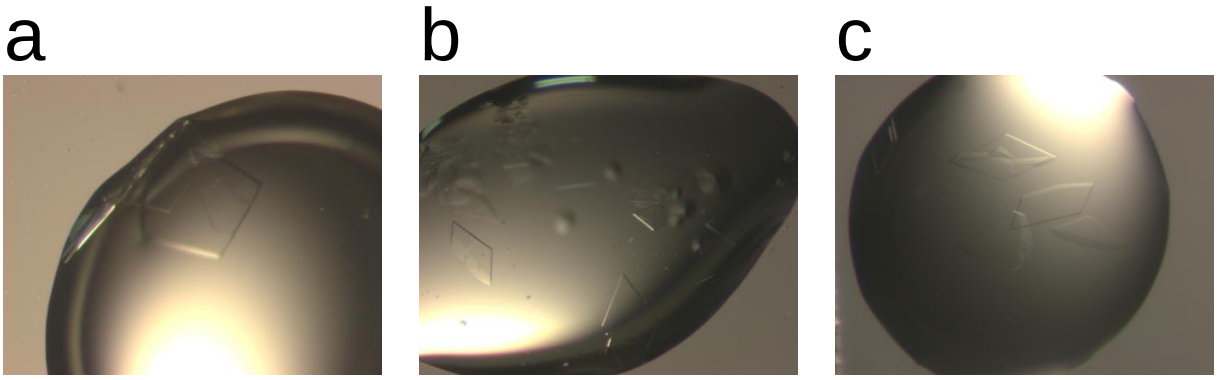


Figure 3-47. Crystallization of p62_1-109. (a) P62_1-109 in NaCl buffer, crystallized in KCl. **(b)** P62_1-109 in CsCl buffer, crystallized in KCl. **(c)** P62_1-109 in NaCl buffer, crystallized in CsCl.

HEWL was purchased as powder and dissolved at a concentration of 50 mg/ml in either purified water, 0.1 M NaAc pH 4.5, or 0.25 M CsCl solution. No further purification was performed. HEWL was crystallized utilizing the hanging drop vapor diffusion method in 24 well plates with 3 µl of protein mixed with 3 µl of mother liquor. Crystallization was attempted in the presence of NaCl, KCl, or CsCl. As cryo-protectant ethylene glycol was used. The precise crystallization and cryo-protectant conditions of crystals used in this study are given in Table 3-4.

Typical crystals of tetragonal HEWL were obtained in each alkali salt, with edge lengths mostly between 200 to 500 µm (Figure 3-48).

Table 3-4. Crystallization and cryo-protectant conditions of HEWL crystals.

#	Protein buffer	Crystallization	Cryo-protectant
1	0.25 M CsCl	50 mM NaAc pH 4.5 1.71 M NaCl	50 mM NaAc pH 4.5 1.71 M NaCl 15 % (w/v) EG[1)]
2	0.25 M CsCl	50 mM NaAc pH 4.5 1.71 M NaCl	50 mM NaAc pH 4.5 0.25 M CsCl 1.46 M NaCl 15 % (w/v) EG
3	H_2O	50 mM NaAc pH 4.5 1.71 M NaCl	50 mM NaAc pH 4.5 1.71 M CsCl 15 % (w/v) EG
4	0.25 M CsCl	50 mM NaAc pH 4.5 1.71 M NaCl	50 mM NaAc pH 4.5 1.71 M CsCl 15 % (w/v) EG
5	H_2O	50 mM NaAc pH 4.5 1.5 M CsCl	50 mM NaAc pH 4.5 1.5 M CsCl 20 % (w/v) EG
6	H_2O	50 mM NaAc pH 4.5 1.5 M KCl	50 mM NaAc pH 4.5 1.5 M KCl 20 % (w/v) EG
7	0.1 M NaAc pH 4.5	0.1 M NaAc pH 4.5 3 M KCl 15 % (w/v) EG	-
8	0.1 M NaAc pH 4.5	0.1 M NaAc pH 4.5 1.71 M NaCl 15 % (w/v) EG	-

1) Ethylene glycol.

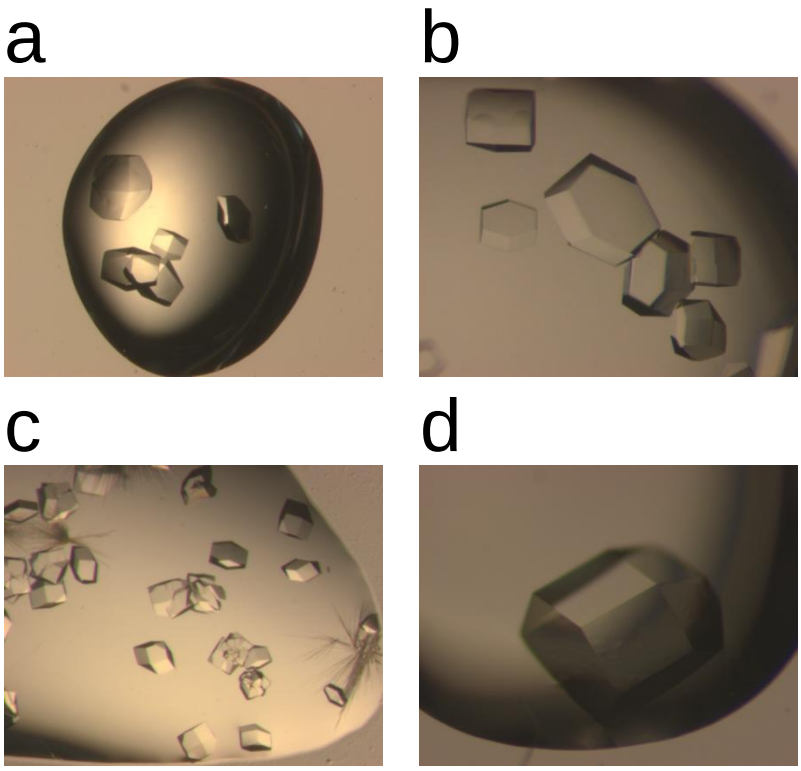


Figure 3-48. Crystallization of HEWL. (a) HEWL dissolved in H_2O, crystallized in NaCl. **(b)** HEWL dissolved in H_2O, crystallized in KCl. **(c)** HEWL dissolved in CsCl, crystallized in NaCl. **(d)** HEWL dissolved in H_2O, crystallized in CsCl.

P34_1-277 purified in KCl buffer was provided by Stefan Peißert. P34_1-277 in CsCl buffer was expressed and purified according to the standard procedure (see 2.2.5 and 2.2.6). The KCl buffer system was established by Dominik Schmitt. BL21 CodonPlus (DE3) RIL cells were used for protein expression. SEC in CsCl buffer appeared to be similar to the purification in KCl buffer and is depicted in Figure 3-49. The composition of all buffers is given below.

Lysis buffer

- 20 mM Tris-HCl pH 8.0
- 0.5 M KCl
- 1 mM TCEP

Elution buffer

- 20 mM Tris-HCl pH 8.0
- 0.3 M KCl
- 0.25 M Imidazole
- 1 mM TCEP

KCl buffer

- 20 mM Tris-HCl pH 8.0
- 0.15 M KCl
- 1 mM TCEP

CsCl buffer

- 20 mM Tris-HCl pH 8.0
- 0.2 M CsCl
- 1 mM TCEP

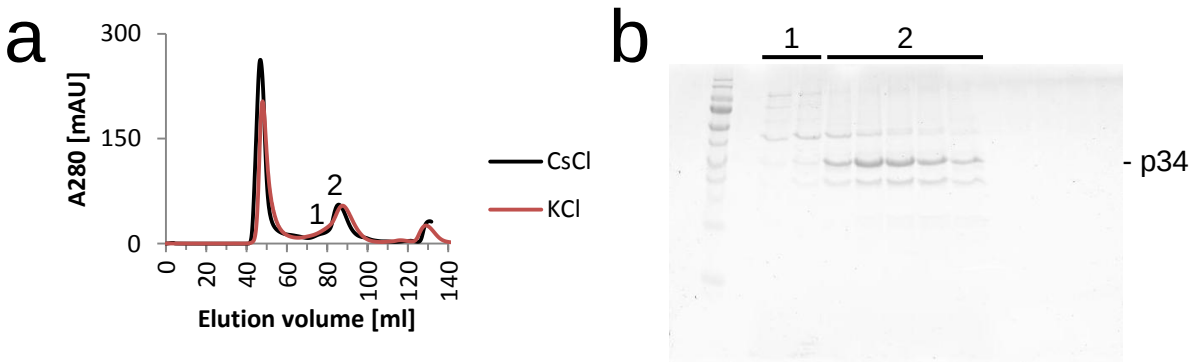


Figure 3-49. Purification of p34_1-277 in CsCl buffer. (a) SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. The curve is overlaid with an elution profile from a purification with KCl buffer. **(b)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (a). The elution profile with KCl buffer was provided by courtesy of Dominik Schmitt.

P34_1-277 in KCl buffer was crystallized at a concentration of 4 mg/ml utilizing the hanging drop vapor diffusion method in 24 well plates. 2 µl of the protein solution was placed on the cover slide without the addition of mother liquor. The protein solution was equilibrated against 1 ml KCl buffer supplemented with NaCl. Crystals were cryo-protected with ethylene glycol. The crystallization and cryo-protectant conditions are given in Table 3-5.

Table 3-5. Crystallization and cryo-protectant conditions of p34_1-277 crystals.

#	Protein buffer	Crystallization	Cryo-protectant
1	20 mM Tris-HCl pH 8.0 0.15 M KCl 1 mM TCEP	20 mM Tris-HCl pH 8.0 0.15 M KCl 1 mM TCEP add 50 mM NaCl[1)]	10 mM Tris-HCl pH 8.0 0.75 M CsCl 25 % (v/v) EG[2)]
2	20 mM Tris-HCl pH 8.0 0.15 M KCl 1 mM TCEP	20 mM Tris-HCl pH 8.0 0.15 M KCl 1 mM TCEP add 0.5 M NaCl[1)]	10 mM Tris-HCl pH 8.0 0.75 M CsCl 25 % (v/v) EG

1) NaCl was supplemented to the mother liquor only.
2) Ethylene glycol.

The crystals had a cubic appearance with edge lengths of about 60 µm (Figure 3-50). Crystallization of p34_1-277 in CsCl buffer was attempted at concentrations of 4 mg/ml and 7.7 mg/ml applying the same procedure as for p34_1-277 in KCl buffer, but no crystals could be obtained.

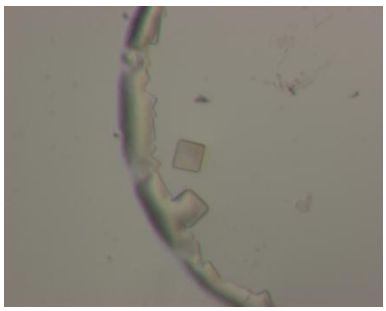


Figure 3-50. Crystallization of p34_1-277. P34_1-277 in KCl buffer, crystallized in KCl buffer.

3.10.2. Data collection, processing, phasing and refinement

Data was collected on single crystals using the rotation method. All datasets were indexed, integrated, and scaled with XDS. One single dataset was collected for each crystal, except for p62_1-109 crystals # 2 and 3. Here, two datasets at different positions of the same crystal were collected, processed with XDS, and then combined and brought to a common scale via XSCALE. Subsequently, the data were merged with Aimless. Data collection and processing statistics are given in Table 3-8 (p62_1-109), Table 3-9 (HEWL), and Table 3-10 (p34_1-277).

The phase problem for each processed dataset was solved by SAD, utilizing SHELXC/D/E, the Phenix AutoSol wizard [477], and the Phenix AutoBuild wizard [478]. For datasets where experimental phasing was not possible and necessary, phases were extracted from solved structures via rigid body refinement with Refmac5. The phasing procedure for each dataset, together with the programs used for the individual steps is given in Table 3-6 (p62_1-109) and Table 3-7 (HEWL). The crystals of p34_1-277 diffracted to only about 4 Å (Table 3-10) and displayed a weak anomalous signal. Experimental phasing was not possible with these

datasets. Merging of both datasets did not improve the anomalous signal to permit structure solution by SAD.

Table 3-6. Phasing procedure for the p62_1-109 datasets.

#	phasing method	substructure determination	initial phases	density modification	automated model building
1	RB[1)] (# 5)	-	-	-	-
2	SAD	HySS (AutoSol)[2)]	Phaser (AutoSol)	RESOLVE (AutoSol)	RESOLVE (AutoBuild)[2)]
3	SAD	HySS (AutoSol)	Phaser (AutoSol)	RESOLVE (AutoSol)	RESOLVE (AutoBuild)
4	SAD	SHELXD	SHELXE	SHELXE	Buccaneer
5	SAD	HySS (AutoSol)	Phaser (AutoSol)	RESOLVE (AutoSol)	RESOLVE (AutoSol)
6	RB (# 4)	-	-	-	-

1) Rigid body refinement using Refmac5. Number in parentheses gives the starting model for rigid body refinement.
2) The program was used through the AutoSol/AutoBuild wizard as specified.

Table 3-7. Phasing procedure for the HEWL datasets.

#	phasing method	substructure determination	initial phases	density modification	automated model building
1	SAD	SHELXD	Phaser (AutoSol)[2)]	RESOLVE (AutoSol)	RESOLVE (AutoBuild)[2)]
2	SAD	SHELXD	Phaser (AutoSol)	RESOLVE (AutoSol)	RESOLVE (AutoSol)
3	SAD	SHELXD	Phaser (AutoSol)	RESOLVE (AutoSol)	RESOLVE (AutoBuild)
4	SAD	SHELXD	Phaser (AutoSol)	RESOLVE (AutoSol)	RESOLVE (AutoSol)
5	SAD	SHELXD	Phaser (AutoSol)	RESOLVE (AutoSol)	RESOLVE (AutoBuild)
6	RB[1)] (# 8)	-	-	-	-
7	RB (# 8)	-	-	-	-
8	SAD	SHELXD	SHELXE	SHELXE	Buccaneer

1) Rigid body refinement using Refmac5. Number in parentheses gives the starting model for rigid body refinement.
2) The program was used through the AutoSol/AutoBuild wizard as specified.

Table 3-8. Data collection and refinement statistics of the p62_1-109 datasets.

#	1	2	3	4	5	6
Beamline	P14 (PETRAIII)	ID29 (ESRF)	ID29 (ESRF)	ID29 (ESRF)	ID29 (ESRF)	BM14 (ESRF)
Wavelength (Å)	1.7712	1.7712	1.7712	1.7712	1.77114	1.7712
Data collection and processing						
Spacegroup	C 222_1	C 222_1	C 222_1	C 222_1	C 222_1	C 222_1
Unit cell parameters						
a/b/c (Å)	34.46/106.52/75.90	39.74/102.30/75.09	40.25/101.76/75.10	34.27/106.78/75.32	34.43/106.79/75.87	34.74/106.17/76.21
α/β/γ (°)	90/90/90	90/90/90	90/90/90	90/90/90	90/90/90	90/90/90
Resolution range (Å)	75.90-1.90	42.27-1.90	42.12-1.90	43.56-1.90	43.67-1.80	43.56-2.50
Highest resolution shell (Å)	1.94-1.90	1.94-1.90	1.94-1.90	1.94-1.90	1.84-1.80	2.60-2.50
R_{merge}	0.080 (0.450)	0.093 (1.420)	0.123 (1.392)	0.071 (0.266)	0.150 (1.359)	0.142 (1.176)
R_{merge} (top intensity bin)	0.055	0.051	0.075	0.044	0.07	0.064
R_{pim}	0.017 (0.114)	0.019 (0.296)	0.025 (0.283)	0.021 (0.080)	0.026 (0.293)	0.041 (0.473)
Number of observations	236591 (9776)	287034 (15642)	300407 (17430)	125536 (7410)	417502 (10544)	63946 (3604)
Number of unique observations	11097 (643)	12135 (692)	12177 (739)	10885 (658)	12670 (487)	5119 (521)
<I/σI>	28.7 (6.9)	21.6 (2.1)	17.4 (2.2)	25.5 (9.2)	22.0 (2.7)	14.4 (1.6)
$CC_{1/2}$	0.999 (0.965)	0.999 (0.841)	0.998 (0.833)	0.999 (0.983)	0.999 (0.851)	0.997 (0.633)
Completeness (%)	97.4 (90.0)	97.8 (92.3)	97.4 (93.6)	96.3 (92.5)	94.9 (65.1)	99.1 (93.0)
Multiplicity	21.3 (15.2)	23.7 (22.6)	24.7 (23.6)	11.5 (11.3)	33.0 (21.7)	12.5 (6.9)
Refinement						
Resolution range (Å)	53.26-1.90	42.27-1.90	42.12-1.90	43.56-1.90	43.67-1.80	43.56-2.50
Number of reflections	10482	11528	11574	10293	12052	4844
R_{work}	0.162	0.193	0.198	0.184	0.183	0.202
R_{free}	0.224	0.223	0.221	0.216	0.215	0.266
R_{free} test set size	595	594	597	572	599	262
Coordinate error (Å)	0.088	0.101	0.1	0.081	0.074	0.242
Wilson B factor ($Å^2$)	23.082	32.232	31.7	17.737	10.914	26.03
B factor ($Å^2$)	28.638	43.584	45.198	23.412	19.978	41.74
Deviations from ideal values						
Bond (Å)	0.019	0.018	0.019	0.019	0.02	0.012
Angle (°)	2.004	1.826	1.93	1.94	2.108	1.598
Chirality (°)	0.121	0.115	0.128	0.107	0.133	0.091
Planarity (Å)	0.012	0.009	0.01	0.01	0.01	0.007
Ramachandran statistics (%)						
Favored	100	98.06	98.15	99.12	100	98.23
Allowed	0	1.94	1.85	0.88	0	1.77
Outlier	0	0	0	0	0	0

(Statistics for highest resolution shell are given in parentheses)

Table 3-9. Data collection and refinement statistics of the HEWL datasets.

#	1	2	3	4	5	6	7	8
Beamline	P14 (PETRAIII)	ID29 (ESRF)	P14 (PETRAIII)	P14 (PETRAIII)	ID29 (ESRF)	ID29 (ESRF)	BM14 (ESRF)	Homesource
Wavelength (Å)	1.7712	1.7712	1.7712	1.7712	1.71075	1.71075	1.7712	1.5418
Data collection and processing								
Spacegroup	$P\ 4_32_12$	$P\ 4_32_12$	$P\ 4_32_12$	$P\ 4_32_12$	$P\ 4_32_12$	$P\ 4_32_12$	$P\ 4_32_12$	$P\ 4_32_12$
Unit cell parameters								
a/b/c (Å)	78.63/78.63/36.76	78.95/78.95/36.87	78.75/78.75/36.81	78.97/78.97/36.81	75.93/75.93/35.71	75.86/75.86/35.69	78.98/78.98/36.88	78.88/78.88/36.96
α/β/γ (°)	90/90/90	90/90/90	90/90/90	90/90/90	90/90/90	90/90/90	90/90/90	90/90/90
Resolution range (Å)	39.31-1.90	39.48-1.90	39.38-1.79	39.48-1.78	37.96-1.75	37.93-1.80	36.88-2.00	19.13-1.47
Highest resolution shell (Å)	1.94-1.90	1.94-1.90	1.82-1.79	1.82-1.78	1.78-1.75	1.84-1.80	2.05-2.00	1.49-1.47
R_{merge}	0.061 (1.650)	0.057 (0.174)	0.104 (0.452)	0.152 (0.460)	0.157 (0.421)	0.094 (0.191)	0.043 (0.076)	0.042 (0.147)
R_{merge} (top intensity bin)	0.031	0.038	0.059	0.11	0.134	0.078	0.032	0.026
R_{pim}	0.013 (0.454)	0.012 (0.039)	0.023 (0.147)	0.033 (0.155)	0.033 (0.166)	0.020 (0.056)	0.009 (0.024)	0.006 (0.028)
Number of observations	187204 (8017)	214048 (12322)	221409 (4371)	222086 (4158)	223578 (4113)	215877 (6750)	194134 (6397)	1040626 (24175)
Number of unique observations	9509 (582)	9678 (619)	11383 (502)	11475 (508)	10988 (556)	10131 (560)	8322 (584)	20426 (914)
$\langle I/\sigma I \rangle$	35.8 (1.5)	43.4 (17.0)	24.9 (4.5)	18.1 (4.5)	18.3 (4.5)	26.2 (8.7)	70.4 (28.0)	84.9 (21.6)
$CC_{1/2}$	1.0 (0.684)	1.0 (0.994)	0.999 (0.881)	0.996 (0.882)	0.996 (0.902)	0.998 (0.985)	1.0 (0.998)	1.0 (0.993)
Completeness (%)	99.5 (96.1)	100.0 (100.0)	98.7 (78.6)	98.9 (80.4)	99.6 (95.2)	99.9 (98.1)	99.8 (98.6)	99.6 (93.1)
Multiplicity	19.7 (13.8)	22.1 (19.9)	19.5 (8.7)	19.4 (8.2)	20.3 (7.4)	21.3 (12.1)	23.3 (11.0)	50.9 (26.4)
Refinement								
Resolution range (Å)	39.31-1.90	39.48-1.90	39.38-1.79	39.48-1.78	37.96-1.75	37.93-1.80	36.88-2.00	19.13-1.47
Number of reflections	9017	9175	10537	10889	10370	9596	7890	19390
R_{work}	0.208	0.155	0.209	0.177	0.175	0.16	0.149	0.141
R_{free}	0.254	0.195	0.257	0.212	0.209	0.202	0.189	0.169
R_{free} test set size	462	469	530	542	583	501	399	994
Coordinate error (Å)	0.131	0.074	0.099	0.077	0.075	0.075	0.077	0.032
Wilson B factor (Å^2)	15.686	15.209	14.394	11.919	11.225	14.272	9.506	6.839
B factor (Å^2)	23.441	18.865	18.518	17.969	16.969	20.277	13.266	10.855
Deviations from ideal values								
Bond (Å)	0.014	0.018	0.018	0.02	0.028	0.026	0.019	0.027
Angle (°)	1.572	1.79	1.784	2.04	2.005	1.962	1.834	2.518
Chirality (°)	0.108	0.128	0.125	0.133	0.156	0.146	0.126	0.17
Planarity (Å)	0.007	0.01	0.01	0.011	0.01	0.01	0.01	0.012
Ramachandran statistics (%)								
Favored	99.21	97.64	98.43	99.21	98.43	99.21	98.43	99.21
Allowed	0.79	2.36	1.57	0.79	1.57	0.79	1.57	0.79
Outlier	0	0	0	0	0	0	0	0

(Statistics for highest resolution shell are given in parentheses)

Table 3-10. Data collection statistics of the p34_1-277 datasets.

#	1	2
Beamline	P14 (PETRAIII)	P14 (PETRAIII)
Wavelength (Å)	1.7712	1.7712
Data collection and processing		
Spacegroup	F 4_132	F 4_132
Unit cell parameters		
a/b/c (Å)	256.62/256.62/256.62	255.94/255.94/255.94
α/β/γ (°)	90/90/90	90/90/90
Resolution range (Å)	49.44-4.30	49.25-4.00
Highest resolution shell (Å)	4.81-4.30	4.47-4.00
R_{merge}	0.627 (4.748)	0.742 (6.513)
R_{merge} (top intensity bin)	0.089	0.098
R_{pim}	0.073 (0.544)	0.061 (0.523)
Number of observations	388277 (111345)	947785 (275126)
Number of unique observations	5340 (1467)	6506 (1790)
<I/σI>	10.0 (1.7)	11.8 (1.7)
$CC_{1/2}$	0.999 (0.599)	0.999 (0.522)
Completeness (%)	99.9 (100.0)	100.0 (100.0)
Multiplicity	72.7 (75.9)	145.7 (153.7)

(Statistics for highest resolution shell are given in parentheses)

The automatically build structures were manually completed and corrected in Coot, and refined with Refmac5. Refinement was performed directly against SAD data, with automated refinement of the anomalous substructure occupancy. Model stereochemistry was checked via the MolProbity server. Refinement and model statistics are given in Table 3-8 (p62_1-109) and Table 3-9 (HEWL).

3.10.3. The cesium substructure

Anomalous density maps generated by direct refinement against the SAD data were used as guidance for placement of the cesium ions. Peaks clearly exceeding the peak height of sulfur atoms from the methionines/cysteines present in the protein were generally attributed to cesium ions. Chloride ions were placed by comparison with anomalous peaks in datasets where no cesium was present. Potassium and chloride ions were distinguished by considering bonding distances, as distances of the former are generally shorter [479, 480]. The sodium ion in the HEWL dataset # 8 was placed after geometrical considerations (Figure 3-61 a) [479].

All cesium site positions that were identified within the different datasets were numbered and are depicted in Figure 3-51 (p62_1-109) and Figure 3-52 (HEWL).

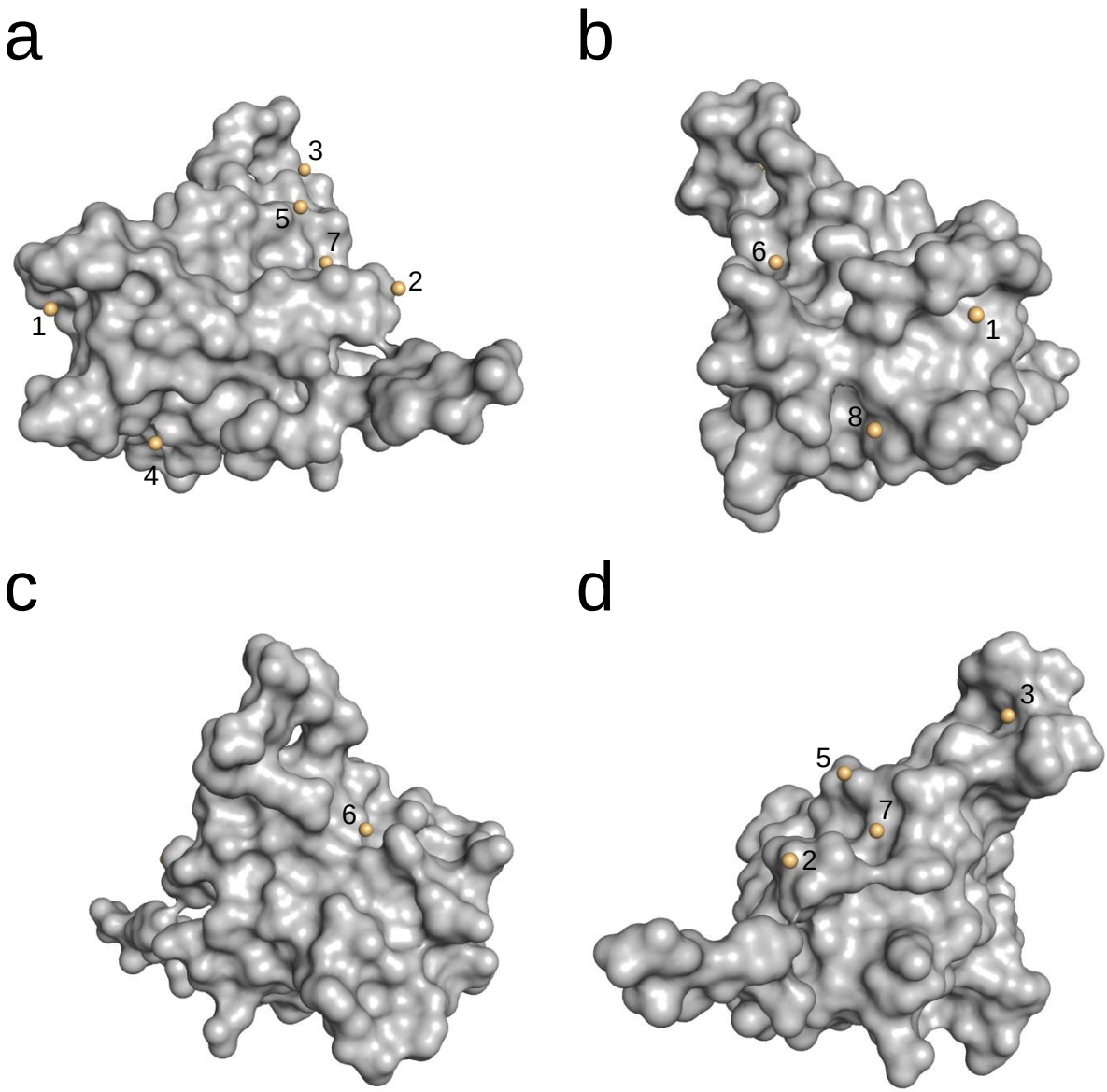


Figure 3-51. Overview of all observed cesium sites in p62_1-109. P62_1-109 is represented as grey surface and shown in four orientations rotated by 90° relative to each other. Cesium ions are represented as spheres and the sites are numbered. **(a)** Front view. **(b)** Side view left. **(c)** Back view. **(d)** Side view right.

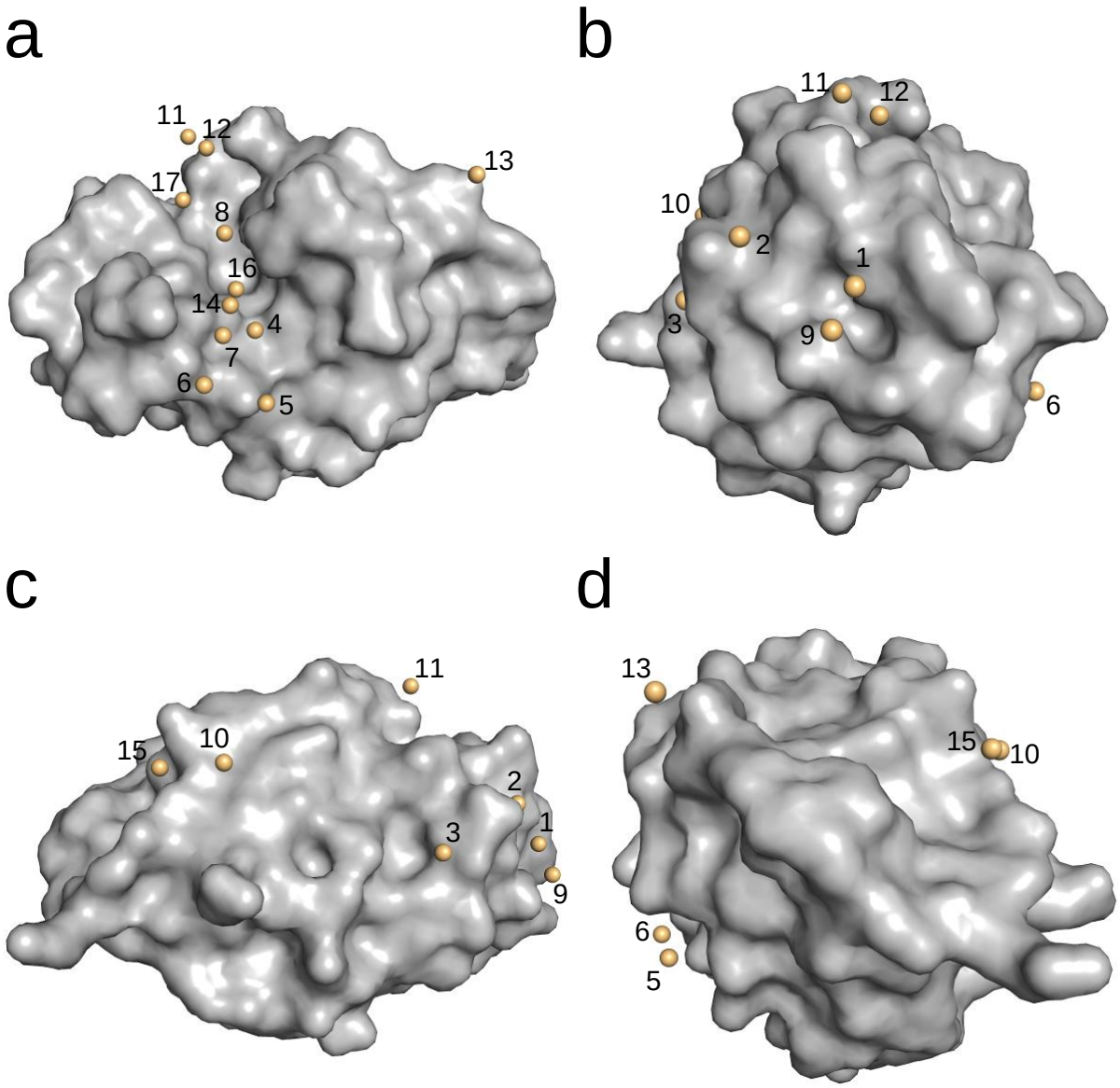


Figure 3-52. Overview of all observed cesium sites in HEWL. HEWL is represented as grey surface and shown in four orientations rotated by 90° relative to each other. Cesium ions are represented as spheres and the sites are numbered. **(a)** Front view. **(b)** Side view left. **(c)** Back view. **(d)** Side view right.

Table 3-11. Occurrence and occupancy of the cesium sites in p62_1-109 after different supplementation of CsCl.

	CsCl supplement			Cesium site[1]										
#	SEC[2]	Crys[3]	Cryo[4]	1[5]	2	3	4	5	6	7	8	Sites[6]	Σ Occ[7]	Occ / site[8]
1	0.25						0.25					1	0.25	0.25
2	0.25		0.25	0.72			0.36					2	1.08	0.54
3			0.75	0.9			0.41		0.27	0.42	0.27	5	2.27	0.45
4	0.25		0.75	0.68	0.67	0.36	0.43	0.38	0.31	0.27	0.33	8	3.43	0.43
5		0.7	0.7	0.54	0.47	0.27	0.37	0.36	0.24	0.25	0.26	8	2.76	0.35

1) Occupancy of observed sites is given. Numbers correspond to sites in Figure 3-51.
2) Supplement of CsCl to SEC buffer in mol/l.
3) Supplement of CsCl to crystallization condition in mol/l.
4) Supplement of CsCl to cryo-protectant in mol/l.
5) Site 1 lies on a crystallographic two-fold axis (Figure 3-55). Doubled occupancy is given.
6) Total number of observed sites.
7) Sum of occupancies of all observed sites.
8) Average occupancy per site.

Table 3-12. Occurrence and occupancy of the cesium sites in HEWL after different supplementation of CsCl.

	CsCl supplement			Cesium site[1]																		
#	Diss[2]	Crys[3]	Cryo[4]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Sites[5]	Σ Occ[6]	Occ / site[7]
1	0.25																				0	0	
2	0.25		0.25	0.44	0.22	0.22	0.21	0.37	0.23	0.19											7	1.88	0.27
3			1.71	0.37	0.81	0.45	0.43	0.89	0.35	0.38	0.27	0.28	0.21	0.71	0.44	0.32	0.4	0.34			15	6.65	0.44
4	0.25		1.71	0.48	0.85	0.5	0.39	0.91	0.46	0.37	0.32	0.3	0.27	0.63	0.38	0.3	0.34	0.3	0.38		16	7.18	0.45
5		1.5	1.5	0.32	0.84	0.29	0.4	0.77	0.21	0.35	0.28	0.29	0.36	0.54	0.45	0.33	0.32	0.38		0.56	16	6.69	0.42

1) Occupancy of observed sites is given. Numbers correspond to sites in Figure 3-52.
2) CsCl concentration in mol/l in which HEWL was dissolved.
3) Supplement of CsCl to crystallization condition in mol/l.
4) Supplement of CsCl to cryo-protectant in mol/l.
5) Total number of observed sites.
6) Sum of occupancies of all observed sites.
7) Average occupancy per site.

The number and occupancies of the individual cesium sites after the different supplementation of CsCl at the stages considered (purification, crystallization, cryo-protection) are summarized in Table 3-11 (p62_1-109) and Table 3-12 (HEWL). Additionally, the overall cesium occupancy and average occupancy per site are given. The cesium substructure and anomalous density of p62_1-109 after different supplementation of CsCl is depicted in Figure 3-53. When 0.25 M CsCl was only used during the SEC (dataset # 1), very low anomalous density could be observed for cesium site 4 (Figure 3-53 a; Figure 3-54 a). Directly compared with p62_1-109 treated only with KCl (dataset # 6), the anomalous peak seems to be higher for # 1 (Figure 3-54), indicating the presence of a cesium ion at this site. However, the resolution of dataset # 6 is lower compared to dataset # 1 (Table 3-8). This might have an influence on the peak height as well, making the placement of the cesium ion in # 1 not unambiguous.

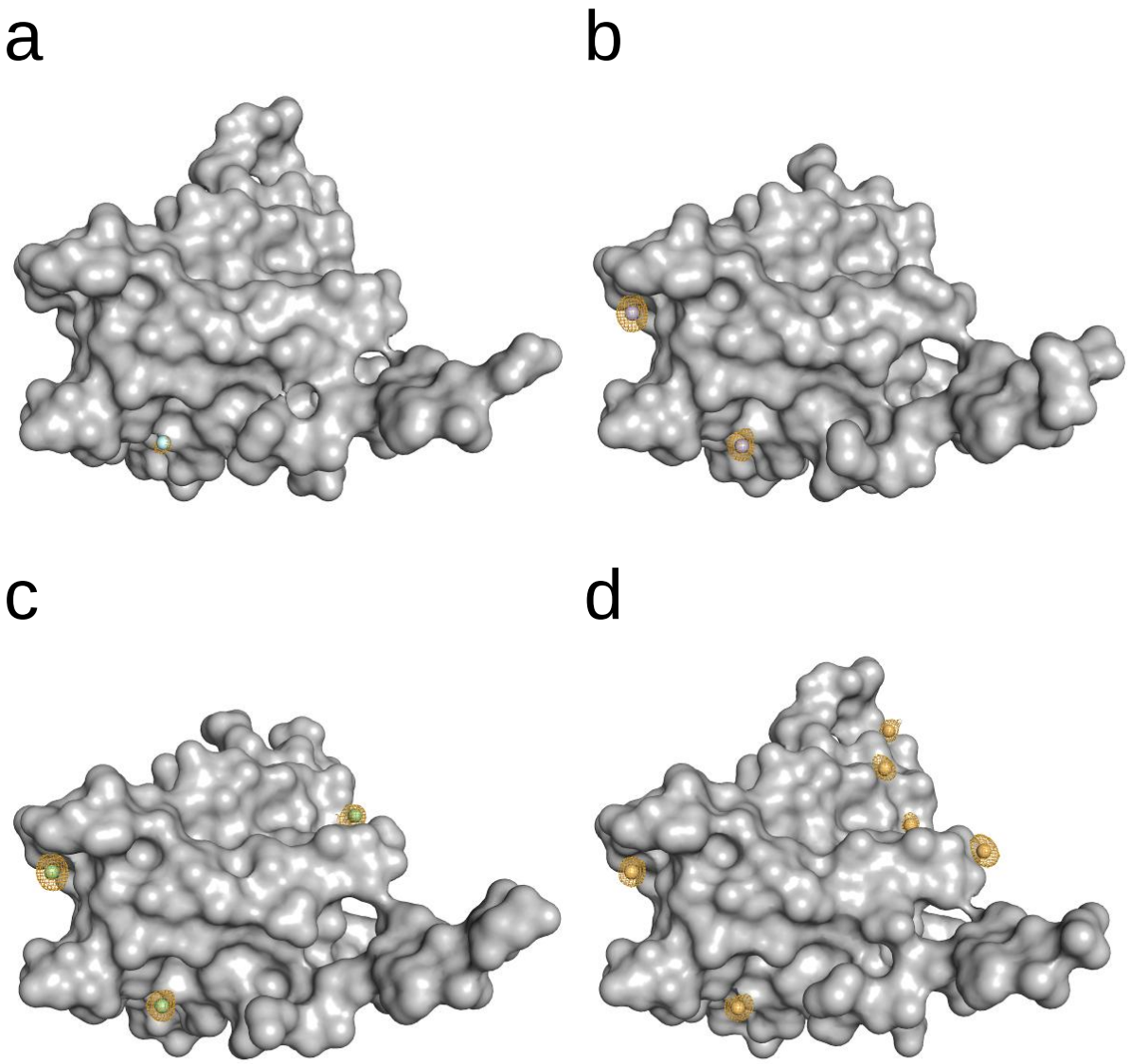


Figure 3-53. Cesium substructure of p62_1-109 after different supplementation of CsCl. P62_1-109 is represented as grey surface, the cesium ions are represented as spheres. Orange meshes display the anomalous density contoured at 3 σ. **(a)** Protein purified in CsCl buffer. **(b)** Protein purified in CsCl buffer and supplemented with 0.25 M CsCl to the cryo-protectant solution. **(c)** Protein purified in NaCl buffer and supplemented with 0.75 M CsCl to the cryo-protectant solution. **(d)** Protein purified in CsCl buffer and supplemented with 0.75 M CsCl to the cryo-protectant solution.

a b

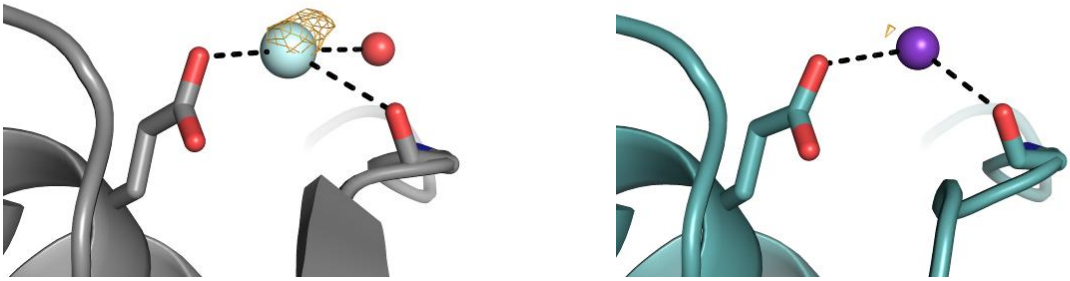


Figure 3-54. Comparison of cesium site 4 in p62_1-109 treated with 0.25 M CsCl during SEC with the corresponding potassium site. Orange meshes display anomalous density contoured at 3 σ. **(a)** Cesium site 4 from Figure 3-53 a. **(b)** Potassium site 4. The protein has been supplemented with 0.9 M KCl in the crystallization solution and 0.9 M KCl in the cryo-protectant solution.

a b

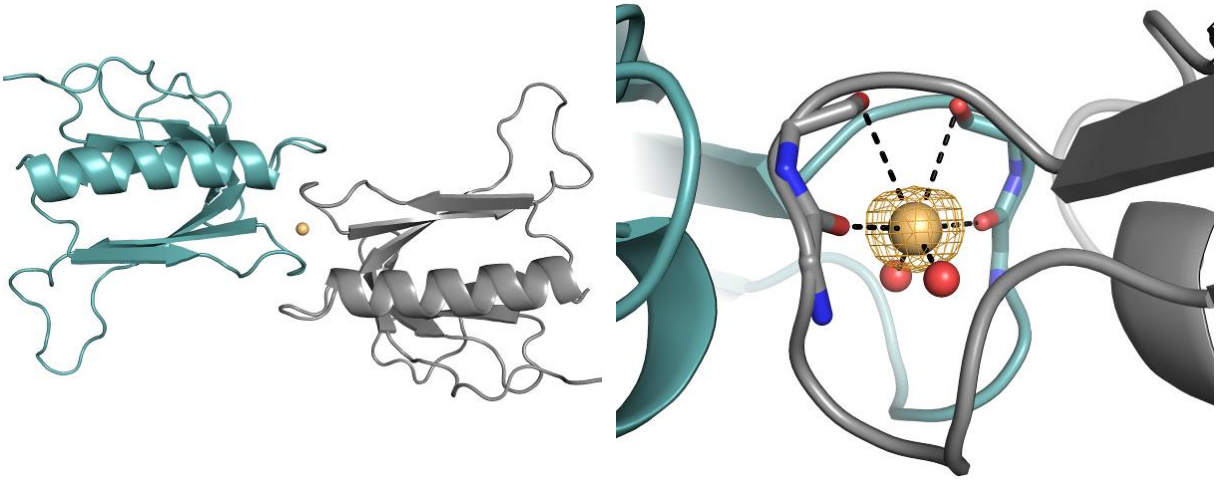


Figure 3-55. Cesium site 1 in p62_1-109 occupies a special position. (a) Cartoon representation of two p62_1-109 molecules related by a crystallographic two-fold axis perpendicular to the paper plane. The cesium ion located on this axis is represented as a sphere. **(b)** Detailed view of cesium site 1. The orange mesh displays the anomalous density contoured at 3 σ.

P62_1-109 crystallized with two slightly different unit cell sizes. On the one hand with a at ~34 Å and b at ~107 Å (datasets # 1, 4, 5, and 6), on the other hand with a at ~40 Å and b at ~102 Å (datasets # 2 and 3) (Table 3-8). The crystals appeared to be very similar and in both cases they grew in all conditions tested, irrespective of the CsCl concentration. The difference in unit cell size goes along with a disordered loop region in datasets # 2 and 3 (Figure 3-56 a). As this loop region coordinates cesium site 3 (Figure 3-56 b), the missing

cesium site 3 in datasets # 2 and 3 (Table 3-11) is most likely due to the slight differences in crystal packing.

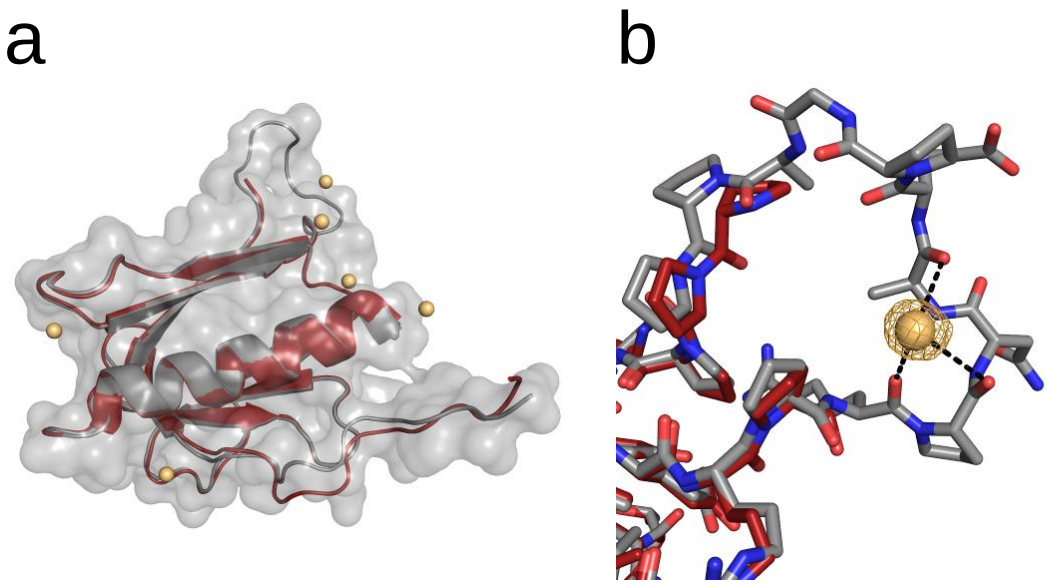


Figure 3-56. Superposition of p62_1-109 models from crystals with different unit cell sizes. (a) Surface and cartoon in grey correspond to dataset # 4. The cesium sites are displayed as spheres. The red cartoon corresponds to dataset # 2. The loop region at the top is not present in the red model, due to disorder. **(b)** Detailed view of the loop region with cesium site 3. The orange mesh displays the anomalous density contoured at 3 σ.

The cesium substructure and anomalous density of HEWL after different CsCl supplementation is depicted in Figure 3-57. The substructure for the protein supplemented with 0.25 M CsCl only during dissolving of the protein powder (dataset # 1) is not shown, as no bound cesium ions could be observed (Table 3-12).

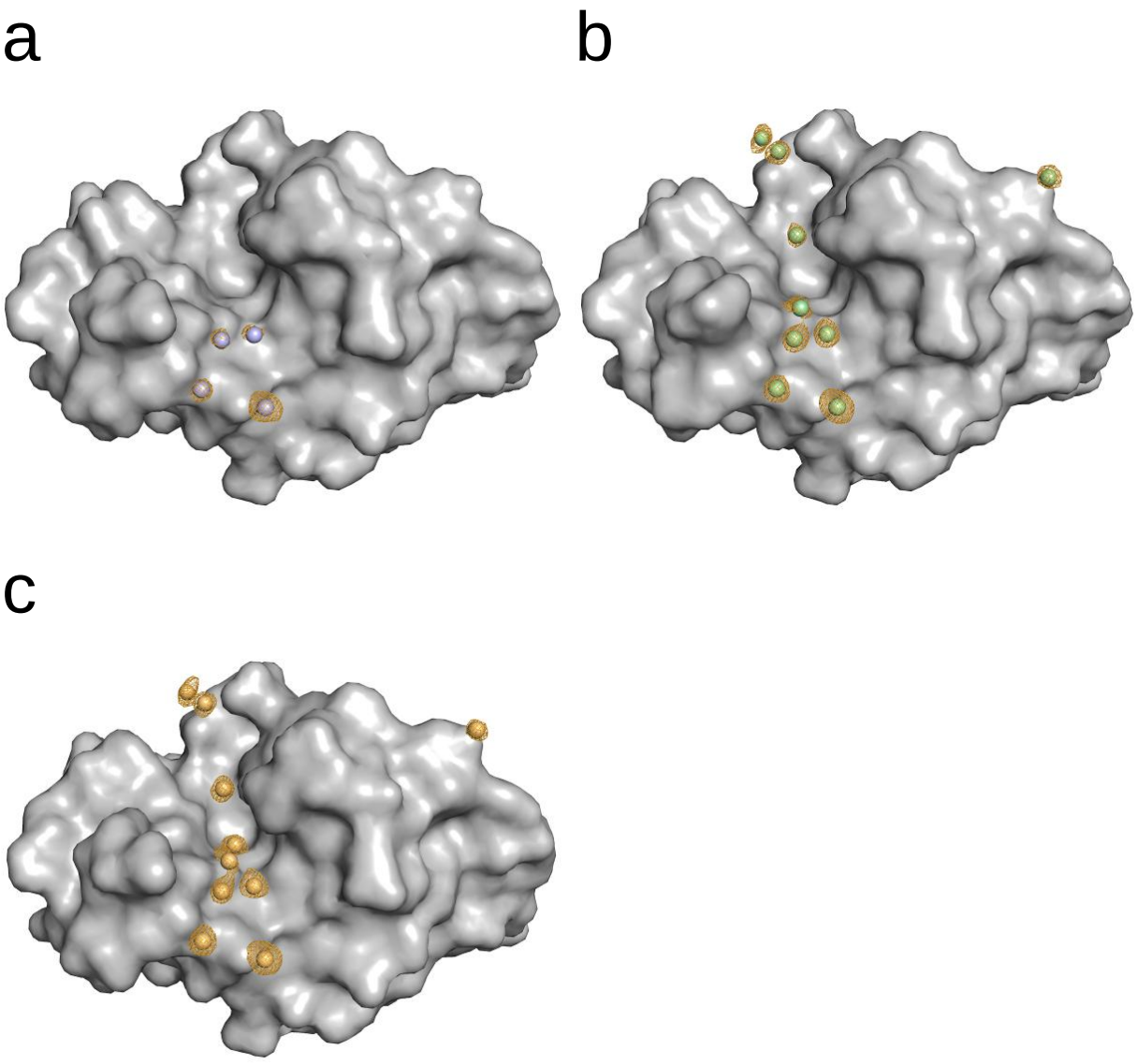


Figure 3-57. Cesium substructure of HEWL after different supplementation of CsCl. HEWL is represented as grey surface, the cesium ions are represented as spheres. Orange meshes display the anomalous density contoured at 3 σ. **(a)** Protein dissolved in 0.25 M CsCl and supplemented with 0.25 M CsCl to the cryo-protectant solution. **(b)** Protein dissolved in water and supplemented with 1.71 M CsCl to the cryo-protectant solution. **(c)** Protein dissolved in 0.25 M CsCl and supplemented with 1.71 M CsCl to the cryo-protectant solution.

3.10.4. Influence of cesium on the overall structure

All protein structures from the different approaches of CsCl supplementation were superimposed to each other, and the calculated RMSD values are given in Table 3-13 (p62_1-109) and Table 3-14 (HEWL).

As seen from the RMSD values, the supplementation with CsCl had hardly any influence on the overall protein structure, as all models of all approaches are very similar. However, the presence of a cesium ion at site 2 had a slight influence on the local protein conformation of HEWL, as it is apparent from the flip of the backbone carbonyl seen in Figure 3-59 (compare c and d).

Table 3-13. RMSD values between the p62_1-109 protein models from all datasets.

#	2	3	4	5	6
1	0.823 (819)	0.989 (850)	0.590 (879)	0.257 (910)	0.609 (867)
2	x	0.679 (839)	0.925 (819)	0.831 (814)	0.899 (807)
3	-	x	1.099 (850)	0.975 (845)	1.107 (838)
4	-	-	x	0.597 (874)	0.525 (894)
5	-	-	-	x	0.637 (862)

RMSD values in Å, calculated by LSQ Superpose via Coot, including all atoms.
Number of aligned atoms is given in parentheses.

Table 3-14. RMSD values between the HEWL protein models from all datasets.

#	2	3	4	5	6	7	8
1	0.475 (990)	0.635 (973)	0.494 (961)	0.786 (961)	0.646 (988)	0.563 (973)	0.558 (962)
2	x	0.466 (983)	0.581 (995)	0.645 (971)	0.774 (998)	0.730 (992)	0.717 (981)
3	-	x	0.400 (969)	0.632 (1011)	0.791 (996)	0.744 (975)	0.720 (979)
4	-	-	x	0.599 (958)	0.703 (984)	0.572 (963)	0.596 (976)
5	-	-	-	x	0.681 (984)	0.874 (963)	0.879 (967)
6	-	-	-	-	x	0.656 (990)	0.654 (994)
7	-	-	-	-	-	x	0.326 (1021)

RMSD values in Å, calculated by LSQ Superpose via Coot, including all atoms.
Number of aligned atoms is given in parentheses.

3.10.5. Substitution of potassium ions with cesium

The question was addressed if ion binding sites are interchangeable between potassium and cesium, or if cesium ions occupy different positions than potassium ions. Therefore, p62_1-109 (dataset # 6) and HEWL (dataset # 6 and 7) were treated with KCl in the absence of CsCl. Figure 3-58 depicts a comparison between the potassium and cesium substructures of p62_1-109. With a comparable KCl concentration only one potential potassium site could be observed where potassium occupies position 4; a detailed view is given in Figure 3-58 c, d.

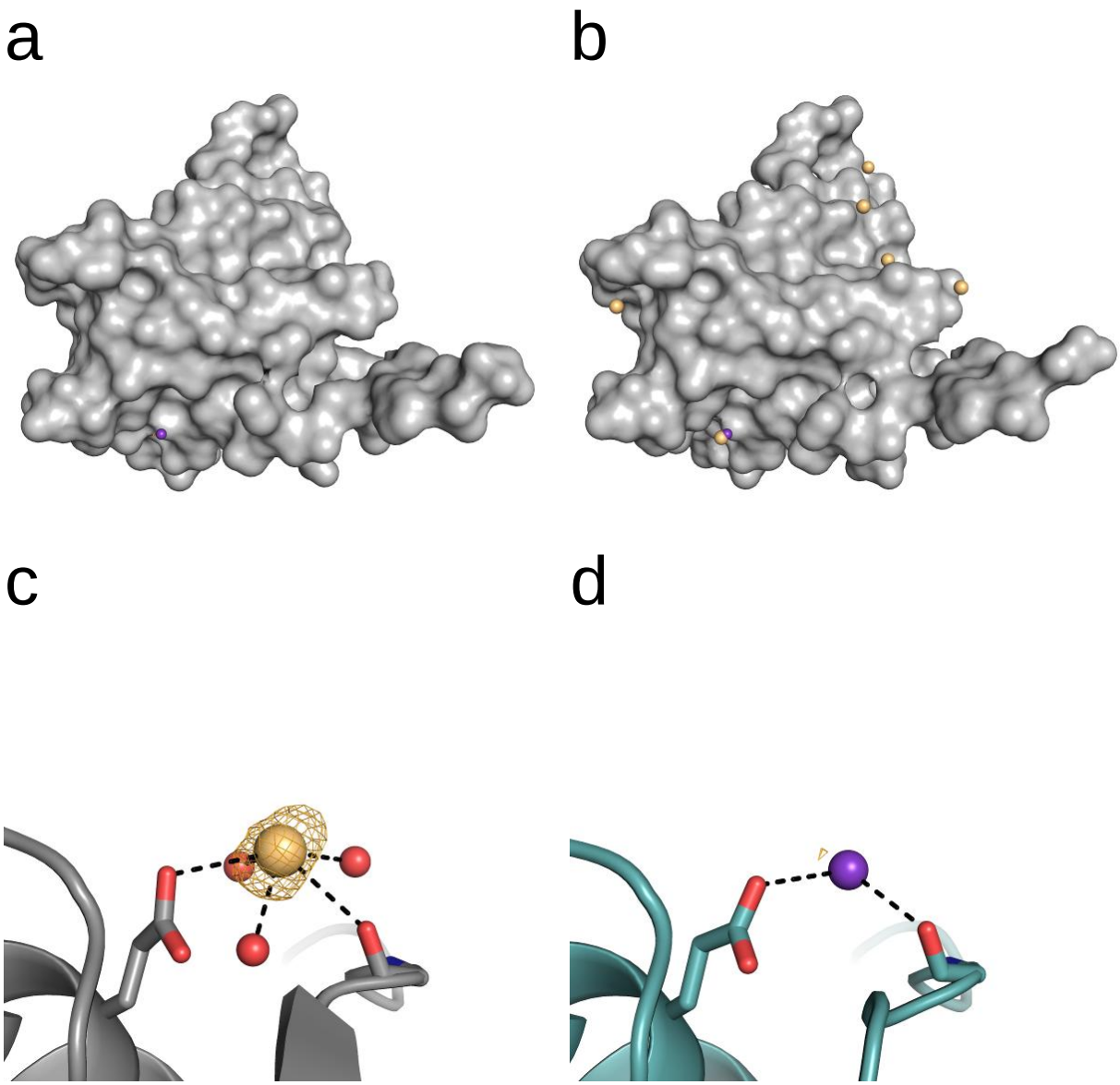


Figure 3-58. Substitution of potassium sites with cesium in p62_1-109. Orange meshes display the anomalous density contoured at 3 σ. **(a)** Potassium substructure of p62_1-109 supplemented with 0.9 M KCl in the crystallization solution and 0.9 M KCl in the cryo-protectant solution. **(b)** Overlay of the potassium substructure with the cesium substructure from p62_1-109 supplemented with 0.7 M CsCl in the crystallization solution and 0.7 M CsCl in the cryo-protectant solution. **(c)** Detailed view of cesium site 4. **(d)** Detailed view of potassium site 4.

The potassium and cesium substructures of HEWL are compared in Table 3-15 and depicted in Figure 3-60. Similar to p62_1-109, fewer potassium sites compared to cesium sites could be observed with similar alkali chloride concentrations. The majority of the sites were conserved between cesium and potassium, yet unique potassium sites were observed as well (Table 3-15).

Table 3-15. Comparison of the potassium and cesium substructure of HEWL.

		Conserved sites[1]																	Not conserved[2]		
#	K/Cs salt[3]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	X-1	X-2	X-3
5	1.5 M CsCl	0.32	0.84	0.29	0.4	0.77	0.21	0.35	0.28	0.29	0.36	0.54	0.45	0.33	0.32	0.38		0.56			
6	1.5 M KCl		0.53		0.55	0.46								0.48					0.84		0.49
7	3 M KCl	0.59	0.59		0.43	0.52	0.57							0.41	0.44		0.47		0.94	0.46	0.43

1) Occupancy of sites conserved between Cs and K. Numbers correspond to positions in Figure 3-52.
2) Occupancy of sites unique to potassium.
3) Concentration of alkali salt supplemented during crystallization and cryo-protection.

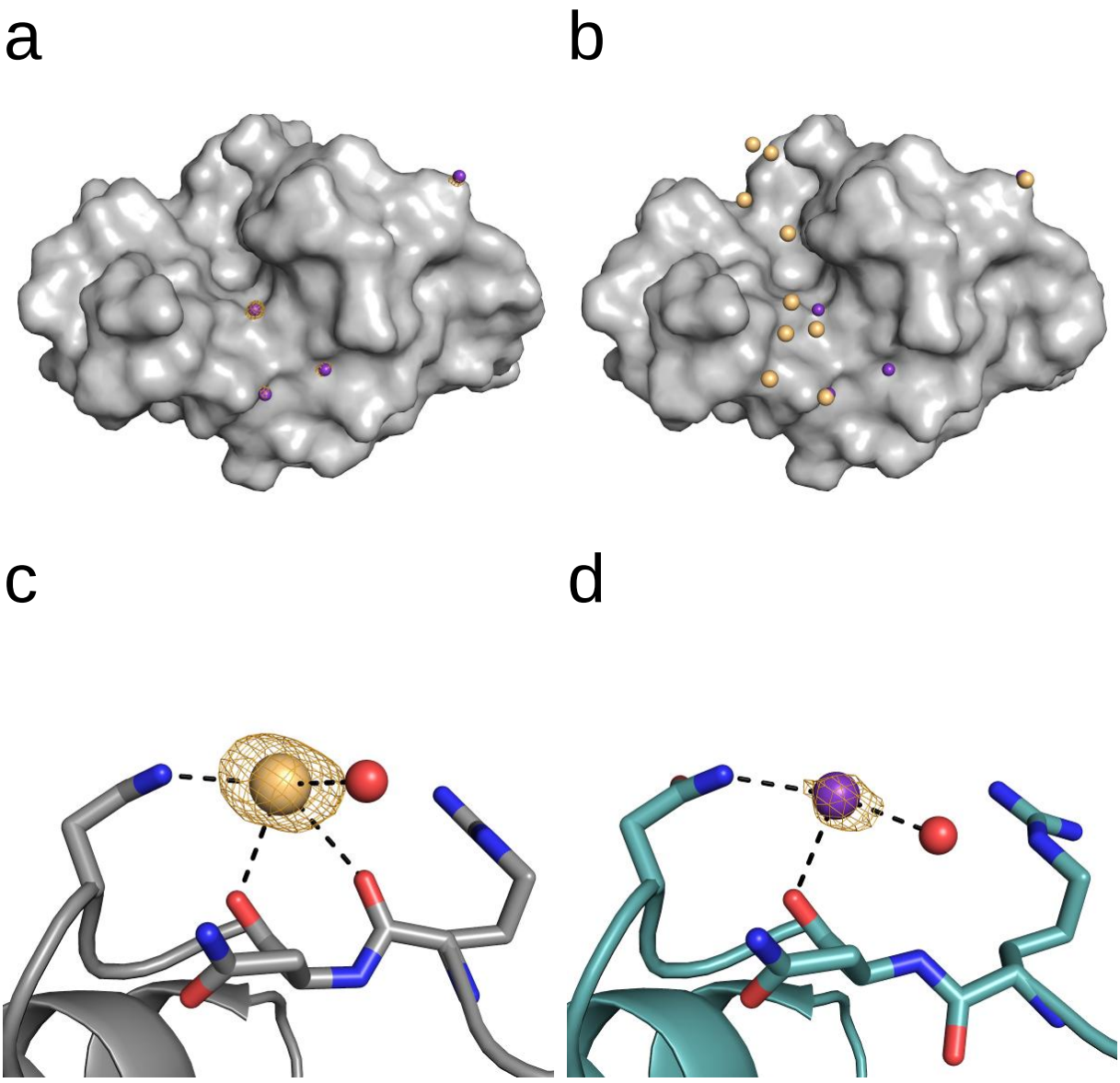


Figure 3-59. Substitution of potassium sites with cesium in HEWL. Orange meshes display the anomalous density contoured at 3 σ. **(a)** Potassium substructure of HEWL supplemented with 1.5 M KCl in the crystallization solution and 1.5 M KCl in the cryo-protectant solution. **(b)** Overlay of the potassium substructure with the cesium substructure from HEWL supplemented with 1.5 M CsCl in the crystallization solution and 1.5 M CsCl in the cryo-protectant solution. **(c)** Detailed view of cesium site 2. **(d)** Detailed view of potassium site 2.

When increasing the concentration of KCl, the number of potassium sites in HEWL increases as well, occupying more sites, which are also occupied by cesium (Table 3-15). Figure 3-60

depicts the potassium substructure of HEWL supplemented with 3 M KCl, compared to the cesium substructure at 1.5 M CsCl.

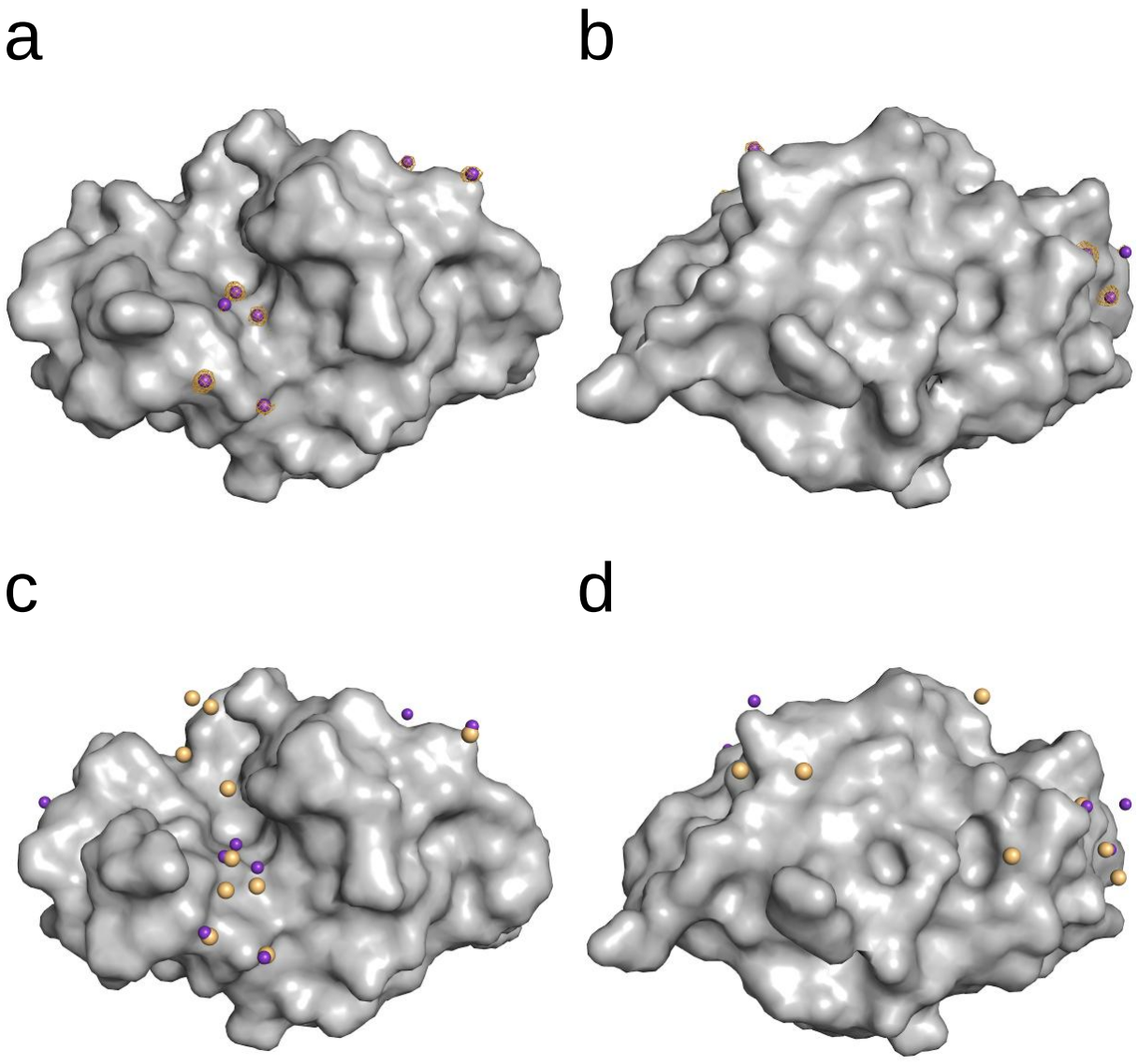


Figure 3-60. Comparison of the cesium substructure to the potassium substructure at high potassium concentrations in HEWL. Orange meshes display the anomalous density contoured at 3 σ. **(a)** Front view of the potassium substructure of HEWL supplemented with 3 M KCl in the crystallization solution. The cryo-protectant was present in the crystallization solution. **(b)** Back view of the potassium substructure of (a). **(c)** Overlay with the cesium substructure of HEWL supplemented with 1.5 M CsCl in the crystallization solution and 1.5 M CsCl in the cryo-protectant solution. **(d)** Back view of the overlay from (c).

In HEWL dataset # 8 one sodium ion could be positioned (Figure 3-61). When compared to HEWL supplemented with KCl or CsCl, the sodium ion could be substituted with potassium (Figure 3-61). Cesium however, was not able to substitute at this position for the sodium ion, most likely due to insufficient flexibility to accommodate for the significantly longer bonding distances of cesium.

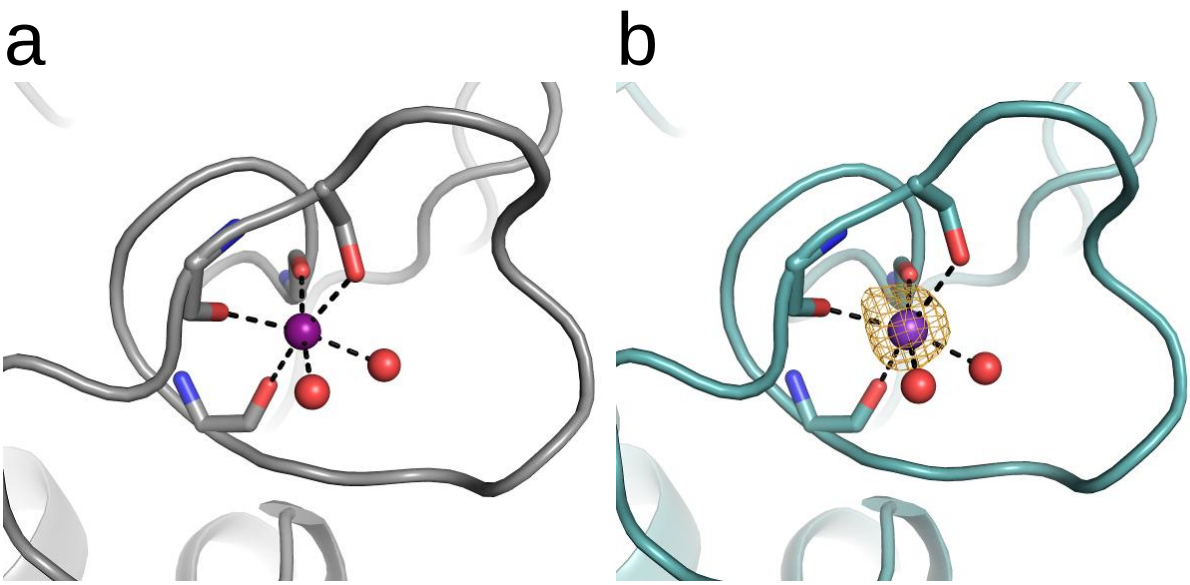


Figure 3-61. Substitution of sodium sites with potassium in HEWL. (a) Sodium site of HEWL supplemented with 1.71 M NaCl in the crystallization solution. The cryo-protectant was present in the crystallization solution. **(b)** Potassium site corresponding to the sodium site in (a). The protein was supplemented with 1.5 M KCl in the crystallization solution and 1.5 M KCl in the cryo-protectant solution. The orange mesh displays the anomalous density contoured at 3 σ.

3.11. Dual complexes

To elucidate the interaction network within TFIIH and to improve stability, purification behavior, and crystallization propensities, co-expression and co-purification of dual complexes of TFIIH subunits and constructs thereof was conducted. Additionally, analysis by native PAGE was pursued to further investigate and confirm interaction sites.

3.11.1. XPB and p52

3.11.1.1. XPB/p52

Full-length XPB and p52 were co-expressed in BL21 CodonPlus (DE3) RIL cells. However, after induction only overexpression of p52 could be observed (Figure 3-62 a). Accordingly, only p52 could be obtained from the IMAC (Figure 3-62 b).

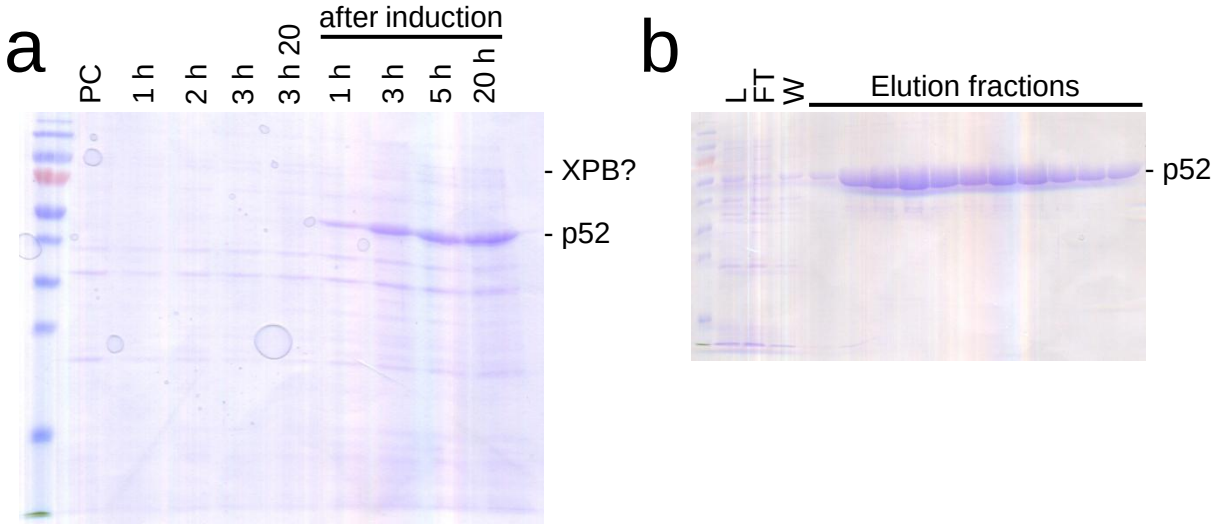


Figure 3-62. Co-expression and purification of XPB and p52 from BL21 CodonPlus (DE3) RIL cells. (a) The amount of XPB and p52 during expression before and after induction was monitored over time via SDS-PAGE. **(b)** SDS-PAGE analysis of IMAC samples. L: lysate; FT: flow through; W: wash.

As co-expression of XPB and p52 in BL21 CodonPlus (DE3) RIL cells failed, co-expression and co-purification in ArcticExpress (DE3) RIL cells was attempted. Purification of XPB/p52 with the standard buffer system is depicted in Figure 3-63. XPB and p52 eluted clearly separately during the SEC. XPB aggregated and was most likely bound by a chaperonin, comparable to the purification of the single protein from ArcticExpress (DE3) RIL cells (see 3.2.2, Figure 3-3).

Overexpression of XPB in the absence of p52 could be observed in BL21 CodonPlus (DE3) RIL cells before, with significant amounts of XPB 3 h after induction (see 3.2.2, Figure 3-2). As expression of p52 might interfere with expression of XPB, p52 expression was induced with a delay of 3 h. Yet, this approach also yielded only p52 expression; the result is depicted in Figure 3-64.

