



**Structural Characterization of the TFIID Subunits
p34 and p44 from *C. thermophilum***

**Strukturelle Charakterisierung der TFIID Untereinheiten
p34 und p44 aus *C. thermophilum***

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

submitted by
Dominik Schmitt

from
Heidelberg

Würzburg 2014

Submitted on:

Office Stamp

Members of the *Promotionskomitee*:

Chairperson: Prof. Dr. Ulrike Holzgrabe

Primary Supervisor: Prof. Dr. Caroline Kisker

Supervisor (2nd): Prof. Dr. Thomas Müller

Supervisor (3rd): Prof. Dr. Hanspeter Naegeli

Supervisor (4th): PD Dr. Robert Hock

Date of Public Defense:

Date of Receipt of Certificate:

*At the last dim horizon, we search among ghostly errors of observations
for landmarks that are scarcely more substantial.*

The search will continue. The urge is older than history.

It is not satisfied and it will not be oppressed.

Edwin Hubble (1889 - 1953)

Astronomer

Abstract

Several important cellular processes, including transcription, nucleotide excision repair and cell cycle control are mediated by the multifaceted interplay of subunits within the general transcription factor II H (TFIIH). A better understanding of the molecular structure of TFIIH is the key to unravel the mechanism of action of this versatile protein complex within these pathways. This becomes especially important in the context of severe diseases like xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy, that arise from single point mutations in some of the TFIIH subunits.

In an attempt to structurally characterize the TFIIH complex, we harnessed the qualities of the eukaryotic thermophile *Chaetomium thermophilum*, a remarkable fungus, which has only recently been recognized as a novel model organism. Homologues of TFIIH from *C. thermophilum* were expressed in *E. coli*, purified to homogeneity and subsequently utilized for crystallization trials and biochemical studies.

The results of the present work include the first crystal structure of the p34 subunit of TFIIH, comprising the N-terminal domain of the protein. The structure revealed a *von Willebrand Factor A* (vWA) like fold, which is generally known to be involved in a multitude of protein-protein interactions. Structural comparison allowed to delineate similarities as well as differences to already known vWA domains, providing insight into the role of p34 within TFIIH. These results indicate that p34 assumes the role of a structural scaffold for other TFIIH subunits via its vWA domain, while likely serving additional functions, which are mediated through its C-terminal zinc binding domain and are so far unknown.

Within TFIIH p34 interacts strongly with the p44 subunit, a positive regulator of the XPD helicase, which is required for regulation of RNA Polymerase II mediated transcription and essential for eukaryotic nucleotide excision repair. Based on the p34 vWA structure putative protein-protein interfaces were analyzed and binding sites for the p34 p44 interaction suggested. Continuous crystallization efforts then led to the first structure of a p34 p44 minimal complex, comprising the N-terminal vWA domain of p34 and the C-terminal C4C4 RING domain of p44. The structure of the p34 p44 minimal complex verified the previous hypothesis regarding the involved binding sites. In addition, careful analysis of the complex interface allowed to identify critical residues, which were subsequently mutated and analyzed with respect to their significance in mediating the p34 p44 interaction, by analytical size exclusion chromatography, electrophoretic mobility shift assays and isothermal titration calorimetry. The structure of the p34 p44 complex also revealed a binding mode of the p44 C4C4 RING domain, which differed from that of other known RING domains in several aspects, supporting the hypothesis that p44 contains a novel variation of this domain.

Zusammenfassung

Zelluläre Prozesse, wie beispielsweise die Transkription, die Nukleotid-Exzisionsreparatur und die Kontrolle des Zellzyklus sind abhängig vom vielschichtigen Zusammenspiel der zehn Protein-Untereinheiten des allgemeinen Transkriptionsfaktors II H (TFIIH). Zur Aufklärung der genauen Funktion dieses Komplexes ist ein besseres Verständnis seiner molekularen Struktur essentiell. Besondere Bedeutung erhält der TFIIH dabei im Hinblick auf verschiedene schwerwiegende Krankheiten, wie z.B. Xeroderma pigmentosum (XP), Cockayne-Syndrom (CS) und Trichothiodystrophie (TTD), die als Folge von einzelnen Punkt-Mutationen in bestimmten Untereinheiten des Komplexes entstehen.

In der vorliegenden Arbeit wurden zur strukturellen Charakterisierung der TFIIH Untereinheiten p34 und p44 die homologen Proteine aus *Chaetomium thermophilum* verwendet. Hierbei handelt es sich um einen eukaryotischen und thermophilen Pilz, der erst kürzlich als neuer und vielversprechender Modellorganismus an Bedeutung gewann. Die TFIIH Homologe aus *C. thermophilum* wurden rekombinant exprimiert, gereinigt und anschließend für Kristallisations-Versuche eingesetzt. Darüber hinaus wurden die Proteine mittels verschiedener biochemischer Verfahren analysiert.

Die erzielten Resultate beinhalten unter anderem die erste Kristall-Struktur der p34 Untereinheit des TFIIH und zeigen eine von *Willebrand Faktor A* (vWA) ähnliche Domäne im N-terminalen Bereich des Proteins. Vergleiche mit bereits bekannten vWA Proteinen liefern Gemeinsamkeiten sowie Unterschiede und erlauben erste Einblicke in die Funktion der p34 Untereinheit innerhalb des TFIIH Komplexes. Die gewonnenen Erkenntnisse legen nahe, dass p34 über seine vWA Domäne anderen TFIIH Untereinheiten als strukturelles Gerüst dient, während die C-terminale Zinkfinger-Domäne des Proteins sehr wahrscheinlich zusätzliche Aufgaben übernimmt, die bisher noch nicht genau bekannt sind.

Innerhalb des TFIIH Komplexes ist p34 eng mit der p44 Untereinheit assoziiert. Letztere ist als positiver Regulator der XPD Helikase bekannt, die im Rahmen der RNA Polymerase II vermittelten Transkription und der eukaryotischen Nukleotid-Exzisionsreparatur eine entscheidende Rolle spielt. Basierend auf der erzielten p34ct vWA Struktur wurden verschiedene Interaktions-Flächen zwischen p34 und p44 analysiert und mögliche Bindestellen für die beiden Proteine ermittelt. Weitere Kristallisations-Experimente ermöglichten schließlich die Aufklärung der Struktur eines p34 p44 Minimal-Komplexes, bestehend aus der N-terminalen vWA Domäne von p34 und der C-terminalen C4C4 RING Domäne von p44. Die gewonnenen Struktur-Daten bestätigten die zuvor ermittelte Bindestelle der beiden Proteine. Eine genauere Untersuchung der Kontakt-Fläche zwischen p34 und p44 lieferte darüber hinaus entscheidende Hinweise auf besonders wichtige Interaktions-Bereiche und Aminosäuren, die im Folgenden mutiert wurden, um deren Bedeutung für die Komplexbildung zu ermitteln. Mit Hilfe der analytischen Größenausschluss-Chromatographie, elektrophoretischer Mobilitäts-Verlagerungs-Assays und isothermaler Titrations-Kalorimetrie konnten hierbei verschiedene Aminosäuren identifiziert werden, die für eine stabile p34 p44 Interaktion erforderlich sind. Ferner zeigte die Struktur des p34 p44 Minimal-Komplexes eine Bindungsweise der p44 C4C4 RING Domäne, die sich von der anderer, bereits bekannter RING Domänen in verschiedenen Punkten unterschied. Diese Erkenntnis bestätigt die zuvor aufgestellte Hypothese, dass es sich im Falle von p44 um eine neue Variante der bereits gut charakterisierten RING Domäne handelt.

Table of Contents

Abstract	A
Zusammenfassung	B
Table of Contents.....	C
I. Introduction	1
I.1 DNA Damage and Repair	1
I.1.1 The Importance of DNA Repair	1
I.1.2 DNA Repair Mechanisms.....	2
I.1.3 Eukaryotic Nucleotide Excision Repair (NER)	3
I.2 Basic Concepts of DNA Transcription	4
I.3 The General Transcription Factor II H (TFIIH).....	5
I.3.1 Molecular Composition of the TFIIH Complex	5
I.3.2 Structural Knowledge on TFIIH Subunits.....	6
I.4 The TFIIH Complex in DNA Repair and Transcription.....	7
I.4.1 TFIIH in Nucleotide Excision Repair.....	7
I.4.2 The Role of TFIIH in Transcription and Cell Cycle Control.....	8
I.4.3 Crosstalk between DNA Repair and Transcription	8
I.5 Diseases Caused by Mutations in TFIIH Subunits	9
I.6 The Fungus <i>C. thermophilum</i>	10
II. Material	11
II.1 Equipment	11
II.1.1 Instruments	11
II.1.2 Consumables.....	13
II.2 Chemicals and Biochemical Substances	14
II.2.1 Chemicals.....	14
II.2.2 Ready-To-Use Kits	15
II.2.3 Crystallization Screens	16
II.2.4 Primers and DNA Substrates	16
II.2.5 Enzymes.....	17
II.3 Cell Culture.....	18
II.3.1 Bacterial and Insect Cell Strains.....	18
II.3.2 Vectors for Protein Expression.....	18
II.3.3 Bacterial Culture Media and Media Supplements	19
II.3.4 Insect Cell Culture Media and Media Supplements	19
II.4 Buffers and Solutions.....	20
II.4.1 Buffers and Solutions for Purification of Proteins	20
II.4.2 Buffers and Solutions for Biochemical Assays.....	20
II.4.3 Buffers and Solutions for Gel Electrophoresis and Sample Staining	21

II.5 Software and Databases	22
II.5.1 Software on Local Hardware	22
II.5.2 Software on the World Wide Web.....	23
II.5.3 Databases	23
III. Methods.....	24
III.1 Molecular Biology	24
III.1.1 Polymerase Chain Reaction (PCR).....	24
III.1.2 DNA Isolation	25
III.1.3 DNA Double Digestion.....	25
III.1.4 DNA Gel Electrophoresis.....	25
III.1.5 DNA Cloning and Ligation	26
III.1.6 Colony-PCR.....	27
III.1.7 Site-Directed Mutagenesis	28
III.1.8 DNA Sequencing.....	28
III.1.9 DNA Hybridization.....	28
III.1.10 Bacmid DNA Generation.....	29
III.1.11 Bacmid DNA Isolation.....	29
III.2 Microbiology	30
III.2.1 Transformation of Bacteria.....	30
III.2.2 Growing of Bacteria for Protein Expression.....	30
III.2.3 Growing and Maintenance of Insect Cell Cultures.....	31
III.2.4 Transfection of Insect Cells.....	31
III.2.5 Plaque Assay.....	31
III.2.6 Virus Amplification.....	32
III.2.7 Growing of Insect Cells for Protein Expression	32
III.3 Protein Biochemistry	33
III.3.1 Cell Lysis.....	33
III.3.2 Purification of Proteins and Chromatography Techniques.....	33
III.3.3 Analysis of Proteins via SDS-PAGE.....	35
III.3.4 Ultra-Filtration and Storage of Protein Samples	35
III.3.5 Analytical Size Exclusion Chromatography (A-SEC)	35
III.3.6 Electrophoretic Mobility Shift Assay (EMSA)	36
III.3.7 Isothermal Titration Calorimetry (ITC)	37
III.4 Spectroscopy.....	38
III.4.1 Determination of DNA and Protein Concentration	38
III.4.2 Thermofluor (TF)	38
III.4.3 Circular Dichroism (CD)	39
III.4.4 Multi Angle Light Scattering (MALS)	39
III.4.5 Mass Spectrometry (MS).....	40

III.5 X-Ray Crystallography	41
III.5.1 General Introduction	41
III.5.2 Principles of X-Ray Crystallography	41
III.5.3 Phase Determination	42
III.5.4 Protein Crystallization	43
III.5.5 Data Collection.....	44
III.5.6 Data Processing and Structure Determination	45
III.5.7 Model Building and Refinement	45
IV. Results and Discussion	46
IV.1 Towards Crystallization of TFIIH Subunits from <i>H. sapiens</i>	46
IV.1.1 Initial Attempts with human XPD, XPB, p44 and p52	46
IV.1.2 The human p44 Protein.....	47
IV.1.3 The p34h p44h Minimal Complex (hMC).....	48
IV.2 Characterization of TFIIH Subunits from <i>C. thermophilum</i>	51
IV.2.1 Purification of p34ct Variants.....	51
IV.2.2 Purification of p44ct Variants.....	53
IV.2.3 Preparation of the p34ct p44ct Minimal Complex (MC).....	55
IV.2.4 DNA Binding Studies with p34ct and p44ct	57
IV.2.5 Thermostability of p34ct and p44ct in Solution.....	59
IV.2.6 Oligomeric State of p34ct and p44ct in Solution	62
IV.3 Structure Determination of DPS - A Successful Failure	64
IV.4 Crystallization of p44ct Variants - Getting Closer	66
IV.5 Crystal Structure of p34ct - A Novel vWA like Domain.....	67
IV.5.1 Crystallization of p34ct.....	67
IV.5.2 Structure Determination of p34ct.....	68
IV.5.3 Crystallization of p34ct 1-277	69
IV.5.4 Structure Determination of p34ct 1-277 - The p34ct vWA like Domain.....	70
IV.5.5 Analysis of p34ct Crystal Packing and Elucidation of Binding Sites for p44ct.....	74
IV.5.6 Comparison of p34ct vWA to other vWA Domains	76
IV.5.7 The Multifaceted Roles of vWA Domains.....	78
IV.5.8 Implications for Protein-Protein Interactions within TFIIH	80
IV.6 Crystal Structure of p34ct p44ct MC - A Novel Binding Mode for RING Domains	81
IV.6.1 Crystallization of p34ct p44ct MC	81
IV.6.2 Structure Determination of p34ct p44ct MC.....	82
IV.6.3 Comparison of the p44ct C4C4 Domain to other RING Domains	84
IV.6.4 Analysis of p34ct p44ct MC Crystal Packing	86
IV.7 Analysis of the p34ct p44ct Complex Interface	87
IV.7.1 Deciphering of Interaction Sites within p34ct p44ct MC.....	87
IV.7.2 Purification of p34ct and p44ct Variants.....	89
IV.7.3 Characterization of p34ct and p44ct Variants.....	90

IV.7.4 Determination of Critical Residues.....	92
IV.7.5 Determination of Binding Affinities	95
V. Conclusion and Outlook.....	97
V.1 The TFIIH Subunits p34 and p44 from <i>C. thermophilum</i>	97
V.2 The p34ct vWA Domain	98
V.3 The p34ct p44ct Minimal Complex.....	99
V.4 Implications from the p34ct p44ct Minimal Complex.....	100
V.5 The Role of the p34ct and p44ct Zinc Finger Domains	101
V.6 The Importance of TFIIH Research.....	102
VI. Literature References	103
VII. Protein Data Bank References	118
VIII. Figures and Tables	119
VIII.1 List of Figures	119
VIII.2 List of Tables	120
IX. Abbreviations.....	121
X. Appendix	124
X.1 Table of Amino Acids.....	124
X.2 Detailed Maps of the Vectors used for Protein Expression.....	125
X.2.1 pBADM-11	125
X.2.2 pETM-11.....	125
X.2.3 pFastBac HTa	126
X.2.4 pFastBac Dual	126
X.3 Detailed Information on p34ct and p44ct Constructs	127
X.3.1 Constructs of p34ct	127
X.3.2 Constructs of p44ct	127
X.4 Secondary Structure Prediction of p34ct using the Phyre ² Algorithm	128
X.5 Secondary Structure Prediction of p44ct using the Phyre ² Algorithm	129
X.6 Data Collection and Refinement Statistics for p34ct.....	130
X.7 Data Collection and Refinement Statistics for p34ct p44ct MC.....	131
Acknowledgment.....	132
Affidavit	133
Curriculum Vitae.....	134

I. Introduction

I.1 DNA Damage and Repair

I.1.1 The Importance of DNA Repair

Each and every type of cellular life is shaped on the basis of a genomic blueprint, which is commonly stored in the form of a four letter chemical code intrinsic to deoxyribonucleic acid (DNA). This ancient and remarkably stable molecule carries all the vital instructions for the formation and proper functioning of both single cells as well as multi-cellular organisms. However, though inherently robust, the integrity of DNA is constantly challenged by a plethora of exogenous and endogenous DNA damaging agents, in turn resulting in a variety of DNA lesions that can have both immediate and long term consequences if left unrepaired (Figure I.1).

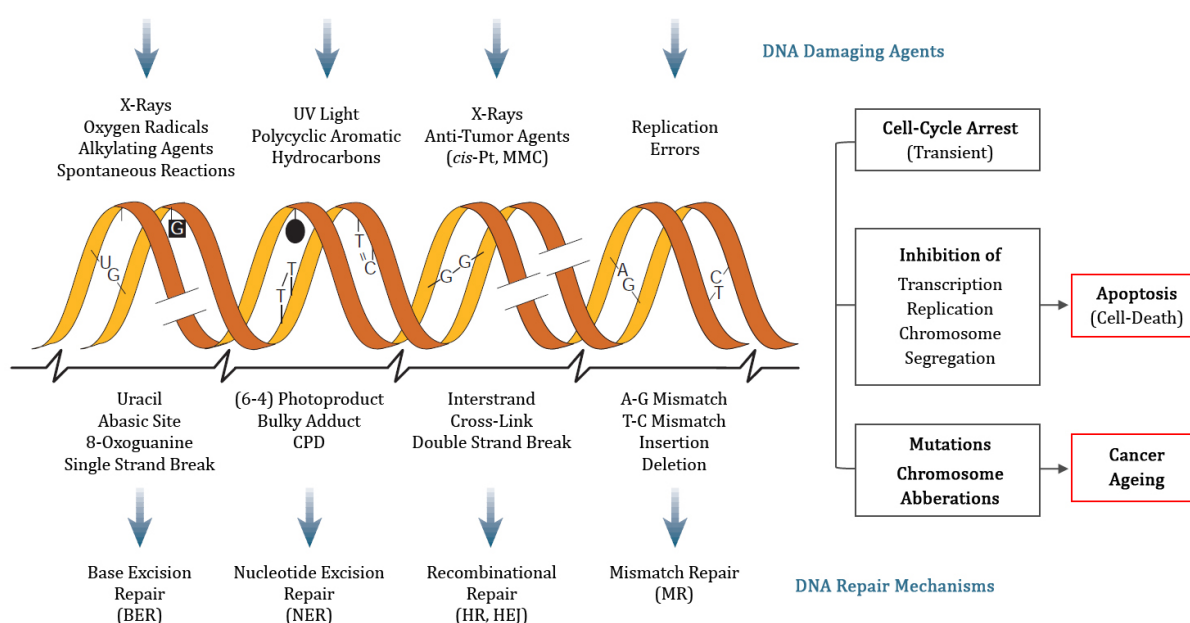


Figure I.1: DNA Damage and Repair Mechanisms. Both exogenous and endogenous DNA damaging agents (top) give rise to various types of DNA damages (middle), which ultimately lead to severe cellular consequences if left unrepaired (middle, right). To counter the adverse effects of the introduced lesions all organisms have evolved DNA repair mechanisms dedicated towards a specific type of damage (bottom). This scheme has been adapted by permission from Macmillan Publishers Ltd: [Nature](#) [1], ©2001.

For instance, the adverse effects of short-wave electromagnetic radiation, like x-rays and UV light, can lead to the formation of pyrimidine dimers and induce both single strand as well as double strand breaks within the helical DNA duplex [1]. Such lesions are practically unavoidable due to the partially short-waved character of natural sun light, to which we are exposed to on a daily basis. Furthermore, genotoxic chemicals present in the environment, our food or cigarette smoke, for example, are sufficiently potent to alter the chemical structure of DNA, introducing severe and often bulky adducts [2]. In addition to these exogenous agents, DNA damage may also arise as a result of metabolic processes, which often release reactive oxygen species (ROS) as by-products. Several ROS molecules like superoxide anions, hydrogen peroxide and hydroxyl radicals can then react with DNA to generate a multitude of oxidative modifications [3]. Moreover, spontaneous decay and hydrolysis of glycosidic bonds alone account for up to 10,000 abasic sites per human cell per day [3].

The cellular machinery itself is also prone to errors and contributes to the number of accumulated lesions. In this regard otherwise normal enzymatic reactions can have unwanted side effects, like misguided DNA methylation via S-Adenosyl methionine (SAM), while subtle imprecision during replication often leads to the incorporation of wrong or incorrectly paired nucleotides. The latter, though the driving force of evolution, alters the sequence of DNA and hence the information stored within, which, in the worst of cases, can result in the loss of genes and their function due to deleterious non-sense or frameshift mutations, respectively.

In a cellular context the ultimate effects of DNA damage are manifold, depending on the amount and type of the lesions introduced (Figure I.1). While minor modifications may simply stall replication and lead to a transient cell-cycle arrest, more severe lesions can irreversibly block transcription and induce apoptosis. Excessive accumulation of malign mutations can in turn promote the development of cancer and may also result in premature aging related diseases [1].

I.1.2 DNA Repair Mechanisms

To overcome the adverse effects of DNA damage, or at least alleviate their consequences, all organisms have evolved mechanisms of DNA repair, which are often dedicated towards a specific type of lesion (Figure I.1). While O⁶-alkylguanine adducts are removed by direct damage reversal (DDR) via alkylguanine transferases, other modifications of DNA bases are usually addressed by the base excision repair (BER) pathway. Here, damaged bases are excised by DNA glycosylases, which hydrolyze the N-glycosidic bond between the base and the sugar phosphate backbone of the DNA to generate an abasic site [3]. The latter is then replaced by a new nucleotide via the action of DNA Polymerase β and the remaining nick is sealed by DNA Ligase III.

In contrast, errors introduced during replicative transcription, like single base pair mismatches or nucleotide insertions and deletions, respectively, are often instantly repaired by the 3' to 5' proofreading activity of the associated DNA polymerases. In addition, translesion synthesis can be employed by specialized polymerases to effectively bypass specific lesions. This allows for the continuation of the otherwise stalled replication process, while the damaged bases are either used as template or looped out [4]. The remaining erroneous incorporations are then subject to the mismatch repair system (MMR), which is in essence conserved from bacteria to humans, although the mechanism of discrimination between the chemically unaltered base and the opposing lesion site differ in prokaryotes compared to eukaryotic species [3].

Structurally more complex and demanding lesions, such as interstrand cross-links or double strand breaks, are targeted by two independent pathways, which have evolved separately to effectively counterbalance the severity of such lesions. While one of these, homologous recombination (HR), is intrinsically accurate, using extensive regions of homology as a template for repair, the nonhomologous end joining (NHEJ) pathway often leads to deletion of a few nucleotides at the site of double strand breaks [3].

More simple by design but remarkably versatile is the nucleotide excision repair (NER) pathway, which is capable of recognizing a vast variety of different DNA damages, ranging from abasic sites and photoproducts to polycyclic aromatic hydrocarbons and other bulky adducts [5]. In prokaryotes only four proteins, UvrA, UvrB, UvrC and UvrD, effectively orchestrate recognition and lesion excision, which is then followed by a DNA Polymerase I and DNA Ligase mediated repair synthesis process [6]. Though the same mechanistic principles apply, the eukaryotic NER pathway is much more complex, involving more than 20 components (see I.1.3).

I.1.3 Eukaryotic Nucleotide Excision Repair (NER)

The nucleotide excision repair (NER) pathway in eukaryotes can be subdivided into two separate processes, which share a common mode of action but operate in different contexts. While global genome NER (GG-NER) is constitutively active and can occur anytime and anywhere in the genome, the transcription coupled NER pathway (TC-NER) is only initiated within the limits of actively transcribed genes [7,8]. However, once the NER machinery is set into motion, the course of events leading to excision and repair of a DNA lesion are generally equivalent in both pathways (Figure I.2).

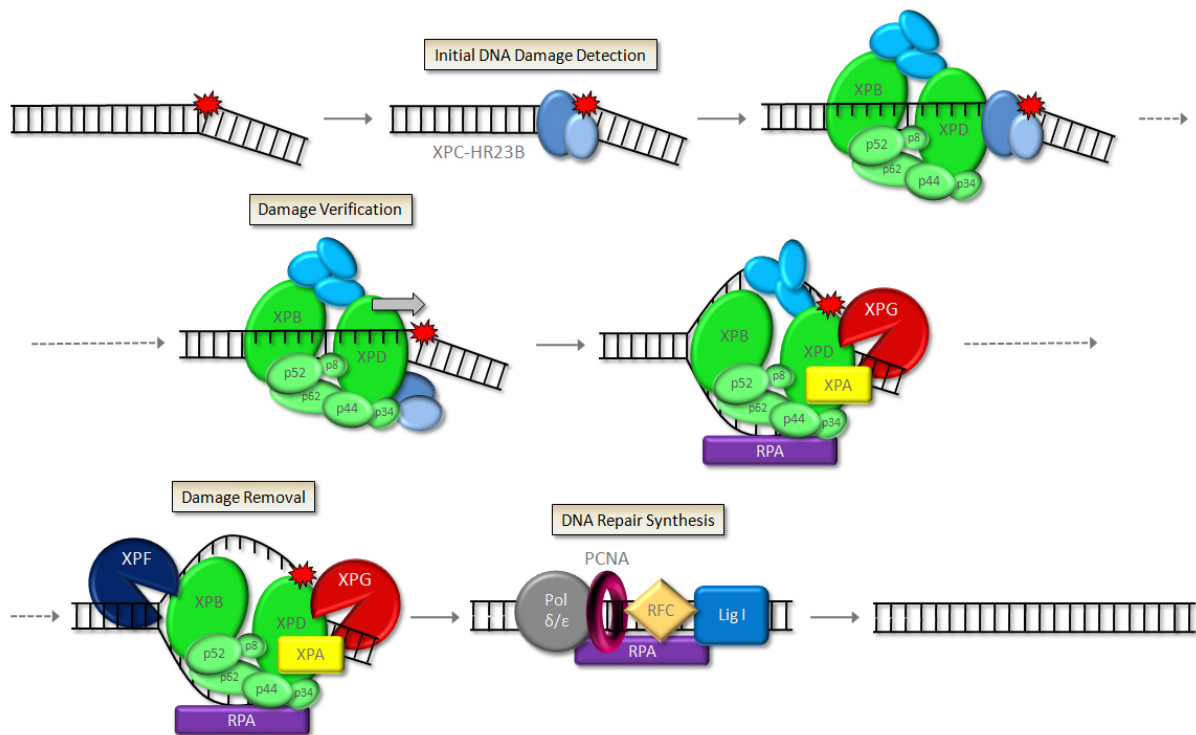


Figure I.2: Mechanism of Eukaryotic Nucleotide Excision Repair. Damage recognition by XPC-HR23B leads to recruitment of the TFIIH complex (green and light blue), which drives DNA strand opening via the action of the XPB and XPD helicases. The endonucleases XPG and ERCC1-XPF then catalyze two incision reactions 3' and 5' to the lesion, which are followed by lesion removal and DNA repair synthesis.

The core NER excision reaction was found to require a minimum set of 15 protein components and could be reconstituted *in vitro* using purified XPC-HR23B, XPA, RPA, XPG, ERCC1-XPF and the TFIIH Complex [9]. Upon initial recognition of a DNA damage by XPC-HR23B, a critical step which is assisted by UV-DDB2 for some lesions [10,11], the essential TFIIH complex is recruited, leading to damage demarcation and DNA double strand unwinding around the lesion [12,13]. This process is mainly driven by the XPB and XPD helicases in TFIIH, while XPD is also thought to play a major role in the course of damage verification [14,15]. With the helicase and ATPase activities of both XPD and XPB tightly regulated by their associated partners p44 and p52, respectively [14], the overall enzymatic activity of TFIIH becomes very complex on a structural level and is, though extensively studied in the past, still far from understood.

Concomitantly to DNA strand opening XPA and RPA as well as the endonucleases XPG and ERCC1-XPF are recruited to catalyze incisions on the 3' and 5' side relative to the lesion, respectively, leading to the actual damage removal. The resulting gap is ultimately filled and sealed by the DNA repair synthesis machinery, consisting of DNA Polymerase δ/ϵ , DNA Ligase I and associated factors [16].

I.2 Basic Concepts of DNA Transcription

Though TFIIH was shown to assume a pivotal role during nucleotide excision repair [17,18], the complex was first identified as one of the basal transcription factors promoting RNA Pol II mediated transcription [19]. Hence the following paragraphs will provide a brief overview on the intricate and vital process of DNA to RNA transcription, thereby displaying the overall function of TFIIH in a broader context.

While the genetic blueprint of every living cell is stored in the form of DNA and codes for a vast set of factors, which in turn shape morphology and function, this information cannot be directly accessed. It first needs to be transformed into a readable and mobile molecule, rendering it readily available to the cellular machinery. This is accomplished by the highly sophisticated and inherently accurate process of DNA transcription, which generates transportable messenger ribonucleic acid (mRNA) copies of each DNA based gene to be used as templates for protein production via the similarly complex mechanism of ribosomal translation. In addition, DNA regions lying within introns or outside of regular genes, are recompiled into non-coding RNA molecules serving a multitude of secondary and regulatory functions [20].

DNA to RNA transcription in eukaryotic cells is mediated by a total of three different RNA Polymerases, each of which accommodates for certain types of RNA products. While RNA Pol I is responsible for the synthesis of larger ribosomal RNA molecules, RNA Pol III also generates smaller multi-purpose transcripts, like transfer RNA (tRNA) and small nuclear RNA (snRNA) [21]. In turn, the majority of mRNA as well as small interfering (siRNA) and micro RNA (miRNA) is synthesized by the combined action of RNA Polymerase II and the various and multifaceted factors supporting its essential role [22].

Apart from a common mode of operation, the transcriptional process can be subdivided into several steps, each of which must meet precise requirements to prevent stalling of the associated RNA polymerase. Starting with the formation of a stable pre-initiation complex, which is followed by initiation and promoter clearance, the actual RNA synthesis is accomplished during the elongation phase and culminates in RNA product release upon termination. However, though all RNA Polymerases are capable of sound RNA synthesis, they are per se unable to initiate transcription or even recognize and attach to a promoter start site. For this they inherently depend on a plethora of assisting proteins and protein complexes, which are commonly referred to as general transcription factors (TF) and include representatives like TFIIA, -B, -D, -E, -F and TFIIH [23,24].

For each round of transcription an astonishing number of 60 proteins assemble in a 3 Mega-Dalton complex at every RNA Polymerase II promoter region, approximately half of which are required for initiation and promoter escape [23]. Although high resolution structural information is still very limited, the 3-dimensional arrangement of a 32 subunit pre-initiation complex (PIC), including TFIIH, has recently been reported based on a combined approach of cryogenic electron microscopy, cross-linking and mass spectrometry results [24]. This study showed the DNA to be sandwiched by RNA Polymerase II on one side and most of the transcription factors on the other, with TFIIH contacting the suspended DNA roughly 30 base pairs downstream of the promoter region via its XPB (Ssl2) helicase. Interestingly, the RNA Polymerase within this PIC was found not to be in contact with the template DNA at all, clearly emphasizing the strong dependence on its associated transcription factors like TFIIH among others [24].

I.3 The General Transcription Factor II H (TFIIH)

I.3.1 Molecular Composition of the TFIIH Complex

The TFIIH complex is a multi-subunit protein assembly involved in at least three major cellular processes: regulation of transcription, cell cycle control and nucleotide excision repair. It consists of 10 different proteins forming two distinct sub-complexes, each assuming a specialized function (Table I.1, Figure I.3). The 7-subunit TFIIH core comprises the two helicases XPD and XPB, as well as p62, p52, p44, p34 and p8. In addition, MAT1, CycH and cdk7 form the cyclin activating kinase complex (CAK), which is attached to the core via a direct bridging interaction of XPD and presumably XPB with MAT1 [25,26].

Recently, it has been suggested that additional steady-state subunits, like XPG and Tfb6 [27,28], are present in TFIIH and ancillary interacting partners were identified [29], emphasizing the complexity of TFIIH and the high level of dynamics to participate in several cellular contexts. Mutations in XPD, XPB and p8, affecting these processes, are known to be causative for xeroderma pigmentosum (XP), Cockayne syndrome (CS) and trichothiodystrophy (TTD), severe diseases associated with TFIIH function [30–32] (see I.5).

Table I.1: Composition and Function of the TFIIH Complex in Human and Yeast

	Human	Yeast	Main Function	Reference	Binding Partners	Reference
Core Subunits	XPD	Rad3	5' to 3' Helicase	[33–35]	p44, MAT1, p8	[25,26,36–38]
	XPB	Ssl2	3' to 5' Helicase	[7,8]	p52, MAT1, cdk7	[26,37,39,40]
	p62	Tfb1			XPD, XPB, p52, p44	[39,41]
	p52	Tfb2	Regulates XPB Activity	[39,40]	XPB, p8	[38,40]
	p44	Ssl1	Regulates XPD Activity	[36,40]	XPD, p34	[40,42]
	p34	Tfb4			p44	[42]
	p8	Tfb5	Regulates XPB Activity	[38]	XPD, p52	[38]
CAK	Mat1	Tfb3	Inhibits XPD Activity	[43]	XPD, XPB, CycH, cdk7	[25,26,37,44]
	CycH	Ccl1			MAT1, cdk7	[26,37,44]
	cdk7	Kin28	Kinase, targets RNA Pol II CTD	[45]	MAT1, CycH	[26,37,44]

The core TFIIH assembly is highly conserved among eukaryotes, with only p8, p52 and p62 lacking in a few rare species [46]. Of all subunits XPD and XPB exhibit the highest sequence conservation, while p62 is most divergent [46]. Within TFIIH the helicase and ATPase activities of XPD and XPB are tightly regulated by their associated partners p44 and p52, respectively [36,40], both of which also bind to other subunits (Table I.1). The interaction of p52 with p8, the smallest of the core subunits, adds to the XPB stimulation and is essential for TFIIH stability and NER activity [47,48]. Moreover, p44 interacts with the p34 subunit, the latter of which leads to a reduced DNA repair capacity when it is depleted from the complex [49].

Over the past two decades the intricate network of regulation within TFIIH has been studied extensively, especially regarding the XPD/p44 and XPB/p52/p8 subunits (see I.4). However, only very little is known with respect to the other core subunits. Apart from its role in stimulation of the XPD helicase [36], an E3 ubiquitin ligase activity has been proposed for the p44 yeast homologue Ssl1 [50]. No enzymatic function is known for the p34 subunit of TFIIH, although recently it has been implicated in the process of mRNA-splicing, based on phylogenetic considerations [46].

I.3.2 Structural Knowledge on TFIIH Subunits

Next to a genetic and biochemical approach, detailed knowledge can be gained from structural studies involving single or assembled protein components. However, though considerable efforts have been made to structurally characterize the TFIIH complex using techniques like cryogenic electron microscopy (cryo-EM), nuclear magnetic resonance spectroscopy (NMR) and x-ray crystallography, all high resolution data obtained so far is limited to only few of its subunits or isolated domains thereof, respectively (Figure I.3).

Despite all experimental set-backs several cryo-EM studies on human and yeast TFIIH revealed a ring like assembly of its subunits around an inner cavity, with the cyclin activating kinase sub-complex (CAK) bridging the central XPB and XPD helicases [51–54]. The latter two are located on separate arms of the U-shaped core and are thought to bracket the engulfed DNA, contacting it from opposite sides [24,54].

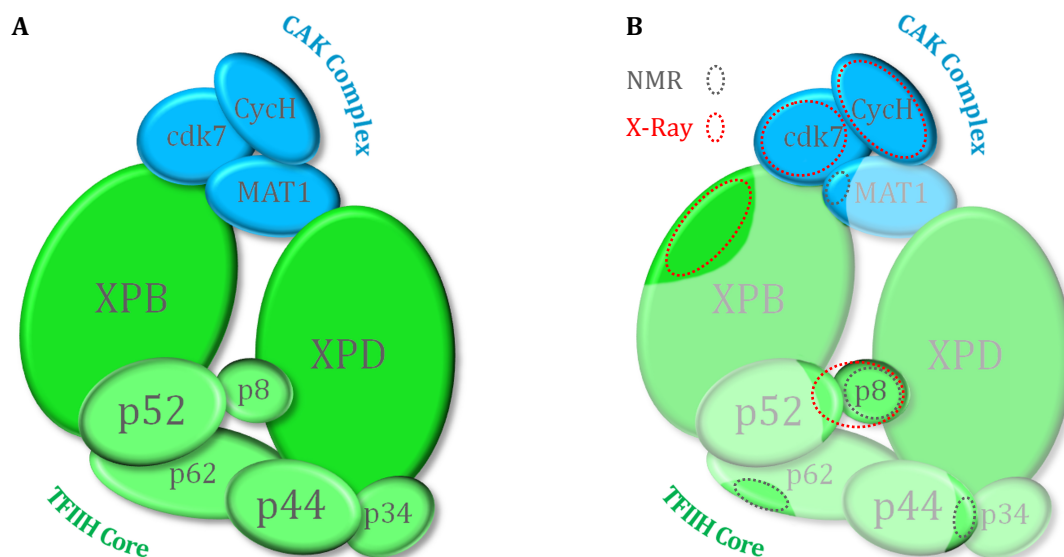


Figure I.3: Composition of the TFIIH Complex and Structural Knowledge on its Subunits. The TFIIH complex consists of 10 different proteins forming 2 sub-complexes (A, green and blue). The 7-subunit TFIIH core (green) comprises the central helicases XPB and XPD as well as their associated factors p62, p52, p44, p34 and p8, while the attached cyclin activating kinase complex (CAK) is formed by MAT1, CycH and cdk7 (blue). Eukaryotic TFIIH subunits for which structural information is available, either via NMR (grey) or x-ray crystallography (red), are highlighted in (B). See main text for further details.

In addition to the holo-TFIIH cryo-EM structures, full-length high resolution data are available for two of the CAK subunits, cdk7 and Cyclin H [55,56], while structural information on the third CAK component, MAT1, is limited to the N-terminal C3HC4 RING domain of the protein [57] (Figure I.3 B). With respect to the TFIIH core subunit architecture, crystal structures of XPD and XPB were determined using homologues from several archaeal species [58–61]. However, though both helicases are present in many archaeal genomes, the latter usually lack all the other TFIIH subunits, suggesting that XPD and XPB function differently and only as part of a non-TFIIH environment within these organisms [62,63]. Currently there is no structural information on a eukaryotic XPD helicase, whereas the C-terminal half of human XPB has recently been determined [64].

The only full-length TFIIH core structure is that of p8, which accounts for the smallest of all known TFIIH subunits [65]. Furthermore, a crystal structure of the yeast homologue of p8 (Tfb5) bound to the C-terminal domain of p52 (Tfb2) represents the only structure of a minimal TFIIH complex solved so far [47,48]. While additional NMR structures of the p62 PH domain and the C-terminal C4C4 RING finger domain of p44 are available [66–68], there is no structural information on the p34 subunit of TFIIH (Figure I.3 B).

I.4 The TFIIH Complex in DNA Repair and Transcription

Shortly after its discovery the TFIIH complex has been shown to be indispensable for both transcription and nucleotide excision repair, suggesting a functional link between these two fundamental processes [18,69]. The following chapters aim to summarize the key findings made throughout the following years with respect to the versatile role of TFIIH within these pathways.

I.4.1 TFIIH in Nucleotide Excision Repair

During the global genome NER pathway a DNA damage is initially detected by the XPC protein, which, once bound to DNA, receives further stabilization via additional binding partners like centrin 2 or HR23B [70,71]. Upon recognition of a DNA lesion the TFIIH complex is recruited through a direct interaction of XPC with both the p62 and XPB subunits in TFIIH [72], subsequently leading to DNA double strand unwinding around the lesion and excision of the damaged DNA by the action of the ERCC1-XPF and XPG endonucleases.

Although NER per se appears to be straightforward, the concomitant presence of two helicases within TFIIH raises the question of how these activities combine to orchestrate the NER process. Both XPD and XPB are SF2 type helicases, yet exhibit opposite polarities [40,73]. Hence early models proposed that each helicase contributes to local unwinding and damage verification, the latter of which could be induced by stalling on their respective translocating strand [74,75]. In this regard it was generally accepted that only the synergistic effect of both helicase activities would lead to proper excision. However, while the ATPase activity of XPB was found to be required to anchor TFIIH to the lesion, its helicase activity turned out to be dispensable for NER *in vitro* [40,76]. On the other hand, the helicase activity of XPD is essential for correct DNA duplex unwinding around the lesion [36,45,77]. Hence, though both XPB and XPD are active helicases, their distinct roles within the NER pathway differ significantly [40].

During the initial stages of NER the TFIIH core is linked to the CAK sub-complex via a direct interaction of MAT1 with the arch domain in XPD [43,78]. However, the presence of CAK negatively regulates the helicase activity of XPD [9,43]. While both sub-complexes are essential for optimal RNA Pol II mediated transcription, only the release of CAK from TFIIH promotes the excision of DNA lesions during the NER pathway [79,80].

Further complication arises from the fact that the activities of XPB and XPD are tightly regulated by their associated partners p52 and p44, respectively, both of which also interact with other TFIIH subunits [36,40]. While p52 contacts the N-terminus of XPB and stimulates its ATPase activity, the N-terminal domain of p44 interacts with the C-terminus of XPD, resulting in an increase in its helicase activity [36,39,40,77]. The tenth and smallest of all TFIIH subunits, p8, adds to the XPB stimulation via binding to the C-terminal end of p52, which shares a common fold with p8 [38,47,48]. Moreover, though dispensable for transcription *in vitro*, p8 was found to confer significant stability to the TFIIH complex, as the cellular concentration of TFIIH decreases considerably when this subunit is mutated [38,81,82]. In turn, the regulatory effect of p44 on the helicase activity of XPD might be assisted by p34, which forms a tight complex with p44 and an overall reduced DNA repair capacity was observed when p34 was depleted from the TFIIH complex [42,49]. However, due to the fact that the majority of biochemical studies focused on other TFIIH subunits, the exact role of p34 remains elusive. Since it is the only TFIIH subunit for which no structural information is available, a detailed structural characterization of p34, possibly in complex with p44 and XPD, would significantly expand our knowledge and shed more light on the molecular basis of overall TFIIH enzymatic activity.

I.4.2 The Role of TFIIH in Transcription and Cell Cycle Control

Apart from its major role during the NER pathway, the TFIIH complex also assumes a central role in the regulation of transcription, especially during the initiation phase, promoter escape and the early elongation steps [27,83]. While most of the studies performed so far focused on the RNA Pol II mediated transcription of genes encoding proteins, TFIIH in addition facilitates the transcription of ribosomal RNA by Pol I [84,85] and probably also the synthesis of tRNA and other small RNAs, which are transcribed by Pol III [86,87].

Upon sequential recruitment of general transcription factors at the promoter start site and formation of a stable pre-initiation complex, the TFIIH core is thought to be responsible for promoter opening, while the CAK sub-complex promotes initiation and elongation by phosphorylation of the C-terminal domain (CTD) of the largest subunit of RNA Pol II [79]. In contrast to the observed activities of XPB and XPD during the NER process, here the ATP-dependent helicase activity of XPB is required for DNA double strand unwinding and promoter escape, while XPD presumably assumes a more passive role [73,88–91].

The XPB helicase functions in close concert with the cdk7 subunit of CAK, which phosphorylates Ser5 and Ser7 within the RNA Pol II CTD and thus controls initiation of transcription [92]. This bivalent kinase activity towards the CTD is in turn modulated by various factors, including MAT1 and Cyclin H, the two binding partners of cdk7 within the TFIIH CAK [37,44]. Hence, while dispensable and even impeditive for NER, the CAK sub-complex is an absolute requirement for the regulation of transcription at the stage of initiation. Moreover, the CAK subunits have been implicated in the process of cell cycle control during the transition from G2 to M phase [93–95], while the activating phosphorylation of several kinases via cdk7 was shown to be required for normal cell cycle progression [96]. In addition, the XPD helicase, which is closely linked to the CAK sub-complex via MAT1, also interacts with MMS19 in the MMXD complex, which in turn is required for chromosome segregation [97]. Taken together, these findings suggest that the elaborate and multifaceted roles of the TFIIH complex, along with their intricate network of regulation, are not confined to DNA repair or transcription alone, but reach out to other cellular processes, which are somehow linked to TFIIH or some of its versatile subunits.

I.4.3 Crosstalk between DNA Repair and Transcription

Though the molecular mechanisms of NER and transcription are intricate enough on their own, more and more evidence suggests that there exists a regular crosstalk between both pathways, further adding to the complexity of these fundamental processes.

In this regard it has been known for some time, that transcriptional activators such as Gal4-VP16 and RAR can stimulate nucleotide excision repair [98], while DNA repair proteins, including the damage sensor XPC, seem to function as co-activators for transcription factors like OCT4 in an embryonic stem cell context [99]. Moreover, several NER factors were shown to be recruited to active promoters and are thought to regulate transcription in the absence of exogenous genotoxins by influencing chromatin remodeling [100,101].

In addition, the TFIIH core subunit p44 has been suggested to mediate the transcriptional response to UV or methyl methanesulfonate (MMS) induced DNA damage via an intrinsic E3 ubiquitin ligase activity, the latter of which has so far only been observed for its homologue Ssl1 in yeast [50]. This close entanglement of NER and transcription factors and their far-reaching impact on other cellular processes ultimately echoes in the complex phenotypes of several diseases that arise from mutations in some of the TFIIH subunits (I.5).

I.5 Diseases Caused by Mutations in TFIIH Subunits

Mutations in the TFIIH complex, which interfere with the enzymatic function or multiple engagements of its versatile subunits, are causative to several major diseases, including the autosomal recessive disorders xeroderma pigmentosum (XP), trichothiodystrophy (TTD) and the combined symptoms of xeroderma pigmentosum and Cockayne syndrome (XP/CS) [30–32]. While many mutations causing these afflictions can be located on a single gene, such as XPD, mutations in the XPB and p8 subunits of TFIIH are also known to result in XP, XP/CS and TTD like phenotypes [102]. In addition, mutations in XPB, p62 and p52, lead to photosensitivity and XP/TTD like phenotypes in *Drosophila* [103–105].

Over the last two decades multiple clinical trials yielded extensive insight into the complex phenotypic patterns associated with the XP, XP/CS and TTD disorders. In parallel, several biochemical and structural studies aimed to elucidate the underlying molecular mechanisms on the level of the proteins involved. Among the observations made it was found that mutations in the XPB and XPD helicases of TFIIH are generally associated with an elevated risk of developing skin, breast and lung cancer [106,107]. In the case of xeroderma pigmentosum most patients are extremely UV sensitive and exhibit a 1,000 fold increase in the risk for skin cancer [108,109]. In addition, XP patients are often prone to progressive neurological degeneration as well as immature sexual development and dwarfism [110].

In contrast, both CS and TTD are not associated with cancer, but instead are accompanied by a spectrum of neurodevelopmental abnormalities. The usual symptoms of CS comprise prenatal growth failure, severe neurological dysfunction, mental retardation and retinal and skeletal abnormalities [111]. Usually CS patients succumb to these ailments in their early teens and only few live to their twenties. Individuals with TTD suffer from various neurological defects, growth retardation, sterility, ichthyosis as well as sulfur-deficient brittle hair and nails, which are caused by reduced levels of cysteine-rich matrix proteins [110,112].

Initially XP, XP/CS and TTD were defined as DNA repair syndromes, affecting the molecular function of TFIIH and associated factors during the nucleotide excision repair pathway. Along that line the elevated skin cancer predisposition in XP patients can be readily explained by the inability of the NER machinery to repair UV-induced DNA lesions. However, some of the observed clinical features are difficult to comprehend on the basis of a malfunctioning NER process alone. Instead, several characteristic features of CS and TTD are likely caused by an insufficient expression of various effector proteins, in turn resulting from drastically reduced cellular levels of TFIIH and hence inefficient transcription [113,114]. In this regard it is well established that most XPD mutations associated with XP, XP/CS or TTD weaken the interaction of XPD and p44 and hence lead to reduced XPD helicase activity during NER [58,77]. However, TTD inducing mutations in XPD, which abolish p44 binding, were also shown to impact transcription [36,73,77]. Moreover, there are only a few patients with mutations in XPB, which would reflect the essential role of this subunit during transcription [73].

In addition to XPD, XPB and p8 also mutations in XPG, the endonuclease required to perform the 3' incision during NER, were shown to induce CS typical phenotypes [115]. Interestingly, though not constitutively associated with the TFIIH complex, XPG seems to stabilize the assembly and prevent dissociation of XPD and the CAK sub-complex, a role especially committed towards transcription [115]. Along with its essential role for both NER and transcription XPG has been suggested as the eleventh subunit of TFIIH [27]. In contrast, mutations in the second NER endonuclease, XPF (FANCD1), have recently been shown to cause Fanconi anemia (FA), a genetic disease associated with a unique interstrand crosslink DNA repair pathway [116].

I.6 The Fungus *C. thermophilum*

In light of the multifaceted functions of TFIID in DNA repair, transcription and beyond, as well as its causative role in the development of several diseases, the 10-subunit complex has been a major research focus since its discovery more than 20 years ago [19]. Still, despite all progress, detailed structural studies on holo-TFIID or single subunits have often been hampered in the past by the difficulty to obtain abundant and pure protein samples, suitable for NMR or crystallographic analysis. In addition, most of the TFIID core and all CAK subunits are absent in bacteria or archaea, which often serve as distant but handy model organisms. On these premises a different strategy would likely be required to pave the way for further achievements with respect to the structural characterization of this essential protein complex.

In the course of x-ray crystallographic studies a thermophilic species is frequently chosen as a model organism, since it generally harbors proteins of excellent stability. This characteristic alleviates some of the common difficulties experienced during expression and purification as well as protein handling and storage of their mesophilic homologues, and in turn may facilitate crystallization and structural characterization. Among eukaryotes, however, only very few species have the capacity to grow at temperatures above 50 °C, greatly limiting the range of thermophilic candidates that could potentially serve as a more suitable model organism [117]. Recently, the structure of another multi-protein assembly, that previously eluded structural characterization, could successfully be described utilizing the genome of *Chaetomium thermophilum* (ct), a eukaryotic and thermophilic fungus, which can thrive at temperatures of up to 60 °C [118].

C. thermophilum is a filamentous ascomycete and belongs to the family of *Chaetomiaceae*. It is commonly found in soil, dung and rotting plants where it contributes significantly to the breakdown of cellulose. The latter includes a decomposition phase at temperatures of 55 °C and higher, which only allows thermophilic organisms to grow [118]. Despite the fact that it was already discovered more than 60 years ago [119], it has only now been recognized as a novel model organism, providing the scientific community with a closely related proteome of good accessibility and high stability. Hence, in an attempt to shed more light on the structure and function of the TFIID complex, the present work focuses on protein homologues from this interesting fungus, which contains all the CAK and TFIID core subunits.

II. Material

II.1 Equipment

II.1.1 Instruments

Alphabetical listing of all major instruments and machines used throughout the study, including their place of origin and manufacturer, respectively.

Instrument	Type / Model		Origin / Manufacturer
Autoclaves	Systec V-150		Systec
Balances	440-35N		Kern & Sohn GmbH
	PB8000-S		Mettler Toledo
	XS105 Dual Range®		Mettler Toledo
	XS6002S Delta Range®		Mettler Toledo
Block-Heater	Rotilabo® Block-Heater H250		Roth
Cell Disrupter	M-110P		Microfluidics
Centrifuges	Avanti J-26 XP	(0 - 1000 ml)	Beckman Coulter
	Centrifuge 5417R	(0 - 2 ml)	Eppendorf
	Centrifuge 5418	(0 - 2 ml)	Eppendorf
	Centrifuge 5810R	(0 - 50 ml)	Eppendorf
Column Body (Gravity Flow)	Econo-Column 2.5 x 20 cm		Bio-Rad
Cuvettes	110-QS (for CD)		Hellma
Dish Washer	Professional G7883 CD		Miele
Electrophoresis Chambers	Mini-PROTEAN® 3 Cell		Bio-Rad
	Mini-PROTEAN® Tetra Cell		Bio-Rad
	Mini-Sub® GT		Bio-Rad
	Sub-Cell® Model 192 Cell		Bio-Rad
FPLC Column Material (AC)	Ni-MAC™	(CV = 1.0 ml)	Novagen
	Protino® Ni-TED	(Beads)	Macherey-Nagel
FPLC Column Material (IEX)	MonoQ™ 5/50 GL	(CV = 1.0 ml)	GE Healthcare
	Resource Q	(CV = 1.0 ml)	GE Healthcare
FPLC Column Material (SEC)	Superdex 70 10/300 GL	(CV ≈ 24 ml)	GE Healthcare
	Superdex 200 10/300 GL	(CV ≈ 24 ml)	GE Healthcare
	Superdex 200 16/60 PG	(CV ≈ 124 ml)	GE Healthcare
FPLC-Systems	ÄKTExpress™ Twin System		GE Healthcare
	ÄKTA™ Purifier with Frac-950		GE Healthcare

Gel-Dryer	GelAir Drying System	Bio-Rad
Gel-Scanners	Gel/Chemi-Doc XR(S) HP Scanjet 3800	Bio-Rad Hewlett-Packard
Ice Maschine	Eismaschine 94774	Ziegra Eismaschinen
Illumination Table	P 265.1	Roth
Incubator	Type B 15	Heraeus
ITC Calorimeters	MicroCal™ VP-ITC MicroCal™ ITC ₂₀₀	GE Healthcare GE Healthcare
Laminar Flow Cabinet	BDK SK-1200	BDK
Magnetic Stirrers	MR 3002 MR Hei-Mix L	Heidolph Heidolph
MALS Setup	DAWN® 8+ HELEOS II® OptiLab® T-rEX™ HC ÄKTA™ Purifier	Wyatt Technology Wyatt Technology GE Healthcare
Microscopes	SteREO Discovery.V12 + AxioCam MRC Stemi 2000-C	Zeiss Zeiss
pH-Electrode	BlueLine 14pH	Schott Instruments
Pipettes	Pipet-Lite LTS (variable sizes)	Rainin
Pipettors	Easypet Pipetus®	Eppendorf Hirschmann Laborgeräte
Power Supplies	PowerPac® Basic™ PowerPac® HC™	Bio-Rad Bio-Rad
Robots	HoneyBee 961/963 Lissy® Autom. Liquid Handling Platform	Zinsser Analytic Zinsser Analytic
Rotors	JLA 8.1000 JLA 16.250 JA 25.50	Beckman Coulter Beckman Coulter Beckman Coulter
Shakers	ISF-1-W ISF-1-X KS-15 A Control mit TH-15 Haube Lab-Therm Thermomix comfort	Kühner Kühner Edmund Bühler GmbH Kühner Eppendorf
Spectrophotometers	BioPhotometer NanoDrop ND-1000	Eppendorf peQLab
Spectropolarimeter	J-810 (for CD)	Jasco

Swiveling Tables	Duomax 1030 TL-10 Type 3013	Heidolph Edmund Bühler GmbH GFL
Thermocycler	MasterCycler epGradient S Stratagene® MX3005P™	Eppendorf Agilent Technologies
Vacuum Pump	Laboport	KNF Neuberger
Vortexer	Vortex Genie 2	Scientific Industries
X-Ray Cryo-System	X-Stream™ 2000	Rigaku
Detector	R-AXIS HTC	Rigaku
Generator	MicroMax-007HF	Rigaku
Optics	VariMax HF	Osmic Inc.

II.1.2 Consumables

Alphabetical listing of all relevant consumable materials used throughout the study, including their place of origin and manufacturer, respectively.

Consumable	Type / Model	Origin / Manufacturer
Cell Culture Plates	6-well	Greiner Bio-One
Cover Slips	22 mm (siliconized)	Jena Biosciences
Cuvettes	Rotilabo®-Einmalküvetten, 1.6 ml	Roth
Cryo-Loops	CryoLoop™	Hampton Research
Crystal Plates	24-well (Hanging Drop) 96-well (Sitting Drop)	Crystalgen Greiner Bio-One
Dialysis Cassettes	Slide-A-Lyzer	Pierce Biotechnology
PCR Plates	96-well PCR Plate	Greiner Bio-One
Petri Dishes	Standard (sterilized)	Roth
Pipette Tips	Pipet-Lite LTS (sterilized)	Rainin
Reaction Tubes	0.2 / 0.5 / 1.5 / 2.0 ml (sterilized) 15 / 50 ml (sterilized)	Sarstedt Greiner Bio-One
Syringes	1.0 / 2.5 / 5.0 / 10.0 ml (sterilized)	B. Braun Melsungen AG
Syringe Filters	Rotilabo® Spritzenfilter, 0.45 / 0.22 µm	Roth
Ultrafiltration Tubes	Vivaspin 20 (MWCO 3,000 - 30,000 Da) Vivaspin 6 (MWCO 3,000 - 30,000 Da)	Sartorius Sartorius
Weighing Dishes	Rotilabo®-Wägeschalen, hexagonal	Roth

II.2 Chemicals and Biochemical Substances

II.2.1 Chemicals

Alphabetical listing of all relevant solid and liquid chemical substances used throughout the study, including their place of origin and manufacturer, respectively.

Chemical	Short	Origin / Manufacturer
1,4-Dithiothreitol	DTT	Roth
2-(Cyclohexylamino)ethanesulfonic acid	CHES	Sigma-Aldrich
2-(N-Morpholino)ethanesulfonic acid	MES	Fluka
2-Methyl-2,4-Pentandiol	MPD	Fluka
3-(Cyclohexylamino)-1-propanesulfonic acid	CAPS	Fluka
5-Bromo-4-chloro-3-indolyl β -D-galactopyranoside	X-Gal	Sigma-Aldrich
Acetic acid	AcOH	Roth
Acrylamide-Solution, Rotiphorese® Gel 30		Roth
Active charcoal		Roth
Agar-Agar		Roth
Agarose		Roth
Agarose HEE0 Ultra Quality		Roth
Agarose MEE0 Ultra Quality		Roth
Ammonium peroxodisulfat	APS	Roth
Ampicillin sodium salt	Amp	Roth
Boric acid		Fluka
Bovine Serum Albumin	BSA	New England Biolabs
Brilliant Blue R-250		Roth
Bromphenol Blue		Sigma-Aldrich
Calcium chloride	CaCl ₂	Fluka
Chloramphenicol	Cam	Roth
Ethanol	EtOH	Roth
Ethidium bromide	EtBr	Roth
Ethylen-diamine-tetraacetic acid	EDTA	Roth
Fetal calf serum	FCS	Sigma-Aldrich
Glycerol		Roth
Hydrochloric acid	HCl	Hartenstein
Imidazole		Roth
Iodoacetamide		Roth
Isopropanol		Roth
Isopropyl- β -D-thiogalactopyranoside	IPTG	Roth
Kanamycin sulfate	Kan	Roth
L-(+)-Arabinose		Roth
LB-medium (Lennox)		Roth

Magnesium chloride	MgCl ₂	Roth
Methanol	MeOH	Roth
Midori Green Advance		Biozym
Nickel sulfate		Roth
Paraffin oil		Fluka
Phosphate buffered saline	PBS	Sigma-Aldrich
Polyethylene glycol 20,000	PEG 20,000	Sigma-Aldrich
Polyethylene glycol 10,000	PEG 10,000	Fluka
Polyethylene glycol 8,000	PEG 8,000	Fluka
Polyethylene glycol 6,000	PEG 6,000	Fluka
Polyethylene glycol 4,000	PEG 4,000	Sigma-Aldrich
Polyethylene glycol 1,000	PEG 1,000	Sigma-Aldrich
Polyethylene glycol 550 monomethyl ether	PEG 550 MME	Sigma-Aldrich
Polyethylene glycol 400	PEG 400	Sigma-Aldrich
Potassium bromide	KBr	Sigma-Aldrich
Potassium chloride	KCl	Roth
Potassium hydroxide	KOH	Roth
Potassium iodide	KI	Sigma-Aldrich
Potassium phosphate dibasic	K ₂ HPO ₄	Roth
Potassium phosphate monobasic	KH ₂ PO ₄	Roth
Sodium chloride	NaCl	Roth
Sodium dodecyl sulfate	SDS	Fluka
Sodium hydroxide	NaOH	Roth
SYPRO-Orange		Invitrogen
Tetramethyl-ethylene-diamine	TEMED	Roth
Tris(2-carboxyethyl)-phosphine	TCEP	Roth
Tris(hydroxymethyl)-aminomethane	TRIS	Roth
Tween® 20		Fluka
Urea		Roth

II.2.2 Ready-To-Use Kits

The following commercially available kits were used in the course of this study, especially in the context of DNA isolation, cloning and crystallization.

Ready-To-Use Kit	Purpose	Origin / Manufacturer
CryoProtX™ MD 1-61	Cryo Optimization	Molecular Dimensions
In-Fusion™ Dry-Down PCR Cloning Kit	Cloning	Clontech
JBScreen Heavy	Heavy Metal Soaking	Jena Biosciences
NucleoSpin® Gel and PCR Clean-Up	Gel-Extraction / PCR Clean-Up	Macherey-Nagel
NucleoSpin® Plasmid	Plasmid Isolation	Macherey-Nagel
Silver Bullets HT HR2-096	Buffer Optimization	Hampton Research

II.2.3 Crystallization Screens

Alphabetical listing of all relevant crystallization and additive screens employed, including their origin and manufacturer, respectively. Most of the screens were prepared in-house using appropriate stock solutions and the automated liquid handling platform Lissy®.

Commercial Screen	Origin / Manufacturer	Prepared by
Additive Screen HR2-183	Hampton Research	-
Crystal Screen I + II	Hampton Research	Lissy® Robot
Index Screen	Hampton Research	Lissy® Robot
Nextal PEG	Quiagen	Lissy® Robot
Nextal pHClear	Quiagen	Lissy® Robot
Nucleix Suite	Quiagen	Lissy® Robot
Protein Complex Suite	Quiagen	Lissy® Robot
Silver Bullets HT HR2-096	Hampton Research	-
Topaz OptiMix I	Fluidigm	-
Topaz OptiMix III	Fluidigm	Lissy® Robot
Topaz OptiMix PEG	Fluidigm	Lissy® Robot
Wizard Screen I + II	Emerald BioSystems	Lissy® Robot

II.2.4 Primers and DNA Substrates

For SLIC cloning of all p34 and p44 variants from *C. thermophilum* the following primers were used. Regions of the respective genes are shown in green, while primer overhangs complementary to the vector target site are indicated in blue. Nucleotides colored in black denote introduced start (ATG) or stop (TGA) codons.

Primer	Sequence (5' → 3')
p34ct-pBADM-F	AATCTTTATTTTCAGGGCGCCATGAGCGCCAGGACGCC
p34ct[1-387]-pBADM-R	CCAAAACAGCCAAGCTCTCGAGTCATCCGTAATCCCCAACGCCAG
p34ct[1-368]-pBADM-R	CCAAAACAGCCAAGCTCTCGAGTCACACCTCACAAAAATGCTGAGACACAC
p34ct[1-277]-pBADM-R	CCAAAACAGCCAAGCTCTCGAGTCAACCGTGCCAAAGCCGAACATG
p44ct-pBADM-F	CTTTATTTTCAGGGCGCCATGGCCGACTCTGATGGCGAGTATG
p44ct[55-534]-pBADM-F	CTTTATTTTCAGGGCGCCATGGCCGCATGGGAGGACATCCAGC
p44ct[75-534]-pBADM-F	CTTTATTTTCAGGGCGCCATGATCACAATTGAGGCCCTCATCGAGGC
p44ct[55-285]-pBADM-R	CCAAAACAGCCAAGCTCTCGAGTCAAATCGTAGCCGCCAGGAACAGC
p44ct[101-285]-pBADM-F	CTTTATTTTCAGGGCGCCATGATCCGTCTGTTGCTGGTGC
p44ct[386-534]-pETM-F	CTTTATTTTCAGGGCGCCATGCTAAGCACCCACTTGGCGCGCTC
p44ct[368-534]-pETM-R	GTGGTGGTGGTGGTGGTCTCGAGCTAGTCGAGCACCATAGCATCCCCATCC
p44ct[386-534]-pETM-NoNH-F	CTTTAAGAAGGAGATATAACCATGCTAAGCACCCACTTGGCGCGCTC
p44ct[368-534]-pETM-CH6-R	CAGTGGTGGTGGTGGTGGTTCGAGCACCATAGCATCCCCATCC
p44ct[368-534]-pETM-NoTag-R	CAGTGGTGGTGGTGGTGGTCTAGTCGAGCACCATAGCATCCCCATCC

To generate site-directed mutants of p34ct and p44ct for interaction studies, the primers below were used. Regions in green correspond to the native gene sequence, while nucleotides in red indicate mutated residues.

Primer	Sequence (5' → 3')
p34ct[W80A]-F	CACATACCAACCGCGCCGTCGCTCTGTACCCTCAGCCCCCGAG
p34ct[W80A]-R	CTCGGGGGGCTGAGGGTACAGAGCGACGGCGCGGTTGGTATGTG
p34ct[K155E]-F	TCGCCCTCGCTCACATTAACGAGACTGCCCTCTCCCTGAC
p34ct[K155E]-R	GTCAGGGAGAGGGCAGTCTCGTTAATGTGAGCGAGGGCGA
p34ct[N222A]-F	GCGCAGTACATCCCCACGATGCGAGCCGTCTTCGCCGCAGCTCAC
p34ct[N222A]-R	GTGAGCTGCGGCGAAGACGGCTGCCATCGTGGGGATGTACTGCGC
p44ct[L405E]-F	CAGGTTGGGTGCTTTGCTTGCAGGGCGCCATTTCCCTGCCTCC
p44ct[L405E]-R	GGAGGCAGGGGAAATGGCGCCTCGCAAGCAAAGCACCCAACCTG
p44ct[F490E]-F	CTTCTGCATCGACTGCGATGTAGAAGCGCATGAGGTCATTACAAATTG
p44ct[F490E]-R	CAATTGTGAATGACCTCATGCGCTTCTACATCGCAGTCGATGCAGAAG
p44ct[Q502E]-F	CACAATTGCCCCGGATGCGAGGCTGACATGCGACCAAAG
p44ct[Q502E]-R	CTTTGGTTCGCATGTCAGCCTCGCATCCGGGGCAATTGTG

All primers were obtained from Biomers.net GmbH in lyophilized form. They were diluted to a concentration of 20 µM using 0.1x TE buffer (II.4.3.1) and stored at - 20 °C.

For DNA binding assays two complementary 50-mer oligonucleotides (Biomers.net GmbH) were used, either alone as single stranded DNA (NDT) or annealed as double stranded DNA substrates (NDT/NDB).

NDT 5' GACTACGTACTGTTACGGCTCCATCTCTACCGCAATCAGGCCAGATCTGC 3'
 NDB 5' GCAGATCTGGCCTGATTGCGGTAGAGATGGAGCCGTAACAGTACGTAGTC 3'

II.2.5 Enzymes

Alphabetical listing of all relevant enzymes used in the course of the study, including their enzyme classification number (EC), their stock concentration and manufacturer.

Enzyme	EC-Number	Concentration	Origin / Manufacturer
DNAse I	3.1.21.1	200 - 300 U/µl	Invitrogen
<i>DpnI</i>	3.1.21.4	20 U/µl	New England Biolabs
<i>HindIII</i>	3.1.21.4	20 U/µl	New England Biolabs
<i>KpnI</i>	3.1.21.4	10 U/µl	New England Biolabs
<i>NcoI</i> -HF™	3.1.21.4	20 U/µl	New England Biolabs
Phusion™ DNA Polymerase	2.7.7.7	2 U/µl	Thermo Scientific
<i>SmaI</i>	3.1.21.4	20 U/µl	New England Biolabs
T4 DNA Polymerase	2.7.7.7	3 U/µl	New England Biolabs
<i>Taq</i> DNA Polymerase	2.7.7.7	5 U/µl	New England Biolabs
TEV Protease	3.4.22.44	var.	Purified in-house
<i>XhoI</i>	3.1.21.4	20 U/µl	New England Biolabs

II.3 Cell Culture

II.3.1 Bacterial and Insect Cell Strains

The following bacterial strains were used in the context of DNA plasmid amplification and protein expression. Competent cells were prepared in-house based on a protocol originally established by Inoue *et al.* [120].

Strain	Genotype	Origin / Manufacturer
DH5 α	<i>E. coli</i> K12, recA1, endA1, hsdR17 (Γ_{K12}^- , m_{K12}^+), supE44, thi-1, gyrA96, relA1, lacF'[proA+B ⁺ , lacI ^q Z Δ M15::Tn10(Tet ^R)]	Clontech Laboratories
DH10Bac	F ⁻ , mcrA, Δ (mrr-hsdRMS-mcrBC), ϕ 80lacZ Δ M15, Δ lacX74, recA1, endA1, araD139, Δ (ara,leu)7697, galU, galK, rpsL, nupG, pMON14272, pMON7124	Invitrogen
BL21-CodonPlus [®] (DE3)-RIL	<i>E. coli</i> B, F ⁻ , ompT, hsdS (r_B^- , m_B^-), dcm ⁺ , Tet ^R , gal λ (DE3), endA, Hte [argU ileY leuW Cam ^R]	Stratagene

The following insect cell strains were used for protein expression of human TFIIH subunits in insect cells.

Strain	Organism	Origin / Manufacturer
<i>Sf21</i>	<i>Spodoptera frugiperda</i>	Dr. Joop van den Heuvel
<i>Sf9</i>	<i>Spodoptera frugiperda</i>	Prof. Thomas Müller

The *Sf21* cells were obtained from a frozen stock maintained at the Protein Sample Production Facility (PSPF) of the Helmholtz Center for Infection Research (HZI) in Braunschweig, while all *Sf9* cells were kindly provided by Prof. Thomas Müller of the University of Würzburg.

II.3.2 Vectors for Protein Expression

The following contains a list of all vectors used for cloning and protein expression, including their properties and origin. Detailed maps for the vectors can be found in Appendix X.2.

Vector	Properties	Origin / Manufacturer
pBADM-11	Vector for expression in <i>E. coli</i> using the araBAD promoter, N-terminal His ₆ -Tag, TEV cleavage site, Amp ^R	EMBL
pETM-11	Vector for expression in <i>E. coli</i> using the T7 Lac promoter, N-terminal His ₆ -Tag, MAD insertion, TEV cleavage site, Kan ^R	EMBL
pFastBac HTa	Vector for expression in <i>Sf21</i> and <i>Sf9</i> insect cells, PH promoter, N-terminal His ₆ -Tag, TEV cleavage site, Amp ^R , Gen ^R	Invitrogen
pFastBac Dual	Vector for expression in <i>Sf21</i> and <i>Sf9</i> insect cells, two multiple cloning sites controlled by PH and p10 promoters, Amp ^R , Gen ^R	Invitrogen

II.3.3 Bacterial Culture Media and Media Supplements

For protein expression and cloning purposes *E. coli* BL21-CodonPlus®(DE3)-RIL, DH5 α and DH10Bac strains were grown using the following liquid or solid culture media, respectively.

LB-Medium	LB-Agar	LB-Agar for Blue/White Screening
10.0 g/l Tryptone	10.0 g/l Tryptone	10.0 g/l Tryptone
5.0 g/l Yeast extract	5.0 g/l Yeast extract	5.0 g/l Yeast extract
5.0 g/l NaCl	5.0 g/l NaCl	5.0 g/l NaCl
	5.0 g/l Agar-Agar	5.0 g/l Agar-Agar
		40 μ g/ml IPTG
		50 μ g/ml X-Gal

To suppress growth of any other than the desired bacteria, both media were sterilized by autoclaving and supplemented with the following antibiotics before use.

Antibiotic	Stock Concentration	Final Concentration
Ampicillin	100 mg/ml (in ddH ₂ O)	100 μ g/ml
Chloramphenicol	34 mg/ml (in EtOH)	34 μ g/ml
Gentamicin	50 mg/ml (in ddH ₂ O)	7 μ g/ml
Kanamycin	50 mg/ml (in ddH ₂ O)	50 μ g/ml
Tetracycline	5 mg/ml (in ddH ₂ O)	10 μ g/ml

II.3.4 Insect Cell Culture Media and Media Supplements

For baculovirus amplification and protein expression in *Sf21* and *Sf9* insect cells, the cells were grown in the following liquid culture media.

Culture Medium	Properties	Origin / Manufacturer
Ex-Cell® 420:	Serum-Free, with L-Glutamine	Sigma Aldrich
Insect-XPRESS™:	Serum-Free, with L-Glutamine	Lonza

Occasionally, the media were supplemented with 5 % fetal calf serum (FCS), especially for transfection of insect cells and Plaque Assays, to improve the rate of transfection and virus infection, respectively.

To suppress growth of any other than the desired cells, the media were prepared in a sterile manner by using a laminar flow cabinet. In addition, for all long term incubation steps in adherent cell cultures, the media were supplemented with the following antibiotics.

Antibiotic	Stock Concentration	Final Concentration
Gentamicin	10 mg/ml (in ddH ₂ O)	10 μ g/ml
Fungizone	250 μ g/ml (in ddH ₂ O)	250 ng/ml

II.4 Buffers and Solutions

II.4.1 Buffers and Solutions for Purification of Proteins

For all purification buffers the pH value was set at 4 °C using either HCl or NaOH. The concentration of buffer, salt and imidazole used varied during the optimization of the purification strategy.

Ni-MAC A:	50 mM Tris pH 8.0 500 mM KCl 20 mM Imidazole 10 % Glycerol 0.5 mM TCEP	Ni-MAC B:	50 mM Tris pH 8.0 500 mM KCl 500 mM Imidazole 10 % Glycerol 0.5 mM TCEP
Ni-TED A:	20 - 50 mM Tris pH 8.0 300 - 500 mM KCl 0.5 mM TCEP	Ni-TED B:	20 - 50 mM Tris pH 8.0 300 - 500 mM KCl 150 - 250 mM Imidazole 0.5 - 1 mM TCEP
IEX A:	20 mM CHES pH 10.0 1 mM TCEP	IEX B:	20 mM CHES pH 9.5 1 M KCl 1 mM TCEP
SEC P1:	20 mM Tris pH 8.0 150 - 200 mM KCl 1 mM TCEP	SEC P2:	20 mM CHES pH 9.5 150 mM KCl 1 mM TCEP

All purified proteins were usually flash frozen and stored in either SEC P1 or SEC P2, respectively, which correspond to the elution buffers during size exclusion chromatography.

II.4.2 Buffers and Solutions for Biochemical Assays

The following buffers were used as reaction and/or running buffer for most biochemical assays, including SEC-MALS experiments (SEC M), analytical size exclusion chromatography runs (SEC A) as well as CD (SEC D) and ITC measurements (SEC T). Their pH was usually set at 4 °C, except for SEC M1, SEC M2 and SEC D, which were set at room temperature.

SEC M1:	20 mM Tris pH 8.0 150 mM KCl 1 mM TCEP	SEC M2:	20 mM CHES pH 9.5 150 mM KCl 1 mM TCEP
SEC A1:	20 mM Tris pH 8.0 150 mM KCl 1 mM TCEP	SEC A2:	20 mM Tris pH 8.0 350 mM KCl 1 mM TCEP
SEC D:	20 mM KH ₂ /K ₂ H-PO ₄ pH 8.1	SEC T:	20 mM CHES pH 9.5 150 mM KCl 1 mM TCEP

II.4.3 Buffers and Solutions for Gel Electrophoresis and Sample Staining

II.4.3.1 DNA Gel Electrophoresis and Staining

Agarose gels for DNA separation were run horizontally at RT. For staining EtBr or Midori Green was used.

5x DNA-SB:	0.01 % (w/v) Bromphenol Blue	Agarose-Gel:	1.0 % Agarose
	50 % (v/v) Glycerol		3.0 µl EtBr / Midori Green
	Add 1x TE		Add 1x TAE to 50.0 ml
1x TE:	10 mM Tris pH 8.0	1x TAE:	40 mM Tris pH 8.0
	1 mM EDTA		20 mM Acetic acid
			2 mM EDTA

II.4.3.2 Protein Gel Electrophoresis via SDS-PAGE and Staining

All SDS-PAGE runs were performed vertically. The gels were usually composed of a 15 % separating gel with a 5 % stacking gel casted on top of the main separating component. For the human p34 p44 minimal complex 20 % separating gels were used, containing 3.3 ml Rotiphorese® 30, accordingly.

15 % PA-Gel:	2.5 ml Rotiphorese® 30	5 % PA-Gel:	330 µl Rotiphorese® 30
	1.3 ml 1.5 M Tris-HCl pH 8.8		250 µl 1.0 M Tris-HCl pH 6.8
	50 µl 10 % SDS		20 µl 10 % SDS
	50 µl 10 % APS		20 µl 10 % APS
	2 µl TEMED		2 µl TEMED
	Add ddH ₂ O to 5.0 ml		Add ddH ₂ O to 2.0 ml
5x SDS-SB:	250 mM Tris-HCl pH 6.8	1x SDS-RB:	192 mM Glycine (14.4 g)
	500 mM DTT		25 mM Tris (3.0 g)
	0.1 % (w/v) Bromphenol Blue		0.1 % SDS (1.0 g)
	10 % (w/v) SDS		Add ddH ₂ O to 1.0 L
	50 % (v/v) Glycerol		
Staining:	50 % (v/v) Methanol	Destaining:	10 % (v/v) Methanol
	10 % (v/v) Acetic acid		5 % (v/v) Acetic acid
	1 % (w/v) Brilliant Blue R-250		

II.4.3.3 Protein Gel Electrophoresis via Native Agarose Gels (NAGE) and Staining

Native agarose gels were run horizontally in the cold room. The pH of the running buffer (RB) was set at 4 °C.

2x NAGE-SB:	100 mM Tris pH 8.0	Agarose-Gel:	0.6 - 0.8 % HEEO Agarose
	0.05 % Bromphenol Blue		0.0 - 3.0 µl Midori Green
	20 % Glycerol		Add 1x TAE to 100 ml
1x NAGE-RB 1:	25 mM Tris pH 8.0 - 9.5	1x NAGE-RB 2:	25 mM CHES pH 9.5
	192 mM Glycine		192 mM Glycine

II.5 Software and Databases

II.5.1 Software on Local Hardware

Alphabetical listing of software that was part of the instrumentation used to carry out experiments.

Software	Version	Instrument	Origin / Developer
Astra 6	v.6.0.6	SEC-MALS Equipment	Wyatt Technologies
Crystal Clear	v.1.4.0 r9	X-Ray Device	Rigaku
MxPro	v.4.10	Stratagene® MX3005P™	Agilent Technologies
ND-1000	v.3.8.1	NanoDrop ND-1000	peQLab
Origin 7 SR4	v.7.0552	MicroCal™ ITC200	GE Healthcare
Origin 7 SR4	v.7.0552	MicroCal™ VP-ITC	GE Healthcare
Quantity One	v.4.2.1	Gel/Chemi-Doc XR(S)	Bio-Rad
SpectraManager II	v.1.53.01	Spectropolarimeter J-810	Jasco
Unicorn	v.5.11	FPLC-System ÄKTExpress™	GE Healthcare
Unicorn	v.5.01	FPLC-System ÄKTA™ Purifier	GE Healthcare
User Interface	v.2.00.00	HoneyBee 961/963	Zinsser Analytic
Zinsser Method Runner	v.7.1.6	Lissy®	Zinsser Analytic

For crystal structure determination, model building and refinement the following software was used.