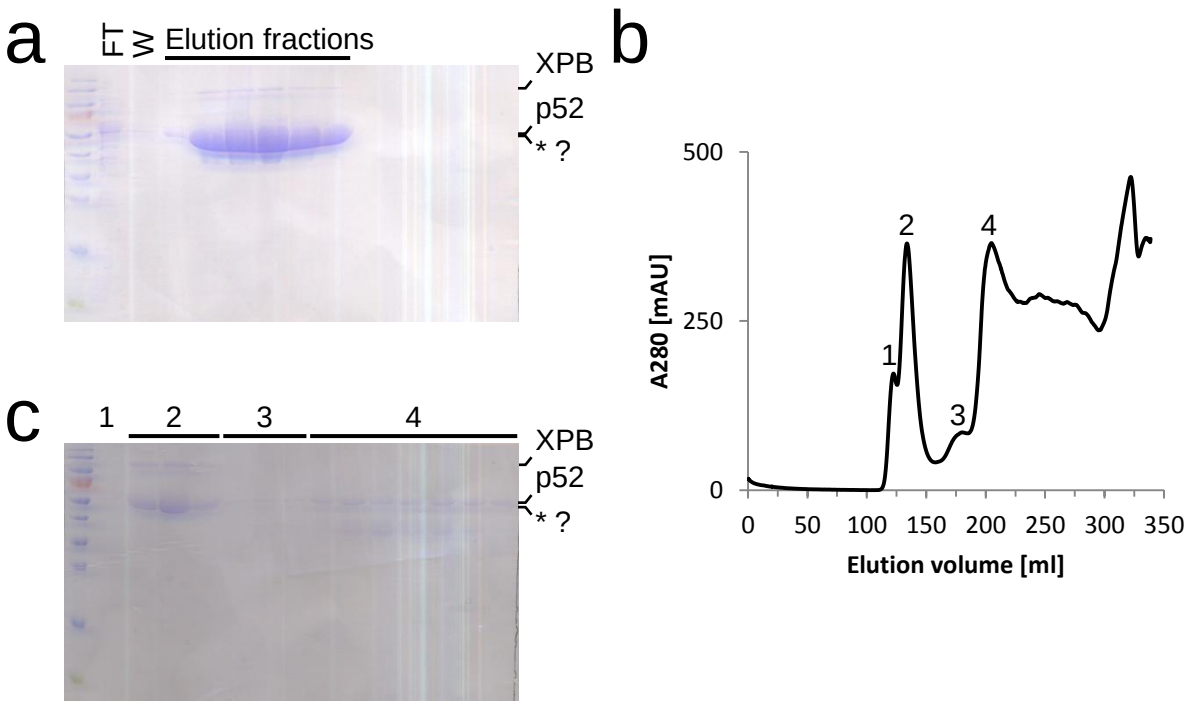


Figure 3-63. Co-purification of XPB and p52 from ArcticExpress (DE3) RIL cells. (a) SDS-PAGE analysis of IMAC samples. FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 26/60 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b). *: chaperonin.

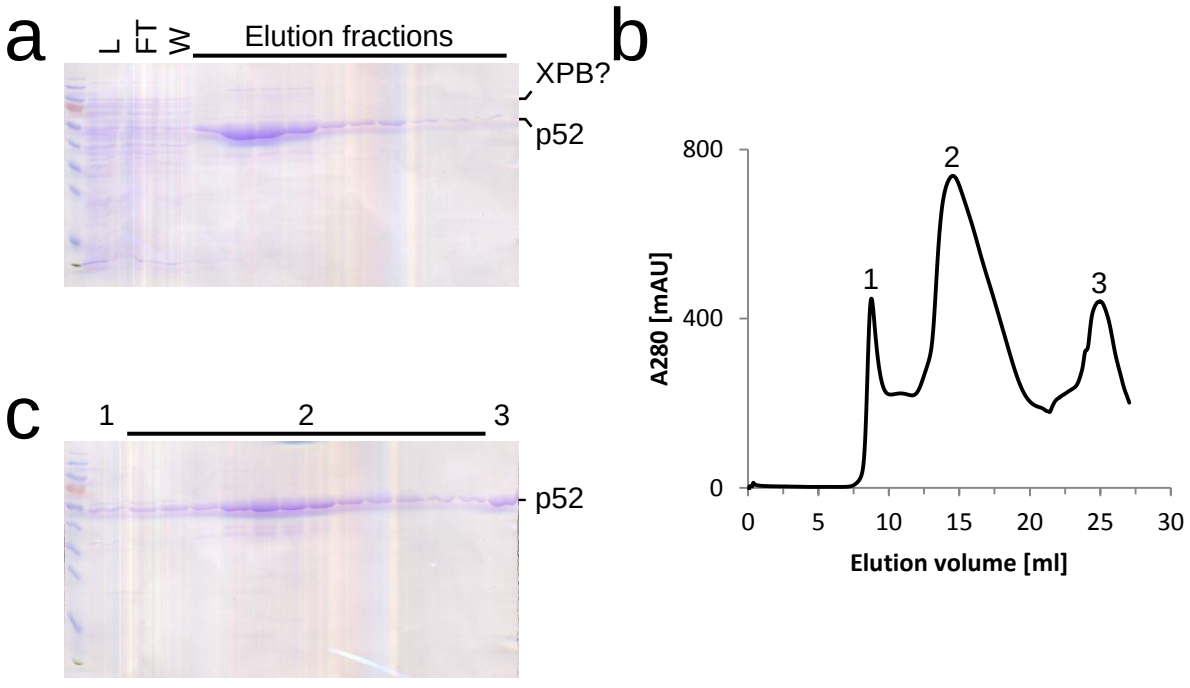


Figure 3-64. Co-purification of XPB and p52 from BL21 CodonPlus (DE3) RIL cells with delayed p52 induction. Expression of p52 was induced 3 h after induction of XPB. **(a)** SDS-PAGE analysis of IMAC samples. L: lysate; FT: flow through; W: wash. **(b)** SEC elution profile from a Superdex 200 10/300 GL column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

In order to address the strong bias on p52 expression, co-expression and co-purification of non-tagged (nT) p52 with a gene optimized XPB was attempted. Co-expression was performed in SoluBL21 cells, and the result from the co-purification using the standard buffer system is shown in Figure 3-65. Similar to previous attempts, only p52 could be obtained. Due to the fact that p52 was non-tagged the yields were much lower compared to purification of p52 by itself, as most likely only unspecifically bound p52 was retained during the IMAC.

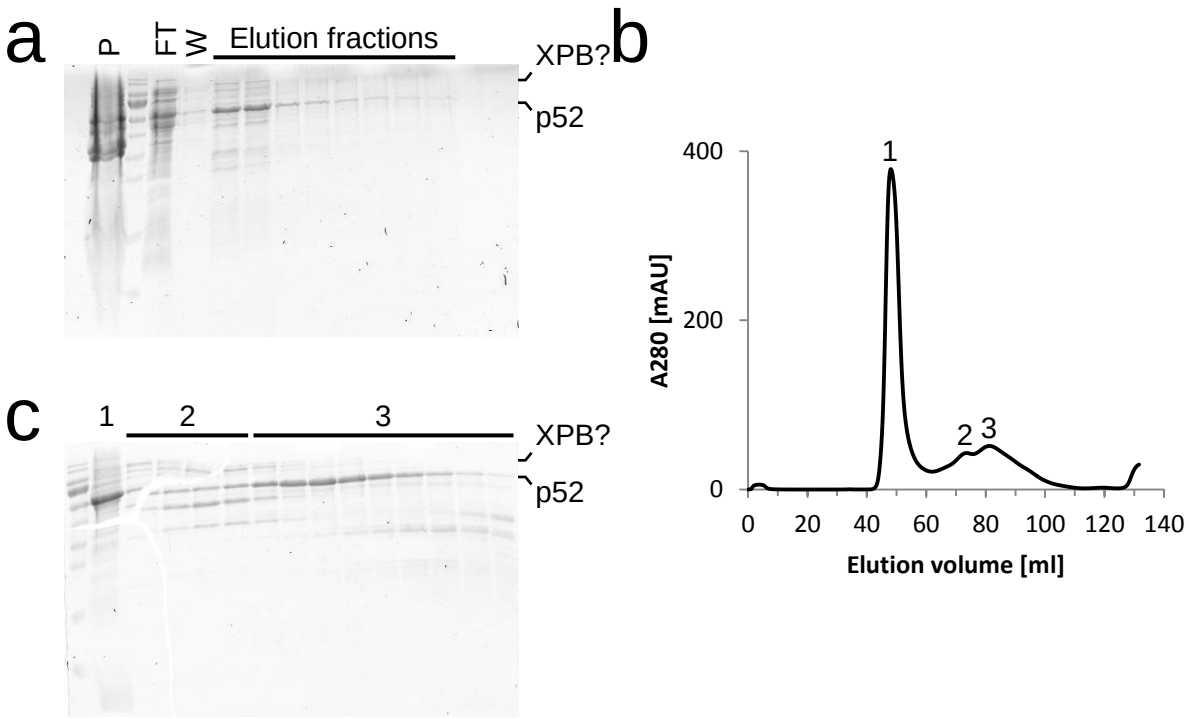


Figure 3-65. Co-purification of gene optimized XPB and p52nT from SoluBL21 cells. (a) SDS-PAGE analysis of IMAC samples. P: pellet; FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

3.11.1.2. XPB_1-345/p52

As the N-terminal extension of XPB interacts with p52 [25], co-expression of XPB_1-345, encompassing the N-terminal extension and the DRD (Figure 3-1) with non-tagged p52 (p52nT) was performed. Purification of XPB_1-345/p52nT from BL21 CodonPlus (DE3) RIL cells with the standard buffer system is shown in Figure 3-66. Based on the SDS-PAGE analysis, only one protein was expressed (Figure 3-66 a, c). However, XPB_1-345 and p52nT run on the same height in an SDS-PAGE (Figure 3-67 b). Furthermore, the peak from the elution volume of XPB_1-345/p52nT was shifted to a lower elution volume compared to the single proteins (Figure 3-67 a). Thus, it was speculated that co-purification of XPB_1-345 and p52nT was successful. Crystallization trials at 12.6 mg/ml with the commercial crystallization screens using the crystallization robot were performed, but no crystals could be obtained.

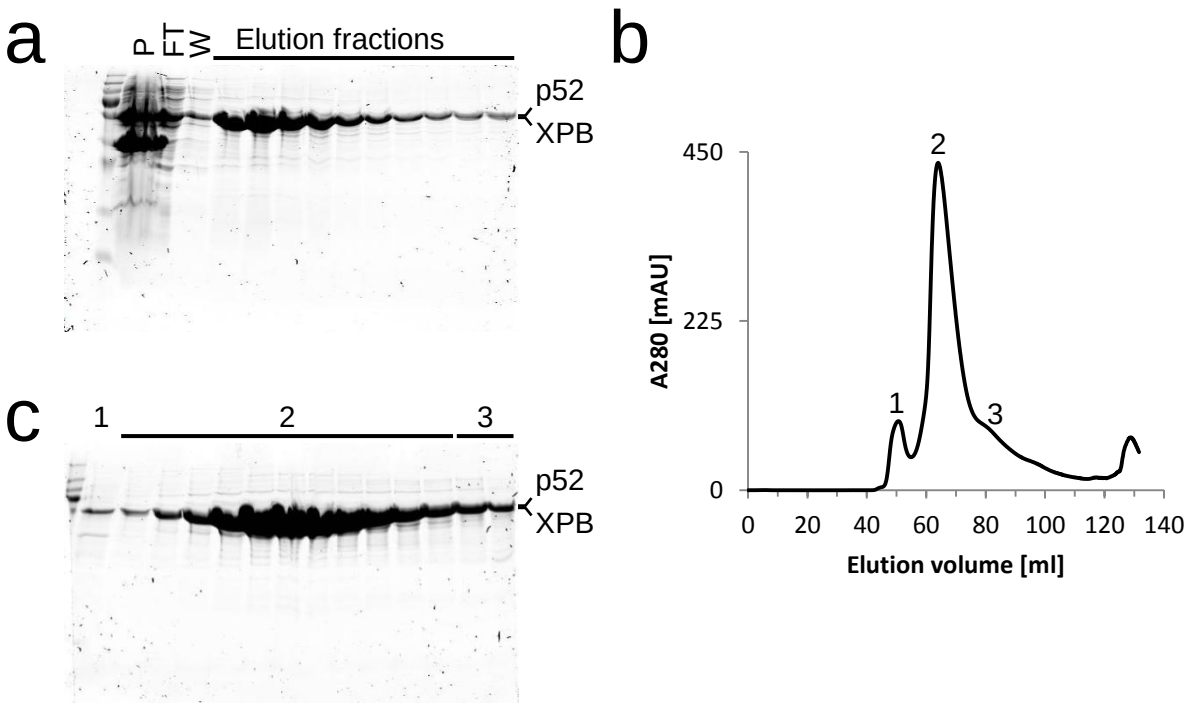


Figure 3-66. Co-purification of XPB_1-345 and p52nT. (a) SDS-PAGE analysis of IMAC samples. P: pellet; FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

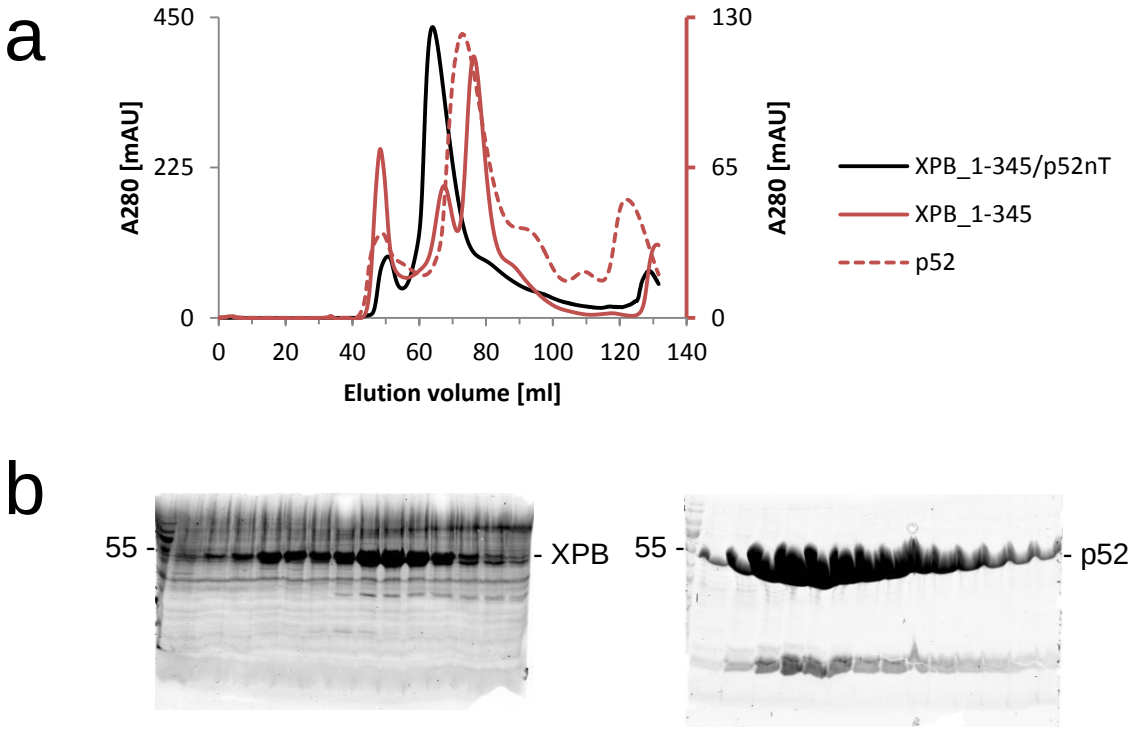


Figure 3-67. Evidence for successful co-purification of XPB_1-345 and p52nT. (a) Overlay of SEC elution profiles from co-purified XPB_1-345/p52nT with elution profiles from the single components. **(b)** Comparison of the SDS-PAGEs of XPB_1-345 and p52nT. The protein marker band for 55 kDa is indicated.

3.11.1.3. XPB_1-345/p52_1-349

To investigate the first XPB binding site of p52, XPB_1-345 was co-expressed and co-purified with the non-tagged construct p52_1-349, lacking the second XPB binding site (Figure 3-33). Purification of XPB_1-345/p52nT_1-349 from BL21 CodonPlus (DE3) RIL cells with the standard buffer system is depicted in Figure 3-68. No stoichiometric complex between XPB_1-345 and p52nT_1-349 could be obtained. The majority of p52nT_1-349 seemed to remain in the flow through and wash fraction (Figure 3-68 a), suggesting only weaker interactions.

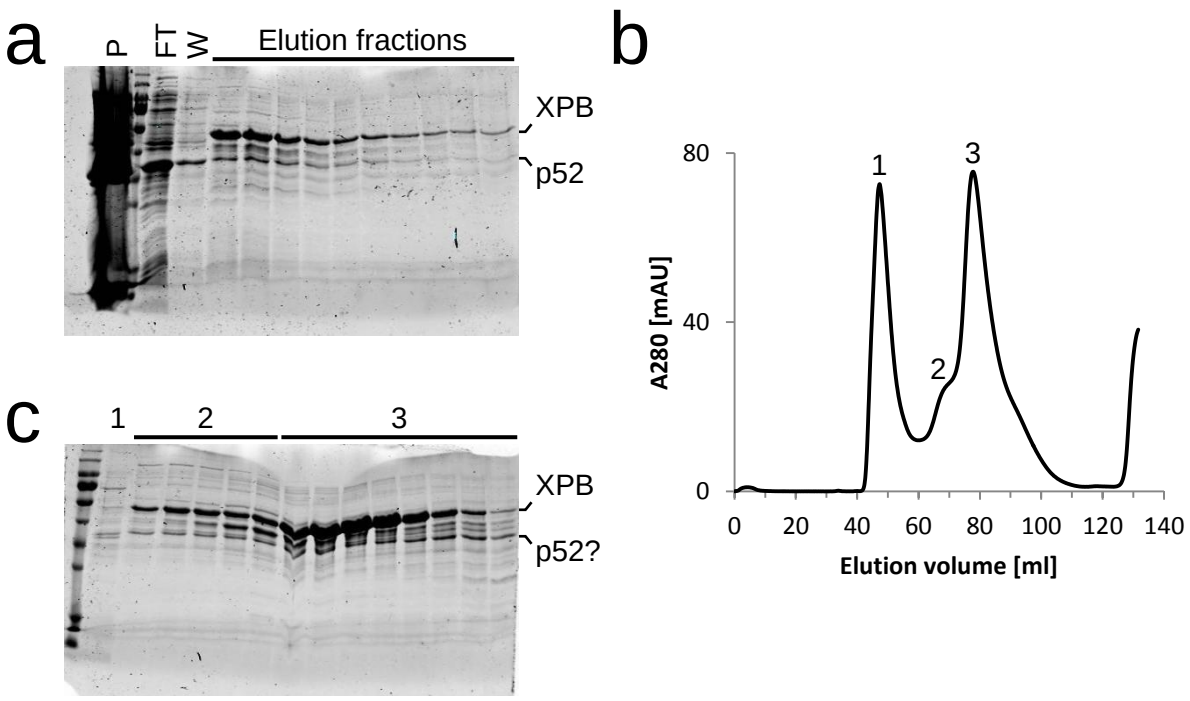


Figure 3-68. Co-purification of XPB_1-345 and p52nT_1-349. (a) SDS-PAGE analysis of IMAC samples. P: pellet; FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b)

3.11.1.4. XPB_1-345/p52_121-E

To investigate the second XPB binding site of p52, XPB_1-345 was co-expressed and co-purified with the non-tagged construct p52_121-E, lacking the first XPB binding site (Figure 3-33). Purification of XPB_1-345/p52nT_121-E from BL21 CodonPlus (DE3) RIL cells with the standard buffer system is depicted in Figure 3-69. Only high amounts of XPB_1-345 could be obtained, with a lower running band, which might correspond to p52nT_121-E. However, the presumed band for p52nT_121-E was highly substoichiometric.

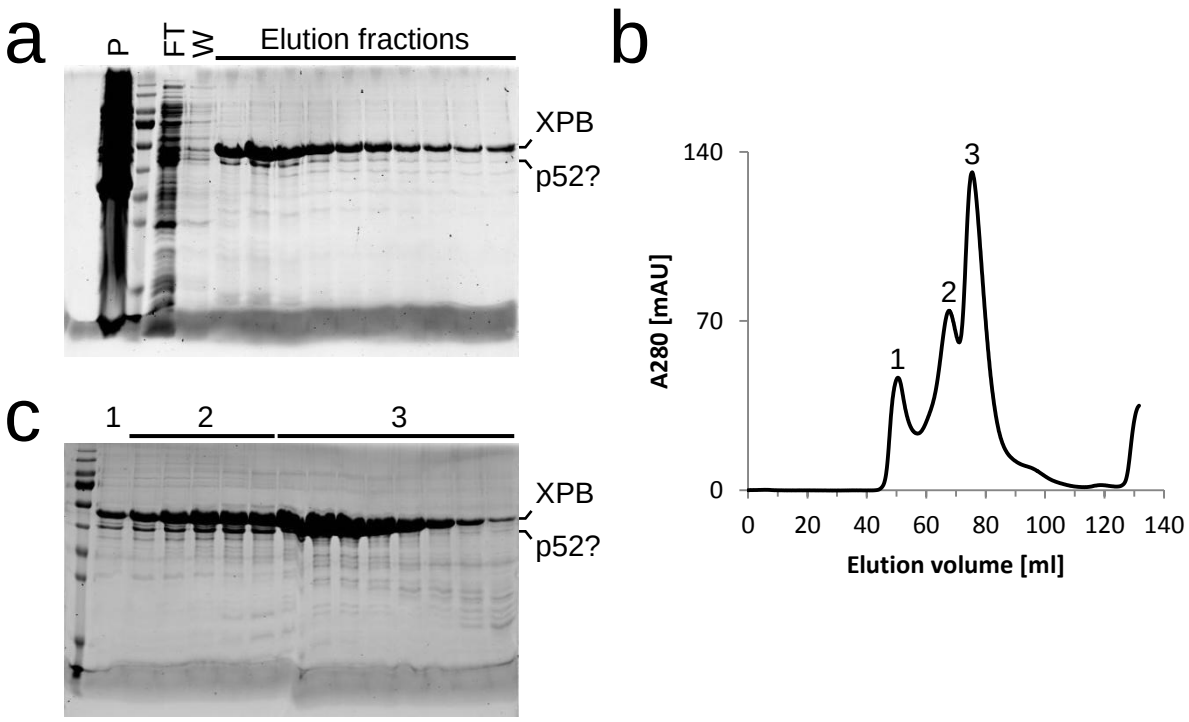


Figure 3-69. Co-purification of XPB_1-345 and p52nT_121-E. (a) SDS-PAGE analysis of IMAC samples. P: pellet; FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

3.11.1.5. XPB_1-345/p52_121-E_dL

As co-purification of XPB_1-345/p52nT_121-E failed (see 3.11.1.4), co-expression and co-purification with the modified construct p52_121-E_dL (Figure 3-33) was attempted. In this approach, both proteins contained a His-tag. The result of the IMAC is shown in Figure 3-70, and in contrast to co-purification with p52nT_121-E (Figure 3-69), only p52_121-E_dL could be obtained.

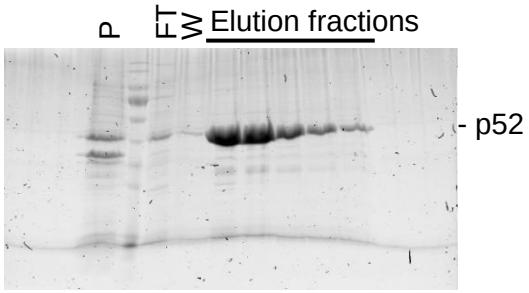


Figure 3-70. SDS-PAGE analysis of IMAC samples of XPB_1-345 co-purified with p52_121-E_dL from BL21 CodonPlus (DE3) RIL cells. P: pellet; FT: flow through; W: wash.

3.11.1.6. XPB_116-345/p52

To narrow down the p52 binding site of XPB and to obtain a complex suitable for crystallization, XPB_116-345 (Figure 3-1) was co-expressed and co-purified with non-tagged p52. Purification of XPB_116-345/p52nT from BL21 CodonPlus (DE3) RIL cells in standard buffer conditions is shown in Figure 3-71. Both proteins co-eluted in stoichiometric amounts in a single peak during the SEC. Importantly, construct XPB_116-345 was insoluble when purified on its own (Figure 3-15). Most likely, XPB_116-345 is kept in solution by p52nT, strongly indicative of complex formation. XPB_116-345/p52nT was subjected to crystallization trials at 7.8 mg/ml with the commercial crystallization screens using the crystallization robot, but no crystals could be obtained.

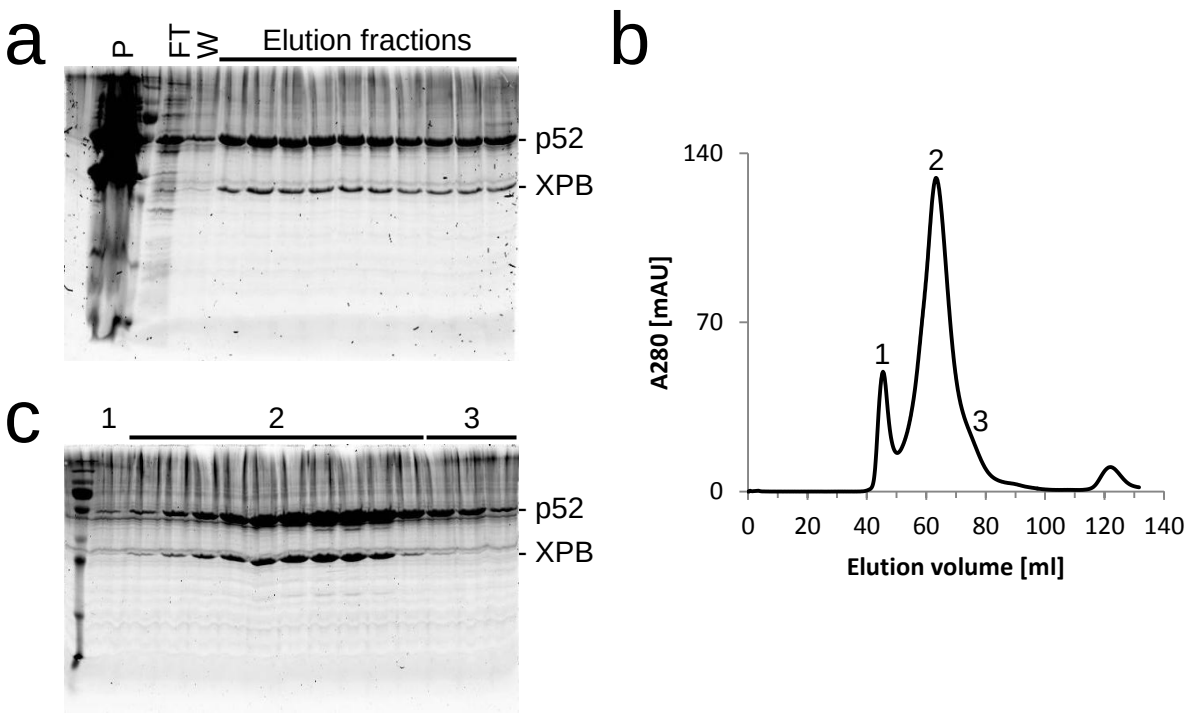


Figure 3-71. Co-purification of XPB_116-345 and p52nT. (a) SDS-PAGE analysis of IMAC samples. P: pellet; FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

3.11.1.7. XPB_116-245/p52

To further narrow down the p52 binding site of XPB, co-expression and co-purification of XPB_116-245 (Figure 3-1) with non-tagged p52 was conducted. XPB_116-245/p52nT was purified from BL21 CodonPlus (DE3) RIL cells using the standard buffer system, and the result is depicted in Figure 3-72. Both proteins co-eluted in stoichiometric amounts in a single peak during the SEC, indicative of complex formation. However, a significant band running closely underneath p52 appeared. It was unclear if it was a contamination or a degradation product of p52. The latter case would imply a weaker stabilization of p52 through

XPB_116-245 compared to XPB_116-345, as for the XPB_116-345/p52nT co-purification (see 3.11.1.6, Figure 3-71), no such band occurred. Only SEC elution fractions without this additional band were pooled and subjected to further studies.

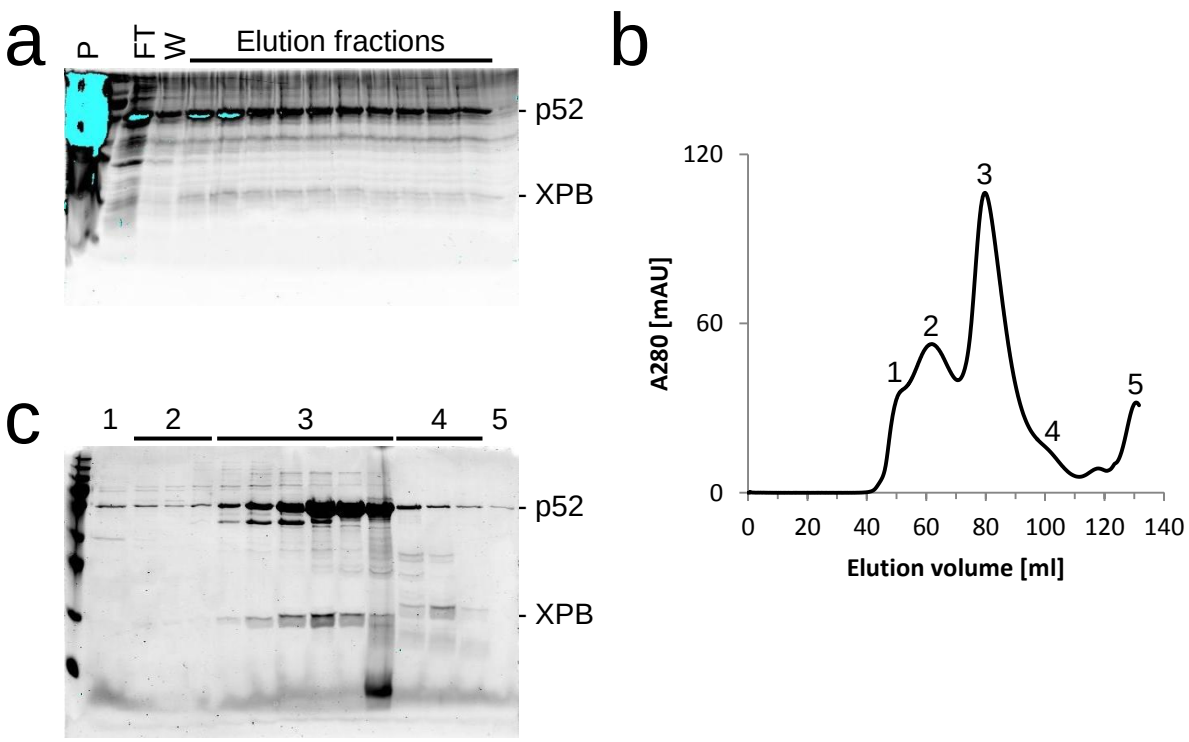


Figure 3-72. Co-purification of XPB_116-245 and p52nT. (a) SDS-PAGE analysis of IMAC samples. P: pellet; FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

XPB_116-245/p52nT was subjected to crystallization trials with the commercial crystallization screens using the crystallization robot. At a concentration of 2.9 mg/ml small needle like crystals could be obtained in a single tray of Wizard™ Classic 1 & 2 HT96 after a time period of about 5 months (Figure 3-73). Various fine screens in 96 well format using the crystallization robot and in the 24 well format were set up. However, the crystals could not be reproduced.

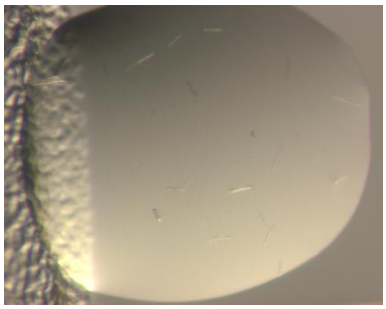


Figure 3-73. Crystallization of XPB_116-245/p52nT. Crystals appeared after 5 months. Reservoir solution: 0.1 M Tris-HCl pH 7.0; 0.2 M $MgCl_2$; 10 % (w/v) PEG 8000.

The crystals were harvested and subjected to synchrotron radiation at beamline ID29 at the ESRF. Protein crystal diffraction could be observed, yet the resolution was limited to about 8-9 Å (Figure 3-74).

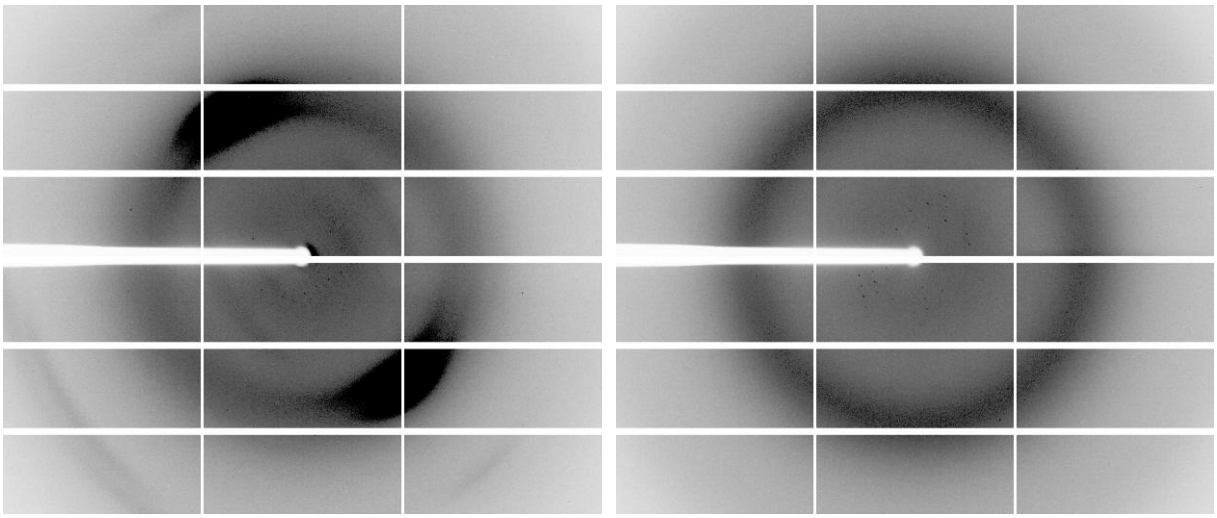


Figure 3-74. X-ray diffraction pattern of an XPB_116-245/p52nT crystal. The crystal was harvested from a crystallization condition consisting of 0.1 M Tris-HCl pH 7.0, 0.2 M $MgCl_2$, and 10 % (w/v) PEG 8000 in the reservoir. Two orientations of the crystal in the beam, 90° apart from each other, are shown. The outmost areas of the detector are omitted.

3.11.1.8. Native PAGE

To further assess the interaction between XPB_1-345 and the two XPB binding sites of p52, the interaction of XPB_1-345 with p52 and p52nT/p8, as well as p52_121-E, lacking the first XPB binding site, was tested via native PAGE (Figure 3-75). Complex formation between XPB_1-345 and p52, as well as p52nT/p8 could be observed. The appearance of a faint band for XPB_1-345 mixed with p52_121-E indicated that the second XPB binding site of p52 might be sufficient for the XPB-p52 interaction.

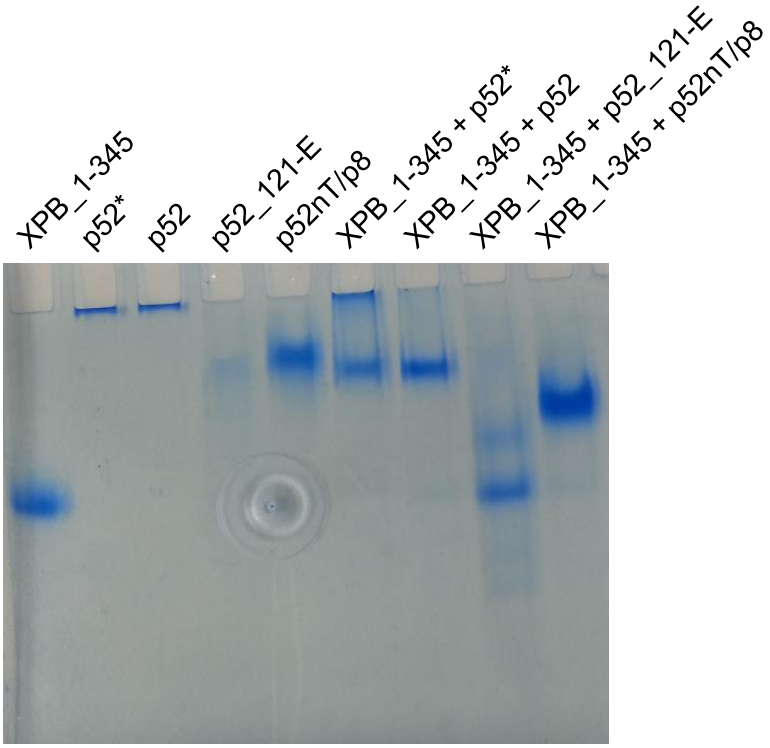


Figure 3-75. Interaction analysis between XPB_1-345 and p52, p52_121-E, or p52nT/p8 via native PAGE. P52*: protein batch stored over night at 4 °C.

3.11.2. p62 and p44

3.11.2.1. p62/p44

Full-length p62 and p44 were co-expressed in BL21 CodonPlus (DE3) RIL cells and initial co-purification with the standard buffer system is shown in Figure 3-76. Both proteins were present, yet it was unclear, if a complex could be obtained, as the peak from the SEC was asymmetric (Figure 3-76 b) and the bands in the SDS-PAGE analysis appeared to be not stoichiometric (Figure 3-76 c). Due to the high purity of the sample, crystallization trials at 9.9 mg/ml with the standard crystallization screens using the crystallization robot were performed yielding no crystals.

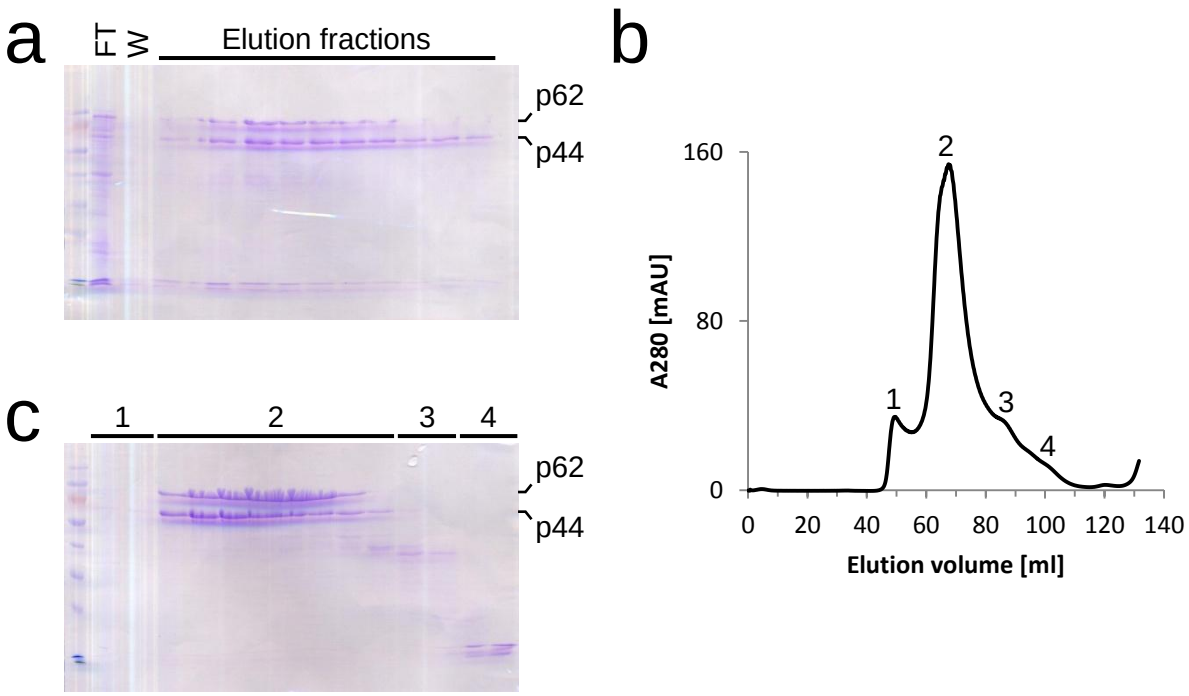


Figure 3-76. Co-purification of p62 and p44. (a) SDS-PAGE analysis of IMAC samples. FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

To test if a complex between p62 and p44 was formed, pooled peak fractions were compared to the single proteins via native agarose gel electrophoresis (Figure 3-77). 5 µl of each protein sample at 40 µM was mixed with 6x OrangeG loading dye in a 5:1 ratio and loaded on the gel. Gel electrophoresis was pursued at 70 V for 3 h at 4 °C in Tris/Glycine buffer. Compared to the single proteins, an additional band appeared for the co-purified proteins, indicative of complex formation (Figure 3-77). P44 was provided by Dominik Schmitt.

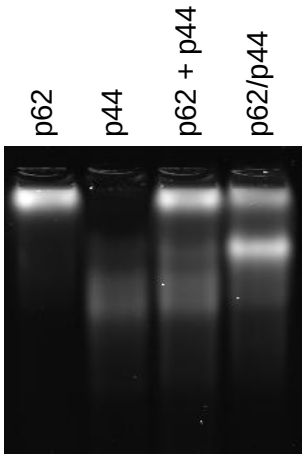


Figure 3-77. Native agarose gel electrophoresis of co-purified p62/p44 and the single components. P62 + p44: mixed single proteins; p62/p44: co-purified proteins.

The composition of the agarose gel is given below.

Native agarose gel

- 0.8 % (w/v) Agarose
- 12.5 mM Tris
- 96 mM Glycine

To prove complex formation and to obtain a stoichiometric p62/p44 complex, co-expression and co-purification with a His-tag only on one of the subunits was attempted. As p62 seemed to be more prominent during initial co-purification trials (Figure 3-76 c), non-tagged p62 (p62nT) was used. Co-purification of p62nT and p44 from BL21 CodonPlus (DE3) RIL cells using the standard buffer conditions is depicted in Figure 3-78. Stoichiometric amounts of p62nT and p44 during the IMAC could be obtained, with excessive p62nT remaining in the flow through and wash fraction (Figure 3-78 a). However, the SEC elution profile exhibited a step wise appearance, presumably representing different oligomeric states of p62nT/p44 (Figure 3-78 b). Fractions from the main peak eluting between 50 and 60 ml were pooled. This protein pool was subjected to crystallization trials at 3.4 mg/ml and 6.7 mg/ml with the commercial crystallization screens using the crystallization robot, but yielded no crystals.

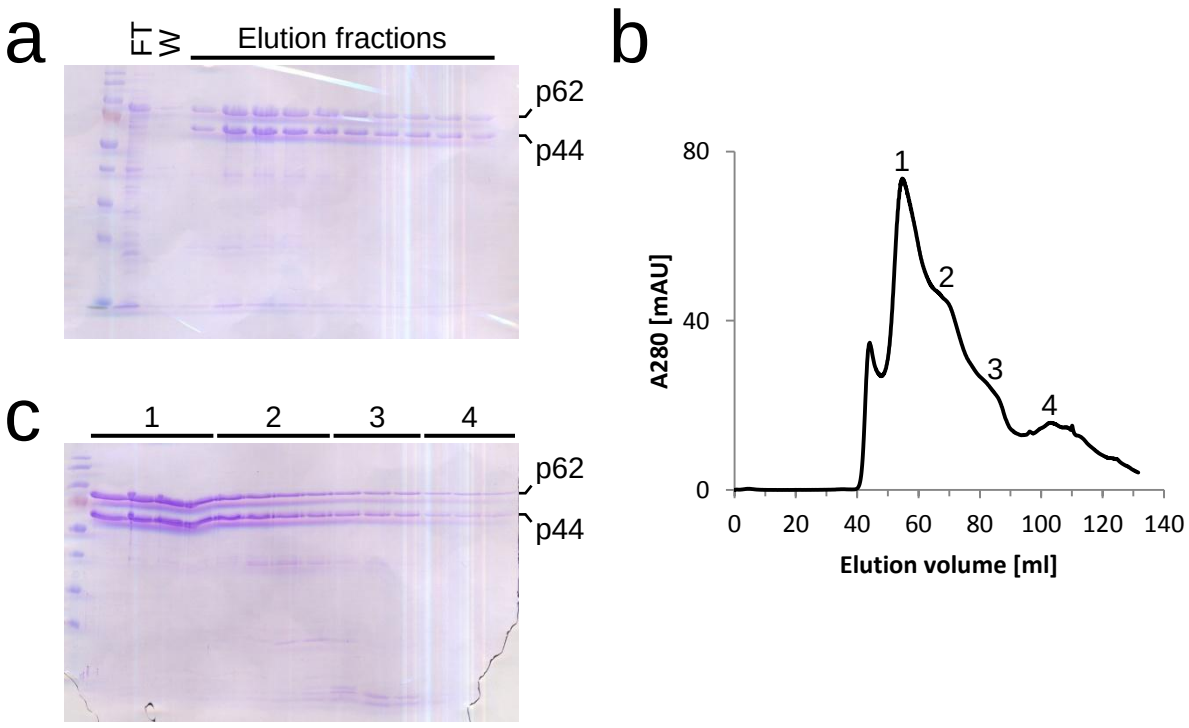


Figure 3-78. Co-purification of p62nT and p44 with standard buffer conditions. (a) SDS-PAGE analysis of IMAC samples. FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

To avoid the different oligomeric states, co-purification of p62nT and p44 under reducing conditions was attempted. Purification with TCEP supplemented to the purification buffers is depicted in Figure 3-79. A stoichiometric protein complex, eluting in one symmetric peak was obtained. Crystallization trials at 11.5 mg/ml with the commercial crystallization screens using the crystallization robot were performed yielding no crystals. Composition of the TCEP containing buffers is given below.

Lysis buffer

- 20 mM Tris-HCl pH 7.5
- 0.3 M NaCl
- 5 mM Imidazole
- 1 mM TCEP

Elution buffer

- 20 mM Tris-HCl pH 7.5
- 0.3 M NaCl
- 0.25 M Imidazole
- 1 mM TCEP

Gel filtration buffer

- 20 mM Tris-HCl pH 7.5
- 0.25 M NaCl
- 1 mM TCEP

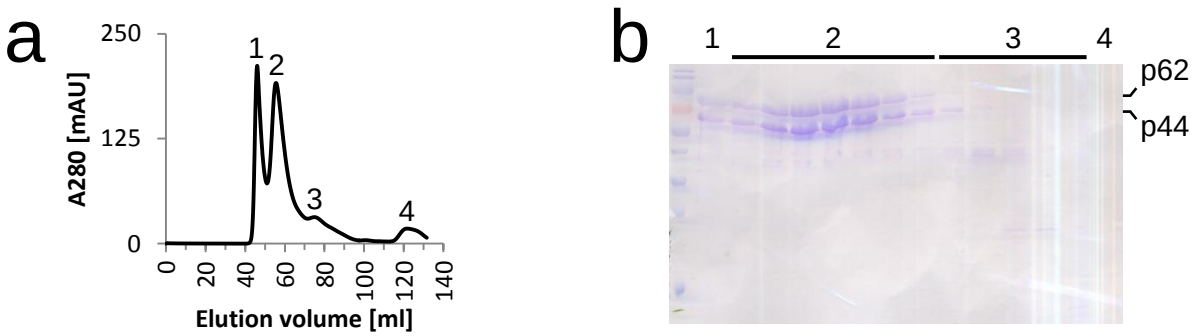


Figure 3-79. Co-purification of p62nT and p44 in the presence of TCEP. (a) SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(b)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (a).

3.11.2.2. p62/p44_1-367

To locate the p62 binding site on p44, construct p44_1-367 lacking the C-terminal ring finger domain (Figure 3-39) was co-expressed and co-purified with non-tagged p62. Purification of p62nT/p44_1-367 from BL21 CodonPlus (DE3) RIL cells with the standard buffer system is depicted in Figure 3-80. A stoichiometric complex, eluting in a single peak could be obtained. Crystallization trials at 8.3 mg/ml were performed with the commercial crystallization screens using the crystallization robot yielding no crystals.

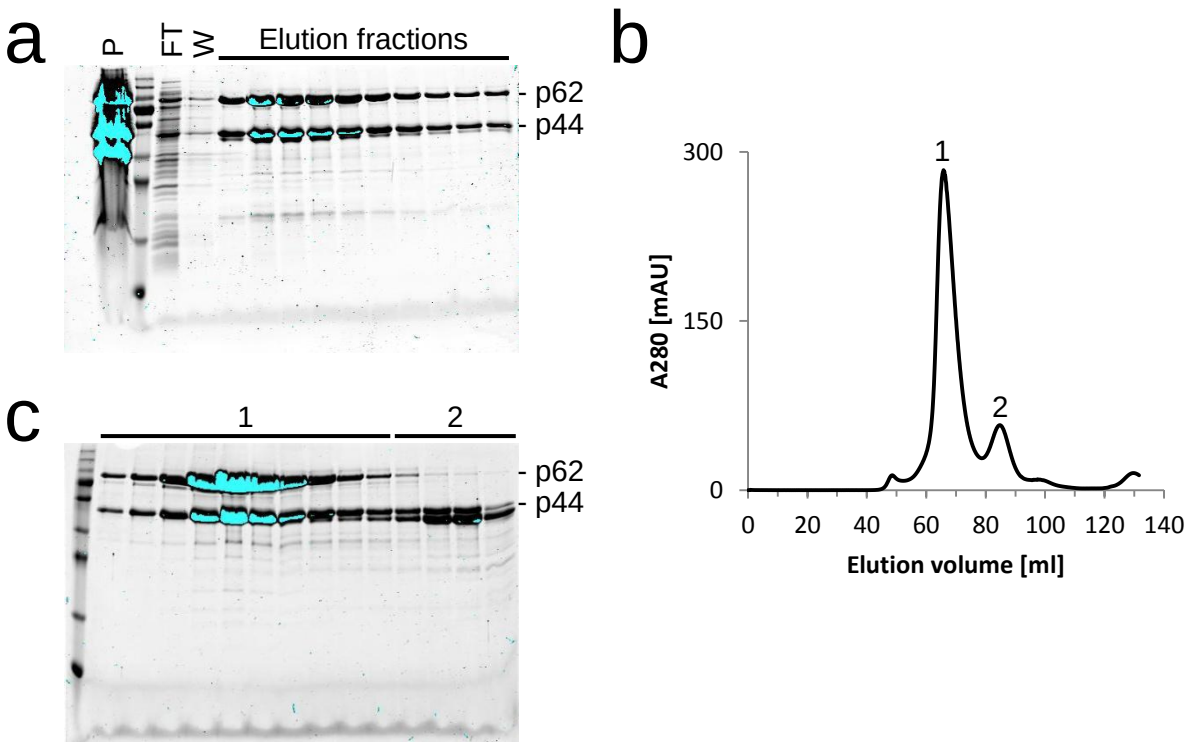


Figure 3-80. Co-purification of p62nT and p44_1-367. (a) SDS-PAGE analysis of IMAC samples. P: pellet; FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

3.11.2.3. p62/p44_286-E

To further investigate the p62 binding site on p44, construct p44_286-E lacking the N-terminal vWA like domain (Figure 3-39) was co-expressed and co-purified with non-tagged p62. The purification of p62nT/p44_286-E from BL21 CodonPlus (DE3) RIL cells with the standard buffer system is shown in Figure 3-81. P62nT and p44_286-E co-eluted during the SEC in stoichiometric amounts in a single peak, indicative of complex formation. However, significant additional bands were visible beneath p62 and p44. As the number and intensity of these bands seemed to increase from IMAC to SEC, degradation might be a possible cause for this observation. P62nT/p44_286-E was subjected to crystallization trials at 7.0 mg/ml, but no crystals could be obtained.

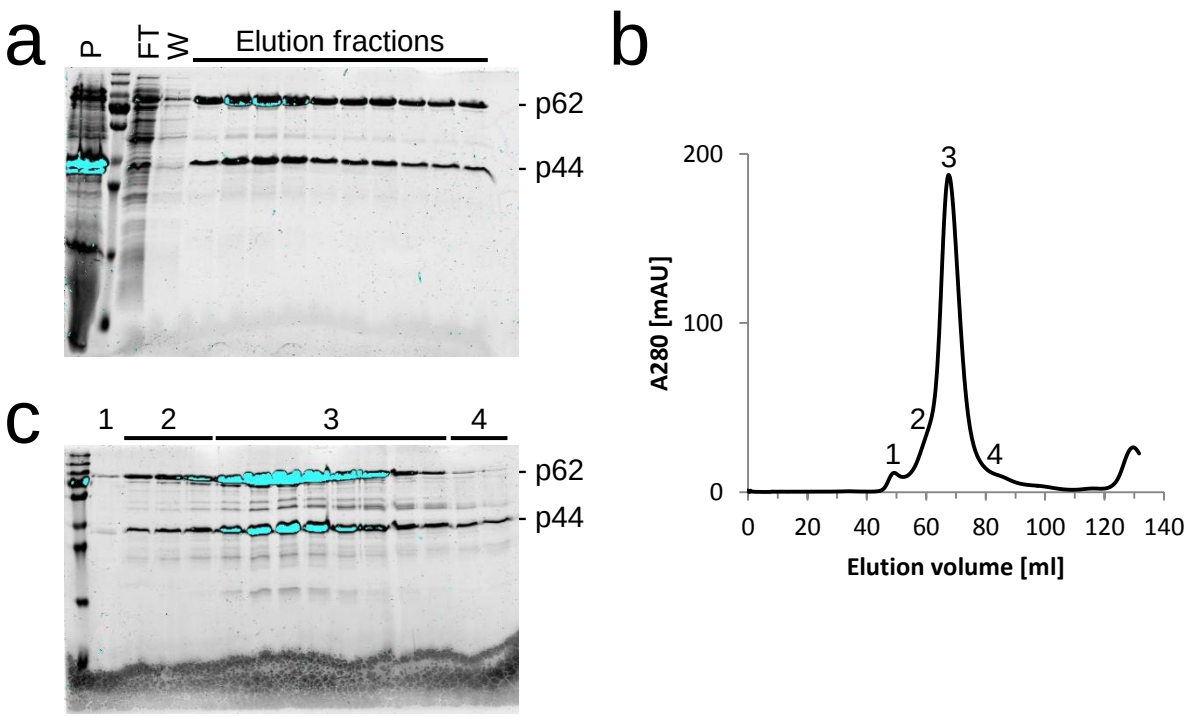


Figure 3-81. Co-purification of p62nT and p44_286-E. (a) SDS-PAGE analysis of IMAC samples. P: pellet; FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

3.11.2.4. p62/p44_286-367

To further narrow down the p62 binding site on p44, construct p44_286-367 lacking both, the N-terminal vWA like domain and the C-terminal ring finger domain (Figure 3-39) was co-expressed and co-purified with non-tagged p62. Purification of p62nT/p44_286-367 from BL21 CodonPlus (DE3) RIL cells with the standard buffer system is depicted in Figure 3-82. Both proteins co-eluted in stoichiometric amounts during the SEC in a single peak, localizing the p62 binding site to the central zinc finger domain of p44. As also observed for p62/p44_286-E (see 3.11.2.3), significant additional bands were visible in the SDS-PAGE. As a quite prominent band could be observed in the SDS-PAGE after the SEC (Figure 3-82 c), the possibility of a stable degradation product was tested by limited proteolysis. However, this approach was not successful, as all protein was rapidly degraded by addition of protease (data not shown). Crystallization trials with p62nT/p44_286-367 at 10.4 mg/ml with the commercial crystallization screens using the crystallization robot were performed, but no crystals could be obtained.

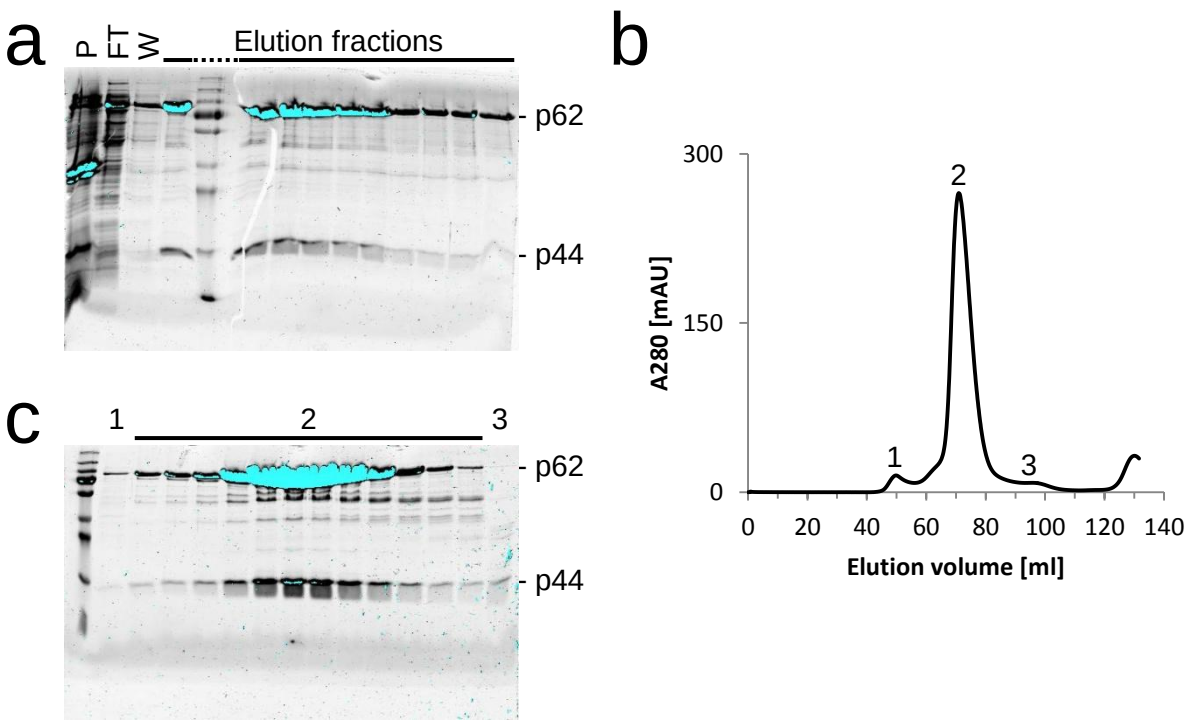


Figure 3-82. Co-purification of p62nT and p44_286-367. (a) SDS-PAGE analysis of IMAC samples. P: pellet; FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

3.11.2.5. p62/p44_1-326 and p62/p44_327-E

To define the p62 binding site on p44 more precisely, two p44 constructs either containing the first half (p44_1-326, Figure 3-39) or second half (p44_327-E, Figure 3-39) of the central zinc finger domain were tested for co-expression and co-purification with non-tagged p62. Purification of p62nT/p44_1-326 from BL21 CodonPlus (DE3) RIL cells with the standard buffer system is depicted in Figure 3-83. Both proteins were present, but eluted clearly separately during the SEC (Figure 3-83 b, c). Purification of p62nT/p44_327-E from BL21 CodonPlus (DE3) RIL cells with the standard buffer system is depicted in Figure 3-84. Here, only small amounts of p62nT could be obtained. Taken together, these results suggest that the splitting of the central zinc finger domain of p44 destroyed secondary structure elements, and thus rendered the interaction with p62 impossible.

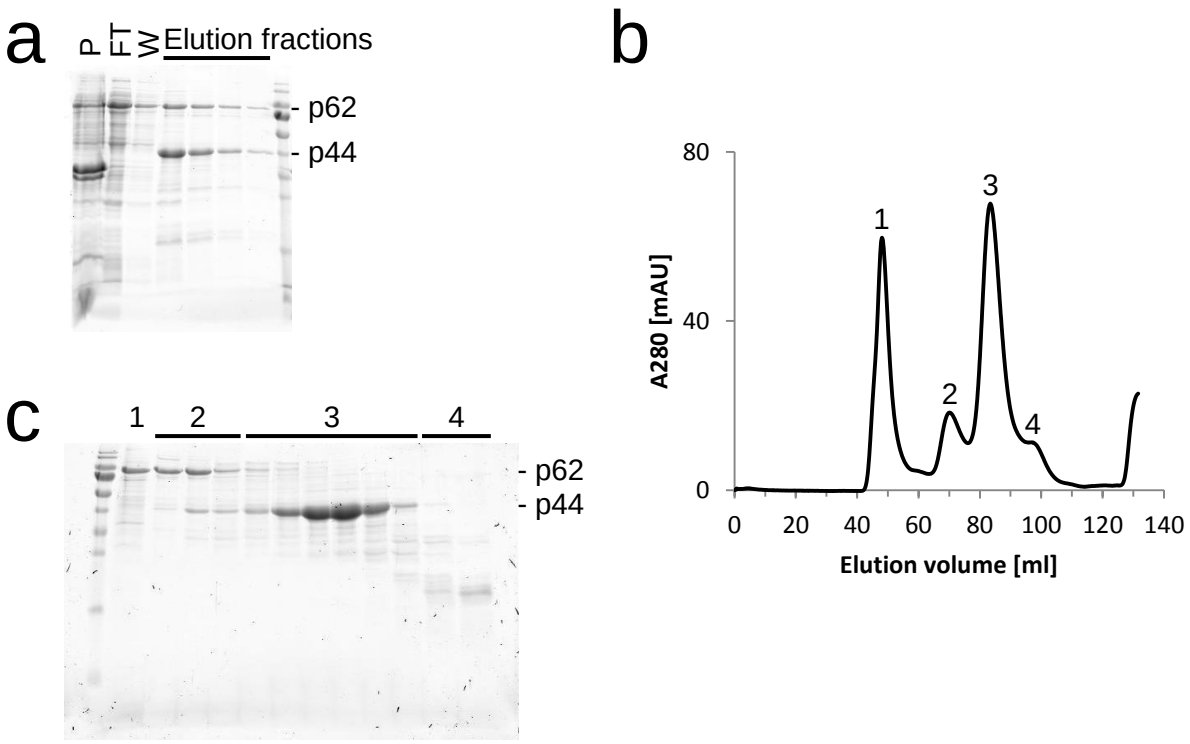


Figure 3-83. Co-purification of p62nT and p44_1-326. (a) SDS-PAGE analysis of IMAC samples. P: pellet; FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

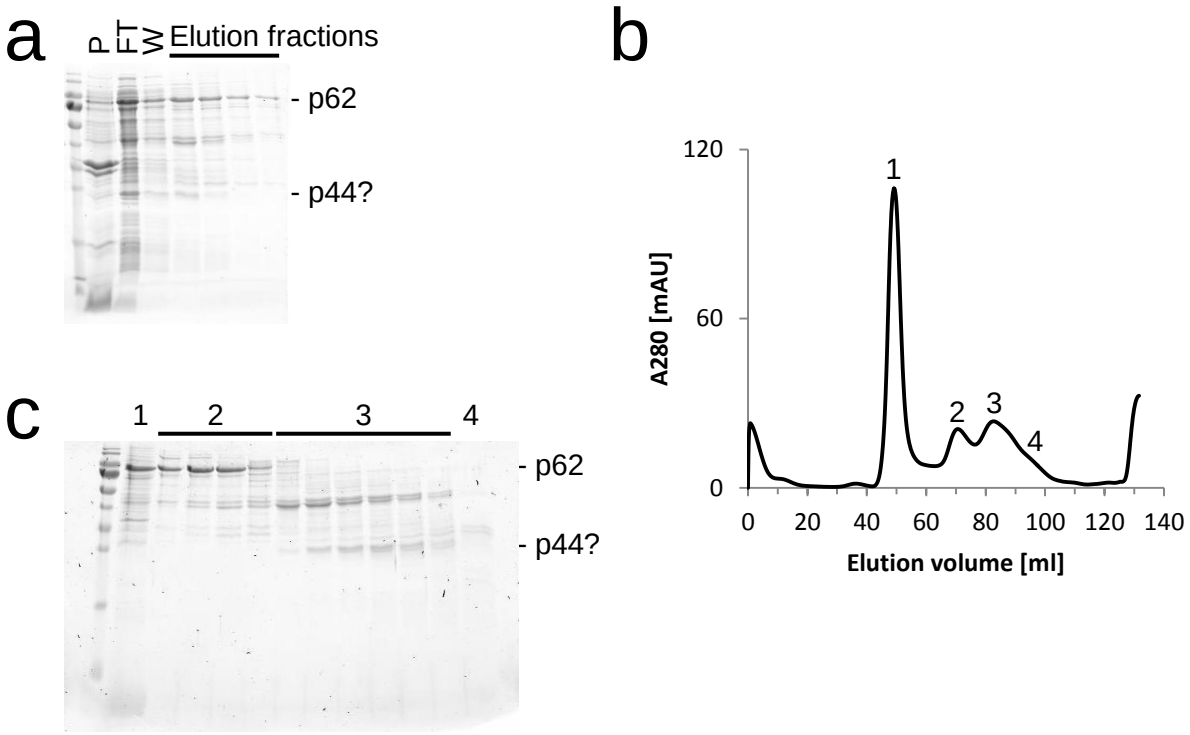


Figure 3-84. Co-purification of p62nT and p44_327-E. (a) SDS-PAGE analysis of IMAC samples. P: pellet; FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

3.11.2.6. p62_435-E/p44_286-367

To confirm the p44 binding site at the C-terminus of p62 and to obtain a minimal complex for crystallization attempts, p44_286-367 was co-expressed and co-purified with non-tagged p62_435-E, a construct encompassing the C-terminal 243 residues of p62 (Figure 3-19). Purification of p62nT_435-E/p44_286-367 from BL21 CodonPlus (DE3) RIL cells with the standard buffer system is depicted in Figure 3-85. A stoichiometric complex, eluting in a single peak could be obtained. Crystallization trials at 4.7 mg/ml with the commercial crystallization screens using the crystallization robot were performed, but no crystals could be obtained.

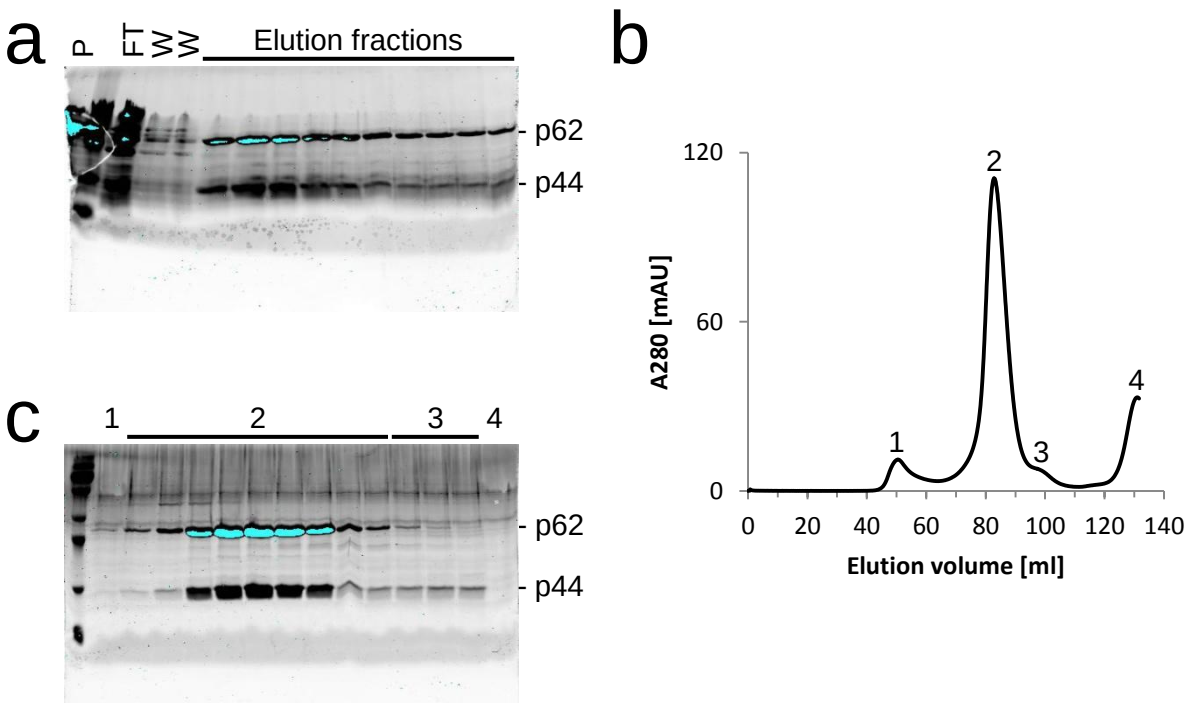


Figure 3-85. Co-purification of p62nT_435-E and p44_286-367. (a) SDS-PAGE analysis of IMAC samples. P: pellet; FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

3.11.2.7. Native PAGE

To test the vWA like domain of p44 for its interaction with p62, the construct p44_1-285, encompassing the vWA like domain (Figure 3-39), was subjected to native PAGE against different p62 constructs. None of the p62 constructs led to a band shift, thus suggesting that the vWA like domain of p44 does not interact with p62 (Figure 3-86). P44_1-285 was provided by Dominik Schmitt.

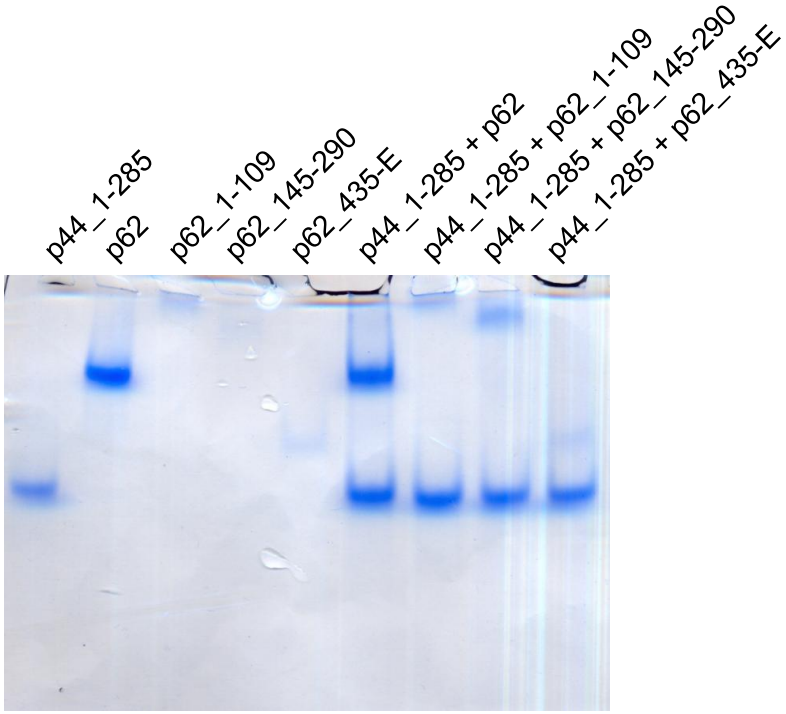


Figure 3-86. Interaction analysis between p44_1-285 and different p62 constructs via native PAGE.

3.11.3. p52 and p34

TFIIH core can be subdivided into two submodules: XPD/p44/p34 and XPB/p52/p8 [79]. How these submodules are linked together is so far not well characterized. One possibility might be an interaction via p52 and p34. In an attempt to firmly establish a possible interaction between p52 and p34, co-expression and co-purification of these two proteins was conducted.

3.11.3.1. p52/p34

P52nT was co-expressed with p34 in BL21 CodonPlus (DE3) RIL cells, and an initial co-purification using the standard buffer system is depicted in Figure 3-87. Both proteins could be obtained, but did not elute as a homogenous peak during the SEC (Figure 3-87 b). In addition, the band intensities of both proteins in the SDS-PAGE analysis of the SEC did not correspond to each other (Figure 3-87 c), indicative that no complex was formed.

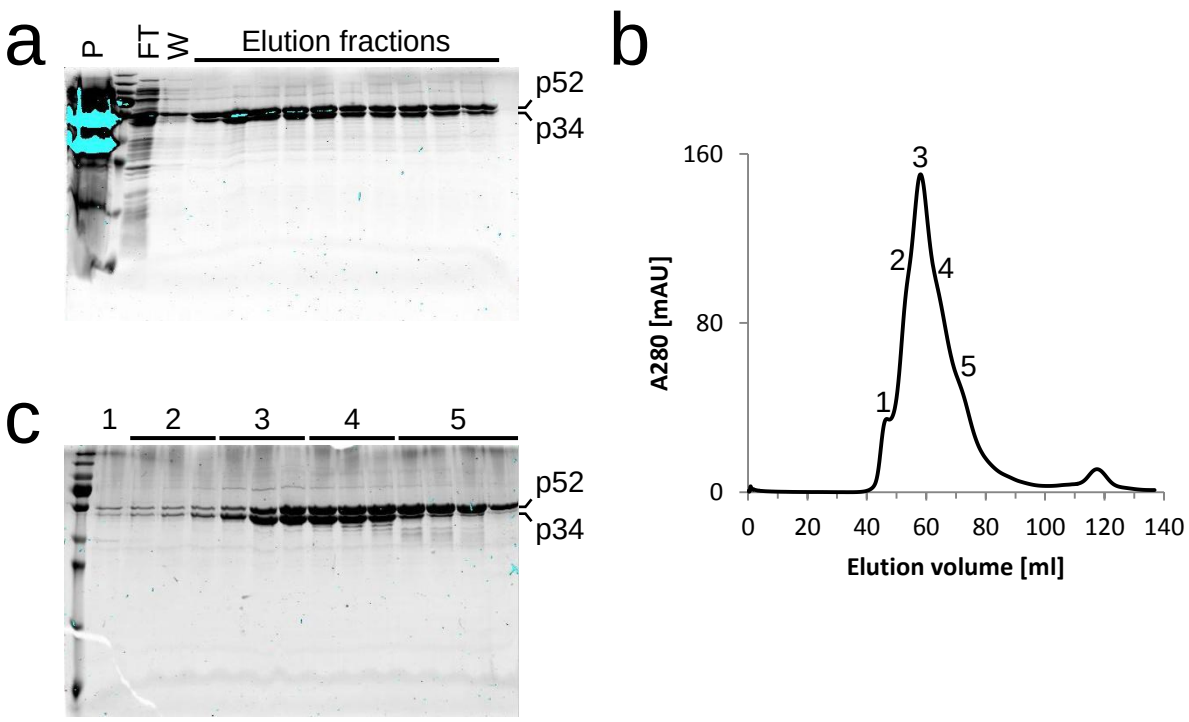


Figure 3-87. Co-purification of p52nT and p34 in standard buffer conditions. (a) SDS-PAGE analysis of IMAC samples. P: pellet; FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

As complex formation between p52 and p34 in standard buffer conditions was ambiguous, co-purification under reducing conditions was attempted utilizing TCEP. Purification of p52nT/p34 from BL21 CodonPlus (DE3) RIL cells in the presence of TCEP is depicted in Figure 3-88. Both proteins co-eluted in stoichiometric amounts in a single peak during the SEC, indicative of complex formation. The composition of the TCEP containing buffer system is given below.

Lysis buffer

- 20 mM Tris-HCl pH 7.5
- 0.3 M NaCl
- 5 mM Imidazole
- 1 mM TCEP

Elution buffer

- 20 mM Tris-HCl pH 7.5
- 0.3 M NaCl
- 0.25 M Imidazole
- 1 mM TCEP

Gel filtration buffer

- 20 mM Tris-HCl pH 7.5
- 0.25 M NaCl
- 1 mM TCEP

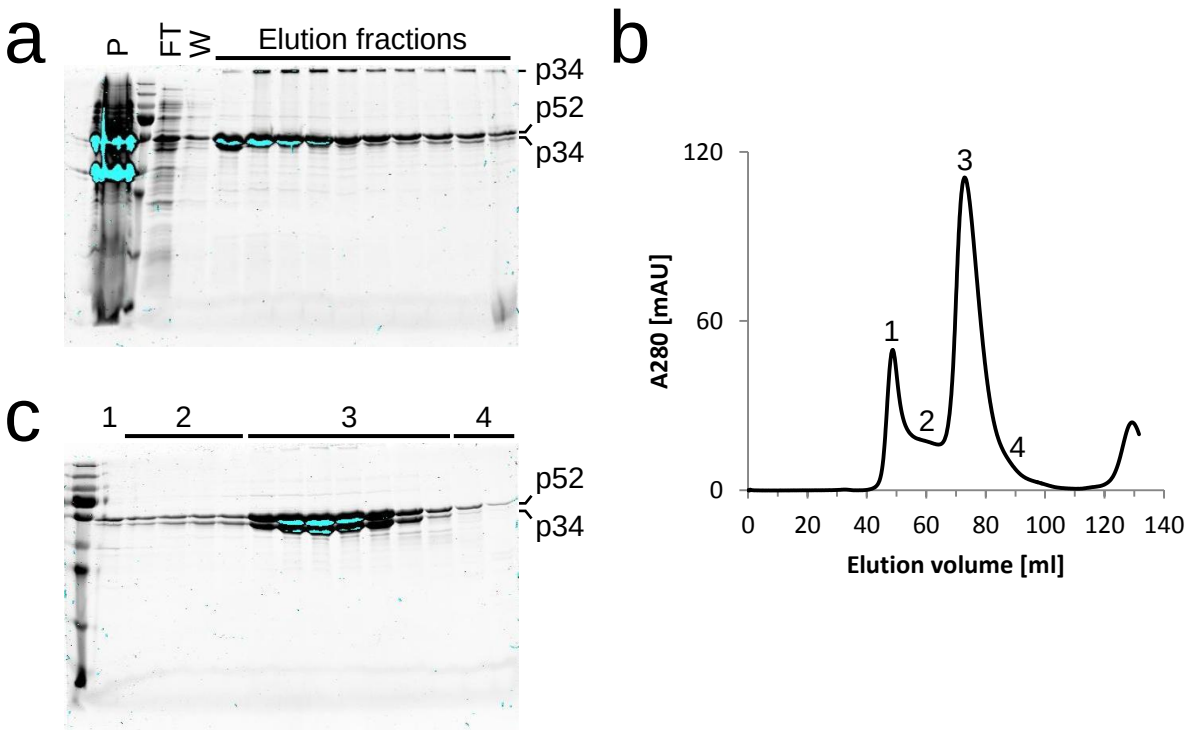


Figure 3-88. Co-purification of p52nT and p34 in presence of TCEP. (a) SDS-PAGE analysis of IMAC samples. P34 partly stuck to the pockets. P: pellet; FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

3.11.3.2. p52/p34_1-277

To further characterize the interaction between p52 and p34, and to narrow down the p52 binding site on p34, co-expression and co-purification of construct p34_1-277, lacking the C-terminal zinc finger domain (Figure 3-40), with non-tagged p52 was attempted. Purification of p52nT/p34_1-277 from BL21 CodonPlus (DE3) RIL cells in the TCEP containing buffer system described above (see 3.11.3.1) is depicted in Figure 3-89. Both proteins co-eluted during the SEC in stoichiometric amounts, indicative of for complex formation. However, the yield and quality of the complex seemed to be inferior compared to the complex utilizing full-length p34.

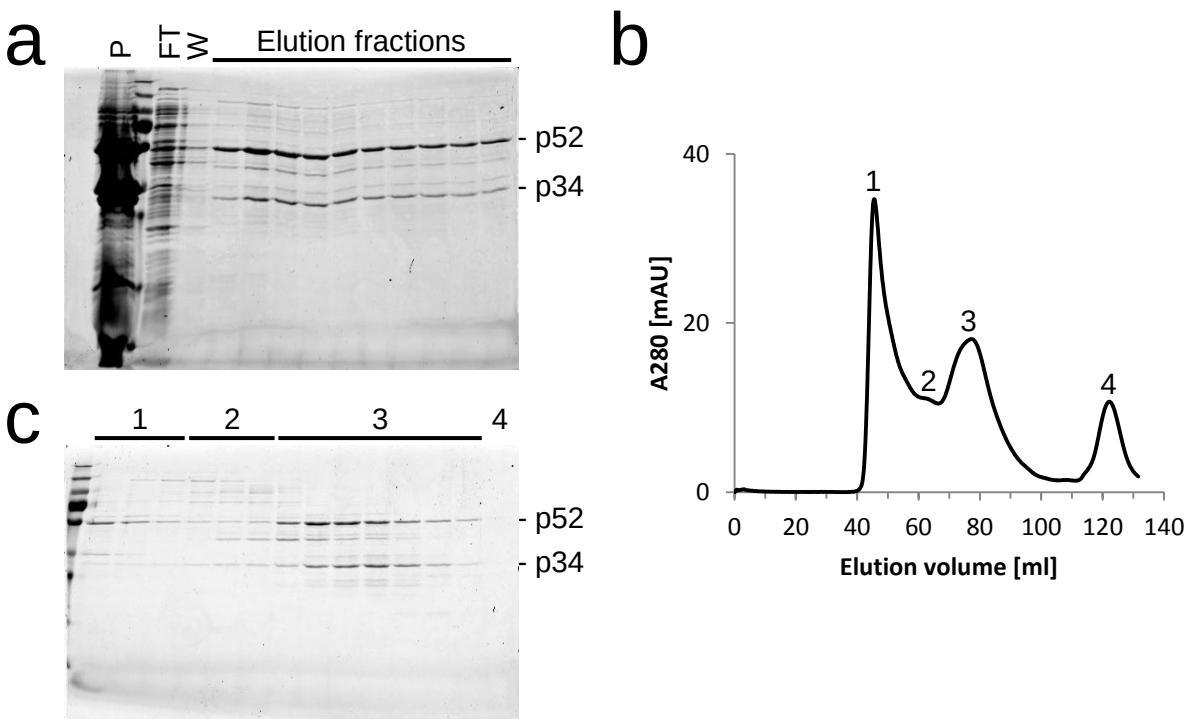


Figure 3-89. Co-purification of p52nT and p34_1-277. (a) SDS-PAGE analysis of IMAC samples. P: pellet; FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

3.11.4. p52 and p8

To further stabilize p52 and to establish the interaction with p8 in the *C. thermophilum* model system, p52 was co-expressed and co-purified with p8. To ensure a stoichiometric complex, p52nT together with tagged p8 was used. P52 was chosen as non-tagged partner based on two considerations: First, during SEC an excess of p8 should readily be separated from p52/p8 compared to an excess of p52 due to the larger size difference. Second, because of the small size, p8 is difficult to visualize on an SDS-PAGE. The occurrence of high amounts of non-tagged p52 would then be an indication of the presence of p8 as well. P52nT/p8 was co-expressed in BL21 CodonPlus (DE3) RIL cells, and the purification with a Hepes buffer system is depicted in Figure 3-90. The composition of the Hepes buffer system is given below.

Lysis buffer

- 20 mM Hepes-NaOH pH 8.0
- 0.375 M NaCl
- 5 mM Imidazole

Elution buffer

- 20 mM Hepes-NaOH pH 8.0
- 0.375 M NaCl
- 0.25 M Imidazole

Gel filtration buffer

- 20 mM Hepes-NaOH pH 8.0
- 0.375 M NaCl

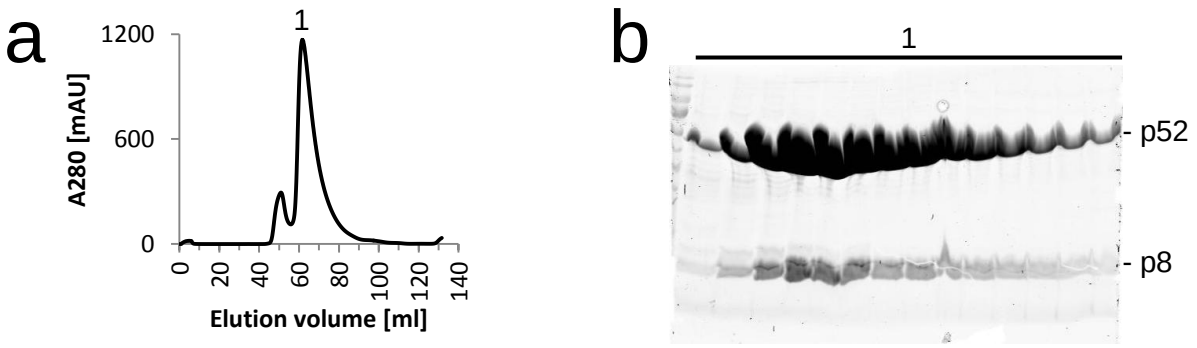


Figure 3-90. Co-purification of p52nT and p8. (a) SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(b)** SDS-PAGE analysis of elution fractions from the SEC. The number corresponds to the peak in (a).

3.11.5. XPB and MAT1

As an interaction between MAT1 and XPB has been suggested [80], co-expression of non-tagged XPB with MAT1_1-248 (Figure 3-43) was attempted. The MAT1_1-248 expression vector was provided by Florian Sauer. MAT1_1-248 has been inserted into a modified pCDFDuet™-1 vector (Novagen), containing the multiple cloning site of the pETM-22 vector (EMBL). Thus, MAT1_1-248 contained an N-terminal thioredoxin tag, followed by a hexa-histidine tag and a 3C cleavage site. Co-expression was performed in BL21 CodonPlus RIL cells, and co-purification of XPBnT and MAT1_1-248 is depicted in Figure 3-91. A clear band on the expected height of MAT1_1-248 was observed, but no band on the expected height of XPB. Most likely, expression of MAT1_1-248 interfered with the expression of XPB, as also observed in the co-expression attempts of XPB with p52 (see 3.11.1.1).

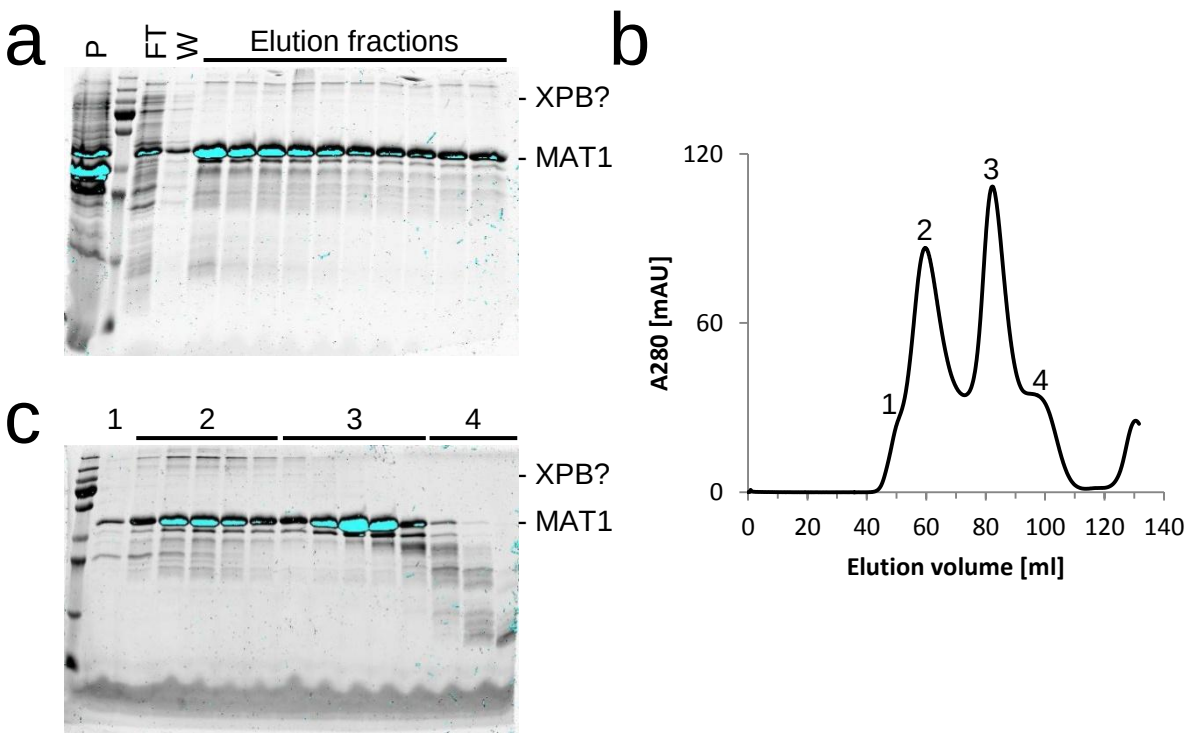


Figure 3-91. Co-purification of XPBnT and MAT1_1-248. (a) SDS-PAGE analysis of IMAC samples. P: pellet; FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

3.12. The TFIIH core interaction network

The different dual complexes established above (see 3.11) were utilized in native PAGE to investigate and elucidate the intricate interaction network within TFIIH core.

3.12.1. Anchoring of XPD

Of critical importance for the structural integrity of TFIIH is the anchoring of XPD. XPD is anchored to TFIIH core via the vWA like domain of p44 [30]. Furthermore, it has been suggested that p62 participates in XPD anchoring [80, 481]. Therefore, XPD was tested for its interaction with the different p62/p44 complexes described above (see 3.11.2). The native PAGE analysis is depicted in Figure 3-92. Band shifts could only be observed for p62/p44 complexes where the vWA like domain of p44 was present. This observation led to two hypotheses: First, p62 does not interact with XPD directly, as full-length p62 was present in each complex tested. Second, the central zinc finger domain and the C-terminal ring finger domain of p44 do not participate in the XPD interaction.

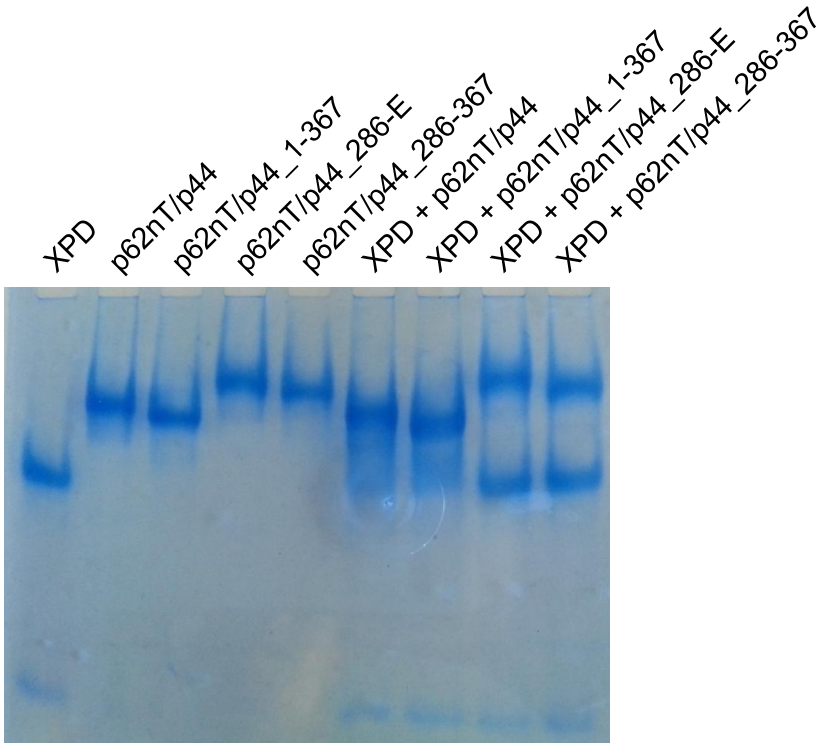


Figure 3-92. Interaction analysis between XPD and different p62/p44 complexes via native PAGE.

3.12.2. The link between the submodules XPD/p44/p34 and XPB/p52/p8

TFIIH core can be subdivided into two functionally associated subcomplexes: XPD/p44/p34 and XPB/p52/p8 [79]. The interactions within these subcomplexes are firmly established, the interactions between these two submodules however, are less well characterized. To shed light on how these subcomplexes associate, several interaction studies were conducted.

As a starting point, the single subunits XPD, XPB_1-345, p34, and p8 were examined for interaction via native PAGE (Figure 3-93). No interaction of p8 with XPD or p34 could be observed, suggesting that the link is not established via p8. Also, no interaction between XPD and XPB_1-345 or between p34 and XPB_1-345 could be detected.

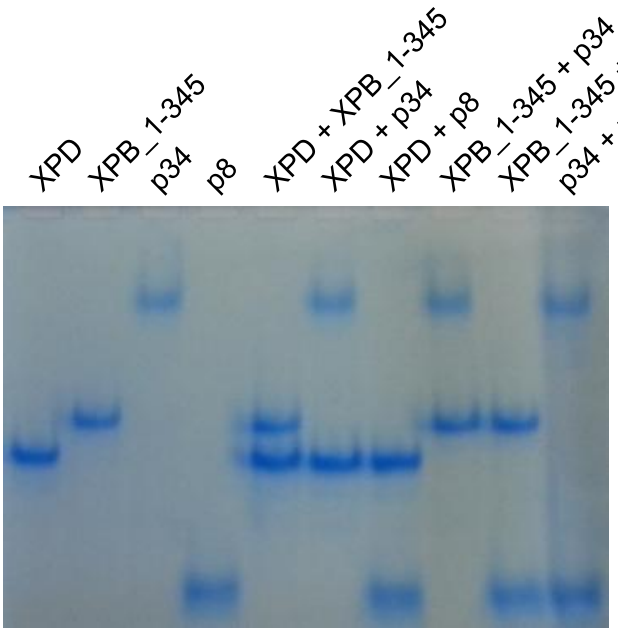


Figure 3-93. Interaction analysis of XPD, XPB_1-345, p34, and p8 via native PAGE.

Furthermore, XPB and XPD were analyzed for their direct interaction, as well as for the interaction with p62nT/p44 or p52nT/p8. The results are depicted in Figure 3-94, indicating no interaction between XPB and XPD. Also, no interaction between XPD and p52nT/p8 or between XPB and p62nT/p44 could be observed. As expected, band shifts for XPB in presence of p52nT/p8 could be observed. Remarkably, when XPB, p52nT/p8, and XPD were mixed, a certain amount of p52nT/p8 seemed to be not shifted, what was not observed when only XPB and p52nT/p8 were mixed. The cause of this observation remained unclear, and possible explanations are highly speculative. Possibilities might be pipetting errors or some influence of XPD on the complex formation between XPB and p52nT/p8.

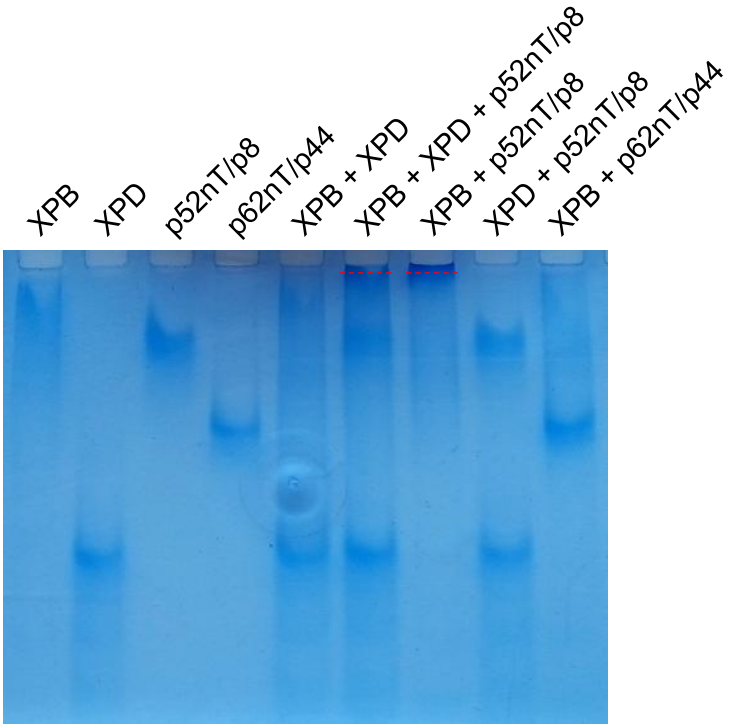


Figure 3-94. Interaction analysis of XPD, XPB, p62nT/p44, and p52nT/p8 via native PAGE. The XPB/p52nT/p8 complex bands are indicated by dashed red lines.

To determine a possible interaction between p62 and p52, p52nT/p8 was tested against the different p62/p44 complexes. Native PAGE analysis did not indicate band shifts for any of these complexes (Figure 3-95), thus not supporting a link via p62-p52, p62-p8, or p52-p44.

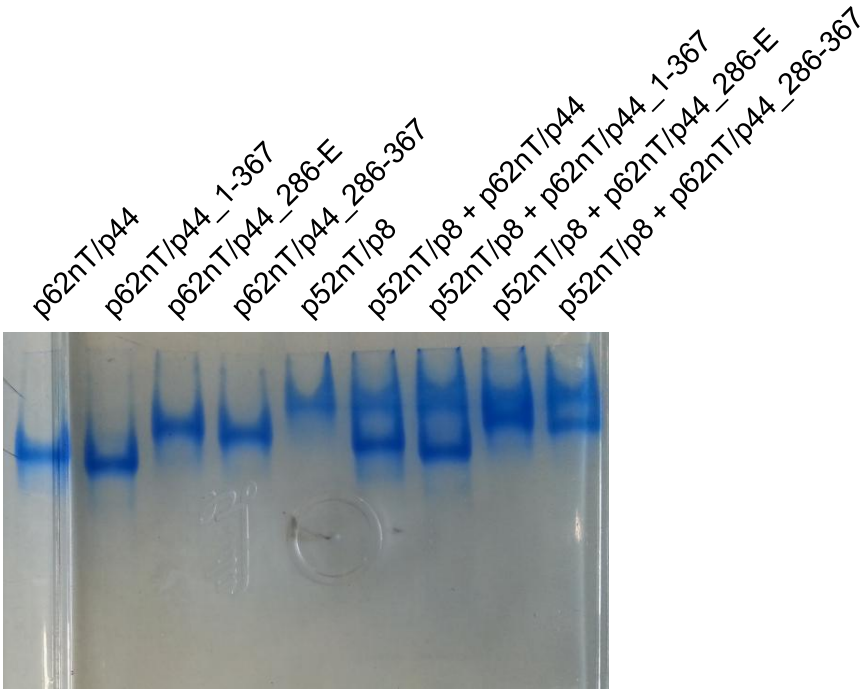


Figure 3-95. Interaction analysis of p52nT/p8 and different p62/p44 complexes via native PAGE.

To assess the role of p34 towards connecting the two submodules, native PAGE analysis was performed based on the observation that p62nT/p44 and p52nT/p8 do not interact (Figure 3-95). The addition of p34 would potentially be able to mediate complex formation between p62nT/p44 and p52nT/p8, resulting in the appearance of a distinct band. The approach is depicted in Figure 3-96. A single band could be observed for the mixture of p62nT/p44, p52nT/p8, and p34, clearly distinct from any of the control bands. This finding strongly suggests that p34 is responsible for linking the two TFIIH core submodules XPB/p52/p8 and XPD/p44/p34. This crucial link is mediated via the interaction of p34 and p52, as established through the co-purification of p52nT/p34 described above (see 3.11.3). In addition, this approach led to the assembly of a higher order TFIIH core complex, containing p62, p52, p44, p34, and p8, termed Core5. Furthermore, the usage of the bridging function of p34 via native PAGE provided a powerful and sensitive tool to further investigate the TFIIH core interaction network.

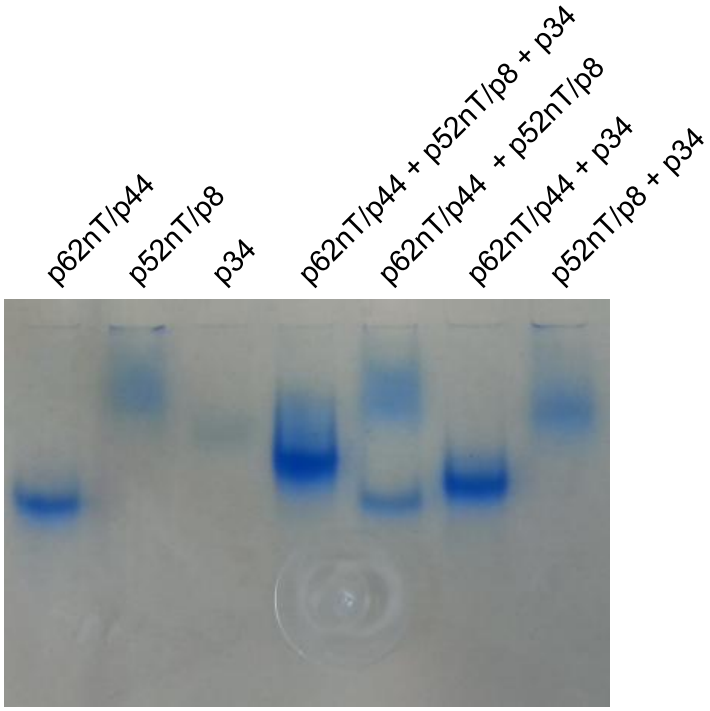


Figure 3-96. Analysis of the ability of p34 to mediate complex formation between p62nT/p44 and p52nT/p8 via native PAGE.

To investigate the p34 link more specifically, the constructs p34_1-277 and p34_278-E (Figure 3-40) were tested for their ability to mediate Core5 formation. P34_1-277 was provided by Dominik Schmitt, p34_278-E was provided by Elisabeth Schönwetter. Native PAGE analysis using p34_1-277 is depicted in Figure 3-97. Successful formation of the Core5 complex could be detected, indicating that the vWA like domain of p34 is sufficient for the interaction. This notion is supported by the band shift observed for the mixture of

p52nT/p8 and p34_1-277 in the native PAGE, and by the observation of a p52nT/p34_1-277 complex above (see 3.11.3.2). Native PAGE analysis using p34_278-E is depicted in Figure 3-98. No formation of the Core5 complex could be detected, however no clear band for p34_278-E could be observed, rendering the interpretation uncertain.

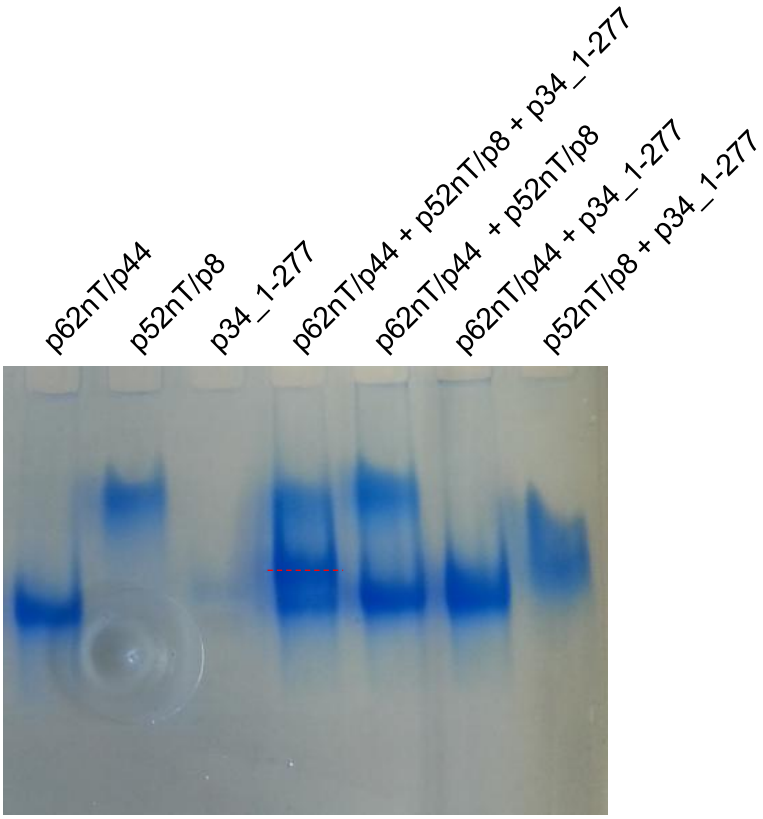


Figure 3-97. Analysis of the ability of p34_1-277 to mediate Core5 formation via native PAGE. The Core5 complex band is indicated by a dashed red line.

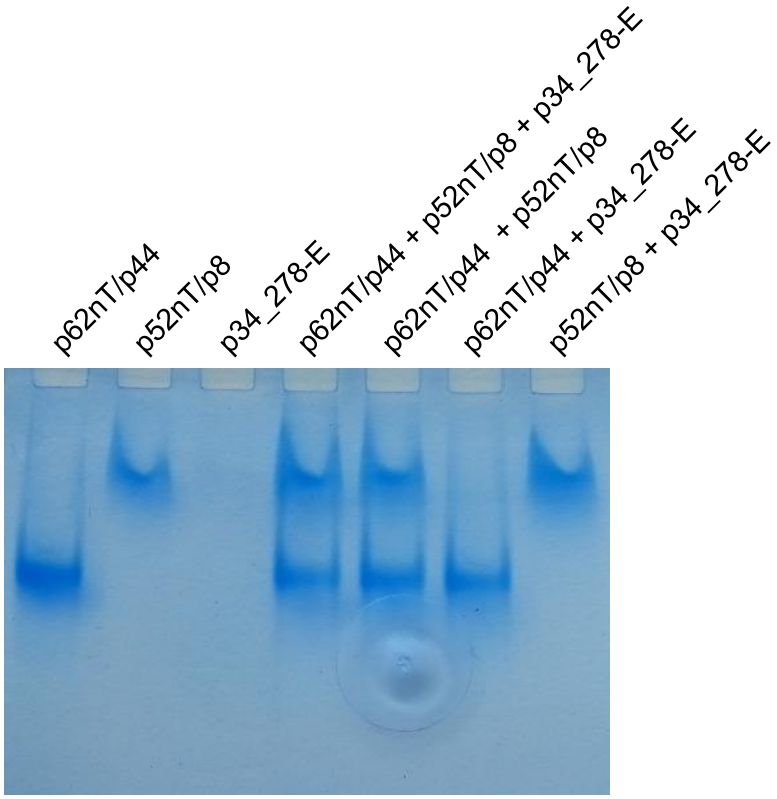


Figure 3-98. Analysis of the ability of p34_278-E to mediate Core5 formation via native PAGE.

To further investigate the contribution of p52 to the submodule linkage, construct p52_121-E (Figure 3-33) was tested for the ability to sustain complex formation mediated by p34 (Figure 3-99). A slight p34 mediated shift could be observed for p52_121-E, indicating that the very N-terminus of p52 is not necessary for interaction with p34. As controls, complex formation between p62nT/p44, p52nT/p8, and p34 as well as between p62nT/p44, p52, and p34 was monitored. In the native PAGE, p62nT/p44/p52nT/p8/p34 resulted in a clear complex band. In contrast, complex formation of p62nT/p44/p52/p34 was weaker and ambiguous. Presumably, this is due to a stabilizing effect of p8 on p52. Then again, it might also indicate a participation of p8 in the p52/p34 interaction. However, no interaction between p34 and p8 could be observed via native PAGE (Figure 3-93).

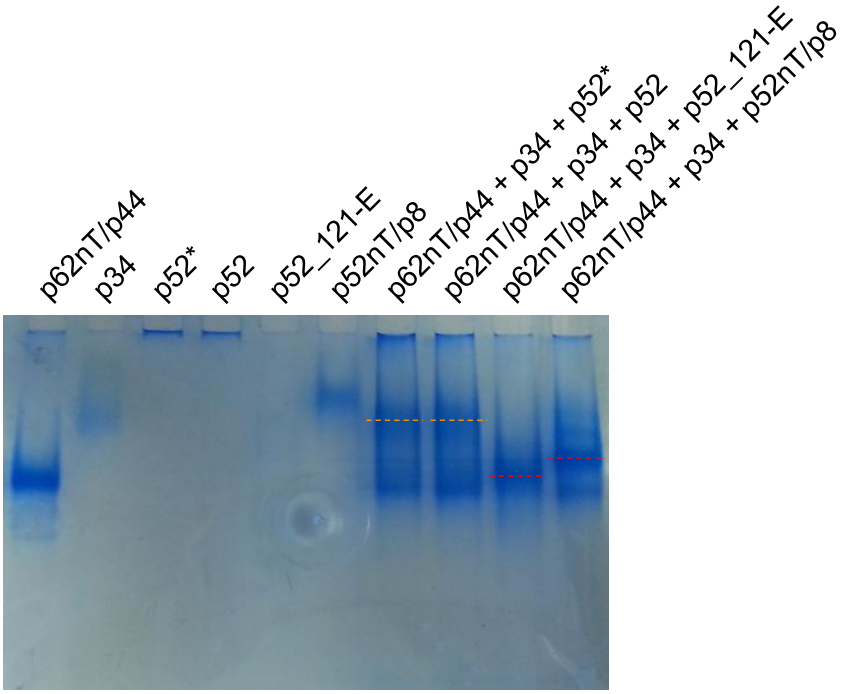


Figure 3-99. Analysis of the ability of p52_121-E to sustain complex formation mediated by p34 via native PAGE. The Core5 and p62nT/p44/p34/p52_121-E complex bands are indicated by dashed red lines. The assumed p62nT/p44/p34/p52 complex bands are indicated by dashed orange lines. P52*: protein batch stored over night at 4 °C.

3.12.3. The extended interface between p44 and p34

The bridging function of p34 established above (see 3.12.2) was utilized for a thorough investigation of the interaction interface between p44 and p34. P34 consists of an N-terminal vWA like domain and a C-terminal zinc finger domain, and it has been shown that the former interacts with the C-terminal ring finger domain of p44 [48]. To assess the role of the C-terminal zinc finger domain of p34 towards its interaction with p44, the p34 variant A151E was utilized. The single amino acid substitution A151E interferes with the binding interface between the vWA like domain of p34 and the C-terminal ring finger domain of p44 and was provided by Elisabeth Schönwetter (Radu, Schönwetter et al., [482]). However, analysis via native PAGE showed that P34_A151E was able to mediate Core5 formation between p62nT/p44 and p52nT/p8 (Figure 3-100). Next, the A151E variant of the shortened construct p34_1-277 (also provided by Elisabeth Schönwetter), lacking the C-terminal zinc finger domain (Figure 3-40), was analyzed towards its ability to mediate Core5 formation (Figure 3-101). In contrast to p34_A151E, no Core5 formation could be observed with this construct. Taken together, these results strongly suggest an extended binding interface between p44 and p34, involving the C-terminal zinc finger domain of p34 in addition to its vWA domain.

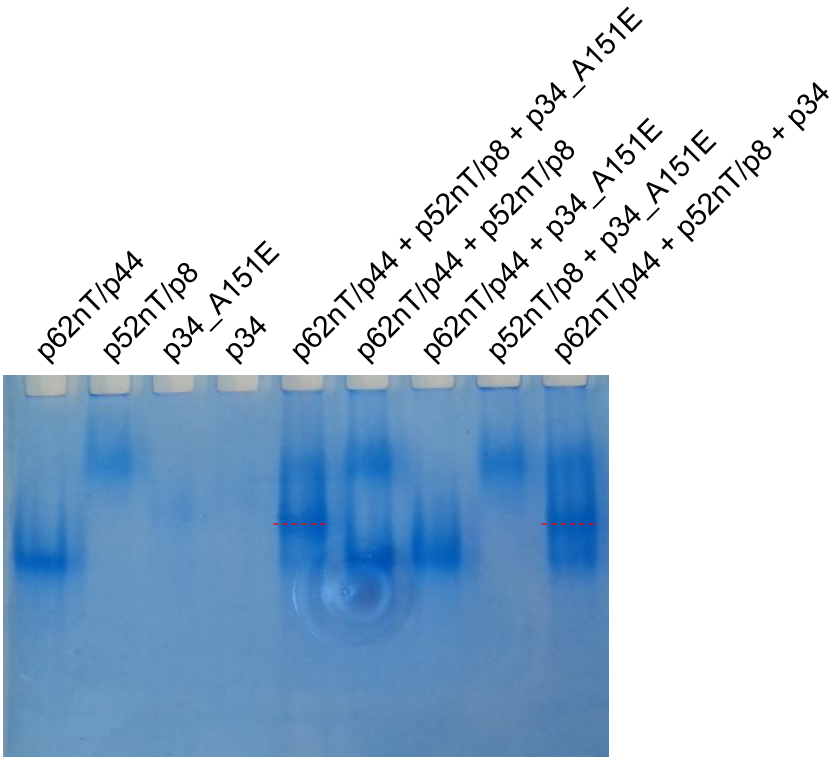


Figure 3-100. Analysis of the ability of the p44 interface variant p34_A151E to mediate Core5 formation via native PAGE. The Core5 complex bands are indicated by dashed red lines.

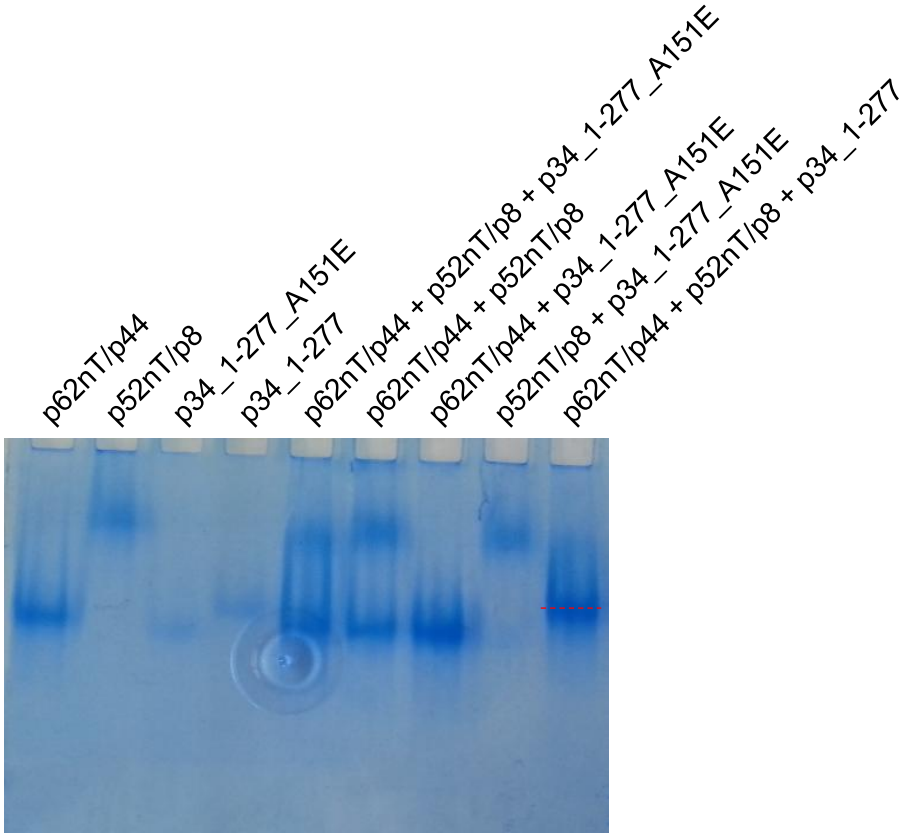


Figure 3-101. Analysis of the ability of the p44 interface mutant p34_1-277_A151E to mediate Core5 formation via native PAGE. The Core5 complex band is indicated by a dashed red line.

Then the question was addressed if the binding interface also extends beyond the C-terminal ring finger domain of p44. For this purpose the p62nT/p44_1-367 complex (see 3.11.2.2), lacking the C-terminal ring finger domain of p44, was utilized. The ability of p62nT/p44_1-367 to sustain Core5 formation with p52nT/p8 mediated by p34 was analyzed via native PAGE (Figure 3-102). Even though attenuated compared to the full-length complex, p62nT/p44_1-367 was able to successfully sustain Core5 formation, as it is apparent from the faint shifted band in lane 5, suggesting that the p34 binding site of p44 is extending beyond the C-terminal ring finger domain of p44.

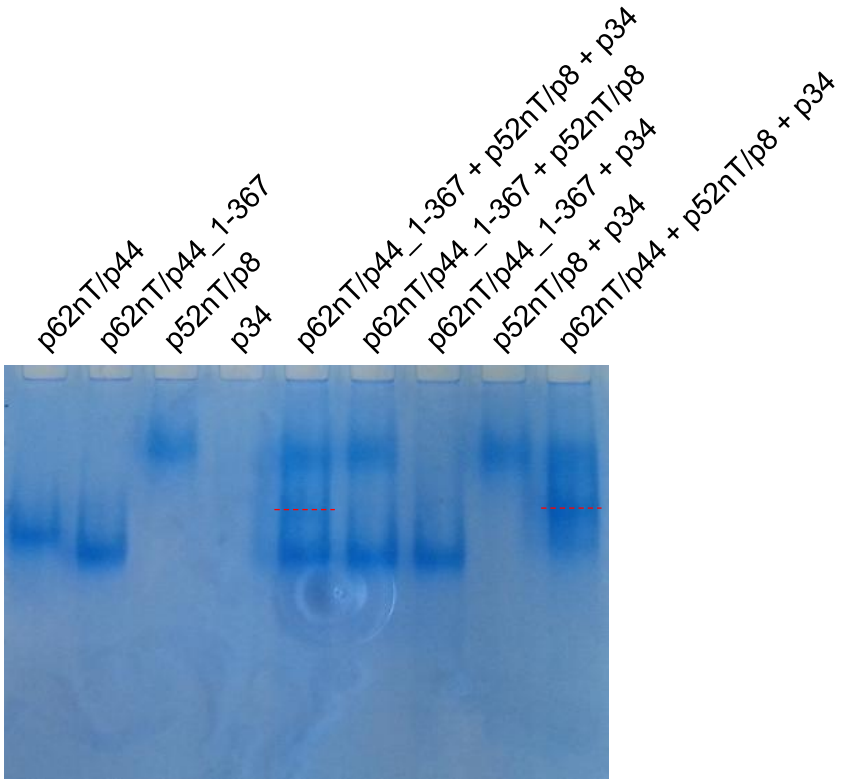


Figure 3-102. Analysis of the ability of p62nT/p44_1-367 to sustain Core5 formation via native PAGE. The Core5 complex bands are indicated by dashed red lines.

To further investigate if the vWA like domain or the central zinc finger domain of p44 contributes to the extended p34 binding interface, construct p44_1-285, only comprising the N-terminal vWA like domain (Figure 3-39), was analyzed for its ability to sustain complex formation with p52nT/p8 mediated by p34 via native PAGE (Figure 3-103). However, p44_1-285 failed to sustain complex formation, and no interaction between p44_1-285 and p34 without p52nT/p8 could be observed. Taken together, these results suggest that the central zinc finger domain of p44 participates in the interaction with p34 in addition to its C-terminal ring finger domain.

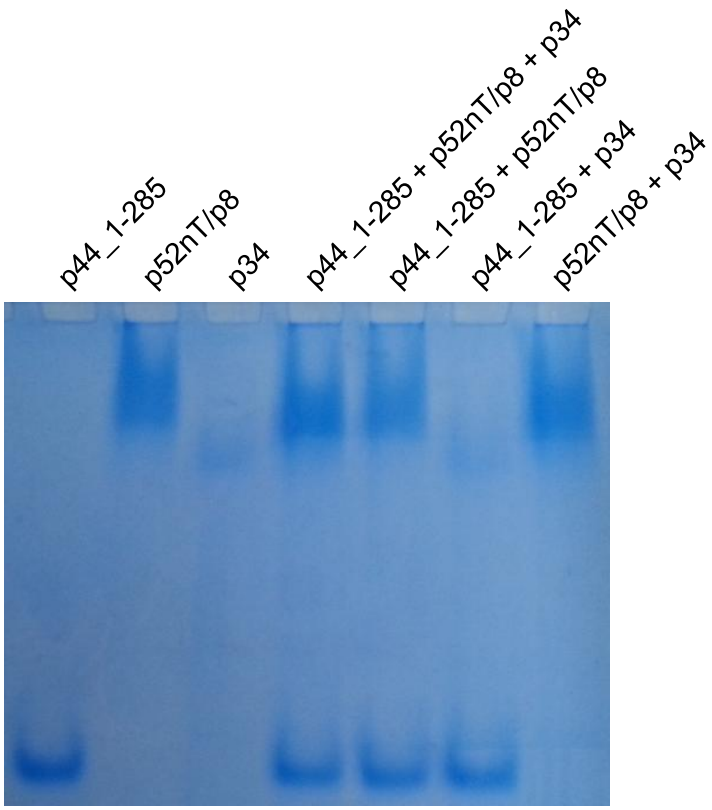


Figure 3-103. Analysis of the ability of p44_1-285 to sustain complex formation with p52nT/p8 mediated by p34 via native PAGE.

3.12.4. Verification of the p52 delta linker variant integrity

The amino acid region 322-344 in p52 has been replaced by a short linker in p52_dL and p52_121-E_dL (Figure 3-33), in order to improve the crystallization propensities of p52. To ensure that this substitution has no impact on the structural and functional integrity of p52, the ability of p52_dL to successfully sustain Core5 formation was analyzed. For this purpose, the bridging function of p34 was utilized, as described (see 3.12.2). P52nT_dL/p8 was expressed and purified as described in section 3.11.4, and subjected to native PAGE in the presence of p62nT/p44 and p34 (Figure 3-104). P52nT_dL/p8 successfully sustained Core5 formation, strongly indicating that the linker region is not required to maintain the integrity of the five subunit complex.

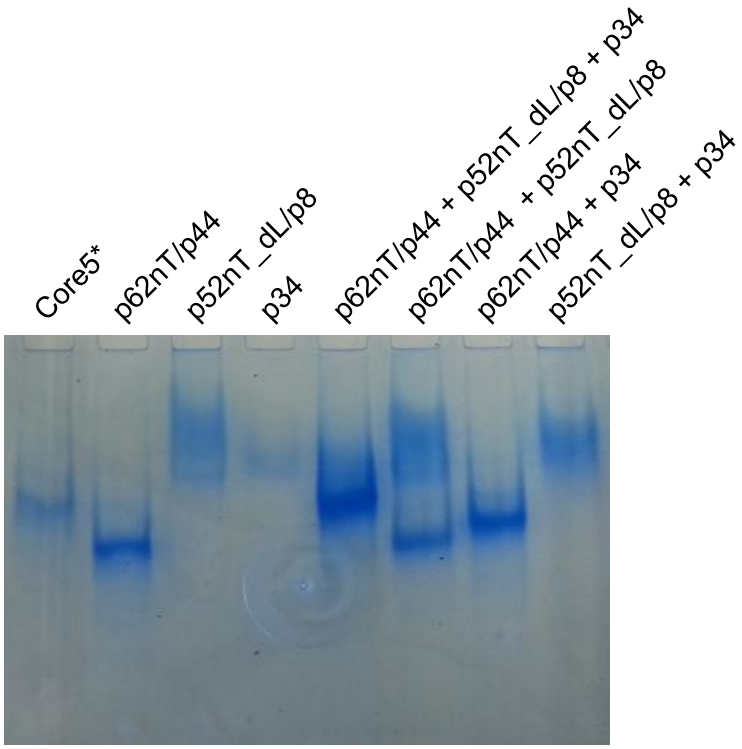


Figure 3-104. Verification of the integrity of the delta linker variant of p52. The ability of p52nT_dL to sustain Core5 formation was monitored via native PAGE. Core5*: complex of p62nT/p44, p52nT/p8, and p34 assembled on a Superdex 200 10/300 GL column.

3.12.5. The disease related XPB F143S variant

To investigate the impact of the disease related XPB mutation F143S on the interaction with p52, construct XPB_1-345_F143S was analyzed by native PAGE in the presence of p52 and p52nT/p8 (Figure 3-105). For the wild-type XPB construct a clear band shift with p52 as well as p52nT/p8 could be observed, confirming complex formation. In contrast, no band shifts for the XPB_1-345_F143S variant could be observed. However, with respect to a single amino acid substitution, the XPB_1-345_F143S variant by itself already displayed a drastically different running behavior compared to the wild-type construct. This difference might indicate that F143S is not a true interface mutant, but may corrupt the fold of the p52 binding region.

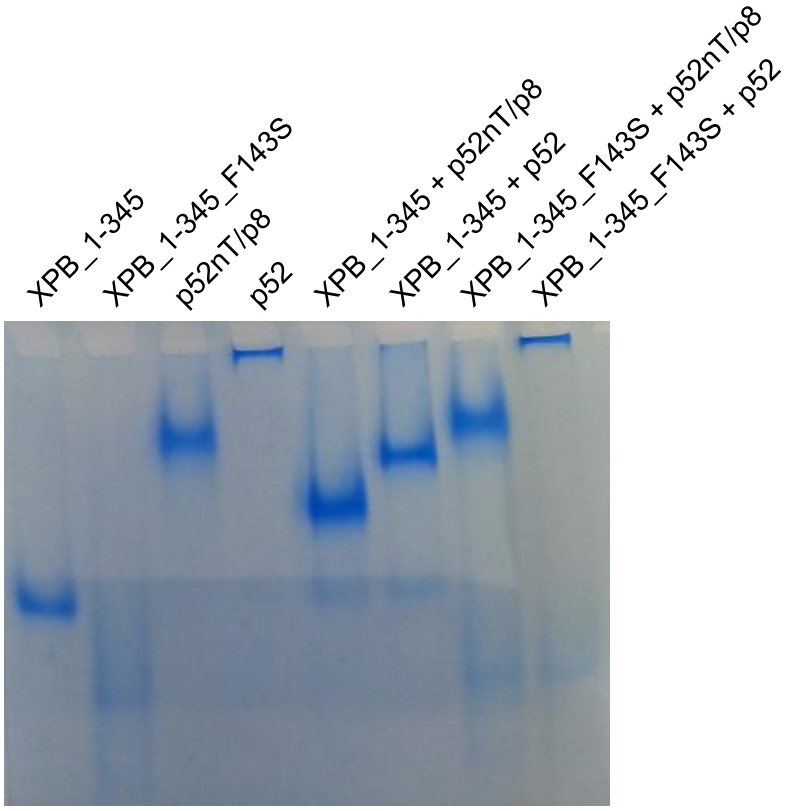


Figure 3-105. Analysis of the influence of the disease related XPB mutation F143S on the interaction between XPB_1-345 and p52 via native PAGE.

3.13. Core complexes

To gain structural insights into the entire TFIIH core, higher order core complexes were assembled. These assemblies were used for structural studies via EM. TFIIH core complexes were reconstituted via SEC, utilizing the established dual complexes and single subunits. In general, two buffer systems were used: a Tris/TCEP and a Ches/TCEP buffer system. The composition of these buffer systems is given below.

Tris/TCEP buffer

- 20 mM Tris-HCl pH 7.5
- 0.25 M NaCl
- 1 mM TCEP

Ches/TCEP buffer

- 20 mM Ches-NaOH pH 9.0
- 0.25 M NaCl
- 1 mM TCEP

Reconstituted core complexes were subjected to EM studies. These studies were performed in collaboration with Christian Spahn (Charité, Berlin) and Thorsten Mielke (Max Planck Institute for Molecular Genetics, Berlin).

3.13.1. Core5

Core5 denotes the five subunit TFIIH core subcomplex, consisting of p62, p52, p44, p34, and p8 as described above. In each of the following assemblies, p8 has been added. However, it was not visible in each SDS-PAGE analysis. As p8 stains weakly and often very low protein amounts were used, it was assumed that p8 is still present even though it is not visible. Furthermore, due to its small size it may have been present in the dye front in some of the cases. P8 labels were omitted in the figures where no p8 bands were visible.

3.13.1.1. Assembly

As Core5 could successfully be reconstituted via native PAGE (Figure 3-96), the assembly of Core5 by means of SEC was attempted. 6.4 nmol of each subunit or dual complex, p62nT/p44, p52nT/p8, and p34, were mixed. The mixture was incubated on ice for 1 h and subjected to SEC with Tris/TCEP buffer (Figure 3-106). A stoichiometric complex could be obtained, and peak fractions were pooled and investigated via EM.

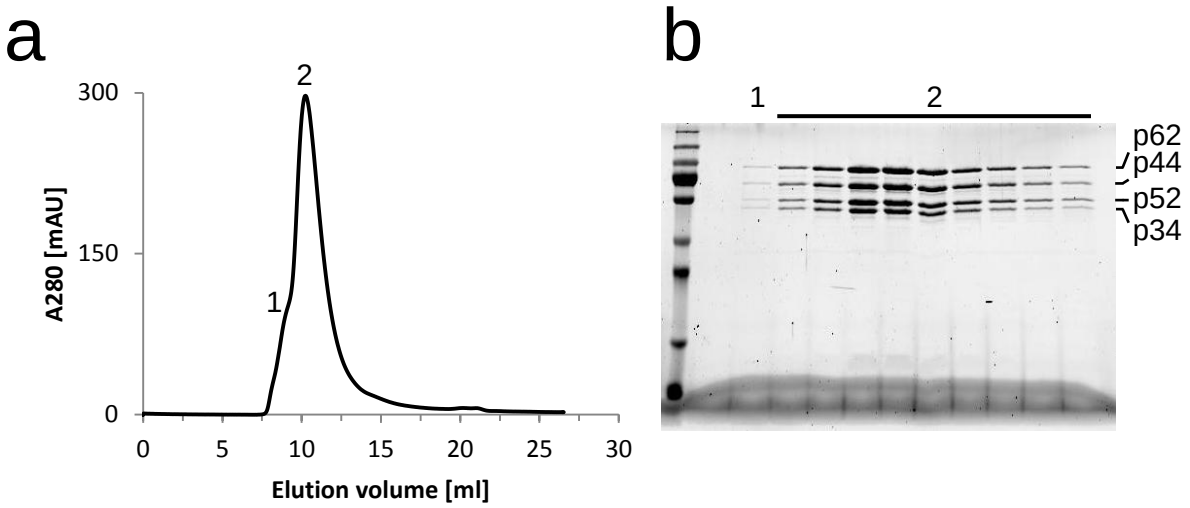


Figure 3-106. Assembly of Core5 via SEC. (a) SEC elution profile from a Superdex 200 10/300 GL column. **(b)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (a).

In order to obtain sufficient material for crystallization attempts, a large scale Core5 assembly was conducted. 58 nmol of each subunit or dual complex, p62nT/p44, p52nT/p8, and p34, were mixed and incubated on ice for 1 h. The mixture was concentrated with a centrifugal filter unit to about 250 µl and applied to a SEC in Tris/TCEP buffer (Figure 3-107). Peak elution fractions were pooled and concentrated to 14.5 mg/ml. Crystallization trials with the commercial crystallization screens using the crystallization robot were performed, but no crystals could be obtained.

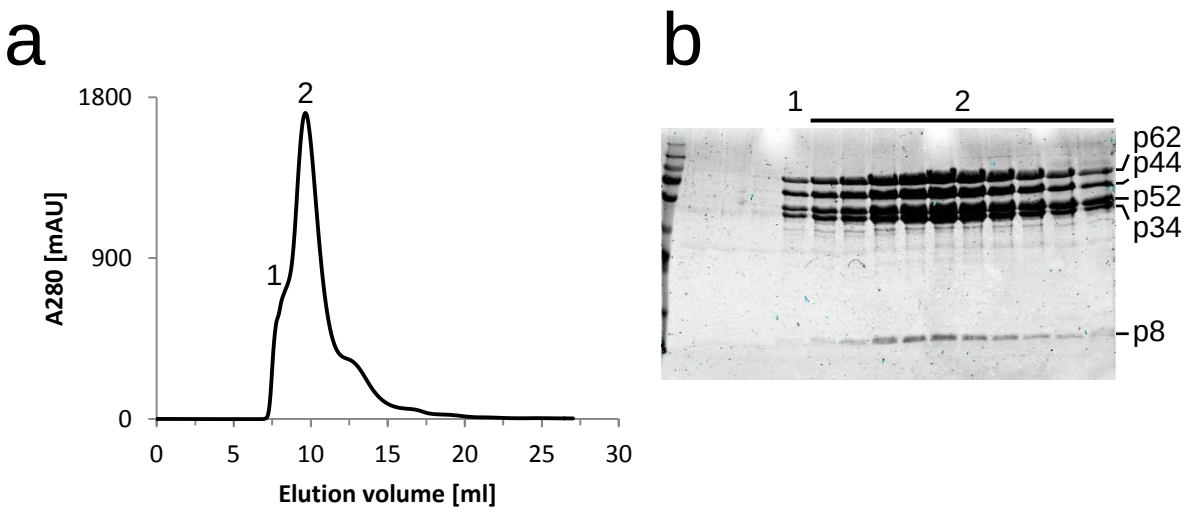


Figure 3-107. Large scale assembly of Core5 via SEC. (a) SEC elution profile from a Superdex 200 10/300 GL column. **(b)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (a).

In an alternative approach, co-purification of Core5 was attempted. P62nT/p44, p52nT/p8, and p34 were expressed in 4 l of LB-medium each. After cell harvest, the cell pellets were combined (17.2 g p62nT/p44 pellet, 15.3 g p52nT/p8 pellet, and 27.8 g p34 pellet) and purification was performed following the standard procedure in the presence of TCEP. The result from co-purification is depicted in Figure 3-108. Due to the quite different expression levels, the purification was dominated by p52nT/p8, impairing formation and isolation of Core5. The buffer system used is given below.

Lysis buffer

- 20 mM Tris-HCl pH 7.5
- 0.3 M NaCl
- 5 mM Imidazole
- 1 mM TCEP

Elution buffer

- 20 mM Tris-HCl pH 7.5
- 0.3 M NaCl
- 0.25 M Imidazole
- 1 mM TCEP

Gel filtration buffer

- 20 mM Tris-HCl pH 7.5
- 0.25 M NaCl
- 1 mM TCEP

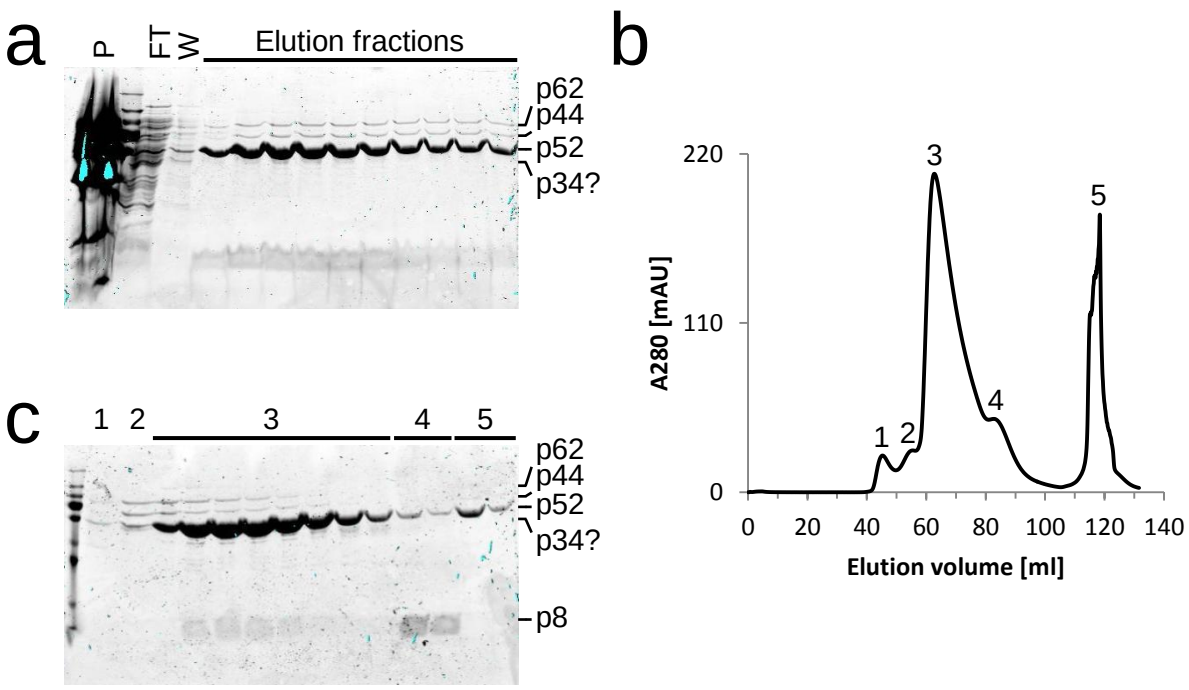


Figure 3-108. Co-purification of Core5 after combination of cell pellets from p62nT/p44, p52nT/p8 and p34. (a) SDS-PAGE analysis of IMAC samples. P: pellet; FT: flow through; W: wash. **(b)** SEC elution profile from a HiLoad 16/600 Superdex 200 pg column. **(c)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (b).

To account for the large molecular weight of Core5 (251 kDa) and a better separation from the void volume, a Superose6 10/300 GL column was used. 7.2 nmol of each subunit or dual complex, p62nT/p44, p52nT/p8, and p34, were mixed, incubated on ice for 1 h, and applied to a Superose 6 10/300 GL column using Tris/TCEP buffer (Figure 3-109 a, b). Peak elution fractions were pooled, concentrated to 100 µl with a centrifugal filter unit, and subjected to another round of SEC (Figure 3-109 c, d). A single peak, containing all subunits could be obtained. P34 however, seemed to be underrepresented.

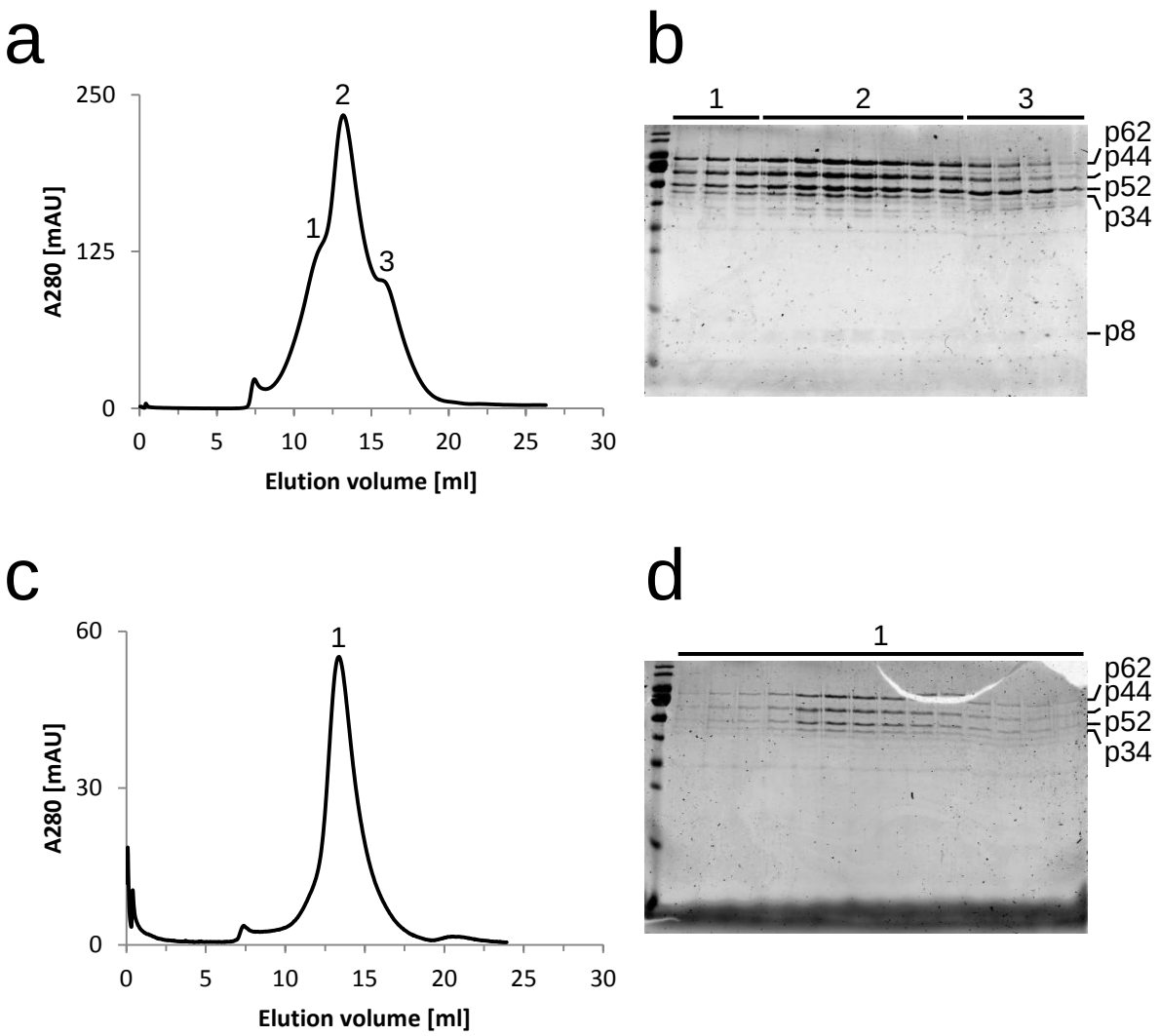


Figure 3-109. Core5 assembly from p62nT/p44, p52nT/p8, and p34 using a Superose 6 10/300 GL column. (a) Elution profile of the first SEC. **(b)** SDS-PAGE analysis of elution fractions from the first SEC. Numbers correspond to peaks in (a). **(c)** Elution profile of the second SEC. Fractions of peak 2 from (a) have been used for the second SEC. **(d)** SDS-PAGE analysis of elution fractions from the second SEC. The number corresponds to the peak in (c).

Since the formation of a p52/p34 complex was established above (see 3.11.3.1), it was possible to reconstitute Core5 via an alternative assembly. Accordingly, p62nT/p44, p52nT/p34, and p8 were used for the assembly. The result of such an alternative assembly is depicted in Figure 3-110. 4.3 nmol of each subunit or dual complex, p62nT/p44, p52nT/p34, and p8, were mixed, incubated on ice for 1 h, and subjected to SEC in Tris/TCEP buffer (Figure 3-110 a, b). Peak fractions were pooled, concentrated to 100 µl with a centrifugal filter unit, and subjected to another round of SEC (Figure 3-110 c, d). A single peak, containing all subunits, was obtained. When directly compared to the prior assembly (Figure 3-109), an improved integration of p34 could be achieved.

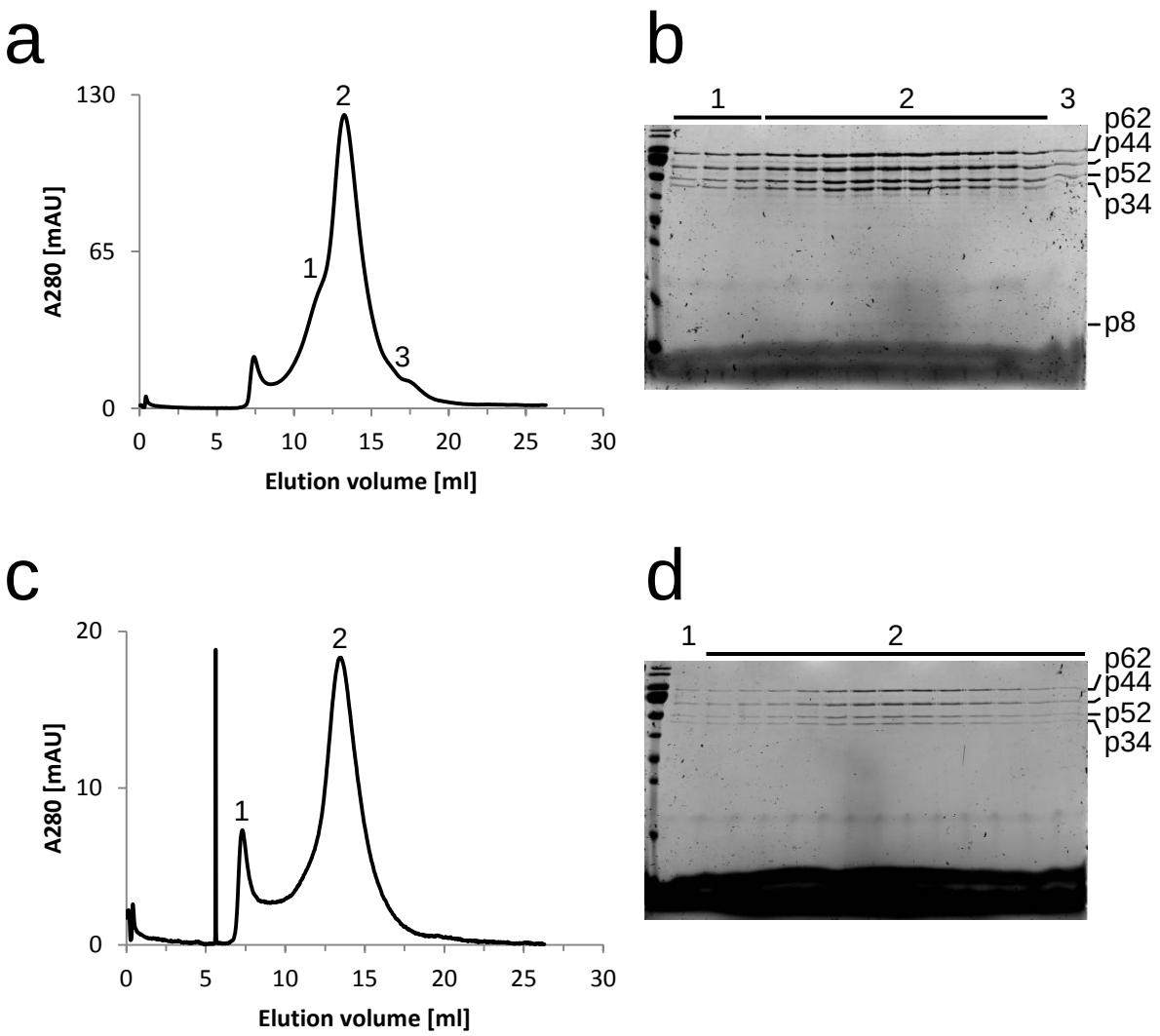


Figure 3-110. Core5 assembly from p62nT/p44, p52nT/p34, and p8 using a Superose 6 10/300 GL column. (a) Elution profile of the first SEC. **(b)** SDS-PAGE analysis of elution fractions from the first SEC. Numbers correspond to peaks in (a). **(c)** Elution profile of the second SEC. Fractions of peak 2 from (a) have been used for the second SEC. **(d)** SDS-PAGE analysis of elution fractions from the second SEC. Numbers correspond to peaks in (c).

EM studies indicated that dilution of the core complexes might be detrimental for complex integrity (Figure 3-139). To circumvent dilution of the sample during SEC and the necessity of concentrating it afterwards, assemblies using a Superose 6 Increase 3.2/300 were pursued. Due to the small bed volume of this column (2.4 ml), dilution of the sample is minimized. An assembly using a Superose 6 Increase 3.2/300 column with Tris/TCEP buffer is depicted in Figure 3-111. 5 nmol of each subunit or dual complex, p62nT/p44, p52nT/p34, and p8, were mixed, incubated on ice for 1 h, and subjected to SEC. A complex with stoichiometric amounts of the subunits could be obtained. The shoulder/peak preceding the main peak seemed to be more pronounced relative to the main peak, when compared to assemblies described above (Figure 3-110 a). Elution fractions from the main peak were

pooled and used for a specific Core7 assembly strategy where preassembled Core5 complex was combined with XPB and XPD (see 3.13.3.1, Figure 3-137).

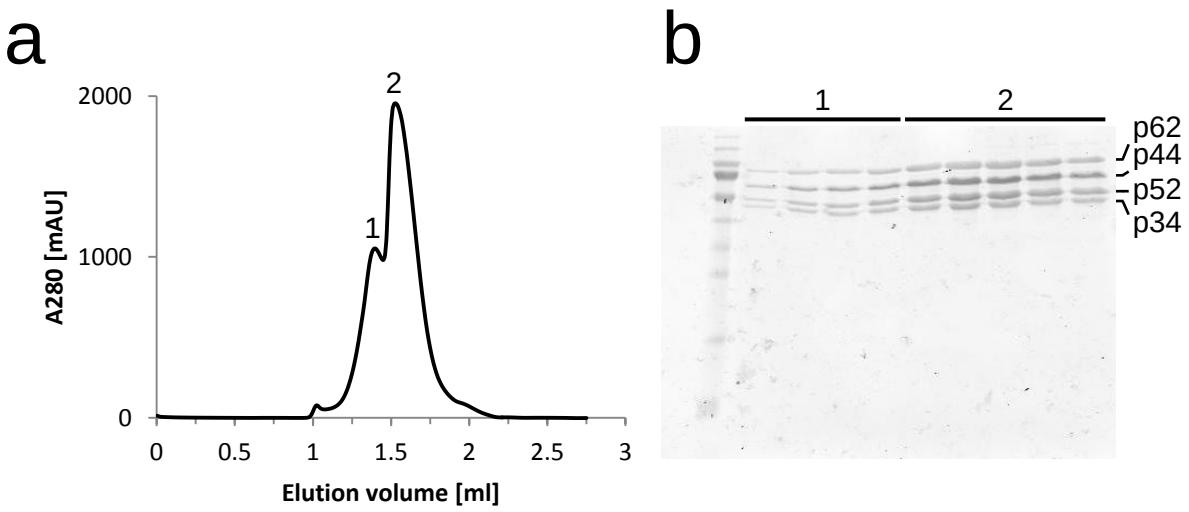


Figure 3-111. Core5 assembly from p62nT/p44, p52nT/p34, and p8 using a Superose 6 Increase 3.2/300 GL column. (a) SEC elution profile. **(b)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (a).

According to native PAGE studies (Figure 3-97) the vWA like domain of p34 was sufficient for the assembly of Core5. Thus, an assembly by means of SEC using p34_1-277 together with p62nT/p44 and p52nT/p8 was attempted (Figure 3-112). Compared to full-length p34 (Figure 3-106), the Core5 formation via p34_1-277 was strongly impaired. This result confirmed the previous finding that the C-terminal zinc finger domain of p34 plays an important role in the TFIIH core interaction network (see 3.12.3). P34_1-277 was provided by Dominik Schmitt.

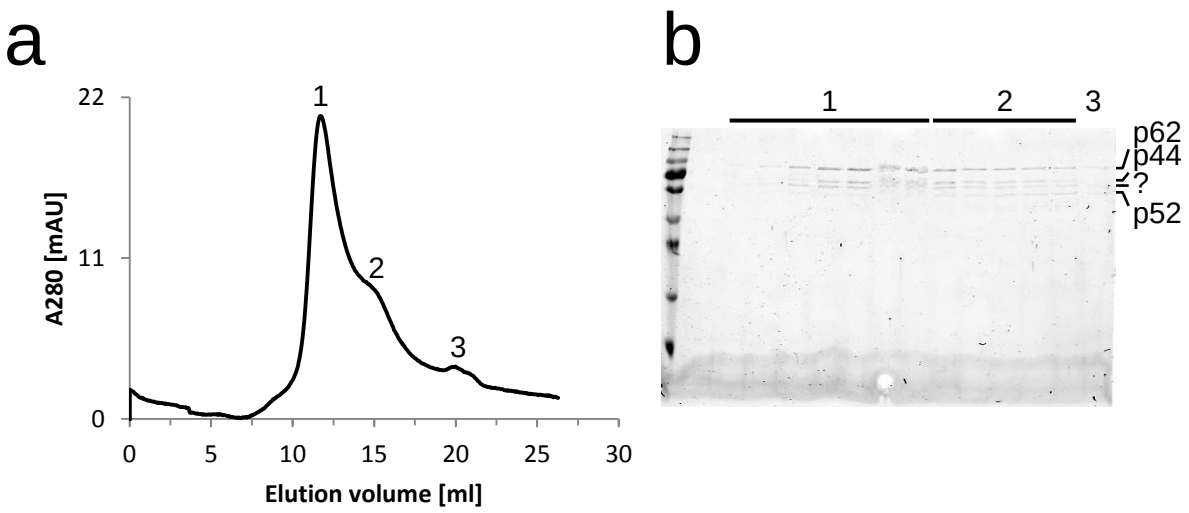


Figure 3-112. Assembly of Core5 using p34_1-277. (a) SEC elution profile from a Superdex 200 10/300 GL column. **(b)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (a).

3.13.1.2. Electron microscopy of Core5

Assembled Core5 (Figure 3-106) was subjected to negative stain experiments using uranyl acetate and imaged via EM (see 2.2.10.1). The stock solution of 0.3 mg/ml Core5 was diluted 1:10 with Tris/TCEP buffer and deposited on the grid. A micrograph is depicted in Figure 3-113, revealing clear particles, with a proper distribution for particle analysis. Particle size corresponds to the expected dimensions of TFIIH core (about 15 nm) [74]. The sample was furthermore imaged via cryo-EM. Here, a 1:8 dilution of the stock solution was deposited. Due to the relatively small size (251 kDa) and the resulting weak contrast, no particles after usual exposure times of 200 ms with a dose of 25 $e^-/Å^2$ could be observed (Figure 3-114 a). However, when increasing the exposure time to 400 ms with a dose of 50 $e^-/Å^2$, particles seem to become visible (Figure 3-114 b).

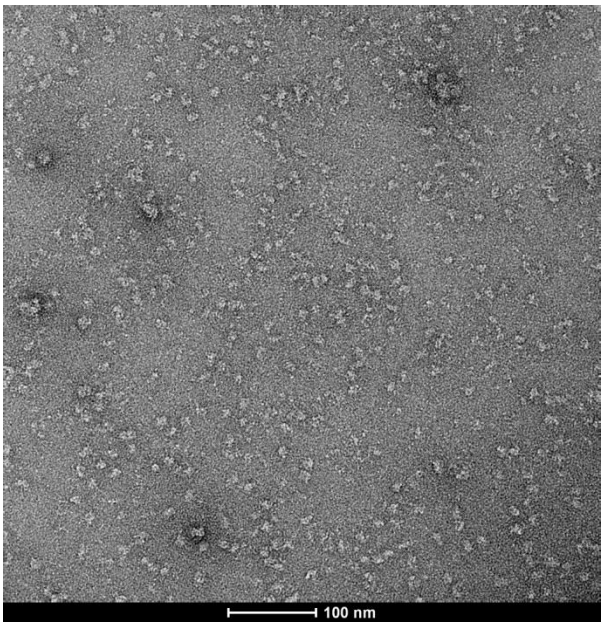


Figure 3-113. Micrograph of negatively stained Core5. A 1:10 dilution of a sample stock of 0.3 mg/ml Core5 was deposited.