Software	Version	Purpose	Reference
ADXV	v.1.9.8	Diffraction Image Viewer	[121]
CCP4 Suite	v.6.2	Multi Purpose	[122]
COOT	v.0.7 pre	Model Building	[123]
iMOSFLM	v.6.2	Data Processing	[124]
O	v.13	Model Building	[125]
Phaser	v.6.2	Molecular Replacement	[126]
Phenix	v.1.8.1069	Model Refinement	[127]
PyMol	v.1.3 / v.1.6	Model Visualization	[128]
SCALA	v.6.2	Data Processing	[129]
SHELX C/D/E	v.6.2	Phasing	[130]
XDS	12/11 - 09/12	Data Processing	[131]

II.5.2 Software on the World Wide Web

For sequence alignments or data analysis the following software on servers of the world wide web was used.

Software	Website	Purpose	Reference
BLAST	www.ncbi.nlm.nih.gov	Sequence Alignment and Search	[132]
DSSP	swift.cmbi.ru.nl	Secondary Structure Assignment	[133]
GENTle	www.magnusmanske.de	Sequence Visualization	[134]
JalView	www.jalview.org	Sequence Alignment Visualization	[135]
LAlign	www.expasy.org	Sequence Alignment	[136]
MUSCLE	www.ebi.ac.uk	Multitple Sequence Alignment	[137]
OligoAnalyzer 3.1	eu.idtdna.com	Oligonucleotide Analysis	[138]
Phyre ²	www.sbg.bio.ic.ac.uk	Secondary Structure Prediction	[139]
ProtParam	www.expasy.org	Protein Parameters Calculation	[140]
Translate	www.expasy.org	Translation of DNA into Amino Acids	

II.5.3 Databases

The databases below were employed for general research on various subjects, with some of them, like NCBI or ExPASy, containing an extensive collection of tools, of which the most important ones are listed in II.5.2.

Database	Website	Purpose	Reference
DALI Server	ekhidna.biocenter.helsinki.fi	Structure Homology Search	[141]
ExPASy	www.expasy.org	Collection of Tools	[142]
NCBI	www.ncbi.nlm.nih.gov	Multipurpose Database	
Protein Data Bank	www.rcsb.org	Protein Structures Database	[143,144]
SMART	smart.embl-heidelberg.de	Domain Annotation Database	[145,146]
UniProt	www.uniprot.org	Protein Knowledge Database	[147]

III. Methods

III.1 Molecular Biology

III.1.1 Polymerase Chain Reaction (PCR)

In preparation for cloning (III.1.5) or site-directed mutagenesis (III.1.7) the gene sequences of both p34ct and p44ct were amplified using the polymerase chain reaction (PCR) method. The technique allows for the amplification of any DNA segment defined by two DNA primers of opposite polarity. Upon heat denaturation of the DNA template and primer annealing DNA polymerases are employed to extend the two primer strands, incorporating the deoxyribonucleotide triphosphates (dNTPs) provided in the reaction mixture in a sequence dependent manner. The process can be automated by using thermostable enzymes that are able to survive repeated cycles of denaturation, primer annealing and extension [148,149].

For PCR amplification of p34ct and p44ct sequences the following standard reaction setup (Table III.1) and PCR program (Table III.2) was used. The initial cloning of full-length p34ct and p44ct was performed by Agnes Elias, based on a cDNA template pool of *C. thermophilum* genomic DNA, which was kindly provided by Ed Hurt from the University of Heidelberg. Any additional amplification, cloning and sequence modification regarding p34ct and p44ct is the result of this work (see detailed information in X.3.1 and X.3.2).

Table III.1: Setup for Standard PCR Reaction.

Component	Stock	Final	1x
Phusion™ HF Buffer	5x	1x	10.0 µl
dNTP Mix	10 mM	200 µM	1.0 µl
Forward-Primer	20 µM	0.4 µM	1.0 µl
Reverse-Primer	20 µM	0.4 µM	1.0 µl
DNA Template	100 ng/µl	50 ng	0.5 µl
Phusion™ DNA Polymerase	2 U/µl	1 U	0.5 µl
ddH ₂ O	-	-	36.0 µl
			Σ 50.0 µl

Table III.2: Program for Standard PCR Reaction.

Step	Temp	Time	Cycles
Initial Denaturation	98 °C	30 s	1
Denaturation	98 °C	30 s	} 30
Annealing	56 - 62 °C	30 s	
Extension	72 °C	90 s	
Final Extension	72 °C	10 min	1
Hold	4 °C	∞	

Upon PCR completion the mixture was treated with 0.5 µl of *DpnI* at 37 °C for 4 - 6 hours or over night, respectively, to digest the methylated template DNA [150]. Each PCR product was then analyzed by DNA gel electrophoresis (III.1.4) and subsequently purified using the NucleoSpin® Gel and PCR Clean-Up kit (II.2.2). The amplified DNA segments were stored at - 20 °C until used for cloning into expression vectors (III.1.5).

III.1.2 DNA Isolation

For isolation of plasmid DNA from *E. coli* DH5 α , the bacteria were transformed with the expression vector of interest and grown in 5 ml LB medium over night, as described in III.2.1. Plasmid purification was performed the next day according to the NucleoSpin[®] Plasmid kit (II.2.2) using 40 - 50 μ l of the provided buffer AE for elution of the DNA. The isolated plasmids were then stored in the elution buffer (5 mM Tris pH 8.5) until used for DNA double digestion (III.1.3) and cloning (III.1.5).

III.1.3 DNA Double Digestion

To allow for correct cloning of PCR amplified DNA segments into isolated plasmids, the vector of choice has to be primed using DNA double digestion. Here, the capability of restriction enzymes to target specific regions in the vector's multiple cloning site is used, leading to sequence dependent DNA incision and linearization [151]. The multiple cloning site of modern vectors is engineered in such a way that several different restriction enzymes can be used, each targeting a different site and hence allowing for highly flexible construct design.

Table III.3: Setup for Standard DNA Double Digestion.

Component	Stock	Final	1x
NEB Buffer 2	10x	1x	2.0 μ l
<i>Nco</i> I	20 U/ μ l	20 U	1.0 μ l
<i>Xho</i> I	20 U/ μ l	20 U	1.0 μ l
Vector	var.	2 - 4 μ g	var.
ddH ₂ O	-	-	var.
			Σ 20.0 μ l

In the present study DNA double digestion utilizing the type-II restriction enzymes *Nco*I and *Xho*I [152] was used to prime both the pBADM-11 and pETM-11 expression vectors for cloning. The 20 μ l reaction mixture (Table III.3) was incubated at 37 °C for 4 - 6 hours and the linearized plasmids subsequently purified using the NucleoSpin[®] Gel and PCR Clean-Up kit (II.2.2). Usually a sample of the reaction product was analyzed by DNA gel electrophoresis (III.1.4) to check for proper double digestion. The linearized vectors were then stored at - 20 °C until used for cloning (III.1.5).

III.1.4 DNA Gel Electrophoresis

To analyze DNA and reaction products from both PCR and double digestion experiments, the samples were separated using agarose gels. The principle is based on the net negative charge of DNA molecules, introduced by their phosphate backbone, allowing them to move towards the anode when a voltage is applied. Due to their differences in size and the drag within the gel matrix, linear DNA fragments are separated according to their size, the latter of which can be compared to a DNA standard applied to the same gel [153].

For DNA gel electrophoresis usually 5 - 10 μ l of samples were mixed with 1 - 2 μ l of 5x DNA-SB and loaded onto 1.0 % agarose gels. The gels were submerged in 1x TAE buffer and run horizontally for 30 - 45 minutes at 90 - 120 volts. As size standard a lane containing 3.0 μ l of GeneRuler™ 1 kb DNA Ladder at a concentration of 70 ng/ μ l (Thermo Scientific) was included, covering a range of 250 - 10,000 bp. To visualize the DNA using the Gel/Chemic-Doc XR(S) system and UV light, the gels were supplemented with 3.0 μ l of either ethidium bromide (Roth) or Midori Green Advance (Biozym) prior to casting (II.4.3.1).

III.1.5 DNA Cloning and Ligation

Generation of all p34ct and p44ct plasmid constructs was accomplished by cloning the PCR amplified genes or parts thereof (III.1.1) into the linearized expression vectors pBADM-11 and pETM-11 (III.1.3). In order to allow for an efficient procedure and simultaneously limiting the amount of enzymes and reaction steps required, cloning of constructs was performed according to the SLIC protocol [154]. Here, the initial PCR primers for gene amplification are designed in such a way as to introduce a 20 - 25 nucleotide extension at each end of the DNA duplex, which is complementary to the target location in the vector's multiple cloning site (II.2.4). In a first reaction step the 3' → 5' exonuclease activity of T4 DNA Polymerase is used to generate sticky overhangs at both ends of the amplified PCR fragments and linearized vectors, respectively, which are then combined in a second step to allow for directed annealing of both DNA segments. The actual ligation of both the insert and vector is then performed *in vivo* by *E. coli* DH5α itself, which is transformed directly with the product of the second reaction step.

Table III.4: Setup for SLIC Cloning (Stage 1)

Component	Stock	Final	1x
NEB Buffer 2	10x	1x	2.0 μl
T4 DNA Polymerase	3 U/μl	1.5 U	0.5 μl
Vector / Insert	var.	2 μg / 1 μg	var.
ddH ₂ O	-	-	var.
			Σ 20.0 μl

Table III.5: Setup for SLIC Cloning (Stage 2)

Component	Stock	Final	1x
T4 DNA Ligase Buffer	5x	1x	2.0 μl
Vector	100 ng/μl	200 ng	2.0 μl
Insert	50 ng/μl	var.	var.
ddH ₂ O	-	-	var.
			Σ 10.0 μl

For cloning of all p34ct and p44ct constructs both the PCR products and target vectors were first treated separately with 1.5 U of T4 DNA Polymerase for 30 minutes at RT (Table III.4) to generate sticky DNA ends. The reaction products were then combined and incubated for another 30 minutes at RT, to allow for proper self-annealing of the insert and vector segments (Table III.5), with the amount of insert calculated using equation 1 below. To complete the SLIC cloning procedure, 50 μl of competent *E. coli* DH5α were transformed with 1 - 2 μl (10 - 20 ng) of product from the second stage and grown over night at 37 °C (III.2.1).

$$ng (Insert) = \frac{ng (Vector) \cdot kb (Insert)}{kb (Vector)} \cdot \frac{2}{1} \quad \text{Equation 1}$$

For selection of positive colonies containing the correct plasmid, the LB agar medium was supplemented with antibiotics, corresponding to the resistance gene provided by the vector transformed. The background of colonies containing an empty plasmid could be further reduced by prolonging both the *DpnI* and double digestion reactions during insert and vector preparation, respectively (III.1.1 and III.1.3).

III.1.6 Colony-PCR

To check for positive colonies after SLIC cloning (III.1.5) and verify the presence of the correct insert, the Colony-PCR assay was used. This technique allows for a quick and efficient screening of multiple colonies by using their intrinsic genomic and plasmid DNA directly as template for a PCR reaction (III.1.1). Primers are designed as such as to allow for the amplification of the cloned gene insert, if present, while no PCR product of the expected size is obtained, if the vector is either empty or not present at all in the template colony used. The latter can be distinguished by choosing primers that attach to the vector's multiple cloning site (MCS) shortly before and after the insert location, yielding only very short DNA amplicons for negative colonies.

Table III.6: Setup for Standard Colony-PCR Reaction.

Component	Stock	Final	1x
ThermoPol® Buffer	10x	1x	2.0 µl
dNTP Mix	10 mM	200 µM	0.4 µl
Forward-Primer	20 µM	0.4 µM	0.4 µl
Reverse-Primer	20 µM	0.4 µM	0.4 µl
DNA Template	-	1 Colony	1 Colony
<i>Taq</i> DNA Polymerase	5 U/µl	1 U	0.2 µl
ddH ₂ O	-	-	16.6 µl
			Σ 20.0 µl

Table III.7: Program for Standard Colony-PCR Reaction.

Step	Temp	Time	Cycles
Initial Denaturation	95 °C	3 min	1
Denaturation	95 °C	30 s	} 35
Annealing	55 °C	30 s	
Extension	72 °C	120 s	
Final Extension	72 °C	10 min	1
Hold	4 °C	∞	

In the present study usually 14 of the colonies grown after SLIC cloning and transformation were picked for verification by Colony-PCR. Each of the colonies was first dipped into the PCR reaction mixture (Table III.6) and then streaked onto a PCR master plate containing LB agar and suitable antibiotics. In this way the template colonies could be preserved for later use.

The primers below were used to amplify the insert including short stretches upstream and downstream from its location within the MCS. For primer extension, the cheaper *Taq* DNA Polymerase was applied and the PCR program adjusted accordingly (Table III.7). The PCR products were analyzed by DNA gel electrophoresis as described in III.1.4 and positive clones were picked to isolate the respective plasmid constructs (III.1.2).

Primer	Sequence (5' → 3')	Target Vector	Target Site
pBAD-For	CCATAGCATTTTTATCCATAAG	pBADM-11	Upstream of MCS
pBAD-Rev	GATTTAATCTGTATCAGGCTG	pBADM-11	Downstream of MCS
T7-Prom	TAATACGACTCACTATAGGG	pETM-11	T7 Promoter
T7-Term	GCTAGTTATTGCTCAGCGG	pETM-11	T7 Terminator

III.1.7 Site-Directed Mutagenesis

For generation of the p34ct 1-277 and p44ct 368-534 mutated variants site-directed mutagenesis of already prepared wild type constructs was performed. The procedure includes a two-stage PCR reaction with primers designed specifically to introduce a 1 to 3 nucleotide exchange in the gene sequence (II.2.4), ultimately leading to incorporation of the desired amino acid at that position.

The PCR setup and program used correspond to the one described in III.1.1, with only a few modifications to allow for a more efficient reaction. To prevent excessive primer dimerization, the initial master mixture was split in two, each including only one of the two primers required. Upon a first stage of only 8 PCR cycles, the two half reactions were combined and an additional round of 25 PCR cycles was performed. The time for primer extension and polymerization was increased to 180 seconds, in order to allow for a complete amplification of the template plasmid, while the temperature for primer annealing was set to 58 °C.

To remove the wild type template from the reaction mixture and reduce the background of negative colonies, the PCR product was treated with *DpnI* at 37 °C over night. As circular plasmids were generated by the site-directed mutagenesis protocol used, 50 µl of competent *E. coli* DH5α could then directly be transformed with 1 - 5 µl of the raw reaction product. The cells were grown over night on LB agar plates supplemented with the respective antibiotics to select for positive transformants, which were checked for the presence of the insert via Colony-PCR (III.1.6). To verify the correct exchange of nucleotides and general integrity of the constructs, the plasmids were purified as described in III.1.2 and finally sent for DNA sequencing (III.1.8).

III.1.8 DNA Sequencing

DNA sequencing of plasmids was used to verify the integrity of the constructs created by SLIC cloning (III.1.5) and site-directed mutagenesis (III.1.7), respectively. The technique is based on the chain termination method introduced by Sanger *et al.* in 1977 [36], which allows to discriminate single nucleotides at the end of an extended primer by introducing fluorescently labeled terminal inhibitor ddNTPs, instead of the usual dNTPs provided. The position information of nucleotides within a sequence is then obtained by separation of the labeled DNA fragments according to their size.

In the present study DNA sequencing was not performed in-house. Instead the plasmids were sent to either SeqLab Sequencing Laboratories Göttingen GmbH or Eurofins MWG Operon GmbH. The results were usually obtained within a few days and visualized using the freeware tool GENTle [134].

III.1.9 DNA Hybridization

For DNA binding studies duplex DNA was prepared by hybridization of two complementary single strands using a simple heat-denaturation and self-annealing procedure. The reaction setup (Table III.8) was heated to 85 °C for 10 min followed by slowly cooling the mixture to RT. The DNA was then stored at - 20 °C until use.

Table III.8: Setup for DNA Hybridization.

Component	Stock	Final	1x
DNA Substrate 1	100 µM	10 µM	5.0 µl
DNA Substrate 2	100 µM	10 µM	5.0 µl
KCl	2 M	100 mM	2.5 µl
TE Buffer	0.1x	< 0.1x	37.5 µl
			Σ 50.0 µl

III.1.10 Bacmid DNA Generation

For the generation of bacmid DNA, which was used to transfect insect cells for virus production (III.2.4), the genes of human XPD, XPB, p52 and p44 were first amplified by PCR (III.1.1) and subsequently cloned into the pFastBac HTa and pFastBac Dual vectors. Prior to cloning, the vectors were linearized by double digestion (III.1.3) using the restriction enzymes *Hind*III and *Nco*I (pFastBac HTa) or *Kpn*I and *Xho*I (pFastBac Dual). Cloning was performed using the In-Fusion™ Dry-Down PCR Cloning Kit (Clontech), according to the protocol provided by the manufacturer.

After In-Fusion™ cloning and transformation of *E. coli* DH5 α the colonies grown were screened as to whether they contained the desired gene using Colony-PCR (III.1.6). Upon plasmid isolation from positive colonies, which yielded an amplicon of the expected size (III.1.2), the correct sequence of each construct was verified by DNA sequencing (III.1.8). To generate bacmid DNA, *E. coli* DH10Bac cells were then transformed with the obtained pFastBac HTa and pFastBac Dual constructs (III.2.1), resulting in the Tn7 mediated transposition of the gene cassette into a bacmid shuttle vector. Following transformation, the cells were grown on LB agar, including IPTG and X-Gal to allow for blue/white screening (II.3.3), at 37 °C for 24 hours. White and hence bacmid positive colonies were then picked and used for bacmid DNA isolation (III.1.11).

III.1.11 Bacmid DNA Isolation

For bacmid DNA isolation, positive (white) colonies from the blue/white screening were seeded each in 5 ml LB medium, including the appropriate antibiotics, and grown over night at 37 °C and 200 rpm. The cells were then lysed using buffers A1 - A3 of the NucleoSpin® Plasmid kit (II.2.2), according to the protocol provided by the manufacturer. After clarification of the lysate at 15,000 g and 4 °C for 10 minutes, the supernatant was transferred into a new tube containing 800 μ l isopropanol to precipitate the DNA. Following an incubation at 4 °C for 10 minutes, the DNA was pelleted by centrifugation at 15,000 g and 4 °C for 15 minutes. The obtained DNA was then washed by adding 500 μ l of a 70 % ethanol solution and pelleted again by centrifugation at 15,000 g and 4 °C for 5 minutes. After removal of the supernatant, the pellet was air dried at room temperature for 5 - 10 minutes and the DNA subsequently dissolved in 30 - 50 μ l of 1x TE buffer at pH 8.0.

The purified Bacmid DNA was stored at - 20 °C and later used to transfect *Sf*21 or *Sf*9 insect cells for virus production and protein expression. As the concentration of purified bacmid DNA could not be determined spectrophotometrically due to a high RNA background in the samples, the molarity was estimated by comparison to a standard of known concentration via DNA agarose gel electrophoresis (III.1.4).

III.2 Microbiology

III.2.1 Transformation of Bacteria

The transformation of bacteria with specifically designed plasmids is a common technique to utilize these workhorses for a set of purposes like vector/insert ligation, plasmid amplification and protein expression. While many bacteria are naturally capable of taking up foreign DNA, *E. coli* K12 strains like DH5 α or BL21 first need to be made competent prior to transformation [155,156]. In our lab this was accomplished utilizing a modified version of the protocol originally established by Inoue *et al.* [120]. Competent bacteria were then flash frozen in liquid nitrogen and stored at - 80 °C until needed.

For standard transformation of *E. coli* DH5 α or BL21-CodonPlus[®](DE3)-RIL 50 μ l of cells were thawed and incubated with 1.0 μ l of plasmid DNA at a concentration of 10 ng/ μ l for 15 - 30 minutes at 4 °C. Following a heat-shock treatment at 42 °C for 60 - 90 seconds the mixture was again cooled on ice for 2 minutes and supplemented with 600 μ l of LB medium at RT. To allow for a first round of duplication the cells were then incubated at 37 °C and 600 rpm for 45 - 60 minutes. After centrifugation at 2,000 g for 2 minutes the cells were resuspended in only 150 μ l of the supernatant and finally plated onto LB agar medium and grown at 37 °C over night or at RT over the weekend, respectively. The LB agar medium was usually supplemented with appropriate antibiotics (II.3.3) to allow for the selection of specific transformants.

III.2.2 Growing of Bacteria for Protein Expression

For protein expression the *E. coli* strain BL21-CodonPlus[®](DE3)-RIL was transformed with the desired p34 or p44 vector construct as described in III.2.1. Single and freshly prepared transformants were then picked and transferred to 5 ml of LB medium containing strain and plasmid specific antibiotics (II.3.3). After an initial incubation at 37 °C and 200 rpm for 6 - 9 hours, 1 - 2 ml of this pre-culture were used to inoculate a larger 200 ml LB culture, which was grown at 37 °C and 200 rpm over night.

The next day 4x 2L of LB medium, again supplemented with antibiotics, were seeded with 40 - 50 ml of one over night culture to initiate a large scale production. Starting at an OD₆₀₀ \approx 0.1 the cells were first grown for 1.5 hours at 37 °C and 200 rpm until early log phase (OD₆₀₀ \approx 0.2 - 0.4). The shakers were then set to 15 °C to allow for more gentle conditions regarding protein expression and stability. During the cooling period the cell growth was continued at 200 rpm until reaching an OD₆₀₀ \approx 0.8 - 1.0, at which protein expression was induced by adding a final concentration of either 0.05 % (v/v) L-(+)-Arabinose (pBADM-11) or 0.5 mM IPTG (pETM-11), respectively. Following induction, the cells were allowed to grow at 200 rpm and 15 °C over night corresponding to 20 - 22 hours post induction and were then harvested by centrifugation at 8,000 g and 4 °C for 15 minutes using 1 liter tubes and the JLA 8.1000 rotor. Finally all cell pellets were stored at - 20 °C until they were used for cell lysis and purification (III.3.1).

III.2.3 Growing and Maintenance of Insect Cell Cultures

Continuous liquid cultures of *Sf21* and *Sf9* insect cells were routinely grown in a volume of 30 ml using either Ex-Cell 420 (Sigma-Aldrich) or Insect-XPRESS (Lonza) as a serum-free culture medium. The cells were grown at 27 °C in 125 ml plastic flasks (Corning) on an orbital shaker platform at a speed of 110 rpm. To keep the cells vital and within their logarithmic growth phase (log-phase), the cultures were split every Monday, Wednesday and Friday, usually seeding a sub-culture at a density of approximately 0.5×10^6 cells/ml.

Before re-seeding or starting any experiments, the cell density was determined by counting a portion of the cells using a typical hemacytometer (Neubauer) and performing a back calculation based on the volume used for filling the counting chamber. In addition, the ratio of live versus dead cells was estimated via trypan blue staining, usually by mixing the cell sample in a 1:1 or 1:3 ratio with a solution containing 0.2 % trypan blue and 0.9 % NaCl prior to counting the cells.

III.2.4 Transfection of Insect Cells

For the transfection of *Sf21* and *Sf9* insect cells with purified bacmid DNA constructs, the cells were diluted in Ex-Cell 420 medium to a density of 0.5×10^6 cells/ml and 2.0 ml were seeded in each well of a 6-well cell culture plate. After an incubation at 27 °C for 30 minutes, which allowed the cells to adhere to the surface of the plate, the medium was drawn off and replaced by a fresh portion of 2.0 ml Ex-Cell 420 in each well.

During another round of incubation at 27 °C for 15 minutes, the transfection mixture was prepared in two separate sterile tubes. First 10 µl of bacmid DNA (~ 200 - 500 ng) were mixed with 40 µl Ex-Cell 420 medium, while 10 µl of the transfection reagent (SuperFect, Quiagen) were added to 140 µl Ex-Cell 420 medium. The two preparations were subsequently combined and incubated at RT for 10 - 15 minutes until the mixture was added to the cells after medium withdrawal. The cells were then incubated with the transfection mixture at 27 °C for 4 hours, allowing the bacmid DNA to enter the cells. Following transfection, the mixture was drawn off and replaced by a fresh portion of 2.0 ml Ex-Cell 420 medium.

After a continuous growth period at 27 °C for 2 - 3 days, the transfection supernatant was collected and cleared from cellular debris by centrifugation at 3,200 g and RT for 10 minutes. The supernatant containing the first virus generation (V_0) was finally stored at 4 °C in the dark until used for virus amplification (III.2.6). To increase the viral titer of the first virus generation, the cells were incubated for 2 more days in 2.0 ml fresh Ex-Cell 420 and the supernatant was ultimately combined with the first harvest.

III.2.5 Plaque Assay

To determine the viral titer of each virus generation, the Plaque Assay was used. First a culture of *Sf21* or *Sf9* insect cells was diluted to 0.625×10^6 cells/ml and 2.0 ml were seeded in each well of a 6-well culture plate. After an incubation at 27 °C for 30 minutes, allowing the cells to adhere to the surface, the medium was drawn off and replaced by a fresh portion of 2.0 ml Ex-Cell 420 medium in each well.

During a second round of incubation at 27 °C for 15 minutes, a serial dilution of the virus stock was prepared in Ex-Cell 420 medium, using a dilution factor of 10 for each step and usually covering a range of 10^{-2} to 10^{-7} . The medium was then drawn off and replaced by 500 µl of the appropriate virus dilution in a dropwise fashion. Following another incubation at 27 °C for 1 - 2 hours, during which the trays were carefully swayed every 15 - 30 minutes, the virus dilution was drawn off. The cells were subsequently covered with an agarose

overlay, which was prepared at a final concentration of 1 % in Ex-Cell 420 medium including 5 % FCS. For the overlay low melting point agarose was used, to avoid damaging the cells by too high temperatures.

After solidification of the agarose overlay at RT for 20 - 30 minutes, the cells were incubated upside down at 27 °C in an incubator with humidification control for 6 days. The viral plaques were finally visualized by staining the live cells with a solution of 0.5 mg/ml neutral red in PBS. For this, 300 µl neutral red solution were added to each well and incubated for at least 4 hours or over night, respectively. After withdrawal of the neutral red solution, the plates were incubated at RT in the dark over night, to allow the plaques formed to become more visible to the naked eye. The next day the plaques were counted and the viral titer calculated based on the number of plaque forming units (pfu) per ml of virus dilution.

III.2.6 Virus Amplification

For virus amplification and generation of viral stocks having higher titers, *Sf21* and *Sf9* insect cells were grown at 27 °C and 110 rpm on an orbital shaker platform until reaching the log-phase (1.0×10^6 cells/ml) and diluted to 0.5×10^6 cells/ml one day prior to infection. The next day a virus volume corresponding to a multiplicity of infection (MOI) of MOI = 0.1 - 0.5 was added at a cell density of 1.0×10^6 cells/ml (equation 2) and the cells were continued to grow at 27 °C and 110 rpm for 5 days.

$$\text{Virus Volume} = \frac{\text{Culture Volume} \cdot \text{Cell Density}}{\text{Virus Titer}} \cdot \text{MOI} \quad \text{Equation 2}$$

The cell count and viability was monitored each day and samples were taken at several points in time to check for proper infection and protein expression. After 5 days, or when the diameter of the cells reached their maximum and the vitality dropped below 80 %, the supernatant containing the released virus was harvested by centrifugation at 3,200 g and 4 °C for 15 minutes. The viral stocks obtained were stored at 4 °C in the dark and their titers were determined via the Plaque Assay (III.2.5).

III.2.7 Growing of Insect Cells for Protein Expression

For protein expression, 30 ml cultures of *Sf21* and *Sf9* insect cells were grown in Ex-Cell 420 (Sigma-Aldrich) or Insect-XPRESS (Lonza) medium to log-phase at 27 °C and 110 rpm. At a density of $1.0 - 1.5 \times 10^6$ cells/ml, the cells were infected with an appropriate volume of virus stock solution, corresponding to an MOI value of MOI = 1.0. The cultures were subsequently continued at 27 °C and 110 rpm and samples were taken at several points in time to check for proper infection and protein expression.

After 3 - 4 days the cells were harvested by soft centrifugation at 180 g for 3 minutes. The pellets were then washed two times in 30 - 50 ml of PBS or the buffer used for purification. After a final round of centrifugation at 1,000 g for 3 minutes, the cells were stored at - 20 °C or directly used for cell lysis. The latter was usually achieved by sonication in a total volume of 30 ml Ni-MAC A (II.4.1), including 20 mM PIPES at pH 7.5 instead of the Tris buffer, for 3 times 1 minute each at an amplitude of 90 % and a step cycle setting of 0.8.

III.3 Protein Biochemistry

III.3.1 Cell Lysis

Following protein expression and cell harvest (III.2.2), the pellets were thawed and resuspended in 2x excess of Ni-MAC A or Ni-TED A, respectively (II.4.1). The suspension was then supplemented by 2.0 μ l DNase I to degrade genomic DNA upon cell lysis, which was performed using the Microfluidics M-110P mechanical cell disrupter for 2 cycles at a pressure setting of 1.5 kbar. The lysate was cleared from debris by centrifugation in the JLA 16.250 rotor at 38,400 g and 4 °C for 1.5 - 2 hours.

The cleared supernatant was then applied to a nickel-coated resin for affinity chromatography (III.3.2.1). Usually a sample of the pellet was also kept for later use, to compare the amount of soluble versus insoluble protein fraction via SDS-PAGE (III.3.3).

III.3.2 Purification of Proteins and Chromatography Techniques

For purification of both p34 and p44 usually a two stage strategy was pursued. While nickel-metal affinity chromatography (AC, III.3.2.1) was primarily used to enrich the His₆-tagged protein variants and remove most contaminants, size exclusion chromatography was performed as a final polishing step (SEC, III.3.2.3). From time to time an additional ion exchange procedure was included between these two chromatography steps to further increase the sample purity (IEX, III.3.2.2). However, due to low expression levels, especially in the case of p34ct, this step was usually omitted in favor of higher protein yields.

III.3.2.1 Affinity Chromatography (AC)

Affinity chromatography (AC) is generally based on the lock-and-key principle, allowing only specifically tagged proteins to bind to a resin, while others not presenting this “key” do not adhere to the same material. This leads to a significant enrichment of the tagged protein versus untagged contaminants and hence is often used as a primary method for protein purification.

In the present study AC via Ni-MAC™ and Protino® Ni-TED resins was used, both relying on the presence of nickel metal ions to “catch” the Hexa-Histidine-Tag (His₆-Tag) of each target protein. The former is based on a Fractogel® bead matrix that carries covalently linked, long and branched polymers. These flexible tentacles coordinate several Ni²⁺ ions proverbially “fishing” for His₆-tagged proteins, generally allowing for high yields. In contrast, the Ni-TED resin is designed to bind less protein in favor of increased purity. Here, the tris-carboxymethyl ethylene diamine chelator (TED) occupies five of the six coordination sites of the Ni²⁺ ions, leaving only one for the binding of proteins. This reduces metal leaching during purification and significantly reduces unspecific adherence of contaminants. In both cases bound proteins can be eluted using imidazole, which mimics the histidine side chain and has a much higher affinity for the Ni²⁺ ions presented.

After cell lysis and centrifugation (III.3.1) the cleared supernatant was applied either to a Ni-MAC™ column (CV = 1 ml) or to 2 - 4 g of Ni-TED beads (CV = 4 - 8 ml). The former was coupled to the ÄKTExpress™ system, while the latter were prepared manually in an Econo-Column (Bio-Rad). Both resins were equilibrated with 10 CV of Buffer A prior to loading of the samples. After loading and washing with 20 CV of Buffer A, the bound proteins were eluted using 20 CV (Ni-MAC™) or 5 - 10 CV (Ni-TED) Buffer B. Usually at least 10 fractions of 2 - 5 ml each were collected and analyzed via SDS-PAGE (III.3.3) to evaluate the procedure.

III.3.2.2 Ion Exchange Chromatography (IEX)

Samples eluted from an affinity chromatography column are often not sufficiently pure for the study at hand. Here, ion exchange chromatography (IEX) can be used to further increase purity by separation of proteins according to their net surface charge in solution. The principle employs column materials that are either negatively (cation exchange) or positively (anion exchange) charged, allowing to capture molecules of opposite overall polarity. The strength of the interaction is dependent on the difference of potential and hence both the protein's pI and the pH of the buffer system used. This allows for separation of samples by gradient elution, progressively increasing the concentration of salt in the buffer, whose ions outcompete the charge of the proteins for binding and displace them from the oppositely charged matrix.

To improve purity of some p34 and p44 samples the anion exchange columns Resource™ Q and Mono™ Q (CV = 1 ml) were used. Proteins eluted from the affinity chromatography step (III.3.2.1) were diluted with IEX A to reach a final salt concentration of 20 - 50 mM, allowing them to bind to the resin. After equilibration of the columns with IEX A and IEX B, using a low-high-low procedure of alternating low (IEX A) and high salt concentration (IEX B), the samples were loaded with IEX A and the flow-through was collected to check for potentially unbound protein. The flow-rate was set to 0.8 ml/min and the elution was finally performed using a linear gradient of up to 100 % IEX B. At least 10 fractions were collected and subsequently analyzed via SDS-PAGE (III.3.3) to verify the success of the separation.

III.3.2.3 Size Exclusion Chromatography (SEC)

Size exclusion chromatography (SEC), also called gel-filtration (GF), was generally used as a final polishing step after an initial AC (III.3.2.1) and/or IEX (III.3.2.2) procedure, respectively. The method takes advantage of the different hydrodynamic volumes of molecules in solution, allowing the user to separate these according to their approximate size. Column materials for SEC are usually based on a Sephadex® matrix or similar composites of cross-linked agarose and dextran. They are prepared as tiny porous beads, which, upon applying a sample at constant flow rate, retard molecules small enough to enter their interior. Larger proteins are unable to enter and pass the material in the inter-particle volume, ultimately travelling faster and eluting earlier from the column. Due to the low sample and large buffer volumes used, the SEC principle often also serves to simultaneously exchange the buffer composition of the sample applied.

In the present study the HiLoad™ Superdex 200 16/60 prep grade SEC column (CV ≈ 124 ml) was used for protein purification. Samples obtained from AC or IEX were first concentrated by ultra-filtration (III.3.4) to volumes of 2 - 5 ml and then applied to the SEC material, which was previously equilibrated using 1.5 CV of either SEC P1 or SEC P2 (II.4.1). Separation was performed at flow-rates of 1.0 ml/min using a total of 1.3 CV of SEC P1 or SEC P2, of which 0.3 CV were prior to and 1.0 CV with fractionation. Upon analysis of the 2 ml fractions collected via SDS PAGE (III.3.3), appropriate fractions were pooled and the samples concentrated by ultra-filtration to the volumes desired (III.3.4).

III.3.3 Analysis of Proteins via SDS-PAGE

For analysis of proteins via sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) usually 20 μ l of sample were mixed with 5 μ l of 5x SDS-SB and loaded onto 15 % or 20 % polyacrylamide (PA) gels, respectively (II.4.3.2). Boiling of samples at 95 °C for 5 min prior to loading was omitted for p34ct 1-277 due to the resulting aggregation and the protein becoming stuck in the pockets of the gel. All SDS-PAGEs were run vertically at 200 V for 50 - 70 minutes, using an SDS containing Tris-Glycine buffer system [157].

After staining of the proteins with Coomassie Brilliant-Blue R-250 for 20 - 30 minutes, unspecific dyeing was removed by destaining and the gels stored in ddH₂O (II.4.3.2) for analysis. For long-term storage the gels were air dried in a sandwich of two thin sheets of cellophane foil.

III.3.4 Ultra-Filtration and Storage of Protein Samples

Before and after size exclusion chromatography and whenever necessary to increase protein molarity the samples were concentrated by ultra-filtration. Usually Vivaspin 20 or Vivaspin 6 units (Sartorius) with a molecular weight cut-off (MWCO) of 10,000 Da or 3,000 Da, respectively, were used. The membranes of such devices are designed to retain molecules of higher molecular weight than the MWCO, while smaller particles, including ions and other buffer components, are allowed to pass, thus reducing the volume and concomitantly increasing the concentration of the sample.

For ultra-filtration of p34 and p44 the proteins were centrifuged intermittently at 3,000 - 3,500 g and 4 °C until the desired volume or molarity was reached. Periodically the tubes were inverted to prevent excessive aggregation and precipitation of the sample at the membrane. From time to time ultra-filtration was also used as a quick method to exchange the protein's buffer composition.

III.3.5 Analytical Size Exclusion Chromatography (A-SEC)

Essentially identical in principle to size exclusion chromatography (III.3.2.3) analytical size exclusion chromatography (A-SEC) experiments were used to estimate the oligomeric state of protein samples and to both analyze and visualize the interaction of binding partners. As SEC separates proteins according to their size, specific elution volumes represent specific molecular masses, based on calibration values derived from a set of known protein standards. Yet, due to the elution volumes depending on the hydrodynamic volume of molecules rather than their actual size, the shape of the proteins in question may drastically influence the calculated mass. Hence the latter should only be regarded as a rough estimate.

Nevertheless, A-SEC provides an excellent tool to qualitatively perform interaction studies. Elution volumes of A-SEC runs using single proteins can be compared to runs where both partners were combined beforehand and only then applied to the column. Reduction of peak intensity for the single proteins and shifting of peaks to earlier elution volumes are good indicators of interaction and complex formation.

In the present study A-SECs were conducted using either Superdex 75 or Superdex 200 10/300 GL columns (CV \approx 24 ml), depending on the size of the proteins applied. The runs were usually performed with 8 - 12 μ M of sample and injection volumes of 400 μ l, applying flow-rates of 0.5 ml/min. The standard running buffers for A-SECs were either SEC A1 or SEC A2, rarely also SEC T was used (II.4.2). Fractions were collected over a total elution range of 0.7 - 1.0 CV in volumes of 0.5 ml each. To evaluate the procedure and verify complex formation the fractions were analyzed via SDS-PAGE (III.3.3).

III.3.6 Electrophoretic Mobility Shift Assay (EMSA)

Apart from analytical gel-filtration experiments (III.3.5) the electrophoretic mobility shift assay (EMSA) is another technique to qualitatively study protein interaction and complex formation. Here, the samples of interest are loaded on a native gel and separated according to their overall surface potential. The running behavior of single proteins can then be compared to that of a combined sample, containing both putative interaction partners. Observable protein band shifts after staining are indicative of complex formation.

In the present study all EMSAs with p34ct and p44ct were performed using 0.6 - 0.8 % native agarose gel electrophoresis (NAGE). This setup not only allowed for the evaluation of the respective protein-protein interaction, but also their binding capability towards DNA. The samples were thawed, diluted to 100 μ M in either SEC A1 or SEC A2 and then mixed to allow for complex formation at a final concentration of 20 μ M protein and 2 μ M DNA, respectively (Table III.9). After an incubation for 20 min at 4 $^{\circ}$ C the mixture was supplemented with 10 μ l of 2x NA-SB and the incubation was continued for an additional 10 - 20 minutes, until the samples were loaded into the wells of an agarose gel. The wells were placed in the center of the gel, to allow the samples to move in both directions.

Table III.9: Setup for EMSA and NAGE Interaction Studies.