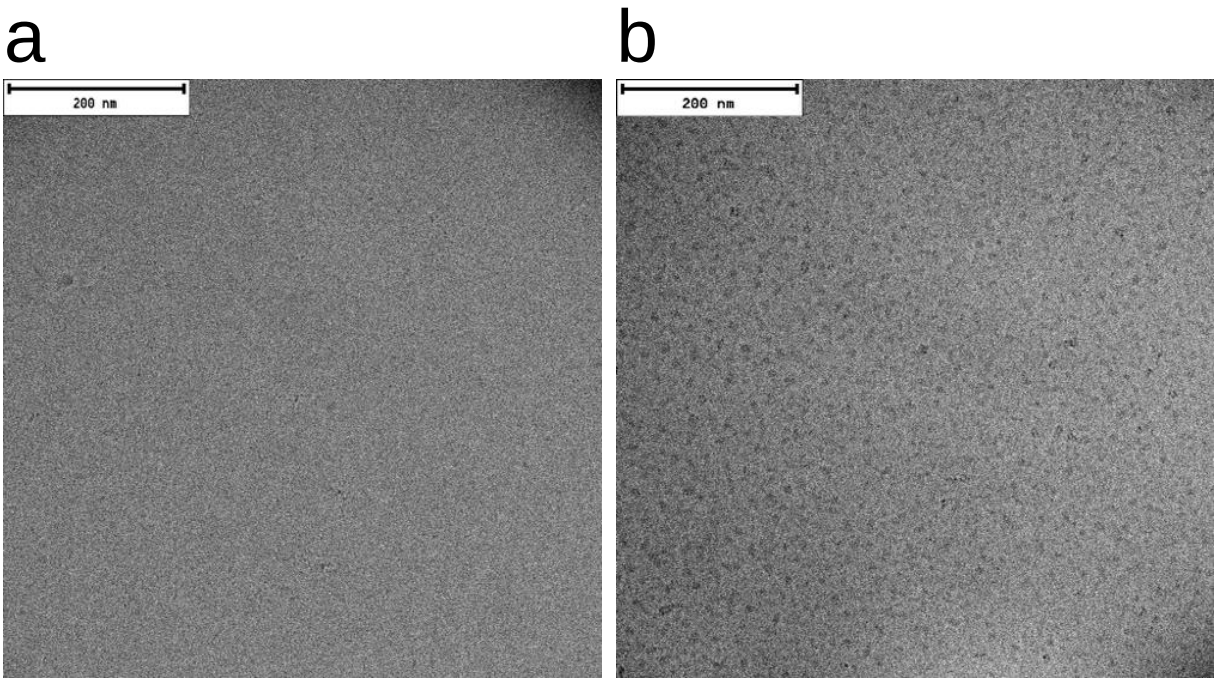


Figure 3-114. Micrographs of vitrified Core5 after different exposure times. A 1:8 dilution of a sample stock of 0.3 mg/ml Core5 was deposited. **(a)** Exposure of 200 ms with a dose of 25 $e^-/Å^2$. **(b)** Exposure of 400 ms with a dose of 50 $e^-/Å^2$.

A dataset of negatively stained Core5 was collected at the Max Planck Institute (MPI) in Berlin. Data processing was performed by Ferdinand Krupp. 96 micrographs were recorded, and 13,164 particles were picked with EMAN2. Grouping of particles into classes was performed with SPARX via the iterative stable alignment and clustering (ISAC) algorithm [483]. This 2D classification resulted in 204 stable classes, which are depicted in Figure 3-115. An initial model was then reconstructed with SPARX. The initial model was

subsequently used for multiparticle refinement [484] against all 13,164 particles with Spider. The final map is depicted in Figure 3-116, with a resolution of about 28 Å. The map displayed a hook-like shape, with two additional weakly attached densities at the top. These two densities might correspond to flexible parts of the complex or originate from heterogeneity in the sample.

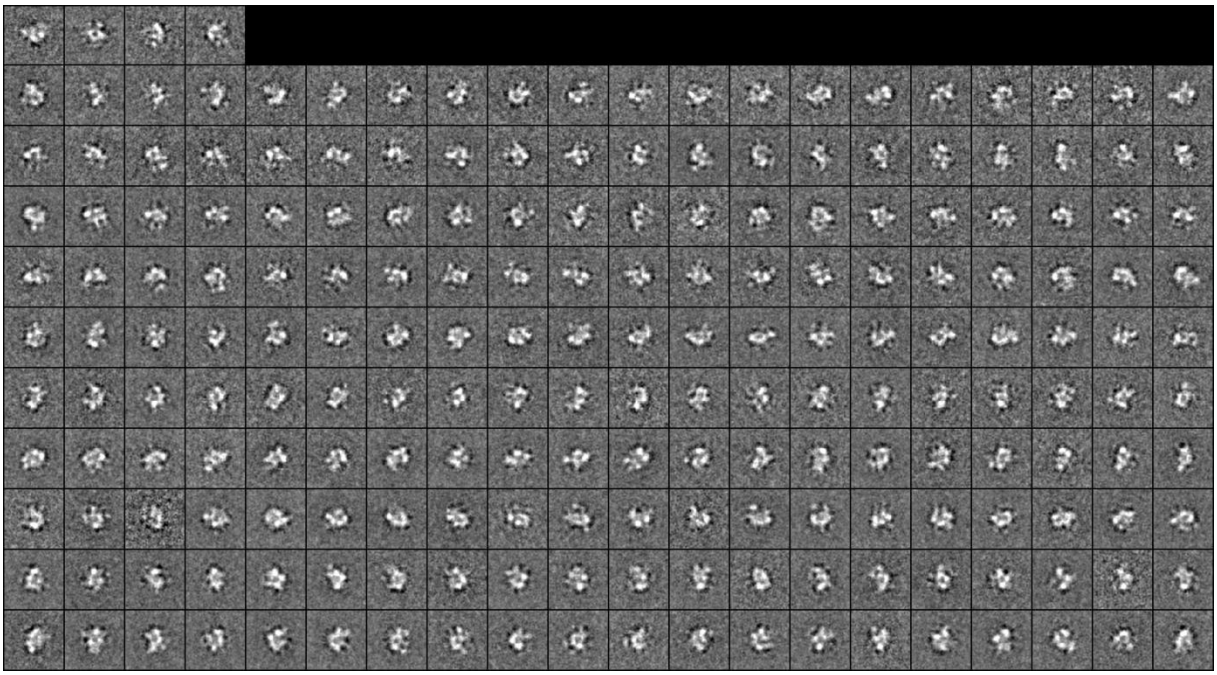


Figure 3-115. Class averages obtained from 2D classification for the negative stain dataset of Core5.

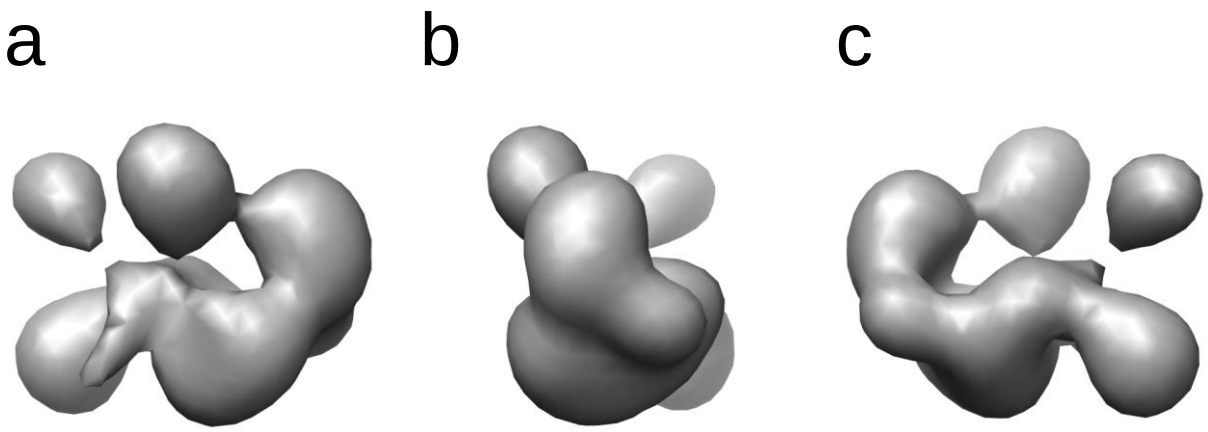


Figure 3-116. Reconstructed 3D map of Core5 from the negative stain dataset. (a) Front view. **(b)** Side view right. **(c)** Back view.

3.13.2. Core6

Core6 denotes a six subunit TFIIH core subcomplex, consisting of XPD, p62, p52, p44, p34, and p8. As already stated above, p8 was not visible via SDS-PAGE due to weak staining or being in the dye front.

3.13.2.1. Assembly

Based on the native PAGE results for Core5 (Figure 3-96), the assembly strategy was extended to include XPD. In a first step the formation of the TFIIH core submodule XPD/p62/p44/p34 was tested via native PAGE (Figure 3-117). A clear shift was observed when XPD was incubated with p62nT/p44, clearly indicating the successful incorporation of XPD.

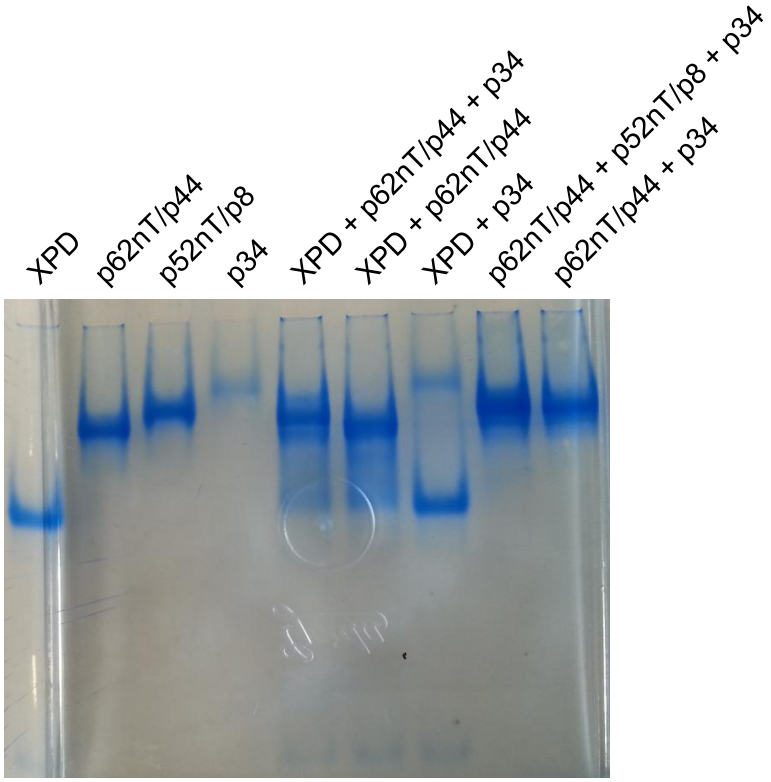


Figure 3-117. Analysis of the formation of the TFIIH core submodule XPD/p62/p44/p34 via native PAGE.

In a second step, the p34 mediated complex formation in the presence and absence of XPD was tested via native PAGE (Figure 3-118). In the presence of XPD the p34 mediated Core5 band appeared to be slightly shifted, indicating the formation of Core6.

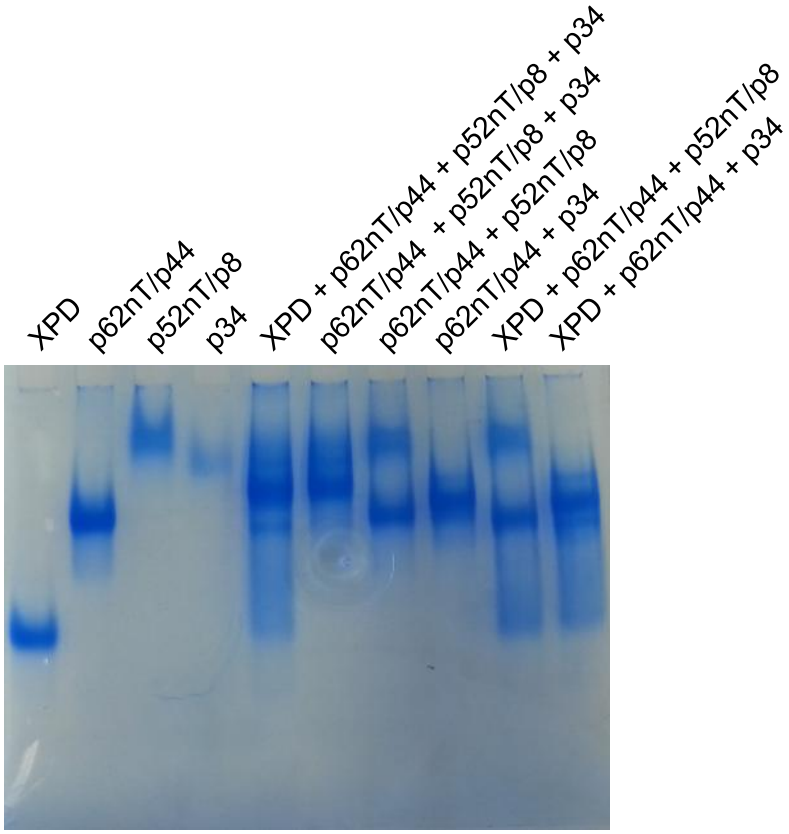


Figure 3-118. Analysis of complex formation between XPD and p34 mediated Core5 via native PAGE.

Following the encouraging results from the native PAGE analysis Core6 assembly via SEC was pursued. 10 nmol of each subunit or dual complex, XPD, p62nT/p44, p52nT/p8, and p34, were mixed. The mixture was incubated on ice for 1 h and applied to a Superdex 200 10/300 GL column using a Hepes/TCEP buffer system. The result from the SEC is depicted in Figure 3-119. A low yield was obtained and the sample eluted as different complexes with different stoichiometries. Elution fractions containing all six subunits were pooled and subjected to EM studies. The Hepes/TCEP buffer system used for the SEC is given below.

Hepes/TCEP buffer

- 20 mM Hepes-NaOH pH 7.5
- 0.15 M NaCl
- 5 mM $MgCl_2$
- 1 mM TCEP

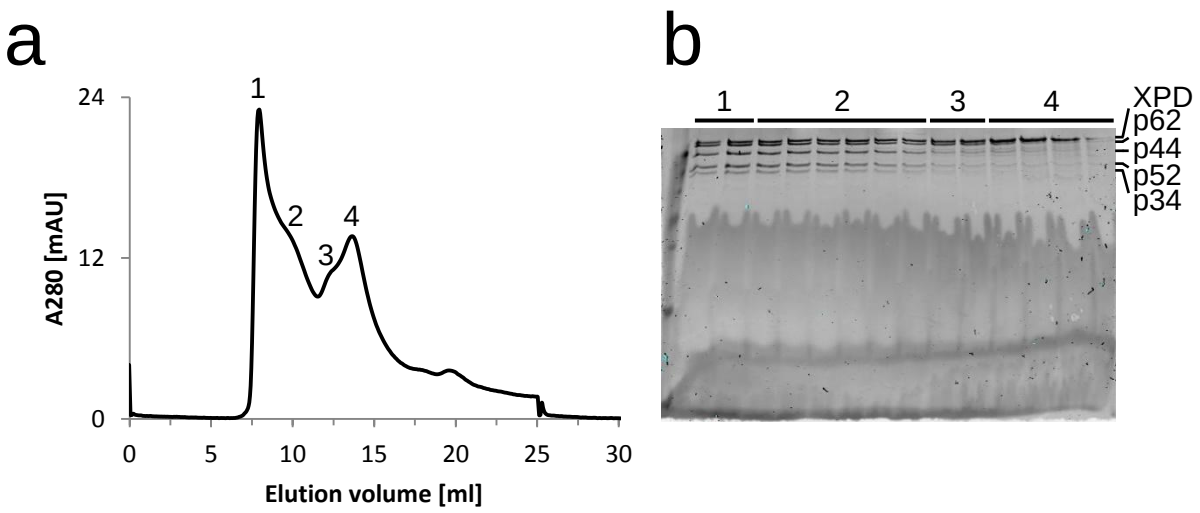


Figure 3-119. Assembly of Core6 via SEC. (a) SEC elution profile from a Superdex 200 10/300 GL column. **(b)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (a).

3.13.2.2. Electron microscopy of Core6

Assembled Core6 (Figure 3-119) was subjected to EM. A stock solution of 0.5 mg/ml Core6 was diluted to about 0.05 mg/ml, deposited, and negatively stained. A micrograph is depicted in Figure 3-120, showing particles with different sizes, indicating heterogeneity and some aggregation.

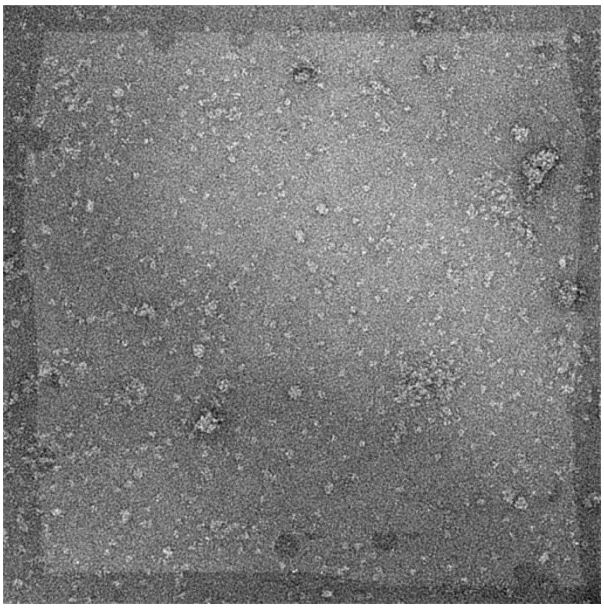


Figure 3-120. Micrograph of negatively stained Core6. The sample was deposited at a concentration of ~0.05 mg/ml.

A dataset of negatively stained Core6 was collected at the MPI in Berlin and processed by Ferdinand Krupp. 409 micrographs were recorded, and 67,806 particles were picked with EMAN2. For 2D classification half of the dataset was used. Particles were grouped with

SPARX via ISAC, resulting in 650 stable classes (Figure 3-121). The 2D classes contained barrel-like contaminations (Figure 3-121, e. g. see 7th row from the bottom, columns 4-12), which were sorted out during multiparticle refinement. An initial model was reconstructed with SPARX and used for multiparticle refinement against all 67,806 particles with Spider. The final map is depicted in Figure 3-122, with a resolution of about 28 Å. The map displayed a stable core, most likely corresponding to Core5. An additional appendix appeared, pointing upwards, presumably corresponding to XPD.

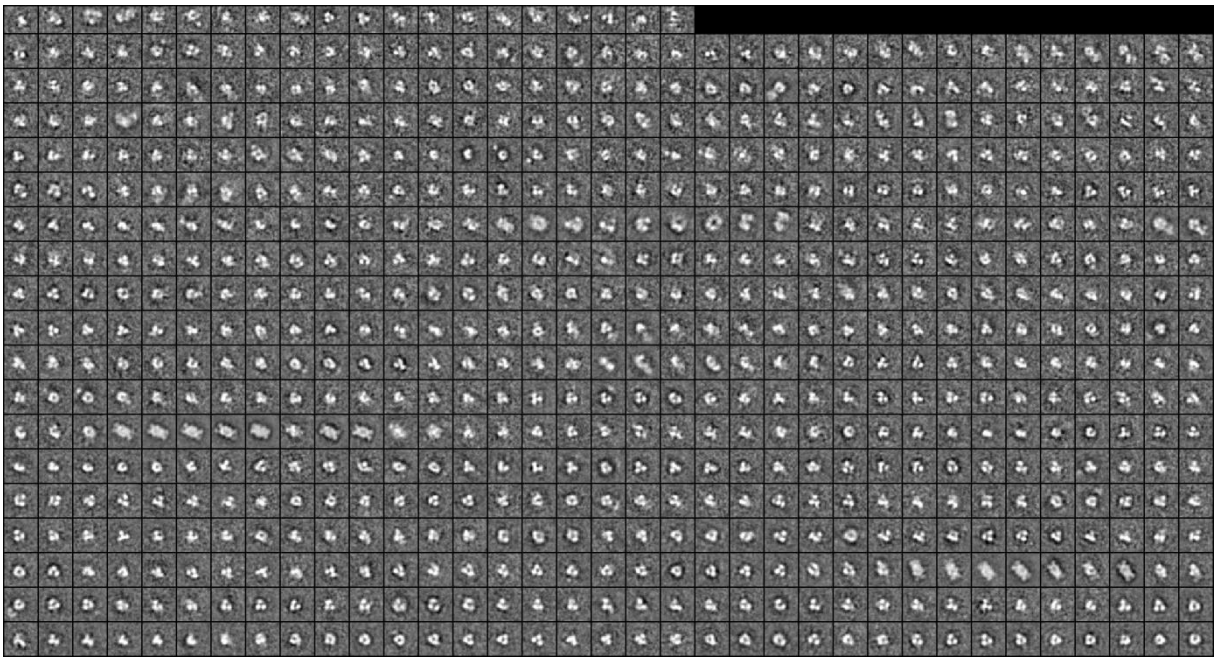


Figure 3-121. Class averages obtained from 2D classification for the negative stain dataset of Core6.

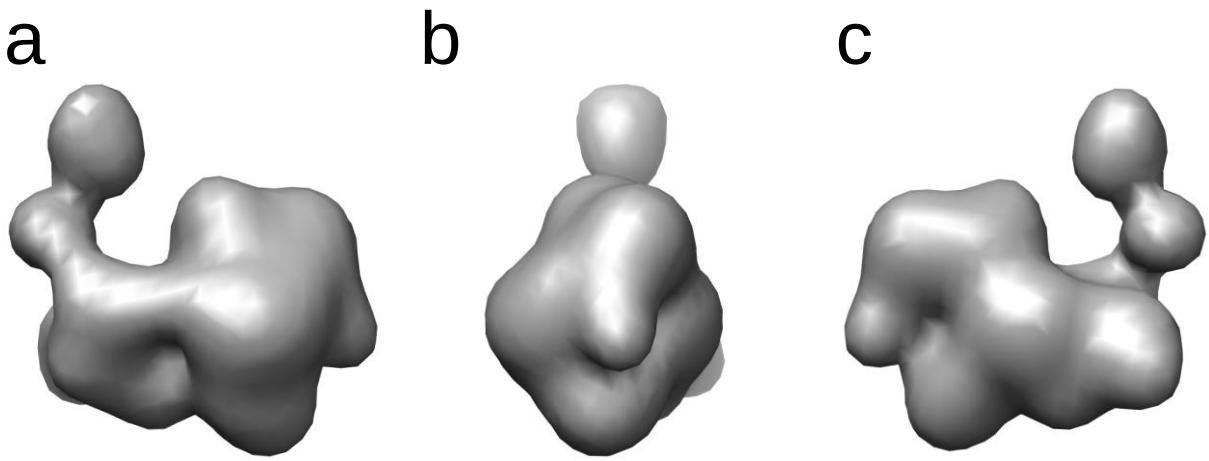


Figure 3-122. Reconstructed 3D map of Core6 from the negative stain dataset. (a) Front view. **(b)** Side view right. **(c)** Back view.

3.13.3. Core7

Core7 denotes the entire seven subunit TFIIH core, consisting of XPB, XPD, p62, p52, p44, p34, and p8. As already stated above, p8 was often not visible via SDS-PAGE due to weak staining or being located in the dye front.

3.13.3.1. Assembly

With the design and successful purification of XPB constructs, initial Core7 assemblies were pursued, utilizing these constructs. Construct XPB_541-E was analyzed for its interaction with p52nT/p8 via native PAGE (Figure 3-123). No clear band shift could be observed. However, the running behavior of the two proteins made the interpretation highly ambiguous, as the bands smeared into each other. Therefore, other XPB constructs were analyzed for initial Core7 assemblies.

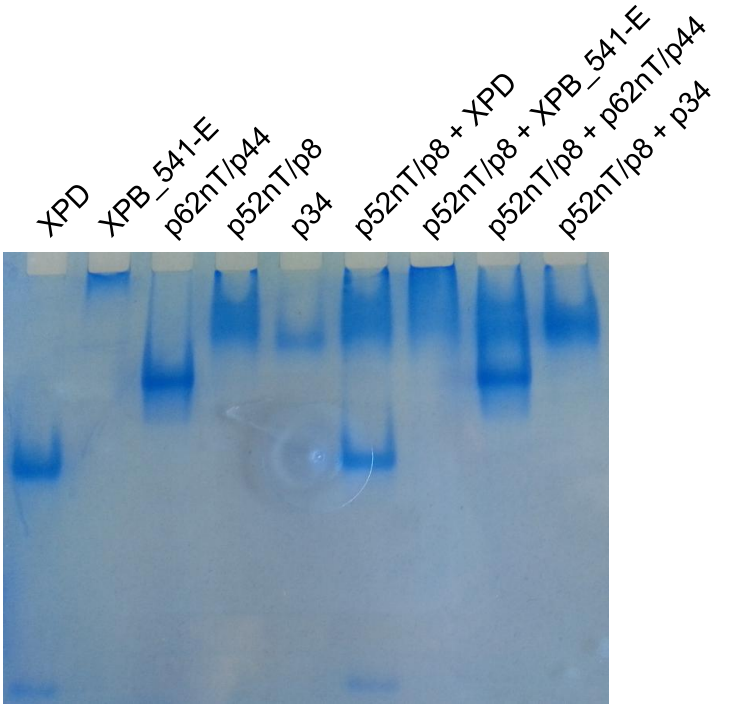


Figure 3-123. Interaction analysis of p52nT/p8 with XPB_541-E and other TFIIH subunits via native PAGE.

As construct XPB_1-345 interacts with p52 (see 3.11.1.2) reconstitution of different core complexes implementing XPB_1-345 was attempted via native PAGE (Figure 3-124). Band shifts of XPB_1-345 and the other subunits could be observed, indicative of successful complex formation. However, the existence of other species made the interpretation uncertain.

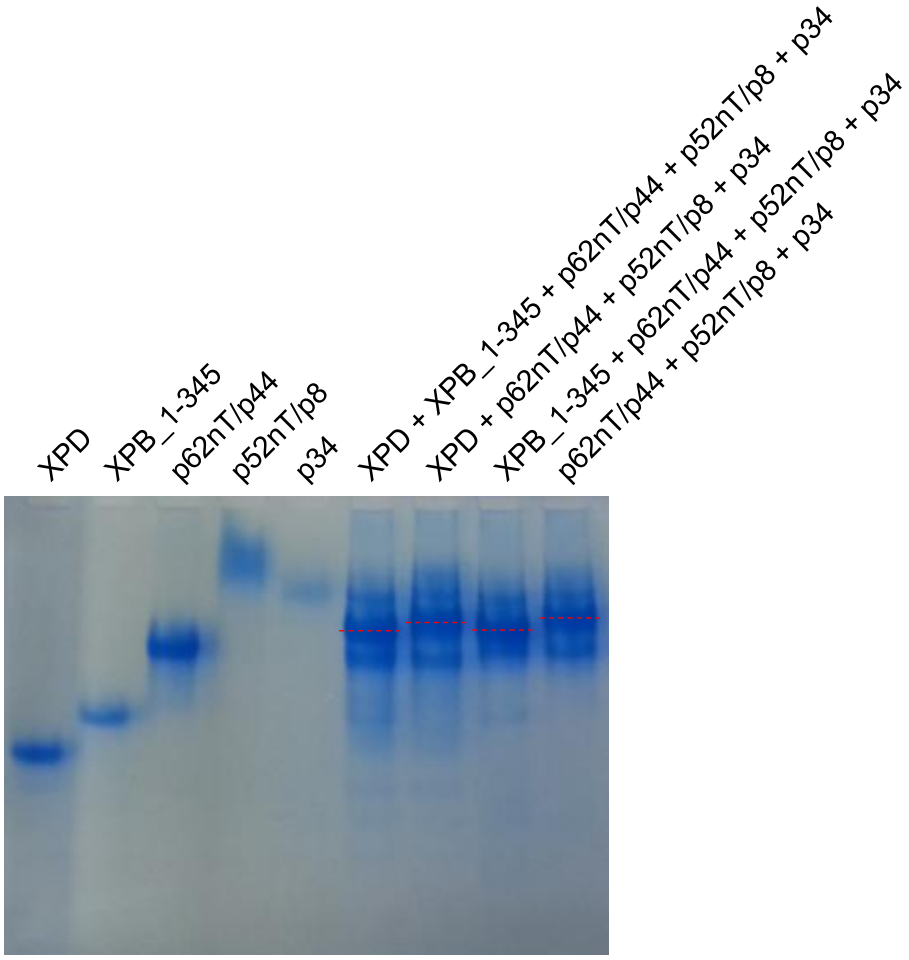


Figure 3-124. Native PAGE analysis of Core7 formation, including the XPB_1-345 construct. The assumed complexes, containing all added subunits, are indicated by dashed red lines.

After the verification of an XPB_116-345/p52nT complex (see 3.11.1.6) further assemblies were pursued. Reconstitution of Core7 via native PAGE using XPB_116-345/p52nT, XPD, p62nT/p44, p34, and p8 was attempted (Figure 3-125). A unique band for the sample containing all seven proteins could be observed, indicative of complex formation.

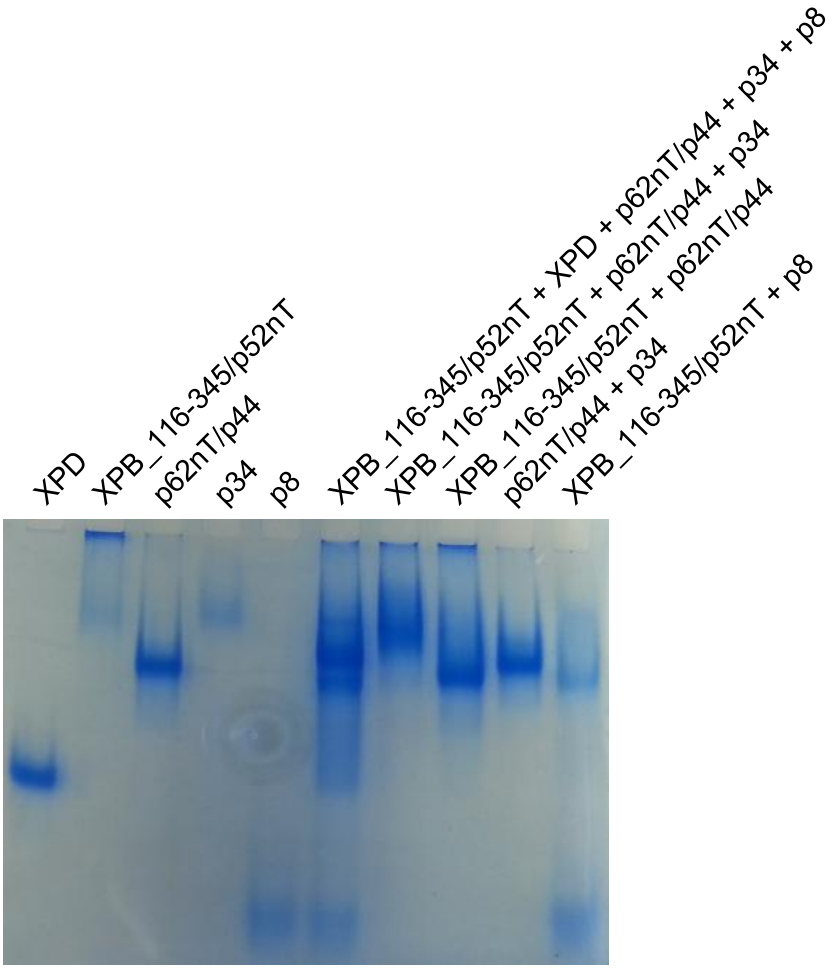


Figure 3-125. Native PAGE analysis of Core7 formation, utilizing the XPB_116-345/p52nT complex.

Reconstitution by SEC was attempted subsequently. 2.8 nmol of each subunit or dual complex, XPB_116-345/p52nT, XPD, p62nT/p44, p34, and p8, were mixed and incubated on ice for 1 h. The mixture was applied to a Superdex 200 10/300 GL column with Tris/TCEP buffer (Figure 3-126). Only a very little amount of protein could be retrieved, but it appeared that all seven proteins eluted together. However, SDS-PAGE analysis only displayed very faint bands, hampering interpretation.

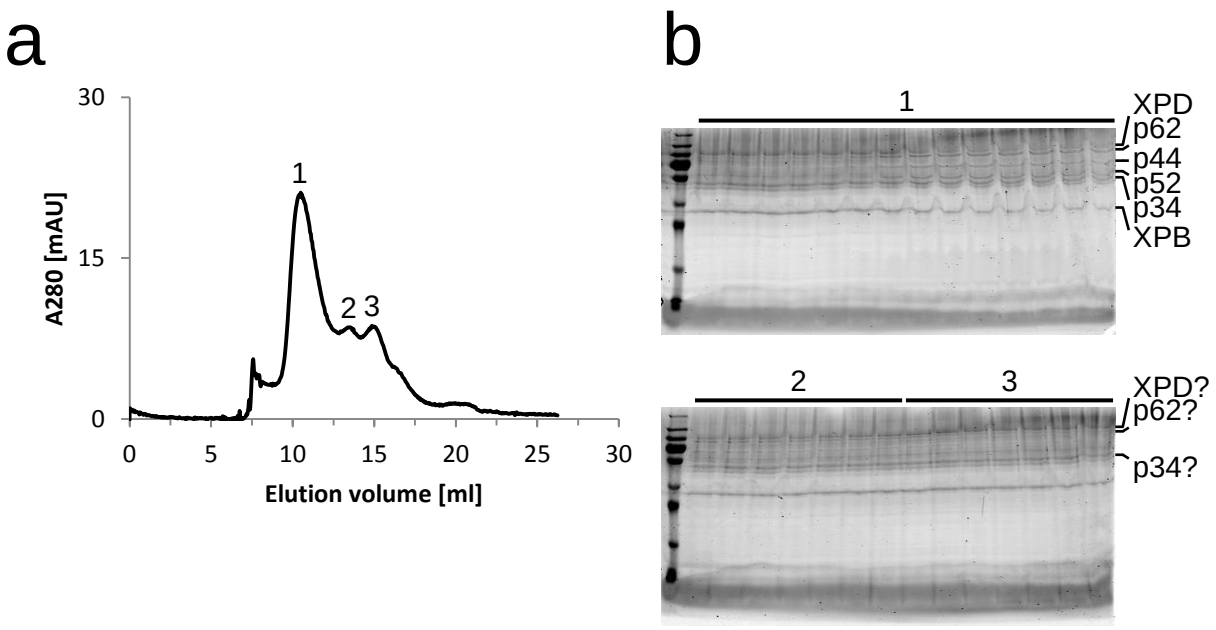


Figure 3-126. Assembly of Core7 via SEC, utilizing XPB_116-345/p52nT. (a) SEC elution profile from a Superdex 200 10/300 GL column. **(b)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (a).

The assembly of a complete Core7, containing the full-length proteins of all core subunits was pursued next. To account for the large molecular weight of Core7 (446 kDa), a Superose 6 10/300 GL column was used for the SEC. Analogous to the p34 mediated Core5 formation (Figure 3-106), p62nT/p44, p52nT/p8, and p34 were combined with XPB and XPD. 3.1 nmol of each subunit or dual complex were mixed, incubated on ice for 1 h, and subjected to SEC in Tris/TCEP buffer (Figure 3-127 a, b). The elution profile displayed three species. The main peak and the preceding slight shoulder contained all subunits, however XPB and p34 seemed to be underrepresented. The main peak elution fractions were pooled, concentrated to 100 µl with a centrifugal filter unit, and subjected to another round of SEC (Figure 3-127 c, d). A single peak containing all subunits could be obtained, however XPB and p34 still seemed to be underrepresented. Peak fractions of the second SEC were pooled, concentrated to about 0.3 mg/ml, and subjected to EM studies.

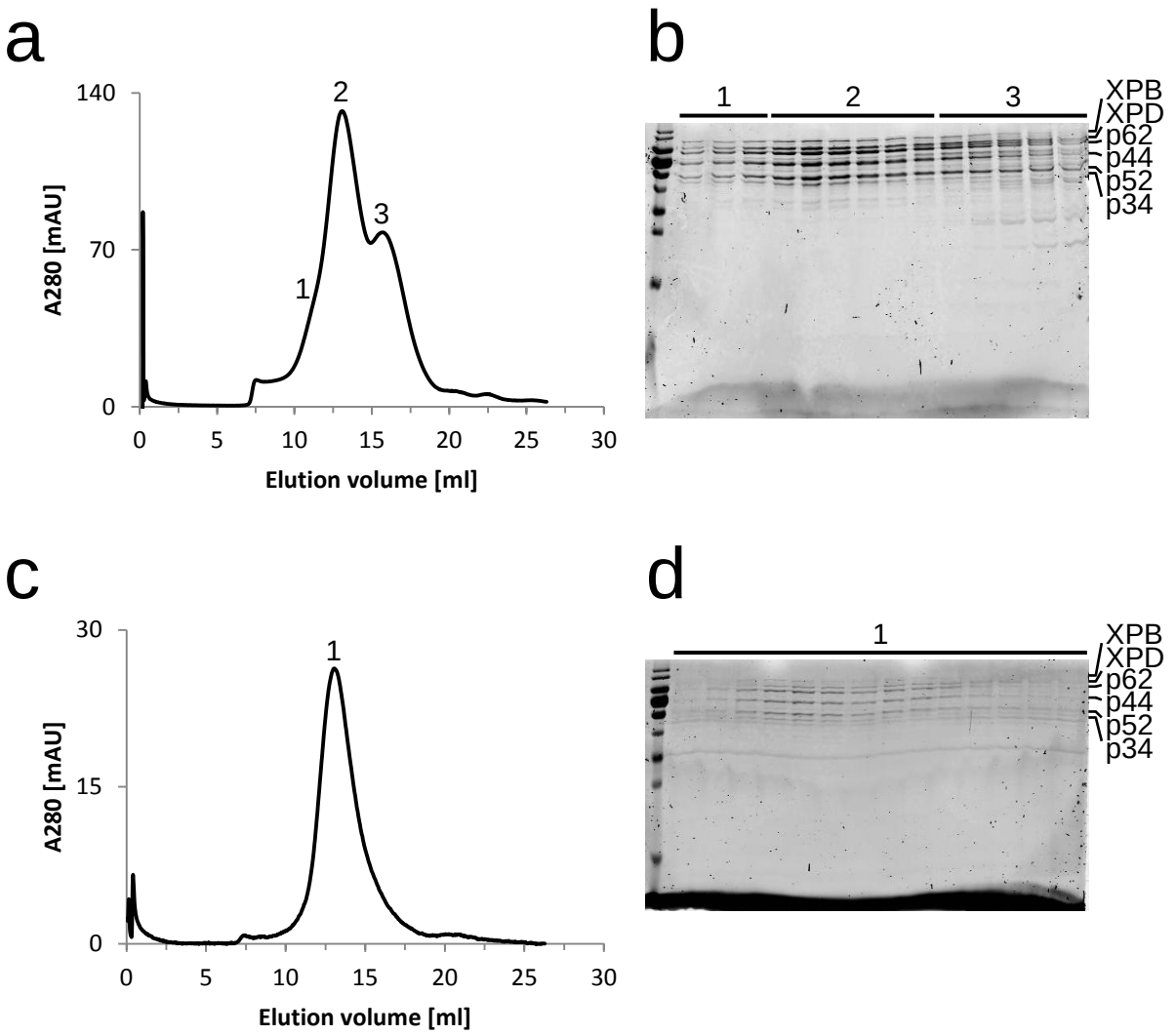


Figure 3-127. Assembly of Core7 via XPB, XPD, p62nT/p44, p52nT/p8, and p34. (a) Elution profile of the first SEC from a Superose 6 10/300 GL column. **(b)** SDS-PAGE analysis of elution fractions from the first SEC. Numbers correspond to peaks in (a). **(c)** Elution profile of the second SEC. Fractions of peak 2 from (a) have been used for the second SEC. **(d)** SDS-PAGE analysis of elution fractions from the second SEC. The number corresponds to the peak in (c).

To overcome the substoichiometric amount of p34 in the Core7 complex, Core7 assembly via formation of the p52nT/p34 complex (see 3.11.3.1) was pursued. Therefore, the dual p52nT/p34 complex was used instead of p52nT/p8, and p8 was added as a single subunit instead of p34. Accordingly, 2.5 nmol of each subunit or dual complex, XPB, XPD, p62nT/p44, p52nT/p34, and p8, were mixed, incubated on ice for 1 h, and subjected to SEC (Figure 3-128 a, b). A similar elution profile as observed before (Figure 3-127) was obtained, but the bands in the SDS-PAGE appeared to be more stoichiometric, with better incorporation of XPB and p34. The main peak fractions were pooled, concentrated to 100 µl, and subjected to another SEC (Figure 3-128 c, d). A single peak could be obtained, seemingly containing all subunits. However, verification via SDS-PAGE was hampered due to the presence of only very faint bands.

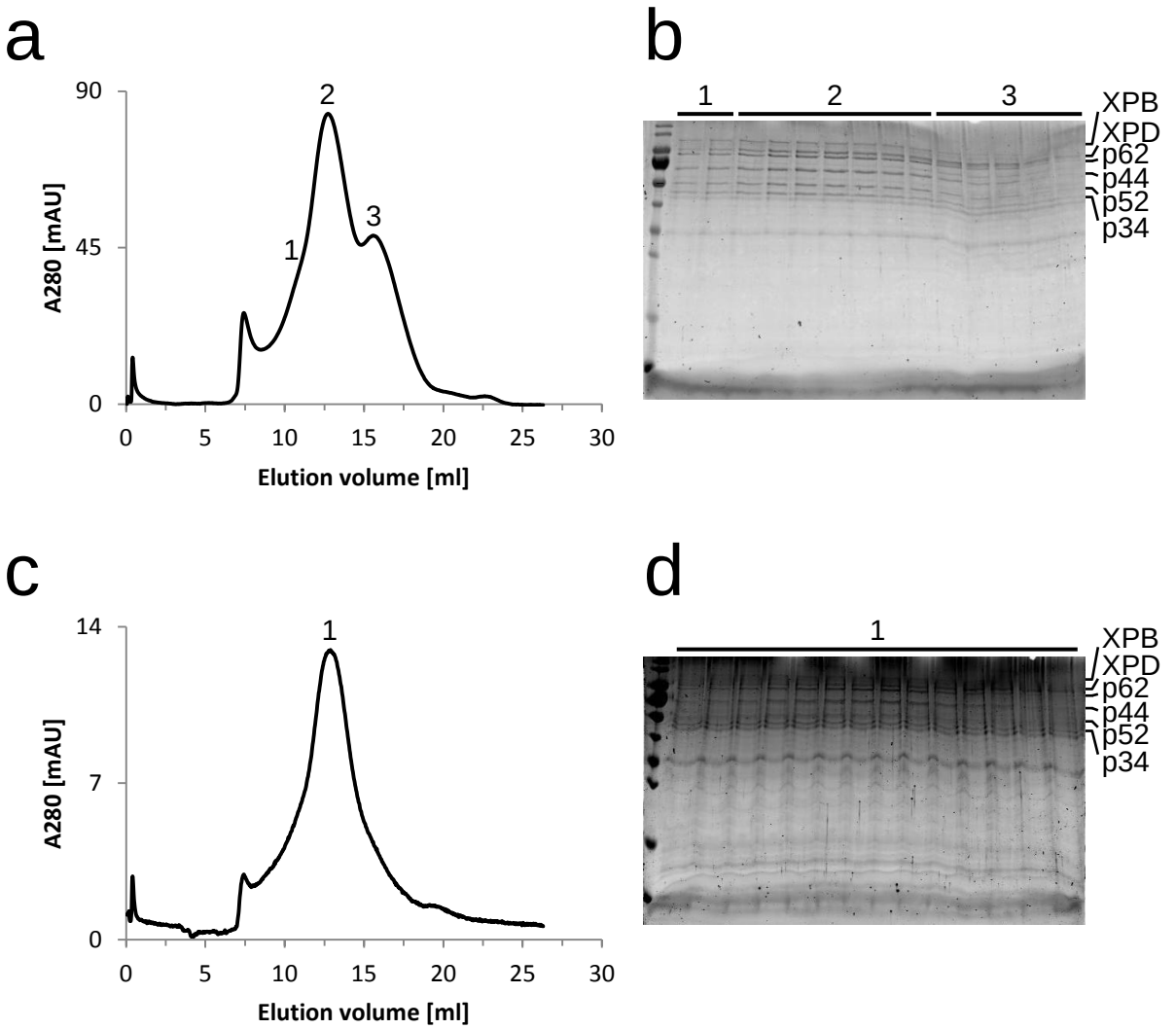


Figure 3-128. Assembly of Core7 via XPB, XPD, p62nT/p44, p52nT/p34, and p8. (a) Elution profile of the first SEC from a Superose 6 10/300 GL column. **(b)** SDS-PAGE analysis of elution fractions from the first SEC. Numbers correspond to peaks in (a). **(c)** Elution profile of the second SEC. Fractions of peak 2 from (a) have been used for the second SEC. **(d)** SDS-PAGE analysis of elution fractions from the second SEC. The number corresponds to the peak in (c).

The influence of a decreased salt concentration on Core7 assembly was analyzed since most enzymatic studies such as ATPase and helicase assays are performed at low mM salt concentrations [22, 100, 184, 186, 485]. 3.1 nmol of each subunit or dual complex, XPB, XPD, p62nT/p44, p52nT/p8, and p34, were mixed and incubated on ice for 1 h. The mixture was subjected to SEC using a buffer with half the NaCl concentration compared to previous purifications (Figure 3-129 a, b). The low salt concentration seems to be detrimental, as the yield of assembled Core7 decreased in comparison to the species of the misassembled complex. Elution fractions of the peak containing all subunits were pooled and concentrated to 100 µl. After applying to a second SEC, no complex could be retrieved any more (Figure 3-129 c). Furthermore, an assembly in the presence of 0.15 M potassium acetate (KAc) was

conducted. Here, 3.3 nmol of each subunit or dual complex, XPB, XPD, p62nT/p44, p52nT/p34, and p8, were mixed and incubated on ice for 1 h. SEC was performed in a Hepes/KAc buffer system (Figure 3-130 a). A similar result to the assembly in the low NaCl buffer was obtained. Peak fractions were again pooled, concentrated, and subjected to a second SEC (Figure 3-130 b). No complex could be retrieved after the second SEC. The composition of the low salt buffer and the Hepes/KAc buffer is given below.

Low salt buffer

- 20 mM Tris-HCl pH 7.5
- 0.125 M NaCl
- 1 mM TCEP

Hepes/KAc buffer

- 20 mM Hepes pH 7.6
- 0.15 M KAc
- 5 % (v/v) glycerol

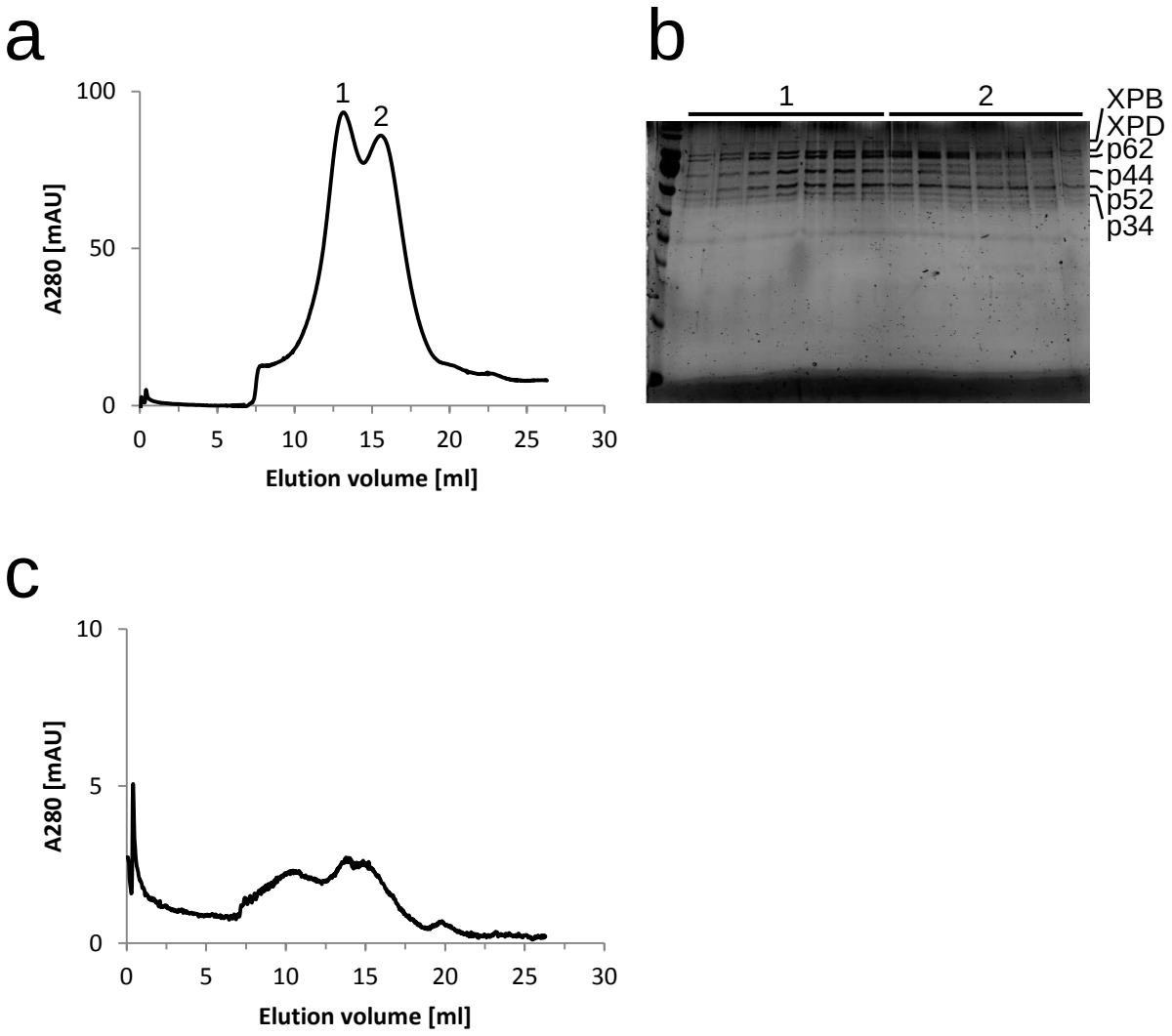


Figure 3-129. Assembly of Core7 in a low salt condition. A buffer containing 0.125 M NaCl has been used. **(a)** Elution profile of the first SEC from a Superose 6 10/300 GL column. **(b)** SDS-PAGE analysis of elution fractions from the first SEC. Numbers correspond to peaks in (a). **(c)** Elution profile of the second SEC. Fractions of peak 1 from (a) have been used for the second SEC.

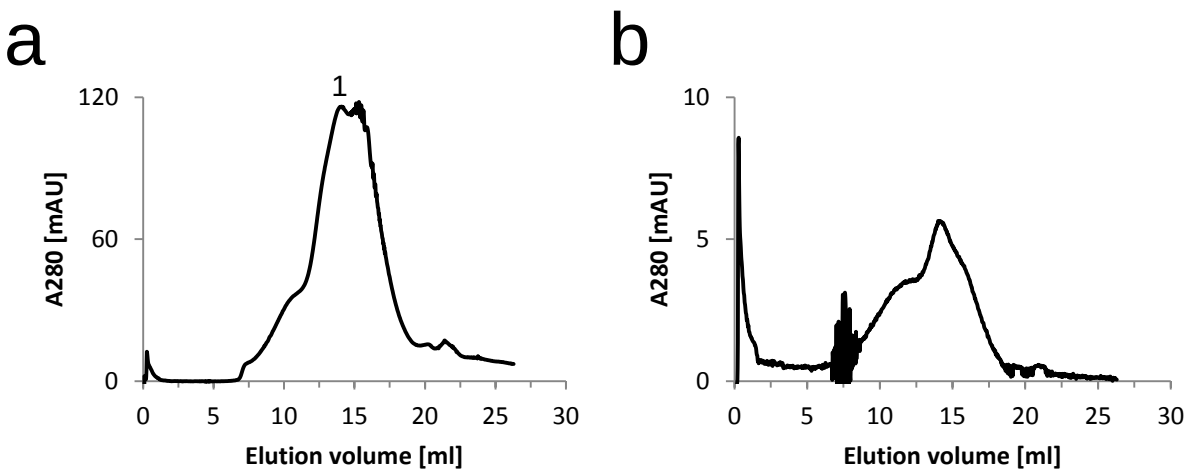


Figure 3-130. Assembly of Core7 in a low potassium acetate condition. A buffer containing 0.15 M KAc has been used. **(a)** Elution profile of the first SEC from a Superose 6 10/300 GL column. **(b)** Elution profile of the second SEC. Fractions of peak 1 from (a) have been used for the second SEC.

As already stated for Core5, dilution of the core complexes might be detrimental for complex integrity (Figure 3-139). Therefore, the assembly of Core7 using a Superose 6 Increase 3.2/300 column was established. The small bed volume of this column strongly decreased dilution during the SEC. Furthermore, this column made it feasible to work with small protein amounts during Core7 assemblies. A Core7 assembly using a Superose 6 Increase 3.2/300 column is depicted in Figure 3-131. 3.3 nmol of each subunit or dual complex, XPB, XPD, p62nT/p44, p52nT/p34, and p8, were mixed and incubated on ice for 1 h. The mixture was concentrated to 50 µl with a centrifugal filter unit and subjected to SEC in Tris/TCEP buffer (Figure 3-131 a, b). A similar elution profile as described above (Figure 3-128 a) was obtained. Main peak fractions were pooled, concentrated to 50 µl, and subjected to a second SEC (Figure 3-131 c, d). The elution profile displayed two peaks. The main peak did not elute at the volume of the main peak from the first SEC, but at the volume of the preceding shoulder from the first SEC. The second smaller peak corresponded to the main peak from the first SEC. A possible cause for this behavior might be the concentrating step between the first and second SEC, which became necessary due to the smaller injection volume for the Superose 6 Increase 3.2/300 column. This possibility was considered and investigated in the following assemblies.

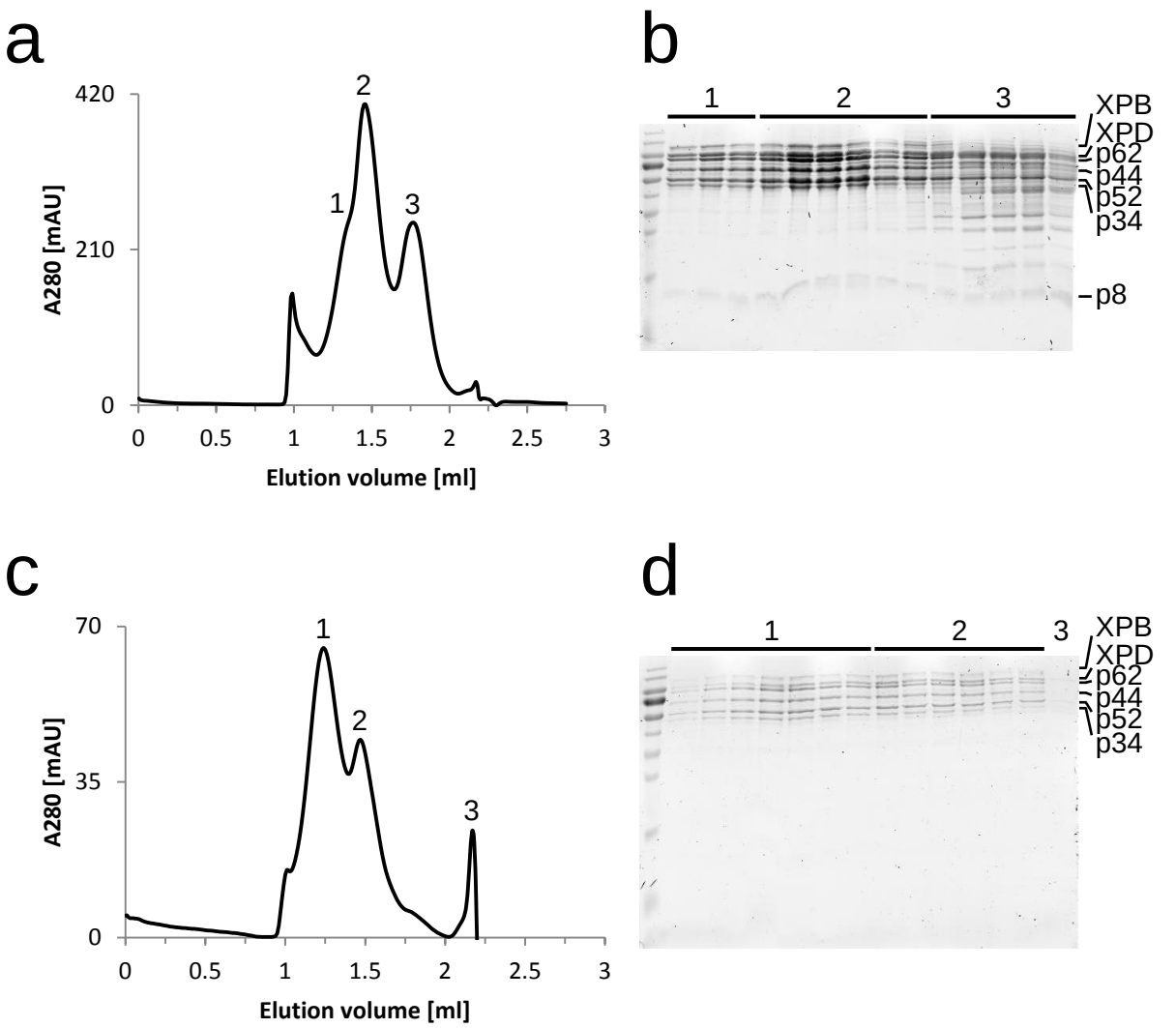


Figure 3-131. Assembly of Core7 using a Superose 6 3.2/300 column. (a) Elution profile of the first SEC. **(b)** SDS-PAGE analysis of elution fractions from the first SEC. Numbers correspond to peaks in (a). **(c)** Elution profile of the second SEC. Fractions of peak 2 from (a) have been used for the second SEC. **(d)** SDS-PAGE analysis of elution fractions from the second SEC. Numbers correspond to peaks in (c).

To assess the effect of concentrating the core TFIIH sample and to circumvent the necessity of a concentration step prior to the second SEC, the assembly strategy was modified. After the first SEC, only a single elution fraction from the peak center was subjected to a second SEC without a concentrating step. Fraction sizes were chosen to be the same as the injection volume for the SEC column (50 µl). To account for the loss of yield during the first SEC and to obtain a sufficient amount of protein in a single elution fraction for the second SEC, the amount of starting material used was increased. However, as this also increased the starting volume, an increased first concentrating step was necessary. Accordingly, 6.5 nmol of each subunit or dual complex, XPB, XPD, p62nT/p44, p52nT/p34, and p8, were mixed and incubated on ice for 1 h. The mixture was concentrated to 50 µl with a centrifugal filter unit and subjected to SEC in Tris/TCEP buffer (Figure 3-132 a, b). Based on the elution

profile four peaks/shoulders could be distinguished, which will be referred to as species in the following. According to the elution volume, species 1, 2, and 4 have been observed before (Figure 3-131 a). Species 3 seemed to be unique for this assembly strategy. Species 1-3 contained all core subunits in stoichiometric amounts, whereas species 4 seemed to be less stoichiometric (Figure 3-132 b). A possible cause for this altered elution profile might be the more pronounced first concentrating step. From each of the four species, the central elution fraction was subjected to a separate SEC (Figure 3-132 c-f). Seemingly, species 1, 2, and 4 were stable and distinct entities, as the main peak of each SEC reappeared at the same elution volume. An exception was species 3, as it reappeared at the elution volume of species 2. The peak fractions of all four SECs were pooled separately and subjected to EM studies.

To explore the effect of a different pH value on Core7 reconstitution, an assembly at a more basic pH of 9 was conducted. The same procedure as described above was applied, using a starting amount of 6.5 nmol for each subunit or dual complex. SEC was then conducted in a basic Ches/TCEP buffer system. Similar to the assembly in Tris/TCEP buffer (Figure 3-132 a), four species were obtained (Figure 3-133 a). Again, all four species were subjected to a second SEC (Figure 3-133 b-e). As described for the Tris/TCEP buffer, the main peak of all species reappeared at the same elution volume, except for species 3. The peak fractions of all four SECs were pooled separately and subjected to EM studies.

The same assembly strategy was repeated in Tris/TCEP buffer again, using the exact same protein batches as for the assembly in Ches/TCEP buffer. In this case, only species 3 was subjected to a second SEC. For better comparison of the two buffer systems, the result is depicted in Figure 3-134, underlining the highly reproducible appearance of the four species with this assembly strategy.

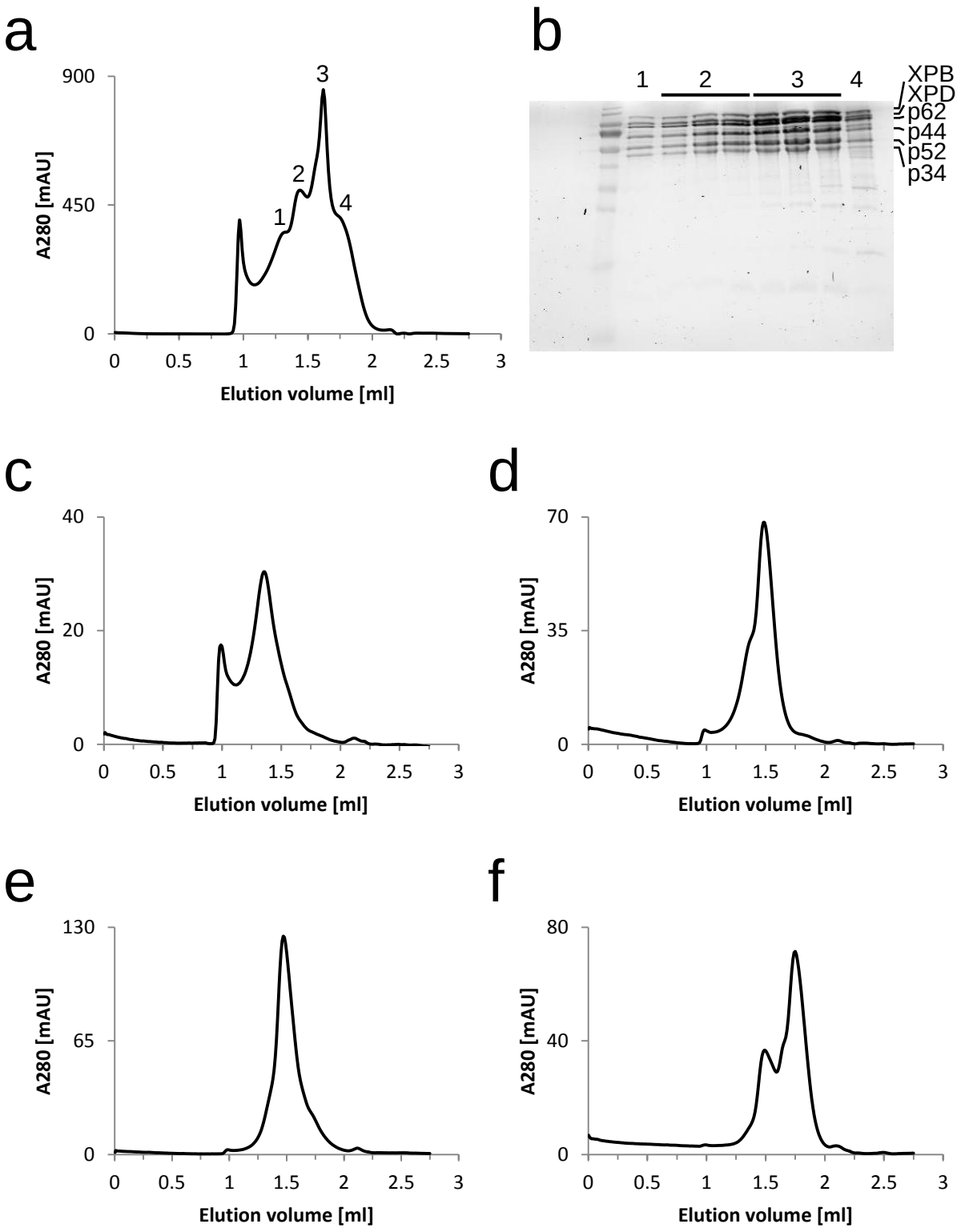


Figure 3-132. Assembly of Core7 at a starting amount of 6.5 nmol, using the Tris/TCEP buffer. **(a)** Elution profile of the first SEC from a Superose 6 Increase 3.2/300 column. **(b)** SDS-PAGE analysis of elution fractions from the first SEC. Numbers correspond to peaks in (a). **(c)** Elution profile of species 1 from (a), subjected to a second SEC. **(d)** Elution profile of species 2 from (a), subjected to a second SEC. **(e)** Elution profile of species 3 from (a), subjected to a second SEC. **(f)** Elution profile of species 4 from (a), subjected to a second SEC.

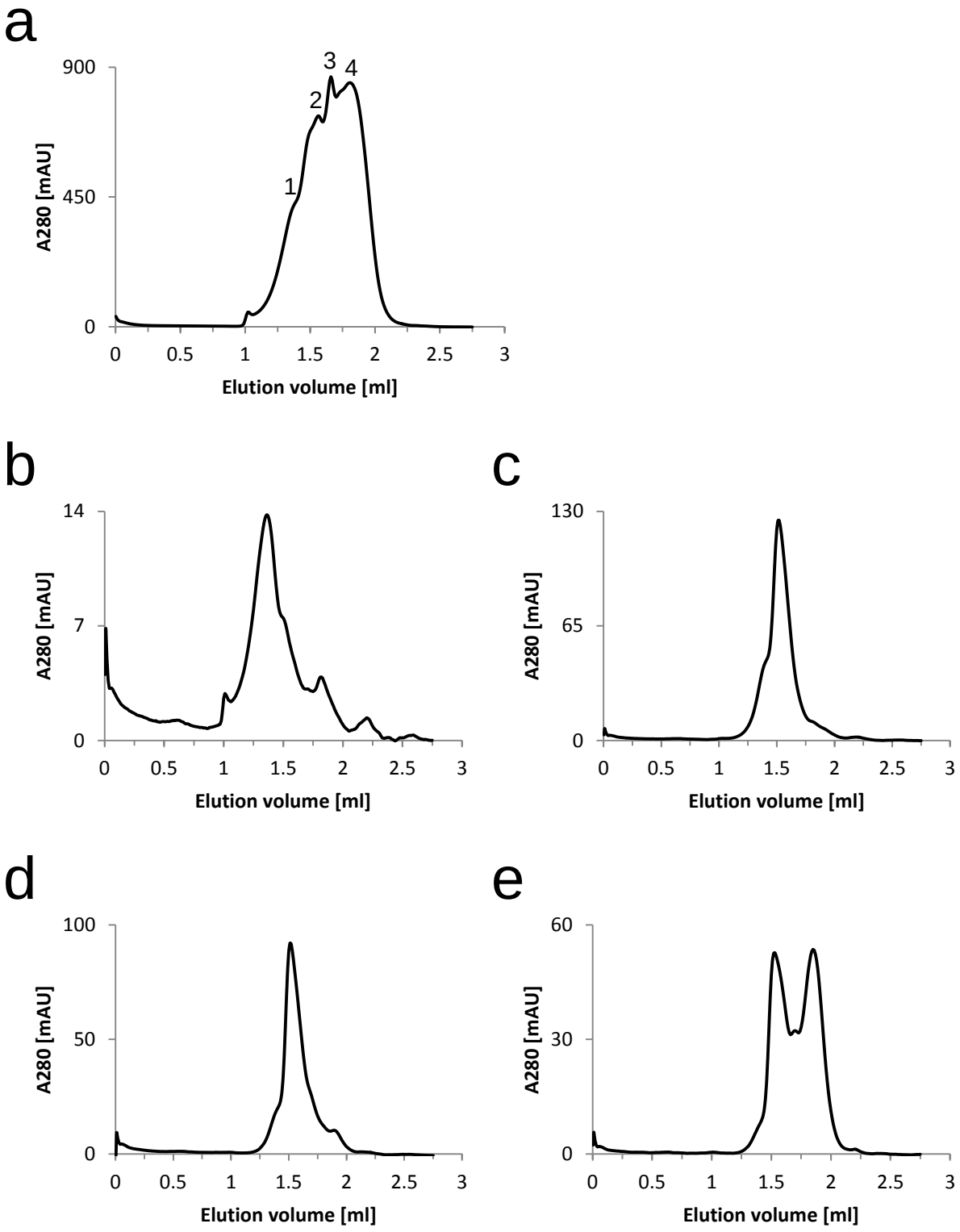


Figure 3-133. Assembly of Core7 at a starting amount of 6.5 nmol, using the Ches/TCEP buffer. **(a)** Elution profile of the first SEC from a Superose 6 Increase 3.2/300 column. **(b)** Elution profile of species 1 from (a), subjected to a second SEC. **(c)** Elution profile of species 2 from (a), subjected to a second SEC. **(d)** Elution profile of species 3 from (a), subjected to a second SEC. **(e)** Elution profile of species 4 from (a), subjected to a second SEC.

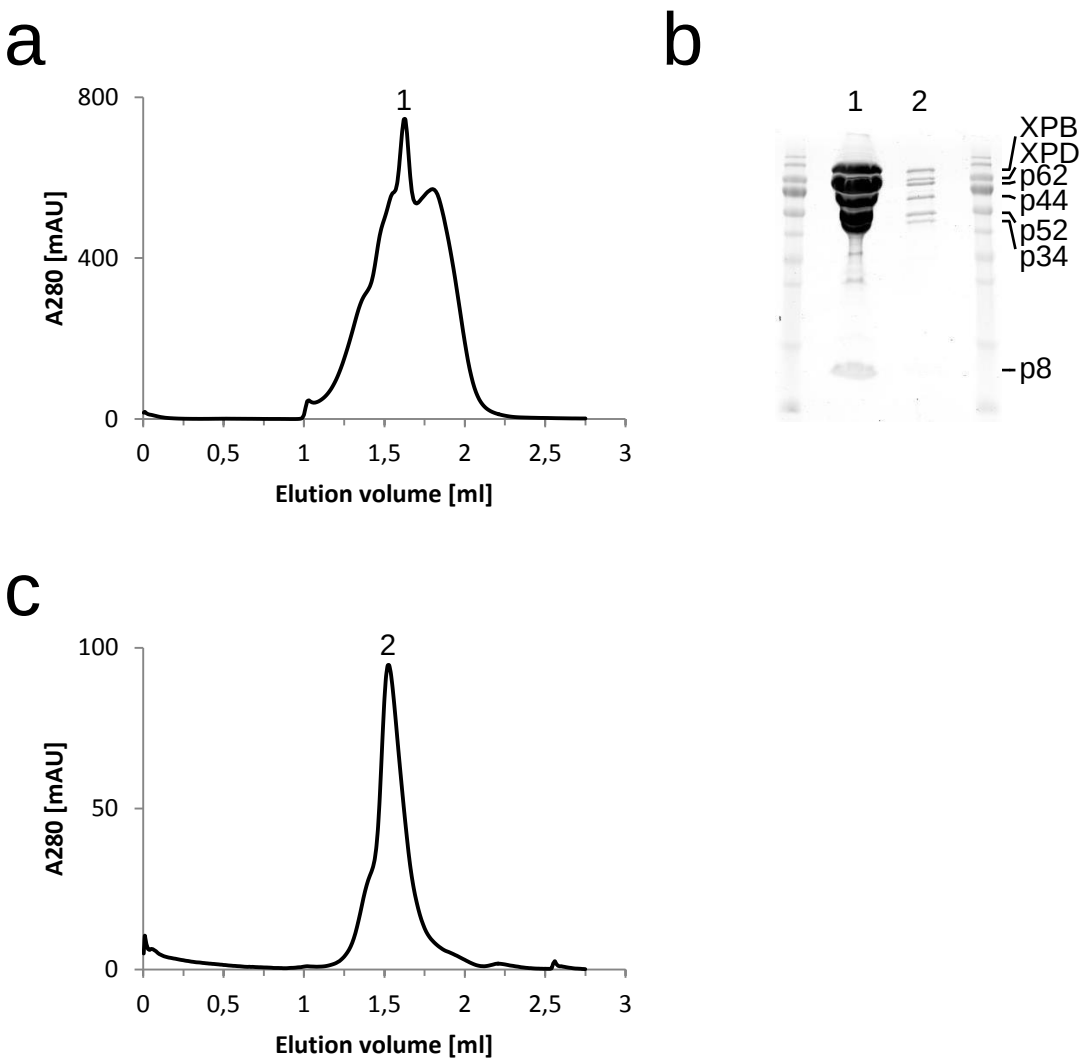


Figure 3-134. Repetition of the Core7 assembly at a starting amount of 6.5 nmol with Tris/TCEP buffer. (a) Elution profile of the first SEC from a Superose 6 Increase 3.2/300 column. **(b)** SDS-PAGE analysis of elution fractions from the first and second SEC. Numbers correspond to peaks in (a) and (c). **(c)** Elution profile of peak 1 (species 3) from (a), subjected to a second SEC.

To further address a possible influence of concentrating steps on the Core7 reconstitution, an assembly procedure devoid of any concentrating steps was developed. For this, the different Core7 components were combined in equimolar ratio, thereby amounting to a total sample volume equal to the injection volume for SEC. Accordingly, 2.2 nmol of each subunit or dual complex, XPB, XPD, p62nT/p44, p52nT/p34, and p8, were mixed, amounting to a total sample volume of 50 µl. The mixture was incubated on ice for 1 h and subjected to SEC in Tris/TCEP buffer (Figure 3-135 a). A similar elution profile as observed before (Figure 3-131 a) was obtained, displaying three species. The central main peak elution fraction (species 2) was subjected to a second SEC (Figure 3-135 b), resulting in a single peak. Peak fractions were pooled and subjected to EM studies.

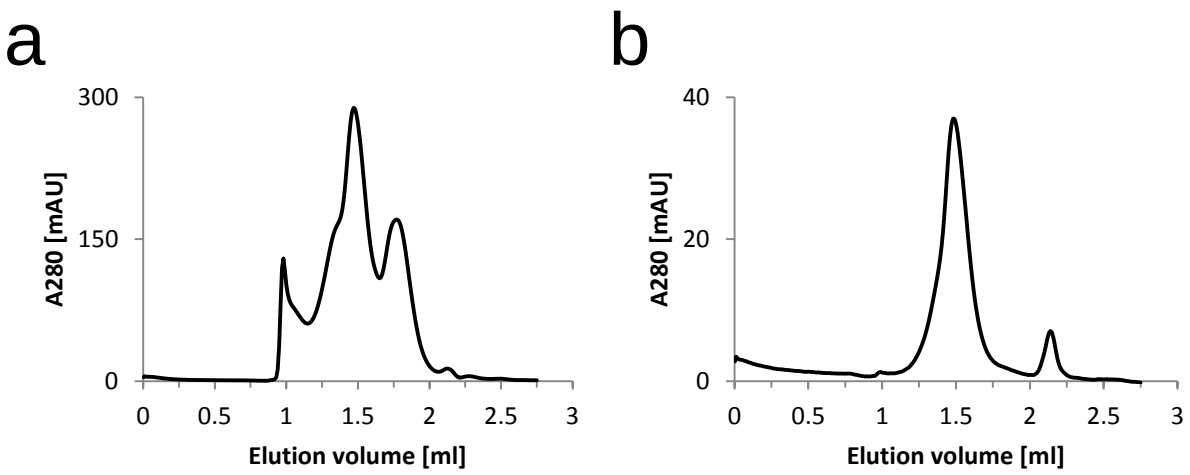


Figure 3-135. Core7 assembly at 2.2 nmol, devoid of any concentrating steps. (a) Elution profile of the first SEC from a Superose 6 Increase 3.2/300 column. **(b)** Elution profile of the second SEC. The central main peak fraction (species 2) from (a) has been used for the second SEC.

To also investigate the preceding shoulder (species 1) by means of EM, a further assembly devoid of any concentrating steps was conducted, using 2.5 nmol of each subunit or dual complex, XPB, XPD, p62nT/p44, p52nT/p34, and p8. Again, an elution profile, featuring three species, was obtained (Figure 3-136 a). According to the SDS-PAGE, species 1 and 2 contained all subunits (Figure 3-136 b). Both species were subjected to a second SEC. In both cases the main peak fractions reappeared at the respective elution volume, indicating distinct, stable entities. Peak fractions from both SECs were pooled separately and species 1 was subjected to EM studies.

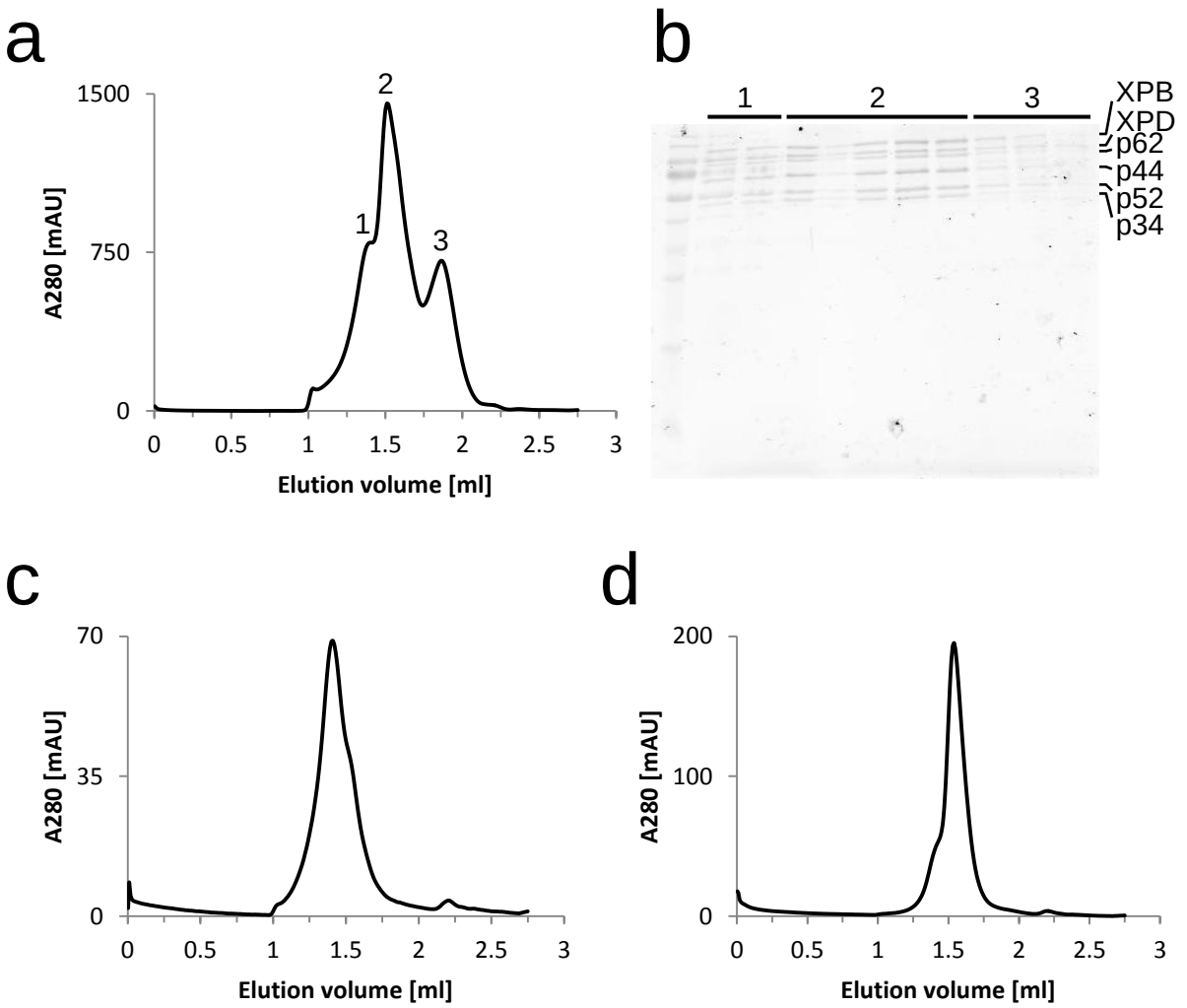


Figure 3-136. Core7 assembly at 2.5 nmol, devoid of any concentrating steps. (a) Elution profile of the first SEC from a Superose 6 Increase 3.2/300 column. **(b)** SDS-PAGE analysis of elution fractions from the first SEC. Numbers correspond to peaks in (a). **(c)** Elution profile of species 1 from (a), subjected to a second SEC. **(d)** Elution profile of species 2 from (a), subjected to a second SEC.

Based on the successful assembly of Core5 (see 3.13.1.1), another assembly strategy was performed to obtain Core7. Preassembled Core5 was used and combined with XPB and XPD. The assembly of the Core5 used for this approach is shown in Figure 3-111. 0.5 nmol of each, Core5, XPB, and XPD, were mixed, incubated on ice for 1 h, and subjected to SEC in Tris/TCEP buffer (Figure 3-137). Similar to the results obtained with the other assembly strategies, three peaks/shoulders were obtained, with the main peak corresponding to species 2 and the preceding shoulder corresponding to species 1 from the assemblies described above. The species 1 and 2 contained all core subunits, whereas the last species (peak 3) seemed to contain an excess of XPB (Figure 3-137 b).

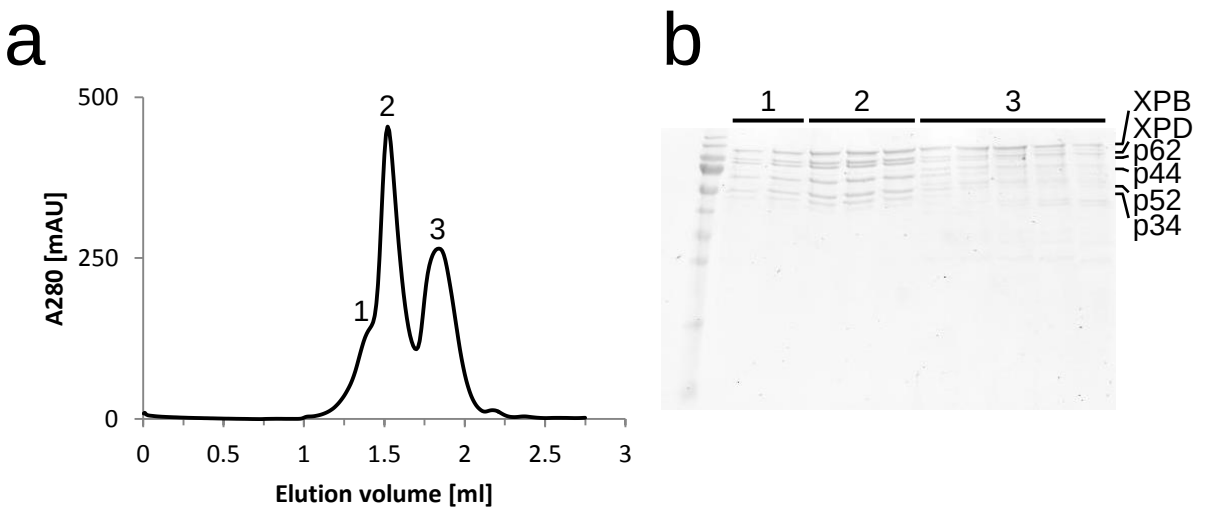


Figure 3-137. Core7 assembly using preassembled Core5. (a) SEC elution profile from a Superose 6 Increase 3.2/300 column. **(b)** SDS-PAGE analysis of elution fractions from the SEC. Numbers correspond to peaks in (a).

3.13.3.2. Electron microscopy of Core7

Core7, assembled via a Superose6 10/300 GL column (Figure 3-127), was subjected to EM studies. A stock solution of 0.3 mg/ml was diluted to about 0.05 mg/ml, negatively stained, and imaged (Figure 3-138). The micrograph displayed a seemingly crowded grid, with no discernible single particles.

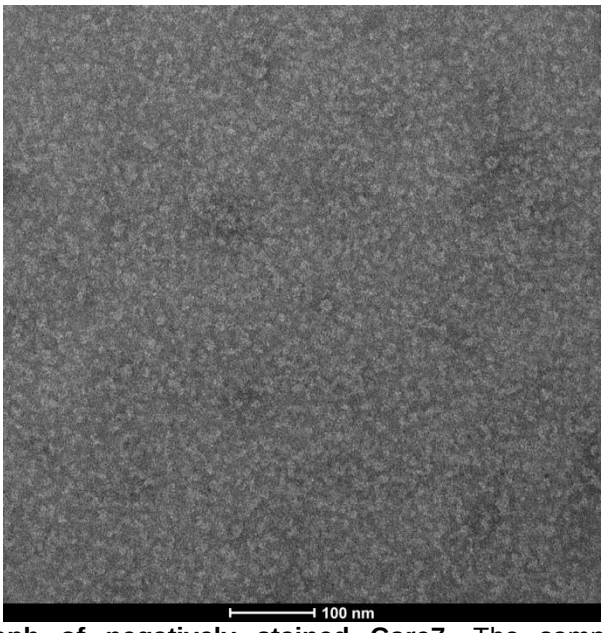


Figure 3-138. Micrograph of negatively stained Core7. The sample was deposited at a concentration of ~0.05 mg/ml.

Optimization of the deposition concentration was quite difficult, as sample dilution seemed to be detrimental for complex integrity. A Core7 stock of 1.2 mg/ml was diluted 1:20 and imaged via negative stain (Figure 3-139 a). Clear, homogenous particles could be observed. As the

particle density was slightly too high, the 1:20 dilution was diluted further to half of the concentration. The resulting 1:40 dilution was deposited and imaged via negative stain (Figure 3-139 b). In stark contrast to the previous micrograph, only very few and quite heterogeneous particles could be retrieved. A one step 1:30 dilution of the protein stock also led to disrupted particles.

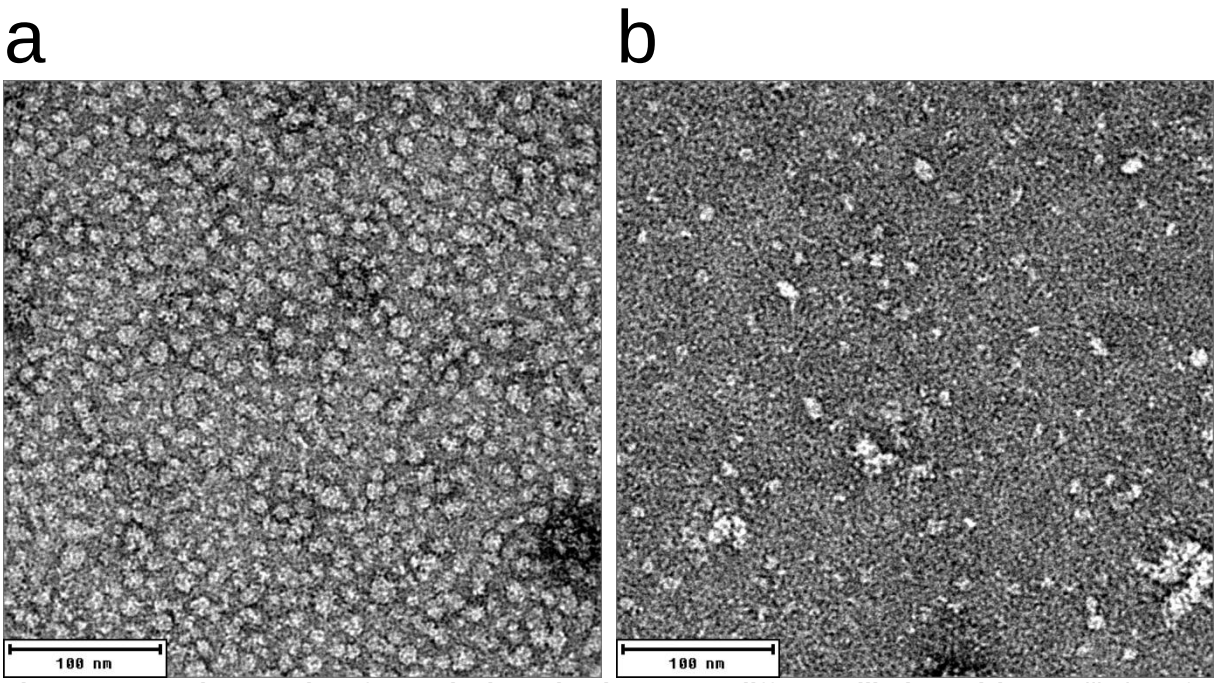


Figure 3-139. Micrographs of negatively stained Core7 at different dilutions. (a) 1:20 dilution of a 1.2 mg/ml Core7 stock. **(b)** 1:2 dilution of the diluted sample from (a).

To circumvent excessive dilution, assemblies using a Superose6 Increase 3.2/300 column with a bed volume of 2.4 ml were performed. To adjust sample concentration, higher amounts of proteins were applied for the assembly. When using a starting amount of 6.5 nmol, four species were obtained in Tris/TCEP buffer (Figure 3-132). For each species, about 0.05 mg/ml were deposited, negatively stained, and imaged. For species 1, particles could be observed, yet with some heterogeneity (Figure 3-140 a). For species 2, the grid appeared to be crowded, with no discernible single particles (Figure 3-140 b). For species 3, particles could also be observed, yet with some heterogeneity (Figure 3-140 c). For species 4, no clear particles could be observed (Figure 3-140 d).

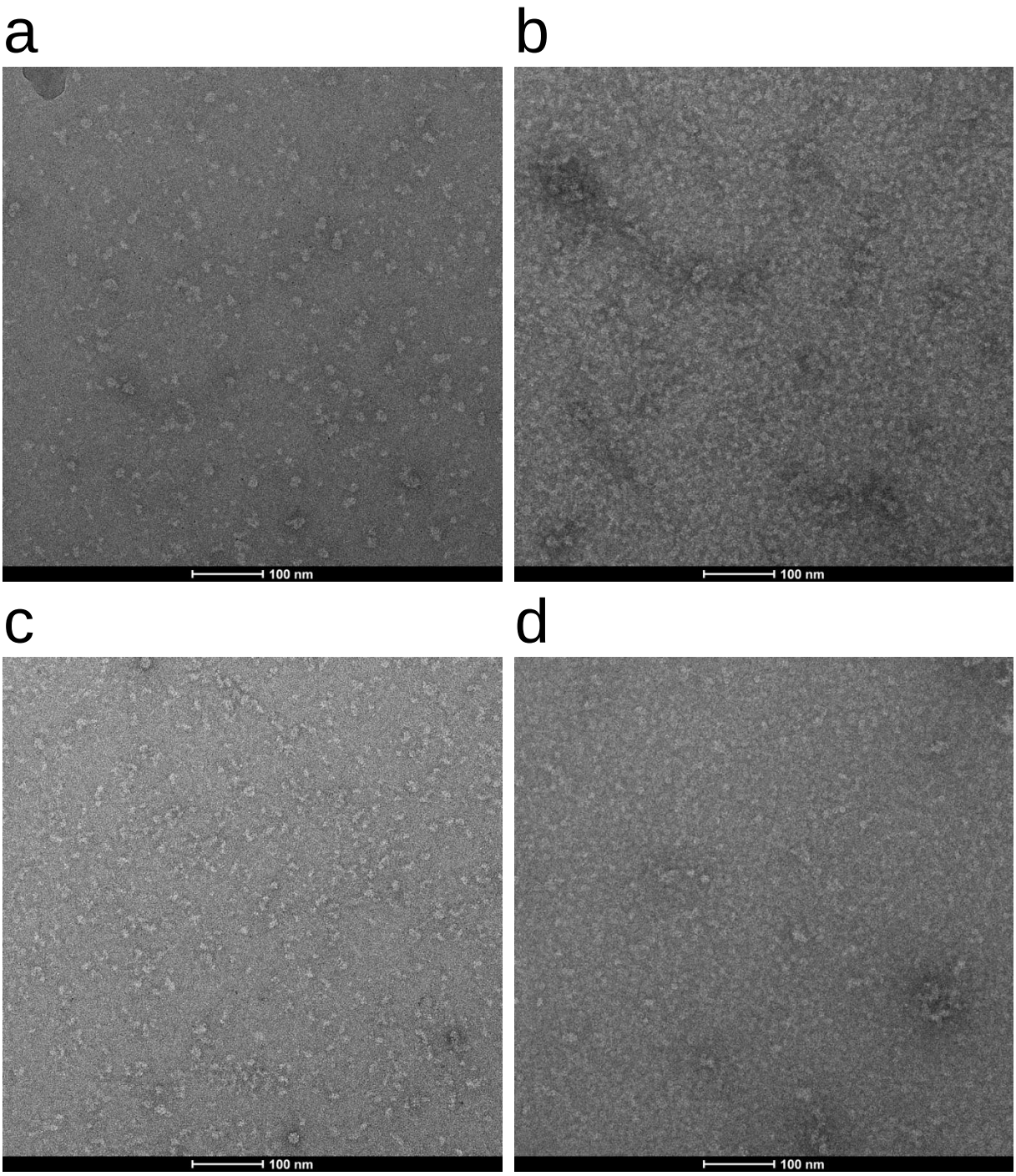


Figure 3-140. Micrographs of negatively stained Core7 assembled with a starting amount of 6.5 nmol in Tris/TCEP buffer. The sample was deposited at a concentration of ~0.05 mg/ml in each case. **(a)** Species 1. **(b)** Species 2. **(c)** Species 3. **(d)** Species 4.

Species 3 was subjected to further data collection. A protein stock of 0.5 mg/ml was diluted 1:32 and crosslinked with 0.5 % glutaraldehyde for 15 min on ice. The sample was deposited, negatively stained, and imaged. A micrograph of the crosslinked sample is shown in Figure 3-141. A dataset of 900 micrographs was recorded at the MPI in Berlin and processed by Ferdinand Krupp. 219 usable micrographs were selected, from which 39,124 particles were picked with EMAN2. For 2D classification half of the dataset was used.

Particles were grouped with SPARX via ISAC, yielding 149 stable classes (Figure 3-142). An initial model was reconstructed with SPARX and subjected to multiparticle refinement against all 13,164 particles with Spider. The final map is depicted in Figure 3-143, with a resolution of about 36 Å. Analogous to Core6 the map displayed a bulky core, with an appendix at the top, presumably corresponding to XPD. An additional protrusion appeared at the back side, presumably corresponding to XPB.

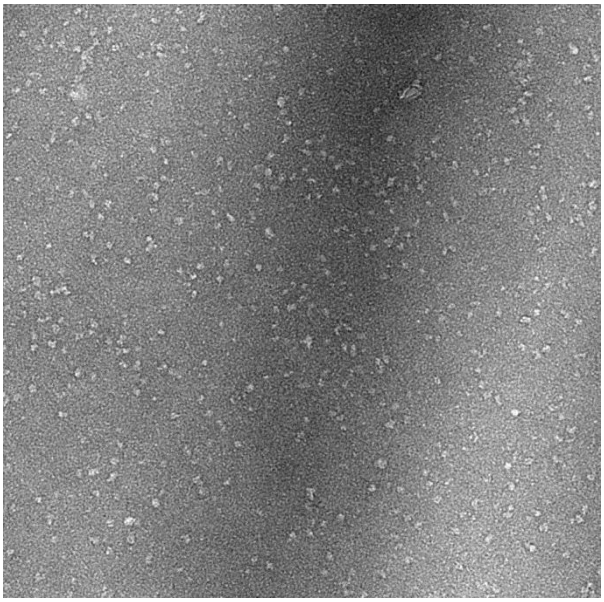


Figure 3-141. Micrograph of negatively stained Core7 crosslinked with 0.5 % glutaraldehyde.

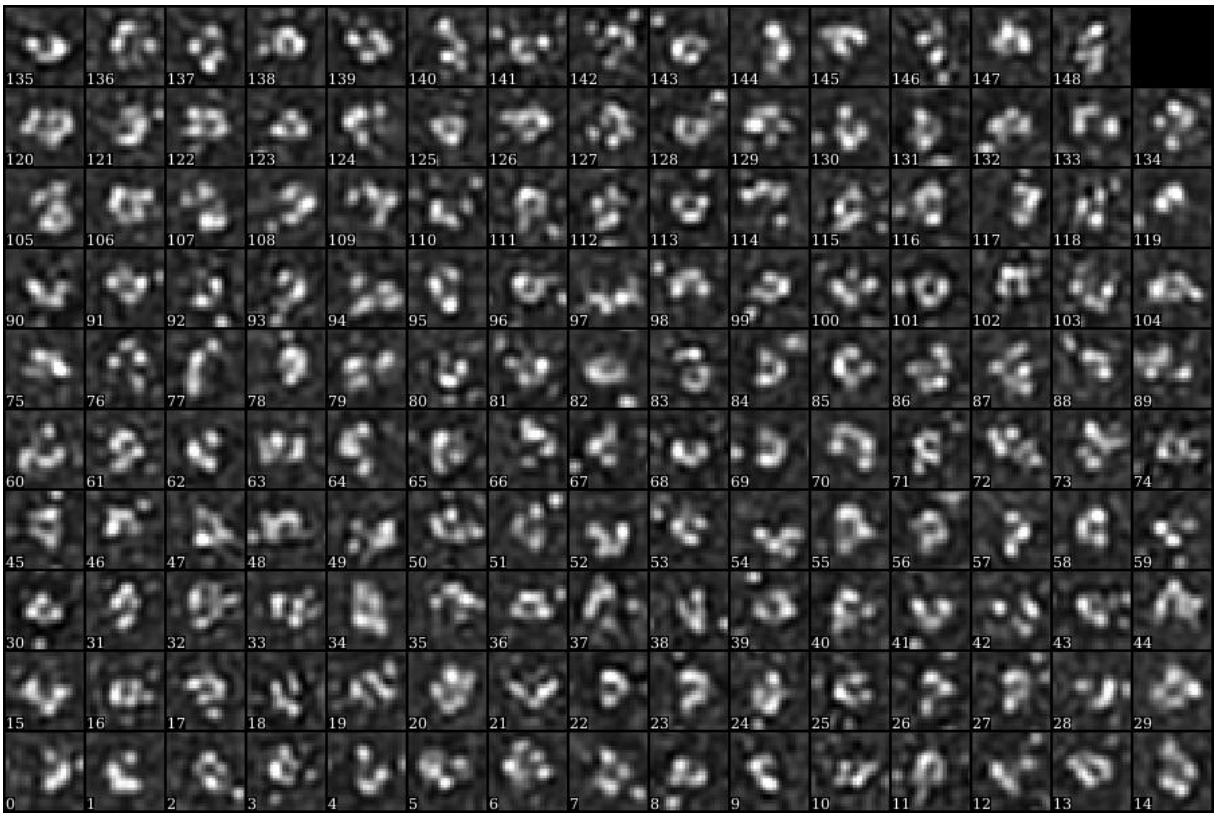


Figure 3-142. Class averages obtained from 2D classification for the negative stain dataset of Core7.

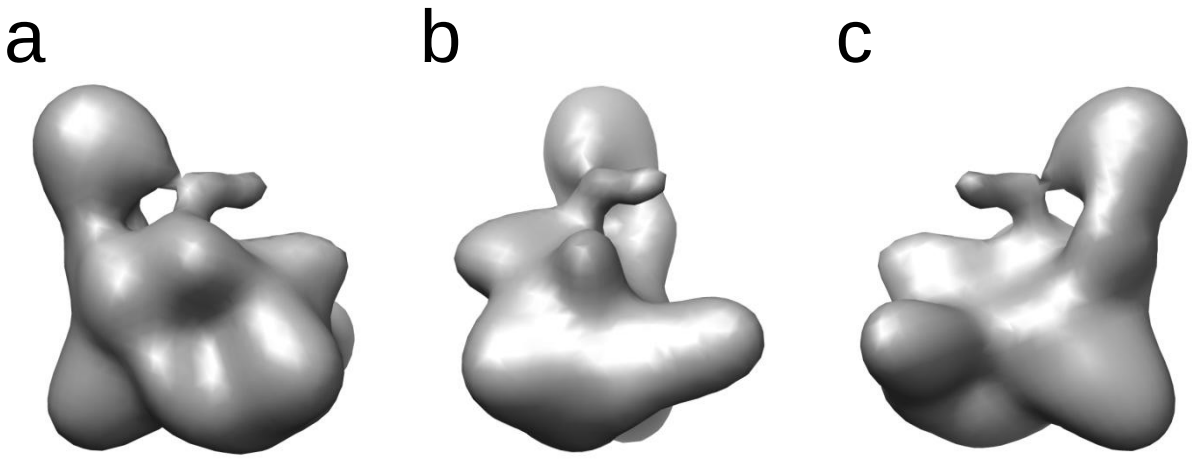


Figure 3-143. Reconstructed 3D map of Core7 from the negative stain dataset. (a) Front view. **(b)** Side view right. **(c)** Back view.

Core7 was assembled in Ches/TCEP buffer using a starting amount of 6.5 nmol (Figure 3-133). As before, four species were obtained, and each species was subjected to negative stain imaging. A micrograph of undiluted stock of about 0.01 mg/ml from species 1 is depicted in Figure 3-144. Particles could be observed, yet with some heterogeneity and aggregation.

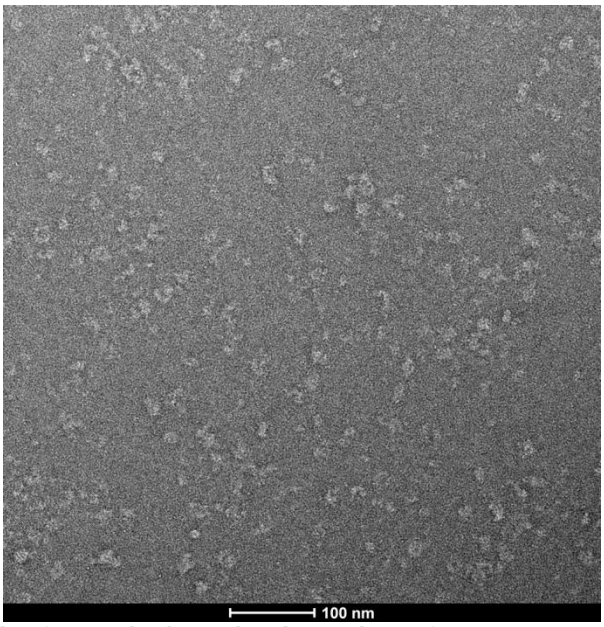


Figure 3-144. Micrograph of negatively stained species 1 from Core7 assembled with a starting amount of 6.5 nmol in Ches/TCEP buffer. The sample was deposited at a concentration of ~0.01 mg/ml.

The elution profile of species 2 displayed a main peak and a preceding shoulder (Figure 3-133 c), and both fractions were imaged separately. When imaged undiluted at about 0.1 mg/ml, the micrographs displayed quite crowded grids for the preceding shoulder (Figure 3-145 a). After dilution to about 0.01 mg/ml only some aggregates remained (Figure 3-145 b). Similar results were obtained with the main peak of species 2. A protein stock of 0.3 mg/ml was diluted to about 0.1 mg/ml and imaged (Figure 3-145 c). Again, the grid appeared to be quite crowded. After dilution to about 0.01 mg/ml it appeared that some particles were present, yet with some heterogeneity and aggregation (Figure 3-145 d). These findings corroborated the notion that dilution might be detrimental for complex integrity as also observed previously (Figure 3-139).

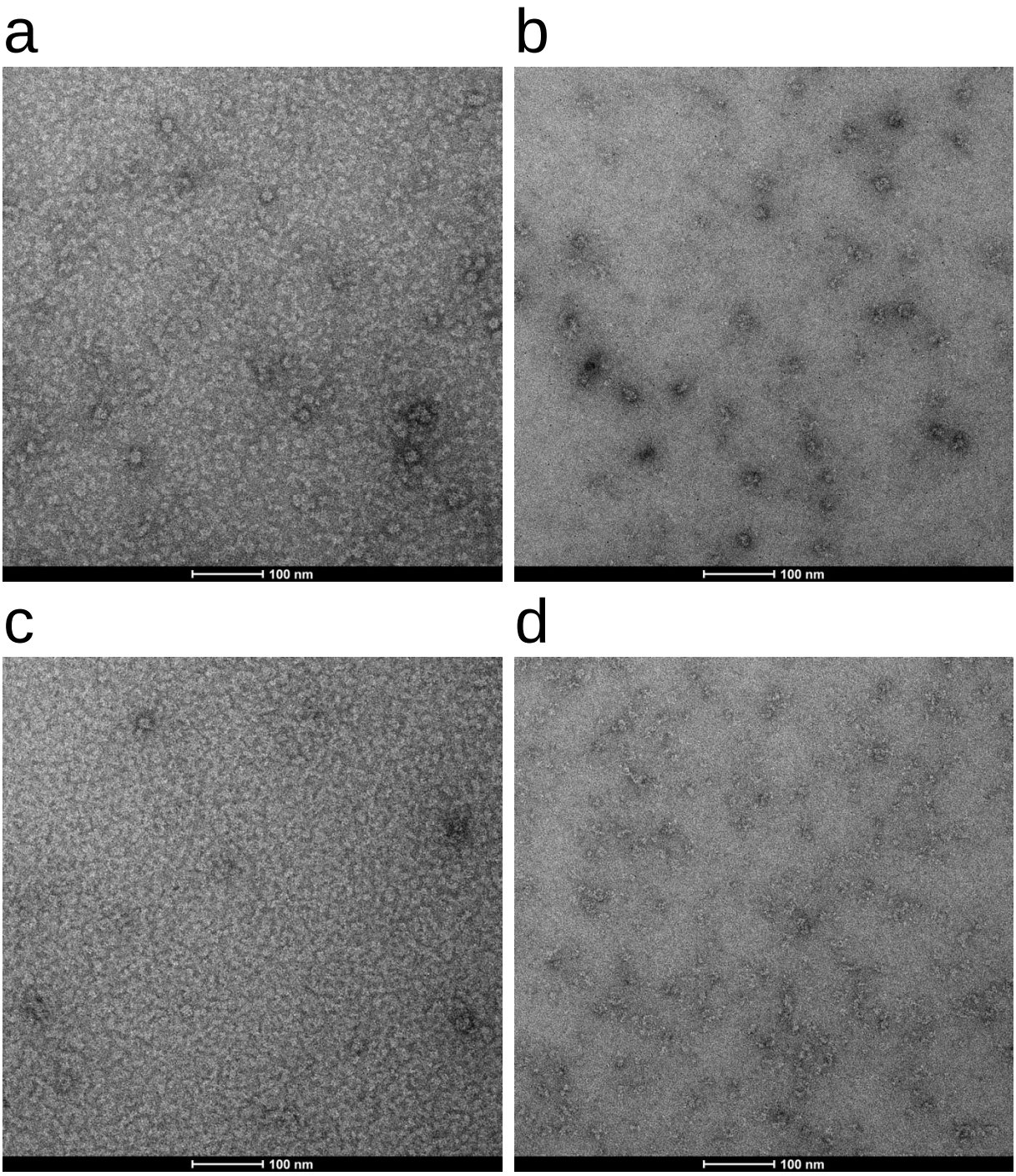


Figure 3-145. Micrographs of negatively stained species 2 from Core7 assembled with a starting amount of 6.5 nmol in Ches/TCEP buffer. (a) Preceding shoulder from the SEC, deposited at a concentration of ~0.1 mg/ml. **(b)** Preceding shoulder from the SEC, deposited at a concentration of ~0.01 mg/ml. **(c)** Main peak from the SEC, deposited at concentration of ~0.1 mg/ml. **(d)** Main peak from the SEC, deposited at a concentration of ~0.01 mg/ml.

The elution profile of species 3 also displayed a preceding shoulder and a main peak (Figure 3-133 d). Both fractions were imaged via negative stain. A stock solution of about 0.1 mg/ml of the preceding shoulder was diluted to about 0.01 mg/ml and deposited (Figure 3-146 a). Seemingly, it only contained aggregates. A main peak stock solution of 0.2 mg/ml was

diluted to about 0.01 mg/ml and deposited (Figure 3-146 b). Some particle-like objects could be obtained, yet mostly too small in size.

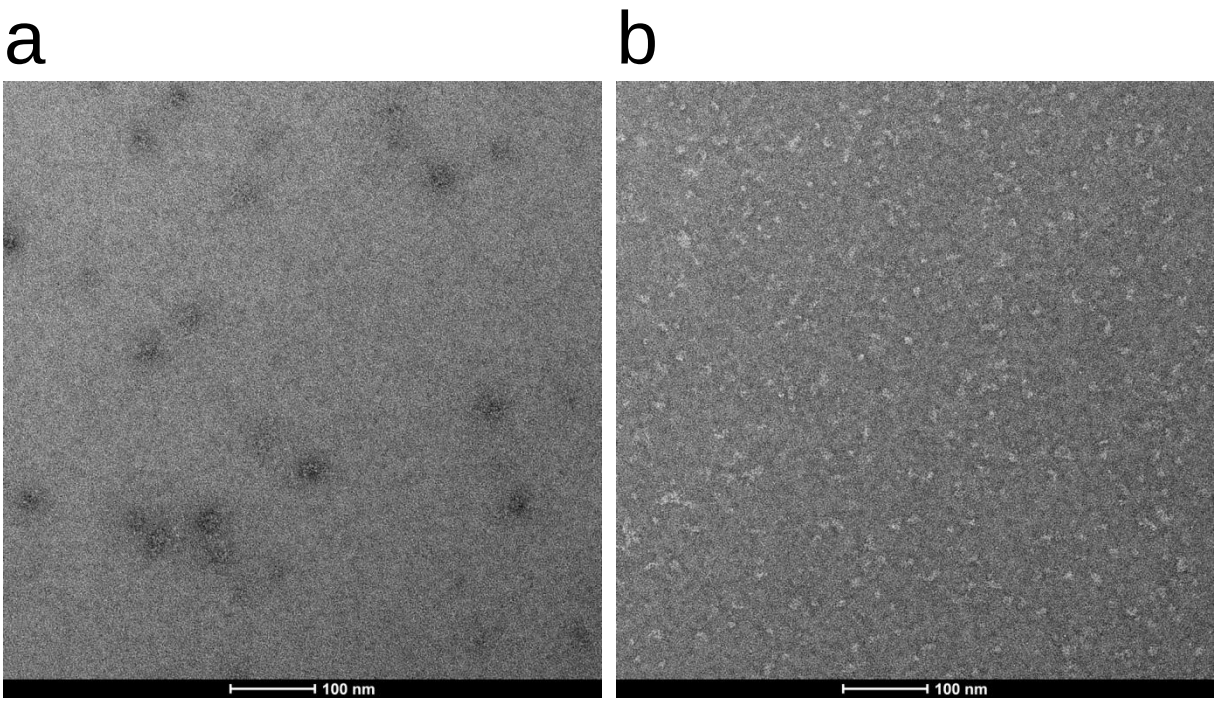


Figure 3-146. Micrographs of negatively stained species 3 from Core7 assembled with a starting amount of 6.5 nmol in Ches/TCEP buffer. (a) Preceding shoulder from the SEC, deposited at a concentration of ~0.01 mg/ml. **(b)** Main peak from the SEC, deposited at a concentration of ~0.01 mg/ml.

The elution profile of species 4 displayed two major peaks. As the elution volume of the first peak corresponded to the elution volume of species 2, the elution fractions of the second peak were investigated. A stock solution of 0.2 mg/ml was diluted to 0.01 mg/ml and deposited (Figure 3-147). The micrographs mainly displayed some aggregation.

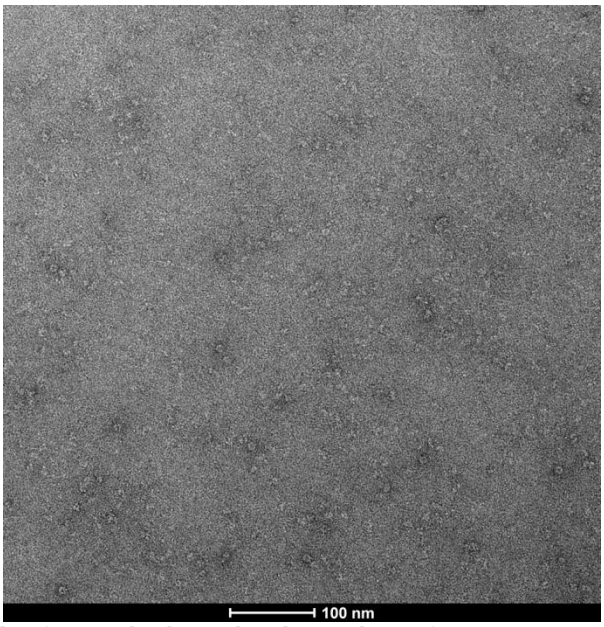


Figure 3-147. Micrograph of negatively stained species 4 from Core7 assembled with a starting amount of 6.5 nmol in Ches/TCEP buffer. The sample was deposited at a concentration of ~0.01 mg/ml.

The preceding shoulder of species 3 was also subjected to cryo-EM. The undiluted stock solution (0.1 mg/ml) was deposited, vitrified, and imaged (Figure 3-148). Clear particles, properly distributed, could be obtained.

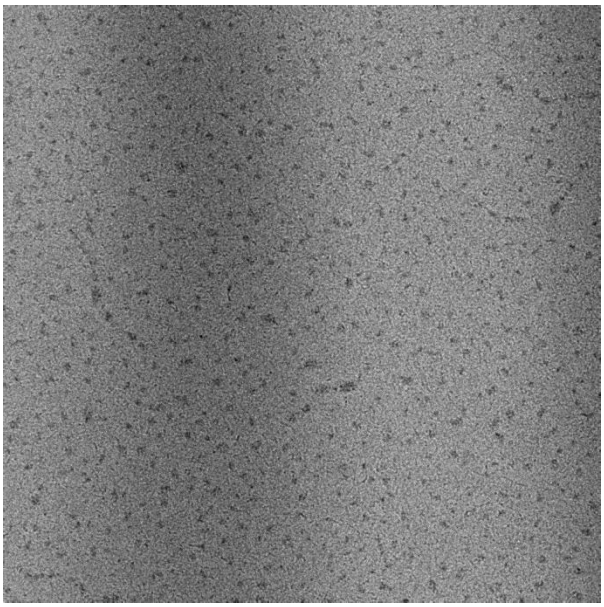


Figure 3-148. Micrograph of vitrified species 3 from Core7 assembled with a starting amount of 6.5 nmol in Ches/TCEP buffer. The preceding shoulder from the SEC was deposited at a concentration of ~0.1 mg/ml.

A dataset of 72 micrographs was collected at the homesource microscope (Tecnai™ G^2 Spirit Twin, FEI) and processed by Ferdinand Krupp. 16,335 particles were picked with EMAN2. Particles were grouped with SPARX via ISAC, resulting in 549 initial classes (Figure

3-149). Classes containing satellites or with low contrast were omitted, resulting in 211 remaining classes. These classes were used for reconstruction of an initial model with SPARX. The initial model was used for multiparticle refinement against all 16,335 particles with Spider. The final map is depicted in Figure 3-150, with a resolution of about 35 Å. The overall shape of the map from the reconstruction with negatively stained Core7 was discernible.

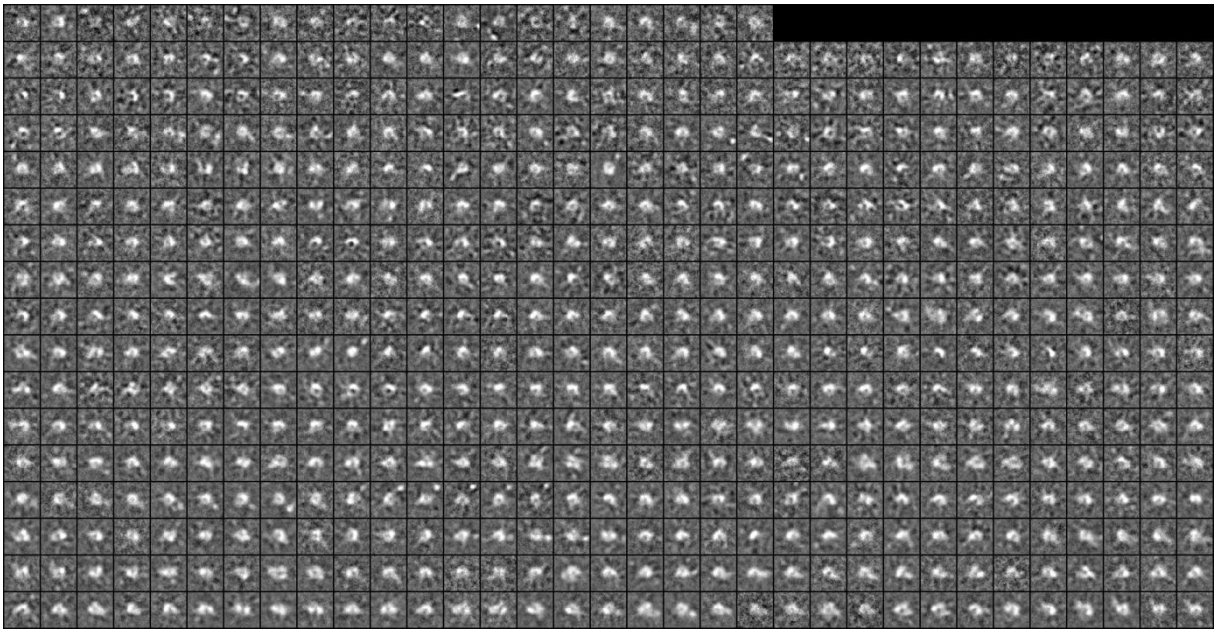


Figure 3-149. Class averages obtained from 2D classification for the cryo-EM dataset of Core7.

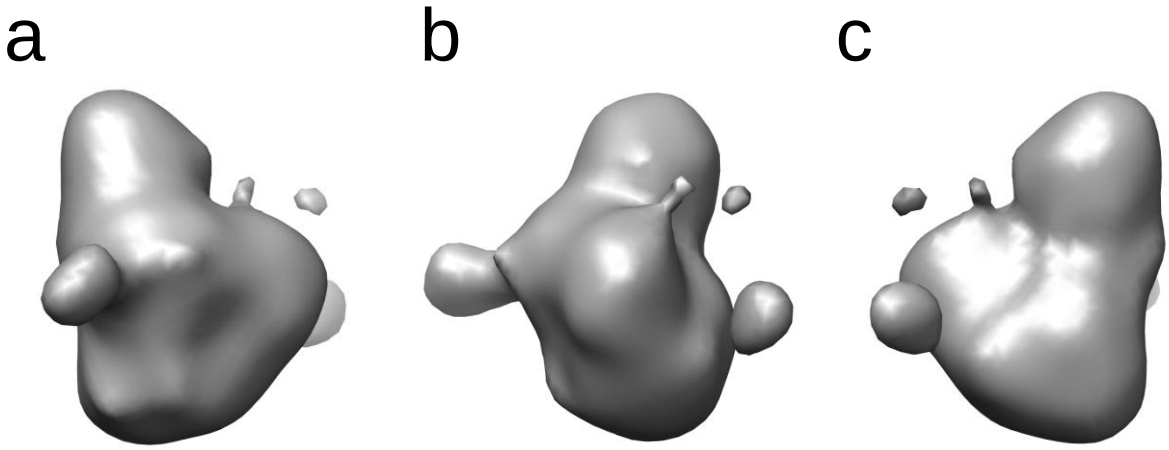


Figure 3-150. Reconstructed 3D map of Core7 from the cryo-EM dataset. (a) Front view. **(b)** Side view right. **(c)** Back view.

Core7 assemblies devoid of any concentrating steps were investigated by means of EM as well. Two species were obtained (Figure 3-136), which were both analyzed. A stock solution of the main peak species (Figure 3-135 b) of about 0.2 mg/ml was diluted to about 0.05 mg/ml and imaged via negative stain (Figure 3-151). Particles seemed to be visible, yet

the grid appeared to be quite crowded. The species from the preceding shoulder from the SEC (Figure 3-136 c) was vitrified undiluted at about 0.1 mg/ml and imaged (Figure 3-152). Properly distributed particles could be observed. When compared to the results from the negative stain, the particle density was much lower, even though a higher concentration was used for cryo imaging. This might be because of the different species used, or the general necessity of higher protein amounts for cryo approaches compared to negative stain [486].

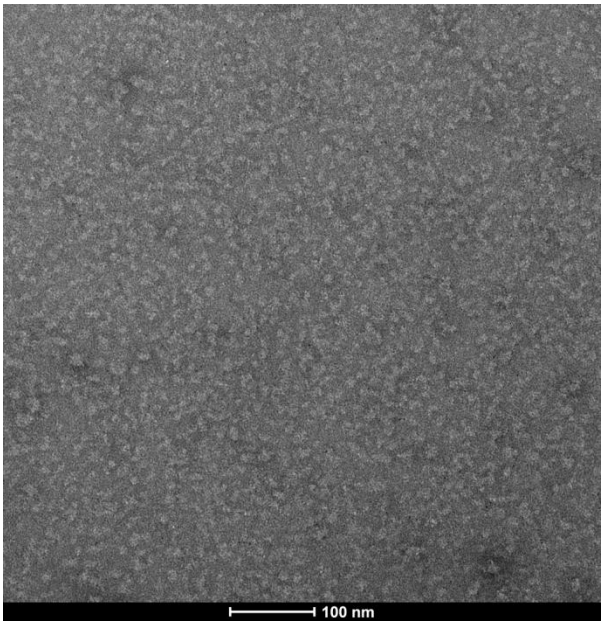


Figure 3-151. Micrograph of negatively stained Core7 assembled without any concentrating steps. The main peak species from the SEC was deposited at a concentration of ~0.05 mg/ml.

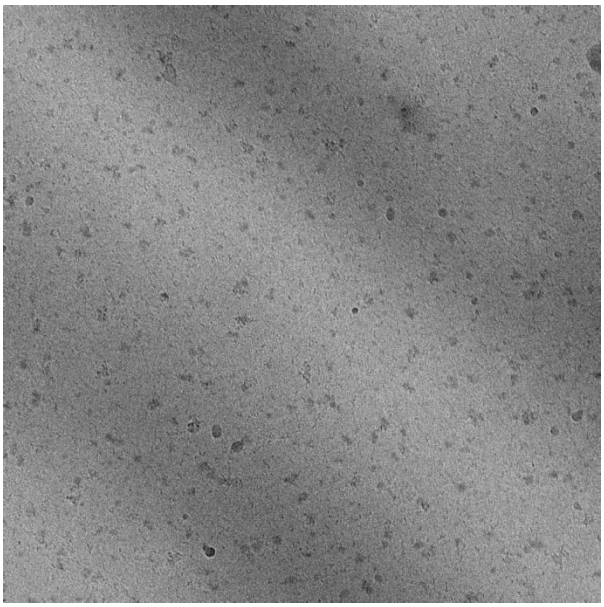


Figure 3-152. Micrograph of vitrified Core7 assembled without any concentrating steps. The species from the preceding shoulder from the SEC was deposited at a concentration of ~0.1 mg/ml.

3.13.4. Core7 and MAT1_1-248

3.13.4.1. Assembly

Incorporation of the CAK subunit MAT1 into the Core7 assembly was attempted. The MAT1_1-248 construct (Figure 3-43) with the tag cleaved off was used, which was provided by Florian Sauer. 3 nmol of each subunit or dual complex, XPB, XPD, p62nT/p44, p52nT/p34, p8, and MAT1_1-248, were mixed and incubated on ice for 1 h. The mixture was concentrated to 50 µl with a centrifugal filter unit and subjected to SEC (Figure 3-153 a, b). Different species were obtained, with the main peak and the preceding shoulder containing all core subunits and MAT1_1-248 in stoichiometric amounts. The preceding shoulder and the main peak were both subjected to another SEC (Figure 3-153 c, d). Both species reappeared at the respective elution volume, indicating the presences of distinct, stable species. Peak fractions were pooled and subjected to EM studies.

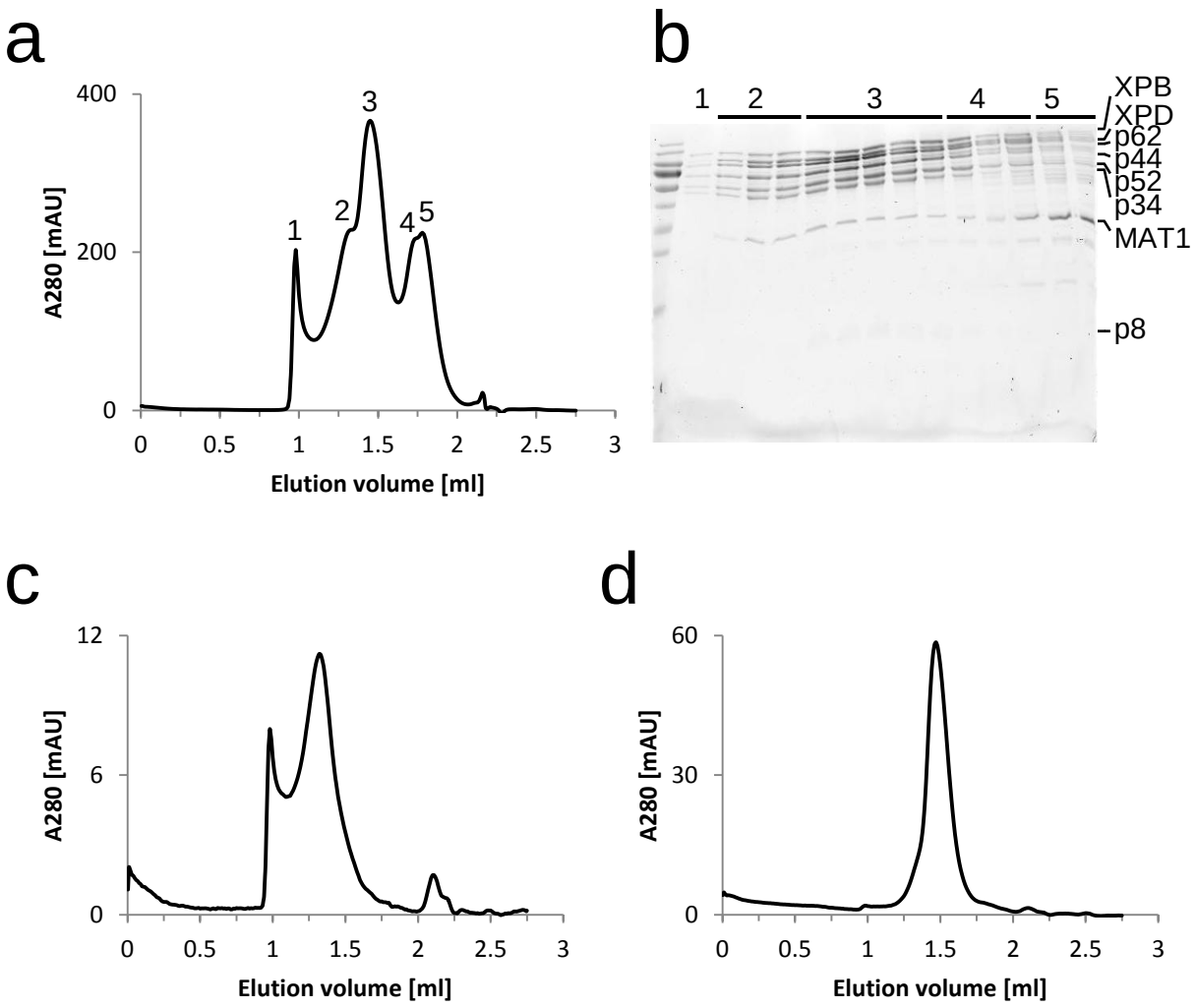


Figure 3-153. Assembly of Core7 and MAT1_1-248. (a) Elution profile of the first SEC from a Superose 6 Increase 3.2/300 column. **(b)** SDS-PAGE analysis of elution fractions from the first SEC. Numbers correspond to peaks in (a). **(c)** Elution profile of species 2 from (a), subjected to a second SEC. **(d)** Elution profile of species 3 from (a), subjected to a second SEC.

3.13.4.2. Electron microscopy of Core7 and MAT1

Assembled Core7/MAT1_1-248 (Figure 3-153) was subjected to EM studies. A stock solution of 0.2 mg/ml was diluted to about 0.05 mg/ml, negatively stained, and imaged (Figure 3-154 a). Properly distributed clear particles could be obtained. Furthermore, undiluted sample was vitrified and imaged as well (Figure 3-154 b). Again, particles were observed. As already observed for Core7, the particle density was much lower compared to negative stain, even though the concentration during deposition was higher.

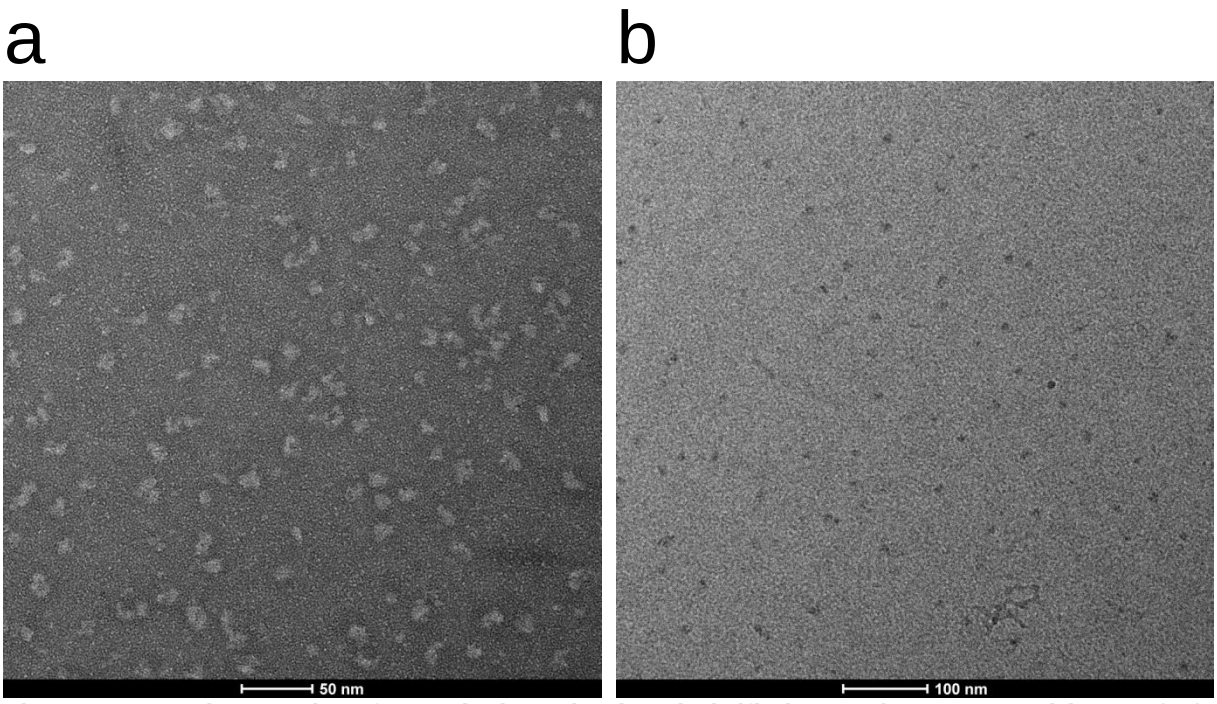


Figure 3-154. Micrographs of negatively stained and vitrified Core7/MAT_1-248. (a) Negatively stained sample, deposited at a concentration of ~0.05 mg/ml. **(b)** Vitrified sample, deposited at a concentration of ~0.2 mg/ml.

3.13.5. Comparison of EM maps

The EM maps obtained from the negative staining of Core5, Core6, and Core7 were compared. Figure 3-155 shows the fitting of the maps into each other. When comparing Core5 to Core6 (Figure 3-155 a), an additional appendix appears at the top, presumably corresponding to XPD. The additional density appeared to be weak, possibly reflecting flexibility or partial loss of XPD in the core complexes. When comparing Core5 to Core7 (Figure 3-155 b), the density at the top became clearer, which might be due to an overall stabilizing effect of XPB. Additionally, a protrusion at the back of the map appeared, which was attributed to XPB. Comparison of Core6 to Core7 (Figure 3-155 c) was in line with the notion from above, as Core6 displayed no counterpart for the protrusion at the back of Core7.

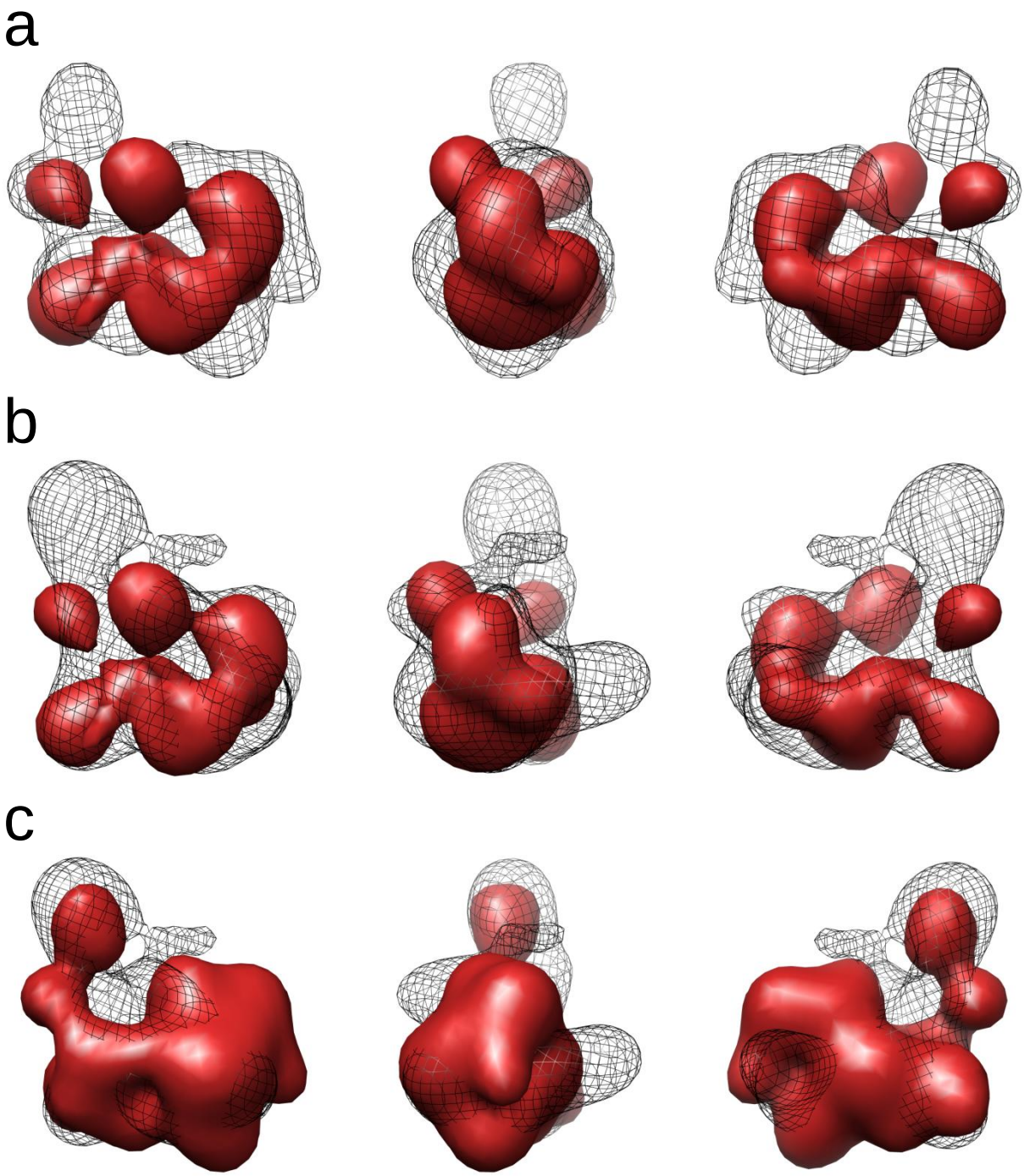


Figure 3-155. Comparison of the EM maps of Core5, Core6, and Core7, obtained by negative staining. The front view (left), side view (middle), and back view (right) are shown. **(a)** Core5 (red surface) compared to Core6 (black mesh). **(b)** Core5 (red surface) compared to Core7 (black mesh). **(c)** Core6 (red surface) compared to Core7 (black mesh).

The Core7 maps obtained from negative stain and cryo-EM were compared (Figure 3-156). It appeared that the cryo-EM map did not support all features of the map from the negative stain. From the shape and volume of the map either XPB or XPD seemed to be missing. The assignment was not unambiguous, therefore both possibilities were pursued. The fit supporting XPD is shown in Figure 3-156 a, the fit supporting XPB is shown in Figure 3-156 b.

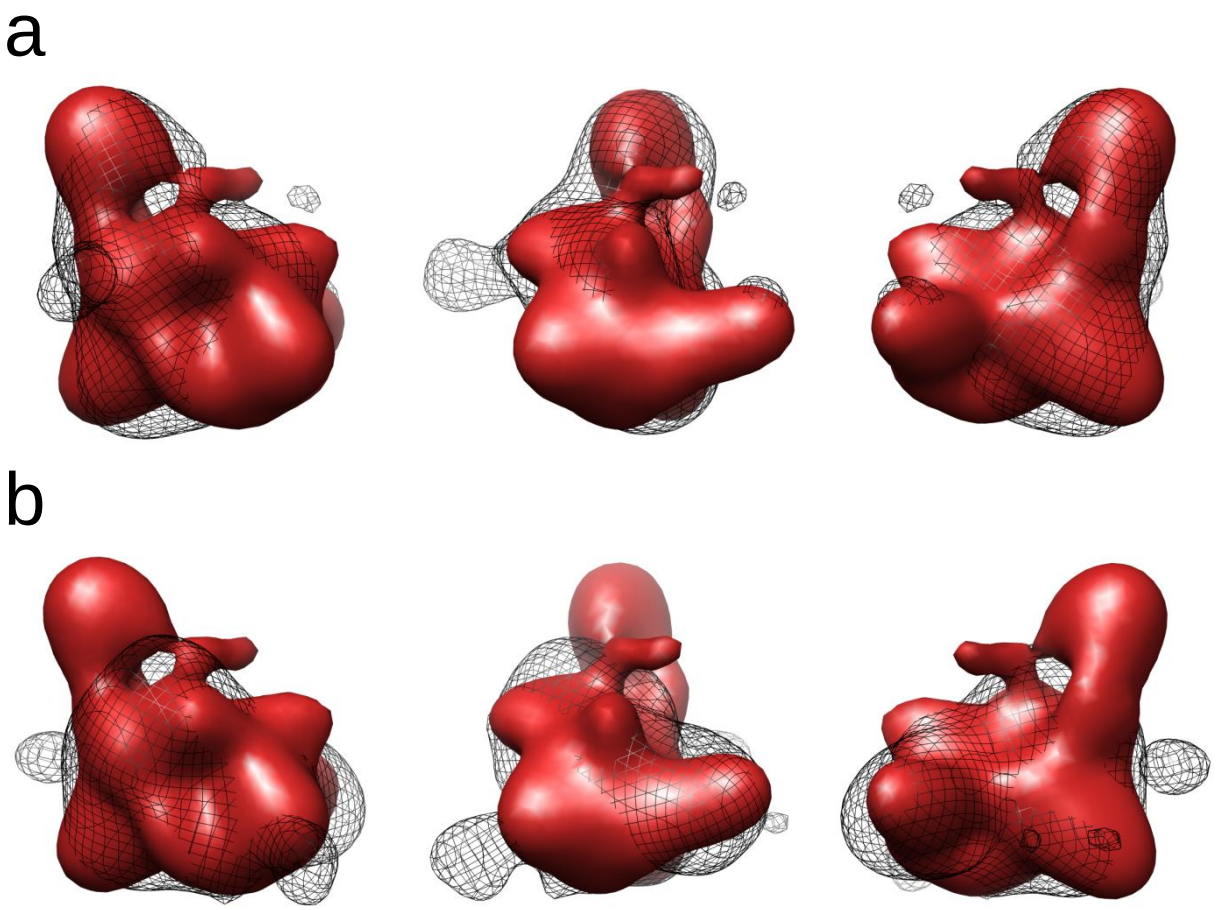


Figure 3-156. Comparison of the negative stain and the cryo-EM map of Core7. The front view (left), side view (middle), and back view (right) are shown. Red surface: negative stain map; black mesh: cryo-EM map. The fitting was not unambiguous, therefore two possibilities are shown. **(a)** Fit supporting XPD. **(b)** Fit supporting XPB.

3.13.6. Fitting of TFIIH core subunits

The TFIIH subunits were fitted into the EM maps obtained via negative stain for Core5, Core6, and Core7 and to the cryo-EM map for Core7. The comparison of these maps with each other (Figure 3-155, Figure 3-156) was used as a guideline. Furthermore, the subunits were placed according to the interactions observed in this work (see 3.11 and 3.12) and also based on the interactions described in the literature. As template for the TFIIH core subunits model 5IVW (human TFIIH core bound to DNA within the PIC) [76] was used. The model was adjusted by moving the different subunits to their suggested position. Due to the limited resolution of these maps, the placement of the subunits was not unambiguous and has to be taken with some caution.

The fitting of the subunits to Core5 is depicted in Figure 3-157. Based on the observed p52/p34 interaction (see 3.11.3), p34 was positioned in the center of the hook-shaped volume, adjacent to p52. The vWA like domain of p44 was placed in a weakly attached additional volume above the hook-shaped volume. This placement implies some flexibility of the vWA like domain of p44, which might well be the case in the absence of its interaction

partner XPD. This position supports the placement of the p44 zinc finger and ring finger domains underneath, directly adjacent to p34. To the left of this position the hook-shaped volume comprises a peripherally attached volume. This volume was attributed to p62, placing it next to the p44 zinc finger domain, in line with the results from our study (see 3.11.2). Finally, the EM map comprises an additional volume behind the vWA like domain of p34 for the placement of the C-terminal zinc finger domain of p34.

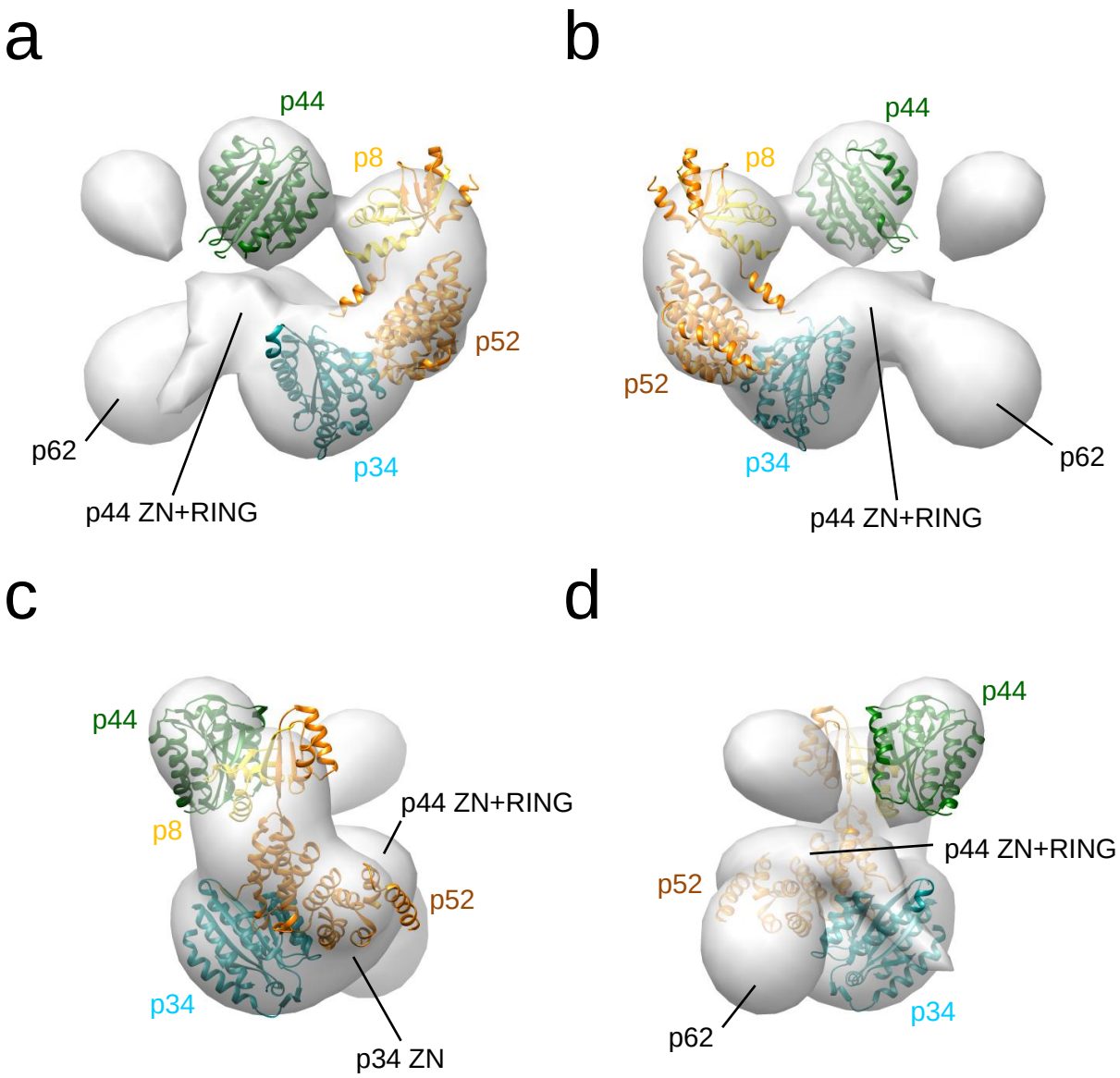


Figure 3-157. Fitting of the TFIIH core subunits to the negative stain EM map of Core5. Positions of p62 and additional domains of p44 and p34 are indicated. **(a)** Front view. **(b)** Back view. **(c)** Side view right. **(d)** Side view left. ZN: zinc finger domain; RING: ring finger domain.

The fitting of the subunits to Core6 is depicted in Figure 3-158. Placement of the subunits was based on the fitting for Core5 (Figure 3-157). Core6 displayed an additional appendix at the top, which was attributed to XPD. Placement of the vWA like domain of p44 near XPD

was ambiguous and not consistent with the placement in Core5. However, the volume for XPD is only weakly defined, proposing flexibility or loss of XPD in some of the assembled complexes. This might then lead to in flexibility of the vWA like domain of p44 as well, as already hypothesized for Core5. Again, the EM volume supported an additional volume for a possible fitting of p62 and the additional domains of p44 and p34.

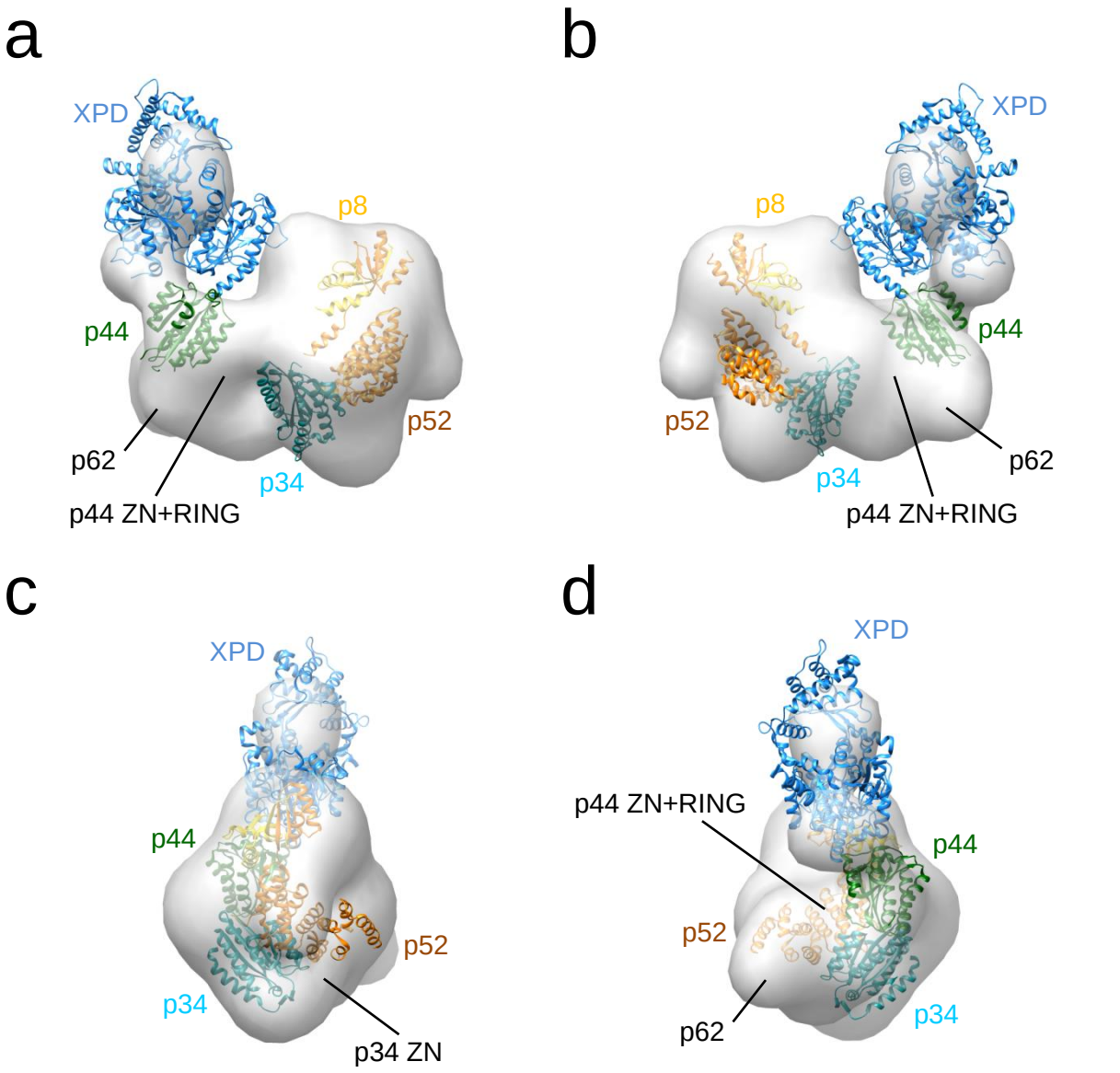


Figure 3-158. Fitting of the TFIIH core subunits to the negative stain EM map of Core6. Positions of p62 and additional domains of p44 and p34 are indicated. **(a)** Front view. **(b)** Back view. **(c)** Side view right. **(d)** Side view left. ZN: zinc finger domain; RING: ring finger domain.

The fitting of the subunits to Core7 from negative stain is depicted in Figure 3-159. The map supported the hypotheses from Core5, also regarding the positioning of the vWA like domain of p44. In line with a seemingly less flexible XPD, an additional volume for the vWA like domain of p44 was present. In comparison to Core5 and Core6, a protrusion at the back of

the map appeared, which was assigned to XPB. This position placed XPB directly adjacent to p52, which is in nice agreement with the interaction of both proteins. As observed for Core5 and Core6, the map contained an additional volume for a possible placement of p62 and the additional domains of p44 and p34.

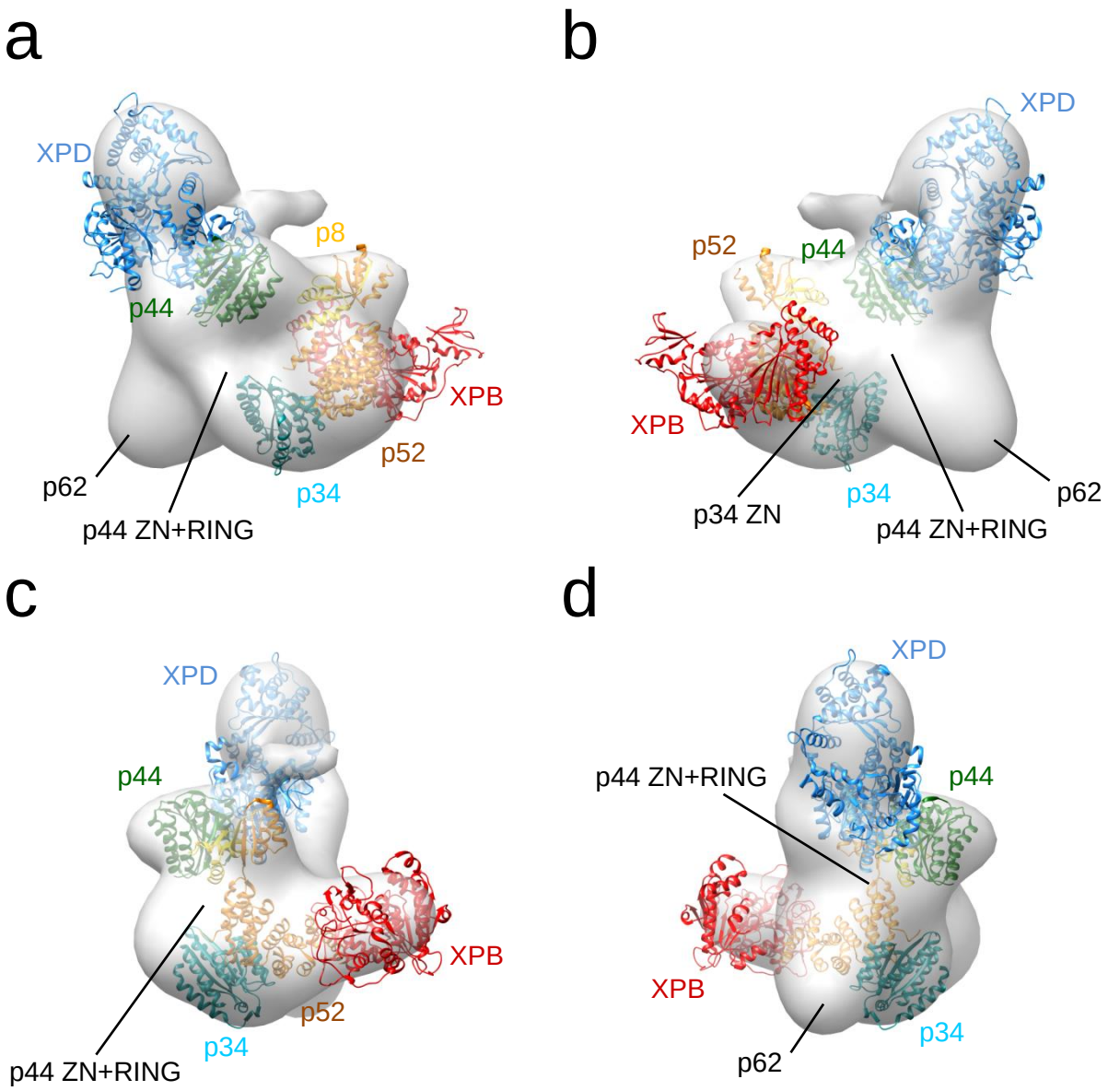


Figure 3-159. Fitting of the TFIIH core subunits to the negative stain EM map of Core7. Positions of p62 and additional domains of p44 and p34 are indicated. **(a)** Front view. **(b)** Back view. **(c)** Side view right. **(d)** Side view left. ZN: zinc finger domain; RING: ring finger domain.