Component	Stock	Final	1x
Protein 1	100 μ M	20 μ M	2.0 μ l
Protein 2	100 μ M	20 μ M	2.0 μ l
DNA	10 μ M	2 μ M	2.0 μ l
SEC A1/A2	1x	1x	4.0 μ l
			Σ 10.0 μ l

All NAGE experiments were performed at 50 V and 4 $^{\circ}$ C for 4 - 8 hours using variations of either NAGE-RB 1 or NAGE-RB 2 as both agarose solvent and running buffer. The pH milieu was varied between pH 8.0 and pH 9.5 to optimize the samples' running behavior (II.4.3.3). While the DNA was visualized using Midori Green Advance (III.1.4), proteins were stained by applying the standard Coomassie procedure (III.3.3).

III.3.7 Isothermal Titration Calorimetry (ITC)

While both A-SECs (III.3.5) and EMSAs (III.3.6) yield only qualitative results regarding interaction studies, isothermal titration calorimetry (ITC) permits the quantification of the data to calculate binding affinities. The method is based on measuring the enthalpy release (ΔH) upon a bimolecular interaction to determine association constants (K_a), stoichiometry (N) and the change in entropy (ΔS). Usually the affinity for a given binding event is given as the inverse (K_d) of the association constant (K_a) and can be derived by:

$$\Delta G = -RT \ln(K_d) = \Delta H - T\Delta S \quad \text{Equation 3}$$

with ΔG being the Gibbs free energy, R the universal gas constant and T the absolute temperature in Kelvin. The experimental setup is such that one binding partner is placed in the measuring cell, while the other is added in small but defined increments via a syringe. Each injection leads to binding and hence heat release until saturation is reached [158]. ITC measurements are usually conducted at c -values between 5 to 500, as given by the equation below (equation 4), with $[cell]$ being the concentration of protein in the measuring cell, N the stoichiometry of the interaction and K_d the expected binding affinity. This allows for an accurate sigmoidal fit of the data points measured and good quality results.

$$c\text{-value} = \frac{[cell]}{K_d} \cdot N \quad \text{Equation 4}$$

In the present study all ITC measurements were carried out using either the MicroCal™ ITC₂₀₀ or VP-ITC instruments, the parameters of which are summarized in Table III.10. The proteins were usually thawed, but sometimes also directly used after purification and stored at 4 °C in between experiments. Their proper concentration was adjusted and checked using the NanoDrop ND-1000 spectrophotometer (III.4.1).

Table III.10: Parameters for ITC Measurements.

Parameter	VP-ITC	ITC ₂₀₀
Injections	30	16 - 20
Initial Delay	5 min	2 min
Injection Spacing	4 min	3 min
Injection Volume	10 μ l	2.5 μ l
Injection Speed	0.5 μ l/s	0.5 μ l/s
Stirring Speed	260 rpm	500 rpm
Reference Power	12 μ cal/s	11 μ cal/s
Concentration in Cell	5 - 10 μ M	30 - 85 μ M
Concentration in Syringe	30 - 75 μ M	250 - 1,600 μ M

To prevent falsification of data due to heat released by dilution effects, all proteins were purified and diluted using identical buffers. While initially the SEC A1 buffer was used, experimental optimization finally led to SEC T being used for all VP-ITC measurements (II.4.2). Data fitting and evaluation was performed with the Origin 7 SR4 software package provided by the manufacturer.

III.4 Spectroscopy

III.4.1 Determination of DNA and Protein Concentration

Both DNA and protein concentrations were determined using the NanoDrop ND-1000 spectrophotometer. The principle is based on the measurement of the absorbance (A) of a given solution at wavelengths of $\lambda = 260$ nm for DNA and $\lambda = 280$ nm for proteins, respectively, resulting predominantly from the aromatic components in both molecules. Concentration values can then be derived by calculation from the following Lambert-Beer relation [159]:

$$c = \frac{A}{\epsilon \cdot d} \quad \text{Equation 5}$$

with c being the molar concentration (M, in mol/l), d the length of the light path through the solution (in cm) and ϵ the extinction coefficient (in $M^{-1} \text{ cm}^{-1}$) of the molecule in question. For proteins the latter was derived by calculation based on their primary sequence using the ProtParam tool [140]. The extinction coefficients of DNA were provided by the ND-1000 software and are generally accepted to be $50 \text{ (ng/}\mu\text{l)}^{-1} \text{ cm}^{-1}$ for dsDNA and $33 \text{ (ng/}\mu\text{l)}^{-1} \text{ cm}^{-1}$ for ssDNA.

III.4.2 Thermofluor (TF)

Thermofluor was used to measure the melting curves of proteins in solution upon thermal unfolding. Originally developed for rapid high-throughput screening of compound libraries and ligand identification [160] it has since been used for a variety of applications. The principle is based on a fluorescent dye, for example SYPRO® Orange (Invitrogen), that binds to hydrophobic amino acids of proteins in solution, which are predominantly exposed upon denaturation. The resulting environmental change leads to an increased absorption at 492 nm and emission at 610 nm of the dye, which can be measured and translated into a protein melting or unfolding curve, respectively [161,162]. In combination with a real-time PCR cycler and multi-well system the method can be used to rapidly screen for protein stability in various aqueous solutions, for example regarding the buffer composition, pH, ionic strength or compound additives [163].

Table III.11: Setup for Thermofluor Experiments.

Component	Stock	Final	1x
Buffer	100 mM	92 mM	23.0 μl
SYPRO® Orange	125x	5x	1.0 μl
Protein	2 mg/ml	0.08 mg/ml	1.0 μl
			Σ 25.0 μl

In the present study thermofluor experiments were prepared at RT in a 96-well PCR plate (Greiner Bio-One) using previously prepared buffer conditions from a master plate (Table III.11). All data were acquired in a Stratagene® MX3005P™ real-time PCR cycler (Agilent Technologies). Starting at 25 °C, the temperature was slowly increased by 1 °C steps up to 95 °C, covering a total of 70 data points for each well. The results were analyzed using an excel sheet provided by the structural genomics consortium (SGC) in Oxford.

III.4.3 Circular Dichroism (CD)

Circular Dichroism (CD) spectroscopy was applied to study the proper folding of proteins in solution. The method employs circular polarized light, which is known to be unequally absorbed by chiral molecules in an aqueous environment, depending on their structural asymmetry. This results in elliptically polarized light, the ellipticity (θ) of which can be calculated from the difference in absorption and plotted against the wavelength (λ). Secondary structure elements such as α -helices and β -strands yield characteristic CD spectra, which can be analyzed to estimate the relative amount of such regions compared to the unfolded parts of a protein [164]. While α -helical superstructures lead to spectral minima at 208 nm and 222 nm and a maximum at 192 nm, the chirality of β -sheets results in a minimum at 215 nm and a maximum at 198 nm. These secondary structure elements can be clearly distinguished from random coils, which contribute to the spectrum with a pronounced minimum at around 200 nm [165]. The technique is especially useful when comparing, for example, protein variants to the wild type molecule, potentially revealing overall structural changes upon mutation of the respective residues.

All CD measurements with p34ct and p44ct were conducted at 20 °C using a 0.1 cm quartz cuvette and the J-810 spectropolarimeter (Jasco). The proteins were thawed and diluted approximately 100 fold to 0.1 mg/ml in 20 mM potassium phosphate at pH 8.1 (SEC D, II.4.2), to reduce noise resulting from absorption of chiral buffer components. The spectra were recorded from $\lambda = 190$ nm to $\lambda = 260$ nm at a scan speed of 20 nm/min, at 2 nm band width, a data pitch of 0.1 nm and 1 second response time. To further improve the overall signal to noise ratio a total of 10 CD spectra were accumulated.

III.4.4 Multi Angle Light Scattering (MALS)

To accurately determine the molecular mass of a protein in solution a SEC-MALS approach was used. Here the common size exclusion chromatography (SEC, III.3.2.3) principle is coupled to both an on-line refractive index (RI) and multi angle light scattering measurement (MALS). While the UV signal of such a SEC-UV/LS/RI setup mainly serves to monitor separation and visualize the protein peaks, the LS and RI components allow to retrieve the molecular mass distribution of the sample. The underlying light scattering equation (LS) can be simplified to the following relation:

$$LS = K_{LS} c W (dn/dc)^2 \quad \text{Equation 6}$$

with K_{LS} being an instrument calibration constant, c the sample concentration (in mg/ml) and W the average molecular weight (in Da). Similarly the refractive index signal (RI) can be expressed as:

$$RI = K_{RI} c (dn/dc) \quad \text{Equation 7}$$

with K_{RI} being a different instrument calibration constant and c again the concentration (in mg/ml). With the dn/dc value generally accepted to be 0.185 ml/g for proteins carrying no post-translational modifications, the molecular weight (W) can be calculated from the ratio of both the LS and RI signals [166,167]. Visualized as a mass distribution over the elution volume, the precise molecular weight of single peaks can be determined, which in turn also allows to estimate the oligomeric state of proteins or protein complexes in solution.

For SEC-MALS measurements the samples were thawed and diluted to a concentration of 10 - 30 μM using either SEC M1 or SEC M2 (II.4.2), which also served as running buffers for the analytical Superdex SD 200 10/300 GL column (GE Healthcare). The injection volume was set to 100 μl and all runs were performed at a flow-rate of 0.5 ml/min without fractionation. To measure the LS and RI signals, respectively, the DAWN® 8+ HELEOS II® and OptiLab® T-rEX™ HC instruments (Wyatt Technologies) were employed and directly coupled to the size exclusion column, which itself was attached to an ÄKTA™ Purifier FPLC system. The recorded UV profile and signals from both MALS and RI detectors were analyzed using the ASTRA software package provided by the Wyatt Technologies Corporation.

III.4.5 Mass Spectrometry (MS)

Mass spectrometry (MS) allows to collect spectra of a sample's atomic composition yielding a specific signature for each molecule. With respect to proteomics this is usually accomplished by enzymatic digestion of the sample of interest and subsequent ionization to generate singly charged protein fragments [168]. The mass-to-charge ratio of such particles can then be accurately determined, for example, by time-of-flight measurements, with lighter ions traveling faster than heavier ones, ultimately reaching the detector of an MS instrument first [169]. By comparison of the collected spectrum with signatures obtained by *in silico* digestion of proteins from a database, proper identification of the molecule in question can be achieved. This example represents only one of the various and powerful applications based on mass spectrometry.

In the present study MS experiments were pursued to identify specific proteins or protein fragments present after purification and in protein crystals, respectively. Usually the samples were first separated via SDS-PAGE prior to analysis (III.3.3) and stained with Coomassie Brilliant Blue R-250. The protein bands of interest were then carefully cut from the gel and stored separately at 4 °C until used. In some cases the samples were treated with a final concentration of 200 mM iodoacetamide for 20 minutes at 4 °C in the dark before separation. Together with the reducing agent (final concentration of 80 mM DDT) present from the sample preparation step using 5x SDS-SB (II.4.3) this procedure led to the alkylation of cysteine residues and hence prevented cysteine oxidation and the formation of disulfide bonds. The actual MS measurements were carried out by the group of Andreas Schlosser, our in-house collaborators, who also helped in evaluating the results.

III.5 X-Ray Crystallography

III.5.1 General Introduction

Macromolecular x-ray crystallography is one of the oldest and most widely used techniques for obtaining 3D structural information on proteins and protein complexes. As of February 2014 a total of 90,143 protein structures were recorded in the Protein Data Bank (PDB), of which an astonishing 89.2 % were solved using conventional x-ray crystallography methods (www.rcsb.org/pdb/statistics/holdings.do).

Although several other methods are available to determine protein structures, these often suffer from certain drawbacks, like size-limitation (NMR), only low to moderate resolution (cryo EM) or adverse effects due to excessive radiation damage (electron crystallography). In contrast, x-ray crystallography usually allows the collection of diffraction data with a minimum of radiation damage at close to atomic resolution. In addition, there is practically no size-limitation imposed on the sample. However, protein crystallization can be a very time consuming and rate-limiting process, often resulting in poor quality crystals. Another disadvantage of the technique is the fact that the structure of the molecule in question is determined in a crystalline state, usually yielding a snapshot of one single conformation. Hence additional methods have to be employed to verify the observations made and the results inferred from a protein crystal structure.

III.5.2 Principles of X-Ray Crystallography

Structure determination by x-ray crystallography is based on the interaction of monochromatic x-rays with the electron clouds of a molecule's atomic composition, resulting in diffraction. The deflected x-rays take on a different path and interfere with each other, leading either to destructive or constructive interference. In 1913 W.L. Bragg showed the latter to occur only if the x-ray's incidence and emergence angle θ meets the following mathematical condition:

$$\text{Bragg's Law} \qquad n \lambda = 2 d_{hkl} \sin(\theta) \qquad \text{Equation 8}$$

with d_{hkl} being the interplanar spacing of a set of reflection planes (hkl), n an integer and λ the wavelength of the electromagnetic wave impinging on the planes [170]. Constructive interference is required to yield strong x-ray diffraction beams, which can be visualized by hitting an appropriate detector array. This results in a specific pattern of unique reflections, each of which being a multi-component wave that can be described as the sum of diffractive contributions from all atoms in the unit cell, called the structure factor F_{hkl} . For a cell containing n atoms the Fourier sum of all the atomic f_{hkl} values for individual atoms would be:

$$\text{The Structure Factor} \qquad F_{hkl} = \sum_{j=1}^n f_j(h, k, l) e^{2\pi i(hx_j + ky_j + lz_j)} \qquad \text{Equation 9}$$

where F_{hkl} is the structure factor of the reflection hkl , f_j the element-specific atomic structure factor, i the imaginary number ($i = \sqrt{-1}$) and x_j , y_j and z_j the coordinates of atom j . Each term of the summation describes the contribution of one atom, which is treated as a single sphere of electron density [171].

Alternatively, the structure factor F_{hkl} can be expressed as the sum of contributions from each volume element of the local electron density ρ in the unit cell:

$$\text{The Structure Factor} \quad F_{hkl} = \int_x \int_y \int_z \rho(x, y, z) e^{2\pi i(hx+ky+lz)} dx dy dz \quad \text{Equation 10}$$

Mathematically the structure factor F_{hkl} is a Fourier transform of the electron density within the unit cell and a Fourier transform as such is reversible. Hence the electron density ρ can be described by the reverse Fourier transform of the structure factor F_{hkl} to obtain the electron density ρ as a Fourier sum [171]:

$$\text{The Electron Density} \quad \rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx+ky+lz)} \quad \text{Equation 11}$$

From this equation the electron density ρ at each 3-dimensional coordinate (x, y, z) can be calculated, if the volume of the unit cell (V), the structure factor amplitudes F_{hkl} and its phase angle (α) are known. However, in an x-ray diffraction experiment only the unit cell volume and the reflection intensities I_{hkl} of each structure factor can be measured. While the structure factor amplitudes F_{hkl} can be calculated from the reflection intensities I_{hkl} based on the following proportional relation, the phase information φ cannot be retrieved.

$$\text{Amplitude} \sim \text{Intensity} \quad |F_{hkl}| \sim \sqrt{I_{hkl}} \quad \text{Equation 12}$$

The loss of the essential phases φ during the course of the experiment is called the “phase problem” in x-ray crystallography. Without the phase information the electron density equation is not fully satisfied and hence the structure of the protein crystal in question cannot be solved.

III.5.3 Phase Determination

The importance of the phase information in relation to mere intensities for calculation of an electron density map is nicely illustrated in Kevin Cowtan’s Book of Fourier [172]. It is also important to note, that each reflection has its own phase, hence the phase problem must be solved for every single reflection used to construct the Fourier sum to approximate $\rho(x, y, z)$.

Several techniques have been developed to determine the phases, all of which are based on assuming some prior knowledge of the structure, from which they arise. While direct methods can be used to bootstrap phases from an initial phase estimate for small molecules up to 200 atoms, a more complicated and indirect approach is required for larger structures. Modified experimental setups like single or multiple isomorphous replacement (SIR, MIR) can be used to calculate phases from differences in the measured intensities upon derivatization of a crystal with heavy atoms. Often suffering from poor isomorphism, this method is augmented by combining it with anomalous scattering (SIRAS, MIRAS) and exploiting anomalous dispersion differences in F_{hkl} , which also help in resolving phase ambiguities [173]. To overcome the problems inherent to the use of different derivatives, multi-wavelength anomalous diffraction (MAD) can be used, in which data of a single crystal is collected at several wavelengths to maximize the anomalous and dispersive effects [173]. In addition, recent advances in phasing and density-modification algorithms also allow to calculate phases from a single anomalous diffraction (SAD) experiment, using high quality data collected at the absorption peak of the anomalous scatterer [174].

Apart from the experimental techniques discussed above, phases can also be determined *in silico* if similar structures of high homology are already available. Serving as search model for molecular replacement (MR) these can be used to determine the phases after correct positioning of the model in the unit cell of the unknown structure. The latter is achieved by first calculating the Patterson function of both the measured data and the model structure to derive two three-dimensional Patterson maps that can be aligned using a rotation function to properly orient the search model. As soon as the correct orientation has been determined a translation function is applied to locate the best position of the model within the unit cell based on a new set of Patterson maps [171,175]. The properly oriented search model can then be used to calculate an estimate for the phases of the unknown structure. This approach is usually successful if the sequence identity between the unknown structure and the homology model is at least 25 - 30 %. In addition, the C α atoms of both structures should have an r.m.s. deviation below 2.0 Å to obtain reliable phase information [173].

In the present study the experimental SIRAS approach was used for phasing and *de novo* solution of the p34ct crystal structure, after initial attempts via SAD and a single potassium iodide derivative failed. Additional x-ray data, revealing the p34ct 1-277 vWA domain at higher resolution and the p34ct p44ct minimal complex, could then be solved by molecular replacement (MR), using the SIRAS based structure as a search model. Details regarding these procedures can be found in the results section of this work (IV.5.2, IV.5.4, IV.6.2).

III.5.4 Protein Crystallization

Protein crystallization is an empirical technique, which is influenced by a variety of factors, like protein and precipitant concentration, pH value, ionic strength and temperature. This bottleneck process in x-ray crystallography involves a “trial and error” approach, screening the protein of interest for nucleation and crystal growth over a multitude of different conditions. If a suitable condition is identified, it is subsequently optimized to yield high quality crystals large enough for x-ray diffraction and analysis.

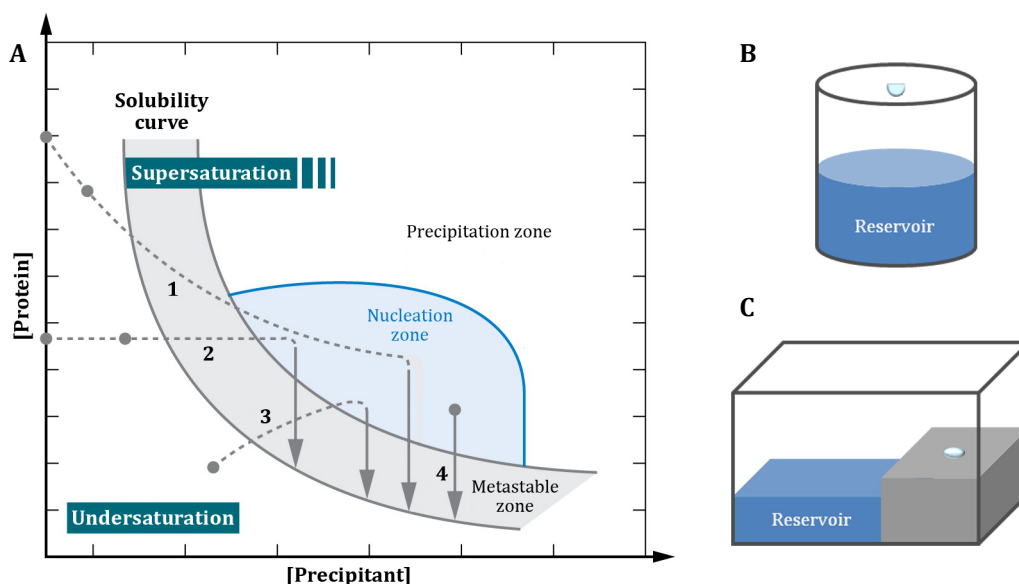


Figure II.1: Phase Diagram and Methods of Protein Crystallization. Proteins can be crystallized from solution applying several methods. The phase diagram (A) based on the two most commonly varied parameters, protein and precipitant concentration, illustrates the principles of Free Interphase Diffusion (1), Microdialysis (2), Vapor Diffusion (3) and Microbatch (4). After successful growth initiation in the nucleation zone (light blue) the protein concentration sharply drops (solid arrows) and crystal growth continues in the metastable zone (light grey). For Vapor Diffusion experiments usually either the hanging drop (B) or the sitting drop (C) setup is applied. The phase diagram (A) is adapted from [176] and originally appeared in [177], [Acta Cryst. \(1998\), D54, 8-15](#). Reproduced with permission of the IUCr, ©1998.

In the present study all protein crystallization trials were performed using the vapor diffusion method. Here, a small amount of protein is mixed with a defined volume of precipitant solution and the droplet placed over a reservoir containing the precipitant solution only. The setup is sealed airtight to prevent evaporation and to allow for the establishment of an equilibrium condition. Due to the higher vapor pressure in the mixed drop compared to the reservoir solution, water will diffuse from the droplet towards the reservoir, resulting in a reduction of the drop's volume. This in turn leads to an increase in both the protein as well as the precipitant concentration in the droplet, moving the system towards a state of supersaturation [176,177]. Reaching the nucleation zone will then potentially yield crystals, which rapidly deplete the solution with respect to its protein content and continue to grow in the metastable zone (Figure II.1 A).

For protein crystallization using the vapor diffusion technique either the "hanging drop" or the "sitting drop" setup were used (Figure II.1 B and C). Trials employing the former were prepared manually using cover slips to hold the protein-precipitant mixture. The cover slips were inverted to seal the paraffin oil coated wells of a 24-well plate (Crystalgen) containing 1 ml of the respective precipitant solution. Experiments with the "sitting drop" setup were prepared using the HoneyBee 961/963 crystallization robot and 96-well plates (Greiner Bio-One). Here usually a total of 0.3 μ l of protein were mixed with an equal volume of precipitant solution and placed on a pedestal above a 40 μ l reservoir. The plates were then sealed with plastic foil.

During the initial screening process roughly 800 - 900 different conditions were tested using "sitting drops" and recipes provided by commercially available screens (II.2.3). Upon successful crystallization of a protein, the respective condition was optimized manually applying the "hanging drop" setup. All crystallization plates were incubated at 20 °C or 4 °C and evaluated regularly to monitor each condition and detect crystal growth.

III.5.5 Data Collection

To measure both the reflection position and intensity of each diffraction event a crystal has to be mounted between an x-ray source and the respective detector. By using a fine tuned goniometer head the crystal can then be rotated in the x-ray beam to allow for the collection of a series of diffraction patterns, which, taken together, yield a data set. The collected data is complete when a sufficient angular range has been covered, depending on the inherent symmetry and space group of the protein crystal in question.

For data collection crystals were fished from the crystallization drop using a stereo microscope (Zeiss) and small nylon loops (Hampton Research). Crystals of p34ct and the p34ct p44ct minimal complex were then flash frozen in liquid nitrogen either directly in their mother liquor, an artificial cryo-solution or by using paraffin oil as a cryo protectant. In order to collect anomalous data some of the p34ct crystals were soaked in a 1.0 μ l drop of 23 % PEG 550 MME, 100 mM Tris pH 8.0 and 100 mM KI for 1 - 5 min prior to flash freezing. All crystals were stored in liquid nitrogen until used for x-ray diffraction experiments.

Besides in-house measurements for testing purposes, data collection of flash frozen crystals was performed at the European Synchrotron Radiation Facility (ESRF) in Grenoble (France) and the Berliner Elektronenspeicherring Synchrotron (BESSY) in Berlin (Germany). All relevant diffraction experiments were conducted at 100 K using beamlines ID 23.2 (ESRF) for the p34ct and p34ct p44ct native data, BL 14.1 (BESSY) for p34ct anomalous KI data and ID 29 (ESRF) for p34ct 1-277 vWA native data at x-ray wavelengths of 0.87 Å, 1.60 Å and 0.92 Å, respectively. More detailed information regarding crystal preparation and data collection can be found in the respective results section within this work (IV.5.1, IV.5.3, IV.6.1).

III.5.6 Data Processing and Structure Determination

Following data collection the measured raw intensities need to be processed to improve their consistency. This can be achieved by first indexing all reflections and then comparing reflections that have the same index but have been recorded from different crystals or on more than one diffraction frame. Re-scaling of these then leads to identical reflections having identical averaged intensities [171]. These processing steps are referred to as indexing, integrating and scaling and are required for a data set to be practically useful in terms of structure determination. The structure factor amplitudes (III.5.2) can then be calculated from the merged and scaled intensities. Additional postrefinement procedures may be applied to recover usable data from reflections that were only partly measured [171].

In the present study all collected diffraction data sets were indexed and processed using either iMOSFLM and SCALA [124,129] or XDS [131]. *De novo* structure solution of p34ct was achieved using SHELX C/D/E [130] and employing a SIRAS approach by combining anomalous p34ct KI data at 4.2 Å with a highly isomorphous p34ct native data set at 3.8 Å to locate the heavy atom substructure. Initially obtained electron density maps allowed to manually build a p34ct preliminary model using the programs O and COOT [125,123] (see III.5.7). Upon collection of higher resolution data, the preliminary model could then be used to solve the structures of both p34ct 1-277 vWA and the p34ct p44ct minimal complex by molecular replacement via Phaser [126].

III.5.7 Model Building and Refinement

Successful determination of a protein structure usually yields an initial and incomplete model, still containing many errors. This is especially the case with a *de novo* solution, the result of which often is a bare electron density map of the unit cell content without any attributed features. To properly establish and correct the initial model, the structure is fitted to the calculated electron density map in real space using molecular modeling tools like COOT and O [123,125]. These not only allow to attune the model to the data observed, but also help to retain the proper geometry of single residues and higher order secondary structure elements. Phases calculated from the generated model can then be applied to improve the initial phase information during a refinement process in reciprocal space, by using the least-squares method and programs like Phenix and Refmac5 [127,178]. Following iterative cycles of model building and refinement, the structure finally converges yielding anything but minor improvements.

The overall progress of iterative refinement in real and reciprocal space can be monitored by comparing the measured structure factor amplitudes $|F_{obs}|$ with the calculated amplitudes $|F_{calc}|$ from the current model. Relating these two values in the following equation yields the so called R-factor, which describes convergence and hence is an indication for the quality of the resulting model [171]:

$$\text{The R-Factor} \quad R = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|} \quad \text{Equation 13}$$

The initial model of the p34ct vWA domain was built in O [125], later corrected and extended in COOT [123] and iteratively refined against the p34ct 1-277 vWA native data to a resolution of 2.7 Å using Phenix [127]. Both COOT and Phenix were also used for model building and refinement of the p34ct p44ct minimal complex structure upon successful molecular replacement with the p34ct 1-277 vWA domain (III.5.6). Visualization of all final models and image generation was performed using the software PyMOL [128].

IV. Results and Discussion

IV.1 Towards Crystallization of TFIID Subunits from *H. sapiens*

The project presented in this work initially started with attempts to purify and crystallize subunits of the human TFIID core complex. Although working with the human proteins turned out to be rather difficult and the results obtained did not meet our expectations, the achievements made will nevertheless be summarized in the paragraphs below. This will help the reader to better understand where the project initiated and what factors finally led to changing the source organism and ultimately succeed in structure determination of at least two TFIID core subunits.

IV.1.1 Initial Attempts with human XPD, XPB, p44 and p52

To shed more light on the structural composition of the human (h) TFIID core complex, attempts were made to purify and crystallize the two central helicases XPDh and XPBh as well as their regulatory binding partners p44h and p52h, respectively. The genes of these TFIID subunits were PCR amplified from a cDNA library and subsequently cloned into pFastBac HTa and pFastBac Dual (Invitrogen) for expression in *Sf21* and *Sf9* insect cells using a baculovirus based system (III.1.10). Concomitantly the proteins were also cloned into the bacterial vectors pETM-11 and pBADM-11, allowing for recombinant expression in *E. coli*. The generation of all pFastBac vector constructs was accomplished by InFusion® Cloning (Clontech), whereas the *E. coli* constructs were created using standard restriction enzyme based cloning procedures.

The obtained pFastBac constructs were then used to generate Bacmid DNA by transformation of competent *DH10Bac* cells (III.1.10). Positive recombinants were identified by blue/white screening and the DNA finally purified via DNA precipitation using buffers from the NucleoSpin® Plasmid Kit (Macherey-Nagel) and Isopropanol (III.1.11). In the following *Sf21* insect cells were transfected with the crude Bacmid DNA extract to generate baculoviruses (III.2.4), which were subsequently used for virus amplification (III.2.6) and protein expression (III.2.7). Viral titers were determined by plaque assay (III.2.5), while the expression of TFIID subunits was analyzed by taking 1 ml samples at different points in time during virus amplification. After cell lysis (by freeze and thaw or sonication) and separation of the pellet from the supernatant the samples were analyzed by SDS-PAGE. Protein expression was verified via western blots using primary antibodies targeted against the His₆-Tag of each construct.

All the TFIID subunits tested could successfully be expressed in *Sf21* and *Sf9* insect cells. However, both XPDh and XPBh as well as p52h were only detectable in the pellet fractions and hence were completely insoluble. The p44h subunit showed faint signals in the supernatant after cell lysis, yet seemed to degrade over time. Attempts to increase solubility by co-expression of XPDh with p44h and XPBh with p52h using the pFastBac Dual constructs failed as well, yielding only insoluble complexes. Expression tests concomitantly performed in *E. coli* confirmed the results obtained with the baculovirus based system. Here, again only p44h could be detected in the soluble fraction, albeit with severe degradation effects, which had already been observed during expression. Nevertheless attempts were made to optimize the strategy and purify the protein (IV.1.2).

IV.1.2 The human p44 Protein

Upon initial expression trials of TFIIH subunits in insect cells and *E. coli* (IV.1.1), a portion of p44h was detected in the soluble fraction, although the bulk of the protein remained insoluble. To improve expression behavior, the focus was shifted towards the cheaper and more accessible *E. coli* system and several truncated p44h variants were designed. The constructs were based on secondary structure prediction results using the Phyre algorithm [139] and reflected the primary domain arrangement of the protein (Figure IV.1).

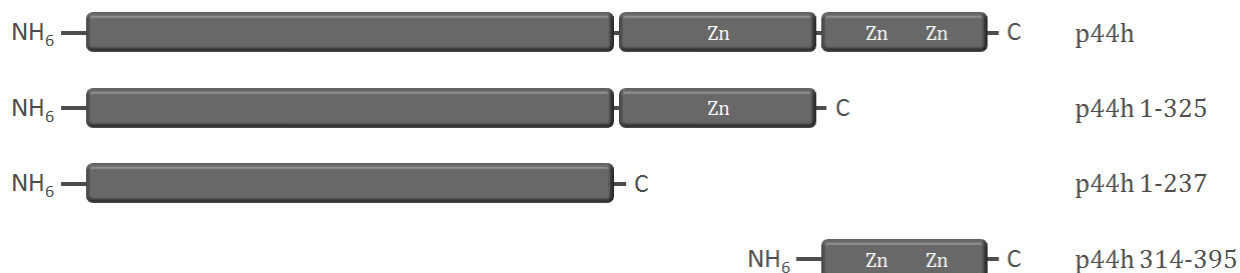


Figure IV.1: Domain Structure of p44h Variants. Full-length p44h consists of an N-terminal domain, a central zinc binding domain and a C-terminal C4C4 RING domain coordinating two additional zinc ions (top panel). The shortened variants primarily used in this work are indicated below the full-length protein. All constructs were cloned into the pBADM-11 vector and fused to an N-terminal His₆-Tag.

For expression in *E. coli* all p44h variants were cloned into the pBADM-11 vector and fused to an N-terminal His₆-Tag. Initial small-scale time course experiments indicated that protein expression usually peaked at around 8 hours post induction, except for p44h 314-395, which reached its maximum yield after 20 hours. Apart from the obvious degradation of p44h full-length, p44h 1-325 and p44h 1-237 already during expression, the level of expression in general was comparable and rather low for all three constructs.

Upon large-scale expression trials in 4x 2L LB-medium the purification of all p44h variants was performed using identical protocols. An affinity chromatography step via Ni-MAC columns (Novagen, III.3.2.1) was followed by concentration of the elution fractions, which were then loaded on a Superdex 200 16/60 column (GE Healthcare, III.3.2.3) for size exclusion chromatography. The SEC elution fractions were analyzed via SDS-PAGE and compared in terms of yield and purity (Figure IV.2).

All p44h variants, but especially p44h 1-325 and 1-237, yielded multiple bands of similar intensity indicative of protein degradation. Keeping the eluates at 4 °C for some time led to further degradation and enrichment of one of the two fragments running slightly below the 25 kDa marker band for all constructs except p44h 314-395 (Figure IV.2). Using various protease inhibitors during purification did not prevent degradation, the latter of which could be confirmed by mass spectrometric analysis.

While p44h 314-395 also showed signs of degradation, always appearing as a double band after SDS-PAGE (Figure IV.2, lane 5), its yield was about 5 times higher compared to the other constructs, likely due to the better

solubility of the p44h C-terminal C4C4 RING domain. This beneficial property turned out to be very useful for the purification of a soluble p34h p44h minimal complex, which will be discussed in the next section (IV.1.3).

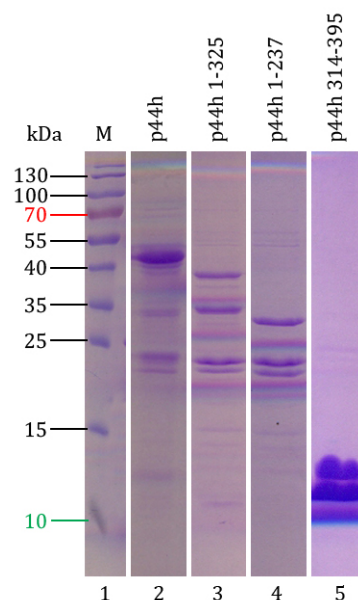


Figure IV.2: Variants of p44h. Shown are p44h full length (2), p44h 1-325 (3), p44h 1-237 (4) and p44h 314-395 (5) upon separation via SDS-PAGE.

IV.1.3 The p34h p44h Minimal Complex (hMC)

The p34 subunit is the natural interaction partner of p44 within TFIIF and similar in primary domain architecture. It consists of an N-terminal domain, which is followed by one zinc-finger motif in its C-terminal region, yet lacks the C4C4 RING domain found at the remote C-terminus of p44 [49,179]. Initially, the interaction was reported in yeast-two-hybrid studies and later confirmed by co-immunoprecipitation, though the p34 subunit generally eluded purification due to its low solubility [41,180]. However, by randomly co-expressing various p34h and p44h constructs of different length in *E. coli*, Fribourg *et al.* discovered that the specific combination of p34h 1-242 and p44h 321-395 yielded a soluble minimal complex [42].

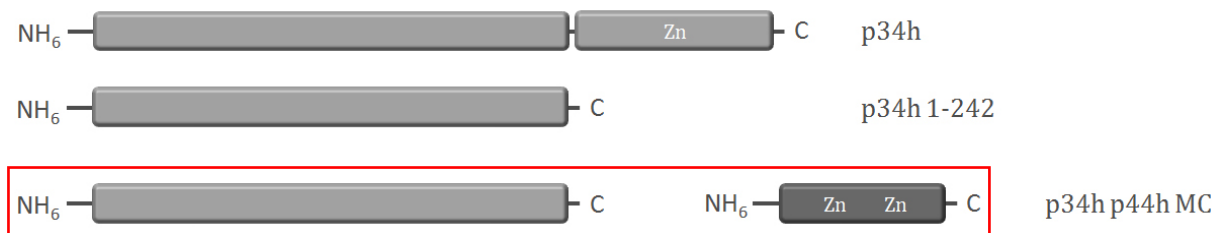


Figure IV.3: Domain Structure of p34h Variants and the p34h p44h MC. Full-length p34h consists of an N-terminal domain and a C-terminal zinc binding domain (top panel). The shortened variant p34h 1-242 (below) was used to generate the p34h p44h minimal complex (boxed in red) by co-expression with the p44h 314-395 construct (see Figure IV.1). All p34h constructs were cloned into the pETM-11 vector and fused to an N-terminal His₆-Tag.

With this knowledge in mind both the p34h full-length and p34h 1-242 variant were cloned into the pETM-11 vector for expression trials in *E. coli* (Figure IV.3). Using BL21(DE3)-RIL cells very high expression levels were observed for both constructs, but neither the full-length nor the shortened p34h variant could be purified on their own, as they were completely insoluble. However, co-expression of p34h 1-242 and p44h 314-395 (Figure IV.3) indeed pulled at least a portion of p34h into the soluble fraction and a pure and stable p34h p44h minimal complex (hMC) could be obtained (IV.1.3.1).

IV.1.3.1 Purification and Crystallization of p34h p44h MC

The p34h p44h minimal complex was purified by either Ni-MAC (Novagen) or Ni-TED (Macherey-Nagel)

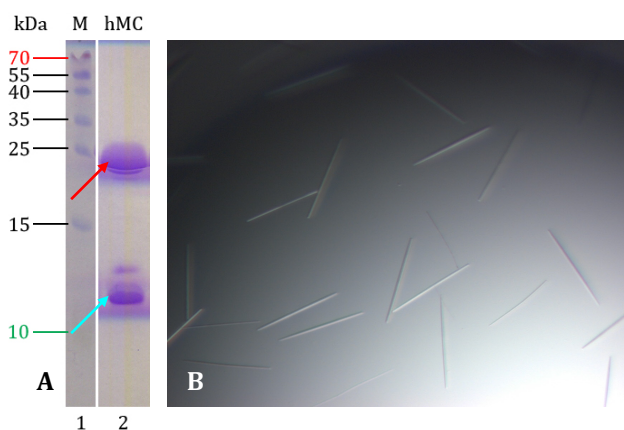


Figure IV.4: The p34h p44h Minimal Complex. Purified sample of the p34h p44h MC analyzed via SDS-PAGE (A, hMC) and crystals of the minimal complex grown in 8 % (w/v) PEG 6,000 and 100 mM Tris-HCl pH 8.0 (B). The proteins in (A) are indicated by red (p34h 1-242) and cyan (p44h 314-395) arrows.

affinity chromatography (III.3.2.1), followed by size exclusion chromatography using a Superdex 200 16/60 column (GE Healthcare, III.3.2.3). While buffers Ni-MAC A/B or Ni-TED A/B were used for the former, the latter was performed in buffer SEC P1 (II.4.1). Upon optimization of the purification strategy all samples obtained were of moderate yields and very clean, as confirmed by SDS-PAGE analysis (Figure IV.4 A).

Initial crystallization trials yielded tiny needle like crystals in several PEG based conditions, which were subsequently optimized using hanging drop vapor diffusion experiments in 24-well plates (Figure IV.4 B).

However, the needles obtained were extremely thin and prone to fragmentation during fishing and cryo-cooling for x-ray diffraction analysis. While at first no diffraction could be observed with the in-house x-ray system, a faint but visible diffraction pattern was detected using high energy synchrotron radiation at the BESSY facility in Berlin. This indicated that the crystals are composed of protein and prompted further optimization trials (IV.1.3.2).