From the comparison of the Core7 maps obtained by negative stain and by cryo-EM it can be speculated that either XPD or XPB was missing in the cryo-EM map (Figure 3-156). The subunits were fitted according to both hypotheses, and the subunit fitting to the negative stain Core7 map was used as a guideline. The fitting of the TFIIH core subunits containing either XPD or XPB as well as the other five subunits into the cryo-EM Core7 map is depicted

in Figure 3-160. Placement of the subunits consistent with the hypothesized positions in the negative stain EM map of Core7 was possible, also leaving some additional volume to accommodate p62 and the additional domains of p44 and p34. When XPB was fitted into the maps, an additional volume at the suggested position of p44 was observed, which could not be reasonably interpreted due to the limited resolution.

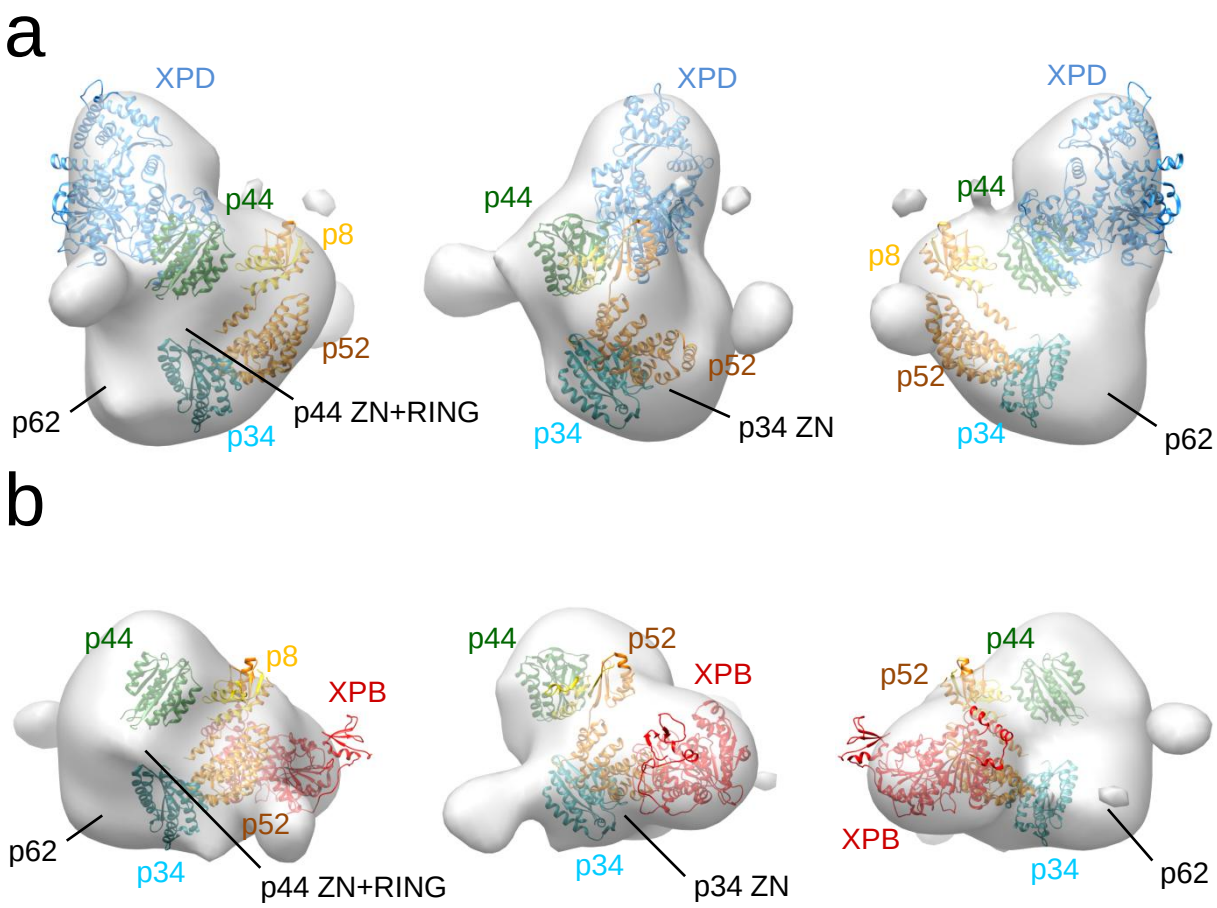


Figure 3-160. Fitting of the TFIIH core subunits to the cryo-EM map of Core7. The front view (left), side view (middle), and back view (right) are shown. Positions of p62 and additional domains of p44 and p34 are indicated. The fitting was not unambiguous, therefore two possibilities are shown. **(a)** Fit supporting XPD. **(b)** Fit supporting XPB. ZN: zinc finger domain; RING: ring finger domain.

4. Discussion

In this work, the purification of several functional TFIIH core subunits could be established and via interaction studies a TFIIH core interaction network could be carved out. The subunits were reconstituted to core complexes and investigated via EM. It was possible to interpret the EM maps and to fit structural models of the TFIIH core subunits into these maps. Based on this analysis a TFIIH core interaction network was generated, which allowed to draw conclusions on the roles of p34 and p62 within this complex.
Additionally, the applicability of cesium for experimental phasing was assessed, utilizing a shortened construct of the p62 subunit.

4.1. Cesium for experimental phasing

4.1.1. Compatibility of CsCl with purification, crystallization, and cryo-protection

In this study, three proteins were tested, p62_1-109, HEWL, and p34_1-277. P62_1-109 represents a novel target, whereas HEWL is a well characterized protein, frequently used for methodic studies. P34_1-277 constitutes a medium resolution target. The replacement of NaCl or KCl with CsCl during protein purification was tested for two proteins, p62_1-109 and p34_1-277. In both cases, the purification and appearance of the SEC elution profile were virtually identical suggesting that CsCl can readily replace NaCl or KCl during the purification process without interfering with the stability of the proteins (see 3.10.1, Figure 3-46 and Figure 3-49), and thus discloses CsCl as a suitable salt component in protein buffers.
The introduction of CsCl during crystallization was analyzed for three different proteins, p62_1-109, HEWL, and p34_1-277. In two cases, p62_1-109 and HEWL, supplementing with CsCl did not lead to an apparent detrimental influence on crystallization (see 3.10.1, Figure 3-47 and Figure 3-48). In the remaining case, p34_1-277, crystallization was impeded, which might be due to the influence of CsCl. However, crystallization of p34_1-277 was a particular case, as crystallization had to be conducted without mixing of the protein solution and reservoir solution. In addition, the overall protein quality after purification in general was less reproducible compared to the other proteins used in this study.
Usage of CsCl in the cryo-protectant solution was analyzed for all three proteins. Seemingly, the presence of CsCl in the cryo-protectant had no detrimental effect, as for two cases diffraction of sufficient quality for experimental phasing could be retrieved (see 3.10.2, Table 3-6 and Table 3-7). For the third case, p34_1-277, only diffraction to approx. 4 Å could be obtained. If the lower resolution compared to prior data sets collected in the absence of CsCl

[47], was due to the introduction of CsCl or due to lower overall protein quality remained unclear at that point.

Taken together, these findings suggest that CsCl is compatible with all three major steps of sample treatment for structural studies via crystallography: purification, crystallization, and cryo-protection. The experimental phasing approach via cesium was successful for two out of three cases, thereby also yielding a novel crystal structure for p62_1-109. However, the results have to be taken with some caution, as the number of test cases is so far limited. To corroborate the findings from these studies further test cases should be included.

4.1.2. Introduction of cesium as anomalous scatterer

Due to the compatibility of CsCl to be included in purification, crystallization, and cryo-protection (see 4.1.1), cesium could be introduced as anomalous scatterer at each of these stages. CsCl can be used in a cryo-soak approach as described for halides by Dauter et al. [487]. This approach has been successfully conducted for p62_1-109, which permitted to solve the phase problem by means of SAD (see 3.10.2, Table 3-6, crystal # 3). Hence, cryo-soaking with CsCl provides a good alternative to cryo-soaking with halides, especially in cases were crystals suffer from treatment with halides, or no bound anions can be obtained due to unfavorable surface charges of the protein. Here, the opposite charge of cesium might be of great benefit.

To further extend this approach, the introduction of cesium during protein purification was considered. P62_1-109 crystal # 1 was treated with 0.25 M CsCl only during the purification step (see 3.10.1, Table 3-3). The anomalous substructure exhibited a single cesium ion, occupying site 4 (see 3.10.3, Figure 3-54 a). Cesium site 4 constituted a highly conserved binding site, also occupied by cesium in all other p62_1-109 approaches (see 3.10.3, Table 3-11). However, the anomalous signal from this single cesium ion was not sufficient to overcome the phase problem.

The direct comparison of p62_1-109 treated with CsCl only during cryo-protection (# 3) with p62_1-109 treated with CsCl during SEC and cryo-protection (# 4) suggests an additive effect of the CsCl treatment. Whereas 5 occupied anomalous sites were detected for # 3, it was possible to identify 8 sites as well as a higher occupancy sum for # 4 (see 3.10.3, Table 3-11). These approaches were also conducted with HEWL. Consistent with the p62_1-109 results, 16 sites and a higher occupancy sum could be observed for HEWL treated with CsCl during SEC and cryo-protection (# 4) compared to HEWL only treated with CsCl during cryo-protection (# 3), where only 15 sites could be detected (see 3.10.3, Table 3-12). The additive effect of CsCl utilized during SEC and for cryo-protection makes it possible to decrease the CsCl concentration in the cryo-soak step, thereby still preserving sufficient anomalous

scatterers for successful phasing (p62_1-109 # 2). This approach might be beneficial for sensitive crystals, unable to withstand high salt concentrations in the cryo-soak step.
Cesium was also used as a component of the crystallization buffer, replacing KCl in case of p62_1-109 and NaCl in case of HEWL (see 3.10.1, Table 3-3 and Table 3-4). For both proteins, crystallization (see 3.10.1, Figure 3-47 and Figure 3-48) and diffraction (see 3.10.2, Table 3-8 and Table 3-9) was not affected. Hence, supplement of CsCl via replacement of related alkali salts poses yet another possibility to introduce cesium.

4.1.3. Phasing procedure with CsCl

Our studies suggest that CsCl provides an elegant, easy-to-use, low cost, and little time consuming approach to overcome the phase problem. CsCl is easily commercially available and much cheaper than for example selenomethionine. Importantly, the handling of CsCl is straightforward, making it suitable as initial standard phasing procedure before applying more sophisticated approaches like protein expression in the presence of selenomethinone. Furthermore, no special equipment is necessary for the application of this approach.
For phasing, a step-wise procedure can be pursued, depending on the demands of the project. First, cryo-soaks with CsCl can be applied as the easiest approach. If the anomalous signal is not sufficient, the cryo-soaks can be combined with supplementing CsCl during SEC, which would also permit to use a reduced CsCl concentration in the cryo-soak if this is detrimental for the crystals. Finally, when the identified crystallization conditions contain alkali chloride salts, cesium can be introduced by replacing the alkali.

4.2. Availability of TFIIH

In the course of this work, protocols for the purification and functional characterization of several TFIIH core subunits, either as single proteins or as dual complexes, were established (Figure 4-1).

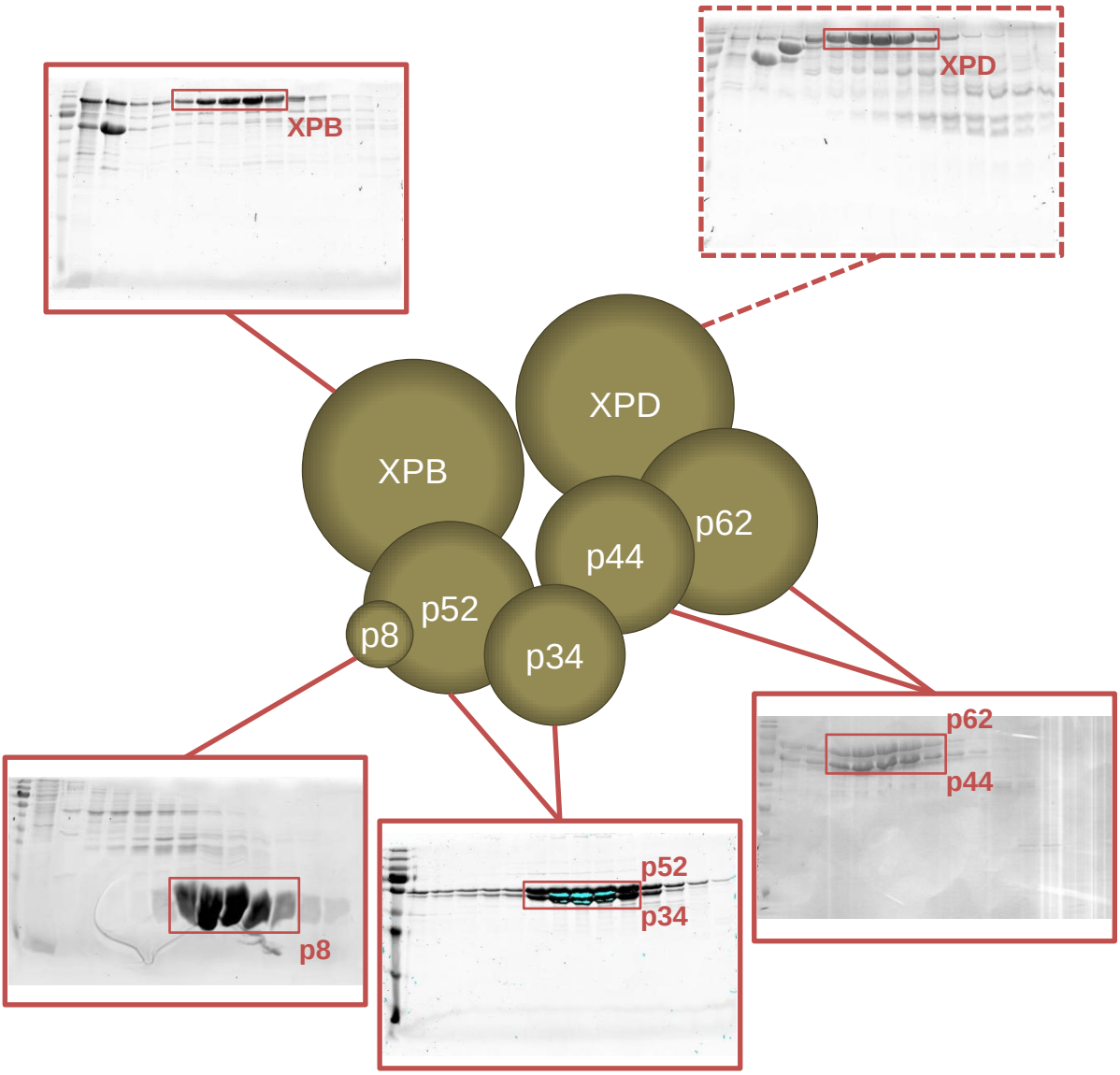


Figure 4-1. Availability of the TFIIH core subunits. The purification of several core subunits was established. Purification of XPD (dashed box) was established by Gudrun Sander and Jochen Kuper.

Of central importance is the availability of functional XPB as a crucial TFIIH member. Activity assays clearly indicated XPBs ATPase activity, which could be stimulated by p52/p8 and specifically inhibited by triptolide (see 3.2.3, Figure 3-11). This clearly demonstrated the full functionality of the protein.

Functional full-length p44 could be generated by means of stabilization with p62. When expressed and purified on its own, p44 displayed several higher oligomers during SEC. A typical elution profile of p44 is depicted in Figure 4-2. In addition, the functional integrity was severely impaired, which can be seen in the interaction study via native agarose gel electrophoresis (see 3.11.2.1, Figure 3-77). Here, singly purified p44 was not able to form a complex with its known binding partner p62, showing that its integrity was compromised. These issues were solved by co-expression and co-purification with p62. The SEC elution profile of p62/p44 displayed a homogenous peak, devoid of any higher oligomers (see

3.11.2.1, Figure 3-79). Furthermore, native agarose gel electrophoresis indicated a clear complex band (see 3.11.2.1, Figure 3-77).

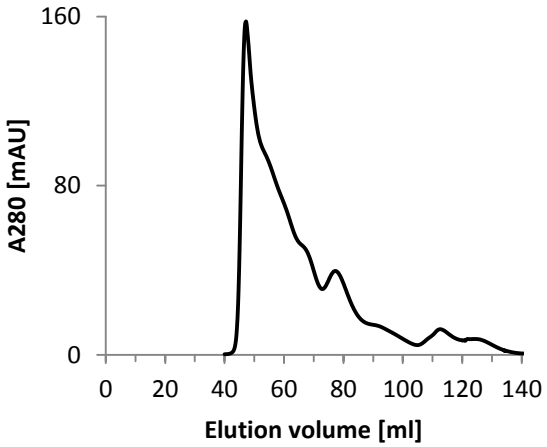


Figure 4-2. SEC elution profile of p44 from a HiLoad 16/600 Superdex 200 pg column. P44 eluted in several higher oligomeric states. The elution profile was provided by courtesy of Dominik Schmitt.

Purification of p34 on its own resulted in low protein yields, also displaying some contaminations (see 3.7.2, Figure 3-41). In contrast, co-expression and co-purification with p52 resulted in high protein yield and quality (see 3.11.3.1, Figure 3-88).

All of these individual steps made it possible to establish a unique and powerful TFIIH core assembly system. This system permits the modular and step wise reconstitution of the TFIIH core, thereby allowing to monitor its individual enzymatic activities and the effects on the enzymatic activity when individual subunits are added. This system thus provides a unique possibility to obtain insights into the intricate interaction network within TFIIH. The most difficult protein to obtain for this approach was XPB. Another critical step was the establishment of fully functional p44 in form of a p62/p44 complex. Full-length p44 is crucial for the overall function of TFIIH since it activates XPD via its N-terminal domain, whereas the C-terminal part of p44 mediates the connection to the remaining TFIIH core via its interaction with p34. Accordingly, full-length p44 was essential for TFIIH core assembly, which was achieved through the p62/p44 complex, providing a p44 fully capable of activating and integrating XPD within TFIIH.

4.3. The TFIIH core interaction network

Via co-purification (see 3.11) and native PAGE interaction studies (see 3.12), a TFIIH core interaction network was developed. A summary of all newly identified or confirmed interactions is given in Figure 4-3.

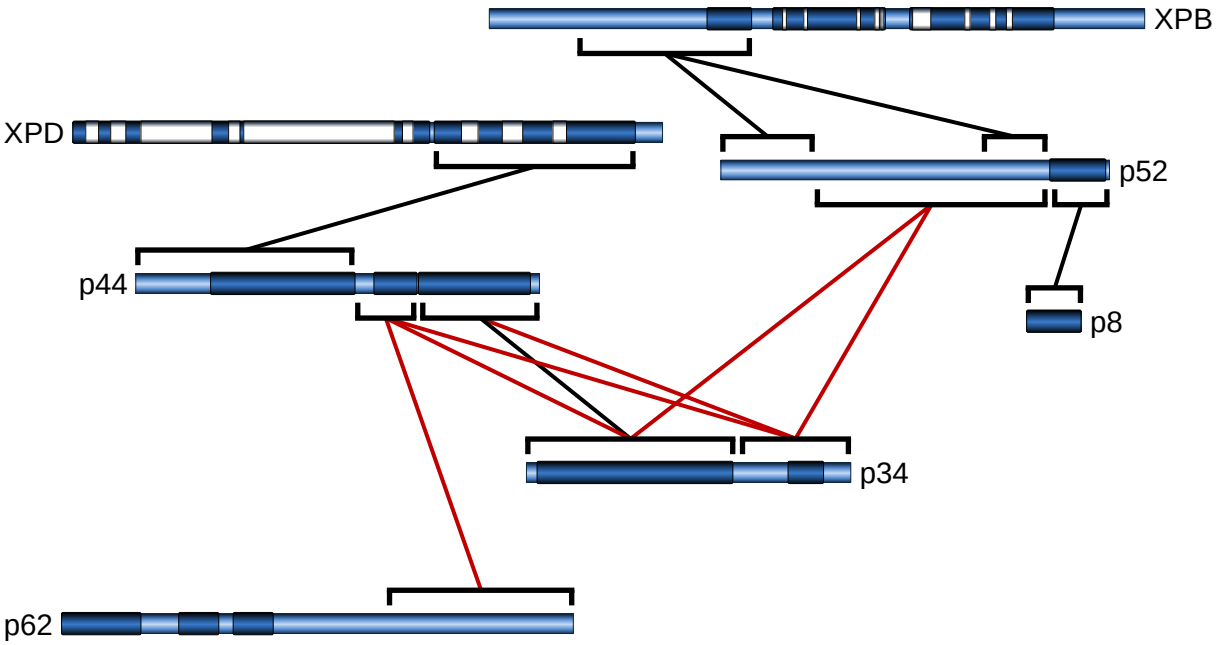


Figure 4-3. The TFIIH core interaction network. Regions interacting with each other are connected by lines. Key interactions established in this work are shown in red. The domain scheme corresponds to Figure 1-1.

The well-established interactions of XPB with p52 and XPD with p44 [22, 25, 30] could be confirmed. Furthermore, the interaction between p52 and p8 [28] could be confirmed, which, together with the XPB/p52 interaction, constitutes the XPB/p52/p8 subcomplex [79]. An interaction between the vWA like domain of p34 and the C-terminal ring finger domain of p44 has been shown before [48], and could be confirmed as well. The involvement of the C-terminus of p62 towards an interaction with p44 has been suggested [46], and was also confirmed. The p62 binding site on p44 could be narrowed down to the central zinc finger domain in p44. An interaction of p34 with p52 has been suggested [80], which could be shown in this study. Our results suggest that both domains in p34 (the vWA like domain and the C-terminal zinc finger domain) seem to be involved in the interaction with p52, whereas the N-terminal part of p52 does not seem to be required for complex formation with p34. The complex network is then further established through the interaction of p34 with p44, which in addition to the interactions described above requires the C-terminal zinc finger domain of p34 and the central zinc finger domain of p44. Via chemical crosslinking/mass spectrometry studies Luo et al. suggested interactions of XPD with p62, p62 with p34, and p52 with p44 [80]. These interactions could not be confirmed. Possible explanations for this discrepancy might be detection of false positives with the crosslinking procedure due to subunit/complex flexibility [488] or very transient interactions, which were not detectable through the assays used in this work. However, transient interactions would contradict a scaffolding role of these interactions.

4.3.1. An important role for p34

An unexpected role for TFIIH integrity could be observed for p34. P34 consists of two domains, an N-terminal vWA like domain, and a C-terminal zinc finger domain [31, 47]. Co-purification experiments (see 3.11.3) and native PAGE analysis (see 3.12.2) revealed that p34 interacts with p52, firmly establishing the link between the submodules XPD/p62/p44 and XPB/p52/p8. Native PAGE analysis further revealed that the vWA like domain of p34 is sufficient for the interaction with p52 (see 3.12.2, Figure 3-97). However, when comparing a co-purification of p52/p34 to p52/p34_1-277 (see 3.11.3, Figure 3-88 and Figure 3-89), complex formation seemed to be weakened for the latter case. Furthermore, reconstitution of Core5 by means of SEC was impeded without the C-terminal zinc finger domain of p34 (see 3.13.1.1, Figure 3-112). These results indicate that the interaction between p52 and p34 is weakened when the C-terminal zinc finger domain of p34 is absent, which can be clearly observed when the more harsh conditions of SEC are utilized. Based on these results it was concluded that also the C-terminal zinc finger domain participates in the interaction with p52. The strong influence of this p34 domain on the Core5 assembly furthermore suggests that this subunit assumes an important role for the integrity of the entire TFIIH.

The vWA like domain of p34 interacts with the ring finger domain of p44 [48]. Native PAGE studies using a p34 interface mutant showed that the C-terminal zinc finger domain of p34 also participates in the interaction with p44 (see 3.12.3, Figure 3-100 and Figure 3-101). Native PAGE furthermore indicated that the interface extends beyond the ring finger domain of p44, including its central zinc finger domain (see 3.12.3, Figure 3-102).

It was thus shown that both domains of p34 interact with p52 and p44, which suggests that p34 is a central scaffolding protein, connecting XPD/p44 and XPB/p52/p8. Egly and Coin proposed two functionally related submodules within TFIIH: XPB/p52/p8 and XPD/p44/p34 [79], and a role for p34 towards the regulation of XPD. Our results rather support the hypothesis that p34 acts as a central scaffold, which also redefines the two submodules within TFIIH: XPB/p52/p8 constitutes the first functionally related submodule, whereas XPD/p44 defines the second and the connection between the two submodules is established via p34 (Figure 4-4).

4.3.2. A role for p62

So far the role of p62 within TFIIH is only poorly characterized. A recent study proposed that p62 acts as an important scaffold for XPD anchoring, as well as a support for the XPD/p44 interaction [80]. Our results clearly contradict this notion. Native PAGE analysis did not support an interaction between XPD and p62 (see 3.12.1). Closer analysis of the p62/p44 interaction positioned the p62 binding site towards the central zinc finger domain of p44 (see 3.11.2). Furthermore, anchoring of p62 via p34 was suggested [80]. However, when the p34

mediated Core5 formation was investigated utilizing p62/p44_1-367 (see 3.12.3, Figure 3-102), complex formation was detectable but strongly reduced. A direct interaction between p62 and p34 should have compensated for the loss of the interface between p34 and the C-terminal ring finger domain of p44, masking this phenotype. Thus, the hypothesis of a p62/p34 interaction is unlikely based on the data obtained from this work. So far we can only suggest that p62 is presumably attached to TFIIH via the central zinc finger domain of p44. It has been shown that the PH domain of p62 competitively interacts with XPC, XPG, TFIIE, and p53 [43, 45, 63, 69, 489]. These results suggest that p62 constitutes the major TFIIH interface to the environment. This hypothesis goes along with a peripheral attachment of p62 to TFIIH, only mediated between p44 and the C-terminus of p62, whereas the PH domain and the BSD tandem within p62 are available to be addressed by environmental factors. This hypothesis is further supported by the fact that no obvious volume for p62 is present in published EM maps (Figure 4-5 c) [76], which may indicate high flexibility or loss of p62. However, no structural model of p62 is available and the resolution of the available EM maps is so far limited, which hampers interpretation and fitting of p62; issues that might also be a reason for the absence of p62. Nevertheless, the supposed flexibility of p62 might well explain the contradictory results of the chemical crosslink/mass spectrometry studies obtained by Luo et al. [80]. Accordingly, p62 would be able to transiently come into close proximity to XPD and p34 without directly interacting with both subunits and get trapped by the crosslink, thereby generating false positives.

The three major roles of TFIIH are global NER, TCR, and transcription [1, 148]. In global NER, initial damage detection is performed by XPC/HR23B/Centrin2 [152]. TFIIH is then recruited to the repair site via the interaction of the p62 PH domain with XPC [42, 44, 45]. In a later stage XPG replaces XPC by binding to the same interface on p62, thereby permitting the progression of the DNA repair process [43, 44]. In TCR, the lesion is recognized via a stalled RNAPII [138]. Thus, XPC is not involved in initial damage detection and TFIIH recruitment. However, it has been shown that XPG binds to stalled RNAPII [248]. In this case, TFIIH might be recruited directly via the interaction of XPG to p62. In transcription, TFIIH is involved in promotor opening and the phosphorylation of the CTD of RNAPII [7, 50]. During PIC formation, the general transcription factors are sequentially recruited, with TFIIH joining last [4]. TFIIH is recruited to the PIC by TFIIE, again through the interaction with the PH domain of p62 [63]. Taken together, p62 is a central component in all three processes and seems to be involved in the TFIIH recruitment during global NER, TCR, and transcription. In addition to XPC, XPG, and TFIIE, it has been shown that p53 also interacts with the PH domain of p62 [69]. The tumor suppressor p53 is a transcriptional activator implicated in cell cycle arrest and apoptosis [490]. P53 induces expression of its target genes by recruiting the transcription machinery [491]. In this model TFIIH would be directed towards

these target genes through the interaction of p53 with p62. The fact that p53 addresses p62 corroborates the notion that p62 provides a major TFIIH interface to the environment. The central importance of this TFIIH interface is underlined by the finding that it is exploited by the herpes simplex virus. Herpes simplex virus possesses a transcriptional activator protein VP16, which induces transcription of viral genes [492]. For this purpose, VP16 recruits TFIIH, again by addressing the PH domain of p62 [68, 70]. Another example is the rift valley fever virus (RVFV). RVFV is an RNA virus and replicates in the cytoplasm [493]. As such, it is independent of the host transcription machinery and these viruses often inhibit transcription of the host cell to avoid an antiviral response [494]. RVFV thus possesses the virulence factor NSs [493]. NSs interacts with the PH domain of p62 [72], thereby blocking the interface for other factors. Furthermore, NSs targets p62 to the ubiquitin proteasome pathway and p62 is rapidly degraded at an early stage of viral infection [495, 496]. According to p62s supposed role as external interface, immediate blockage and destruction of p62 would effectively hinder the recruitment of TFIIH to transcription start sites, thereby shutting down host cell transcription.

The provision of one central interface, competitively addressed by different proteins, presumably aids in organizing the participation of TFIIH in its different processes. On the one hand it might minimize the interference of the different pathways by hindering recruitment of a TFIIH complex, which is already engaged in another pathway, due to a blocked p62 interface. On the other hand it might give the cell the possibility to actively manage the distribution of TFIIH to the different pathways by overexpressing the respective recruiting factor, which then outcompetes the other recruiting factors.

4.3.3. Organization of the TFIIH core

Based on the conclusions decribed above, the TFIIH core is proposedly subdivided into three functional submodules: the XPD/p44 submodule, the XPB/p52/p8 submodule, and the external interface submodule. P34 constitutes the central scaffold, anchoring the XPD/p44 and the XPB/p52/p8 submodules. A schematic model of this TFIIH core organization is depicted in Figure 4-4.

In principle the XPD/p44 and XPB/p52/p8 submodules are organized into an enzymatic subunit (XPD or XPB) and a second subunit, which serves as a stimulator for the enzymatic subunit and as anchor to the remaining TFIIH core (p44 or p52). The XPB/p52/p8 submodule possesses an additional regulatory protein with p8. The two submodules fulfill different tasks during NER and transcription.

During NER, the XPD/p44 submodule is responsible for DNA unwinding around the lesion via the helicase activity of XPD [22]. The ATPase/helicase activity is stimulated by p44, constituting the regulator of this submodule [30]. During transcription, the XPD/p44

submodule plays an important scaffolding role by anchoring the CAK complex to the TFIIH core. This is achieved by the interaction of MAT1 with XPD [38]. The enzymatic activity of XPD is dispensable during transcription [24]. Furthermore, p44 serves as an anchor for p62 (the external interface submodule) (Figure 4-3) [46].

The role of the XPB/p52/p8 submodule during NER is the anchoring of TFIIH to the repair site through an interaction of XPB with the dsDNA. Interestingly, it has been shown that this anchoring is dependent on the ATPase activity of XPB [15]. In transcription, the XPB/p52/p8 submodule is supposed to be responsible for promotor opening via XPBs enzymatic activity [7]. However, it has also been suggested that XPB is not involved in promotor opening at all and that it solely acts as a regulatory transcriptional block, which is released by its own ATPase activity [342]. In any case, the ATPase activity of XPB is stimulated through p52 and p8, constituting the regulatory subunits of the XPB/p52/p8 submodule [22, 26].

The supposed external interface submodule is constituted by the p62 subunit. This submodule is presumably the major interface to the environment and responsible for recruitment of TFIIH to the different pathways via external recruiting factors as described above. Furthermore, TFIIH can be recruited to specific target gene sites, for example via the interaction of the tumor suppressor p53 with the external interface submodule [69].

Finally, p34 constitutes the central scaffold of the TFIIH core. It interacts with the “anchor proteins” p44 and p52 of the XPD/p44 and XPB/p52/p8 submodules, respectively (Figure 4-3) [48, 80]. Both domains of p34 interact with p52 and p44. It therefore fulfils a key role concerning the structural integrity of TFIIH.

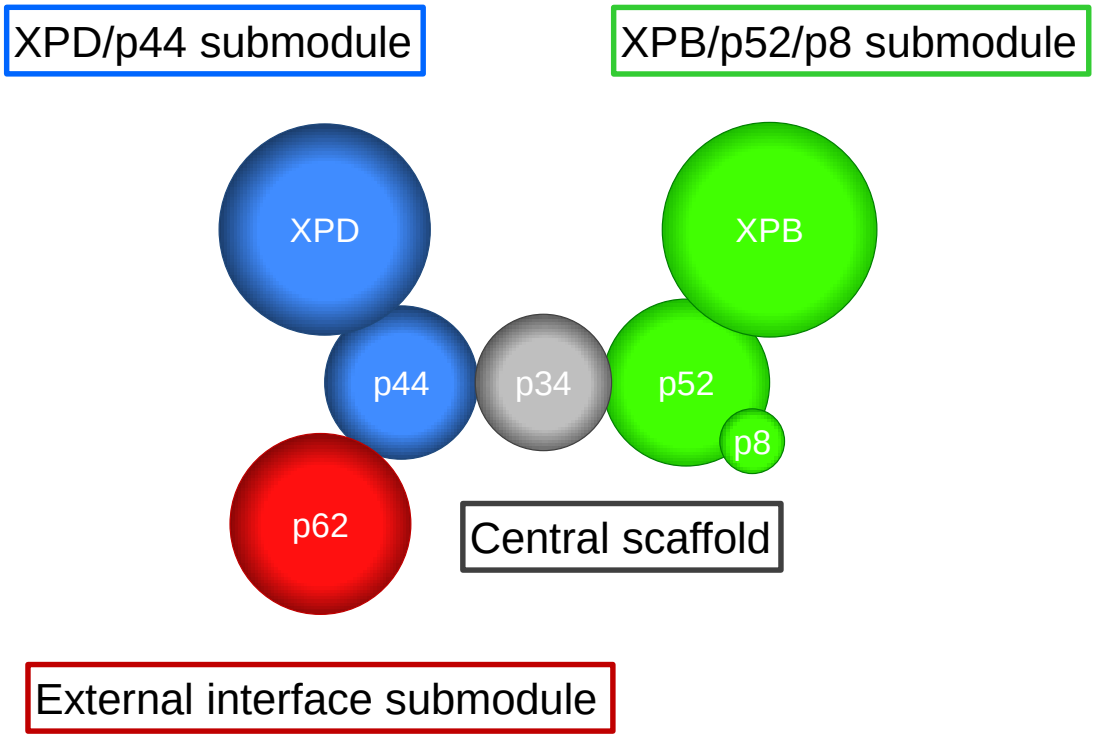


Figure 4-4. Model of the TFIIH core organization. Submodules are indicated by different coloring.

4.4. TFIIH core assembly

4.4.1. Implications for activity assays

TFIIH core assemblies were conducted in low salt conditions, containing either 0.125 M KCl or 0.15 M KAc (see 3.13.3.1, Figure 3-129 and Figure 3-130). In both cases the assembly was severely impaired, and the TFIIH core was seemingly dissociating. This suggests that TFIIH is not stable at low salt concentrations, which could have significant implications for the assessment of the enzymatic activities of TFIIH. So far the enzymatic activities of TFIIH, like the ATPase or helicase activity, have been generally investigated at low salt concentrations containing 0-50 mM KCl (see for example [22, 100, 184, 186, 485]). Considering the results from this work, it is well possible that TFIIH dissociated when these conditions were utilized and that the reported results are not obtained from a complete TFIIH complex but rather from the separate subunits. To investigate and properly address this issue, development of activity assays at higher salt concentrations seem to be compulsory.

4.4.2. Fitting of TFIIH core subunits

The TFIIH core subunits could be fitted to the EM volumes of Core5, Core6, and Core7 (see 3.13.6). However, the limited resolution (~28-36 Å) hampered the interpretation and made the placement of the subunits ambiguous. As a cornerstone the differences between the EM volumes were beneficial to position the XPB and XPD subunits. The interaction network established above (Figure 4-3) was used as a guideline and an interpretation obeying all of these interactions could be achieved. Furthermore, the proposed roles for p34 (see 4.3.1) and p62 (see 4.3.2) were considered. The vWA like domain of p34 should be directly positioned in between p52 and the C-terminal domains of p44. Furthermore, behind the vWA like domain of p34 an additional volume was present for the placement of the C-terminal zinc finger domain of p34. At this position this domain would again be in contact with p52 as well as the C-terminal domains of p44. The proposed role of p62, as described above, is the external interface of TFIIH. As such, it would be peripherally attached, available to external factors. The EM volumes supported a bulge present at the periphery, which was hypothesized to correspond to the position of p62. Additionally, at this position p62 would be located next to the assumed position of the central zinc finger domain of p44, in line with the p62 binding site of p44 that was identified in this work. As stated above, this interpretation has to be taken with some caution due to the limited resolution of the EM maps. To address this issue, higher resolution EM maps of TFIIH core are urgently required.

4.4.3. Comparison with EM map EMD-8131

The EM maps obtained for Core7 were compared with map EMD-8131 [76], the TFIIH core map with the highest resolution available to date. Comparison of the Core7 EM map obtained from negative stain fitted to EMD-8131 is depicted in Figure 4-5. The direct comparison revealed that the overall size and volume of both maps correspond to each other (Figure 4-5 a). However, distinct differences were observed as well. EMD-8131 displays density at the top right, which is absent in the Core7 map. On the other hand, the Core7 map indicates additional density at the bottom left and at the back in form of a protrusion. A comparison of the positions of the fitted subunits from this work with the published model suggests a significantly different position for XPB (Figure 4-5, compare b and c). The Core7 EM map from this work represents the isolated Core7, while EMD-8131 represents Core7 in the context of the PIC. Thus, the different positions of XPB might reflect different conformational states of TFIIH engaged in transcription compared to isolated TFIIH. With respect to EMD-8131 and the fitted subunits, it appears that the EM map does not support clear density for the placement of the p62 subunit (Figure 4-5 c). The Core7 EM map from this work seems to possess additional volume in form of a bulge at the bottom left, enabling the accommodation of p62.

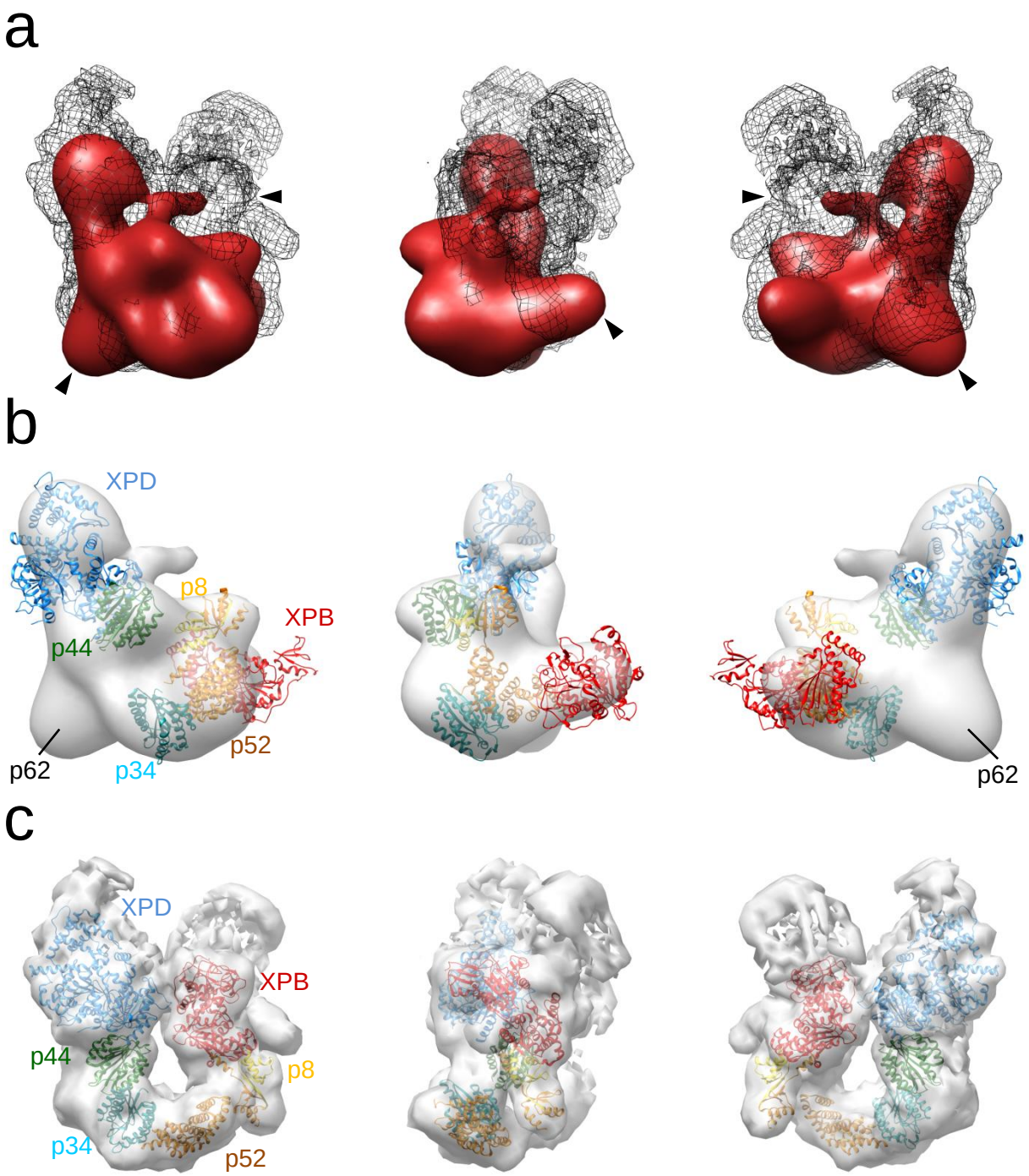


Figure 4-5. Comparison of the negative stain EM map of Core7 with EMD-8131. The front view (left), side view (middle), and back view (right) are shown. **(a)** EM map of Core7 (red surface) fitted to EMD-8131 (black mesh). The most distinct differences between the maps are indicated by arrows. **(b)** EM map of Core7 with the fitted core subunits. The un-occupied space may allow the positioning of the p62 subunit as indicated. **(c)** EM map EMD-8131 with the fitted core subunits (5IVW).

The cryo-EM Core7 map from this work was compared with EMD-8131 as well. As described above, the interpretation of the cryo-EM Core7 map was ambiguous (see 3.13.6, Figure 3-160), with seemingly either XPB or XPD missing in the complex. Due to the limited resolution it could not be distinguished, which subunit might be absent and both possibilities were considered and superimposed with EMD-8131 (Figure 4-6). A reason for the missing

volume in the Core7 cryo-EM map might be the loss of one of these subunits during sample deposition. For the deposition of negatively stained Core7 a crosslinked sample was used, whereas the sample for cryo-EM was deposited without a crosslinker. It can easily be envisioned that the complex without a crosslinker was less stable, resulting in the loss of one of the subunits. As the Core7 complex from negative stain seemingly contained all subunits, whereas the Core7 complex from cryo-EM lacked one subunit, the following discussion was based on the EM maps obtained from negative staining.

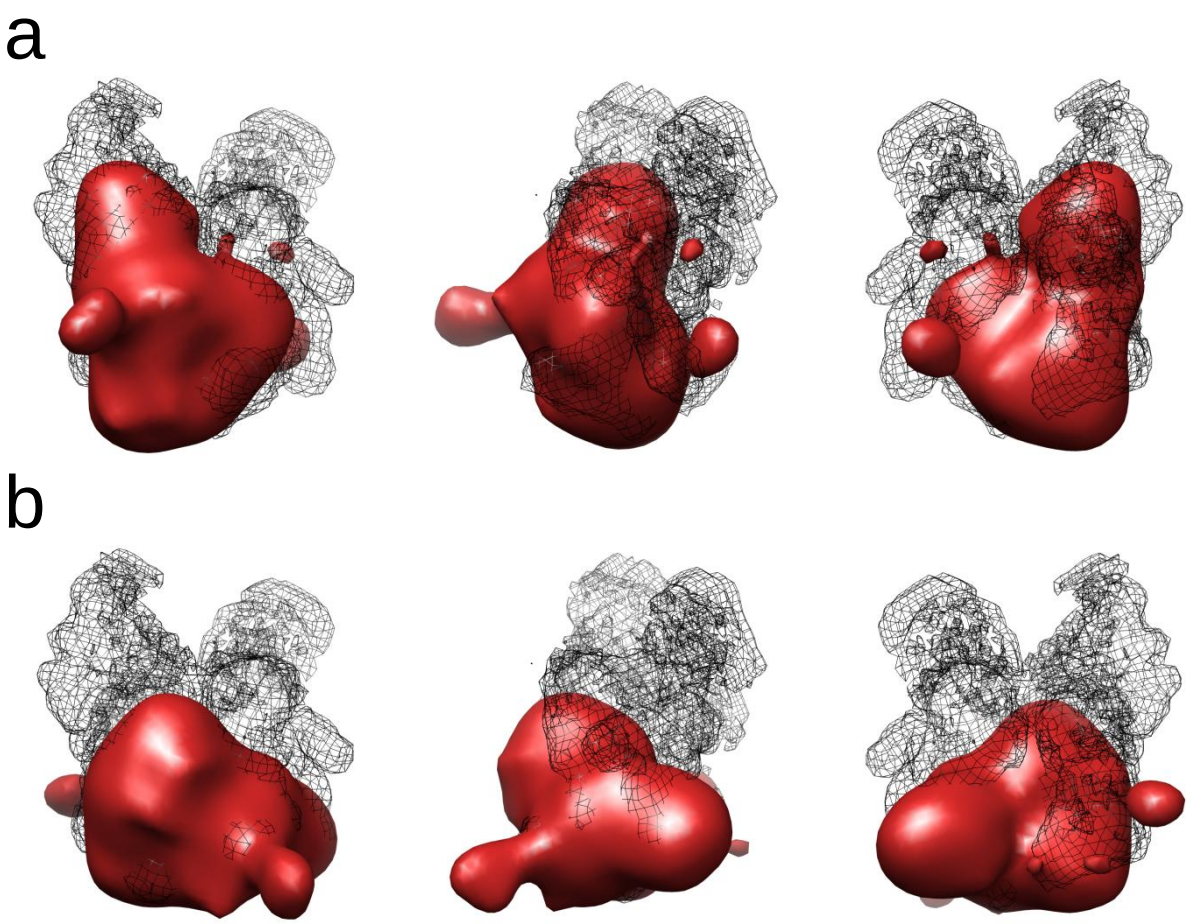


Figure 4-6. Comparison of the cryo-EM map of Core7 with EMD-8131. The front view (left), side view (middle), and back view (right) are shown. Red surface: Core7 cryo-EM map; black mesh: EMD-8131. **(a)** Fit supporting XPD. **(b)** Fit supporting XPB.

4.5. TFIIH conformations

The comparison of the Core7 EM structure obtained from this work with EMD-8131 (Figure 4-5) indicated the existence of different TFIIH conformations. Two conformations might be distinguished, a closed conformation (corresponding to the published structure 5IVW) and a possible open conformation (corresponding to the structure from this work). These hypothesized conformations are schematically depicted in Figure 4-7 a. The major difference between these conformations would be the relative position of XPB and XPD to each other.

In the closed conformation XPB and XPD are in close proximity, whereas in the open conformation XPB and XPD would be much further apart. As the closed conformation was derived from TFIIH incorporated into the PIC [76], it was concluded that TFIIH participates in transcription in a closed conformation. In contrast, the open conformation was derived from an isolated TFIIH core. This poses the question about a possible role of an open conformation and how the transition between these conformations would be triggered. A possible factor for such a conformational change might be the CAK complex. CAK is anchored to TFIIH core via XPD [38]. Furthermore, an interaction between MAT1 and XPB has been suggested [80], and EM studies placed CAK in the proximity of XPB [75]. Accordingly, XPD and XPB might be tethered together by CAK, keeping TFIIH in the closed conformation. This would also be in line with the occurrence of the closed conformation in the PIC, as the CAK is required for transcription [50]. Consequently, the release of CAK might enable the transition to the open conformation. In fact, the open conformation was derived from an isolated TFIIH core where no CAK was present. Following this hypothesis, the suggested open conformation might play a role in the progression of NER, as the CAK is released during NER [185]. In NER, XPD unwinds the DNA around the lesion, creating a bubble [22]. To permit this activity, it has to move along the DNA. However, as XPD is part of the TFIIH core complex during NER, it is restricted in its movement. One possibility is the movement of the entire complex. With respect to XPB, ATPase assays revealed a clear preference of XPB for dsDNA, as the ATPase activity could only be stimulated in the presence of dsDNA (see 3.2.3). Based on these results we conclude that XPB most likely binds to a dsDNA region during NER and does not engage with the ssDNA region of the repair bubble. A plausible scenario is therefore that XPB binds to a defined position upstream of the repair site after recruitment of TFIIH via XPC/HR23B/Centrin2. This engagement with the dsDNA is supported by the findings from the study of Fishburn et al., where the authors showed that SSL2 (XPB) possesses a dsDNA dependent translocase activity [21]. Furthermore, cryo-EM studies of the PIC placed XPB downstream of the actual promotor opening side, in contact with dsDNA at all stages of transcription initiation [76, 78]. At this stage of the NER pathway; i.e. after the damage has been initially recognized and TFIIH has been recruited, the transcription factor would be in a closed conformation with CAK being present. It has been shown that CAK inhibits DNA binding of TFIIH [186], most likely by blocking XPD. The arrival of XPA leads to the dissociation of CAK, thereby enabling XPD to engage the DNA, leading to the activation of XPD [37, 185]. In order to move downstream to unwind the DNA around the lesion, TFIIH would then adopt the open conformation. In this state XPB and XPD would be much further apart, granting XPD the necessary freedom to move downstream and act as a helicase. Presumably, XPB stays bound at a fixed position during this process, still binding to the dsDNA region upstream of the repair bubble. The

energy for the conformational change of TFIIH might well be provided by the ATPase activity of XPB. This model is schematically depicted in Figure 4-7 b.

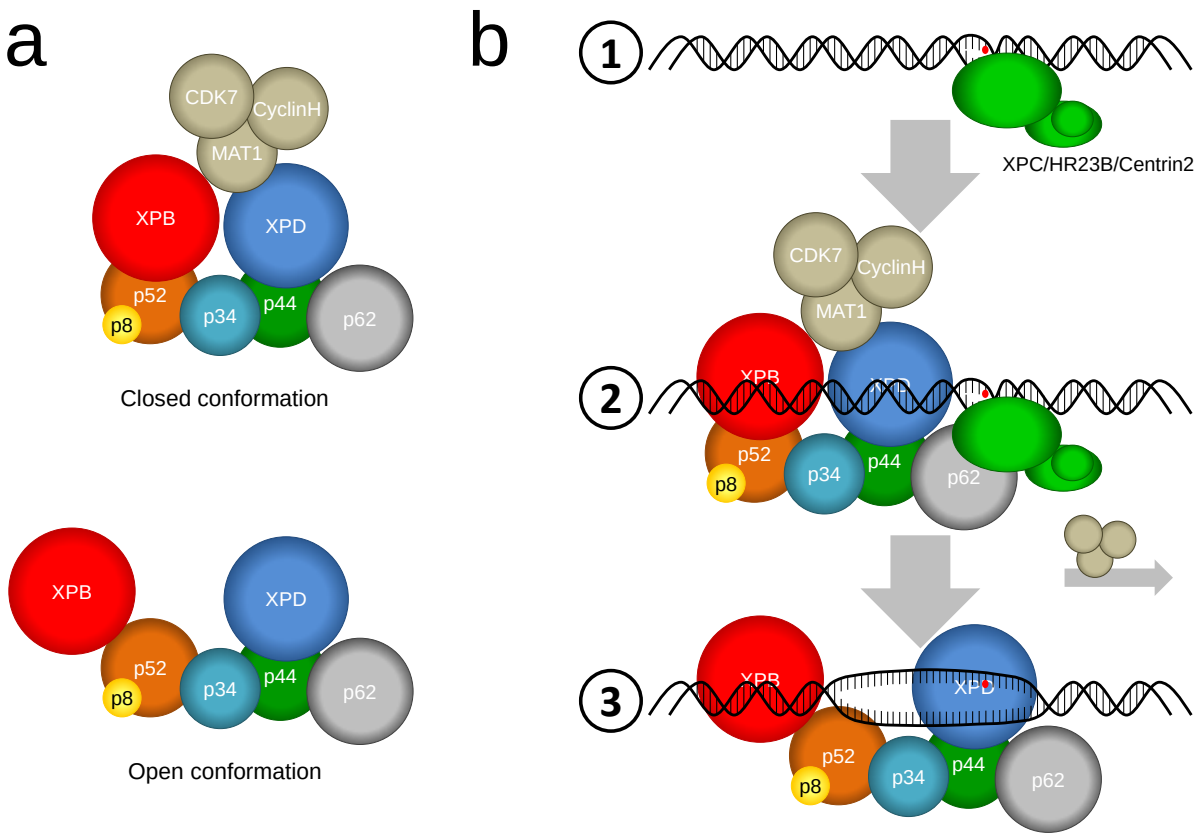


Figure 4-7. Model of the conformational flexibility of TFIIH. (a) Schematic diagram of the closed and the open conformation. **(b)** Proposed role of the different conformations in NER. A DNA lesion is recognized by XPC/HR23B/Centrin2 (1). TFIIH is recruited in its closed conformation via p62. In this conformation XPB assumes a defined position, where it anchors TFIIH to the DNA by binding to the dsDNA upstream of the damage (2). The entry of XPA (not shown) leads to the release of CAK, which permits the transition of the TFIIH core to the open conformation, providing the necessary freedom for XPD to move downstream for DNA unwinding. XPB remains stationary and bound to the dsDNA region upstream of the repair bubble (3).

4.6. Global NER revised

The results from this work were combined with different previously published results and integrated into a unified model to dissect and explain the global NER mechanism. This revised model is depicted in Figure 4-8. Damaged DNA is recognized by the XPC/HR23B/Centrin2 complex, which binds to the unpaired DNA bases on the strand opposite to the lesion [152, 153, 163]. Depending on the lesion this process can be facilitated by UV-DDB [173, 174]. XPC recruits TFIIH in a closed conformation through the interaction with p62 [42, 44, 45], which constitutes the interface submodule of TFIIH. It has been suggested that XPC also interacts with XPB [42]. This interaction may be required to precisely position XPB upstream of the DNA damage. XPB then anchors TFIIH by clamping the DNA with its RED motif and ThM domain [14, 15]. XPA is subsequently recruited,

resulting in the release of the CAK complex [185]. TFIIH transitions into the open conformation, providing the necessary mobility for XPD to unwind the DNA around the lesion. This leads to the activation of XPDs helicase activity [37, 38]. In this model XPB would stay bound to the dsDNA at a fixed position upstream of the repair site. The damage is verified by XPD and/or XPA [66, 198, 205]. The repair bubble is stabilized through binding of RPA to the undamaged single strand [191, 194]. At this stage, the repair bubble is precisely defined by TFIIH. On the one hand, XPB is fixed upstream of the repair bubble close to the 5' end of the damaged strand. On the other hand, p62 is fixed downstream of the repair bubble close to the 3' end of the damaged strand. This arrangement is crucial for the recruitment of XPG. It has been shown that XPG is important for the recruitment of the DNA resynthesis machinery (RFC, PCNA, and DNA polymerase δ, ε, or κ) [233]. Considering the fact that XPG performs the 3' incision within the damaged strand and that the DNA resynthesis machinery has to start synthesizing from the 5' incision site strongly suggests a position of XPG spanning the entire repair bubble from the 5' to the 3' end. It has been suggested that XPG interacts with p62 and XPB [43, 481, 497]. Accordingly, p62 leads XPG to the 3' end and XPB draws XPG to the 5' end of the bubble. The endonuclease XPF-ERCC1 is recruited to the 5' end via XPA [206, 209]. It has been shown that the presence of XPF-ERCC1 hinders the recruitment of the DNA resynthesis machinery [233]. Consequently, XPF-ERCC1 is most likely released directly after performing the 5' incision. Presumably, XPF-ERCC1 stays bound to the 5' end of the damaged strand [225]. At this stage XPG is still bound to the repair site. Release of XPF-ERCC1 makes the 5' site of the repair bubble available for the DNA resynthesis machinery, which is then recruited via XPG and RPA [233]. Most likely, DNA resynthesis is initiated before the 3' incision and with the NER machinery still in place to avoid the exposure of potentially dangerous ssDNA gaps [213]. Finally, XPG performs the 3' incision and is released together with the excised strand and TFIIH [217, 225, 226], and the single stranded gap is refilled by the DNA polymerase [229].

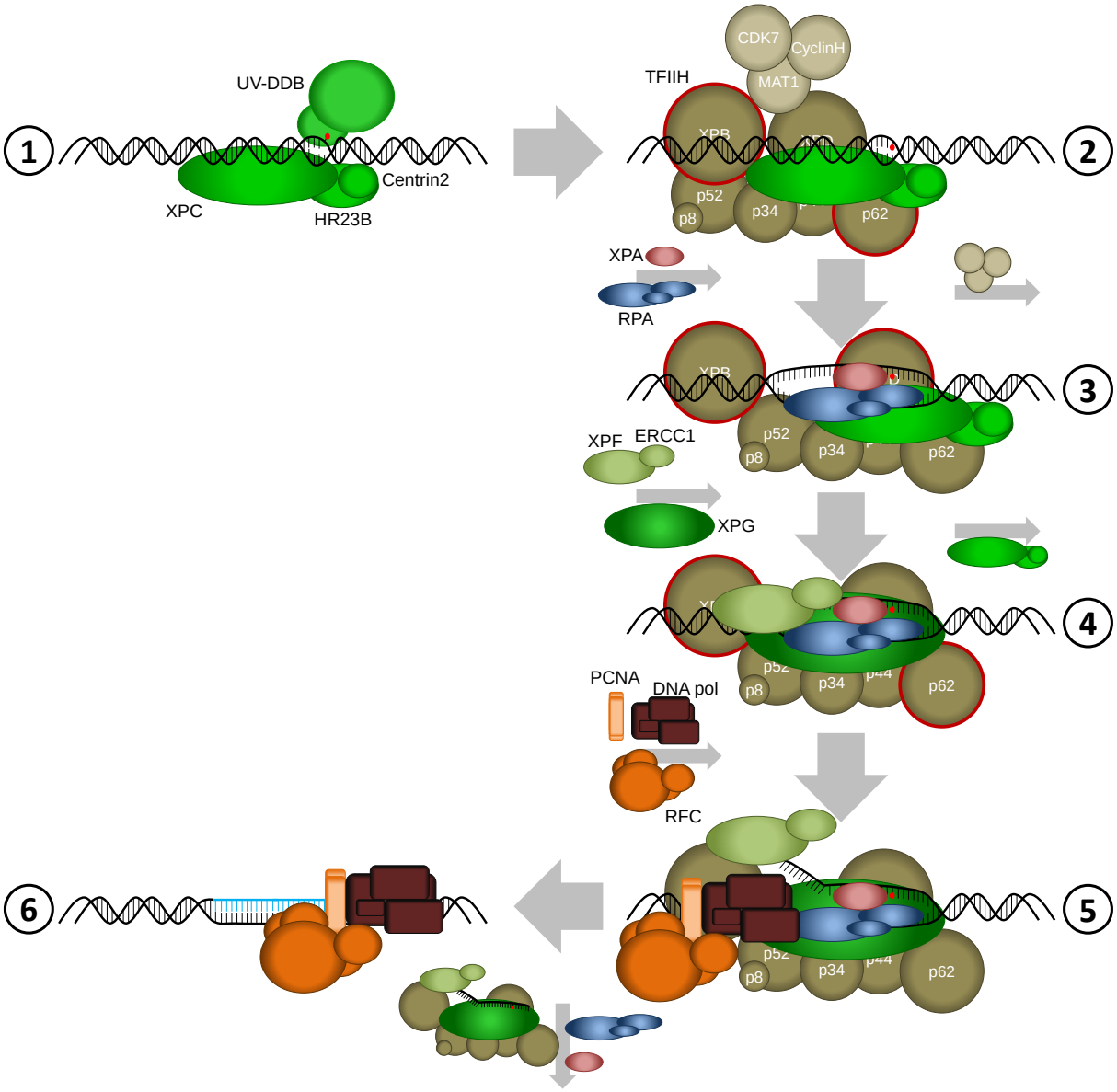


Figure 4-8. Revised model of the global NER mechanism. TFIIH subunits important for a particular step are highlighted. Damaged DNA is recognized by XPC/HR23B/Centrin2 with or without the help of UV-DDB (1). TFIIH in the closed conformation is recruited by XPC via p62. XPB is properly positioned upstream of the damaged site via interaction with XPC (2). XPA and RPA are recruited, leading to the release of CAK. Furthermore, TFIIH adopts an open conformation, enabling XPD to unwind the DNA around the damage (3). The endonucleases XPG and XPF-ERCC1 are recruited. XPG is properly positioned via interaction with p62 and XPB, thereby spanning the entire repair bubble (4). XPF-ERCC1 makes an incision 5' of the lesion into the damaged strand and is released. RFC, PCNA, and DNA polymerase δ, ε, or κ are recruited via XPG and RPA. Subsequently, XPG makes an incision 3' to the lesion (5). The excised strand is released bound by TFIIH, XPG, and XPF-ERCC1 and the single stranded gap is filled by DNA synthesis (6).

5. References

1. Compe, E. and J.M. Egly, *Nucleotide Excision Repair and Transcriptional Regulation: TFIIH and Beyond.* Annual Review of Biochemistry, Vol 85, 2016. **85**: p. 265-290.
2. Lehmann, A.R., *DNA repair-deficient diseases, xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy.* Biochimie, 2003. **85**(11): p. 1101-1111.
3. Giglia-Mari, G., et al., *A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A.* Nat Genet, 2004. **36**(7): p. 714-9.
4. Compe, E. and J.M. Egly, *TFIIH: when transcription met DNA repair.* Nature Reviews Molecular Cell Biology, 2012. **13**(6): p. 343-354.
5. Adamczewski, J.P., et al., *MAT1, cdk7 and cyclin H form a kinase complex which is UV light-sensitive upon association with TFIIH.* Embo j, 1996. **15**(8): p. 1877-84.
6. Rossignol, M., I. Kolb-Cheynel, and J.M. Egly, *Substrate specificity of the cdk-activating kinase (CAK) is altered upon association with TFIIH.* Embo j, 1997. **16**(7): p. 1628-37.
7. Tirode, F., et al., *Reconstitution of the transcription factor TFIIH: Assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7.* Molecular Cell, 1999. **3**(1): p. 87-95.
8. Weeda, G., et al., *A presumed DNA helicase encoded by ERCC-3 is involved in the human repair disorders xeroderma pigmentosum and Cockayne's syndrome.* Cell, 1990. **62**(4): p. 777-91.
9. Fan, L. and K.T. DuPrez, *XPB: An unconventional SF2 DNA helicase.* Progress in Biophysics & Molecular Biology, 2015. **117**(2-3): p. 174-181.
10. White, M.F., *Structure, function and evolution of the XPD family of iron-sulfur-containing 5 '-> 3 ' DNA helicases.* Biochemical Society Transactions, 2009. **37**: p. 547-551.
11. Byrd, A.K. and K.D. Raney, *Superfamily 2 helicases.* Front Biosci (Landmark Ed), 2012. **17**: p. 2070-88.
12. Gorbalenya, A.E. and E.V. Koonin, *Helicases: amino acid sequence comparisons and structure-function relationships.* Current Opinion in Structural Biology, 1993. **3**(3): p. 419-429.
13. Singleton, M.R., M.S. Dillingham, and D.B. Wigley, *Structure and mechanism of helicases and nucleic acid translocases.* Annu Rev Biochem, 2007. **76**: p. 23-50.
14. Fan, L., et al., *Conserved XPB core structure and motifs for DNA unwinding: Implications for pathway selection of transcription or excision repair.* Molecular Cell, 2006. **22**(1): p. 27-37.
15. Oksenych, V., et al., *Molecular insights into the recruitment of TFIIH to sites of DNA damage.* Embo Journal, 2009. **28**(19): p. 2971-2980.
16. Fuss, J.O. and J.A. Tainer, *XPB and XPD helicases in TFIIH orchestrate DNA duplex opening and damage verification to coordinate repair with transcription and cell cycle via CAK kinase.* DNA Repair, 2011. **10**(7): p. 697-713.
17. Oksenych, V. and F. Coin, *The long unwinding road XPB and XPD helicases in damaged DNA opening.* Cell Cycle, 2010. **9**(1): p. 90-96.
18. Fan, L., et al., *XPD helicase structures and activities: Insights into the cancer and aging phenotypes from XPD mutations.* Cell, 2008. **133**(5): p. 789-800.
19. Liu, H., et al., *Structure of the DNA repair helicase XPD.* Cell, 2008. **133**(5): p. 801-812.
20. Wolski, S.C., et al., *Crystal structure of the FeS cluster-containing nucleotide excision repair helicase XPD.* Plos Biology, 2008. **6**(6): p. 1332-1342.
21. Fishburn, J., et al., *Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation.* Proceedings of the National Academy of Sciences of the United States of America, 2015. **112**(13): p. 3961-3966.

22. Coin, F., V. Oksenych, and J.M. Egly, *Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair.* Molecular Cell, 2007. **26**(2): p. 245-256.
23. Winkler, G.S., et al., *TFIIH with inactive XPD helicase functions in transcription initiation but is defective in DNA repair.* J Biol Chem, 2000. **275**(6): p. 4258-66.
24. Kuper, J., et al., *In TFIIH, XPD Helicase Is Exclusively Devoted to DNA Repair.* Plos Biology, 2014. **12**(9).
25. Jawhari, A., et al., *p52 mediates XPB function within the transcription/repair factor TFIIH.* Journal of Biological Chemistry, 2002. **277**(35): p. 31761-31767.
26. Coin, F., et al., *P8/TTD-A as a repair-specific TFIIH subunit.* Molecular Cell, 2006. **21**(2): p. 215-226.
27. Vitorino, M., et al., *Solution structure and self-association properties of the p8 TFIIH subunit responsible for trichothiodystrophy.* Journal of Molecular Biology, 2007. **368**(2): p. 473-480.
28. Kainov, D.E., et al., *Structural basis for group A trichothiodystrophy.* Nature Structural & Molecular Biology, 2008. **15**(9): p. 980-984.
29. Zhou, Y., H.P. Kou, and Z.G. Wang, *Tfb5 interacts with Tfb2 and facilitates nucleotide excision repair in yeast.* Nucleic Acids Research, 2007. **35**(3): p. 861-871.
30. Coin, F., et al., *Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH.* Nature Genetics, 1998. **20**(2): p. 184-188.
31. Humbert, S., et al., *p44 and p34 subunits of the BTF2/TFIIH transcription factor have homologies with SSL1, a yeast protein involved in DNA repair.* Embo j, 1994. **13**(10): p. 2393-8.
32. Fribourg, S., et al., *Structural characterization of the cysteine-rich domain of TFIIH p44 subunit.* J Biol Chem, 2000. **275**(41): p. 31963-71.
33. Tremeau-Bravard, A., C. Perez, and J.M. Egly, *A role of the C-terminal part of p44 in the promoter escape activity of transcription factor IIH.* J Biol Chem, 2001. **276**(29): p. 27693-7.
34. Kellenberger, E., et al., *Solution structure of the C-terminal domain of TFIIH p44 subunit reveals a novel type of C4C4 ring domain involved in protein-protein interactions.* Journal of Biological Chemistry, 2005. **280**(21): p. 20785-20792.
35. Kim, J.S., et al., *Crystal Structure of the Rad3/XPD Regulatory Domain of Ssl1/p44.* Journal of Biological Chemistry, 2015. **290**(13): p. 8321-8330.
36. Takagi, Y., et al., *Ubiquitin ligase activity of TFIIH and the transcriptional response to DNA damage.* Mol Cell, 2005. **18**(2): p. 237-43.
37. Sandrock, B. and J.M. Egly, *A yeast four-hybrid system identifies Cdk-activating kinase as a regulator of the XPD helicase, a subunit of transcription factor IIH.* Journal of Biological Chemistry, 2001. **276**(38): p. 35328-35333.
38. Abdulrahman, W., et al., *ARCH domain of XPD, an anchoring platform for CAK that conditions TFIIH DNA repair and transcription activities.* Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(8): p. E633-E642.
39. Doerks, T., et al., *BSD: a novel domain in transcription factors and synapse-associated proteins.* Trends Biochem Sci, 2002. **27**(4): p. 168-70.
40. Jawhari, A., et al., *Domain architecture of the p62 subunit from the human transcription/repair factor TFIIH deduced by limited proteolysis and mass spectrometry analysis.* Biochemistry, 2004. **43**(45): p. 14420-14430.
41. Gervais, V., et al., *TFIIH contains a PH domain involved in DNA nucleotide excision repair.* Nature Structural & Molecular Biology, 2004. **11**(7): p. 616-622.
42. de Jesus, B.M.B., et al., *Dissection of the Molecular Defects Caused by Pathogenic Mutations in the DNA Repair Factor XPC.* Molecular and Cellular Biology, 2008. **28**(23): p. 7225-7235.
43. Lafrance-Vanasse, J., et al., *Structural and functional characterization of interactions involving the Tfb1 subunit of TFIIH and the NER factor Rad2.* Nucleic Acids Research, 2012. **40**(12): p. 5739-5750.

44. Lafrance-Vanasse, J., et al., *Structural and functional evidence that Rad4 competes with Rad2 for binding to the Tfb1 subunit of TFIIH in NER.* Nucleic Acids Research, 2013. **41**(4): p. 2736-2745.
45. Okuda, M., et al., *Structural Insight into the Mechanism of TFIIH Recognition by the Acidic String of the Nucleotide Excision Repair Factor XPC.* Structure, 2015. **23**(10): p. 1827-1837.
46. Matsui, P., J. DePaulo, and S. Buratowski, *An interaction between the Tfb1 and Ssl1 subunits of yeast TFIIH correlates with DNA repair activity.* Nucleic Acids Res, 1995. **23**(5): p. 767-72.
47. Schmitt, D.R., et al., *The Structure of the TFIIH p34 Subunit Reveals a Von Willebrand Factor A Like Fold.* Plos One, 2014. **9**(7).
48. Fribourg, S., et al., *Dissecting the interaction network of multiprotein complexes by pairwise coexpression of subunits in E. coli.* J Mol Biol, 2001. **306**(2): p. 363-73.
49. Fisher, R.P., *Secrets of a double agent: CDK7 in cell-cycle control and transcription.* Journal of Cell Science, 2005. **118**(22): p. 5171-5180.
50. Glover-Cutter, K., et al., *TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II.* Mol Cell Biol, 2009. **29**(20): p. 5455-64.
51. Ganuza, M., et al., *Genetic inactivation of Cdk7 leads to cell cycle arrest and induces premature aging due to adult stem cell exhaustion.* Embo j, 2012. **31**(11): p. 2498-510.
52. Lolli, G., et al., *The crystal structure of human CDK7 and its protein recognition properties.* Structure, 2004. **12**(11): p. 2067-79.
53. Malumbres, M., *Cyclin-dependent kinases.* Genome Biol, 2014. **15**(6): p. 122.
54. Hanks, S.K. and T. Hunter, *Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification.* Faseb j, 1995. **9**(8): p. 576-96.
55. Fisher, R.P. and D.O. Morgan, *A novel cyclin associates with MO15/CDK7 to form the CDK-activating kinase.* Cell, 1994. **78**(4): p. 713-24.
56. Andersen, G., et al., *The structure of cyclin H: common mode of kinase activation and specific features.* Embo j, 1997. **16**(5): p. 958-67.
57. Andersen, G., et al., *The crystal structure of human cyclin H.* FEBS Lett, 1996. **397**(1): p. 65-9.
58. Kim, K.K., et al., *Three-dimensional structure of human cyclin H, a positive regulator of the CDK-activating kinase.* Nat Struct Biol, 1996. **3**(10): p. 849-55.
59. Busso, D., et al., *Distinct regions of MAT1 regulate cdk7 kinase and TFIIH transcription activities.* J Biol Chem, 2000. **275**(30): p. 22815-23.
60. Gervais, V., et al., *Solution structure of the N-terminal domain of the human TFIIH MAT1 subunit: new insights into the RING finger family.* J Biol Chem, 2001. **276**(10): p. 7457-64.
61. Sainsbury, S., C. Bernecky, and P. Cramer, *Structural basis of transcription initiation by RNA polymerase II.* Nature Reviews Molecular Cell Biology, 2015. **16**(3): p. 129-143.
62. Hilario, E., et al., *Structure of the C-terminal half of human XPB helicase and the impact of the disease-causing mutation XP11BE.* Acta Crystallographica Section D-Biological Crystallography, 2013. **69**: p. 237-246.
63. Okuda, M., et al., *Structural insight into the TFIIE-TFIIH interaction: TFIIE and p53 share the binding region on TFIIH.* Embo Journal, 2008. **27**(7): p. 1161-1171.
64. Okuda, M. and Y. Nishimura, *Extended String Binding Mode of the Phosphorylated Transactivation Domain of Tumor Suppressor p53.* Journal of the American Chemical Society, 2014. **136**(40): p. 14143-14152.
65. Okuda, M., et al., *The Interaction Mode of the Acidic Region of the Cell Cycle Transcription Factor DP1 with TFIIH.* Journal of Molecular Biology, 2016. **428**(24): p. 4993-5006.