IV.1.3.2 Crystal Analysis and preliminary X-Ray Diffraction of p34h p44h MC

Upon progressing micro- and macro-seeding attempts, a few larger and more stable crystals could be grown from p34h p44h MC samples, allowing for closer inspection. The crystals assumed a flat bar like shape, having lengths of more than 200 μm , a width of approximately 10 - 20 μm and generally depths of only a few μm (Figure IV.5 B). In addition, most crystals appeared to be partly hollow under the microscope, probably contributing to their still fragile character. Nevertheless, analysis of some of the thickest crystals obtained using a micro-focus synchrotron beamline at the ESRF in Grenoble now allowed the collection of a series of much stronger diffraction images. One of the putative p34h p44h MC crystals tested diffracted up to 3.2 \AA , yet a dataset collected could only be processed to roughly 6 \AA due to the inherently high anisotropy of the crystal's diffraction pattern (Figure IV.5 C).

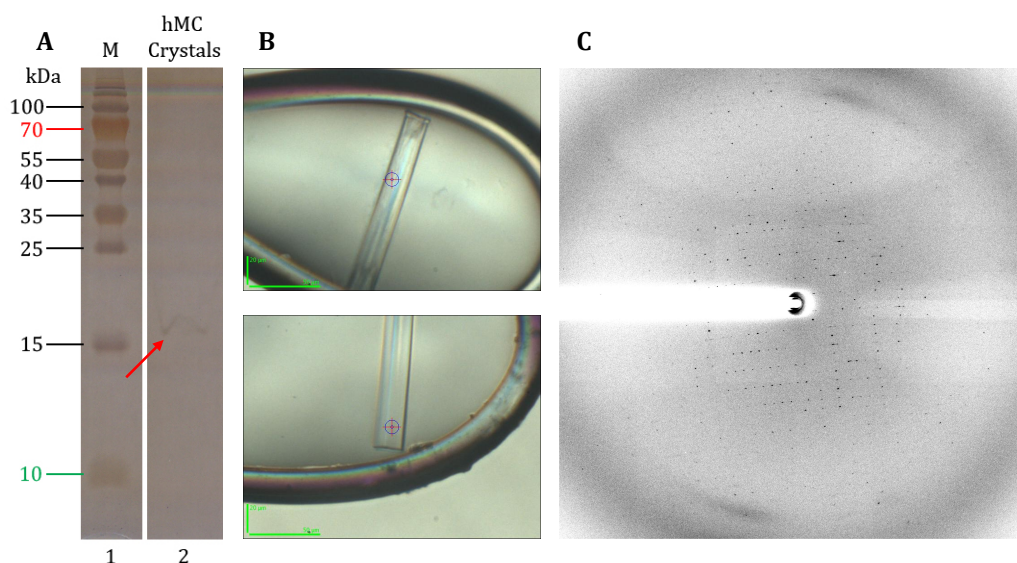


Figure IV.5: Analysis of p34h p44h MC Crystals and X-Ray Diffraction. Panel (A) shows crystals grown from p34h p44h MC samples as analyzed by SDS-PAGE and silver staining. Only a single protein band was detectable in the form of a downward smear between the 25 kDa and 15 kDa marker band (A, red arrow). Several fished and cryo frozen crystals (B) were tested and preliminary x-ray diffraction data could be collected (C). The size axes in (B) indicate 20 μm (y-axis) and 50 μm (x-axis), respectively.

Apart from better diffraction properties growing thicker crystals also allowed for the investigation of their actual contents. To visualize the proteins forming the crystal lattice via SDS-PAGE and silver staining a total of 15 - 20 of the putative p34h p44h MC crystals were fished. After extensive washing in reservoir solution, the crystals were supplemented with 5x SDS-SB, boiled at 95 $^{\circ}\text{C}$ and separated on a 20 % PA-Gel (II.4.3.2).

Interestingly, silver staining of the resulting gel revealed only one of the two protein bands expected. In addition, the band did not run at a molecular weight either corresponding to p34h 1-242 or p44h 314-395. Instead it appeared slightly above the 15 kDa marker band (Figure IV.5 A, red arrow), although downward smearing of the protein may indicate a somewhat higher molecular weight than observed.

Possible contaminant proteins of similar molecular weight could barely be detected by SDS-PAGE in all the p34h p44h MC samples used (compare Figure IV.4 A). However, further analysis of putative p34h p44h MC crystals by mass spectrometry indicated the crystals to consist predominantly of a catabolite activator protein originating from *E. coli* (CAP, CRP), several structures of which are already available [181]. With a molecular weight of about 23 kDa, this contaminant, or a degradation product thereof, might indeed correspond to the protein observed by silver staining of the crystals grown (Figure IV.5 A). The very low concentration of the CRP contaminant in all purified p34h p44h MC samples also explains the crystals' strong resistance towards optimization. In fact, it is kind of paradox to realize, that in this case it was necessary to crystallize a protein in order to identify it as an actual contaminant of an otherwise highly pure and stable p34h p44h complex sample, which consistently eluded crystallization.

Finally the project centered around the human TFIIH proteins was set aside in favor of TFIIH homologues from the eukaryotic fungus *Chaetomium thermophilum* (IV.2). This interesting organism turned out to be a very promising candidate for obtaining pure and stable TFIIH subunits in a way previously not achievable.

IV.2 Characterization of TFIIH Subunits from *C. thermophilum*

With the human TFIIH core proteins proving difficult to work with and the p34h p44h minimal complex resisting crystallization attempts (IV.1), alternative options were investigated to obtain TFIIH subunits more suitable for structure determination. Due to the fact that most of the TFIIH core subunits are absent in archaea and bacteria, which usually serve as distant but handy model organisms, another eukaryotic system had to be identified, which is more amenable to structural studies. Very recently the structure of a nuclear pore multi-protein assembly, that so far had eluded characterization, could successfully be described utilizing the genome of *Chaetomium thermophilum* (ct) [20], a eukaryotic and thermophile fungus. Despite the fact that it was already discovered more than 60 years ago [21], it has only now been recognized as a novel model organism, providing the scientific community with a closely related proteome of good accessibility and high stability. Hence, in an attempt to shed more light on the structure and function of TFIIH, we started to work with homologues from this fungus, which contains all the CAK and TFIIH core subunits.

The following chapters describe the very first purification of several p34 (IV.2.1) and p44 (IV.2.2) variants from *C. thermophilum* including the p34ct p44ct minimal complex (IV.2.3), which was designed analogous to its human counterpart. In addition, the proteins will be characterized with respect to yields and purity, as well as their thermostability (IV.2.5) and oligomeric state (IV.2.6) in solution.

IV.2.1 Purification of p34ct Variants

Both the p34ct and p44ct subunits were cloned from *C. thermophilum* cDNA as described in III.1.1 and III.1.5. Initial expression trials with p34ct full-length (amino acids 1-429) surprisingly yielded not only moderate to good expression levels but also protein in the soluble fraction. This was in contrast to the human homologue of p34ct, which was completely insoluble when expressed on its own in *E. coli* (IV.1.3). Upon large-scale expression in 4x 2L LB-medium, p34ct could be purified reproducibly via Ni-MAC or Ni-TED affinity chromatography followed by SEC (III.3.2). The yields were usually in the range of approximately 0.6 - 1.3 mg per liter of *E. coli* cell culture, with only negligible amounts of protein in the insoluble fraction.

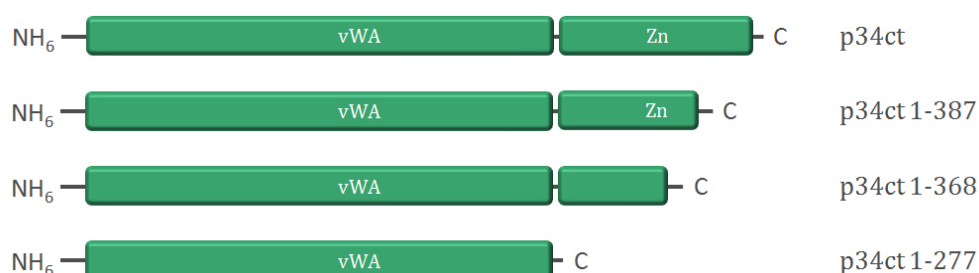


Figure IV.6: Domain Structure of p34ct Variants. Full-length p34ct consists of an N-terminal vWA like domain and a C-terminal zinc binding domain (top panel). The shortened variants shown below were designed to improve the crystallization behavior. All constructs were cloned into the pBADM-11 vector and fused to an N-terminal His₆-Tag.

Crystallization of p34ct full-length and preliminary structure determination (IV.5.1 and IV.5.2) later showed that the C-terminus of the protein is likely to interfere with crystal packing, leading to an overall reduced resolution in terms of x-ray diffraction. Hence, a new set of p34ct variants were designed to improve crystallization behavior, all of which were truncated starting from their C-terminal end. (Figure IV.6).

These constructs were based on secondary structure predictions of the C-terminal domain of p34ct using the Phyre² server [139]. While a putatively unstructured extension was removed from the remote C-terminus with p34ct 1-387, the p34ct 1-368 variant stopped right behind a β -strand, which was predicted to be the last element with a defined secondary structure character in the sequence. The latter, however, disrupted the C-terminal C4 zinc binding motif, by leaving out the last C-x-x-C pair. A construct completely devoid of the C-terminal domain was p34ct 1-277, which comprised the N-terminal domain only. Detailed results on the secondary structure prediction including all boundaries of the respective p34ct variants can be found in appendix X.4.

In the following, attempts were made to purify all truncated variants using the same protocols previously established for full-length p34ct (III.3.2). Although expression levels were significantly reduced, samples of both p34ct 1-387 and p34ct 1-277 could be obtained, their yields ranging from 0.2 to 0.7 mg per liter of *E. coli* cell culture. In contrast, the p34ct 1-368 variant eluded purification due to no detectable expression. Here the deliberate disruption of the zinc binding motif presumably led to the C-terminal domain not being properly folded, likely resulting in early degradation within *E. coli*.

While samples of p34ct full-length were generally obtained with high purity, the p34ct 1-277 variant consistently showed a few contaminants, two of which were observed at higher and one at lower molecular weight (Figure IV.7, red arrows).

An even higher degree of impurity was detected with the p34ct 1-387 construct (not shown). In both cases likely the overall reduced protein levels led to a general decrease in purity, with respect to the used Ni-MAC and Ni-TED resin, in an increasingly abundant contaminant background. In support of this, typical large-scale expression trials yielded roughly 30 % more *E. coli* wet cell mass when compared to full-length p34ct, indicative of significantly reduced expression rates with the truncated variants.

Although several attempts were made to increase the purity of p34ct 1-277, only minor improvements could be noted. While an additional IEX step (III.3.2.2) in between AC and SEC helped to remove some of the contaminants, it also further reduced the yields and increased the aggregation tendency of the protein. Nevertheless, p34ct and p34ct 1-277 were sufficiently pure to permit further studies. Both proteins showed distinct peaks during analytical SEC (Figure IV.8) and readily yielded crystals (IV.5.1 and IV.5.3).

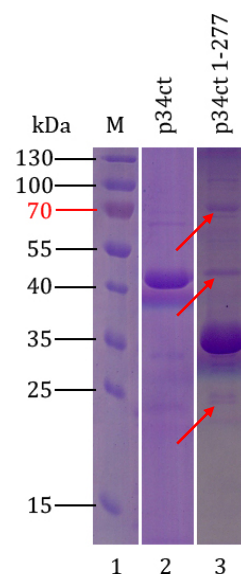


Figure IV.7: Variants of p34ct. Shown are elution fractions after SD 200 SEC, analyzed via SDS-PAGE.

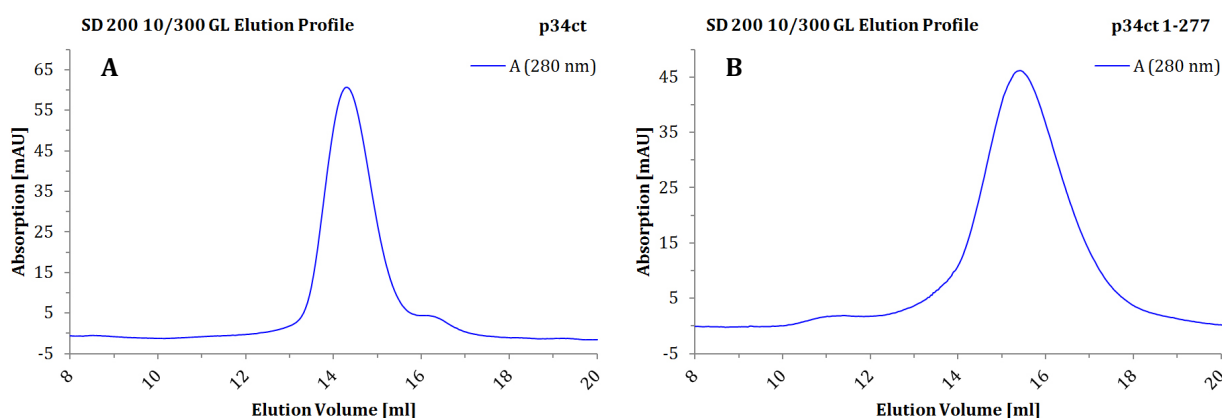


Figure IV.8: SEC Elution Profiles of p34ct Samples. Depicted are elution profiles of p34ct full-length (A) and p34ct 1-277 (B) after analytical size exclusion chromatography using an SD 200 10/300 GL column. Injected concentrations (volumes) were 8 μ M (500 μ l) in (A) and 10 μ M (400 μ l) in (B).

IV.2.2 Purification of p44ct Variants

While p34ct could be purified with relative ease and the samples obtained were homogenous and stable in solution, p44ct turned out to be the more problematic candidate. Though initial trials with full-length p44ct showed good expression levels, the protein revealed a severe tendency to form higher oligomers already during purification, which was performed analogous to p34ct (IV.2.1). Hence, several shortened p44ct variants were designed early on, lacking either putative unstructured regions at the N-terminus of the protein and/or both its C-terminal zinc binding domains (Figure IV.10 and Appendix X.5).

In the following attempts were made to purify all of these p44ct variants, except for p44ct 75-285, which was omitted in favor of the more promising p44ct 101-285 variant. Both the N-terminally truncated proteins as well as the constructs lacking the C-terminal zinc binding domains showed good expression levels when over-expressed in *E. coli*, similar to the full-length protein. However, several different properties became apparent during purification. Here, the two N-terminally truncated variants (p44ct 55-534 and 75-534, Figure IV.6) were prone to degradation, yielding a product of roughly 30 - 32 kDa as analyzed via SDS-PAGE (not shown). This would account for the remaining N-terminal domain of the protein, having a size of 29.4 kDa as in the case of p44ct 55-285 (X.3.2). Neither p44ct 1-285 and p44ct 55-285 nor p44ct full-length showed similar signs of degradation. Hence, fragmentation in p44ct 55-534 and 75-534 likely resulted starting from their C-terminus, leading to digestion of both C-terminal zinc binding domains. One could thus hypothesize that the latter are probably shielded in the full-length protein by the first 55 - 75 amino acids of its N-terminal domain and thereby protected from degradation.

Another interesting finding could be made regarding the prominent oligomerization tendency of p44ct. While full-length p44ct consistently showed several higher oligomers during size exclusion chromatography (Figure IV.11 A), both its N-terminal (p44ct 1-285) and C-terminal C4C4 RING domain (p44ct 368-534) showed distinct and homogenous peaks (Figure IV.11 B and D). Hence, the protein's aptitude to form higher oligomers is likely mediated via its central zinc binding domain, the latter of which might also contribute to the DNA binding capability observed with full-length p44ct (IV.2.4).

In terms of protein yield, the p44ct 1-285 and p44ct 368-534 variants were comparable to full-length p44ct. Usually around 0.8 - 1.3 mg of these proteins could be obtained per liter of *E. coli* cell culture grown. This was surpassed, however, by the p44ct 101-285 construct, which was designed later on and comprised the vWA like region of p44ct's N-terminal half only. Up to 2.3 mg of protein per liter of cell culture could be purified using this variant, which also eluted in a single and highly homogenous peak during preparative SEC (Figure IV.11 C). In general the purity of all these p44ct samples was reasonably good after SEC, as analyzed via SDS-PAGE (Figure IV.9).

Taken together, both p44ct 1-285 and p44ct 101-285 proved to be the most promising candidates with respect to crystallization (IV.4), while the p44ct 368-534 variant, though it did not yield crystals on its own, could be used to generate the p34ct p44ct minimal complex (IV.2.3).

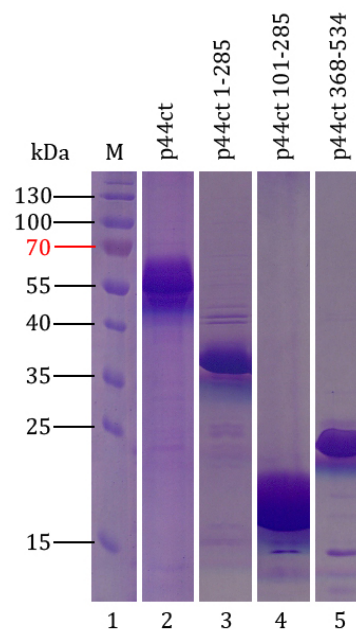


Figure IV.9: Variants of p44ct. Shown are elution fractions of p44ct full length (2) and the best working shortened variants (3 - 5) as analyzed by SDS-PAGE after SD 200 SEC.

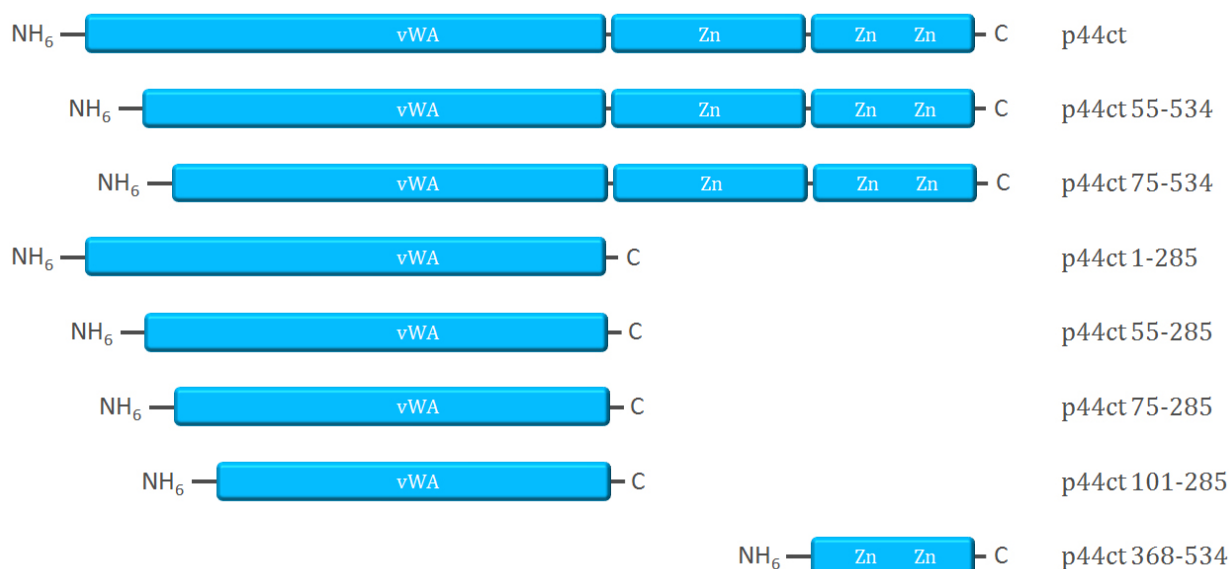


Figure IV.10: Domain Structure of p44ct Variants. Full-length p44ct consists of an N-terminal vWA like domain, a central zinc binding domain and a C-terminal C4C4 RING domain coordinating an additional two zinc ions (top panel). The shortened variants shown below were primarily designed to reduce the oligomerization tendency of p44ct full-length, while p44ct 368-534 was also used for co-expression with p34ct 1-277 to generate the p34ct p44ct minimal complex (see Figure IV.12). All constructs were cloned into the pBADM-11 vector and fused to an N-terminal His₆-Tag, except for p44ct 368-534, which was prepared as several pETM-11 variants, bearing either an N-terminal, C-terminal or no His₆-Tag.

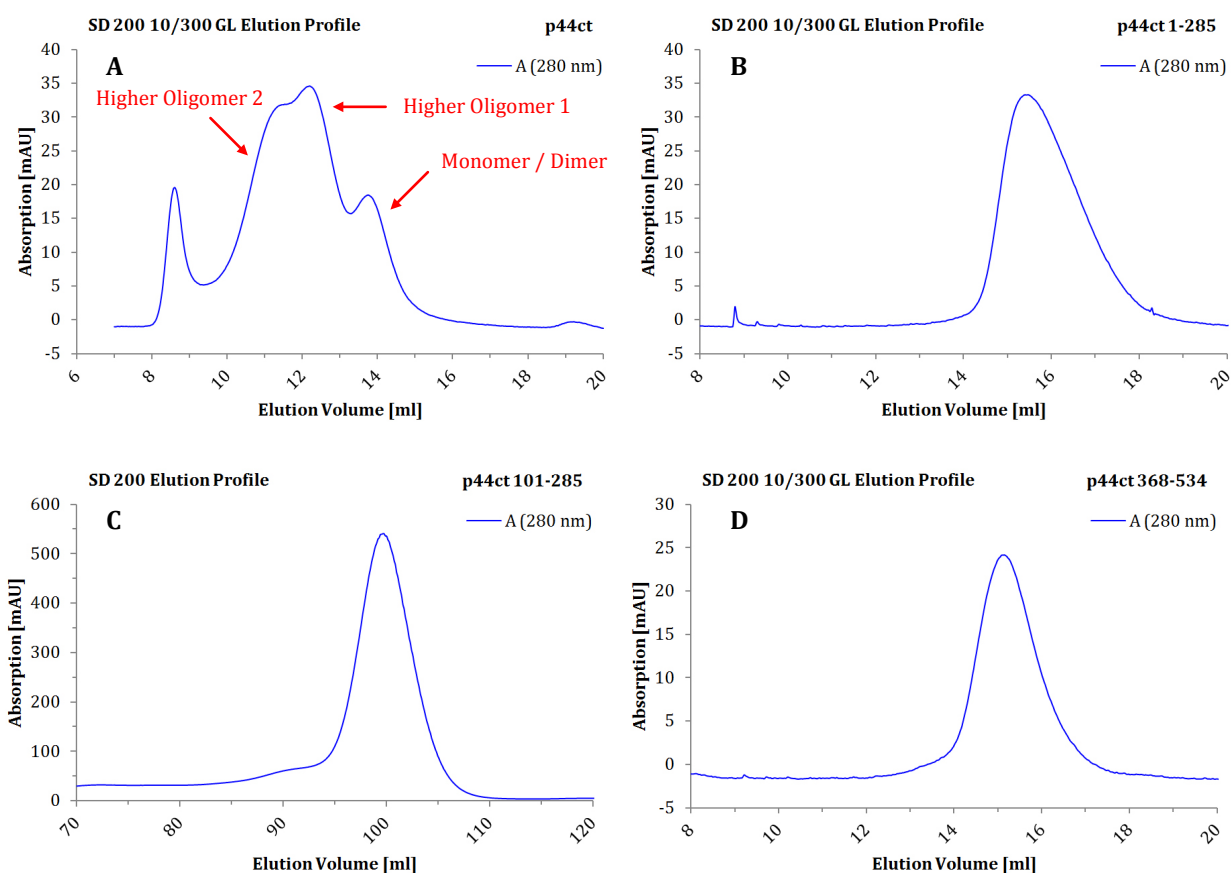


Figure IV.11: SEC Elution Profiles of p44ct Samples. Shown are elution profiles of p44ct full-length (A), p44ct 1-285 (B), p44ct 101-285 (C) and p44ct 368-534 (D) after size exclusion chromatography using an SD 200 10/300 GL (A, B and D) or an SD 200 16/60 prep grade column (C). Injected samples (volumes) were 8 μ M (500 μ l) in (A), 8 μ M (400 μ l) in (B and D) and concentrated Ni-TED eluate (\sim 3 ml) in (C).

IV.2.3 Preparation of the p34ct p44ct Minimal Complex (MC)

In analogy to the human minimal complex, which already behaved well in solution but did not crystallize as expected (IV.1.3), the p34ct p44ct minimal complex (MC) was designed. It consisted of the p34ct N-terminal domain (residues 1-277) and the p44ct C-terminal C4C4 RING domain (residues 368-534), the latter of which was previously characterized via NMR studies using a human p44 construct containing this domain [68,67]. The complex domain boundaries, originally defined by Fribourg et al. [42] for the human proteins, were transferred to the *C. thermophilum* sequence and slightly adjusted to accommodate our structural knowledge of p34ct 1-277 (IV.5.4) and secondary structure prediction results on p44ct (Appendix X.5, Figure IV.12).

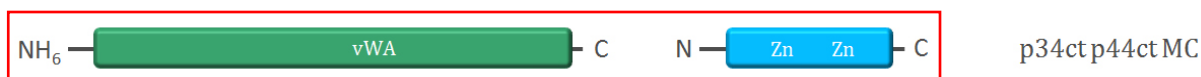


Figure IV.12: Domain Structure of p34ct p44ct MC. The p34ct p44ct minimal complex (boxed in red) was generated by co-expression of both the p34ct 1-277 vWA like domain (green, Figure IV.6) and the p44ct 368-534 C4C4 RING domain (cyan, Figure IV.10). While the p34ct 1-277 construct was fused to an N-terminal His₆-Tag for affinity chromatography, the p44ct C4C4 domain was usually not tagged.

A strong interaction of p34ct and p44ct was already obvious from analytical SEC experiments performed with the full-length proteins. However, the pronounced tendency of p44ct to form higher oligomers (IV.2.2) led to complexes of several oligomeric states, as indicated by the shift of the p44ct elution profile to earlier elution volumes (Figure IV.13 A). Co-expression of both p34ct and p44ct in *E. coli* and subsequent purification also yielded a similar pattern of higher oligomers (not shown). Considering the results obtained with the truncated variants of p44ct (IV.2.2) this again points to the central zinc binding domain of the protein as primary mediator for oligomerization.

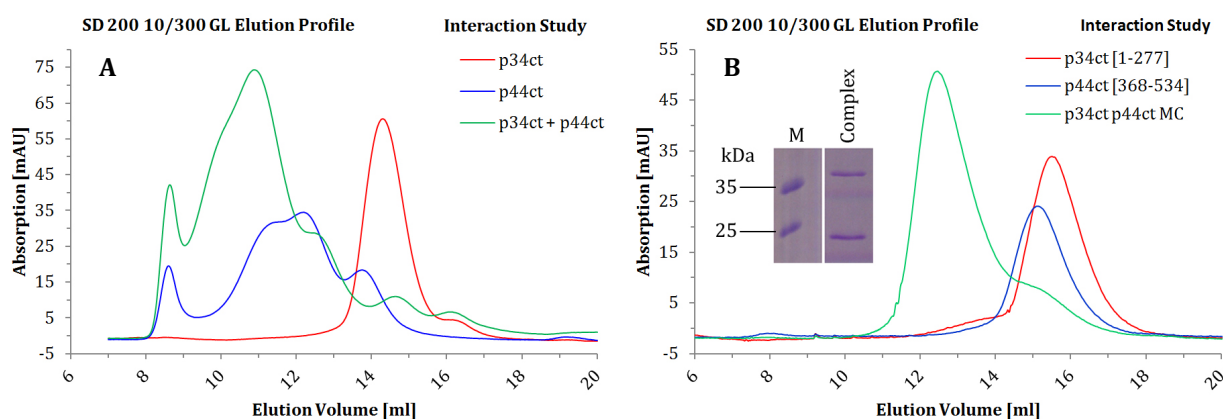


Figure IV.13: Interaction Studies with p34ct and p44ct. The interaction of p34ct and p44ct is shown for the full-length proteins (A) and the p34ct 1-277 and p44ct 368-534 variants (B). The latter were used to generate the p34ct p44ct minimal complex. Complex formation was visualized by analytical SEC using the SD 200 10/300 GL column and verified via SDS-PAGE, as indicated by the inlay in (B).

When p34ct 1-277 and p44ct 368-534 were purified separately and re-combined *in vitro* to generate the minimal complex, a clear complex formation could be observed. Interaction studies via analytical SEC yielded a sharp homogenous peak for the complex, while no additional oligomers were observed (Figure IV.13 B). In order to obtain a stable p34ct p44ct minimal complex of optimal homogeneity for crystallization trials, both p34ct 1-277 and p44ct 368-534 were subsequently either co-expressed in *E. coli* or purified together by prior mixing of *E. coli* cell pellets obtained from separate expression attempts.

The co-purification of p34ct p44ct MC was performed via Ni-TED affinity chromatography (Macherey-Nagel) followed by SEC via an SD 200 16/60 column (GE Healthcare), using buffers Ni-TED A/B (300 mM KCl) and SEC P2, respectively (II.4.1 and III.3.2). SDS-PAGE analysis after size exclusion chromatography indicated a high purity of the sample, except for a few minor contaminants at higher molecular weight (Figure IV.14 A). The latter may be remnants of the impurities already seen for p34ct 1-277 alone (Figure IV.7). As previously observed with the minimal complex reconstituted *in vitro* (Figure IV.13 B), the co-purified p34ct p44ct MC yielded a single and homogenous peak during preparative SEC, though a portion of the sample also eluted in the form of aggregates (Figure IV.14 B). However, the latter could successfully be separated from the bulk of the main complex and did not re-appear during later SEC experiments (not shown).

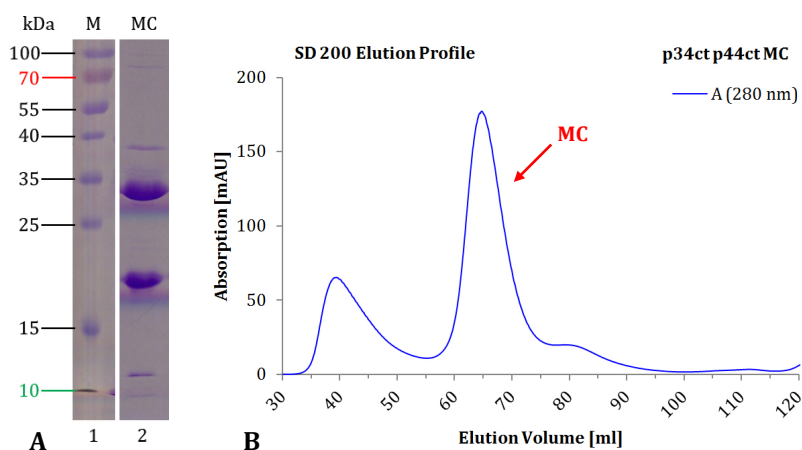


Figure IV.14: The p34ct p44ct Minimal Complex. Shown is the purified p34ct p44ct minimal complex (MC) after SEC as analyzed via SDS-PAGE (A) and its respective SEC elution profile (B). For purification the preparative SD 200 16/60 column was used.

Interestingly, the p34ct p44ct minimal complex could only be co-purified when the p44ct C4C4 RING domain was not tagged. When p44ct 368-534 included the His₆-Tag it seemed to effectively displace p34ct 1-277 from the Ni-TED affinity resin, for unknown reasons. Hence, because of this result and the fact that protein yields were usually limited by the lower p34ct expression levels, the complex was captured by using the His₆-Tag with p34ct 1-277 only. The yields obtained for co-purified p34ct p44ct MC samples were usually in the range of 0.8 - 1.4 mg per liter of *E. coli* cell culture.

IV.2.4 DNA Binding Studies with p34ct and p44ct

With the exception of its two helicases, XPD and XPB, studies towards the DNA binding capability of TFIIH subunits are still very limited. So far, neither p34 nor p44 have been shown to interact with DNA *in vitro*, though in the case of Tfb4 (yeast homologue of p34) nucleic acid binding via a basic Lys/Arg-rich C-terminal region has recently been suggested [88]. Given the DNA oriented role of TFIIH in transcription as well as DNA repair and the fact that both p34 and p44 do contain zinc-finger motifs, which are often involved in binding of nucleic acids, DNA binding by any of these two subunits is not unlikely. Hence, we investigated DNA binding of p34ct and p44ct towards either a single stranded or a double stranded 50-mer DNA substrate in electrophoretic mobility shift assays (EMSAs) using native agarose gel electrophoresis (NAGE, III.3.6). In an attempt to decipher which of their respective domains might confer the DNA binding capability, we also tested several of the truncated p34ct and p44ct variants. The results obtained are summarized in Figure IV.15 and will be discussed on the next page.

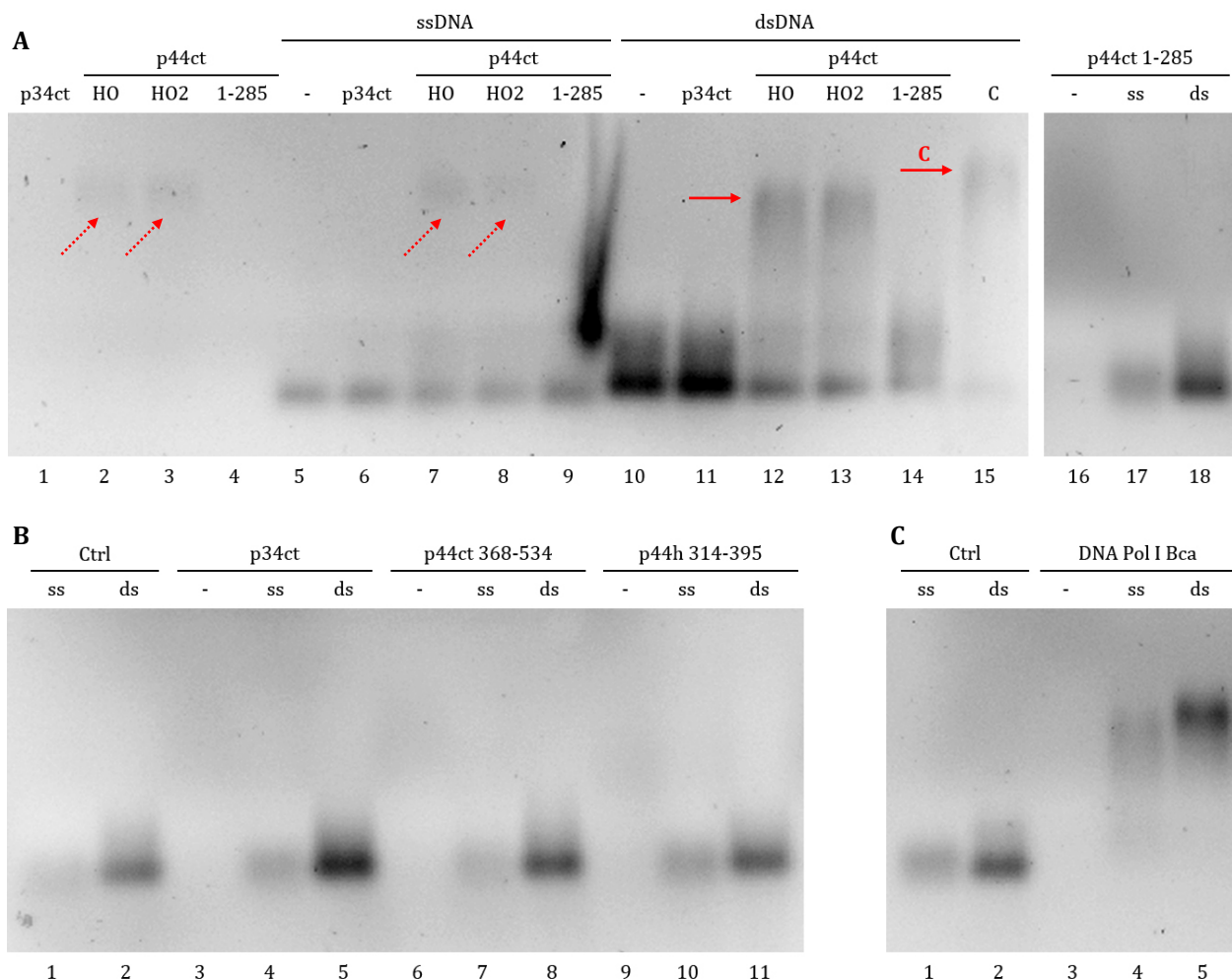


Figure IV.15: DNA Binding Studies with p34ct and p44ct Variants. Shown are EMSAs performed via NAGE using 0.8 % agarose gels after staining of the DNA with Midori Green. In panel (A) full-length p34ct and two different p44ct higher oligomers (HO and HO2) as well as p44ct 1-285 were first analyzed without DNA (lanes 1 - 4) and then against either a 50-mer ssDNA (lanes 5 - 9) or dsDNA (lanes 10 - 14) substrate. Binding of a pre-incubated p34ct p44ct complex to dsDNA is shown in lane 15 (A), while both p44ct 1-285 and p34ct were repeated in lanes 16 - 18 (A) and lanes 3 - 5 (B), respectively. DNA binding by the C-terminal C4C4 RING domain of both *C. thermophilum* and human p44 is assayed in lanes 6 - 11 (B). All DNA bound to proteins is indicated by dotted (intrinsic) or filled (bound in assay) red arrows (see text for details). DNA Polymerase I from *Bacillus caldotenax* was used as positive control for DNA binding (C).

Neither p34ct full-length (Figure IV.15 A, lanes 1, 6 and 11) nor p34ct 1-277 (not shown) were able to bind either single stranded (ss) or double stranded (ds) DNA, while DNA Polymerase I from *Bacillus caldotenax*, which was used as a positive control, bound both ssDNA and dsDNA (Figure IV.15 C). With the results being confirmed in several independent experiments (Figure IV.15 B, lanes 3 - 5) this ruled out the involvement of p34ct's C-terminal zinc finger motif in nucleic acid binding, at least under the conditions tested.

In contrast, with full-length p44ct some bound DNA could already be detected when analyzing the protein in absence of additional DNA substrates (Figure IV.15 A, lanes 2 and 3). This was partly suspected, as purified samples of p44ct usually showed an elevated absorbance at 260 nm, increasing its 260 nm versus 280 nm ratio from a typical range of 0.6 - 0.7 to values of 0.8 - 1.0. While in EMSAs the addition of ssDNA alone did not increase the signal intensity for the intrinsic DNA bound to p44ct (dotted arrows), incubation of full-length p44ct with dsDNA yielded stronger protein-DNA complexes (Figure IV.15 A, lanes 12 and 13, filled arrow). Furthermore, a pre-incubated p34ct p44ct complex retained the DNA binding capability of full-length p44ct (Figure IV.15 A, lane 15, filled arrow C). Interestingly, neither the p44ct N-terminal (1-285) nor its C-terminal C4C4 RING domain (368-534) alone were able to bind DNA (Figure IV.15 A, lanes 16 - 18 and B, lanes 6 - 8). The latter could also be confirmed for the human p44h 314-395 counterpart (Figure IV.15 B, lanes 9 - 11). Though solid proof is lacking, these findings indirectly point to the central zinc binding domain of p44ct as mediator for DNA binding.

Taken together, the EMSAs performed with several p34ct and p44ct variants as well as p44h 314-395 verified dsDNA binding by full-length p44ct, while p34ct did not interact with the DNA substrates presented. In addition, the lack of DNA binding by p44ct 1-285 and its C-terminal C4C4 RING domain implies the central zinc finger motif of p44ct to contact nucleic acids, although the latter interaction might be supported and enhanced by other parts of the protein.

IV.2.5 Thermostability of p34ct and p44ct in Solution

With *C. thermophilum* being a eukaryotic thermophile we investigated the actual thermostability of several p34ct and p44ct variants in solution using the Thermofluor assay (III.4.2). All proteins analyzed were purified as described in IV.2.1 and IV.2.2 using buffer SEC P1 with either 150 mM or 200 mM KCl (II.4.1), which will in the following be referred to as reference buffer.

Both full-length proteins were found to be moderately thermostable over a wide range of buffer conditions, with melting temperatures in the reference buffer usually around 56 °C for p34ct and 50 °C in the case of p44ct (Table IV.1). While the former showed only minor variation in terms of unfolding within pH 7.5 - 9.0, p44ct turned out to be increasingly unstable at higher pH (Figure IV.16 A, B and Table IV.1). In addition, the melting temperature of p44ct was significantly reduced above an ionic strength of 200 mM KCl whereas p34ct retained the same stability from 50 mM KCl up to 500 mM KCl (Figure IV.16 C and D). This prompted several attempts to purify and store p44ct using low salt conditions, which helped to increase the protein stability, but also reduced the overall purity of the sample, ultimately resulting in the crystallization and structure determination of a co-purified contaminant (IV.3).

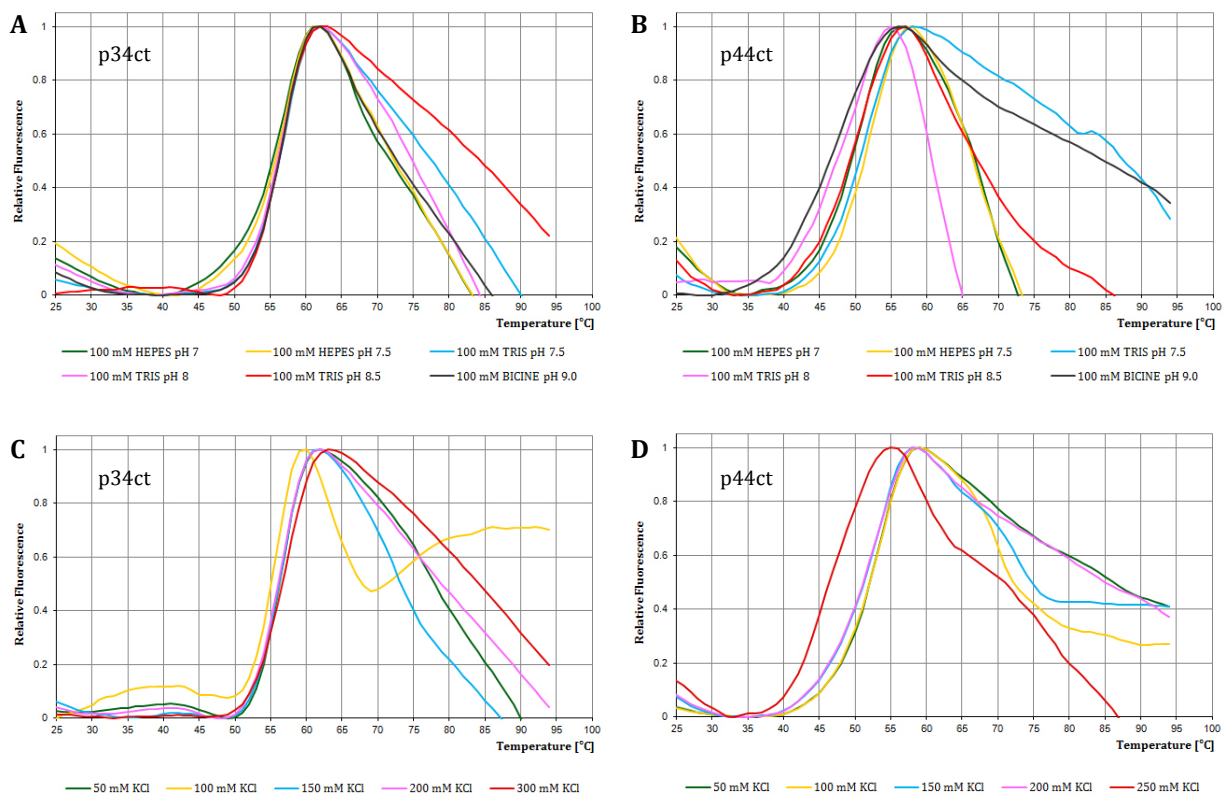


Figure IV.16: Thermofluor Analysis of p34ct and p44ct Full-Length Proteins. Shown are the thermal unfolding curves of p34ct and p44ct full-length in different buffer conditions at a buffer concentration of approximately 100 mM (A, B) and against varying salt concentrations (C, D). The reference buffer used for protein purification and storage consisted of 20 mM Tris-HCl, 200 mM KCl and 1 mM TCEP.

As both p34ct and p44ct are zinc binding proteins, we also examined the effects of divalent cation salts on thermal stability. While both proteins were indifferent towards $MgCl_2$ and $CaCl_2$ up to a tested concentration of 200 mM, using more than 10 μM of $NiCl_2$ or $ZnCl_2$ either precipitated p34ct and p44ct or significantly reduced their overall melting temperatures (data not shown). However, protein stability was not affected when p34ct and p44ct were treated with equimolar amounts of $NiCl_2$ and $ZnCl_2$ (around 1 - 2 μM).

Like full-length p34ct the p34ct 1-277 variant retained a melting temperature of 56 °C in the reference buffer. Interestingly, however, the unfolding curves here often showed a biphasic character with a much weaker early transition starting below 25 °C followed by a second more prominent unfolding signal at the expected onset temperature of 45 - 50 °C (Figure IV.17 A). The first transition was notably reduced in high pH buffers like CHES or CAPS with CHES at pH 9.5 yielding the best results in terms of the overall unfolding profile, albeit at a somewhat reduced melting temperature of 51 °C (Figure IV.17 A and Table IV.1).

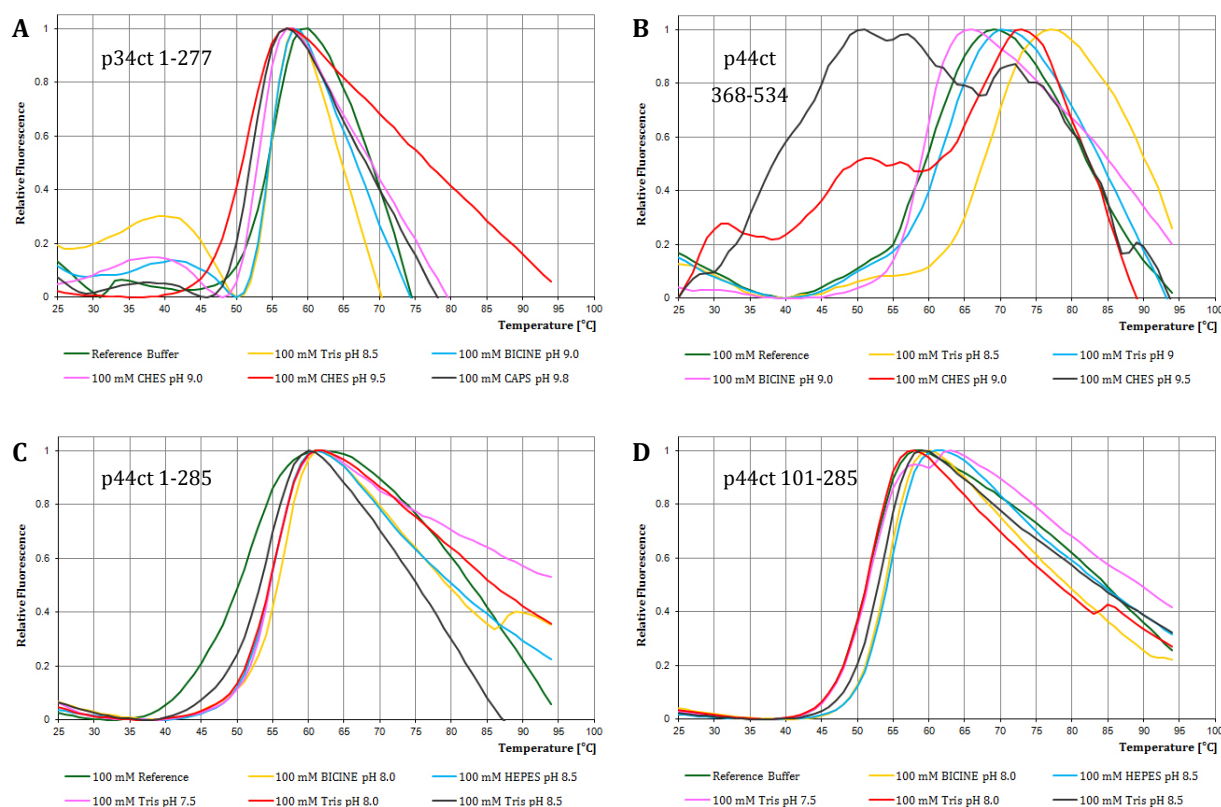
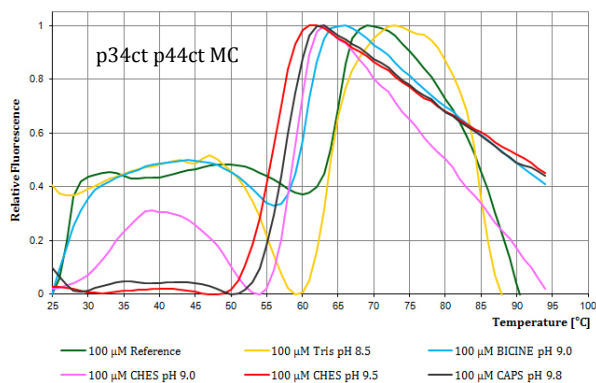


Figure IV.17: Thermofluor Analysis of Shortened p34ct and p44ct Variants. Shown are the thermal unfolding curves of p34ct 1-277 (A), p44ct 368-534 (B), p44ct 1-285 (C) and p44ct 101-285 (D) in different buffer conditions at a buffer concentration of approximately 100 mM. The reference buffer used for protein purification and storage consisted of 20 mM Tris-HCl, 150 mM KCl and 1 mM TCEP.

While CHES at pH 9.5 was considered the best alternative for p34ct 1-277, unfolding curves using either CHES or CAPS based high pH buffer systems could not be determined for p44ct 368-534 (Figure IV.17 B). The exact reason for this is unclear as both buffers did not visibly precipitate the protein and p44ct 368-534 could also be purified in a buffer including 20 mM CHES pH 9.5 for ITC studies (IV.7.5). Presumably the high pH environment resulted in a more open conformation of the very small C4C4 RING domain, already allowing for saturation with SYPRO® Orange prior to the experiment. No additional dye binding upon thermal unfolding thus might prevent the detection of a proper unfolding signal. In general p44ct 368-534 showed a very high thermostability of 61 °C in the reference buffer and 70 °C in Tris pH 8.5 (Figure IV.17 B and Table IV.1).

Apart from the highly stable C-terminal C4C4 RING domain also the N-terminal variants p44ct 1-285 and p44ct 101-285 revealed an improved stability compared to full-length p44ct, though the increase is not as prominent. Promising results for both proteins were obtained with several buffers at pH 8.0 - 8.5, while CHES pH 9.5 only slightly decreased the reference melting temperature of 53 °C (Figure IV.17 C, D and Table IV.1).

**Table IV.1:** Thermofluor Melting Temperatures.

Protein	Reference	Tris pH 8.5	CHES pH 9.5
p34ct	56 °C	56 °C	55 °C
p34ct 1-277	56 °C	56 °C	51 °C
p44ct	50 °C	50 °C	41 °C
p44ct 1-285	53 °C	55 °C	52 °C
p44ct 101-285	53 °C	54 °C	52 °C
p44ct 368-534	61 °C	70 °C	n/d
p34ct p44ct MC	65 °C	64 °C	56 °C

Figure IV.18: Thermofluor Analysis of p34ct p44ct MC. Depicted are thermal unfolding curves in different buffer conditions. The reference buffer contained 20 mM Tris-HCl, 150 mM KCl and 1 mM TCEP.

In addition to the single protein samples, Thermofluor melting curves were also recorded for the p34ct p44ct minimal complex consisting of p34ct 1-277 and p44ct 368-534 (Figure IV.18). Similarly to p34ct 1-277 alone (Figure IV.17 A) a biphasic unfolding profile was observed for p34ct p44ct MC with an early onset below 25 °C for most buffer conditions, except for the high pH buffers CHES pH 9.5 and CAPS pH 9.8. The latter yielded the best profile with respect to thermal unfolding, albeit with an overall melting temperature almost 10 °C below the best values observed for the second transition in other buffers (Figure IV.18 and Table IV.1). However, the single unfolding event in CHES pH 9.5 still revealed an increase of 5 °C compared to the melting temperature of p34ct 1-277 alone, indicative of the stabilizing effect introduced by p44ct 368-534. Hence, as CHES at pH 9.5 allowed for a late onset of unfolding and generally conferred moderate to good stability, this buffer system was used for most experiments with p34ct 1-277 and p34ct p44ct MC.

IV.2.6 Oligomeric State of p34ct and p44ct in Solution

Although many studies, including Cryo-EM experiments [86,87], suggest the 7-subunit core of the general transcription factor TFIID to consist of single subunits, the true oligomeric state of each subunit within TFIID has not been resolved. Recently, Kainov *et al.* reported dimeric forms of both Tfb2 (p52) and Tfb5 (p8) that undergo a dimer to hetero-tetramer transition upon interaction with each other [17,88]. The authors also proposed that the N-terminal domain of Tfb2 is responsible for homo-dimerization, while each C-terminal domain interacts with one Tfb5 molecule. As many proteins are known to self-associate in the absence of their respective binding partners [182], we analyzed the molar mass and hence the oligomeric state of several p34ct and p44ct variants in solution, including the p34ct p44ct minimal complex.

Initially, analytical SEC experiments (A-SEC, III.3.5) were used to evaluate the composition of purified p34ct and p44ct samples. However, due to the elution volumes inherently dependent on the hydrodynamic volume of separated particles rather than their actual molecular weight, this only allowed for semi-quantitative results with low accuracy. To achieve a high precision molar mass determination, analytical size exclusion chromatography was coupled to an on-line multi angle light scattering (MALS) and differential refractive index (dRI) measurement (SEC-MALS, III.4.4). This not only allowed to detect the usual UV signal, but also to compare it to the MALS and dRI readout as well as the molar mass distribution calculated from the data.

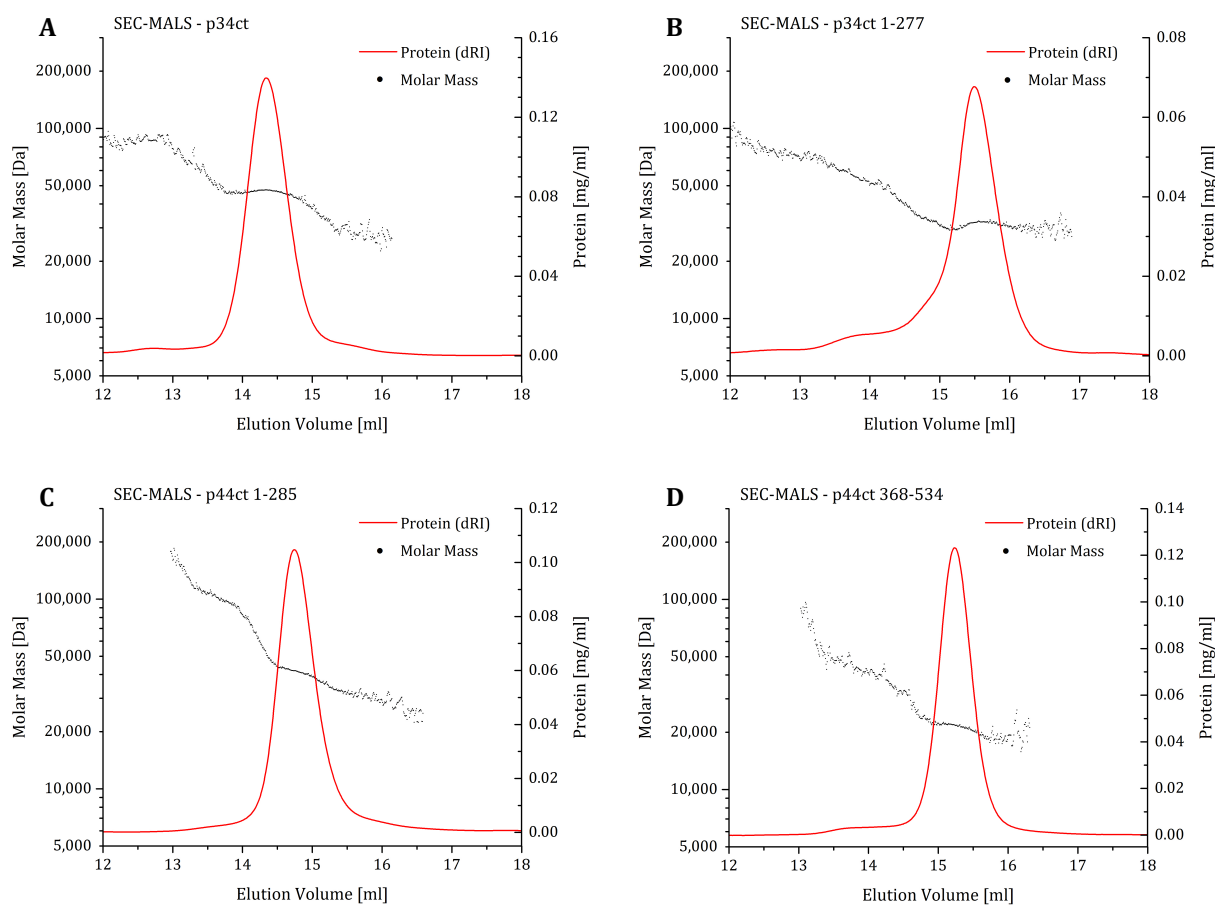


Figure IV.19: Molar Mass Analysis of p34ct and p44ct Variants. Depicted are SEC-MALS profiles of p34ct (A), p34ct 1-277 (B), p44ct 1-285 (C) and p44ct 368-534 (D). The differential refractive index (dRI) and multi angle light scattering (MALS) detection was coupled to size exclusion chromatography (SEC) via a SD 200 10/300 GL column. Sample concentrations are given in mg/ml and shown in red as a function of the differential refractive index (dRI). The calculated molar mass is indicated as a scatter plot, with one data point per measurement and second.

Table IV.2: Results from Molar Mass Analysis of p34ct and p44ct Variants.

Protein	Buffer	Expected MW [Da]	Measured MW [Da]	Oligomeric State
p34ct	CHES pH 9.5	46,767	46,890	Monomer
p34ct 1-277	CHES pH 9.5	31,941	32,020	Monomer
p44ct 1-285	TRIS pH 8.0	34,974	41,600	Monomer
p44ct 368-534	CHES pH 9.5	21,085	21,650	Monomer
p34ct p44ct MC (mixed)	CHES pH 9.5	53,026	96,860	Hetero-Tetramer
p34ct p44ct MC (co-purified)	CHES pH 9.5	49,899	101,700	Hetero-Tetramer

Combined SEC-MALS measurements with both full-length p34ct and p34ct 1-277 unambiguously showed that these proteins are monomers in solution, which was not expected from standard A-SEC experiments (Figure IV.19 A and B). Similarly both p44ct 1-285 and p44ct 368-534 yielded sharp and distinct peaks with a calculated mass distribution indicative of monomers (Figure IV.19 C and D). The molar masses determined for all of these proteins were markedly close to the expected values, except for p44ct 1-285, which was observed to be 19 % larger compared to the calculated molar mass of its monomer (Table IV.2). With respect to p44ct 1-285 and p44ct 368-534 the results obtained again point to the central zinc binding domain of p44ct as mediator for the oligomerization of the full-length protein (compare IV.2.2 and IV.2.3), as both its N-terminal and C-terminal domains do not dimerize on their own under the conditions tested.

Interestingly, while p34ct 1-277 and p44ct 368-534 are monomers in solution, their complex seems to assume a higher oligomeric assembly. Both the minimal complex reconstituted *in vitro* by mixing of single protein samples (Figure IV.20 A) and p34ct p44ct MC generated by co-purification (Figure IV.20 B) reveal a molar mass distribution indicative of a tetramer (Table IV.2). While the mixed complex sample is slightly off, presumably due to the lower concentrations used and the complex formation not being complete, the determined molar mass for the co-purified complex is only 1.9 % above the expected value for a hetero-tetramer consisting of two p34ct p44ct MC hetero-dimers. This is consistent with observations made in the p34ct p44ct MC crystal structure and will be further discussed in the respective results section (IV.6.4).

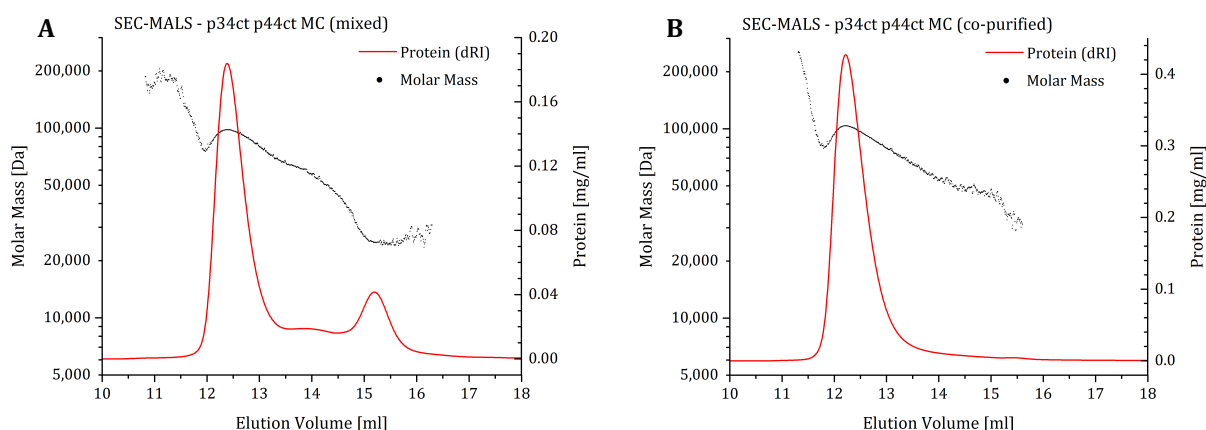


Figure IV.20: Molar Mass Analysis of p34ct p44ct MC. Shown are SEC-MALS profiles of p34ct p44ct MC reconstituted *in vitro* by mixing of p34ct 1-277 and p44ct 368-534 samples (A) and p34ct p44ct MC (B) obtained by co-purification. The differential refractive index (dRI) and multi angle light scattering (MALS) detection was coupled to size exclusion chromatography (SEC) via a SD 200 10/300 GL column. Sample concentrations are given in mg/ml and shown in red as a function of the differential refractive index (dRI). The calculated molar mass is indicated as a scatter plot, with one data point per measurement and second.

IV.3 Structure Determination of DPS - A Successful Failure

Along with p34ct also p44ct was subjected to crystallization trials, covering more than 850 unique conditions. While the very first purified batches of p44ct, which contained higher salt concentrations, did not yield any crystals, tiny crystals could be observed in a few conditions using p44ct purified with 150 mM imidazole for Ni-TED elution and only 150 mM KCl in the purification and storage buffers. During follow-up experiments, a lot of effort was put into optimizing these crystals and grow them to sizes suitable for x-ray diffraction analysis, which turned out to be more difficult than expected.

Interestingly, although still very small (longest axis < 50 μm), the crystals obtained diffracted to 2.5 \AA using the home source x-ray generator. In the following native diffraction data sets of up to 2.0 \AA could be collected at the electron synchrotron BESSY in Berlin with crystals grown in 30 % PEG 400, 100 mM Tris-HCl pH 8.0 and 200 mM sodium citrate pH 5.6 (Figure IV.21 A). We also performed an x-ray energy scan around the zinc absorption edge and obtained a strong signal for the presence of zinc in the crystals tested. This led us to believe that the crystals were composed of p44ct, which is known to contain a total of 3 zinc binding sites.

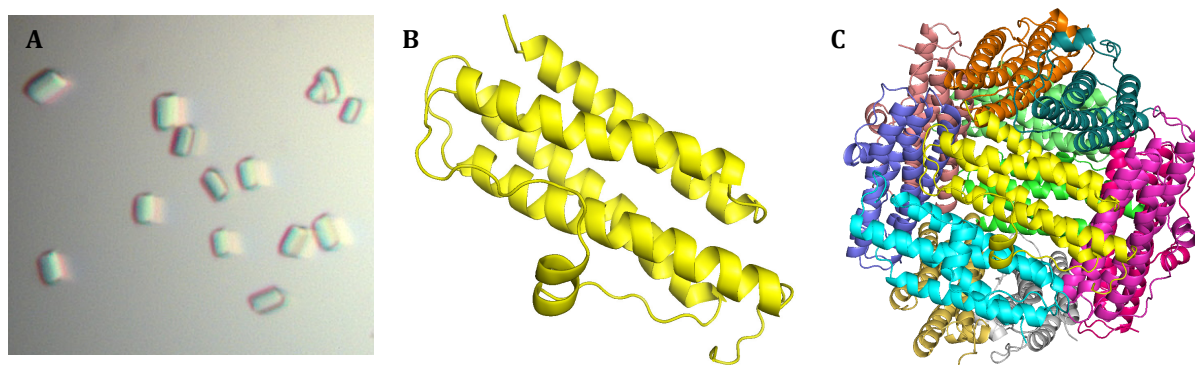


Figure IV.21: Crystals and overall Structure of DPS. Crystals of DPS were grown in 30 % (v/v) PEG 400, 100 mM Tris pH 8.0 and 200 mM sodium citrate pH 5.6 (A). The DPS monomer consists of a single domain predominantly formed by 4 long α -helices (B). In solution a total of 12 monomers accumulate to form a spherical dodecamer assembly with both metal ion as well as DNA binding properties (C).

In an attempt to solve the structure of the putative p44ct crystals we prepared a seleno-methionine (Se-Met) derivative of the protein, which ultimately yielded even bigger crystals using the previously identified crystallization conditions. Yet, what we also observed was a very prominent impurity in the Se-Met sample during purification, with an apparent molecular weight of approximately 17 kDa, as visualized by SDS-PAGE. This contamination band was almost as strong as the actual p44ct band on the gel (Figure IV.22). The same impurity could also be identified with the native p44ct samples previously purified. However, here p44ct was present in abundance, whereas the 17 kDa contaminant was only barely visible (Figure IV.22).

Using crystals grown from a p44ct Se-Met sample, the structure could be solved via a single anomalous diffraction experiment utilizing the selenium atoms to localize the heavy atom substructure and solve the phase problem. The derivative crystal's space group was $P2_12_12_1$ with unit cell parameters of $a = 100.5 \text{ \AA}$, $b = 105.1 \text{ \AA}$ and $c = 202.0 \text{ \AA}$ and angles of $\alpha = \beta = \gamma = 90^\circ$. One of the data sets collected at beamline ID 23.1 of the ESRF in Grenoble could be processed to roughly 2.3 \AA and revealed a ferritin-like protein, which could later be identified as DPS, a "DNA binding protein from starved cells" of *E. coli*. The DPS monomer shows a typical ferritin-like fold consisting of 4 relatively long α -helices, which accumulate to form a spherical dodecameric assembly in solution (Figure IV.21 B and C).

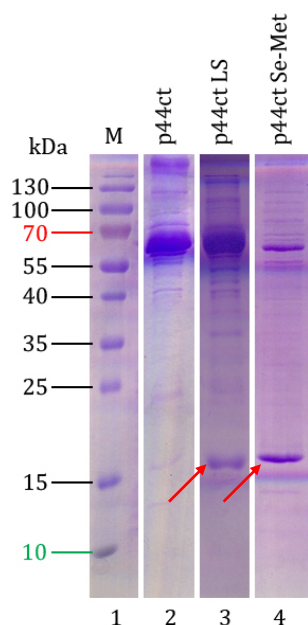


Figure IV.22: Samples of p44ct obtained from different Purification Strategies. Shown are elution fractions of p44ct full-length as analyzed by SDS-PAGE after SD 200 SEC. The protein was purified using either high salt (2), low salt (3) or as Se-Met derivative (4). The DPS impurity is indicated by red arrows.

The DPS structure could be solved within a time frame of only 8 weeks from purification to *de novo* electron density map determination. Unfortunately the protein, which inherits both metal ion as well as DNA binding properties, had already been structurally characterized almost 15 years ago [183]. However, this more or less accidentally purified and crystallized impurity nicely demonstrates how fast and efficient macro-molecular crystallography can be, while at the same time being extremely difficult with respect to obtaining the structure of the protein of interest. Hence, the structure determination of DPS was termed a successful failure as it provided for a good proof of principle, but yet did not yield any new and exciting insights.

IV.4 Crystallization of p44ct Variants - Getting Closer

While multiple crystallization attempts with full-length p44ct only culminated in the structure determination of DPS (IV.3), similar trials with different p44ct shortened variants turned out to be more promising. Here, a degradation product of both p44ct 55-534 and p44ct 75-534 yielded tiny needle like crystals in a few ethylene glycol or alcohol based conditions. Similarly shaped crystals were observed with p44ct 1-285 using either several PEGs or ammonium phosphate as precipitants. However, neither of these conditions yielded crystals suitable for x-ray diffraction experiments and optimization attempts were unsuccessful.

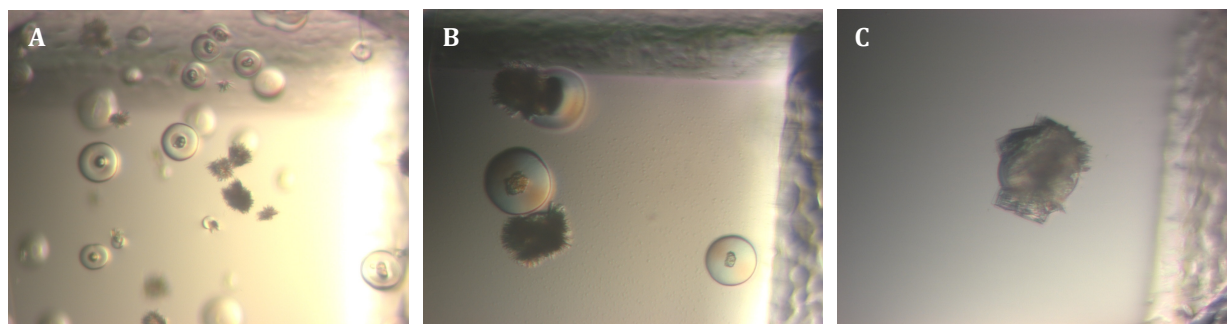


Figure IV.23: Crystals of the p44ct 101-285 vWA Domain. Crystalline aggregates and needle clusters were grown from p44ct 101-285 samples in 3 M NaCl and 100 mM HEPES pH 7.5 (A) or Tris pH 8.5 (B), while single platelet clusters were obtained with 1.5 M NaOAc and 100 mM HEPES pH 7.0 (C).

Upon further confinement of construct borders crystalline aggregates as well as needle or platelet clusters could be grown with p44ct 101-285 (Figure IV.23). The latter variant only comprised the already annotated *von Willebrand Factor A* (vWA) like domain of the protein, which is predicted to be similar in secondary structure to the p34ct N-terminal domain (compare X.4 and X.5). While the crystallites and needle clusters grown in 3 M NaCl (Figure IV.23 A and B) were nicely reproducible and also partly responded to optimization attempts, the single platelet clusters obtained from either a 1.5 M sodium acetate (Figure IV.23 C) or a 2.0 M sodium formate condition (not shown) eluded reproduction.

Preliminary x-ray diffraction data collected at the ESRF BM 14 U beamline with crystals grown from p44ct 101-285 samples confirmed these to contain protein. Though poor in overall crystal quality, two of the non-reproducible platelet clusters even showed single diffraction spots up to a resolution of 3.5 Å. However, although this seemed very promising, attempts to optimize the crystallization conditions to obtain bigger and higher quality crystals by using micro- and macro-seeding only led to minor improvements. Eventually the p44ct crystallization project was set aside in favor of full-length p34ct and p34ct 1-277, which yielded highly reproducible single crystals of a quality more suitable for structure determination (IV.5).

IV.5 Crystal Structure of p34ct - A Novel vWA like Domain

Although TFIIH has been studied extensively in the past, high resolution structural knowledge on most TFIIH subunits is still limited due to the difficulty to obtain abundant and pure protein samples suitable for NMR or crystallographic analysis. The following sections describe the first crystallization and structure determination of the p34 subunit of TFIIH from *C. thermophilum* (ct). The structure was solved using crystals grown from full-length p34ct as well as p34ct 1-277 and comprises the N-terminal domain of the protein, which is known to strongly interact with the C-terminal C4C4 RING domain of p44 [21,22].

IV.5.1 Crystallization of p34ct

Initial crystallization trials with full-length p34ct yielded several conditions supporting crystal growth, most of which were based on polyethylene glycol (PEG) as precipitant. Upon optimization rather big crystals could be grown repeatedly using 0.5 - 2.0 μ l of protein in 20 mM Tris-HCl pH 8.0, 150 mM KCl and 1 mM TCEP, at a concentration of 10 - 12 mg/ml, which was added to 0.5 μ l of reservoir solution. The latter was composed of 100 mM Tris pH 8.0 and varying amounts of 18 - 32 % PEG 550 MME, depending on the volume of protein solution per volume of reservoir in the mixture.

Crystals of p34ct usually grew within 3 days at 20 °C to sizes of more than 50 x 200 x 200 μ m (Figure IV.24 B). Interestingly, however, all crystals were either attached to a solid surface or a skin of denatured protein on top of the crystallization drop, which both seemed to serve as scaffold for nucleation. Furthermore, the p34ct crystals always revealed a rubber-like phenotype, rendering them rather elastic than fragile. This effect became more pronounced with prolonged crystal growth and incubation, up to a point where the crystals assumed a virtually indestructible cohesiveness. The reason for this, though, is not clear. While oxidative modifications can be ruled out due to the presence of either DTT or TCEP as reducing agents, flexible protein parts, which could be unevenly distributed within the crystal, might confer additional stability apart from the usual protein contacts forming the crystal lattice. The presence of a slightly smaller putative degradation product might be an additional factor (Figure IV.24 A, red arrow). Consistent with these shortcomings the p34ct crystals yielded x-ray diffraction patterns up to a maximum resolution of 3.8 Å only (Figure IV.24 C).

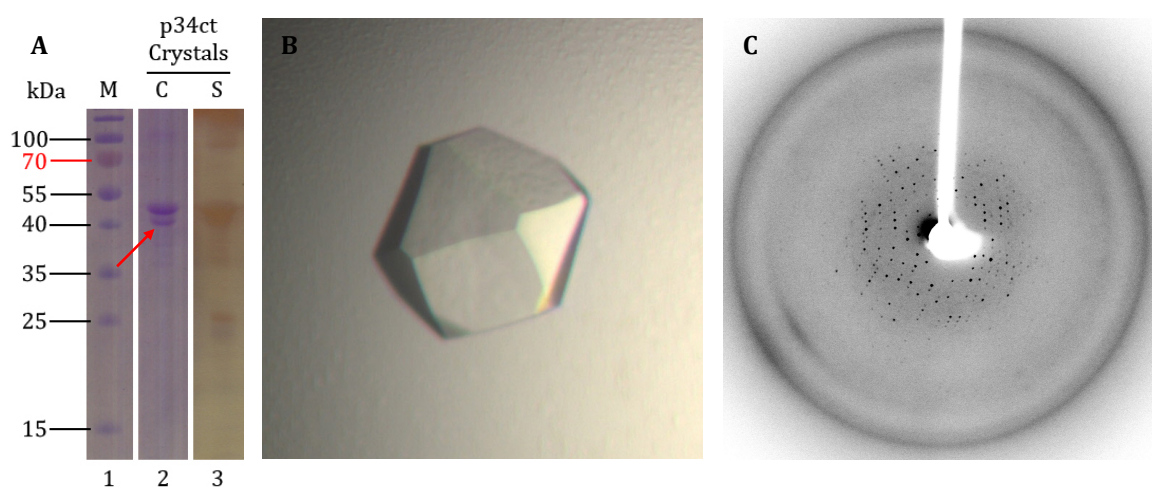


Figure IV.24: Analysis of p34ct Crystals and X-Ray Diffraction. Panel (A) shows crystals grown from p34ct samples as analyzed by SDS-PAGE using either Coomassie (lane 2) or silver staining (lane 3). Apart from full-length p34ct at least one additional band was detected at lower molecular weight (A, red arrow). Typical p34ct crystals were of cuboctahedral shape and grew in roughly 20 % PEG 550 MME and 100 mM Tris-HCl pH 8.0 (B). Strong x-ray diffraction patterns were observed up to a resolution of 3.8 Å (C).

IV.5.2 Structure Determination of p34ct

Typical crystals of p34ct grew in the highly symmetric space group $F4_132$ with cubic unit cell parameters of $a = b = c = 257.3 \text{ \AA}$ and all angles $\alpha = \beta = \gamma = 90^\circ$. The crystals usually diffracted strongly up to a moderate resolution of $5 - 6 \text{ \AA}$, regardless of their age and size. Upon extensive optimization using various techniques, like additive and cryo solution screening as well as micro- and macro-seeding trials, significantly bigger crystals could be grown, yielding more intense reflections. However, the overall resolution limit largely remained the same, presumably due to the inherent crystal flaws discussed in IV.5.1. The exceptions were a few huge single crystals, which reached the 4 \AA border in terms of diffraction quality. This prompted us to continue our efforts and attempt to solve the p34ct structure.

Initially we tried to use the intrinsic zinc signal of the protein's C-terminal zinc binding domain to solve the phase problem, collecting several $4 - 5 \text{ \AA}$ data sets at the zinc absorption edge. However, this approach failed due to only weak anomalous signal intensities. Attempts to purify and crystallize a seleno-methionine derivative of p34ct were equally unsuccessful, as the heavy metal incorporation seemed to drastically lower the expression level of the protein, resulting in highly impure samples. We finally succeeded in achieving a *de novo* solution using derivative data of a potassium iodide (KI) soaked crystal at a resolution of 4.2 \AA . Combining the p34ct KI data with a highly isomorphous p34ct native data set at 3.8 \AA in a SIRAS approach using the SHELX C/D/E programs [130] eventually allowed to determine the p34ct crystal structure.