66. Kuper, J., et al., *Functional and structural studies of the nucleotide excision repair helicase XPD suggest a polarity for DNA translocation.* Embo Journal, 2012. **31**(2): p. 494-502.
67. Constantinescu-Aruxandei, D., et al., *Mechanism of DNA loading by the DNA repair helicase XPD.* Nucleic Acids Research, 2016. **44**(6): p. 2806-2815.
68. Di Lello, P., et al., *NMR structure of the amino-terminal domain from the Tfb1 subunit of TFIIH and characterization of its phosphoinositide and VP16 binding sites.* Biochemistry, 2005. **44**(21): p. 7678-7686.
69. Di Lello, P., et al., *Structure of the Tfb1/p53 complex: Insights into the interaction between the p62/Tfb1 subunit of TFIIH and the activation domain of p53.* Molecular Cell, 2006. **22**(6): p. 731-740.
70. Langlois, C., et al., *NMR structure of the complex between the Tfb1 subunit of TFIIH and the activation domain of VP16: Structural similarities between VP16 and p53.* Journal of the American Chemical Society, 2008. **130**(32): p. 10596-10604.
71. Chabot, P.R., et al., *Structural and Functional Characterization of a Complex between the Acidic Transactivation Domain of EBNA2 and the Tfb1/p62 Subunit of TFIIH.* Plos Pathogens, 2014. **10**(3).
72. Cyr, N., et al., *A Omega XaV motif in the Rift Valley fever virus NSs protein is essential for degrading p62, forming nuclear filaments and virulence.* Proceedings of the National Academy of Sciences of the United States of America, 2015. **112**(19): p. 6021-6026.
73. Lecoq, L., et al., *Structural characterization of interactions between transactivation domain 1 of the p65 subunit of NF-kappaB and transcription regulatory factors.* Nucleic Acids Res, 2017.
74. Schultz, P., et al., *Molecular structure of human TFIIH.* Cell, 2000. **102**(5): p. 599-607.
75. Gibbons, B.J., et al., *Subunit architecture of general transcription factor TFIIH.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(6): p. 1949-1954.
76. He, Y., et al., *Near-atomic resolution visualization of human transcription promoter opening.* Nature, 2016. **533**(7603): p. 359-65.
77. Robinson, P.J., et al., *Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex.* Cell, 2016. **166**(6): p. 1411-1422.
78. Murakami, K., et al., *Structure of an RNA polymerase II preinitiation complex.* Proc Natl Acad Sci U S A, 2015. **112**(44): p. 13543-8.
79. Egly, J.M. and F. Coin, *A history of TFIIH: Two decades of molecular biology on a pivotal transcription/repair factor.* DNA Repair, 2011. **10**(7): p. 714-721.
80. Luo, J., et al., *Architecture of the Human and Yeast General Transcription and DNA Repair Factor TFIIH.* Molecular Cell, 2015. **59**(5): p. 794-806.
81. Diderich, K., M. Alanazi, and J.H.J. Hoeijmakers, *Premature aging and cancer in nucleotide excision repair-disorders.* DNA Repair, 2011. **10**(7): p. 772-780.
82. Hebra, F. and M. Kaposi, *Xeroderma*, in *On diseases of the skin, including the exanthemata*, W. Tay, Editor. 1874, The New Sydenham Society: London. p. 252-258.
83. DiGiovanna, J.J. and K.H. Kraemer, *Shining a Light on Xeroderma Pigmentosum.* Journal of Investigative Dermatology, 2012. **132**(3): p. 785-796.
84. Cleaver, J.E., *Defective Repair Replication of DNA in Xeroderma Pigmentosum.* Nature, 1968. **218**(5142): p. 652-656.
85. Moncollin, V., et al., *Purification of a factor specific for the upstream element of the adenovirus-2 major late promoter.* Embo j, 1986. **5**(10): p. 2577-84.
86. Weber, C.A., et al., *Molecular cloning and biological characterization of a human gene, ERCC2, that corrects the nucleotide excision repair defect in CHO UV5 cells.* Mol Cell Biol, 1988. **8**(3): p. 1137-46.
87. Conaway, R.C. and J.W. Conaway, *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, 1989. **86**(19): p. 7356-60.

88. Sung, P., et al., *RAD3 protein of Saccharomyces cerevisiae is a DNA helicase.* Proc Natl Acad Sci U S A, 1987. **84**(24): p. 8951-5.
89. Sung, P., et al., *The RAD3 gene of Saccharomyces cerevisiae encodes a DNA-dependent ATPase.* Proc Natl Acad Sci U S A, 1987. **84**(17): p. 6045-9.
90. Weber, C.A., et al., *ERCC2: cDNA cloning and molecular characterization of a human nucleotide excision repair gene with high homology to yeast RAD3.* Embo j, 1990. **9**(5): p. 1437-47.
91. Weeda, G., et al., *Molecular cloning and biological characterization of the human excision repair gene ERCC-3.* Mol Cell Biol, 1990. **10**(6): p. 2570-81.
92. Feaver, W.J., O. Gileadi, and R.D. Kornberg, *Purification and characterization of yeast RNA polymerase II transcription factor b.* J Biol Chem, 1991. **266**(28): p. 19000-5.
93. Gerard, M., et al., *Purification and interaction properties of the human RNA polymerase B(II) general transcription factor BTF2.* J Biol Chem, 1991. **266**(31): p. 20940-5.
94. Feaver, W.J., et al., *CTD kinase associated with yeast RNA polymerase II initiation factor b.* Cell, 1991. **67**(6): p. 1223-30.
95. Flores, O., H. Lu, and D. Reinberg, *Factors involved in specific transcription by mammalian RNA polymerase II. Identification and characterization of factor IIH.* J Biol Chem, 1992. **267**(4): p. 2786-93.
96. Serizawa, H., R.C. Conaway, and J.W. Conaway, *A carboxyl-terminal-domain kinase associated with RNA polymerase II transcription factor delta from rat liver.* Proc Natl Acad Sci U S A, 1992. **89**(16): p. 7476-80.
97. Lu, H., et al., *Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II.* Nature, 1992. **358**(6388): p. 641-5.
98. Fischer, L., et al., *Cloning of the 62-kilodalton component of basic transcription factor BTF2.* Science, 1992. **257**(5075): p. 1392-5.
99. Schaeffer, L., et al., *DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor.* Science, 1993. **260**(5104): p. 58-63.
100. Schaeffer, L., et al., *The ERCC2/DNA repair protein is associated with the class II BTF2/TFIIH transcription factor.* Embo j, 1994. **13**(10): p. 2388-92.
101. Roy, R., et al., *The MO15 cell cycle kinase is associated with the TFIIH transcription-DNA repair factor.* Cell, 1994. **79**(6): p. 1093-101.
102. Marinoni, J.C., et al., *Cloning and characterization of p52, the fifth subunit of the core of the transcription/DNA repair factor TFIIH.* Embo j, 1997. **16**(5): p. 1093-102.
103. Ranish, J.A., et al., *Identification of TFB5, a new component of general transcription and DNA repair factor IIH.* Nat Genet, 2004. **36**(7): p. 707-13.
104. Berg, J.M., J.L. Tymoczko, and L. Stryer, *DNA, RNA, and the Flow of Genetic Information*, in *Biochemistry*. 2012, W. H. Freeman and Company. p. 113-143.
105. Garfinkel, L. and S.D. Stellman, *Smoking and lung cancer in women: findings in a prospective study.* Cancer Res, 1988. **48**(23): p. 6951-5.
106. Wang, J.S. and J.D. Groopman, *DNA damage by mycotoxins.* Mutat Res, 1999. **424**(1-2): p. 167-81.
107. D'Orazio, J., et al., *UV radiation and the skin.* Int J Mol Sci, 2013. **14**(6): p. 12222-48.
108. Hoeijmakers, J.H.J., *Genome maintenance mechanisms for preventing cancer.* Nature, 2001. **411**(6835): p. 366-374.
109. Cadet, J., et al., *Oxidative damage to DNA: formation, measurement, and biological significance.* Rev Physiol Biochem Pharmacol, 1997. **131**: p. 1-87.
110. Lindahl, T., *Instability and decay of the primary structure of DNA.* Nature, 1993. **362**(6422): p. 709-15.
111. Beckman, K.B. and B.N. Ames, *Oxidative decay of DNA.* J Biol Chem, 1997. **272**(32): p. 19633-6.
112. Swenberg, J.A., et al., *Endogenous versus Exogenous DNA Adducts: Their Role in Carcinogenesis, Epidemiology, and Risk Assessment.* Toxicological Sciences, 2011. **120**: p. S130-S145.
113. Friedberg, E.C., *DNA damage and repair.* Nature, 2003. **421**(6921): p. 436-40.

114. Krokan, H.E., et al., *Base excision repair of DNA in mammalian cells.* FEBS Lett, 2000. **476**(1-2): p. 73-7.
115. Fortini, P., et al., *The base excision repair: mechanisms and its relevance for cancer susceptibility.* Biochimie, 2003. **85**(11): p. 1053-71.
116. Moldovan, G.L. and A.D. D'Andrea, *How the fanconi anemia pathway guards the genome.* Annu Rev Genet, 2009. **43**: p. 223-49.
117. Ho, T.V. and O.D. Scharer, *Translesion DNA synthesis polymerases in DNA interstrand crosslink repair.* Environ Mol Mutagen, 2010. **51**(6): p. 552-66.
118. Muniandy, P.A., et al., *DNA interstrand crosslink repair in mammalian cells: step by step.* Crit Rev Biochem Mol Biol, 2010. **45**(1): p. 23-49.
119. van Gent, D.C., J.H. Hoeijmakers, and R. Kanaar, *Chromosomal stability and the DNA double-stranded break connection.* Nat Rev Genet, 2001. **2**(3): p. 196-206.
120. Valerie, K. and L.F. Povirk, *Regulation and mechanisms of mammalian double-strand break repair.* Oncogene, 2003. **22**(37): p. 5792-812.
121. Jiricny, J., *The multifaceted mismatch-repair system.* Nat Rev Mol Cell Biol, 2006. **7**(5): p. 335-46.
122. Fishel, R., *Mismatch repair.* J Biol Chem, 2015. **290**(44): p. 26395-403.
123. Gillet, L.C.J. and O.D. Scharer, *Molecular mechanisms of mammalian global genome nucleotide excision repair.* Chemical Reviews, 2006. **106**(2): p. 253-276.
124. Hubscher, U. and S. Spadari, *DNA replication and chemotherapy.* Physiol Rev, 1994. **74**(2): p. 259-304.
125. Friedberg, E.C., A.R. Lehmann, and R.P. Fuchs, *Trading places: how do DNA polymerases switch during translesion DNA synthesis?* Mol Cell, 2005. **18**(5): p. 499-505.
126. Yang, W. and R. Woodgate, *What a difference a decade makes: insights into translesion DNA synthesis.* Proc Natl Acad Sci U S A, 2007. **104**(40): p. 15591-8.
127. Goodman, M.F. and R. Woodgate, *Translesion DNA polymerases.* Cold Spring Harb Perspect Biol, 2013. **5**(10): p. a010363.
128. Nair, D.T., et al., *Replication by human DNA polymerase-iota occurs by Hoogsteen base-pairing.* Nature, 2004. **430**(6997): p. 377-80.
129. Nair, D.T., et al., *Rev1 employs a novel mechanism of DNA synthesis using a protein template.* Science, 2005. **309**(5744): p. 2219-22.
130. Nair, D.T., et al., *Human DNA polymerase iota incorporates dCTP opposite template G via a G.C + Hoogsteen base pair.* Structure, 2005. **13**(10): p. 1569-77.
131. Lone, S., et al., *Human DNA polymerase kappa encircles DNA: implications for mismatch extension and lesion bypass.* Mol Cell, 2007. **25**(4): p. 601-14.
132. Biertumpfel, C., et al., *Structure and mechanism of human DNA polymerase eta.* Nature, 2010. **465**(7301): p. 1044-8.
133. Prakash, S., R.E. Johnson, and L. Prakash, *Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function.* Annu Rev Biochem, 2005. **74**: p. 317-53.
134. Kuraoka, I., et al., *Effects of endogenous DNA base lesions on transcription elongation by mammalian RNA polymerase II - Implications for transcription-coupled DNA repair and transcriptional mutagenesis.* Journal of Biological Chemistry, 2003. **278**(9): p. 7294-7299.
135. Kathe, S.D., G.P. Shen, and S.S. Wallace, *Single-stranded breaks in DNA but not oxidative DNA base damages block transcriptional elongation by RNA polymerase II in HeLa cell nuclear extracts.* Journal of Biological Chemistry, 2004. **279**(18): p. 18511-18520.
136. Saxowsky, T.T. and P.W. Doetsch, *RNA polymerase encounters with DNA damage: Transcription-coupled repair or transcriptional mutagenesis?* Chemical Reviews, 2006. **106**(2): p. 474-488.
137. Marietta, C. and P.J. Brooks, *Transcriptional bypass of bulky DNA lesions causes new mutant RNA transcripts in human cells.* Embo Reports, 2007. **8**(4): p. 388-393.
138. Spivak, G. and A.K. Ganesan, *The complex choreography of transcription-coupled repair.* DNA Repair, 2014. **19**: p. 64-70.

139. Doetsch, P.W., *Translesion synthesis by RNA polymerases: occurrence and biological implications for transcriptional mutagenesis.* Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis, 2002. **510**(1-2): p. 131-140.
140. Saxowsky, T.T., et al., *8-Oxoguanine-mediated transcriptional mutagenesis causes Ras activation in mammalian cells.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(48): p. 18877-18882.
141. Mitchell, J.R., J.H.J. Hoeijmakers, and L.J. Niedernhofer, *Divide and conquer: nucleotide excision repair battles cancer and ageing.* Current Opinion in Cell Biology, 2003. **15**(2): p. 232-240.
142. Duensing, A., et al., *A role of the mitotic spindle checkpoint in the cellular response to DNA replication stress.* Journal of Cellular Biochemistry, 2006. **99**(3): p. 759-769.
143. Rodriguez, R. and M. Meuth, *Chk1 and p21 cooperate to prevent apoptosis during DNA replication fork stress.* Molecular Biology of the Cell, 2006. **17**(1): p. 402-412.
144. Hoeijmakers, J.H., *DNA damage, aging, and cancer.* N Engl J Med, 2009. **361**(15): p. 1475-85.
145. Ljungman, M., et al., *Inhibition of RNA polymerase II as a trigger for the p53 response.* Oncogene, 1999. **18**(3): p. 583-592.
146. Ljungman, M. and D.P. Lane, *Opinion - Transcription - guarding the genome by sensing DNA damage.* Nature Reviews Cancer, 2004. **4**(9): p. 727-737.
147. Latonen, L. and M. Laiho, *Cellular UV damage responses - Functions of tumor suppressor p53.* Biochimica Et Biophysica Acta-Reviews on Cancer, 2005. **1755**(2): p. 71-89.
148. Marteijn, J.A., et al., *Understanding nucleotide excision repair and its roles in cancer and ageing.* Nature Reviews Molecular Cell Biology, 2014. **15**(7): p. 465-481.
149. Scharer, O.D., *Nucleotide Excision Repair in Eukaryotes.* Cold Spring Harbor Perspectives in Biology, 2013. **5**(10).
150. Hanawalt, P.C. and G. Spivak, *Transcription-coupled DNA repair: two decades of progress and surprises.* Nature Reviews Molecular Cell Biology, 2008. **9**(12): p. 958-970.
151. Masutani, C., et al., *Purification and Cloning of a Nucleotide Excision-Repair Complex Involving the Xeroderma-Pigmentosum Group-C Protein and a Human Homolog of Yeast Rad23.* Embo Journal, 1994. **13**(8): p. 1831-1843.
152. Sugasawa, K., et al., *Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair.* Molecular Cell, 1998. **2**(2): p. 223-232.
153. Araki, M., et al., *Centrosome protein centrin 2/caltractin 1 is part of the xeroderma pigmentosum group C complex that initiates global genome nucleotide excision repair.* Journal of Biological Chemistry, 2001. **276**(22): p. 18665-18672.
154. Sugasawa, K., et al., *A multistep damage recognition mechanism for global genomic nucleotide excision repair.* Genes & Development, 2001. **15**(5): p. 507-521.
155. Reardon, J.T., D. Mu, and A. Sancar, *Overproduction, purification, and characterization of the XPC subunit of the human DNA repair excision nuclease.* Journal of Biological Chemistry, 1996. **271**(32): p. 19451-19456.
156. Sugasawa, K., et al., *HHR23B, a human Rad23 homolog, stimulates XPC protein in nucleotide excision repair in vitro.* Molecular and Cellular Biology, 1996. **16**(9): p. 4852-4861.
157. Sugasawa, K., et al., *Two human homologs of Rad23 are functionally interchangeable in complex formation and stimulation of XPC repair activity.* Molecular and Cellular Biology, 1997. **17**(12): p. 6924-6931.
158. Nishi, R., et al., *Centrin 2 stimulates nucleotide excision repair by interacting with xeroderma pigmentosum group C protein.* Molecular and Cellular Biology, 2005. **25**(13): p. 5664-5674.
159. Hey, T., et al., *The XPC-HR23B complex displays high affinity and specificity for damaged DNA in a true-equilibrium fluorescence assay.* Biochemistry, 2002. **41**(21): p. 6583-6587.

160. Hoogstraten, D., et al., *Versatile DNA damage detection by the global genome nucleotide excision repair protein XPC.* J Cell Sci, 2008. **121**(Pt 17): p. 2850-9.
161. Camenisch, U., et al., *Two-stage dynamic DNA quality check by xeroderma pigmentosum group C protein.* Embo Journal, 2009. **28**(16): p. 2387-2399.
162. Kong, M.W., et al., *Single-Molecule Imaging Reveals that Rad4 Employs a Dynamic DNA Damage Recognition Process.* Molecular Cell, 2016. **64**(2): p. 376-387.
163. Min, J.H. and N.P. Pavletich, *Recognition of DNA damage by the Rad4 nucleotide excision repair protein.* Nature, 2007. **449**(7162): p. 570-6.
164. Puumalainen, M.R., et al., *Xeroderma pigmentosum group C sensor: unprecedented recognition strategy and tight spatiotemporal regulation.* Cellular and Molecular Life Sciences, 2016. **73**(3): p. 547-566.
165. Gunz, D., M.T. Hess, and H. Naegeli, *Recognition of DNA adducts by human nucleotide excision repair - Evidence for a thermodynamic probing mechanism.* Journal of Biological Chemistry, 1996. **271**(41): p. 25089-25098.
166. Lee, J.H., et al., *NMR structure of the DNA decamer duplex containing double T*G mismatches of cis-syn cyclobutane pyrimidine dimer: implications for DNA damage recognition by the XPC-hHR23B complex.* Nucleic Acids Research, 2004. **32**(8): p. 2474-2481.
167. Kusumoto, R., et al., *Diversity of the damage recognition step in the global genomic nucleotide excision repair in vitro.* Mutation Research-DNA Repair, 2001. **485**(3): p. 219-227.
168. Chu, G. and E. Chang, *Xeroderma Pigmentosum Group-E Cells Lack a Nuclear Factor That Binds to Damaged DNA.* Science, 1988. **242**(4878): p. 564-567.
169. Wakasugi, M., et al., *Damaged DNA-binding protein DDB stimulates the excision of cyclobutane pyrimidine dimers in vitro in concert with XPA and replication protein A.* Journal of Biological Chemistry, 2001. **276**(18): p. 15434-15440.
170. Hwang, B.J. and G. Chu, *Purification and Characterization of a Human Protein That Binds to Damaged DNA.* Biochemistry, 1993. **32**(6): p. 1657-1666.
171. Fujiwara, Y., et al., *Characterization of DNA recognition by the human UV-damaged DNA-binding protein.* Journal of Biological Chemistry, 1999. **274**(28): p. 20027-20033.
172. Tang, J.Y., et al., *Xeroderma pigmentosum p48 gene enhances global genomic repair and suppresses UV-induced mutagenesis.* Molecular Cell, 2000. **5**(4): p. 737-744.
173. Wittschieben, B.O., S. Iwai, and R.D. Wood, *DDB1-DDB2 (Xeroderma Pigmentosum Group E) Protein Complex Recognizes a Cyclobutane Pyrimidine Dimer,Mismatches, Apurinic/Apyrimidinic Sites, and Compound Lesions in DNA.* Journal of Biological Chemistry, 2005. **280**(48): p. 39982-39989.
174. Scrima, A., et al., *Structural Basis of UV DNA-Damage Recognition by the DDB1-DDB2 Complex.* Cell, 2008. **135**(7): p. 1213-1223.
175. Fischer, E.S., et al., *The Molecular Basis of CRL4(DDB2/CSA) Ubiquitin Ligase Architecture, Targeting, and Activation.* Cell, 2011. **147**(5): p. 1024-1039.
176. Wakasugi, M., et al., *DDB accumulates at DNA damage sites immediately after UV irradiation and directly stimulates nucleotide excision repair.* Journal of Biological Chemistry, 2002. **277**(3): p. 1637-1640.
177. Fitch, M.E., et al., *In vivo recruitment of XPC to UV-induced cyclobutane pyrimidine dimers by the DDB2 gene product.* J Biol Chem, 2003. **278**(47): p. 46906-10.
178. Moser, J., et al., *The UV-damaged DNA binding protein mediates efficient targeting of the nucleotide excision repair complex to UV-induced photo lesions.* DNA Repair, 2005. **4**(5): p. 571-582.
179. Naegeli, H. and K. Sugasawa, *The xeroderma pigmentosum pathway: Decision tree analysis of DNA quality.* DNA Repair, 2011. **10**(7): p. 673-683.
180. Li, R.Y., et al., *Interactions of the transcription/DNA repair factor TFIIH and XP repair proteins with DNA lesions in a cell-free repair assay.* Journal of Molecular Biology, 1998. **281**(2): p. 211-218.

181. Yokoi, M., et al., *The xeroderma pigmentosum group C protein complex XPC-HR23B plays an important role in the recruitment of transcription factor IIH to damaged DNA.* Journal of Biological Chemistry, 2000. **275**(13): p. 9870-9875.
182. Volker, M., et al., *Sequential assembly of the nucleotide excision repair factors in vivo.* Molecular Cell, 2001. **8**(1): p. 213-224.
183. Riedl, T., F. Hanaoka, and J.M. Egly, *The comings and goings of nucleotide excision repair factors on damaged DNA.* Embo Journal, 2003. **22**(19): p. 5293-5303.
184. Winkler, G.S., et al., *Novel functional interactions between nucleotide excision DNA repair proteins influencing the enzymatic activities of TFIIH, XPG, and ERCC1-XPF.* Biochemistry, 2001. **40**(1): p. 160-165.
185. Coin, F., et al., *Nucleotide excision repair driven by the dissociation of CAK from TFIIH.* Molecular Cell, 2008. **31**(1): p. 9-20.
186. Li, C.L., et al., *Tripartite DNA Lesion Recognition and Verification by XPC, TFIIH, and XPA in Nucleotide Excision Repair.* Molecular Cell, 2015. **59**(6): p. 1025-1034.
187. Evans, E., et al., *Mechanism of open complex and dual incision formation by human nucleotide excision repair factors.* Embo Journal, 1997. **16**(21): p. 6559-6573.
188. Coverley, D., et al., *Requirement for the replication protein SSB in human DNA excision repair.* Nature, 1991. **349**(6309): p. 538-41.
189. Li, L., et al., *An interaction between the DNA repair factor XPA and replication protein A appears essential for nucleotide excision repair.* Mol Cell Biol, 1995. **15**(10): p. 5396-402.
190. Matsunaga, T., et al., *Replication protein A confers structure-specific endonuclease activities to the XPF-ERCC1 and XPG subunits of human DNA repair excision nuclease.* J Biol Chem, 1996. **271**(19): p. 11047-50.
191. Bochkarev, A., et al., *Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA.* Nature, 1997. **385**(6612): p. 176-181.
192. Tapias, A., et al., *Ordered conformational changes in damaged DNA induced by nucleotide excision repair factors.* Journal of Biological Chemistry, 2004. **279**(18): p. 19074-19083.
193. Liu, Y., et al., *Interactions of human replication protein A with single-stranded DNA adducts.* Biochem J, 2005. **385**(Pt 2): p. 519-26.
194. Krasikova, Y.S., et al., *Localization of xeroderma pigmentosum group A protein and replication protein A on damaged DNA in nucleotide excision repair.* Nucleic Acids Research, 2010. **38**(22): p. 8083-8094.
195. Pugh, R.A., C.G. Wu, and M. Spies, *Regulation of translocation polarity by helicase domain 1 in SF2B helicases.* Embo Journal, 2012. **31**(2): p. 503-514.
196. Sugasawa, K., et al., *Two-Step Recognition of DNA Damage for Mammalian Nucleotide Excision Repair: Directional Binding of the XPC Complex and DNA Strand Scanning.* Molecular Cell, 2009. **36**(4): p. 642-653.
197. Naegeli, H., L. Bardwell, and E.C. Friedberg, *Inhibition of Rad3 DNA Helicase Activity by DNA Adducts and Abasic Sites - Implications for the Role of a DNA Helicase in Damage-Specific Incision of DNA.* Biochemistry, 1993. **32**(2): p. 613-621.
198. Mathieu, N., N. Kaczmarek, and H. Naegeli, *Strand- and site-specific DNA lesion demarcation by the xeroderma pigmentosum group D helicase.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(41): p. 17545-17550.
199. Mathieu, N., et al., *DNA Quality Control by a Lesion Sensor Pocket of the Xeroderma Pigmentosum Group D Helicase Subunit of TFIIH.* Current Biology, 2013. **23**(3): p. 204-212.
200. Robins, P., et al., *Complementation of DNA repair in xeroderma pigmentosum group A cell extracts by a protein with affinity for damaged DNA.* Embo j, 1991. **10**(12): p. 3913-21.
201. Jones, C.J. and R.D. Wood, *Preferential binding of the xeroderma pigmentosum group A complementing protein to damaged DNA.* Biochemistry, 1993. **32**(45): p. 12096-104.

202. Asahina, H., et al., *The XPA protein is a zinc metalloprotein with an ability to recognize various kinds of DNA damage.* Mutat Res, 1994. **315**(3): p. 229-37.
203. Missura, M., et al., *Double-check probing of DNA bending and unwinding by XPA-RPA: an architectural function in DNA repair.* Embo Journal, 2001. **20**(13): p. 3554-3564.
204. Camenisch, U., et al., *Recognition of helical kinks by xeroderma pigmentosum group A protein triggers DNA excision repair.* Nature Structural & Molecular Biology, 2006. **13**(3): p. 278-284.
205. Koch, S.C., et al., *Structural insights into the recognition of cisplatin and AAF-dG lesion by Rad14 (XPA).* Proceedings of the National Academy of Sciences of the United States of America, 2015. **112**(27): p. 8272-8277.
206. Li, L., et al., *Specific association between the human DNA repair proteins XPA and ERCC1.* Proc Natl Acad Sci U S A, 1994. **91**(11): p. 5012-6.
207. Li, L., et al., *Mutations in XPA that prevent association with ERCC1 are defective in nucleotide excision repair.* Mol Cell Biol, 1995. **15**(4): p. 1993-8.
208. Saijo, M., et al., *Sequential binding of DNA repair proteins RPA and ERCC1 to XPA in vitro.* Nucleic Acids Research, 1996. **24**(23): p. 4719-4724.
209. Tsodikov, O.V., et al., *Structural basis for the recruitment of ERCC1-XPF to nucleotide excision repair complexes by XPA.* Embo Journal, 2007. **26**(22): p. 4768-4776.
210. Orelli, B., et al., *The XPA-binding domain of ERCC1 Is Required for Nucleotide Excision Repair but Not Other DNA Repair Pathways.* Journal of Biological Chemistry, 2010. **285**(6): p. 3705-3712.
211. He, Z., et al., *RPA involvement in the damage-recognition and incision steps of nucleotide excision repair.* Nature, 1995. **374**(6522): p. 566-9.
212. de Laat, W.L., et al., *DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair.* Genes & Development, 1998. **12**(16): p. 2598-2609.
213. Staresincic, L., et al., *Coordination of dual incision and repair synthesis in human nucleotide excision repair.* Embo Journal, 2009. **28**(8): p. 1111-1120.
214. Bardwell, A.J., et al., *Specific Cleavage of Model Recombination and Repair Intermediates by the Yeast Rad1-Rad10 DNA Endonuclease.* Science, 1994. **265**(5181): p. 2082-2085.
215. Matsunaga, T., et al., *Human DNA repair excision nuclease. Analysis of the roles of the subunits involved in dual incisions by using anti-XPG and anti-ERCC1 antibodies.* J Biol Chem, 1995. **270**(35): p. 20862-9.
216. Sijbers, A.M., et al., *Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease.* Cell, 1996. **86**(5): p. 811-22.
217. O'Donovan, A., et al., *XPG endonuclease makes the 3' incision in human DNA nucleotide excision repair.* Nature, 1994. **371**(6496): p. 432-5.
218. O'Donovan, A., et al., *Isolation of active recombinant XPG protein, a human DNA repair endonuclease.* J Biol Chem, 1994. **269**(23): p. 15965-8.
219. Evans, E., et al., *Open complex formation around a lesion during nucleotide excision repair provides a structure for cleavage by human XPG protein.* Embo Journal, 1997. **16**(3): p. 625-638.
220. Mu, D., D.S. Hsu, and A. Sancar, *Reaction mechanism of human DNA repair excision nuclease.* Journal of Biological Chemistry, 1996. **271**(14): p. 8285-8294.
221. Huang, J.C., et al., *Human Nucleotide Excision Nuclease Removes Thymine Dimers from DNA by Incising the 22nd Phosphodiester Bond 5' and the 6th Phosphodiester Bond 3' to the Photodimer.* Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**(8): p. 3664-3668.
222. Svoboda, D.L., et al., *DNA repair by eukaryotic nucleotide excision nuclease. Removal of thymine dimer and psoralen monoadduct by HeLa cell-free extract and of thymine dimer by Xenopus laevis oocytes.* J Biol Chem, 1993. **268**(3): p. 1931-6.

223. Huang, J.C., et al., *Substrate spectrum of human excinuclease: repair of abasic sites, methylated bases, mismatches, and bulky adducts.* Proc Natl Acad Sci U S A, 1994. **91**(25): p. 12213-7.
224. Moggs, J.G., et al., *Analysis of incision sites produced by human cell extracts and purified proteins during nucleotide excision repair of a 1,3-intrastrand d(GpTpG)-cisplatin adduct.* Journal of Biological Chemistry, 1996. **271**(12): p. 7177-7186.
225. Hu, J.C., et al., *Nucleotide Excision Repair in Human Cells FATE OF THE EXCISED OLIGONUCLEOTIDE CARRYING DNA DAMAGE IN VIVO.* Journal of Biological Chemistry, 2013. **288**(29): p. 20918-20926.
226. Kemp, M.G., et al., *Mechanism of Release and Fate of Excised Oligonucleotides during Nucleotide Excision Repair.* Journal of Biological Chemistry, 2012. **287**(27): p. 22889-22899.
227. Kemp, M.G. and J. Hu, *PostExcision Events in Human Nucleotide Excision Repair.* Photochem Photobiol, 2017. **93**(1): p. 178-191.
228. Podust, L.M., et al., *Assembly of DNA polymerase delta and epsilon holoenzymes depends on the geometry of the DNA template.* Nucleic Acids Res, 1994. **22**(15): p. 2970-5.
229. Shivji, M.K., et al., *Nucleotide excision repair DNA synthesis by DNA polymerase epsilon in the presence of PCNA, RFC, and RPA.* Biochemistry, 1995. **34**(15): p. 5011-7.
230. Araujo, S.J., et al., *Nucleotide excision repair of DNA with recombinant human proteins: definition of the minimal set of factors, active forms of TFIIH, and modulation by CAK.* Genes & Development, 2000. **14**(3): p. 349-359.
231. Ogi, T. and A.R. Lehmann, *The Y-family DNA polymerase kappa (pol kappa) functions in mammalian nucleotide-excision repair.* Nature Cell Biology, 2006. **8**(6): p. 640-642.
232. Ogi, T., et al., *Three DNA Polymerases, Recruited by Different Mechanisms, Carry Out NER Repair Synthesis in Human Cells.* Molecular Cell, 2010. **37**(5): p. 714-727.
233. Mocquet, V., et al., *Sequential recruitment of the repair factors during NER: the role of XPG in initiating the resynthesis step.* Embo Journal, 2008. **27**(1): p. 155-167.
234. Overmeer, R.M., et al., *Replication protein A safeguards genome integrity by controlling NER incision events.* Journal of Cell Biology, 2011. **192**(3): p. 401-415.
235. Moser, J., et al., *Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase III alpha in a cell-cycle-specific manner.* Molecular Cell, 2007. **27**(2): p. 311-323.
236. Mellon, I., G. Spivak, and P.C. Hanawalt, *Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene.* Cell, 1987. **51**(2): p. 241-9.
237. Mellon, I. and P.C. Hanawalt, *Induction of the Escherichia coli lactose operon selectively increases repair of its transcribed DNA strand.* Nature, 1989. **342**(6245): p. 95-8.
238. Donahue, B.A., et al., *Transcript cleavage by RNA polymerase II arrested by a cyclobutane pyrimidine dimer in the DNA template.* Proc Natl Acad Sci U S A, 1994. **91**(18): p. 8502-6.
239. Cline, S.D., et al., *Malondialdehyde adducts in DNA arrest transcription by T7 RNA polymerase and mammalian RNA polymerase II.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(19): p. 7275-7280.
240. Brueckner, F., et al., *CPD damage recognition by transcribing RNA polymerase II.* Science, 2007. **315**(5813): p. 859-862.
241. Selby, C.P. and A. Sancar, *Cockayne syndrome group B protein enhances elongation by RNA polymerase II.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(21): p. 11205-11209.
242. van den Boom, V., et al., *DNA damage stabilizes interaction of CSB with the transcription elongation machinery.* J Cell Biol, 2004. **166**(1): p. 27-36.
243. Selby, C.P. and A. Sancar, *Human transcription-repair coupling factor CSB/ERCC6 is a DNA-stimulated ATPase but is not a helicase and does not disrupt the ternary*

transcription complex of stalled RNA polymerase II. Journal of Biological Chemistry, 1997. **272**(3): p. 1885-1890.
244. Tantin, D., A. Kansal, and M. Carey, *Recruitment of the putative transcription-repair coupling factor CSB/ERCC6 to RNA polymerase II elongation complexes.* Molecular and Cellular Biology, 1997. **17**(12): p. 6803-6814.
245. van Gool, A.J., et al., *The Cockayne syndrome B protein, involved in transcription-coupled DNA repair, resides in an RNA polymerase II-containing complex.* Embo j, 1997. **16**(19): p. 5955-65.
246. Citterio, E., et al., *ATP-dependent chromatin remodeling by the Cockayne syndrome B DNA repair-transcription-coupling factor.* Molecular and Cellular Biology, 2000. **20**(20): p. 7643-7653.
247. Beerens, N., et al., *The CSB protein actively wraps DNA.* Journal of Biological Chemistry, 2005. **280**(6): p. 4722-4729.
248. Sarker, A.H., et al., *Recognition of RNA polymerase II and transcription bubbles by XPG, CSB, and TFIIH: Insights for transcription-coupled repair and Cockayne syndrome.* Molecular Cell, 2005. **20**(2): p. 187-198.
249. Tantin, D., *RNA polymerase II elongation complexes containing the cockayne syndrome group B protein interact with a molecular complex containing the transcription factor IIH components xeroderma pigmentosum B and p62.* Journal of Biological Chemistry, 1998. **273**(43): p. 27794-27799.
250. Fousteri, M., et al., *Cockayne syndrome A and B proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled RNA polymerase II in vivo.* Molecular Cell, 2006. **23**(4): p. 471-482.
251. Selby, C.P., et al., *RNA polymerase II stalled at a thymine dimer: Footprint and effect on excision repair.* Nucleic Acids Research, 1997. **25**(4): p. 787-793.
252. Tornaletti, S., D. Reines, and P.C. Hanawalt, *Structural characterization of RNA polymerase II complexes arrested by a cyclobutane pyrimidine dimer in the transcribed strand of template DNA.* Journal of Biological Chemistry, 1999. **274**(34): p. 24124-24130.
253. Lee, K.B., et al., *Transcription-coupled and DNA damage-dependent ubiquitination of RNA polymerase II in vitro.* Proc Natl Acad Sci U S A, 2002. **99**(7): p. 4239-44.
254. Woudstra, E.C., et al., *A Rad26-Def1 complex coordinates repair and RNA pol II proteolysis in response to DNA damage.* Nature, 2002. **415**(6874): p. 929-33.
255. Somesh, B.P., et al., *Multiple mechanisms confining RNA polymerase II ubiquitylation to polymerases undergoing transcriptional arrest.* Cell, 2005. **121**(6): p. 913-923.
256. Anindya, R., O. Aygun, and J.Q. Svejstrup, *Damage-induced ubiquitylation of human RNA polymerase II by the ubiquitin ligase Nedd4, but not Cockayne syndrome proteins or BRCA1.* Molecular Cell, 2007. **28**(3): p. 386-397.
257. Malik, S., et al., *Elongating RNA polymerase II is disassembled through specific degradation of its largest but not other subunits in response to DNA damage in vivo.* J Biol Chem, 2008. **283**(11): p. 6897-905.
258. Harreman, M., et al., *Distinct ubiquitin ligases act sequentially for RNA polymerase II polyubiquitylation.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(49): p. 20705-20710.
259. Wilson, M.D., M. Harreman, and J.Q. Svejstrup, *Ubiquitylation and degradation of elongating RNA polymerase II: The last resort.* Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms, 2013. **1829**(1): p. 151-157.
260. Sigurdsson, S., A.B. Dirac-Svejstrup, and J.Q. Svejstrup, *Evidence that transcript cleavage is essential for RNA polymerase II transcription and cell viability.* Mol Cell, 2010. **38**(2): p. 202-10.
261. Saltzman, A.G. and R. Weinmann, *Promoter specificity and modulation of RNA polymerase II transcription.* Faseb j, 1989. **3**(6): p. 1723-33.
262. Sawadogo, M. and A. Sentenac, *RNA polymerase B (II) and general transcription factors.* Annu Rev Biochem, 1990. **59**: p. 711-54.
263. Gnatt, A.L., et al., *Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution.* Science, 2001. **292**(5523): p. 1876-82.

264. Borukhov, S. and E. Nudler, *RNA polymerase: the vehicle of transcription.* Trends Microbiol, 2008. **16**(3): p. 126-34.
265. Decker, K.B. and D.M. Hinton, *Transcription regulation at the core: similarities among bacterial, archaeal, and eukaryotic RNA polymerases.* Annu Rev Microbiol, 2013. **67**: p. 113-39.
266. Plaschka, C., et al., *Architecture of the RNA polymerase II-Mediator core initiation complex.* Nature, 2015. **518**(7539): p. 376-380.
267. Roeder, R.G. and W.J. Rutter, *Multiple Forms of DNA-Dependent Rna Polymerase in Eukaryotic Organisms.* Nature, 1969. **224**(5216): p. 234-&.
268. Chambon, P., *Eukaryotic nuclear RNA polymerases.* Annu Rev Biochem, 1975. **44**: p. 613-38.
269. Sentenac, A., *Eukaryotic RNA polymerases.* CRC Crit Rev Biochem, 1985. **18**(1): p. 31-90.
270. Archambault, J. and J.D. Friesen, *Genetics of eukaryotic RNA polymerases I, II, and III.* Microbiol Rev, 1993. **57**(3): p. 703-24.
271. Cramer, P., et al., *Structure of eukaryotic RNA polymerases.* Annu Rev Biophys, 2008. **37**: p. 337-52.
272. Hoogstraten, D., et al., *Rapid switching of TFIIH between RNA polymerase I and II transcription and DNA repair in vivo.* Molecular Cell, 2002. **10**(5): p. 1163-1174.
273. Iben, S., et al., *TFIIH plays an essential role in RNA polymerase I transcription.* Cell, 2002. **109**(3): p. 297-306.
274. Assfalg, R., et al., *TFIIH is an elongation factor of RNA polymerase I.* Nucleic Acids Research, 2012. **40**(2): p. 650-659.
275. Nonnekens, J., et al., *Mutations in TFIIH causing trichothiodystrophy are responsible for defects in ribosomal RNA production and processing.* Human Molecular Genetics, 2013. **22**(14): p. 2881-2893.
276. Listerman, I., et al., *Extragenic accumulation of RNA polymerase II enhances transcription by RNA polymerase III.* Plos Genetics, 2007. **3**(11): p. 2268-2277.
277. Barski, A., et al., *Pol II and its associated epigenetic marks are present at Pol III-transcribed noncoding RNA genes.* Nature Structural & Molecular Biology, 2010. **17**(5): p. 629-U132.
278. Reeder, R.H. and R.G. Roeder, *Ribosomal RNA synthesis in isolated nuclei.* J Mol Biol, 1972. **67**(3): p. 433-41.
279. Bowman, L.H., et al., *Location of the initial cleavage sites in mouse pre-rRNA.* Mol Cell Biol, 1983. **3**(8): p. 1501-10.
280. Sollner-Webb, B. and J. Tower, *Transcription of cloned eukaryotic ribosomal RNA genes.* Annu Rev Biochem, 1986. **55**: p. 801-30.
281. Hannan, K.M., R.D. Hannan, and L.I. Rothblum, *Transcription by RNA polymerase I.* Front Biosci, 1998. **3**: p. d376-98.
282. Fromont-Racine, M., et al., *Ribosome assembly in eukaryotes.* Gene, 2003. **313**: p. 17-42.
283. Henras, A.K., et al., *The post-transcriptional steps of eukaryotic ribosome biogenesis.* Cellular and Molecular Life Sciences, 2008. **65**(15): p. 2334-2359.
284. Learned, R.M., S. Cordes, and R. Tjian, *Purification and characterization of a transcription factor that confers promoter specificity to human RNA polymerase I.* Mol Cell Biol, 1985. **5**(6): p. 1358-69.
285. Clos, J., D. Buttgereit, and I. Grummt, *A purified transcription factor (TIF-IB) binds to essential sequences of the mouse rDNA promoter.* Proc Natl Acad Sci U S A, 1986. **83**(3): p. 604-8.
286. Bell, S.P., et al., *Functional cooperativity between transcription factors UBF1 and SL1 mediates human ribosomal RNA synthesis.* Science, 1988. **241**(4870): p. 1192-7.
287. Jantzen, H.M., et al., *Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins.* Nature, 1990. **344**(6269): p. 830-6.
288. Schnapp, A. and I. Grummt, *Transcription complex formation at the mouse rDNA promoter involves the stepwise association of four transcription factors and RNA polymerase I.* J Biol Chem, 1991. **266**(36): p. 24588-95.

289. Grummt, I., *Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus.* Genes & Development, 2003. **17**(14): p. 1691-1702.
290. Sanij, E. and R.D. Hannan, *The role of UBF in regulating the structure and dynamics of transcriptionally active rDNA chromatin.* Epigenetics, 2009. **4**(6): p. 374-82.
291. Comai, L., N. Tanese, and R. Tjian, *The TATA-binding protein and associated factors are integral components of the RNA polymerase I transcription factor, SL1.* Cell, 1992. **68**(5): p. 965-76.
292. Eberhard, D., et al., *A TBP-containing multiprotein complex (TIF-IB) mediates transcription specificity of murine RNA polymerase I.* Nucleic Acids Res, 1993. **21**(18): p. 4180-6.
293. Heix, J., et al., *Cloning of murine RNA polymerase I-specific TAF factors: conserved interactions between the subunits of the species-specific transcription initiation factor TIF-IB/SL1.* Proc Natl Acad Sci U S A, 1997. **94**(5): p. 1733-8.
294. Hanada, K., et al., *RNA polymerase I associated factor 53 binds to the nucleolar transcription factor UBF and functions in specific rDNA transcription.* Embo j, 1996. **15**(9): p. 2217-26.
295. Seither, P., et al., *Constitutive and strong association of PAF53 with RNA polymerase I.* Chromosoma, 1997. **106**(4): p. 216-225.
296. Miller, G., et al., *hRRN3 is essential in the SL1-mediated recruitment of RNA Polymerase I to rRNA gene promoters.* Embo j, 2001. **20**(6): p. 1373-82.
297. Yuan, X., et al., *Multiple interactions between RNA polymerase I, TIF-IA and TAF(I) subunits regulate preinitiation complex assembly at the ribosomal gene promoter.* EMBO Rep, 2002. **3**(11): p. 1082-7.
298. Titov, D.V., et al., *XPB, a subunit of TFIIH, is a target of the natural product triptolide.* Nature Chemical Biology, 2011. **7**(3): p. 182-188.
299. Weinmann, R. and R.G. Roeder, *Role of DNA-dependent RNA polymerase 3 in the transcription of the tRNA and 5S RNA genes.* Proc Natl Acad Sci U S A, 1974. **71**(5): p. 1790-4.
300. Schramm, L. and N. Hernandez, *Recruitment of RNA polymerase III to its target promoters.* Genes Dev, 2002. **16**(20): p. 2593-620.
301. Nikitina, T.V., L.I. Tischenko, and W.A. Schulz, *Recent insights into regulation of transcription by RNA polymerase III and the cellular functions of its transcripts.* Biol Chem, 2011. **392**(5): p. 395-404.
302. White, R.J., *Transcription by RNA polymerase III: more complex than we thought.* Nature Reviews Genetics, 2011. **12**(7): p. 459-463.
303. Carey, M.F., S.P. Gerrard, and N.R. Cozzarelli, *Analysis of RNA polymerase III transcription complexes by gel filtration.* J Biol Chem, 1986. **261**(9): p. 4309-17.
304. Gabrielsen, O.S., et al., *Two polypeptide chains in yeast transcription factor tau interact with DNA.* J Biol Chem, 1989. **264**(13): p. 7505-11.
305. Huet, J. and A. Sentenac, *The TATA-binding protein participates in TFIIIB assembly on tRNA genes.* Nucleic Acids Res, 1992. **20**(24): p. 6451-4.
306. Kassavetis, G.A., et al., *The role of the TATA-binding protein in the assembly and function of the multisubunit yeast RNA polymerase III transcription factor, TFIIIB.* Cell, 1992. **71**(6): p. 1055-64.
307. Hernandez, N., *Tbp, a Universal Eukaryotic Transcription Factor.* Genes & Development, 1993. **7**(7b): p. 1291-1308.
308. Kassavetis, G.A., et al., *S. cerevisiae TFIIIB is the transcription initiation factor proper of RNA polymerase III, while TFIIIA and TFIIIC are assembly factors.* Cell, 1990. **60**(2): p. 235-45.
309. Kassavetis, G.A. and E.P. Geiduschek, *Transcription factor TFIIIB and transcription by RNA polymerase III.* Biochem Soc Trans, 2006. **34**(Pt 6): p. 1082-7.
310. Raha, D., et al., *Close association of RNA polymerase II and many transcription factors with Pol III genes.* Proc Natl Acad Sci U S A, 2010. **107**(8): p. 3639-44.
311. Zylber, E.A. and S. Penman, *Products of RNA polymerases in HeLa cell nuclei.* Proc Natl Acad Sci U S A, 1971. **68**(11): p. 2861-5.

312. Sims, R.J., 3rd, S.S. Mandal, and D. Reinberg, *Recent highlights of RNA-polymerase-II-mediated transcription.* Curr Opin Cell Biol, 2004. **16**(3): p. 263-71.
313. Darnell, J.E., Jr., *Reflections on the history of pre-mRNA processing and highlights of current knowledge: a unified picture.* Rna, 2013. **19**(4): p. 443-60.
314. Jurado, A.R., et al., *Structure and function of pre-mRNA 5'-end capping quality control and 3'-end processing.* Biochemistry, 2014. **53**(12): p. 1882-98.
315. Merrick, W.C., *Eukaryotic protein synthesis: still a mystery.* J Biol Chem, 2010. **285**(28): p. 21197-201.
316. Hershey, J.W.B., N. Sonenberg, and M.B. Mathews, *Principles of Translational Control: An Overview.* Cold Spring Harb Perspect Biol, 2012. **4**(12).
317. Bhat, M., et al., *Targeting the translation machinery in cancer.* Nat Rev Drug Discov, 2015. **14**(4): p. 261-78.
318. Maldonado, E., et al., *Factors involved in specific transcription by mammalian RNA polymerase II: role of transcription factors IIA, IID, and IIB during formation of a transcription-competent complex.* Mol Cell Biol, 1990. **10**(12): p. 6335-47.
319. Burley, S.K. and R.G. Roeder, *Biochemistry and structural biology of transcription factor IID (TFIID).* Annu Rev Biochem, 1996. **65**: p. 769-99.
320. Bleichenbacher, M., S. Tan, and T.J. Richmond, *Novel interactions between the components of human and yeast TFIIA/TBP/DNA complexes.* Journal of Molecular Biology, 2003. **332**(4): p. 783-793.
321. Louder, R.K., et al., *Structure of promoter-bound TFIID and model of human pre-initiation complex assembly.* Nature, 2016. **531**(7596): p. 604-9.
322. Li, Y., et al., *RNA polymerase II initiation factor interactions and transcription start site selection.* Science, 1994. **263**(5148): p. 805-7.
323. Flores, O., et al., *The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex.* Proc Natl Acad Sci U S A, 1991. **88**(22): p. 9999-10003.
324. Buratowski, S., et al., *Five intermediate complexes in transcription initiation by RNA polymerase II.* Cell, 1989. **56**(4): p. 549-61.
325. Conaway, R.C. and J.W. Conaway, *General initiation factors for RNA polymerase II.* Annu Rev Biochem, 1993. **62**: p. 161-90.
326. Maxon, M.E., J.A. Goodrich, and R. Tjian, *Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance.* Genes Dev, 1994. **8**(5): p. 515-24.
327. Ohkuma, Y. and R.G. Roeder, *Regulation of TFIIH ATPase and kinase activities by TFIIE during active initiation complex formation.* Nature, 1994. **368**(6467): p. 160-3.
328. Flanagan, P.M., et al., *A mediator required for activation of RNA polymerase II transcription in vitro.* Nature, 1991. **350**(6317): p. 436-8.
329. Kim, Y.J., et al., *A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II.* Cell, 1994. **77**(4): p. 599-608.
330. Aranda, A. and A. Pascual, *Nuclear hormone receptors and gene expression.* Physiological Reviews, 2001. **81**(3): p. 1269-1304.
331. Kornberg, R.D., *Mediator and the mechanism of transcriptional activation.* Trends in Biochemical Sciences, 2005. **30**(5): p. 235-239.
332. Malik, S. and R.G. Roeder, *Dynamic regulation of pol II transcription by the mammalian Mediator complex.* Trends in Biochemical Sciences, 2005. **30**(5): p. 256-263.
333. Conaway, R.C. and J.W. Conaway, *Function and regulation of the Mediator complex.* Curr Opin Genet Dev, 2011. **21**(2): p. 225-30.
334. Jeronimo, C. and F. Robert, *Kin28 regulates the transient association of Mediator with core promoters.* Nat Struct Mol Biol, 2014. **21**(5): p. 449-55.
335. Esnault, C., et al., *Mediator-dependent recruitment of TFIIH modules in preinitiation complex.* Mol Cell, 2008. **31**(3): p. 337-46.
336. Chen, W. and R.G. Roeder, *Mediator-dependent nuclear receptor function.* Semin Cell Dev Biol, 2011. **22**(7): p. 749-58.

337. Coin, F., et al., *Mutations in XPB and XPD helicases found in xeroderma pigmentosum patients impair the transcription function of TFIIH.* Embo Journal, 1999. **18**(5): p. 1357-1366.
338. Douziech, M., et al., *Mechanism of promoter melting by the xeroderma pigmentosum complementation group B helicase of transcription factor IIH revealed by protein-DNA photo-cross-linking.* Mol Cell Biol, 2000. **20**(21): p. 8168-77.
339. Kim, T.K., R.H. Ebright, and D. Reinberg, *Mechanism of ATP-dependent promoter melting by transcription factor IIH.* Science, 2000. **288**(5470): p. 1418-1421.
340. Grunberg, S., L. Warfield, and S. Hahn, *Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening.* Nature Structural & Molecular Biology, 2012. **19**(8): p. 788-796.
341. Lin, Y.C., W.S. Choi, and J.D. Gralla, *TFIIH XPB mutants suggest a unified bacterial-like mechanism for promoter opening but not escape.* Nature Structural & Molecular Biology, 2005. **12**(7): p. 603-607.
342. Alekseev, S., et al., *Transcription without XPB Establishes a Unified Helicase-Independent Mechanism of Promoter Opening in Eukaryotic Gene Expression.* Mol Cell, 2017. **65**(3): p. 504-514.e4.
343. Stefanini, M., et al., *A New Nucleotide-Excision-Repair Gene Associated with the Disorder Trichothiodystrophy.* American Journal of Human Genetics, 1993. **53**(4): p. 817-821.
344. Hashimoto, S. and J.M. Egly, *Trichothiodystrophy view from the molecular basis of DNA repair/transcription factor TFIIH.* Human Molecular Genetics, 2009. **18**: p. R224-R230.
345. Vermeulen, W., et al., *Sublimiting concentration of TFIIH transcription/DNA repair factor causes TTD-A trichothiodystrophy disorder.* Nature Genetics, 2000. **26**(3): p. 307-313.
346. Corden, J.L., et al., *A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II.* Proc Natl Acad Sci U S A, 1985. **82**(23): p. 7934-8.
347. Sogaard, T.M. and J.Q. Svejstrup, *Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator.* J Biol Chem, 2007. **282**(19): p. 14113-20.
348. Wong, K.H., Y. Jin, and K. Struhl, *TFIIH Phosphorylation of the Pol II CTD Stimulates Mediator Dissociation from the Preinitiation Complex and Promoter Escape.* Molecular Cell, 2014. **54**(4): p. 601-612.
349. Cho, E.J., et al., *mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain.* Genes Dev, 1997. **11**(24): p. 3319-26.
350. Ho, C.K. and S. Shuman, *Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme.* Mol Cell, 1999. **3**(3): p. 405-11.
351. Rochette-Egly, C., et al., *Stimulation of RAR alpha activation function AF-1 through binding to the general transcription factor TFIIH and phosphorylation by CDK7.* Cell, 1997. **90**(1): p. 97-107.
352. Chen, D., et al., *Activation of estrogen receptor alpha by S118 phosphorylation involves a ligand-dependent interaction with TFIIH and participation of CDK7.* Mol Cell, 2000. **6**(1): p. 127-37.
353. Keriel, A., et al., *XPD mutations prevent TFIIH-dependent transactivation by nuclear receptors and phosphorylation of RARalpha.* Cell, 2002. **109**(1): p. 125-35.
354. Compe, E., et al., *Dysregulation of the peroxisome proliferator-activated receptor target genes by XPD mutations.* Mol Cell Biol, 2005. **25**(14): p. 6065-76.
355. Akoulitchev, S., S. Chuikov, and D. Reinberg, *TFIIH is negatively regulated by cdk8-containing mediator complexes.* Nature, 2000. **407**(6800): p. 102-6.
356. Dvir, A., R.C. Conaway, and J.W. Conaway, *Promoter escape by RNA polymerase II. A role for an ATP cofactor in suppression of arrest by polymerase at promoter-proximal sites.* J Biol Chem, 1996. **271**(38): p. 23352-6.

357. Dvir, A., R.C. Conaway, and J.W. Conaway, *A role for TFIIH in controlling the activity of early RNA polymerase II elongation complexes.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(17): p. 9006-9010.
358. Kugel, J.F. and J.A. Goodrich, *Promoter escape limits the rate of RNA polymerase II transcription and is enhanced by TFIIE, TFIIH, and ATP on negatively supercoiled DNA.* Proc Natl Acad Sci U S A, 1998. **95**(16): p. 9232-7.
359. Kumar, K.P., S. Akoulitchev, and D. Reinberg, *Promoter-proximal stalling results from the inability to recruit transcription factor IIH to the transcription complex and is a regulated event.* Proc Natl Acad Sci U S A, 1998. **95**(17): p. 9767-72.
360. Moreland, R.J., et al., *A role for the TFIIH XPB DNA helicase in promoter escape by RNA polymerase II.* J Biol Chem, 1999. **274**(32): p. 22127-30.
361. Schachter, M.M., et al., *A Cdk7-Cdk4 T-Loop Phosphorylation Cascade Promotes G1 Progression.* Molecular Cell, 2013. **50**(2): p. 250-260.
362. Fesquet, D., et al., *The Mo15 Gene Encodes the Catalytic Subunit of a Protein-Kinase That Activates Cdc2 and Other Cyclin-Dependent Kinases (Cdks) through Phosphorylation of Thr161 and Its Homologs.* Embo Journal, 1993. **12**(8): p. 3111-3121.
363. Poon, R.Y., et al., *The cdc2-related protein p40MO15 is the catalytic subunit of a protein kinase that can activate p33cdk2 and p34cdc2.* Embo j, 1993. **12**(8): p. 3123-32.
364. Matsuoka, M., et al., *Activation of cyclin-dependent kinase 4 (cdk4) by mouse MO15-associated kinase.* Mol Cell Biol, 1994. **14**(11): p. 7265-75.
365. Morgan, D.O., *Principles of Cdk Regulation.* Nature, 1995. **374**(6518): p. 131-134.
366. Desai, D., Y. Gu, and D.O. Morgan, *Activation of human cyclin-dependent kinases in vitro.* Mol Biol Cell, 1992. **3**(5): p. 571-82.
367. Harper, J.W. and S.J. Elledge, *The role of Cdk7 in CAK function, a retro-retrospective.* Genes Dev, 1998. **12**(3): p. 285-9.
368. Tassan, J.P., et al., *Cell cycle analysis of the activity, subcellular localization, and subunit composition of human CAK (CDK-activating kinase).* J Cell Biol, 1994. **127**(2): p. 467-78.
369. Van Houten, B., J. Kuper, and C. Kisker, *Role of XPD in cellular functions: To TFIIH and beyond.* DNA Repair, 2016. **44**: p. 136-142.
370. Chen, J., et al., *Xpd/Ercc2 regulates CAK activity and mitotic progression.* Nature, 2003. **424**(6945): p. 228-232.
371. Li, X.M., O. Urwyler, and B. Suter, *Drosophila Xpd Regulates Cdk7 Localization, Mitotic Kinase Activity, Spindle Dynamics, and Chromosome Segregation.* Plos Genetics, 2010. **6**(3).
372. Seroz, T., et al., *Cloning of a human homolog of the yeast nucleotide excision repair gene MMS19 and interaction with transcription repair factor TFIIH via the XPB and XPD helicases.* Nucleic Acids Research, 2000. **28**(22): p. 4506-4513.
373. Ito, S., et al., *MMXD, a TFIIH-Independent XPD-MMS19 Protein Complex Involved in Chromosome Segregation.* Molecular Cell, 2010. **39**(4): p. 632-640.
374. Gari, K., et al., *MMS19 links cytoplasmic iron-sulfur cluster assembly to DNA metabolism.* Science, 2012. **337**(6091): p. 243-5.
375. Stehling, O., et al., *MMS19 assembles iron-sulfur proteins required for DNA metabolism and genomic integrity.* Science, 2012. **337**(6091): p. 195-9.
376. Vashisht, A.A., et al., *The Association of the Xeroderma Pigmentosum Group D DNA Helicase (XPD) with Transcription Factor IIH Is Regulated by the Cytosolic Iron-Sulfur Cluster Assembly Pathway.* Journal of Biological Chemistry, 2015. **290**(22): p. 14218-14225.
377. Yeom, E., S.T. Hong, and K.W. Choi, *Crumbs interacts with Xpd for nuclear division control in Drosophila.* Oncogene, 2015. **34**(21): p. 2777-2789.
378. Bradford, P.T., et al., *Cancer and neurologic degeneration in xeroderma pigmentosum: long term follow-up characterises the role of DNA repair.* Journal of Medical Genetics, 2011. **48**(3): p. 168-176.

379. Dollfus, H., et al., *Ocular manifestations in the inherited DNA repair disorders.* Surv Ophthalmol, 2003. **48**(1): p. 107-22.
380. Robbins, J.H., et al., *Neurological Disease in Xeroderma-Pigmentosum - Documentation of a Late Onset Type of the Juvenile Onset Form.* Brain, 1991. **114**: p. 1335-1361.
381. Andrews, A.D., S.F. Barrett, and J.H. Robbins, *Xeroderma Pigmentosum Neurological Abnormalities Correlate with Colony-Forming Ability after Ultraviolet-Radiation.* Proceedings of the National Academy of Sciences of the United States of America, 1978. **75**(4): p. 1984-1988.
382. Le May, N., et al., *NER Factors Are Recruited to Active Promoters and Facilitate Chromatin Modification for Transcription in the Absence of Exogenous Genotoxic Attack.* Molecular Cell, 2010. **38**(1): p. 54-66.
383. Singh, A., et al., *TFIIH Subunit Alterations Causing Xeroderma Pigmentosum and Trichothiodystrophy Specifically Disturb Several Steps during Transcription.* American Journal of Human Genetics, 2015. **96**(2): p. 194-207.
384. Cockayne, E.A., *Dwarfism with retinal atrophy and deafness.* Arch Dis Child, 1936. **11**(61): p. 1-8.
385. Venema, J., et al., *The genetic defect in Cockayne syndrome is associated with a defect in repair of UV-induced DNA damage in transcriptionally active DNA.* Proc Natl Acad Sci U S A, 1990. **87**(12): p. 4707-11.
386. Stefanini, M., et al., *Genetic analysis of twenty-two patients with Cockayne syndrome.* Hum Genet, 1996. **97**(4): p. 418-23.
387. van Gool, A.J., et al., *Cockayne syndrome: defective repair of transcription?* Embo Journal, 1997. **16**(14): p. 4155-4162.
388. Lehmann, A.R., *The xeroderma pigmentosum group D (XPD) gene: one gene, two functions, three diseases.* Genes & Development, 2001. **15**(1): p. 15-23.
389. Oh, K.S., et al., *Phenotypic heterogeneity in the XPB DNA helicase gene (ERCC3): Xeroderma pigmentosum without and with Cockayne syndrome.* Human Mutation, 2006. **27**(11): p. 1092-1103.
390. Nance, M.A. and S.A. Berry, *Cockayne Syndrome - Review of 140 Cases.* American Journal of Medical Genetics, 1992. **42**(1): p. 68-84.
391. Colella, S., et al., *Alterations in the CSB gene in three Italian patients with the severe form of Cockayne syndrome (CS) but without clinical photosensitivity.* Human Molecular Genetics, 1999. **8**(5): p. 935-941.
392. Brooks, P.J., *Blinded by the UV light: How the focus on transcription-coupled NER has distracted from understanding the mechanisms of Cockayne syndrome neurologic disease.* DNA Repair, 2013. **12**(8): p. 656-671.
393. Kraemer, K.H., et al., *Xeroderma pigmentosum, trichothiodystrophy and Cockayne syndrome: A complex genotype-phenotype relationship.* Neuroscience, 2007. **145**(4): p. 1388-1396.
394. Faghri, S., et al., *Trichothiodystrophy: a systematic review of 112 published cases characterises a wide spectrum of clinical manifestations.* Journal of Medical Genetics, 2008. **45**(10): p. 609-621.
395. de Boer, J., et al., *A mouse model for the basal transcription/DNA repair syndrome trichothiodystrophy.* Mol Cell, 1998. **1**(7): p. 981-90.
396. Dubaele, S., et al., *Basal transcription defect discriminates between xeroderma pigmentosum and trichothiodystrophy in XPD patients.* Molecular Cell, 2003. **11**(6): p. 1635-1646.
397. Botta, E., et al., *Reduced level of the repair/transcription factor TFIIH in trichothiodystrophy.* Human Molecular Genetics, 2002. **11**(23): p. 2919-2928.
398. Theil, A.F., J.H.J. Hoeijmakers, and W. Vermeulen, *TTDA: Big impact of a small protein.* Experimental Cell Research, 2014. **329**(1): p. 61-68.
399. Compe, E., et al., *Neurological defects in trichothiodystrophy reveal a coactivator function of TFIIH.* Nature Neuroscience, 2007. **10**(11): p. 1414-1422.
400. Berneburg, M., et al., *The cancer-free phenotype in trichothiodystrophy is unrelated to its repair defect.* Cancer Research, 2000. **60**(2): p. 431-438.

401. Cleaver, J.E., et al., *A summary of mutations in the UV-sensitive disorders: Xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy.* Human Mutation, 1999. **14**(1): p. 9-22.
402. Burglen, L., et al., *The gene encoding p44, a subunit of the transcription factor TFIIH, is involved in large-scale deletions associated with Werdnig-Hoffmann disease.* American Journal of Human Genetics, 1997. **60**(1): p. 72-79.
403. Taylor, E.M., et al., *Xeroderma pigmentosum and trichothiodystrophy are associated with different mutations in the XPD (ERCC2) repair/transcription gene.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(16): p. 8658-8663.
404. Broughton, B.C., et al., *Two individuals with features of both xeroderma pigmentosum and trichothiodystrophy highlight the complexity of the clinical outcomes of mutations in the XPD gene.* Human Molecular Genetics, 2001. **10**(22): p. 2539-2547.
405. Berneburg, M., et al., *UV damage causes uncontrolled DNA breakage in cells from patients with combined features of XP-D and Cockayne syndrome.* Embo Journal, 2000. **19**(5): p. 1157-1166.
406. Theron, T., et al., *Transcription-associated breaks in xeroderma pigmentosum group D cells from patients with combined features of xeroderma pigmentosum and Cockayne syndrome.* Molecular and Cellular Biology, 2005. **25**(18): p. 8368-8378.
407. Coin, F., et al., *Phosphorylation of XPB helicase regulates TFIIH nucleotide excision repair activity.* Embo Journal, 2004. **23**(24): p. 4835-4846.
408. Hwang, J.R., et al., *A 3'->5'XPB helicase defect in repair/transcription factor TFIIH of xeroderma pigmentosum group B affects both DNA repair and transcription.* Journal of Biological Chemistry, 1996. **271**(27): p. 15898-15904.
409. Cruz-Becerra, G., et al., *Analysis of Drosophila p8 and p52 mutants reveals distinct roles for the maintenance of TFIIH stability and male germ cell differentiation.* Open Biol, 2016. **6**(10).
410. Evans, P.R. and G.N. Murshudov, *How good are my data and what is the resolution?* Acta Crystallogr D Biol Crystallogr, 2013. **69**(Pt 7): p. 1204-14.
411. Cowtan, K., *Fitting molecular fragments into electron density.* Acta Crystallogr D Biol Crystallogr, 2008. **64**(Pt 1): p. 83-9.
412. Winn, M.D., et al., *Overview of the CCP4 suite and current developments.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 235-42.
413. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.
414. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.
415. Mindell, J.A. and N. Grigorieff, *Accurate determination of local defocus and specimen tilt in electron microscopy.* J Struct Biol, 2003. **142**(3): p. 334-47.
416. Tang, G., et al., *EMAN2: an extensible image processing suite for electron microscopy.* J Struct Biol, 2007. **157**(1): p. 38-46.
417. Grosse-Kunstleve, R.W. and P.D. Adams, *Substructure search procedures for macromolecular structures.* Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 11): p. 1966-73.
418. Carragher, B., et al., *Leginon: an automated system for acquisition of images from vitreous ice specimens.* J Struct Biol, 2000. **132**(1): p. 33-45.
419. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
420. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
421. Evans, P., *Scaling and assessment of data quality.* Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 1): p. 72-82.
422. Schrodinger, LLC, *The PyMOL Molecular Graphics System, Version 1.7.6*. 2015.
423. Murshudov, G.N., et al., *REFMAC5 for the refinement of macromolecular crystal structures.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 355-67.

424. Terwilliger, T.C., *Maximum-likelihood density modification.* Acta Crystallogr D Biol Crystallogr, 2000. **56**(Pt 8): p. 965-72.
425. Terwilliger, T.C., *Automated main-chain model building by template matching and iterative fragment extension.* Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 1): p. 38-44.
426. Sheldrick, G.M., *Experimental phasing with SHELXC/D/E: combining chain tracing with density modification.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 479-85.
427. Hohn, M., et al., *SPARX, a new environment for Cryo-EM image processing.* J Struct Biol, 2007. **157**(1): p. 47-55.
428. Frank, J., B. Shimkin, and H. Dowse, *Spider—A modular software system for electron image processing.* Ultramicroscopy, 1981. **6**(4): p. 343-357.
429. Kabsch, W., *XDS.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 125-32.
430. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0.* Bioinformatics, 2007. **23**(21): p. 2947-8.
431. Papadopoulos, J.S. and R. Agarwala, *COBALT: constraint-based alignment tool for multiple protein sequences.* Bioinformatics, 2007. **23**(9): p. 1073-9.
432. Rice, P., I. Longden, and A. Bleasby, *EMBOSS: the European Molecular Biology Open Software Suite.* Trends Genet, 2000. **16**(6): p. 276-7.
433. Gasteiger, E., et al., *Protein Identification and Analysis Tools on the ExPASy Server*, in *The Proteomics Protocols Handbook*, J.M. Walker, Editor. 2005, Humana Press: Totowa, NJ. p. 571-607.
434. Zhang, Y., *I-TASSER server for protein 3D structure prediction.* BMC Bioinformatics, 2008. **9**: p. 40.
435. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.
436. Kelley, L.A., et al., *The Phyre2 web portal for protein modeling, prediction and analysis.* Nat Protoc, 2015. **10**(6): p. 845-58.
437. Kallberg, M., et al., *Template-based protein structure modeling using the RaptorX web server.* Nat Protoc, 2012. **7**(8): p. 1511-22.
438. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Res, 2000. **28**(1): p. 235-42.
439. Consortium, T.U., *UniProt: the universal protein knowledgebase.* Nucleic Acids Res, 2017. **45**(D1): p. D158-d169.
440. Li, M.Z. and S.J. Elledge, *Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC.* Nat Methods, 2007. **4**(3): p. 251-6.
441. Wang, W. and B.A. Malcolm, *Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange Site-Directed Mutagenesis.* Biotechniques, 1999. **26**(4): p. 680-2.
442. Bachman, J., *Site-directed mutagenesis.* Methods Enzymol, 2013. **529**: p. 241-8.
443. Brunelle, J.L. and R. Green, *One-dimensional SDS-polyacrylamide gel electrophoresis (1D SDS-PAGE).* Methods Enzymol, 2014. **541**: p. 151-9.
444. Porath, J., et al., *Metal chelate affinity chromatography, a new approach to protein fractionation.* Nature, 1975. **258**(5536): p. 598-9.
445. Block, H., et al., *Immobilized-metal affinity chromatography (IMAC): a review.* Methods Enzymol, 2009. **463**: p. 439-73.
446. Porath, J., *Immobilized metal ion affinity chromatography.* Protein Expr Purif, 1992. **3**(4): p. 263-81.
447. Porath, J. and P. Flodin, *Gel filtration: a method for desalting and group separation.* Nature, 1959. **183**(4676): p. 1657-9.
448. Duong-Ly, K.C. and S.B. Gabelli, *Gel filtration chromatography (size exclusion chromatography) of proteins.* Methods Enzymol, 2014. **541**: p. 105-14.
449. Duong-Ly, K.C. and S.B. Gabelli, *Using ion exchange chromatography to purify a recombinantly expressed protein.* Methods Enzymol, 2014. **541**: p. 95-103.
450. Cummings, M.D., M.A. Farnum, and M.I. Nelen, *Universal screening methods and applications of ThermoFluor.* J Biomol Screen, 2006. **11**(7): p. 854-63.

451. Boivin, S., S. Kozak, and R. Meijers, *Optimization of protein purification and characterization using Thermofluor screens.* Protein Expr Purif, 2013. **91**(2): p. 192-206.
452. Drenth, J., *Principles of Protein X-Ray Crystallography*. 3 ed. 2010, New York: Springer Science+Business Media
453. Rupp, B., *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*. 2010, New York: Garland Science, Taylor & Francis Group.
454. Boistelle, R. and J.P. Astier, *Crystallization mechanisms in solution.* Journal of Crystal Growth, 1988. **90**(1): p. 14-30.
455. Asherie, N., *Protein crystallization and phase diagrams.* Methods, 2004. **34**(3): p. 266-72.
456. Giacovazzo, C., et al., *Fundamentals of Crystallography*. 3 ed. IUCr Texts on Crystallography. 2011, New York: Oxford University Press.
457. Bergfors, T., *Seeds to crystals.* J Struct Biol, 2003. **142**(1): p. 66-76.
458. Taylor, G.L., *Introduction to phasing.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 325-38.
459. Dauter, Z., M. Dauter, and E. Dodson, *Jolly SAD.* Acta Crystallogr D Biol Crystallogr, 2002. **58**(Pt 3): p. 494-506.
460. Hendrickson, W.A., *Anomalous Diffraction in Crystallographic Phase Evaluation.* Q Rev Biophys, 2014. **47**(1): p. 49-93.
461. Merritt, E.A. *X-ray Anomalous scattering*. 2012; Available from: http://www.bmsc.washington.edu/scatter.
462. Wang, J.W., et al., *SAD phasing by combination of direct methods with the SOLVE/RESOLVE procedure.* Acta Crystallographica Section D-Biological Crystallography, 2004. **60**: p. 1244-1253.
463. Schroder, R.R., *Advances in electron microscopy: A qualitative view of instrumentation development for macromolecular imaging and tomography.* Arch Biochem Biophys, 2015. **581**: p. 25-38.
464. McMullan, G., A.R. Faruqi, and R. Henderson, *Direct Electron Detectors.* Methods Enzymol, 2016. **579**: p. 1-17.
465. Kuhlbrandt, W., *Biochemistry. The resolution revolution.* Science, 2014. **343**(6178): p. 1443-4.
466. Cheng, Y., *Single-Particle Cryo-EM at Crystallographic Resolution.* Cell, 2015. **161**(3): p. 450-7.
467. Orlova, E.V. and H.R. Saibil, *Structural Analysis of Macromolecular Assemblies by Electron Microscopy.* Chem Rev, 2011. **111**(12): p. 7710-48.
468. Skiniotis, G. and D.R. Southworth, *Single-particle cryo-electron microscopy of macromolecular complexes.* Microscopy (Oxf), 2016. **65**(1): p. 9-22.
469. Milne, J.L., et al., *Cryo-electron microscopy--a primer for the non-microscopist.* Febs j, 2013. **280**(1): p. 28-45.
470. Böttcher, B., *Electron Cryomicroscopy and Image Reconstruction of Viral Nanoparticles*, in *Viral Nanotechnology*. 2015, CRC Press. p. 27-60.
471. Brenner, S. and R.W. Horne, *A negative staining method for high resolution electron microscopy of viruses.* Biochim Biophys Acta, 1959. **34**: p. 103-10.
472. Böttcher, B., *Transmission Electron Microscopy: Preparation of Specimens*, in *eLS*. 2012, John Wiley & Sons, Ltd.
473. Nogales, E., *The development of cryo-EM into a mainstream structural biology technique.* Nat Methods, 2016. **13**(1): p. 24-7.
474. Carroni, M. and H.R. Saibil, *Cryo electron microscopy to determine the structure of macromolecular complexes.* Methods, 2016. **95**: p. 78-85.
475. Amlacher, S., et al., *Insight into Structure and Assembly of the Nuclear Pore Complex by Utilizing the Genome of a Eukaryotic Thermophile.* Cell, 2011. **146**(2): p. 277-289.
476. He, Q.L., et al., *Covalent Modification of a Cysteine Residue in the XPB Subunit of the General Transcription Factor TFIIH Through Single Epoxide Cleavage of the Transcription Inhibitor Triptolide.* Angewandte Chemie-International Edition, 2015. **54**(6): p. 1859-1863.

477. Terwilliger, T.C., et al., *Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard.* Acta Crystallogr D Biol Crystallogr, 2009. **65**(Pt 6): p. 582-601.
478. Terwilliger, T.C., et al., *Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard.* Acta Crystallogr D Biol Crystallogr, 2008. **64**(Pt 1): p. 61-9.
479. Zheng, H., et al., *Data mining of metal ion environments present in protein structures.* J Inorg Biochem, 2008. **102**(9): p. 1765-76.
480. Carugo, O., *Buried chloride stereochemistry in the Protein Data Bank.* BMC Struct Biol, 2014. **14**: p. 19.
481. Iyer, N., et al., *Interactions involving the human RNA polymerase II transcription/nucleotide excision repair complex TFIIH, the nucleotide excision repair protein XPG, and Cockayne syndrome group B (CSB) protein.* Biochemistry, 1996. **35**(7): p. 2157-2167.
482. Radu, L., et al., *The intricate network between the p34 and p44 subunits is central to the activity of the transcription/DNA repair factor TFIIH.* Nucleic Acids Research, 2017.
483. Yang, Z., et al., *Iterative stable alignment and clustering of 2D transmission electron microscope images.* Structure, 2012. **20**(2): p. 237-47.
484. Loerke, J., J. Giesebrecht, and C.M. Spahn, *Multiparticle cryo-EM of ribosomes.* Methods Enzymol, 2010. **483**: p. 161-77.
485. Winkler, G.S., et al., *Affinity purification of human DNA repair/transcription factor TFIIH using epitope-tagged xeroderma pigmentosum B protein.* J Biol Chem, 1998. **273**(2): p. 1092-8.
486. Thompson, R.F., et al., *An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology.* Methods, 2016. **100**: p. 3-15.
487. Dauter, Z., M. Dauter, and K.R. Rajashankar, *Novel approach to phasing proteins: derivatization by short cryo-soaking with halides.* Acta Crystallogr D Biol Crystallogr, 2000. **56**(Pt 2): p. 232-7.
488. Phizicky, E.M. and S. Fields, *Protein-protein interactions: methods for detection and analysis.* Microbiol Rev, 1995. **59**(1): p. 94-123.
489. Di Lello, P., et al., *p53 and TFIIEalpha share a common binding site on the Tfb1/p62 subunit of TFIIH.* Proc Natl Acad Sci U S A, 2008. **105**(1): p. 106-11.
490. Levine, A.J., *p53, the cellular gatekeeper for growth and division.* Cell, 1997. **88**(3): p. 323-31.
491. Ptashne, M. and A. Gann, *Transcriptional activation by recruitment.* Nature, 1997. **386**(6625): p. 569-77.
492. Wysocka, J. and W. Herr, *The herpes simplex virus VP16-induced complex: the makings of a regulatory switch.* Trends Biochem Sci, 2003. **28**(6): p. 294-304.
493. Bouloy, M. and F. Weber, *Molecular biology of rift valley Fever virus.* Open Virol J, 2010. **4**: p. 8-14.
494. Lyles, D.S., *Cytopathogenesis and inhibition of host gene expression by RNA viruses.* Microbiol Mol Biol Rev, 2000. **64**(4): p. 709-24.
495. Kalveram, B., O. Lihoradova, and T. Ikegami, *NSs Protein of Rift Valley Fever Virus Promotes Posttranslational Downregulation of the TFIIH Subunit p62.* Journal of Virology, 2011. **85**(13): p. 6234-6243.
496. Kainulainen, M., et al., *Virulence Factor NSs of Rift Valley Fever Virus Recruits the F-Box Protein FBXO3 To Degrade Subunit p62 of General Transcription Factor TFIIH.* Journal of Virology, 2014. **88**(6): p. 3464-3473.
497. Bardwell, A.J., et al., *Yeast nucleotide excision repair proteins Rad2 and Rad4 interact with RNA polymerase II basal transcription factor b (TFIIH).* Mol Cell Biol, 1994. **14**(6): p. 3569-76.