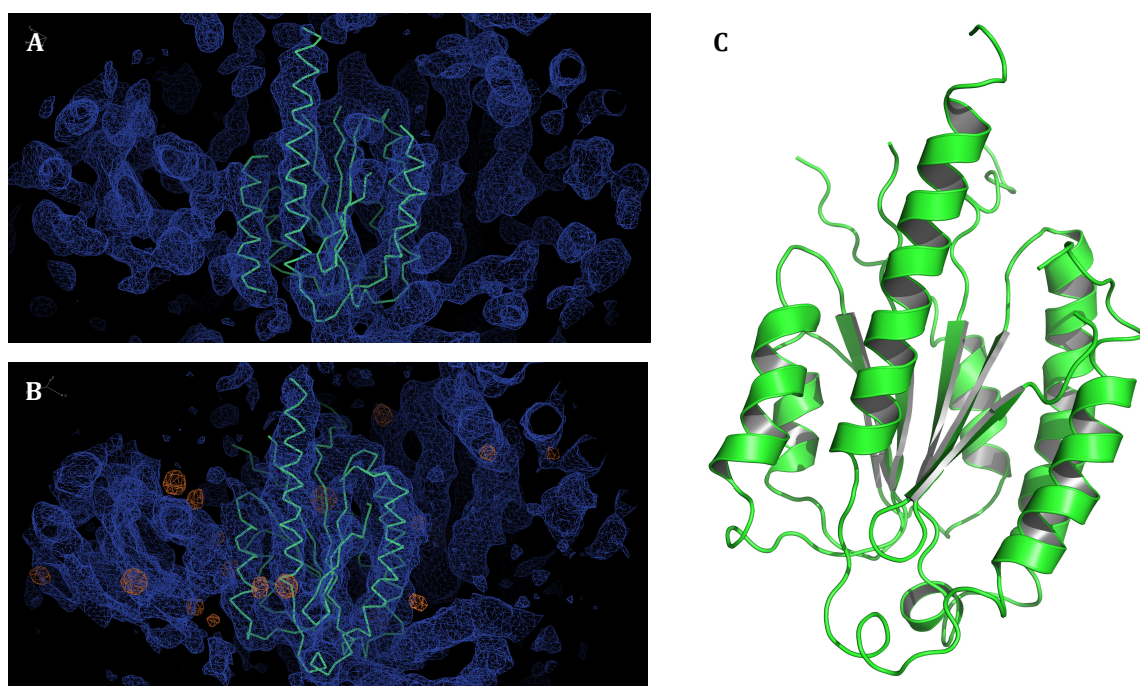


Figure IV.25: Electron Density of p34ct Crystals and Preliminary Model. Initial electron density maps obtained via SHELX C/D/E and SIRAS phasing (A) were used to build a preliminary model of the p34ct N-terminal domain (B and C). The model could subsequently be refined to a resolution of 3.8 \AA using Phenix [127], reaching R-factors of $R_{\text{work}} = 0.3824$ and $R_{\text{free}} = 0.3998$. While the $2\text{Fo}-\text{Fc}$ maps are colored in blue (A, B) the anomalous density of the iodide ions used for phasing is shown in orange (B).

The initial low resolution electron density map obtained revealed approximately half of the protein, with one molecule in the asymmetric unit (Figure IV.25 A). Based on secondary structure prediction results using the Phyre² algorithm [139], it was clear that p34ct contains an N-terminal domain and an additional C-terminal domain including a C4 zinc-finger motif, the latter of which could not be located in the electron density map.

Analysis of the p34ct crystals via SDS-PAGE (Figure IV.24 A) and mass spectrometry (data not shown), however, indicated that the crystals are composed predominantly of the full-length protein. This led us to conclude, that the p34ct C-terminal domain is disordered within the crystal lattice and likely not involved in any crystal contacts, a finding which might explain some of the obvious flaws observed with all the p34ct crystals in the first place (IV.5.1). Using C α -traces of structural homologues, suggested by the Phyre² analysis, we were able to manually build a preliminary model of the p34ct N-terminal domain using the programs O and COOT [123,125]. The model was subsequently refined to a resolution of 3.8 Å using Phenix [127], resulting in R-factors of R_{work} = 0.3824 and R_{free} = 0.3998 (Figure IV.25 B and C). More detailed information on data collection and refinement statistics can be found in Appendix X.6.

While the overall quality of the preliminary model was still rather poor due to the limited resolution and low phasing power, the p34ct preliminary structure marked a decisive breakthrough and provided the basis for solving the p34ct 1-277 structure later on at higher resolution (IV.5.4).

IV.5.3 Crystallization of p34ct 1-277

With the diffraction power of the initial p34ct crystals stuck at a resolution of 3.8 Å, attempts were made to increase crystal quality, allowing to complete the hitherto preliminary model at a higher resolution. Based on the knowledge gained through the p34ct preliminary structure, we assumed that the intrinsic flexibility of the C-terminal domain of the protein, which was completely disordered in the p34ct full-length crystals, adversely affected crystal packing. Hence the p34ct 1-277 variant was designed, comprising the N-terminal domain only and thus removing the flexible and unstructured part of the protein.

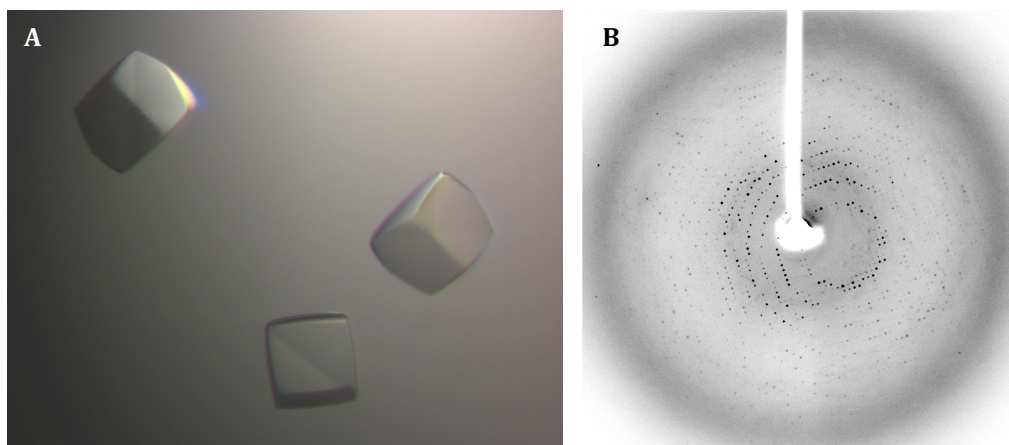


Figure IV.26: Crystals of p34ct 1-277 and X-Ray Diffraction. Typical crystals of p34ct 1-277 were of prism-like shape and grew directly from the protein solution over a reservoir of protein buffer including additionally 50 - 500 mM NaCl (A). Strong x-ray diffraction was observed up to a resolution of 2.7 Å (B).

During purification of p34ct 1-277 it became immediately obvious that lacking the C-terminal zinc domain rendered the N-terminal domain much more insoluble compared to the full-length protein. The yields were significantly reduced, which in turn adversely affected the overall purity of the samples obtained (IV.2.1). In addition, the protein showed an increased tendency to precipitate from solution. Hence, when subjecting p34ct 1-277 to various crystallization trials under standardized as well as known conditions previously optimized for full-length p34ct, more than 80 % of the drops contained precipitated protein.

To alleviate the solubility issue, attempts were made to crystallize p34ct 1-277 directly from the protein solution, without the need for additional precipitants. This approach was successful and crystals were readily obtained, assuming the same cubic space group $F4_132$ with almost exactly the same unit cell dimensions of $a = b = c = 257.1 \text{ \AA}$ and all angles $\alpha = \beta = \gamma = 90^\circ$. Upon optimization, relatively large single crystals were grown using 2.0 - 4.0 μl of protein in 20 mM Tris-HCl pH 8.0, 150 mM KCl and 1 mM TCEP at a concentration of 10 - 12 mg/ml. In order to control the speed of crystallization, the solution was placed over a reservoir containing the protein buffer, including an additional 50 - 500 mM NaCl to establish an osmotic gradient of variable strength. The prism-like cubic crystals of p34ct 1-277 then usually appeared over night and reached sizes of 100 - 200 μm in all three dimensions within 3 - 5 days (Figure IV.26 A).

For data collection the p34ct 1-277 crystals were washed in 1.0 μl protein buffer, soaked by first sequentially adding 2 times 1.0 μl and then transferring them to 1.0 μl of the same cryo solution containing a combination of 10 mM Tris-HCl pH 8.0, 75 mM KCl, 12.5 % di-ethylene-glycol, 6.25 % ethylene-glycol, 6.25 % MPD, 6.25 % 1,2-propanediol and 6.25 % glycerol. This complex procedure allowed to further boost the diffraction quality of flash frozen p34ct 1-277 crystals, which otherwise were limited to a resolution of 3.5 \AA (Figure IV.26 B). In the following a highly redundant data set was collected at beamline ID 29 at the ESRF in Grenoble, which could be processed to a maximum resolution of 2.7 \AA .

IV.5.4 Structure Determination of p34ct 1-277 - The p34ct vWA like Domain

Upon collection of higher resolution data, the structure of the p34ct 1-277 N-terminal domain could be solved by molecular replacement using Phaser [126] with the truncated p34ct preliminary model previously built serving as template for phasing.

Similarly to full-length p34ct roughly 17 amino acids at the N-terminus of p34ct 1-277, in addition to the hexa-histidine tag and 2 rather long linker regions (residues 86 to 105 and 169 to 197), were not accounted for in the electron density map (Figure IV.27). However, the overall data quality improved significantly compared to the preliminary structure and allowed us to rebuild the p34ct N-terminal domain with much higher accuracy. Besides refitting the $\text{C}\alpha$ -trace backbone and all secondary structure elements, most amino acid side chains now showed distinct density and could be properly modeled.

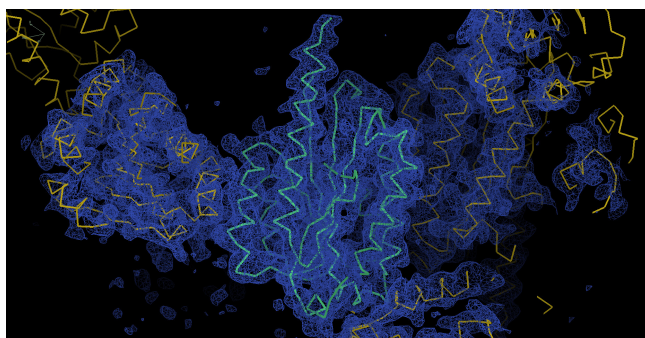


Figure IV.27: Electron Density of p34ct 1-277. Shown is the 2Fo-Fc map in blue with the $\text{C}\alpha$ -trace of the refined p34ct 1-277 model colored in green. Symmetry related molecules forming crystal contacts at three different faces of the domain are depicted in yellow.

The preliminary model was corrected and extended in COOT [123] and refined against the p34ct 1-277 data to a resolution of 2.7 \AA using Phenix [127]. After several rounds of iterative model building and refinement, the final model comprised 209 out of 277 residues and yielded R-factors of $R_{\text{work}} = 0.2220$ and $R_{\text{free}} = 0.2355$ with good overall stereochemistry values. In total 98.5 % of all residues were located in favored regions in the Ramachandran plot and no outliers were observed. However, the average Wilson B-factor of 92.8 was found to

be relatively high, presumably due to 68 residues which are disordered in the structure. Detailed information on data collection and refinement statistics for p34ct 1-277 is made available in Appendix X.6.

The overall structure of the p34ct N-terminal domain is shown in Figure IV.28. It consists of a central 6-stranded β -sheet with 5 parallel (β 1, β 2, β 4, β 5 and β 6) and one shorter anti-parallel β -strand (β 3), which is surrounded by a total of 3 α -helices on either side (α 2, α 3, α 8 on one side and α 5, α 6, α 7 on the other side). Two additional, much shorter α -helices are located at the C-terminal end of the central β -sheet of the domain (α 1 and α 4, Figure IV.28 A). Surprisingly, the overall fold reveals a *von Willebrand Factor A* (vWA) like architecture (Figure IV.29), with high structural similarity to the A1 domain of *von Willebrand Factor*, a blood glycoprotein involved in hemostasis and platelet aggregation [184]. This more than 2,000 amino acid long multi-domain polypeptide is an extracellular matrix component and best known for its causative role in the *von Willebrand* disease [185], the most common inherited bleeding disorder in humans.

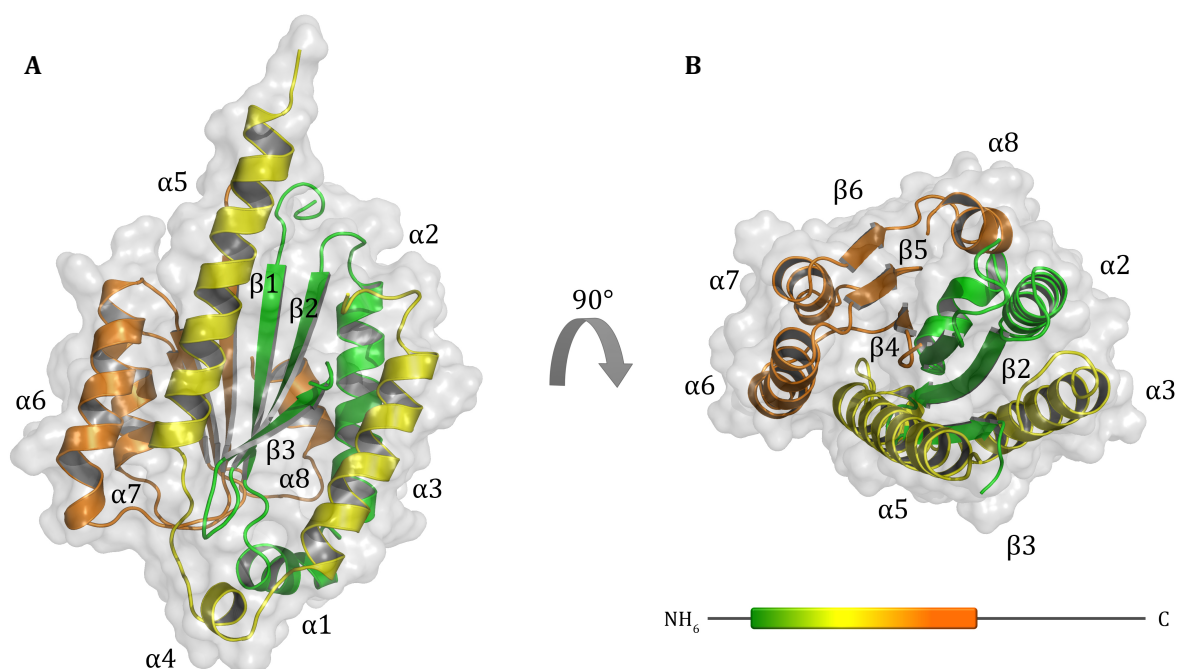


Figure IV.28: Structure of the p34ct vWA like Domain. Front view of the p34ct 1-277 vWA domain (A) and view from the top (B) with the color coding corresponding to the N-terminal (green), central (yellow) and C-terminal (orange) part of the domain, respectively. The structure assumes a typical vWA like fold, consisting of a central β -sheet, which is sandwiched by 3 α -helices on either side. The location of the vWA domain within full-length p34ct is indicated in the equally colored bar representation (B).

The presence of a vWA like domain at the N-terminus of p34 has previously not been reported. In line with this our searches using the primary sequence of both human and *C. thermophilum* p34 did not return a vWA like feature via BLAST and the SMART domain annotation database, respectively [132,145]. However, a Phyre² secondary structure analysis did predict the vWA like fold and suggested several vWA containing structural homologues to p34 [139]. Interestingly, also the N-terminal half of p44, the binding partner of p34 within TFIIF, is predicted to contain a vWA like domain.

While other TFIIF subunits, including p44, are relatively well conserved between different organisms, there seems to be higher variability in p34. Here, the sequence identity is 93 % between human and mouse, whereas it is only 28 % when compared to *C. thermophilum*. Interestingly, the *C. thermophilum* proteins are also larger than their human orthologues, including extensions at both their N- and C-termini and flexible linker insertions. In total the size of p44 and p34 increases by 35 % and 39 %, respectively.

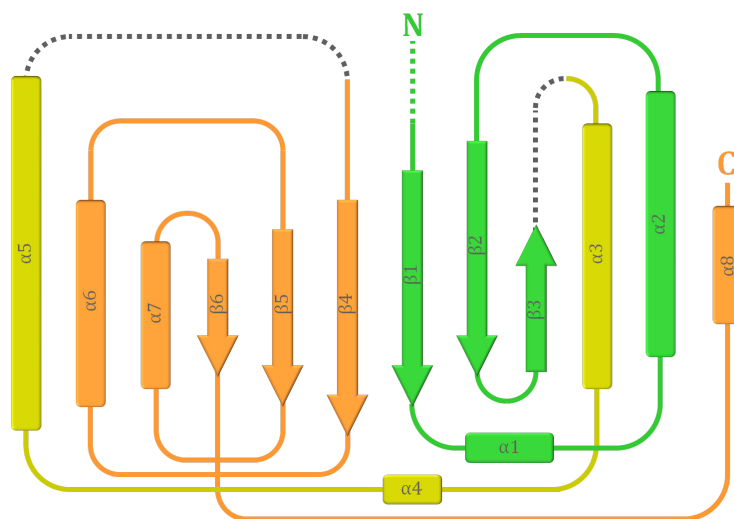


Figure IV.29: Topology of the p34ct vWA Domain. Drawing a topology diagram of the vWA fold shows that its architecture can be distinguished from a typical Rossmann fold by β -strand β 3, which is anti-parallel in all vWA domains, instead of strictly parallel.

Upon closer inspection, p34ct reveals the presence of one variable linker and three linker insertions when compared to its human, mouse, yeast and plant homologues (Figure IV.30). The variable linker is located in a region connecting β -strand β 3 to helix α 3 and is disordered in our crystal structure (Figure IV.30, Linker 1), whereas the three insertions encompass residues 166 - 199, 275 - 320 and 391 - 414 and are not present in the other species. The first of these is located in the vWA domain (Figure IV.30, Insertion 1). It connects helix α 5 back to the β -sheet core and is again disordered in the p34ct structure. The second insertion bridges the N-terminal vWA domain the C-terminal region (Figure IV.30, Insertion 2). It spans more than 40 residues and dislodges the highly conserved C-terminal domain quite distinctively from the vWA fold, most likely introducing a high degree of flexibility between the N-terminal and C-terminal parts of the protein. The third insertion is located at the far C-terminus beyond the C4 zinc-binding motif (Figure IV.30, Insertion 3).

Additional residual electron density patches were observed in the p34ct full-length data, indicating the possible position of the flexible C-terminal domain of p34ct relative to its rigid vWA fold. The zinc binding domain is probably located at the top and side of the vWA domain, shifted upwards between helices α 7 and α 8 (Figure IV.31). Considering the spacious N-C-linker region, this location might be different in p34 proteins of other organisms, which, based on their primary sequence, possess a more compact arrangement of both their N-terminal vWA and C-terminal zinc binding domains.

Attempts were made to model the C-terminal region of p34ct upon molecular replacement of the initial full-length length data with the p34ct 1-277 vWA structure. However, the efforts eventually failed due to only marginal improvements in the residual density patches observed.

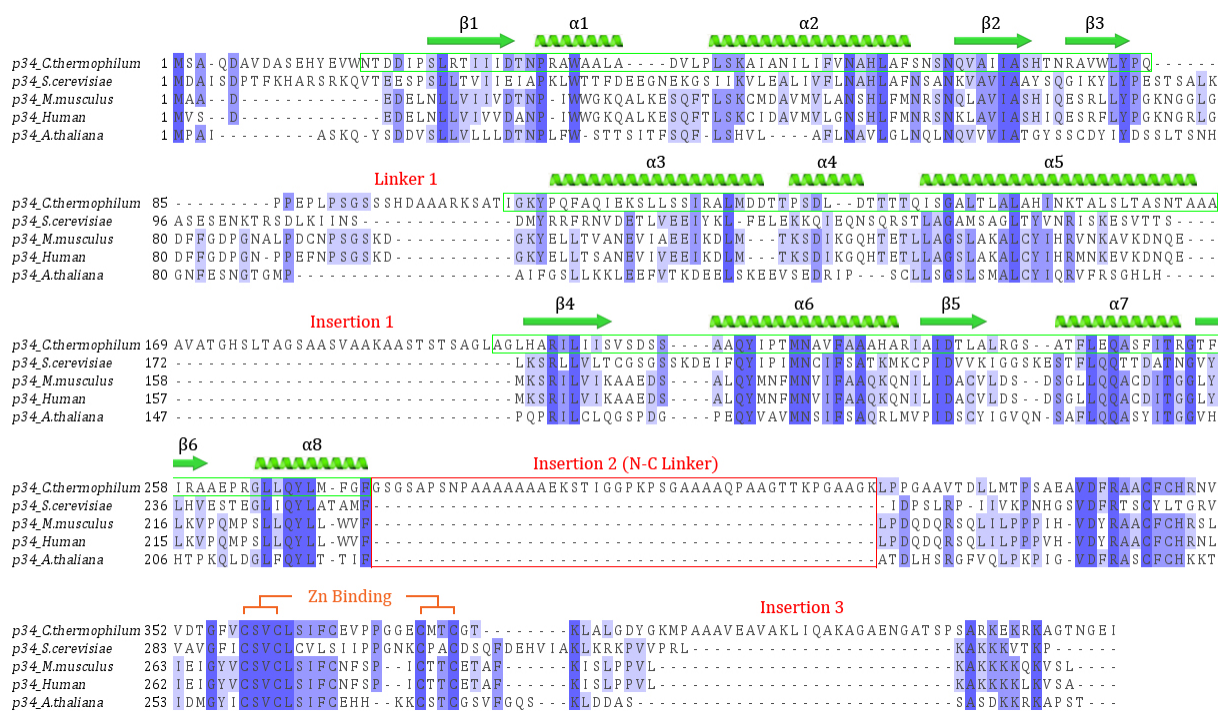


Figure IV.30: Multiple Sequence Alignment of p34 Proteins. The multiple sequence alignment of p34 proteins from five different species was performed with MUSCLE [137] and visualized via JalView [135]. Regions visible in the p34ct vWA structure are boxed in green and secondary structure elements are indicated above the sequence, with arrows representing β -strands while coils are used for α -helices. Conserved residues between species are colored in different shades of blue, depending on the degree of conservation. The N-C-Linker present in *C. thermophilum*, but absent in *S. cerevisiae*, *M. musculus*, human and *A. thaliana*, is highlighted in red. The conserved C-terminal C4 zinc finger motif is indicated in orange.

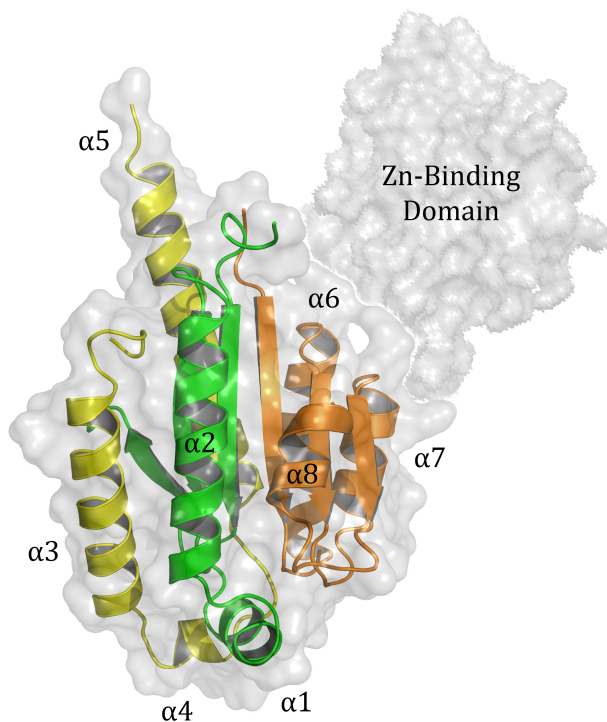


Figure IV.31: Location of the p34ct C-terminal Domain. The schematic contour model illustrates the approximate position of the C-terminal zinc binding domain of p34ct, which could be tentatively located at the top and back side of the vWA fold, shifted upwards between helices $\alpha 7$ and $\alpha 8$.

IV.5.5 Analysis of p34ct Crystal Packing and Elucidation of Binding Sites for p44ct

Closer analysis of the p34ct crystal packing and potential binding surfaces using the PDBePISA service [186] revealed 3 putative dimerization sites for the vWA domain, involving a total of 20, 17 and 11 residues, respectively (Figure IV.32). The first interface would encompass helices $\alpha 2$ and $\alpha 8$, covering an area of 698 \AA^2 and is mainly hydrophobic. The second involves residues in $\alpha 1$ and $\alpha 4$ as well as their respective connecting loops to helices $\alpha 2$ and $\alpha 3$. This interface buries 619.5 \AA^2 and is predominantly formed via salt bridges between several arginine and aspartic acid residues. The last interface involves helix $\alpha 6$, covering an area of 476.7 \AA^2 , and is again strongly hydrophobic.

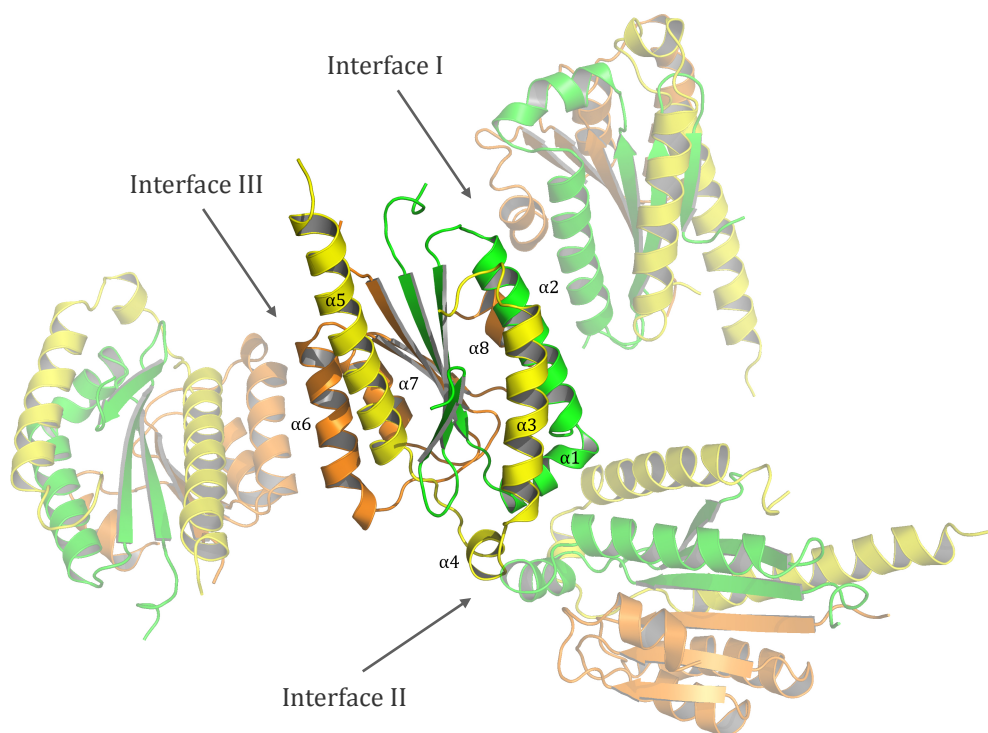


Figure IV.32: Crystal Packing of the p34ct vWA Domain. Crystal contacts within both full-length p34ct and p34ct 1-277 crystals are formed via three putative dimerization interfaces in the vWA domain. The interfaces mainly involve residues in α -helices $\alpha 2$ and $\alpha 8$ (I), $\alpha 1$ and $\alpha 4$ (II) and $\alpha 6$ (III).

Given the calculated free energy gain upon dimer formation of -8.0 , -3.1 and -7.9 kcal/mol, respectively, both hydrophobic contacts are most likely to be formed in solution. However, as their complex significance score equals 0.0 and all potential interfaces also correspond to observed crystal contacts, they may be artificial and just a result of crystal packing. In support of this hypothesis the SEC-MALS analysis of both full-length p34ct and p34ct 1-277 yielded molar masses of 46,890 Da and 32,020 Da, respectively, closely matching the predicted monomeric values of 46,767 Da for p34ct and 31,941 Da in the case of p34ct 1-277 (Figure IV.19 and Table IV.2). Hence, under the conditions tested, p34ct is monomeric in solution, exposing several hydrophobic surface patches, which would, however, lead to a significant free energy gain upon saturation.

At that stage we started to more closely investigate how the interaction of p34 and p44 could be envisioned. From previous studies it was known that residues 321-395 of human p44 (375-534 in *C. thermophilum*) are sufficient for the formation of a tight p34 p44 complex [42]. In addition, an NMR structure of the C-terminal domain of human p44 revealed the fold of its C4C4 RING domain [68] (Figure IV.33 B). The authors also investigated, which parts of the domain are important for interaction with p34 and observed that the conservation of the first zinc binding site is required for the formation of a stable complex (Figure IV.33 B).

A single coordinating cysteine residue mutated at that position seemed to disrupt the local fold of the domain and completely abolished binding [68]. In addition, Phe-374, located in helix $\alpha 2$ close to the first zinc binding site, seems to be important for the interaction with p34 since its mutation to a serine disrupted complex formation [68] (Figure IV.33 B). Overall, the authors suggested, that the interaction is more likely to be of hydrophobic than electrostatic character [68]. Interestingly, however, the overall surface charge of the human p44 C4C4 domain is highly acidic, with a total of 10 aspartic and glutamic acid residues, all of which are solvent exposed (Figure IV.33 B), and only 2 positively charged residues (Arg, Lys).

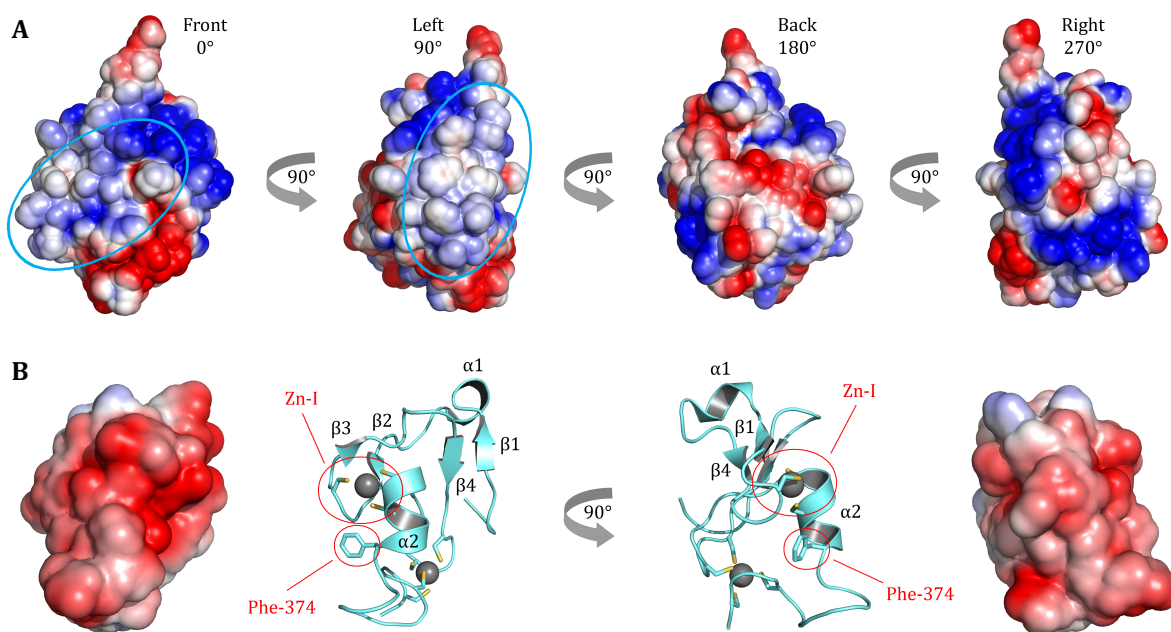


Figure IV.33: Electrostatic Surface Potential of the p34ct vWA Domain. The electrostatic potential contoured at ± 3.0 kbT/ec is mapped onto the p34ct vWA surface in blue (positive), red (negative) and white (neutral) and suggests two putative binding sites for the p44ct C4C4 domain (circled in cyan). An NMR structure of the human C-terminal domain of p44 (1Z60) is shown in two orientations, both as a cartoon model and with the electrostatic surface potential contoured at ± 3.0 kbT/ec mapped onto its surface (B). The zinc ions bound to the C4C4 motif in p44 are shown in grey and the two regions most likely involved in p34 binding are circled in red. The front orientation in (A) equals that in Figure IV.28 A.

With most of the residues around the first zinc binding site and especially helix $\alpha 2$, including Phe-374, highly conserved, we thus proposed that these two sites contact p34 at a mainly hydrophobic and potentially slightly basic region [187]. Two predominantly hydrophobic patches can be found on the front and left side of the p34 vWA domain ($\beta 3$, $\alpha 5$, $\alpha 6$, $\alpha 7$ in Figure IV.28 A and Figure IV.33 A). Binding of p44ct C4C4 to the back of the vWA domain ($\beta 6$, $\alpha 8$) is unlikely, as there is a stretch of solvent exposed, highly negatively charged residues. Sterical hindrance due to the C-terminal domain of p34 might be an additional factor in disfavoring this site. Similarly, we excluded binding to the right face of the domain ($\alpha 2$, $\alpha 3$), as most residues in helix $\alpha 3$ and the hydrophobic part of $\alpha 2$ are not very well conserved (Figure IV.30 and Figure IV.33 A). Hence, the most likely interface was found at the front and/or left side of the vWA domain ($\beta 3$, $\alpha 5$, $\alpha 6$, $\alpha 7$), with most surface exposed residues in helices $\alpha 6$ and $\alpha 7$ being strongly hydrophobic and highly conserved (Figure IV.30 and Figure IV.33 A). Along that line, also the lower portion of helix $\alpha 5$ shows a high degree of conservation and provides a slightly positive contribution to the overall surface potential, possibly mitigating part of the negative contribution by p44 C4C4. Taken together, the front and left side of the vWA domain ($\beta 3$, $\alpha 5$, $\alpha 6$, $\alpha 7$) were considered the most attractive targets for binding of the p44 C4C4 domain (Figure IV.33 A) [187].

IV.5.6 Comparison of p34ct vWA to other vWA Domains

The vWA domain is a general purpose mediator of protein-protein interactions found in more than 6,000 proteins within the eukaryotic kingdom, as annotated in the SMART non redundant protein database [145]. While the majority of vWA containing molecules are cell adhesion or extracellular matrix components, like integrins, vWF and vWA collagens, the vWA fold shows up in a lot of other, completely unrelated cellular processes. Some vWA domains, for example, are part of ion channel subunits, protease inhibitors and relatives thereof, the anthrax toxin receptor family as well as the complement system [188,189]. The most ancient vWA folds are all found in intracellular proteins and are often involved in multi-protein complexes. Homologues of vWA proteins are also widespread in both archaea and bacteria, suggesting an early common ancestor prior to separation into the three major branches of cellular life [190].

With the N-terminal domain of p34ct assuming a vWA like fold, it was not surprising that a closer homology investigation using the DALI server yielded 140 hits with a Z-score higher than 10.0 and r.m.s.d. values between 2.1 and 3.3 Å [141]. Interestingly, most molecules with the highest scores either correspond to the name giving human *von Willebrand Factor* (vWF) A1 domain itself [191] or complexes of this domain from mouse and human origin with a variety of different proteins [192–195]. Besides its capability to mediate protein-protein interactions, the human vWF A1 domain was also identified as a DNA binding protein, specifically to bind to DNA aptamers, which were designed to provide antithrombotic treatment [196]. In case of the yeast homologue of p34, Tfb4, nucleic acid binding via a basic Lys/Arg-rich C-terminal region has recently been suggested [197]. Yet, our data could not confirm this observation since neither p34ct 1-277 nor full-length p34ct were able to bind to single stranded or double stranded 50-mer DNA substrates (IV.2.4). However, it cannot be ruled out that this differs among species. In p34ct the C-terminal Lys/Arg-rich stretch is disrupted by a linker insertion and not as basic as seen in p34/Tfb4 from *S. cerevisiae* (Figure IV.30).

Superposition of the p34ct 1-277 vWA structure with molecules suggested by the DALI search reveals that the overall fold is highly conserved, even among rather distant homologues involved in several different cellular pathways. For example, the vWA domain of the cell adhesion molecule Integrin $\alpha 1$ [198], the *von Willebrand Factor* A1 domain [191] and the 26S proteasome regulatory subunit Rpn10 [199] all share the central 6-stranded β -sheet core, surrounded by 6 α -helices (Figure IV.34 A, B and C, respectively).

However, there are also some remarkable differences present in p34ct that have not been observed in the vWA domains of most other proteins. Helix $\alpha 3$ is significantly longer in p34ct, while helix $\alpha 8$ is relatively short and often found to be longer in other vWA containing proteins, except for Rpn10 and the vWA domain of the DNA dependent helicase Ku70 [200], which share a C-terminal α -helix similar to that in p34ct vWA. Furthermore, the two rather prominent helices at the C-terminal end of the central β -sheet ($\alpha 1$ and $\alpha 4$) are often replaced by short loops in other vWA folds (Figure IV.34). The most striking feature, however, is helix $\alpha 5$, which is about twice as long compared to the same helix in other vWA proteins, considerably protruding outwards from the top side of the domain (Figure IV.34 and Figure IV.35). The drastic extension of this helix might also explain the presence of the rather long linker region required to provide the connection back to the β -sheet core, as observed in p34ct (Figure IV.30). Hence, as this linker is absent from p34 sequences of other organisms, it remains unclear, whether helix $\alpha 5$ is similarly prominent in these species. Based on our p34ct vWA structure, multiple sequence alignments and secondary structure prediction results (not shown), we propose helix $\alpha 5$ to be at least one or two turns shorter in human, mouse and *S. cerevisiae* p34ct (Tfb4), or otherwise shifted more to the center of the domain to permit the connection to $\beta 4$ (Figure IV.30).

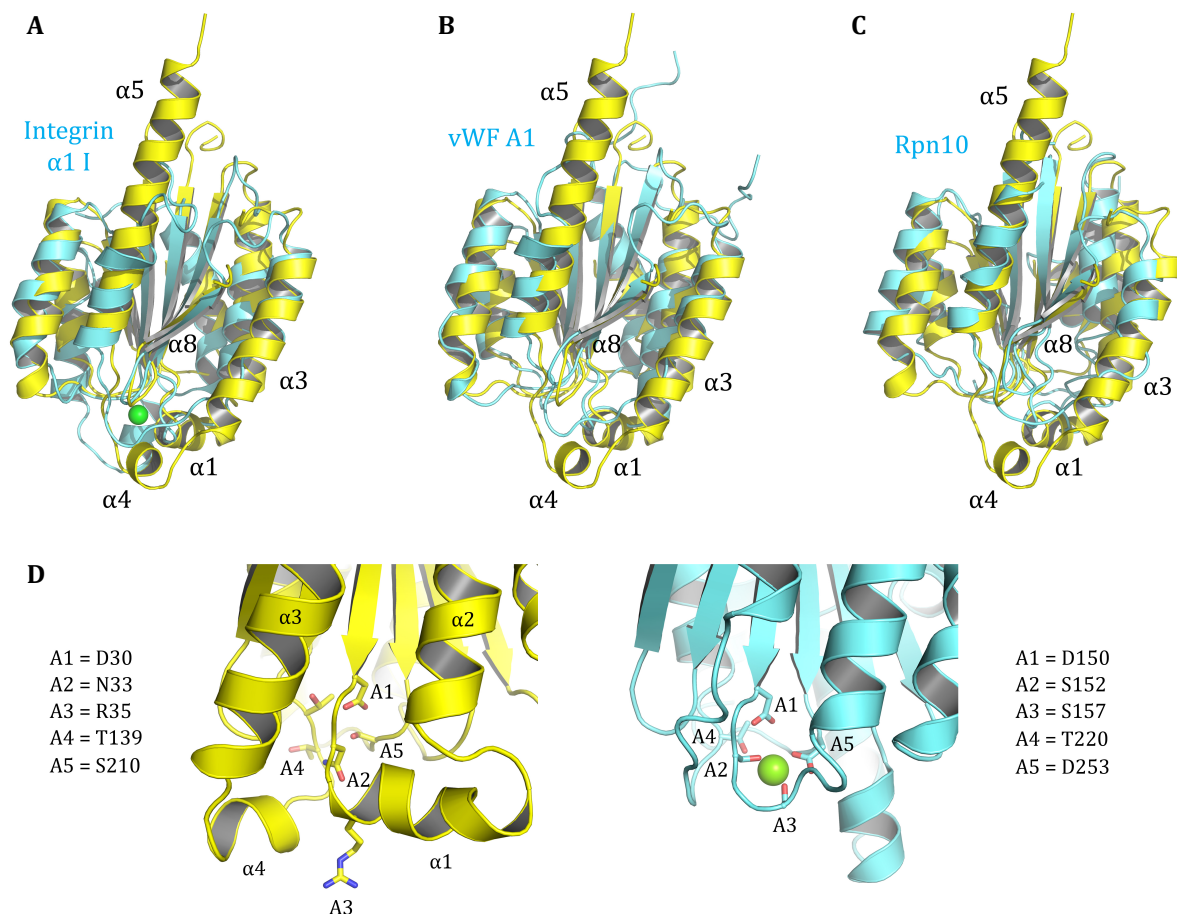


Figure IV.34: Superposition of p34ct vWA with other vWA Domains. Shown is the superposition of the human Integrin $\alpha 1$ I domain (A, 1PT6), the human vWF A1 domain (B, 1AUQ) and Rpn10 from *Schizosaccharomyces pombe* (C, 2X5N, each in cyan) with the p34ct 1-277 vWA structure (yellow). The putative metal coordination site in p34ct is enlarged (D, yellow) and compared to the MIDAS motif in Integrin $\alpha 1$ (D, blue), with the 5 MIDAS elements (A1 - A5) shown in stick representation and the bound Mg^{2+} ion in Integrin $\alpha 1$ (1PT6) as green sphere.