6. Appendix

6.1. Abbreviations

A280	Absorption at 280 nm
AA/BAA	Acrylamide/Bis-acrylamide
ADP	Adenosine diphosphate
AIEX	Anion exchange chromatography
Amp	Ampicillin
APS	Ammonium persulfate
ATP	Adenosine triphosphate
BSA	Bovine serum albumin
BSD	BTF2-like transcription factors, synapse-associated and DOS2-like
CAK	CDK activating kinase
Cam	Chloramphenicol
CDK1 - CDK7	Cyclin dependent kinase 1 - 7
cDNA	Complementary DNA
Ches	N-Cyclohexyl-2-aminoethanesulfonic acid
CHO	Chinese hamster ovary
CPD	Cyclobutane pyrimidine dimer
CS	Cockayne syndrome
CSA / CSB	Cockayne syndrome group A / B
CTD	C-terminal domain
DDB1 / DDB2	DNA damage-binding protein 1 / 2
dL	Delta linker
DNA	Deoxyribonucleic acid
DNAse	Deoxyribonuclease
dNTP	Deoxynucleoside triphosphate
DRD	Damage recognition domain
dsDNA	Double stranded DNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
EMBL	European Molecular Biology Laboratory
ERCC1 - ERCC3	Excision repair cross-complementation group 1 - 3
ESRF	European Synchrotron Radiation Facility
EtOH	Ethanol
FPLC	Fast protein liquid chromatography

fs	Frameshift
HD1 / HD2	Helicase domain 1 / 2
Hepes	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HEWL	Hen egg white lysozyme
HR23B	UV excision repair protein RAD23 homolog B
IDA	Iminodiacetic acid
IEX	Ion exchange chromatography
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ISAC	Iterative stable alignment and clustering
Kan	Kanamycin
kb	Kilo bases
MAT1	Ménage à trois 1
MES	2-(N-morpholino)ethanesulfonic acid
mRNA	Messenger RNA
MWCO	Molecular weight cutoff
NAD+ / NADH	Nicotinamide adenine dinucleotide
NER	Nucleotide excision repair
NR	Nuclear receptor
nt	Nucleotides
nT	Non-tagged
OD_{600}	Optical density at 600 nm
opti	Codon optimized cDNA sequence
PAGE	Polyacrylamide gel electrophoresis
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PH	Pleckstrin homology
PIC	Preinitiation complex
Pipes	1,4-piperazinediethanesulfonic acid
PMSF	Phenylmethylsulfonyl fluoride
RFC	Replication factor C
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
RNAPI - RNAPIII	RNA polymerase I - III
RPA	Replication protein A

rpm	Rounds per minute
rRNA	Ribosomal RNA
SAD	Single wavelength anomalous diffraction
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
SLIC	Sequence and ligation independent cloning
ssDNA	Single stranded DNA
SSL1 / SSL2	Suppressor of stem-loop protein 1 / 2
TBP	TATA binding protein
TCEP	Tris(2-carboxyethyl)phosphine
TCR	Trancription coupled repair
TED	Tris-carboxymethyl ethylene diamine
TEMED	Tetramethylethylenediamine
TEV	Tobacco etch virus
Tfb	Transcription factor b
Tfb1 - Tfb5	Transcription factor b subunit 1 - 5
TFIIA - TFIIH / TFIIS	Transcription factor IIA - IIH / IIS
ThM	Thumb motif
Tris	Tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
TTD	Trichothiodystrophy
UV	Ultraviolet
UV-DDB	Ultraviolet radiation-DNA damage-binding protein
vWA	von Willebrand factor A
XP	Xeroderma pigmentosum
XPA - XPG	Xeroderma pigmentosum group A - G
XRCC1	X-ray repair cross-complementing protein 1

6.2. Standard Thermofluor screen

All buffers present in the standard Thermofluor screen are listed in Table 6-1. Concentration of all buffers is 0.1 M.

Table 6-1. Composition of the standard Thermofluor screen.

Buffer	pH
Citric acid	4.5
Acetic acid	4.6
Bis-Tris	5.5
Mes	5.5
Bis-Tris propane	6.0
Sodium cacodylate	6.0
Ada	6.5
Bis-Tris	6.5
Imidazole	6.5
Mes	6.5
Pipes	6.5
Sodium cacodylate	6.5
K/Na phosphate	6.8
Ada	7.0
Bis-Tris	7.0
Bis-Tris propane	7.0
Hepes	7.0
Mops	7.0
Pipes	7.0
Tris	7.0
Hepes	7.5
K/Na phosphate	7.5
Mops	7.5
Pipes	7.5
Tris	7.5
Bicine	8.0
Hepes	8.0
Imidazole	8.0
Taps	8.0
Tris	8.0
Bicine	8.5
Glycyl-glycine	8.5
Hepes	8.5
Tris	8.5
Bicine	9.0
Ches	9.0
Taps	9.0
Tris	9.0
Ches	9.5
Caps	9.8

6.3. Secondary structure predictions and sequence alignments

For secondary structure predictions three different web servers have been used: RaptorX, Phyre2, and I-TASSER. For sequence alignments between the protein sequences from

Homo sapiens and *Chaetomium thermophilum* three different web servers have been used: EMBOSS Needle, ClustalW2, and COBALT.

6.3.1. XPB

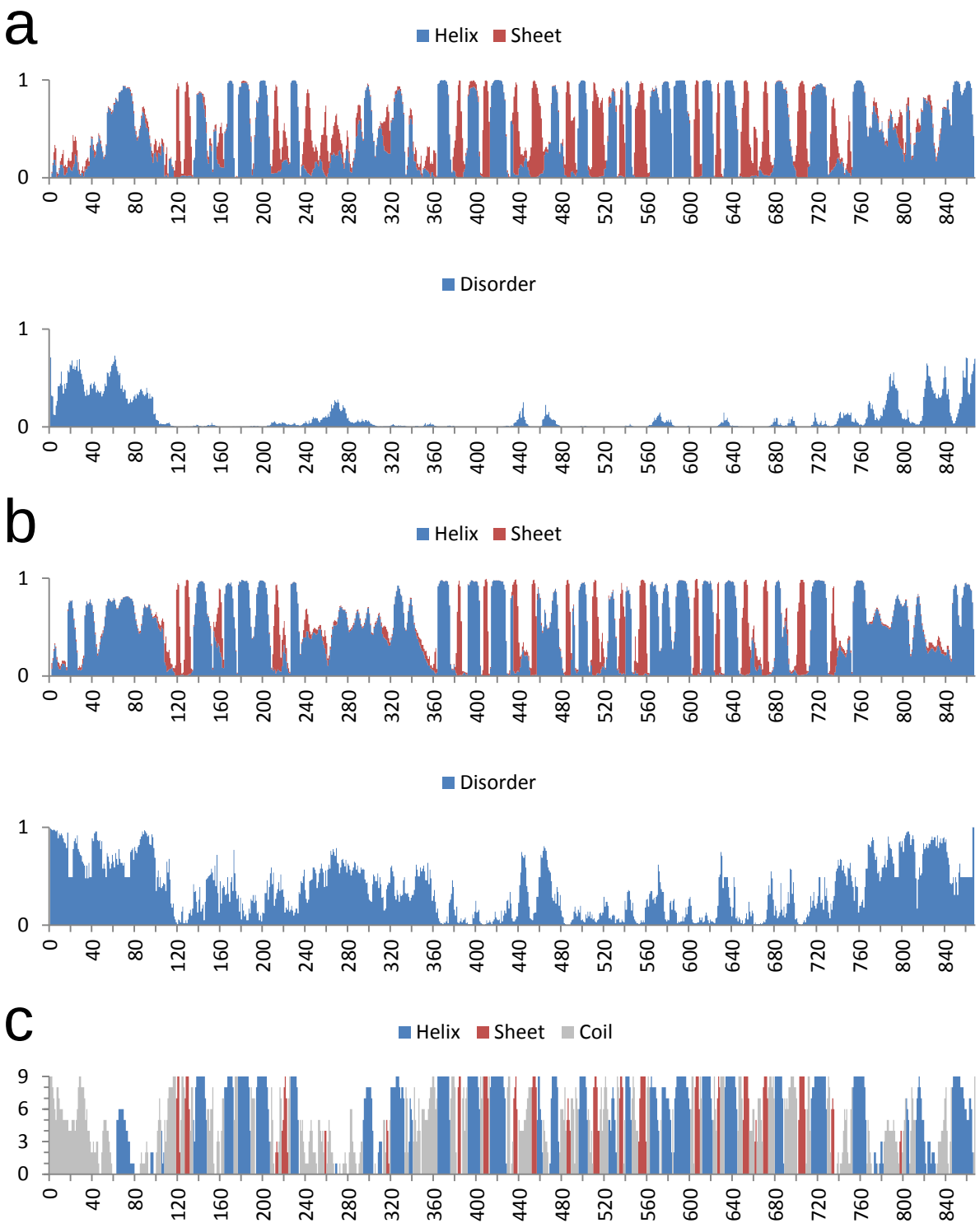
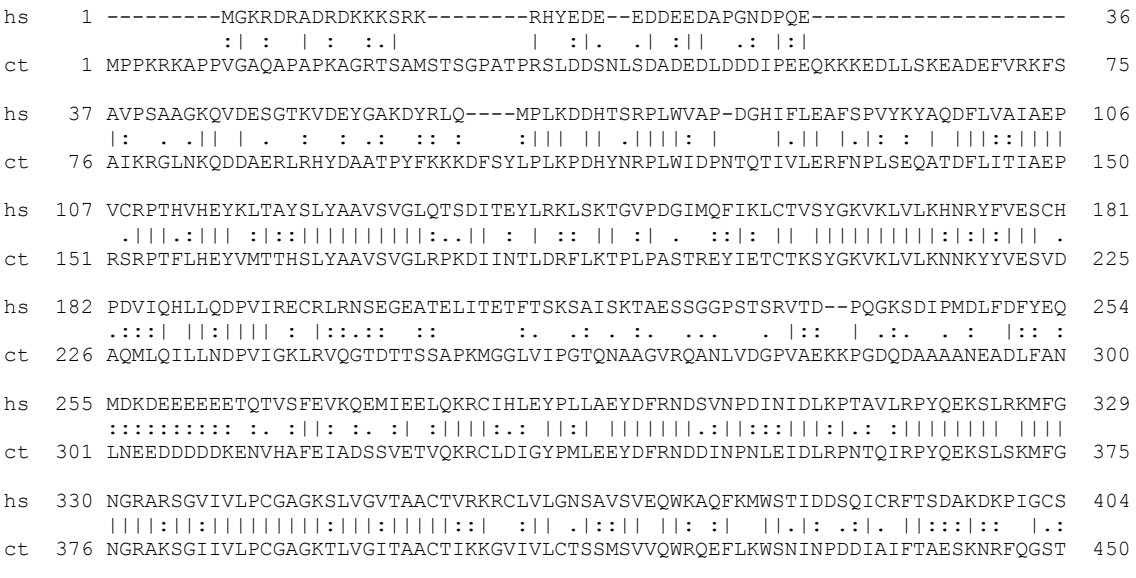


Figure 6-1. Secondary structure prediction of XPB. (a) RaptorX. **(b)** Phyre2. **(c)** I-TASSER.

Figure 6-2. Alignment between the human (hs) and *C. thermophilum* (ct) amino acid sequence of XPB with EMBOSS Needle.



```
hs  405 -VAISTYSMLGHTTKRSWEAERVMEWLKTQEWGLMILDEVHTIPAKMFRRVLTIVQAHCKLGLTATLVREDDKIV  478
         : ::|||||: :: :|| :::::|::|: :|||||:|||||.:||:||||||:: :::|.|||||||||:|||||:|
ct  451 GIIVTTYSMVTNSRERSHDSKKMMDFLRGREWGLMLLDEVHVVPAEMFRRVISSIKSHAKLGLTATLLREDDRIS  525

hs  479 DLNFLIGPKLYEANWMELQNNGYIAKVQCAEVWCPMSPEFYREYVAIKTKKRILLYTMNPNKFRACQFLIKFHER  553
        .||||||||||||||||||.::|:|||||||||||||..||| ||:  ::: :  ||:|||.||:|||:||::||
ct  526 HLNFLIGPKLYEANWMELSQQGHIAKVQCAEVWCPMPTEFYDEYLRANARMKRTLYAMNPRKFQACQYLINYHEA  600

hs  554 RNDKIIVFADNVFALKEYAIRLNKPYIYGPTSQGERMQILQNFKHNPKINTIFISKVGDTSFDLPEANVLIQISS  628
        |.|||||:|::::||:||::|:| :||| |||.||||:|:||:|||::||:|:||:||||:|||||. ||||||
ct  601 RGDKIIVFSDELYSLKQYALKLKKVFIYGGTSQAERMQVLENFQHNPEVNTLFLSKIGDTSLDLPEATCLIQISS  675

hs  629 HGGSRRQEAQRLGRVLRAKKGMVAEEYNAFFYSLVSQDTQEMAYSTKRQRFLVDQGYSFKVITKLAGMEEE-DLA  702
        | |||||||||||:||||:    | :|||||||||:|||||  ||:||| |||||||:|||||:||.:|:  |||
ct  676 HFGSRRQEAQRLGRILRAKR-RNDEGFNAFFYSLVSKDTQEMYYSSKRQAFLVDQGYAFKVITHLANIEQTPDLA  749

hs  703 FSTKEEQQQLLQKVLAATDLDAEEEVVAGEFGSRS---------SQASRRFGTMSSMSGADDTVYMEYHSSRSKA  768
        ||| :| ::|||:.|. .: .|||:| :.:: .|:         | . |   | :|.:||.:| .|:| :.: .:|
ct  750 FSTPQEVRELLQRTLVDNEKGAEEDVETDDLFGRTGRKKKAGALSGVRRTAGMLSELSGGQDMAYIEQNKAANRA  824

hs  769 PSKH---------------VHPLFKRFRK--------------  782
         .::                 .:||::.:
ct  825 LKQQGAGGGKSSKAAAMAEKSSFFKKIEREKEKSRAAARAARQ  867
```

Figure 6-3. Alignment between the human (hs) and *C. thermophilum* (ct) amino acid sequence of XPB with ClustalW2.

```
hs    1 MGKRDRAD---------------------------------RDKKKSRKRHYEDEEDDEEDAPGNDPQEAVPSAA   42
ct    1 MPPKRKAPPVGAQAPAPKAGRTSAMSTSGPATPRSLDDSNLSDADEDLDDDIPEEQKKKEDLLSKEADEFVRKFS   75

hs   43 ----GKQVDESGTKVDEYGA-------KDYRLQMPLKDDHTSRPLWVAPDGH-IFLEAFSPVYKYAQDFLVAIAE  105
ct   76 AIKRGLNKQDDAERLRHYDAATPYFKKKDFSY-LPLKPDHYNRPLWIDPNTQTIVLERFNPLSEQATDFLITIAE  149

hs  106 PVCRPTHVHEYKLTAYSLYAAVSVGLQTSDITEYLRKLSKTGVPDGIMQFIKLCTVSYGKVKLVLKHNRYFVESC  180
ct  150 PRSRPTFLHEYVMTTHSLYAAVSVGLRPKDIINTLDRFLKTPLPASTREYIETCTKSYGKVKLVLKNNKYYVESV  224

hs  181 HPDVIQHLLQDPVIRECRLRNSEGEATEL----ITETFTSKSAISKTAESSGGPSTSRVTDPQGKSDIPMDLFDF  251
ct  225 DAQMLQILLNDPVIGKLRVQGTDTTSSAPKMGGLVIPGTQNAAGVRQANLVDGPVAEK--KPGDQDAAAANEADL  297

hs  252 YEQMDKDEEEEEETQTVSFEVKQEMIEELQKRCIHLEYPLLAEYDFRNDSVNPDINIDLKPTAVLRPYQEKSLRK  326
ct  298 FANLNEEDDDDDKENVHAFEIADSSVETVQKRCLDIGYPMLEEYDFRNDDINPNLEIDLRPNTQIRPYQEKSLSK  372

hs  327 MFGNGRARSGVIVLPCGAGKSLVGVTAACTVRKRCLVLGNSAVSVEQWKAQFKMWSTIDDSQICRFTSDAKDKPI  401
ct  373 MFGNGRAKSGIIVLPCGAGKTLVGITAACTIKKGVIVLCTSSMSVVQWRQEFLKWSNINPDDIAIFTAESKNRFQ  447

hs  402 GC-SVAISTYSMLGHTTKRSWEAERVMEWLKTQEWGLMILDEVHTIPAKMFRRVLTIVQAHCKLGLTATLVREDD  475
ct  448 GSTGIIVTTYSMVTNSRERSHDSKKMMDFLRGREWGLMLLDEVHVVPAEMFRRVISSIKSHAKLGLTATLLREDD  522

hs  476 KIVDLNFLIGPKLYEANWMELQNNGYIAKVQCAEVWCPMSPEFYREYVAIKTKKRILLYTMNPNKFRACQFLIKF  550
ct  523 RISHLNFLIGPKLYEANWMELSQQGHIAKVQCAEVWCPMPTEFYDEYLRANARMKRTLYAMNPRKFQACQYLINY  597

hs  551 HERRNDKIIVFADNVFALKEYAIRLNKPYIYGPTSQGERMQILQNFKHNPKINTIFISKVGDTSFDLPEANVLIQ  625
ct  598 HEARGDKIIVFSDELYSLKQYALKLKKVFIYGGTSQAERMQVLENFQHNPEVNTLFLSKIGDTSLDLPEATCLIQ  672

hs  626 ISSHGGSRRQEAQRLGRVLRAKKGMVAEEYNAFFYSLVSQDTQEMAYSTKRQRFLVDQGYSFKVITKLAGMEEE-  699
ct  673 ISSHFGSRRQEAQRLGRILRAKRRN-DEGFNAFFYSLVSKDTQEMYYSSKRQAFLVDQGYAFKVITHLANIEQTP  746

hs  700 DLAFSTKEEQQQLLQKVLAATDLDAEEEVVAGEFGSRS---------SQASRRFGTMSSMSGADDTVYMEYHSSR  765
ct  747 DLAFSTPQEVRELLQRTLVDNEKGAEEDVETDDLFGRTGRKKKAGALSGVRRTAGMLSELSGGQDMAYIEQNKAA  821

hs  766 SKAPSKH---------------VHPLFKRFRK--------------  782
ct  822 NRALKQQGAGGGKSSKAAAMAEKSSFFKKIEREKEKSRAAARAARQ  867
```

Figure 6-4. Alignment between the human (hs) and *C. thermophilum* (ct) amino acid sequence of XPB with COBALT.

6.3.2. p62

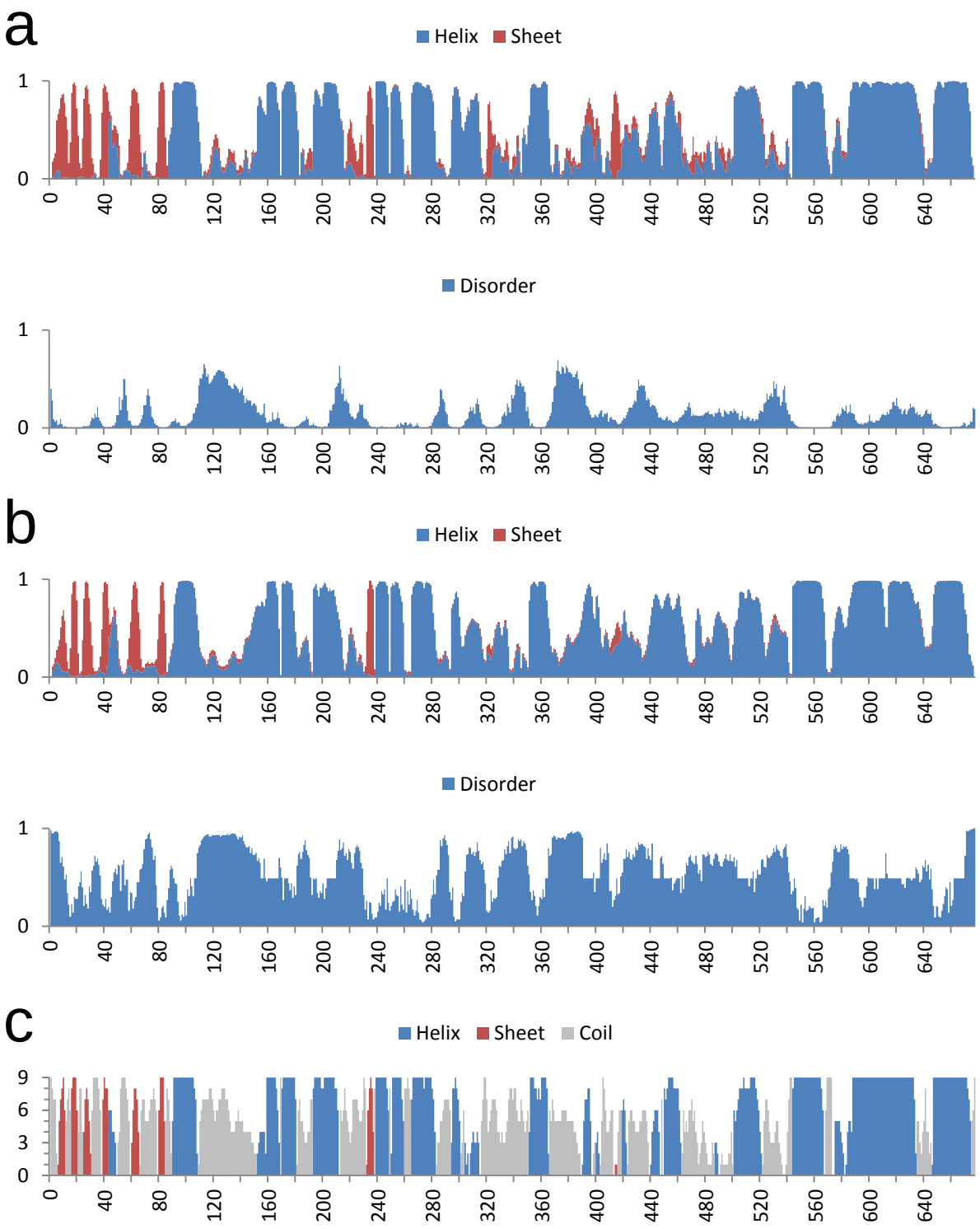


Figure 6-5. Secondary structure prediction of p62. (a) RaptorX. **(b)** Phyre2. **(c)** I-TASSER.

```
hs    1 MATSSEEVLLIVKKVRQKKQDGALYLMAER--IAWAP---EGKDRFTISHMYADIKCQKISPEGKAKIQLQLVLH   70
ct    1 -------MSIPRSQTTYKKKEGILTLTEDRKFLIWTPLPATGPP--TVSLALDNITNLQQTPPGSAKVILKFTER   66

hs   71 AGDTTN---------FHFSNESTAVKERDAVKDLLQQLL----------PK------------------------  102
ct   67 PRPNAEPGAPPPQYMFQFTHPTDARAEANAIRDLLSQLLAAARENDPNVPKPAAAAAAGGGGGGQSGATNGAPAA  141

hs  103 --------FKRKANK-ELEEKNRMLQEDPVLFQLYKDLVV--SQVISAE----EFWANRLNVNATDSSSTSNHKQ  162
ct  142 PASQAPRWFDDAALKADIDLQQSLMKKDKDLAQTYAEALALKPESIPTETFNAQFWSTRVGLLRAHAIELHQKKG  216

hs  163 DVGISAAFLADVRPQT-DGCNGLRYNLTSDIIESIFRTYPAVKMKYAENVPHNMTEKEFWTRFFQSHYFHRDR--  234
ct  217 AYNV----LSTIKPRTEDG--QLKLSISPEQIQMILQQHPIVRRIYNDNVP-KLSESEFWSRFFLSKLFKKLRGE  284

hs  235 ---LNTGSKDLFAECAKIDEK-GLKTMVSLGVKNP-LLDLTALEDKPLDEGYGISSVPSASNSKSI--KENSNAA  302
ct  285 RYTENDNPDPLFDKFLEADNSFGLASKIVAAQKVPHIIDLEANEE---NQG-GFKS----GNRKDVEMRPRANVP  351

hs  303 IIKRFNHHSAMVLAAGLRKQEAQNEQTSEPSNMDGNSGDADCFQPAVKRAKLQES------------IEYEDLGK  365
ct  352 IIKTLNSLSEKIMA----------------------NVAPADLLDPDPVTARVPASGLTEDDTRTFAQLSLRDLQG  405

hs  366 NNSVKTIALNLKKSDRYYHGPTPIQS-LQYATSQDI------------------------------INSFQSIR  408
ct  406 DAETARIILNVREQNQFFQNQAAAAAGLDGQASEDAKVFEHQVPSEVLFDVHADLETLDEDGAGGIDLHGSIGID  480

hs  409 QEMEA-----YTPKLTQVLSSSAASST------------ITALSPGGALMQGGTQQAINQMVPNDIQSELKHLYV  466
ct  481 DESESDPEGQGTSKVPHVGSRAARKQAQKQILEGMRKKRLETYGTGGAATSSLDDDTSPMSIPPDIAQRCYLTNA  555

hs  467 AVGELLRHFWSCFPVNTPFLEEKVVKMKSNLERFQVTKLCPFQEKIRR------------QYLSTNLVSHIEEML  529
ct  556 TTVEFLRQFWSCFLSTAP--DSQTANGRSNSSE-QMQQLAYYAETLRRSRERIDALAAEAEVIRQQVVERRKREI  627

hs  530 QTAYNKLH---TWQSRRLMKKT----------------------------  548
ct  628 MEHYERTRRKIRWQPPRGGRESVMALFEATLEAVRAALAVWGAVSGKGTV  677
```

Figure 6-6. Alignment between the human (hs) and *C. thermophilum* (ct) amino acid sequence of p62 with EMBOSS Needle.

```
hs    1 MATSSEEVLLIVKKVRQKKQDGALYLMAER--IAWAP-EGKDRFTISHMYADIKCQKISPEGKAKIQLQLVL---   69
ct    1 -------MSIPRSQTTYKKKEGILTLTEDRKFLIWTPLPATGPPTVSLALDNITNLQQTPPGSAKVILKFTERPR   68

hs   70 ----HAGDTTNFHFSNE--STAVKERDAVKDLLQQL---------------------------------------   99
ct   69 PNAEPGAPPPQYMFQFTHPTDARAEANAIRDLLSQLLAAARENDPNVPKPAAAAAAGGGGGGQSGATNGAPAAPA  143

hs  100 ----LPKFKRKANKELEEKNRMLQEDPVLFQLYKD---LVVSQVIS---AEEFWANRLNVNATDSSSTSNHKQDV  164
ct  144 SQAPRWFDDAALKADIDLQQSLMKKDKDLAQTYAEALALKPESIPTETFNAQFWSTRVGLLRAHAIELHQKKGAY  218

hs  165 GISAAFLADVRPQTDGCNGLRYNLTSDIIESIFRTYPAVKMKYAENVPHNMTEKEFWTRFFQSHYFHRDRL----  235
ct  219 NV----LSTIKPRTED-GQLKLSISPEQIQMILQQHPIVRRIYNDNVPK-LSESEFWSRFFLSKLFKKLRGERYT  287

hs  236 -NTGSKDLFAECAKIDEK-GLKTMVSLGVKNP-LLDLTALEDKPLDEGYGISSVPSASNSKSI--KENSNAAIIK  305
ct  288 ENDNPDPLFDKFLEADNSFGLASKIVAAQKVPHIIDLEANEE---NQG-GFKS----GNRKDVEMRPRANVPIIK  354

hs  306 RFNHHSAMVLAAGLRKQEAQNEQTSEPSNMDGNSGDADCFQPAVKRAKLQES------------IEYEDLGKNNS  368
ct  355 TLNSLSEKIMA----------------------NVAPADLLDPDPVTARVPASGLTEDDTRTFAQLSLRDLQGDAE  408

hs  369 VKTIALNLKKSDRYYHGPTPIQS-LQYATSQDI------------------------------INSFQSIRQEM  411
ct  409 TARIILNVREQNQFFQNQAAAAAGLDGQASEDAKVFEHQVPSEVLFDVHADLETLDEDGAGGIDLHGSIGIDDES  483

hs  412 EAY-----TPKLTQVLSSSAASST------------ITALSPGGALMQGGTQQAINQMVPNDIQSELKHLYVAVG  469
ct  484 ESDPEGQGTSKVPHVGSRAARKQAQKQILEGMRKKRLETYGTGGAATSSLDDDTSPMSIPPDIAQRCYLTNATTV  558

hs  470 ELLRHFWSCFPVNTPFLEEKVVKMKSNLERFQVTKLCPFQEKIRR------------QYLSTNLVSHIEEMLQTA  532
ct  559 EFLRQFWSCFLSTAP--DSQTANGRSNSSE-QMQQLAYYAETLRRSRERIDALAAEAEVIRQQVVERRKREIMEH  630

hs  533 YNKLHT---WQSRRLMKKT----------------------------  548
ct  631 YERTRRKIRWQPPRGGRESVMALFEATLEAVRAALAVWGAVSGKGTV  677
```

Figure 6-7. Alignment between the human (hs) and *C. thermophilum* (ct) amino acid sequence of p62 with COBALT.

6.3.3. p52

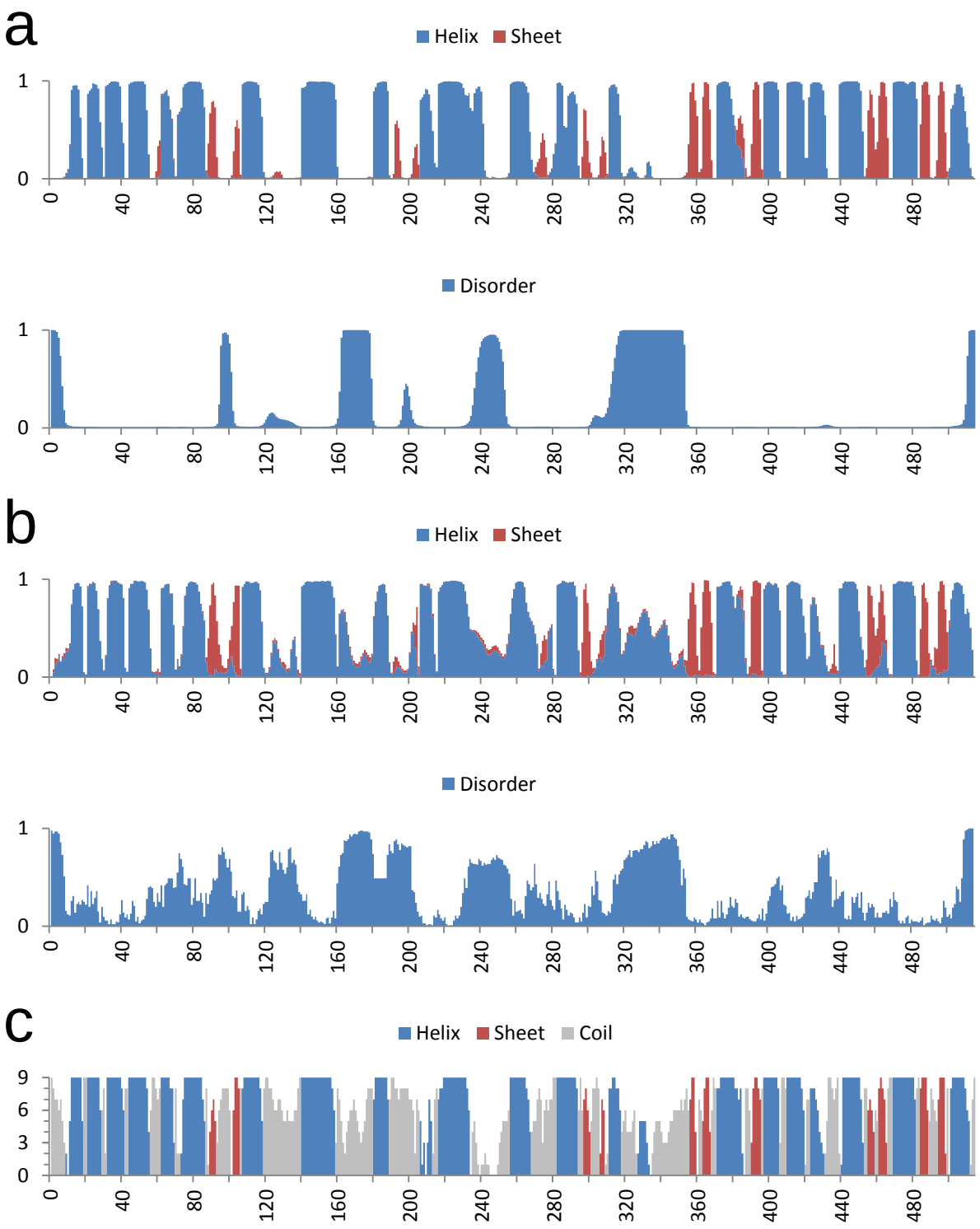


Figure 6-8. Secondary structure prediction of p52. (a) RaptorX. **(b)** Phyre2. **(c)** I-TASSER.

```
hs    1 --MESTPSRGLNRVHLQCRNLQEFLGGLSPGVLDRLYGHPATCLAVFRE-LPSLAKNWVMRMLFLEQPLPQAAVA   72
          :.:.|         |...|.::|..|......:||..|::..|:||. ||.|||.:|..:|::.||:..:.:.
ct    1 MSIPAVP---------QSYQLSDYLEKLPGTTFRKLYQQPSSAFAIFRRMLPPLAKVFVQALLYMPQPMLLSDLD   66

hs   73 LWVKKEFSKAQEESTGLLSGLRIWHTQLLPGGL---QGLILNPIFRQNLRIALLGGGKAWSDDTSQLGPDKHARD  144
          :||:.|....::.:..:|..|.|  .|:.|.|.   |.:.|...||.:||:||.||....|.......|.....|
ct   67 VWVRPEAKMHRDRALSILRSLHI--VQITPPGKDRPQEVQLTTNFRNSLRLALEGGAAHNSFGVPSSLPVDPRID  139

hs  145 VPSLDKYAEERWEVVLHFMV----------------GSPSAAVSQDLAQLLSQAGLMKSTEPGEPP--CITSAGF  201
          :..||.||.::||.:||::|               |.|.|:|..     |..||.:....|....  .||.|||
ct  140 IAFLDNYARKKWEDILHYVVSSVPVHGDAGPGGMGGGGPKASVKD-----LLLAGRLVERRPDTKTGIGITQAGF  209

hs  202 QFLLLDTPAQLWYFMLQYLQTA-QSRGM-----------------DLVEILSFLFQLSFSTLGKDYSVEGMSDSL  258
          .|||.:..||:|..:|.:|:.| |::..                 |.:|:|||||.|:...||:.|..:.:|::.
ct  210 TFLLQEANAQVWTLLLLWLEAADQAKAAAAAASGVDPKNAPPTKPDSIEMLSFLFMLASLELGRAYDTDALSETR  284

hs  259 LNFLQHLREFGLVFQRKRKSRRYYPTRLAINLSSGVS-----------------------GAGGTVHQPGFIVV  309
          .|.|..|.:|||::..:..:|:|:||||||..|:|..|                       ||..:.|: |.|::
ct  285 RNMLPALVDFGLIYIPREDTRQYFPTRLATTLTSSASALRSVSSGFTAATNNTANDASSLGGADPSAHK-GSIII  358

hs  310 ETNYRLYAYTESELQIALIALFSEMLYRFPNMVVAQVTRESVQQAIASGITAQQIIHFLRTRAHPVMLKQT----  380
          ||||||||||.|.||||::|||:.:..||..||..::||||::::||:.||||.|||.:|.:.||..|::..
ct  359 ETNYRLYAYTSSPLQIAVLALFTHLNMRFAGMVTGRLTRESIRRAISFGITADQIISYLASHAHEQMVRAAAAAG  433

hs  381 -PVLPPTITDQIRLWELERDRLRFTEGVLYNQFLSQVDFELLLAHARELGVLVFENSAKRLMVVTPAGHSDVKRF  454
           ||||||:.||||||:||.:|:|.:.|.|:..|.:..::..|..:|.|:||||:.:..||:...:.  ...:|.:
ct  434 RPVLPPTVVDQIRLWQLENERMRTSPGFLFKDFENVEEYMALAGYAEEIGVLVWRSDRKRMFFASK--FEQLRDY  506

hs  455 WKRQKHSS  462
          .|.:|...
ct  507 LKSRKKEG  514
```

Figure 6-9. Alignment between the human (hs) and *C. thermophilum* (ct) amino acid sequence of p52 with EMBOSS Needle.

```
hs    1 MESTPSRGLNRVHLQCRNLQEFLGGLSPGVLDRLYGHPATCLAVFR-ELPSLAKNWVMRMLFLEQPLPQAAVALW   74
          | |      .  |. :|.::|  |.  .: :|| :|:::.:|:||  ||.||| :|  :|:: ||:  : : :|
ct    1 -MSIP------AVPQSYQLSDYLEKLPGTTFRKLYQQPSSAFAIFRRMLPPLAKVFVQALLYMPQPMLLSDLDVW   68

hs   75 VKKEFSKAQEESTGLLSGLRIWHTQLLPGG-LQGLILNPIFRQNLRIALLGGGKAWSDDTSQLGPDKHARDVPSL  148
          |: | .  ::.: .:| .|:| :      .  | : |.. ||:.||:|| ||.   | ....  | .   |:. |
ct   69 VRPEAKMHRDRALSILRSLHIVQITPPGKDRPQEVQLTTNFRNSLRLALEGGAAHNSFGVPSSLPVDPRIDIAFL  143

hs  149 DKYAEERWEVVLHFMVGSP-----------SAAVSQDLAQLLSQAGLMKSTEPGEPP--CITSAGFQFLLLDTPA  210
          |:||.::|| :||::|.|            ... .:  .: |  || : . .|.  .   ||.||| ||| :: |
ct  144 DNYARKKWEDILHYVVSSVPVHGDAGPGGMGGGGPKASVKDLLLAGRLVERRPDTKTGIGITQAGFTFLLQEANA  218

hs  211 QLWYFMLQYLQTAQ--------SRGMD----------LVEILSFLFQLSFSTLGKDYSVEGMSDSLLNFLQHLRE  267
          |:| ::| :|::|:        : |:|           :|:||||| |:   ||: |..:.:|::  |:|  | :
ct  219 QVWTLLLLWLEAADQAKAAAAAASGVDPKNAPPTKPDSIEMLSFLFMLASLELGRAYDTDALSETRRNMLPALVD  293

hs  268 FGLVFQRKRKSRRYYPTRLAINLSSGVS-----------------------GAGGTVHQPGFIVVETNYRLYAYT  319
          |||::  :..:|:|:||||| .|:|..|                       |....  : | |::||||||||
ct  294 FGLIYIPREDTRQYFPTRLATTLTSSASALRSVSSGFTAATNNTANDASSLGGADPSAHKGSIIIETNYRLYAYT  368

hs  320 ESELQIALIALFSEMLYRFPNMVVAQVTRESVQQAIASGITAQQIIHFLRTRAHPVMLKQT-----PVLPPTITD  389
          .| ||||::|||:.:  ||..||..::||||:::||: ||||:||| :| ::||  |:: :     ||||||:.|
ct  369 SSPLQIAVLALFTHLNMRFAGMVTGRLTRESIRRAISFGITADQIISYLASHAHEQMVRAAAAAGRPVLPPTVVD  443

hs  390 QIRLWELERDRLRFTEGVLYNQFLSQVDFELLLAHARELGVLVFENSAKRLMVVTPAGHSDVKRFWKRQKHSS  462
          |||||:||.:|:| : |.|:::| .  ::  | .:|.|:||||:... ||::..:   ..::: : | :|:..
ct  444 QIRLWQLENERMRTSPGFLFKDFENVEEYMALAGYAEEIGVLVWRSDRKRMFFAS--KFEQLRDYLKSRKKEG  514
```

Figure 6-10. Alignment between the human (hs) and *C. thermophilum* (ct) amino acid sequence of p52 with ClustalW2.

```
hs    1 MESTPSRGLNRVHLQCRNLQEFLGGLSPGVLDRLYGHPATCLAVFR-ELPSLAKNWVMRMLFLEQPLPQAAVALW   74
ct    1 M-SIPAVP------QSYQLSDYLEKLPGTTFRKLYQQPSSAFAIFRRMLPPLAKVFVQALLYMPQPMLLSDLDVW   68

hs   75 VKKEFSKAQEESTGLLSGLRIWHTQLLPGG--LQGLILNPIFRQNLRIALLGGGKAWSDDTSQLGPDKHARDVPS  147
ct   69 VRPEAKMHRDRALSILRSLHIVQI-TPPGKDRPQEVQLTTNFRNSLRLALEGGAAHNSFGVPSSLPVDPRIDIAF  142

hs  148 LDKYAEERWEVVLHFMVGSP------------SAAVSQDLAQLLSQAGLM-KSTEPGEPPCITSAGFQFLLLDTP  209
ct  143 LDNYARKKWEDILHYVVSSVPVHGDAGPGGMGGGGPKASVKDLLLAGRLVERRPDTKTGIGITQAGFTFLLQEAN  217

hs  210 AQLWYFMLQYLQTAQS--------RGMDL----------VEILSFLFQLSFSTLGKDYSVEGMSDSLLNFLQHLR  266
ct  218 AQVWTLLLLWLEAADQAKAAAAAASGVDPKNAPPTKPDSIEMLSFLFMLASLELGRAYDTDALSETRRNMLPALV  292

hs  267 EFGLVFQRKRKSRRYYPTRLAINLSSGVSGAGGTVHQP----------------------GFIVVETNYRLYAY  318
ct  293 DFGLIYIPREDTRQYFPTRLATTLTSSASALRSVSSGFTAATNNTANDASSLGGADPSAHKGSIIIETNYRLYAY  367

hs  319 TESELQIALIALFSEMLYRFPNMVVAQVTRESVQQAIASGITAQQIIHFLRTRAHPVMLKQT-----PVLPPTIT  388
ct  368 TSSPLQIAVLALFTHLNMRFAGMVTGRLTRESIRRAISFGITADQIISYLASHAHEQMVRAAAAAGRPVLPPTVV  442

hs  389 DQIRLWELERDRLRFTEGVLYNQFLSQVDFELLLAHARELGVLVFENSAKRLMVVTPAGHSDVKRFWKRQKHSS  462
ct  443 DQIRLWQLENERMRTSPGFLFKDFENVEEYMALAGYAEEIGVLVWRSDRKRMFFASK--FEQLRDYLKSRKKEG  514
```

Figure 6-11. Alignment between the human (hs) and *C. thermophilum* (ct) amino acid sequence of p52 with COBALT.

6.3.4. p44

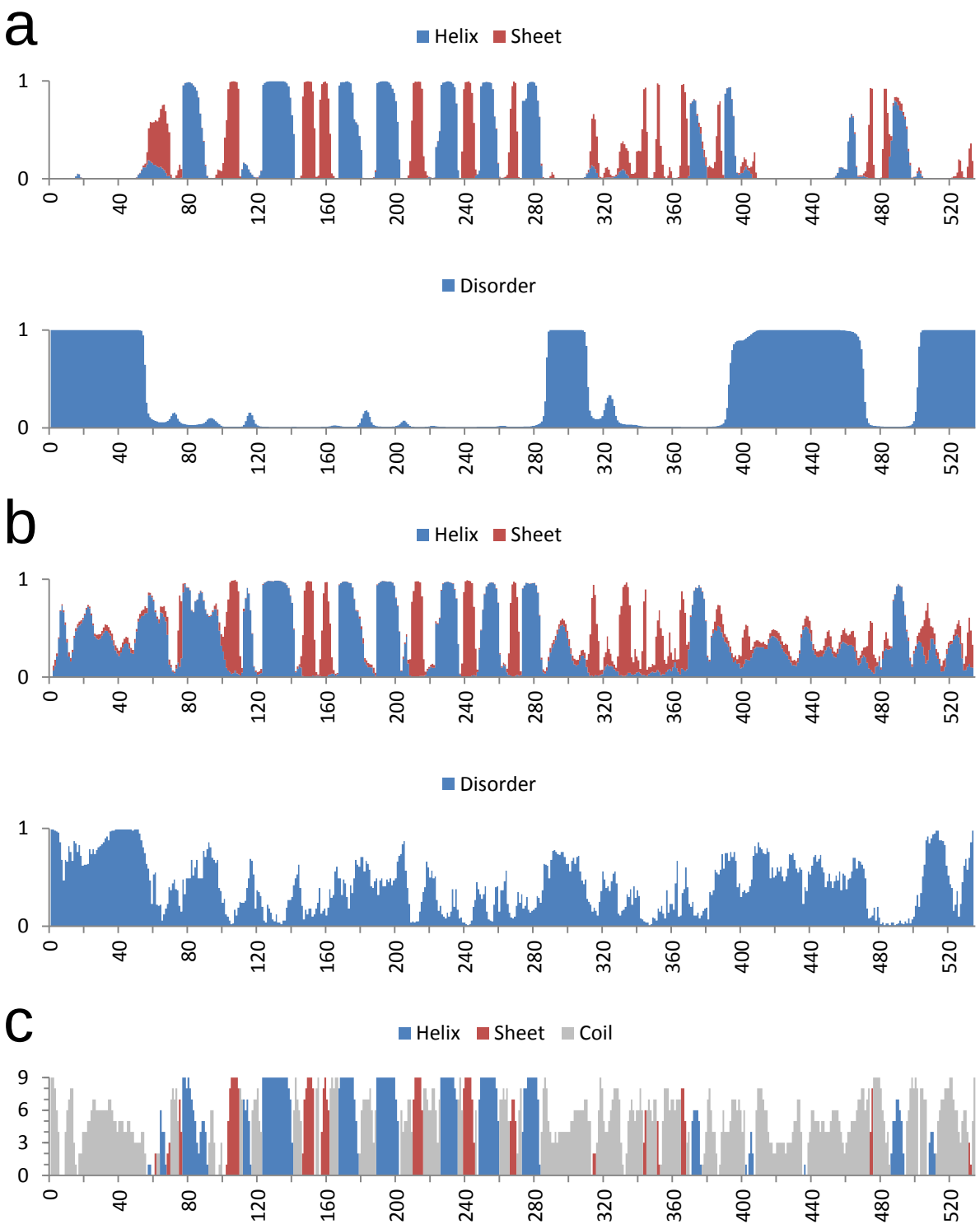


Figure 6-12. Secondary structure prediction of p44. (a) RaptorX. **(b)** Phyre2. **(c)** I-TASSER.

```
hs    1 -----------------------------------------------MDEEPERTK-RWEGGYERTWEILKEDESGS   29
                                                      .......|| .|| ..:|:||.:.|.|.||
ct    1 MADSDGEYVEDLSDDELHDHRPAEAGPHGARSKAGAGKKRDGKKGKKGSSRHTKAAWE-DIQRSWENVVETEDGS   74

hs   30 LKATIEDILFKAKRKRVFEHHGQVRLGMMRHLYVVVDGSRTMEDQDLKPNRLTCTLKLLEYFVEEYFDQNPISQI  104
        :  |||.::...||:|:.....::.|::|||.:|:|.|..|.::||.|||...||.....||.|||:||||||:
ct   75 I--TIEALIEAEKRRRLMRDTTPLQRGIIRHLVLVLDMSFAMAEKDLLPNRYLLTLNYAVDFVREYFEQNPISQM  147

hs  105 GIIVTKSKRAEKLTELSGNPRKHITSLK-KAVDMTCHGEPSLYNSLSIAMQTLKHMPGHTSREVLIIFSSLTTCD  178
        |||..:...|.:::::||||..||..|: .|......|.|||.|:|.:....|.|.|.|.:|||||::.:|.:.|
ct  148 GIIAMRDGIAVRVSDMSGNPADHIERLRFWAEHQEPQGNPSLQNALEMCRGALYHTPSHGTREVLIVYGALLSSD  222

hs  179 PSNIYDLIKTLKAAKIRVSVIGLSAEVRVCTVLARETG----GTYHVILDESHYKELLTHHVSPPPASSSS----  245
        |.:|::.|..|...:|||:|:||:|:|.||..|...|.    .||.|.|.|.|::||......||||::||
ct  223 PGDIHETISNLVKDRIRVTVVGLAAQVAVCAELCTRTNHGDDSTYAVALHEQHFRELFLAATIPPPATASSATDK  297

hs  246 ------------ECSLIRMGFPQHTIASLSDQDAKPSFSMAHLDGNTEPGLTLGGYFCPQCRAKYCELPVECKIC  308
                     :.||:.||||..|:||.|    ..|....|...:.|      ||.|.:||||.|.||.||..|
ct  298 NGANGNANAASTDASLLMMGFPSRTLASAS----HVSLCACHSRPSRE------GYLCTRCRAKVCRLPAECPAC  362

hs  309 GLTLVSAPHLARSYHHLFPLDAFQEIPLEEYNGER--FCYGC-----------------------QGE-------  351
        ||||:.:.||||||||||..:.|:...|....:  .|:.|                         ||:
ct  363 GLTLILSTHLARSYHHLFPLKGWVEVSWAEARKSKQVGCFACLAPFPLPPAPGSEKTGKEPTQKTQGQAQQPPQE  437

hs  352 -------------------------------LKDQHVYVCAVCQNVFCVDCDVFVHDSLHCCPGC-----------  385
                                       :.:...|.|..|...||:|||||.|:.:|.||||
ct  438 RQGSSSNSNNAKKTTGISLATALPEARAVGVSESGRYKCPTCGKHFCIDCDVFAHEVIHNCPGCQADMRPKQDAS  512

hs  386 IHKIPAPSGV------------  395
        .:.|...:|:
ct  513 SNNIGPANGLNNVVDGDAMVLD  534
```

Figure 6-13. Alignment between the human (hs) and *C. thermophilum* (ct) amino acid sequence of p44 with EMBOSS Needle.

```
hs    1 ----------------------------------------------MDEEPERTKR-WEGGYERTWEILKEDESGS   29
ct    1 MADSDGEYVEDLSDDELHDHRPAEAGPHGARSKAGAGKKRDGKKGKKGSSRHTKAAWED-IQRSWENVVETEDGS   74

hs   30 LKATIEDILFKAKRKRVFEHHGQVRLGMMRHLYVVVDGSRTMEDQDLKPNRLTCTLKLLEYFVEEYFDQNPISQI  104
ct   75 I--TIEALIEAEKRRRLMRDTTPLQRGIIRHLVLVLDMSFAMAEKDLLPNRYLLTLNYAVDFVREYFEQNPISQM  147

hs  105 GIIVTKSKRAEKLTELSGNPRKHITSLKK-AVDMTCHGEPSLYNSLSIAMQTLKHMPGHTSREVLIIFSSLTTCD  178
ct  148 GIIAMRDGIAVRVSDMSGNPADHIERLRFWAEHQEPQGNPSLQNALEMCRGALYHTPSHGTREVLIVYGALLSSD  222

hs  179 PSNIYDLIKTLKAAKIRVSVIGLSAEVRVCTVLARETGG----TYHVILDESHYKELLTHHVSPPPA--------  241
ct  223 PGDIHETISNLVKDRIRVTVVGLAAQVAVCAELCTRTNHGDDSTYAVALHEQHFRELFLAATIPPPATASSATDK  297

hs  242 --------SSSSECSLIRMGFPQHTIASLSDQDAKPSFSMAHLDGNTEPGLTLGGYFCPQCRAKYCELPVECKIC  308
ct  298 NGANGNANAASTDASLLMMGFPSRTLASASHV----SLCACHSRPSRE------GYLCTRCRAKVCRLPAECPAC  362

hs  309 GLTLVSAPHLARSYHHLFPLDAFQEIPLEEYNGERF--CYGCQGEL-----------------------------  352
ct  363 GLTLILSTHLARSYHHLFPLKGWVEVSWAEARKSKQVGCFACLAPFPLPPAPGSEKTGKEPTQKTQGQAQQPPQE  437

hs  353 -------------------------------KDQHVYVCAVCQNVFCVDCDVFVHDSLHCCPGCIHKIPAPSGV-  395
ct  438 RQGSSSNSNNAKKTTGISLATALPEARAVGVSESGRYKCPTCGKHFCIDCDVFAHEVIHNCPGCQADMRPKQDAS  512

hs  396 ----------------------  395
ct  513 SNNIGPANGLNNVVDGDAMVLD  534
```

Figure 6-14. Alignment between the human (hs) and *C. thermophilum* (ct) amino acid sequence of p44 with COBALT.

6.4. Codon optimized XPB cDNA sequence

An alignment between the wild-type and the codon optimized cDNA sequence of XPB is given below (Figure 6-15).

```
wt      1 ATGCCTCCCAAGCGAAAGGCCCCTCCTGTGGGCGCGCAAGCTCCCGCACCCAAGGCTGGACGCACTTCTGCCATG   75
          |||||.||.||.||.||.||.||.|||||.||.||.||.||.||.||.||.||.||.||.||.||.||.|||
opti    1 ATGCCGCCGAAACGCAAAGCACCACCTGTTGGTGCACAGGCACCGGCGCCTAAAGCCGGTCGTACCTCGGCGATG   75

wt     76 TCTACTTCCGGGCCGGCTACTCCCCGCAGCCTGGATGATTCAAATTTGTCGGACGCCGATGAAGACCTCGATGAC  150
          ...||.||.||.||.||.|||||.||||||.|.|||||||...|||.||...|||||.||||||||.||.||||||.
opti   76 AGCACCTCAGGTCCTGCGACTCCGCGCAGCTTAGATGATAGTAATCTGAGCGACGCAGATGAAGATCTGGATGAT  150

wt    151 GATATCCCAGAGGAGCAGAAGAAGAAAGAGGATCTGCTGAGCAAGGAAGCCGATGAATTTGTCAGGAAGTTCTCG  225
          ||.||||||||.||||||||.||||||||.||.||||.|.|||||.|||||.||.||.||||||..|.||.||....
opti  151 GACATCCCAGAAGAGCAGAAAAAGAAAGAAGACCTGTTAAGCAAAGAAGCGGACGAGTTTGTGCGTAAATTTAGC  225

wt    226 GCAATAAAGCGCGGGCTCAACAAGCAAGACGACGCTGAGCGTCTCCGGCATTATGATGCTGCCACTCCGTACTTC  300
          ||||||.||||||.||.||.||.||||||||.||.||.||||||.||.||.||||||.||||||.|||||||||
opti  226 GCAATTAAGCGTGGTCTGAATAAGCAAGATGATGCCGAGCGCCTTCGCCATTACGATGCCGCCACCCCGTACTTC  300

wt    301 AAGAAGAAAGACTTCTCCTACCTCCCCCTCAAACCCGACCACTACAATCGCCCTTTGTGGATTGACCCCAATACG  375
          ||||||.||.||||||.||.||.||.||.||.||||||.||||||.||.||.||||||||.||||||.||.||.||||||.
opti  301 AAGAAAAAGGACTTTTCGTATCTGCCGCTGAAACCAGACCATTATAACCGCCCTTTATGGATCGATCCGAATACC  375

wt    376 CAAACCATCGTACTGGAACGGTTCAACCCTCTGTCGGAGCAGGCAACTGATTTTCTCATCACGATTGCTGAGCCC  450
          ||.||||||||.||.||.||.||.||.||...||...||.||.||.||.||||||.||.||.|||||||||||||||.
opti  376 CAGACCATCGTGCTCGAGCGCTTTAATCCGTTGAGCGAACAAGCCACCGATTTCCTGATTACGATTGCTGAGCCG  450

wt    451 AGATCGCGCCCAACGTTTCTGCACGAGTATGTCATGACGACTCACAGCCTCTATGCTGCAGTCTCGGTTGGTCTT  525
          .|....||.||.||.||.|||||||||||||||.||||||||.|||...||.||.||.||.||.||.|||||||||
opti  451 CGCAGCCGTCCGACTTTCCTGCACGAGTATGTGATGACGACCCACTCTCTGTACGCCGCCGTGTCAGTTGGTCTT  525

wt    526 CGGCCCAAAGATATCATAAACACCCTTGATCGCTTTTTGAAAACGCCACTTCCTGCATCGACTCGGGAGTATATA  600
          ||.||.|||||||||||.||.|||||||||||||||.||||||||.|||||||||.||....||.||.||.||.||.
opti  526 CGTCCGAAAGATATCATTAATACCCTTGATCGCTTTCTGAAAACCCCACTTCCGGCGAGTACGCGCGAATACATT  600

wt    601 GAGACGTGCACAAAGAGCTACGGCAAGGTGAAGCTAGTTCTCAAGAACAACAAGTATTACGTAGAGAGCGTGGAC  675
          ||.||.|||||.||||...|||||||||||||||.||.||||||.||.|||||||||||||||||||.||...|||.||.
opti  601 GAAACTTGCACCAAGTCATACGGCAAGGTGAAACTGGTTCTGAAAAACAACAAGTATTACGTCGAATCCGTCGAT  675

wt    676 GCCCAGATGCTGCAGATCCTGCTAAACGATCCTGTCATTGGCAAGCTGCGCGTTCAGGGAACTGACACGACCAGC  750
          ||.|||||.|||||||||.||||.||.||||||.||.||||||.||.||||||||||.||.||.||||||.||||||.
opti  676 GCGCAGATGTTGCAGATCTTGCTGAATGATCCGGTGATTGGAAAACTGCGCGTTCAAGGCACCGACACCACCAGT  750

wt    751 TCTGCTCCCAAGATGGGGGGCCTCGTTATCCCCGGCACCCAGAACGCCGCTGGCGTCCGGCAGGCAAACCTGGTG  825
          ...||.||.||.|||||.|||.|.||.||.||.||||||.|||||.|||||.|||||.||.||.||.||.||||||.
opti  751 AGCGCCCCAAAAATGGGCGGCTTGGTGATTCCTGGCACACAGAATGCCGCGGGCGTTCGCCAAGCGAATCTGGTA  825

wt    826 GACGGCCCGGTTGCCGAAAAGAAGCCCGGTGACCAGGACGCTGCTGCTGCCAACGAGGCCGACTTGTTCGCCAAT  900
          ||.||.||||||||.||||||||.||.|||||.|||||.||.||.||.||.||.||.||.||.||..|||||||.||.
opti  826 GATGGACCGGTTGCGGAAAAGAAACCTGGTGATCAGGATGCGGCCGCGGCTAATGAAGCTGATCTGTTCGCTAAC  900

wt    901 TTAAATGAGGAAGATGATGATGACGACAAAGAAAACGTCCATGCCTTTGAGATTGCGGATTCGTCTGTCGAGACG  975
          |||||.||.||||||||.|||||.||||||||.|||||.|||||.|||||.|||||||||||||||.|||||.||.||.|||
opti  901 TTAAACGAAGAAGATGACGATGATGACAAAGAGAACGTTCATGCGTTTGAGATTGCGGACTCGTCCGTAGAAACG  975

wt    976 GTCCAGAAGCGCTGCCTTGACATTGGCTACCCCATGCTGGAAGAGTATGATTTCAGAAATGACGACATAAACCCC 1050
          ||.||.||.|||||||||||||||||||||.||.|||||||||||||||.||.||||.|.||||||.||||||.|||||||
opti  976 GTGCAAAAACGCTGCCTTGACATTGGCTATCCGATGCTGGAAGAGTACGACTTCCGTAATGATGACATCAACCCC 1050

wt   1051 AACCTGGAGATTGATCTCCGCCCCAATACCCAGATTCGTCCCTACCAAGAGAAGAGCCTCAGCAAGATGTTTGGC 1125
          |||.|.||||||||.||.||||||.||||||.||||||.||.||.||||||.||.||....||.||.||||||||.||.
opti 1051 AACTTAGAGATTGACCTGCGCCCGAATACTCAGATCCGCCCGTACCAGGAAAAATCTCTGAGTAAGATGTTCGGT 1125

wt   1126 AATGGCCGAGCGAAAAGCGGTATCATTGTTCTGCCATGCGGCGCCGGCAAGACGCTGGTCGGCATCACCGCTGCC 1200
          ||.||.||.||.|||||||||||||||||||.||||||.||||||||.||||||.||...|.||.||.||.||.||.||.
opti 1126 AACGGGCGTGCCAAAAGCGGTATCATTGTGCTGCCCTGCGGCGCAGGCAAAACCTTAGTAGGAATTACTGCCGCG 1200

wt   1201 TGCACCATCAAGAAGGGCGTCATTGTGCTTTGCACCAGCTCCATGTCTGTCGTCCAATGGCGTCAAGAATTCCTC 1275
          ||||||||.||||||.||||||.||.||.||||..|||||||.||.||.||.||||||||||||.||..|.
opti 1201 TGCACCATTAAGAAAGGCGTTATCGTGCTGTGTACAAGCAGCATGTCCGTGGTGCAGTGGCGTCAAGAGTTTTTG 1275
```

```
wt   1276 AAGTGGTCCAACATCAATCCCGATGATATTGCCATCTTCACCGCGGAGAGCAAGAATAGGTTTCAAGGGAGCACT 1350
          ||.||||||||||||||||.|||||.|||||.||.||.||.||.|||...||.||..|.|||||.||....||.
opti 1276 AAATGGTCCAACATCAATCCTGATGACATTGCGATTTTTACGGCCGAGTCGAAAAACCGTTTTCAGGGCTCAACC 1350

wt   1351 GGTATCATTGTCACCACGTACTCGATGGTCACAAACAGTCGTGAGCGATCGCATGACTCTAAGAAGATGATGGAT 1425
          ||.||.|||||.||.||.|||||||||||||||.|||..||||||.||.||.|||||.||.||.||||||||||.
opti 1351 GGGATTATTGTGACGACATACTCGATGGTCACCAACTCTCGTGAACGCTCACATGATTCGAAAAAGATGATGGAC 1425

wt   1426 TTCTTGAGGGGCAGAGAATGGGGTTTGATGTTGCTCGACGAGGTACACGTCGTGCCGGCCGAGATGTTCCGGCGC 1500
          |||.|..|.||..|.|||||||..|.|||.||||.||.||.||.|||||.||.||.||.||.||||||||.|||
opti 1426 TTCCTTCGTGGTCGTGAATGGGGCCTCATGCTGCTGGATGAAGTCCACGTGGTACCAGCGGAAATGTTCCGTCGC 1500

wt   1501 GTCATCTCGTCCATCAAGTCACACGCCAAGCTCGGTTTGACGGCGACACTTCTTCGTGAGGACGATAGGATTTCG 1575
          ||.||.||.||.|||||.||.|||||.||.||.||..||||.|||||.||..|.|||||.||.|||.|.|||||||
opti 1501 GTAATTTCCTCAATCAAATCTCACGCAAAACTGGGCCTGACTGCGACGCTGTTGCGTGAAGATGATCGCATTTCG 1575

wt   1576 CATCTTAACTTCTTGATCGGGCCCAAGCTTTATGAGGCCAACTGGATGGAGCTCTCACAGCAGGGCCACATTGCG 1650
          |||.|.||.||..||||.||.||||||.|.|||||.||.||.||||||||.||....||||||.||.||.|||||||
opti 1576 CATTTGAATTTTCTGATTGGACCCAAATTGTATGAAGCAAATTGGATGGAACTGAGTCAGCAAGGGCATATTGCG 1650

wt   1651 AAGGTGCAGTGTGCTGAGGTTTGGTGCCCGATGCCTACAGAGTTCTACGACGAGTACCTGCGCGCCAATGCCCGG 1725
          ||.||.||||||||.||.||.||||||||.|||||.||||||||.||.||.||.||.||.||.|||||.||.||.
opti 1651 AAAGTCCAGTGTGCAGAAGTCTGGTGCCCCATGCCGACAGAGTTTTATGATGAATATCTCCGTGCCAACGCACGC 1725

wt   1726 ATGAAGCGCACGTTATACGCGATGAACCCGCGCAAGTTCCAGGCTTGCCAGTACCTCATCAACTACCATGAAGCT 1800
          |||||.|||||.||.||||||.|||||.||.||.||||||.|||||.||.||||||.||.||||||||.||.|||||||
opti 1726 ATGAAACGCACTTTGTACGCCATGAATCCCCGTAAGTTTCAGGCCTGTCAGTATCTTATCAACTATCACGAAGCT 1800

wt   1801 CGCGGAGACAAAATTATTGTCTTCTCGGACGAGCTCTACTCTCTGAAGCAGTATGCCCTGAAGCTCAAGAAAGTC 1875
          |||||.||||||||||.|||||.||.||.||||||||....||||||.||||||.||||||||.||||||.||.|||
opti 1801 CGCGGCGACAAAATTATCGTCTTTTCCGATGAGCTCTATAGCCTGAAACAGTACGCCCTGAAACTCAAAAAGGTC 1875

wt   1876 TTCATCTACGGCGGCACCAGCCAGGCCGAACGTATGCAGGTCTTGGAGAACTTCCAACACAACCCCGAGGTGAAC 1950
          ||||||||.||.||||||....||.||||||||||||||||||.||.||.||.||||||||.||||||.||.||.|||
opti 1876 TTCATCTATGGTGGCACGTCGCAAGCCGAACGTATGCAGGTTTTAGAAAATTTCCAACATAACCCGGAAGTAAAC 1950

wt   1951 ACTCTTTTCCTCTCCAAGATCGGCGATACTTCTCTCGATCTGCCTGAGGCAACCTGCCTGATCCAGATCTCTTCT 2025
          ||.||.||.||||||.||.||.||.||||||.||..|.||||||||||.||.||||||.||.||.|||||||...
opti 1951 ACACTGTTTCTCTCTAAAATTGGTGATACCTCATTAGATCTGCCTGAAGCGACCTGTCTCATTCAGATCTCTAGC 2025

wt   2026 CACTTCGGCTCGCGTCGTCAGGAGGCGCAGCGCCTGGGCCGCATCCTTCGGGCGAAGAGGCGTAACGACGAAGGG 2100
          ||.||||||....||||||||.||.||.||||||..|.||.||.||.||||||.||.|||.|.||.||||||||||||
opti 2026 CATTTCGGTAGCCGTCGTCAAGAAGCACAGCGTTTAGGTCGTATTCTTCGCGCAAAGCGCCGCAACGACGAAGGG 2100

wt   2101 TTCAACGCCTTCTTCTACTCCCTCGTCTCAAAGGACACCCAAGAAATGTACTACTCCTCTAAGCGTCAAGCGTTC 2175
          ||||||||.||||||.||...|.|.||.||.||.||||||.||||||||||||||....||.||||||.||.|||||||
opti 2101 TTCAACGCGTTCTTTTATAGCTTAGTGTCCAAAGACACTCAAGAAATGTACTATAGTTCGAAGCGCCAGGCGTTC 2175

wt   2176 CTTGTCGACCAGGGCTACGCCTTCAAGGTAATCACCCACCTCGCGAATATCGAACAAACTCCCGACTTGGCCTTC 2250
          .|.||.||.||||||.||||||.||.||.||||||.||.||||||.||.||.||||||.||||||.||.||..|.||.||.
opti 2176 TTAGTGGATCAGGGATACGCGTTTAAAGTAATTACGCACCTTGCCAACATCGAGCAAACGCCGGATCTCGCGTTT 2250

wt   2251 TCCACGCCCCAGGAAGTCCGCGAACTCCTCCAGCGCACGCTGGTCGACAACGAGAAGGGCGCCGAGGAGGATGTC 2325
          ..|||.||.||||||||.||||||.||||||||||||.||.||.||.||||||.||.||.||.||.||.||.
opti 2251 AGCACCCCGCAGGAAGTTCGCGAGCTCCTGCAGCGCACGCTTGTAGATAATGAGAAAGGTGCAGAAGAAGACGTT 2325

wt   2326 GAGACCGACGATCTCTTCGGCCGGACAGGTCGCAAGAAGAAGGCGGGCGCCCTCTCCGGTGTGCGCCGCACGGCC 2400
          ||||||||.||||||.||.||.||.||.||.||||||.||||||.||.||.|||..|....||.||||||.||||||.||.
opti 2326 GAGACCGATGATCTGTTTGGGCGCACCGGGCGCAAAAAGAAAGCCGGTGCGTTAAGTGGCGTGCGTCGCACAGCG 2400

wt   2401 GGCATGCTGAGCGAGCTCAGCGGCGGCCAGGACATGGCGTACATCGAGCAGAACAAGGCTGCCAACCGCGCACTC 2475
          ||.|||.|.||.||.||.||.||||||||.||||||||.||.||.||.||.||||||||.||.||.||.||||||..|.
opti 2401 GGAATGTTAAGTGAACTGAGTGGCGGCCAAGACATGGCATATATTGAACAGAACAAAGCAGCTAATCGCGCCTTG 2475

wt   2476 AAGCAGCAGGGTGCTGGTGGCGGTAAGAGCAGCAAGGCGGCTGCGATGGCCGAGAAGAGTTCTTTCTTCAAGAAG 2550
          ||.||||||.||.||||||.||||||.....||||||.||.||.||.||.||.||.|||.....||||||.||||||.
opti 2476 AAACAGCAAGGCGCTGGCGGCGGCAAGTCTTCAAAGGCCGCGGCGATGGCCGAAAAGTCGAGCTTCTTTAAGAAA 2550

wt   2551 ATCGAGCGCGAGAAGGAGAAGAGCAGGGCCGCGGCCCGCGCGGCGAGGCAGTAA 2604
          ||.||||||||||.||.||.....|.||.||.||.||.||.|||.|.||||||
opti 2551 ATTGAGCGCGAGAAAGAAAAATCACGTGCTGCAGCTCGTGCAGCGCGTCAGTAA 2604
```

Figure 6-15. Alignment between the wild-type (wt) and codon optimized (opti) cDNA sequence of XPB.

6.5. Further expression constructs

In course of this work, further expression constructs of the TFIIH core subunits have been designed and attempted. Table 6-2 gives an overview of these constructs and the results from the conducted steps.

Table 6-2. Further expression constructs.

Construct	Vector	Cloning	Expression/Purification	Crystallization	Comments
XPB					
XPB_K392R	pBADM-11	Done	bound to chaperone		Walker A mutant
XPB_F143S	pBADM-11	Done			Disease related mutant
XPB_T163P	pBADM-11	Done			Disease related mutant
XPBopti_K392R	pBADM-11	Done			Walker A mutant
XPBopti_F143S		Primers ordered			Disease related mutant
XPBopti_T163P		Primers ordered			Disease related mutant
XPB_1-345_T163P	pBADM-11	Done			Disease related mutant
XPB_60-E	pBADM-11	Done	bound to chaperone		
XPB_60-768	pBADM-11	Done	bound to chaperone		
XPB_60-345	pBADM-11	Done	Successful	No crystals	
XPB_116-E	pBADM-11	Done	bound to chaperone		
XPB_286-747	pBADM-11	Done	bound to chaperone		DRD + HD1 + HD2
XPB_293-654	pBADM-11	Done	bound to chaperone		Corresponds to Archaeoglobus fulgidus XPB
XPB_346-747	pBADM-11	Done	bound to chaperone		HD1 + HD2
XPB_541-777	pBADM-11	Done	Successful		Corresponds to 4ERN
p62					
p62_dL	pETM-11	Done	Successful	No crystals	Residues 109-145 replaced by SNGNG linker
p52					
p52nT_M455C	pETM-11	Done			Introduction of cysteine to enable disulfide bridge to p8
p52_1-321_dAA	pETM-11	Done			Deletion of A121 + A122 in p52_1-321
p52nT_248-464	pETM-11	Done	Co-expressed with XPB_60-345; no homogenous species after SEC		
p52_1-464		Primers ordered			
p52_1-454		Primers ordered			
p52_1-435		Primers ordered			
p52_1-429		Primers ordered			
p52_1-235		Primers ordered			
p52_19-E		Primers ordered			With and without tag
p52_248-E		Primers ordered			With and without tag
p52_326-E		Primers ordered			
p52_345-E		Primers ordered			
p52_354-E		Primers ordered			
p44					
p44_LL	pBADM-11	Done	Co-expressed with p62nT; successful	No crystals	Residues 410-468 replaced by SNGNG linker ("long loop")
p44nT_286-E	pBADM-11	Done			
p44nT_286-367	pBADM-11	Done			
p34					
p34nT_1-277	pETM-11	Done			

7. Acknowledgment

I would like to express my gratitude to Prof. Caroline Kisker. Having studied to become a teacher at school initially, I had a hard time to get a doctoral position. But Caroline gave me this opportunity without prejudices. I am thankful for her support and trust, giving me the freedom to shape my research projects after my own visions and ideas.

I would like to thank Prof. Thomas Müller and Dr. Katrin Paeschke for their supervision, advice, time, and discussions during our meetings. I am really grateful that they shared their experience with me.

Dr. Jochen Kuper deserves a special acknowledgment. His brilliant mind is admirable and he is a role model for me. His ideas and advice really helped me in pushing my research forward. I am very thankful for his support not only scientifically but also personally.

I would like to thank Dr. Franz-Xaver Kober and Dr. Maria Hirschbeck for introducing me to all the basic laboratory techniques necessary in our field. I really learned a lot from them. I am grateful that they involved me in their research projects, resulting into fruitful collaborations in both cases. Furthermore, I would like to thank Prof. Hermann Schindelin for introducing me into the fascinating field of crystallography and all the advice regarding model building and interpretation of electron density maps.

I am very grateful to all members (present and past) of the TFIIH team for their contributions to this work and the fruitful teamwork. Those are Agnes Elias, Dr. Dominik Schmitt, Elisabeth Schönwetter, Dr. Florian Sauer, Gudrun Sander, Dr. Jochen Kuper, and Stefan Peißert.

The electron microscopy studies were only possible through different collaborations. In this respect, I would like to thank Prof. Christian Spahn, Dr. Thorsten Mielke, Ferdinand Krupp, Justus Loerke, and Jörg Bürger. Furthermore, I thank Prof. Bettina Böttcher. I owe special thanks to Christian Kraft for all his help in grid preparation and imaging at the electron microscope.

Finally, I thank all past and present lab members of the structural biology division for the great working atmosphere, numerous synchrotron trips, meetings, discussions, and lab-outings, making my graduation such a great time and experience.

8. Curriculum vitae

9. List of publications

Koelmel, W., Kuper, J., Kisker, C. (in preparation) Usage of cesium for experimental phasing.

Skinner, J. J.*, Wang, S.*, Lee, J., Ong, C., Sommese, R, Sivaramakrishnan, S., **Koelmel, W.**, Hirschbeck, M., Schindelin, H., Kisker, C., Lorenz, K., Sosnick, T., Rosner, M. R. (in revision) A conserved salt bridge theft triggered by phosphorylation regulates the protein interactome.

Radu, L.*, Schoenwetter, E.*, Braun, C., Marcoux, J., **Koelmel, W.**, Schmitt, D. R., Kuper, J., Sanglier-Cianferani, S., Egly, J. M., Poterszman, A.; Kisker, C. (2017) The intricate network between the p34 and p44 subunits is central to the activity of the transcription/DNA repair factor TFIIH. *Nucleic Acids Research*, doi: 10.1093/nar/gkx743

Jung, L. A.*, Gebhardt, A.*, **Koelmel, W.**, Ade, C. P., Walz, S., Kuper, J., von Eyss, B., Letschert, S., Redel, C., d'Artista, L., Biankin, A., Zender, L., Sauer, M., Wolf, E., Evan, G., Kisker, C., Eilers, M. (2017) OmoMYC blunts promoter invasion by oncogenic MYC to inhibit gene expression characteristic of MYC-dependent tumors. *Oncogene* **36**(14), 1911-1924.

Kober, F. X.*, **Koelmel, W.***, Kuper, J., Drechsler, J., Mais, C., Hermanns, H. M., Schindelin, H. (2013) The crystal structure of the protein-disulfide isomerase family member ERp27 provides insights into its substrate binding capabilities. *Journal of Biological Chemistry* **288**(3), 2029-2039.

*These authors contributed equally

10. Affidavit

Affidavit

I hereby confirm that my thesis entitled “Structural and functional characterization of TFIIH from *Chaetomium thermophilum*” 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.

Place, Date

Signature

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation „Strukturelle und funktionale Charakterisierung von TFIIH aus *Chaetomium thermophilum*“ eigenständig, d.h. insbesondere selbständig 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.

Ort, Datum

Unterschrift