The *von Willebrand Factor A* like domain in general is reminiscent of a typical Rossmann fold, which is found in many nucleotide binding proteins [201], with the major difference being that all vWA domains contain one anti-parallel, instead of strictly parallel β -strands (Figure IV.29). While no vWA domain is known to bind nucleotides, many of them are capable of coordinating metal ions via a noncontiguous sequence motif termed the metal-ion-dependent adhesion site (MIDAS). For example, all vWA domains in Integrin α subunits and most Integrin β representatives contain a perfectly conserved MIDAS motif, which is formed by 3 loops at the C-terminal end of the central β -sheet of the vWA fold [188].

The MIDAS motif consists of 5 residues in the order D-x-S-x-S ... T ... D, which will in the following be referred to as A1 to A5. In p34 from *C. thermophilum* (ct), homo sapiens (h) and several other species the aspartic acid at position A1 is perfectly conserved. However, in both p34ct and p34h the MIDAS motif is disrupted by a helical insertion ($\alpha 1$) directly after A1, which is often rich in one or two tryptophanes, leading to non-conservative substitutions in A2 and A3 (Figure IV.34 D). In case of the p44 vWA domain A1 and A2, and in some species even A3, are conserved (Appendix X.5). Here, another insertion of different sequence character is present after A2 or A3, respectively, which is likely also helical, based on secondary structure prediction results (compare Appendices X.4 and X.5). In addition, p44 seems to carry substitutions in A4 and A5, of which one is non-conservative. Hence, neither p34 nor p44 are likely to bind a metal ion via the MIDAS motif. This is consistent with our p34ct vWA structure not including any bound ligands.

IV.5.7 The Multifaceted Roles of vWA Domains

The typical vWA domain is known for its capability to mediate protein-protein interactions. However, the surface used and the type of contact made can vary among the numerous vWA domains found in different proteins. In integrins, for example, the binding of the ligand is dependent on the presence of divalent cations and thus a functional MIDAS motif. It usually results in allosteric effects that can lead to long-distance conformational changes and rearrangement of domains, as seen for the Integrin β I subunit [202]. Other vWA domains seem to be rather rigid molecules, that do not undergo large conformational changes but rather serve as scaffolding proteins. Several structures of vWA domains in complex with their respective ligands or binding partners are available from the Protein Data Bank (www.rcsb.org), providing interesting insights into how the vWA fold contacts other protein molecules.

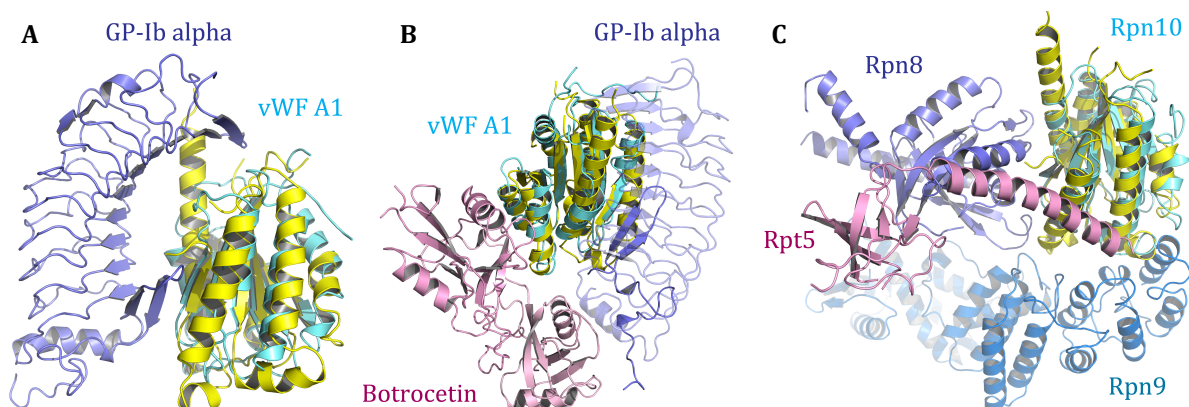


Figure IV.35: Superposition of p34ct vWA with Complex Structures of vWA Domain Homologues.

The p34ct 1-277 structure (yellow) is superposed onto the human vWF A1 domain (cyan) in complex with Glycoprotein Ib alpha alone (A, 1SQ0) and in complex with human Glycoprotein Ib alpha and the snake venom Botrocetin from *Bothrops jararaca* (B, 1U0N). Superposition of p34ct 1-277 with Rpn10 in contact with its binding partners in the 26S proteasome regulatory particle (4C0V) is shown in (C). The latter is based on a cryo-EM reconstruction using data collected on the *Saccharomyces cerevisiae* 26S core particle. Color coding of the superposed molecules shown in A, B and C is the same as used in Figure IV.34.

For the *von Willebrand Factor* A1 domain it has been shown that Glycoprotein Ib alpha (GP-Ib- α) binds the vWA fold at two distinct sites and thereby generates a tight interaction. One is formed by a β -strand addition of residues in the C-terminal flank of GP-Ib- α with vWA β 3 and several leucine rich repeats contacting helix α 3, which would correspond to α 5 in p34ct vWA. The second prominent interface is at the top of the domain, involving loops α 1 β 2, β 3 α 2 and α 3 β 4 in vWF A1 [203] (Figure IV.35 A).

Another complex structure of the A1 domain bound to GP-Ib- α and the snake venom Botrocetin shows the latter contacting helices α 4 and α 5 of the vWA fold (α 6 and α 7 in p34ct, Figure IV.35 B). Interestingly, the binding of Botrocetin to the side of vWF A1 allows the toxin to move around slightly upon the incoming GP-1b- α , forming additional contacts with the glycoprotein, thereby significantly reducing its off-rate [194]. In this case the vWA fold serves as a scaffold to mediate the interaction of two partners, whose affinity for each other would normally be too low for a stable engagement.

Another vWA domain involved in multiple protein-protein interactions is that of the 26S proteasome subunit Rpn10 (Figure IV.34 C). The complete architecture of the proteasome regulatory particle has recently been solved using a cryo-EM approach at a resolution of approximately 10 Å [204–206]. Fitting of already known high resolution structures, including Rpn10 [199], into the electron density determined, revealed both the location and some of the interaction partners of the Rpn10 subunit (Figure IV.35 C). The Rpn10 vWA domain

predominantly interacts with Rpn8 via an interface similar to that observed for the GP-Ib- α C-terminal flank, involving β 3 and helix α 4 in Rpn10 (α 5 in p34ct). Additional contacts are made with Rpt5 and Rpn9, the former via helices α 2 and α 3 (same in p34ct) and the latter by Rpn9 probably contacting several loop regions in Rpn10 at the bottom of the vWA domain (Figure IV.35 C). Although this may be the exception, Rpn10 is thus a prime example of a vWA domain engaged in multiple interactions, while simultaneously serving as a ubiquitin receptor via its C-terminal extension and UIM motif, presumably relaying the ubiquitylation signal to other proteasomal subunits.

Given the high structural similarity to both vWF A1 and Rpn10, one can hypothesize that p34ct and likely also p44ct fulfill comparable roles within TFIIH. Through its ability to contact several subunits simultaneously the vWA domain might serve as a general scaffold to stabilize the overall complex integrity, while at the same time both proteins could perform auxiliary functions mediated via their C-terminal zinc binding motifs.

IV.5.8 Implications for Protein-Protein Interactions within TFIIH

In light of the multifaceted roles of already known vWA domains, it is tempting to speculate how both p44 and p34 function within the context of the general transcription factor TFIIH. The vWA fold of p44 shares high homology with the 26S proteasomal subunit Rpn10 [188,207], yielding 25 % identity and 59 % similarity within a stretch of 144 amino acids using LAlign [136]. Furthermore, p34ct vWA is structurally very similar to the Rpn10 vWA domain of *Schizosaccharomyces pombe*, with a DALI Z-score of 16.8 and an r.m.s. deviation of 2.6 Å (Figure IV.34 C). As Rpn10 is located at the outer rim of the 19S regulatory particle, indirectly stabilizing its lid-base interaction [204] and probably relaying signals from the outside to other proteasomal subunits, it is conceivable that both p44 and p34 assume similar roles in TFIIH.

It is well established that the p44 vWA domain contacts a C-terminal region in XPD, thereby stimulating its helicase activity [36]. In turn p34 vWA binds strongly to the C-terminal C4C4 RING domain of p44 [42], which is essential for transcriptional activity [208], potentially forming a regulatory triad with XPD and p44. However, only very little is known about the function of p44's central and p34's C-terminal zinc finger domain, respectively. Based on the results discussed above, it is very likely that the vWA domains of both p44 and p34 are involved in multiple contacts, presumably providing stability and serving as a platform and mediator for other subunits, while their zinc finger motifs may enable them to perform additional roles. In line with this hypothesis, co-immunoprecipitation experiments performed almost 20 years ago, using *in-vitro* translated polypeptides, suggested p34 to interact with the two helicases XPD and XPB of core-TFIIH [41]. The authors also showed that the N-terminal region of p34 (residues 1 - 245) is sufficient to co-precipitate with both XPD and XPB. However, to our knowledge, this observation has never been confirmed by other methods. In addition, results from the same group imply an interaction between p44 and p62 based on yeast two hybrid studies [41], a finding we could verify in preliminary *in vitro* experiments using purified proteins from *C. thermophilum* (data not shown).

Recent cryo-EM studies using TFIIH subunits from yeast identified a homogenous and stable minimal core complex consisting of Tfb1, Tfb2, Tfb4, Tfb5 and Ssl1 (corresponding to p62, p52, p34, p8 and p44), which was large enough for single particle analysis [54]. Interestingly, however, both p52 and p8 were previously reported to interact exclusively with each other and with XPB via p52 [197]. So far only Jawhari *et al.* have suggested that p62 interacts weakly with p52, but lacked to show the data supporting their claim [39]. Hence, one could speculate which other factors are responsible to anchor the p52/p8 subunits into the minimal-core complex. Is the putatively weak interaction of p52 with p62 sufficient, or might one or both vWA domains of p44 and p34 be involved in stabilizing otherwise weakly interacting complex subunits, as seen similarly with the vWF A1 / GP-Ib-a / Botrocetin complex (Figure IV.35 B)? In support of this notion, it was shown that a stable 5-subunit minimal complex of Rad3, Tfb1, Tfb2, Tfb4 and Ssl1 (corresponding to XPD, p62, p52, p34 and p44) can only be obtained when Tfb4 (p34) is included during co-expression [53]. When Tfb4 was omitted, the complex lacked Tfb2 (p52), which suggests that Tfb4 (p34) is required for the integration of Tfb2 (p52) into the minimal core complex in yeast.

Further analysis will be required to decipher this intricate network of interplay and regulation between TFIIH subunits and beyond in the context of transcription, DNA repair, cell cycle control and several other cellular processes, that are linked to TFIIH or some of its versatile subunits.

IV.6 Crystal Structure of p34ct p44ct MC - A Novel Binding Mode for RING Domains

IV.6.1 Crystallization of p34ct p44ct MC

With the p34ct 1-277 vWA structure in hand (IV.5.4) and based on our previous experience with the human p34h p44h minimal complex (IV.1.3), we re-evaluated our approach to achieve a complex structure. In the following the p34ct p44ct minimal complex (MC) was designed, comprising the p34ct 1-277 vWA domain and the remote C-terminal C4C4 RING domain intrinsic to p44ct 368-534. The complex was either reconstituted *in vitro* or obtained by co-purification of both proteins, of which only p34ct 1-277 was tagged (IV.2.3).

Initial crystallization trials with p34ct p44ct MC immediately led to reproducible but very fragile needle shaped crystals (Figure IV.36 A), which resembled the flat needles previously obtained for the human minimal complex (Figure IV.4 B and Figure IV.5 B). Along the same line, also the p34ct p44ct MC crystals showed only poor diffraction and eluded proper optimization regarding crystal size and quality.

As the initial p34ct p44ct MC needles crystallized in a PEG based mixture containing CHES pH 9.5 as a buffer component, the first optimization trials focused on using these conditions. Furthermore, a ThermoFluor analysis showed, that both p34ct 1-277 as well as the reconstituted minimal complex exhibited a very good unfolding profile in CHES at high pH, with melting temperatures of $T_m = 51\text{ }^\circ\text{C}$ and $T_m = 56\text{ }^\circ\text{C}$, respectively (Figure IV.17 A and Figure IV.18). However, melting curves of the p44ct domain alone could not be recorded for CHES pH 9.5 or similar high pH buffers (Figure IV.17 B), presumably due to its small and accessible fold being in a more open conformation, leading to a prior and satiable binding of the SYPRO Orange compound. Further analysis using circular dichroism spectroscopy later supported this finding (Figure IV.45 B).

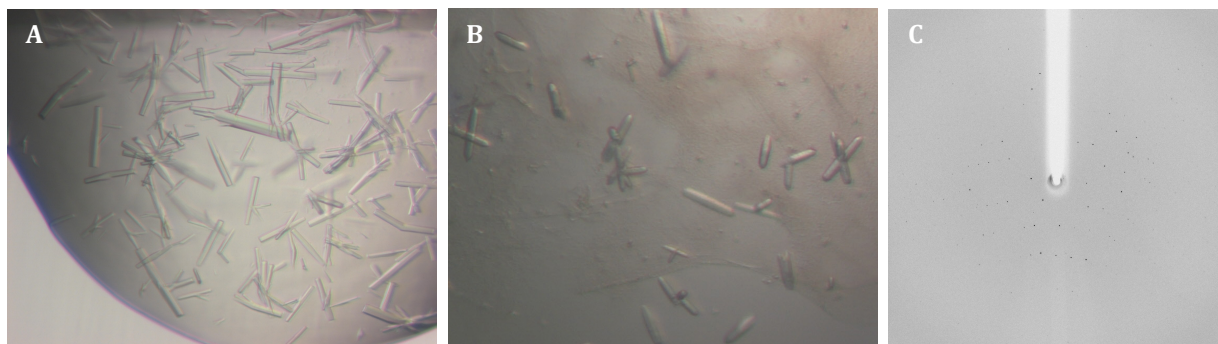


Figure IV.36: Crystals of p34ct p44ct MC and X-Ray Diffraction. Typical crystals of p34ct p44ct MC were reproducible flat needles, which grew in 10 % PEG 8,000 and 100 mM CHES pH 9.5 (A). In addition, tiny bar shaped crystals were observed once in 15 % PEG 20,000 and 100 mM MES pH 6.5 (B), yielding faint x-ray diffraction patterns up to a resolution of 3.7 Å (C).

Upon dissatisfying progress with the CHES based optimization trials, several intensive re-screening attempts were carried out to identify additional conditions supporting crystal growth. However, only very few novel crystallization conditions could be detected. One of these contained tiny bar shaped crystals among some brownish precipitate (Figure IV.36 B). This condition, consisting of 15 % PEG 20,000 and 100 mM MES pH 6.5, was markedly different from those previously identified, especially regarding its pH, which was shifted by 3 magnitudes towards the acidic. Of these crystals a total of 4 could be fished and taken to a micro focus synchrotron beamline at the ESRF in Grenoble for data collection (IV.6.2). Unfortunately, the tiny crystals grown in this condition entirely eluded reproduction during further optimization trials.

IV.6.2 Structure Determination of p34ct p44ct MC

Due to their limited size the tiny bar shaped crystals of p34ct p44ct MC yielded only faint diffraction patterns. Nevertheless, complete data sets were collected, one of which could be processed to a resolution of 3.7 Å (Figure IV.36 C). The space group determined was P6₃22 with cell parameters of $a = b = 138.2$ Å, $c = 95.2$ Å and $\alpha = \beta = 90^\circ$ and $\gamma = 120^\circ$. Ultimately, the quality of the data collected was sufficient to solve the structure of the p34ct p44ct Minimal Complex via molecular replacement, using the previously built p34ct 1-277 vWA domain at a resolution of 2.7 Å as template model for phasing (IV.5.4).

The initial electron density maps after molecular replacement revealed the presence of the p34ct 1-277 vWA domain in its previously determined conformation including some additional and intermittent density patches close to $\alpha 5$, $\alpha 6$ and $\beta 3$ of the vWA fold. The latter clearly accounted for the bound C4C4 RING domain of p44ct, which was subsequently built using the human p44 NMR structure (PDB ID: 1Z60) and the weak anomalous density signals for the two zinc ions bound to the domain as guidelines for modeling [68,67]. Apart from both N- and C-termini as well as a long linker insertion, the central part of the p44ct C4C4 domain could be well modeled (see Appendix X.5).

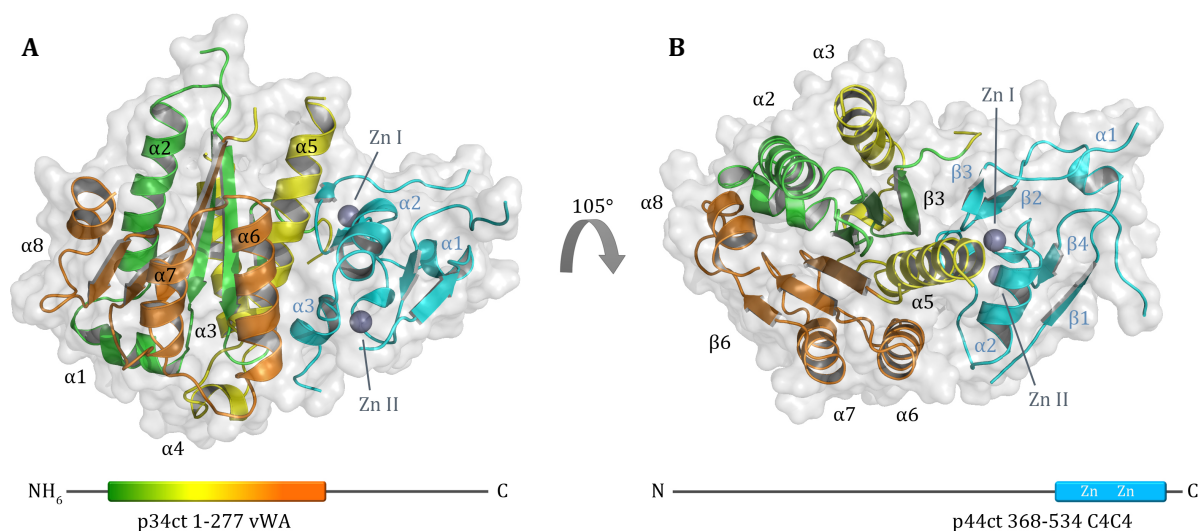


Figure IV.37: Structure of the p34ct p44ct Minimal Complex. Side view of p34ct p44ct MC (A) and view from the top (B) with the color coding corresponding to the N-terminal (green), central (yellow) and C-terminal (orange) part of the p34ct 1-277 vWA domain, respectively, and the p44ct C4C4 RING domain colored in cyan. Contacts with p34ct 1-277 vWA are mainly formed by residues around the first zinc binding site (Zn I) in p44ct C4C4, especially via $\alpha 2$ and the $\beta 2\beta 3$ loop (B). The location of both domains within the full-length proteins is indicated in the equally colored bar representation (A and B, below).

The p34ct p44ct MC structure was refined to a resolution of 3.7 Å using Phenix, yielding final R-factors of $R_{\text{work}} = 0.2783$ and $R_{\text{free}} = 0.3238$ with good overall stereochemistry values, considering the low resolution of the data collected. In total 92.4 % and 7.2 % of all residues were in the favored and allowed regions of the Ramachandran plot, respectively, with only 0.4 % Ramachandran and 2.2 % rotamer outliers observed. The final model comprised 198 out of 277 residues in p34ct vWA and 77 out of the 167 amino acids in the p44ct C4C4 domain, with the remaining residues not being modeled due to their disordered nature and presumably high flexibility in the p34ct p44ct MC crystals. Nevertheless, the results allowed for the first time to delineate the tight mode of interaction between the p34ct and p44ct subunits of TFIIF (Figure IV.37 and section IV.7). Further details with respect to data collection and refinement statistics can be found in Appendix X.7.

The p44ct C4C4 RING domain interacts with the p34ct vWA fold at a mainly hydrophobic and slightly basic surface, comprising one of the two binding sites previously predicted (IV.5.5 and Figure IV.33 A). Contacts are predominantly made via α -helices $\alpha 5$, $\alpha 6$ and the extension of β -strand $\beta 3$ in p34ct, involving residues around the first zinc binding site (Zn I), α -helix $\alpha 2$ and the $\beta 2\beta 3$ loop forming Zn I in the p44ct C4C4 domain (Figure IV.37). The latter is inserted into a largely hydrophobic groove between helices $\alpha 5$ and $\beta 3$ in p34ct, tightly locking p44ct to the p34ct vWA fold and bringing its first zinc binding site (Zn I) in very close proximity to helix $\alpha 5$ in p34ct (Figure IV.37 B). This explains why complex formation is completely abolished upon mutation of a single cysteine residue involved in the coordination of the first zinc ion [68]. The resulting disruption of the local fold presumably leads to a loosening of the $\beta 2\beta 3$ loop in p44ct, preventing it from proper insertion. Further details regarding the interaction interface will be discussed in section IV.7.1.

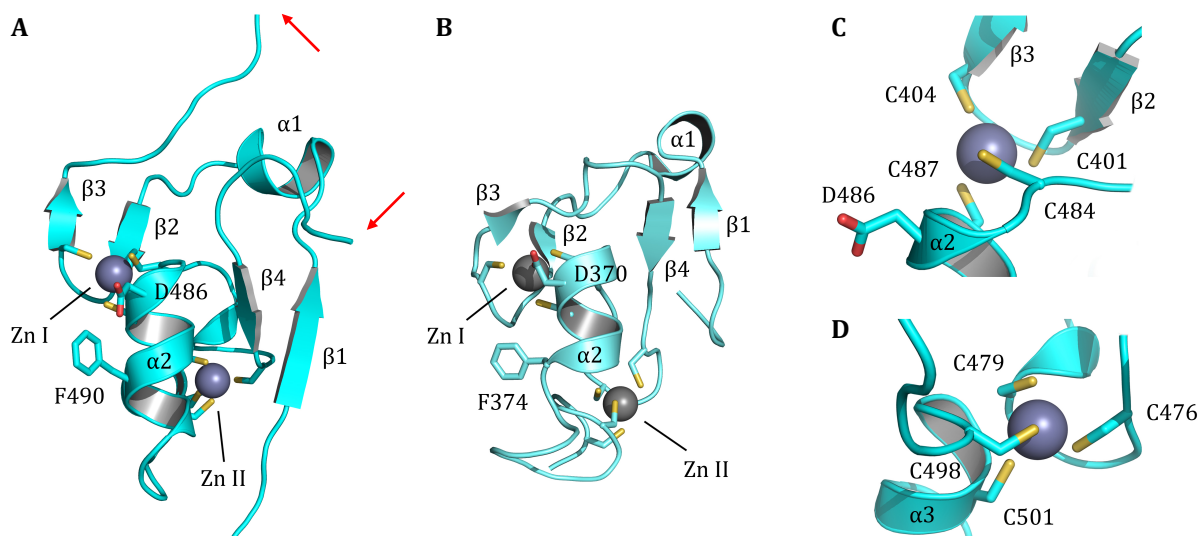


Figure IV.38: Comparison of p44ct C4C4 to p44h C4C4. The p44ct C4C4 RING domain (A) assumes the same overall fold as the human C4C4 NMR structure (B, 1Z60), except for a 50 amino acid linker insertion between $\beta 3$ and $\beta 4$, which is disordered in p34ct p44ct MC (A, red arrows). Shown in stick representation are the highly conserved F490/F374 and D486/D370 residues, involved in p34ct vWA binding, as well as the 8 cysteine residues, coordinating the two zinc ions bound to the domain (grey spheres). For clarity, both zinc binding sites in the p44ct C4C4 RING domain are enlarged in (C, Zn I) and (D, Zn II).

Comparing our model of p44ct 368-534 to the human NMR structure reveals the same overall fold of the C4C4 RING domain (Figure IV.38), mutually supporting both findings. Especially helix $\alpha 2$ as well as the local environment around the two zinc binding sites show a strong degree of structural conservation. In addition, a few highly conserved residues, like F374 and D370 in human p44 C4C4, which were previously suggested to be involved in complex formation [68], are also present in the p44ct C4C4 domain (Figure IV.38) and are oriented towards the vWA domain in the p34ct p44ct MC structure (Figure IV.37).

Consistent with this side of the C4C4 domain contacting p34ct there is more structural variability observable in the regions pointing away from the complex interface, like helix $\alpha 1$ and the short $\beta 1\beta 4$ sheet in p44ct C4C4, which seem to be twisted and in a more open conformation (Figure IV.38 A). This could be directly related to the 50 - 55 amino acid linker insertion in the p44ct C4C4 domain compared to its human counterpart, marking the most striking difference of both domains. The insertion is located between β -strands $\beta 3$ and $\beta 4$ in p44ct C4C4 (Figure IV.38 A, red arrows) and is presumed to be highly flexible, due to the disorder in the p34ct p44ct MC crystal structure. In contrast, $\beta 3$ and $\beta 4$ in human p44 are connected by a very short loop of only 4 residues, rendering the p44h C4C4 domain much more compact (Figure IV.38 B).

IV.6.3 Comparison of the p44ct C4C4 Domain to other RING Domains

While, apart from the linker insertion, the p44ct C4C4 domain is comparable to that of its human counterpart, including several conserved residues involved in mediating the p34 interaction within TFIIH (IV.6.2), both domains exhibit a high similarity to the fold of other RING modules. A close relation to the C3HC4 type of RING domains has previously been suggested [67]. In addition, homology matching of our p44ct C4C4 structure to others in the protein data bank using the DALI server [141] also revealed a high similarity to the zinc binding regions of several C3H2C3 type RING finger proteins (Figure IV.39).

Typically RING domains are mediators of protein-protein interaction, which have been implicated in a variety of cellular processes [209,210]. They are found in a number of E3 ubiquitin ligases and are often involved in regulating protein levels via E2 conjugating enzyme dependent ubiquitination of target substrates [211,212]. Usually they coordinate two zinc ions through a variable combination of cysteine and histidine residues, leading to the C3HC4 and C3H2C3 classification, based on the sequential order of residues involved. However, also NOT4, a C4C4 type RING protein, has already been shown to bind to an E2 activating enzyme and exhibit E3 ubiquitin ligase activity [213], suggesting a similar role for the p44 RING domain. Though a ubiquitin ligase activity has recently been suggested for the p44 homologue Ssl1 from yeast [50], no direct interaction of either p34 nor p44 with any component of the ubiquitination pathway has been reported to date.

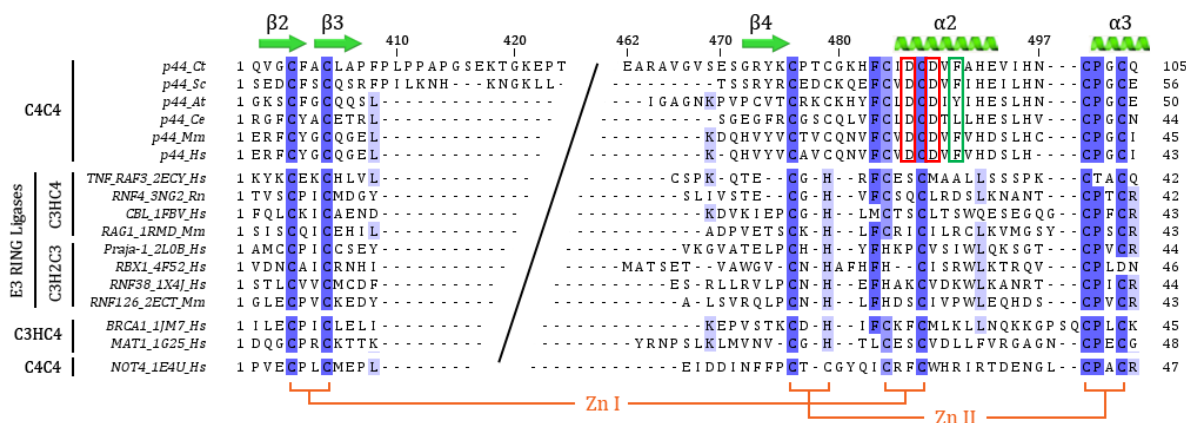


Figure IV.39: Multiple Sequence Alignment of RING Domains. The RING domain of p44 proteins from *C. thermophilum* (Ct), *S. cerevisiae* (Sc), *A. thaliana* (At), *C. elegans* (Ce), *M. musculus* (Mm) and human (Hs) are compared to similar motifs of structural homologues (E3 RING Ligases) from the DALI database [141] as well as the RING domains of BRCA1, MAT1 and NOT4. The type of RING motif is indicated on the far left and PDB IDs are shown behind the name of each protein of comparison. For clarity, 38 residues of the linker region only present in *C. thermophilum* have been omitted. Secondary structure elements in the p44ct C4C4 domain are depicted above the sequence, along with proper residue numbering. Green arrows represent β-strands while coils indicate α-helices. Conserved residues between all species are colored in different shades of blue, depending on the degree of conservation. The C4C4, C3HC4 or C3H2C3 zinc binding motifs are indicated in orange. The alignment was generated using MUSCLE [137], manually modified and visualized via JalView [135].

Among all known RING domains the structural conservation is mainly located around the first zinc binding site, especially the β2β3 hairpin and helix α2, while more variability can be observed for the second zinc binding area and adjacent regions (not shown). Interestingly, the structurally highly conserved β2β3 hairpin as well as helix α2 are also the predominant elements engaged in interaction with the p34 subunit of TFIIH, suggesting a similar binding mode compared to other RING domains (Figure IV.37). Indeed these regions are often involved in protein-protein contacts mediated by different RING finger molecules [214,215].

However, multiple sequence alignments of p44 C4C4 with RING motifs from various structural homologues indicate several striking differences, which are predominantly located in the secondary structure elements forming the p34ct p44ct interface. Especially noteworthy are two highly conserved aspartic acid residues in helix $\alpha 2$ of all p44 C4C4 domains (Figure IV.39, boxed in red), which are replaced by polar and hydrophobic amino acids, respectively, in other RING domains. Considering that the second of these is likely to be involved in electrostatic contacts with a prominent lysine in p34ct vWA, thereby capping the first zinc binding site in p44 C4C4 (IV.7.1), the exchange marks a significant chemical shift in this otherwise structurally comparable interaction surface. In addition, the phenylalanine F490 in p44ct, which is similarly conserved and located in an area characterized by exclusively hydrophobic residues (Figure IV.39, boxed in green), is replaced by a significantly smaller or highly charged amino acid in most other RING domains. As F490 is essential for complex formation between p34 vWA and p44 C4C4 (IV.7.4) [68], this difference adds to the switch in surface potential within helix $\alpha 2$ of all p44 RING domains.

While remarkable substitutions are observable in helix $\alpha 2$ of p44 C4C4 compared to other RING domains, putative general differences in the binding mode of the $\beta 2\beta 3$ loop and the first zinc binding site are difficult to decipher, due to the more divergent nature of residues found in these regions (Figure IV.39). Apart from the highly conserved cysteines forming this site and the obvious structural homology, the underlying amino acid sequence exhibits no clear correlation, even among the closely related p44 RING domains.

Taken together there are several remarkable differences detectable that distinguish the p44 C4C4 domain from other known RING finger molecules, including the C4C4 domain of NOT4, which otherwise shares the same zinc coordinating motif (Figure IV.39). Along with the fact that p44 has not yet been verified to interact with any of its putative partners in the ubiquitination pathway, this suggests the p44 C4C4 domain to represent a novel variation within the otherwise structurally highly homologous RING domains. Though this hypothesis has already been raised before [68], our structural data on p34ct p44ct MC and the analysis of its complex interface (IV.6.2, IV.7.1 and IV.7.4) further support this assumption. It will be interesting to investigate how the tight p34 p44 interaction regulates the overall enzymatic activity of TFIIF and what specific role the p44 C4C4 RING domain assumes in this context.

IV.6.4 Analysis of p34ct p44ct MC Crystal Packing

While both p34ct 1-277 and p44ct 368-534 were found to be monomers in solution, the p34ct p44ct minimal complex surprisingly yielded a molar mass indicative of a hetero-tetramer during SEC-MALS analysis (IV.2.6). Upon determination of the p34ct p44ct MC structure (IV.6.2) and closer investigation of the contact interfaces via the PDBePISA server [186] a strong hetero-tetrameric assembly could indeed be observed.

The p34ct p44ct interaction itself is mediated via 33 residues in p34ct 1-277 vWA and 27 residues in the p44ct C4C4 domain, in total burying an area of 1,037.6 Å² and resulting in a free energy gain of -10.4 kcal/mol with a calculated complex significance score of CSS = 0.511. The zinc ion of the first zinc binding site in p44ct C4C4 further contributes to the free energy gain, adding an additional -13.0 kcal/mol due to its close proximity to α5 in p34ct vWA and effective solvent shielding (Figure IV.37). Overall this yields a combined free energy gain of ΔG = -23.4 kcal/mol for the p34ct p44ct hetero-dimer, strongly supporting the expected functional relevance of complex formation.

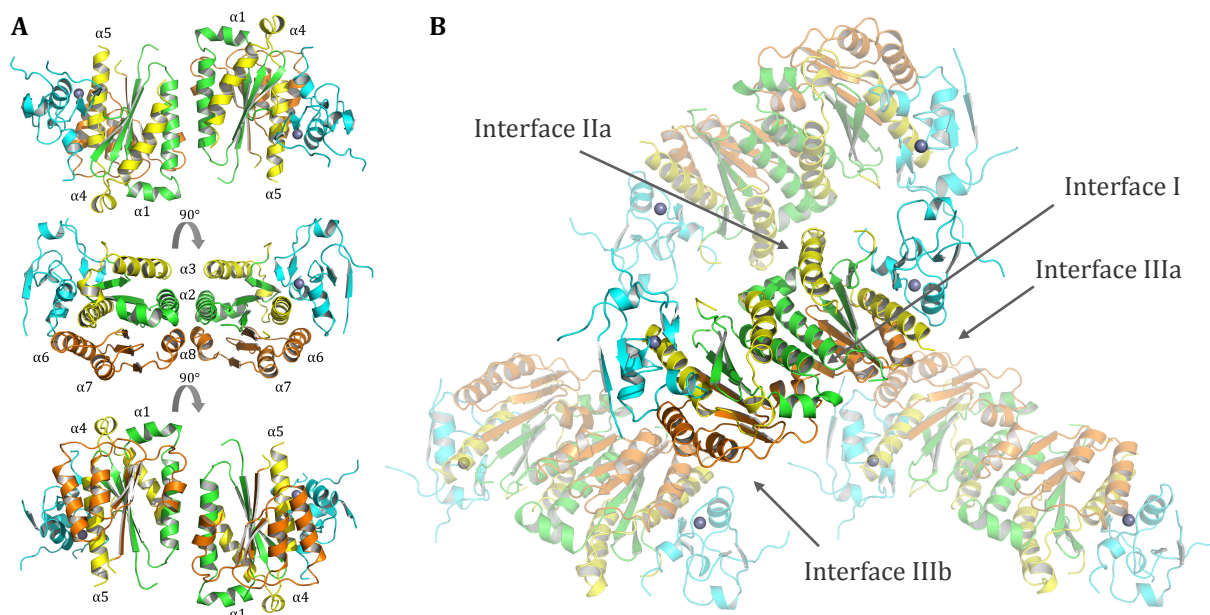


Figure IV.40: Crystal Packing of the p34ct p44ct Minimal Complex. Panel (A) shows 3 orientations of the hetero-tetramer of p34ct p44ct MC, which is formed by dimerization of two p34ct vWA domains (colored green, yellow and orange) via Interface I (helices α2 and α8) upon p44ct C4C4 binding (cyan). Within the crystal lattice, the tetrameric assembly forms two additional crystal contacts (Interfaces IIa/b and Interfaces IIIa/b) via each of its vWA domains (B). For clarity interface IIb and the zinc ion not involved in p34ct binding (Zn II) have been omitted from (A and B).

Apart from the obvious hetero-dimer the p34ct p44ct minimal complex itself also forms dimers within the crystal lattice, which are mediated solely by the p34ct vWA domain (Figure IV.40). The most significant of these involves an additional 27 residues in each molecule and buries an area of 861.3 Å², yielding a calculated free energy gain of -12.9 kcal/mol with a CSS-value of 0.339 (Interface I). Here, contacts between both vWA domains are predominantly hydrophobic and formed via helices α2 and α8, resulting in a similar interface as previously observed for the strongest putative dimerization surface in the p34ct vWA structure (IV.5.5). While p34ct vWA dimers alone were not favored in solution, the p34ct p44ct hetero-tetramer readily assembles (IV.2.6). Presumably the solvent shielding of a large portion of p34ct vWA by p44ct C4C4 binding, at a site opposite to the dimerization interface, shifts the equilibrium towards the p34ct vWA dimer, yielding an overall free energy gain of ΔG = -59.7 kcal/mol for p34ct p44ct hetero-tetramer formation.

IV.7 Analysis of the p34ct p44ct Complex Interface

IV.7.1 Deciphering of Interaction Sites within p34ct p44ct MC

As already mentioned in section IV.6.2 the interaction of p34ct 1-277 vWA and p44ct 368-534 C4C4 is established via α -helices $\alpha 5$, $\alpha 6$ and the extension of β -strand $\beta 3$ in p34ct vWA, involving residues around the first zinc binding site (Zn I), α -helix $\alpha 2$ and the $\beta 2\beta 3$ loop in the p44ct C4C4 RING domain (Figure IV.41). The latter inserts itself into a hydrophobic pocket, bringing a leucine (L405) in close proximity to several proline and one tryptophan residue (W80) in the p34ct vWA $\beta 3$ extension (Figure IV.42 A). Next to these the zinc ion bound to Zn I is sandwiched between the $\beta 2\beta 3$ loop and helix $\alpha 2$ in p44ct C4C4 and $\alpha 5$ in p34ct vWA, effectively shielding it from the surrounding solvent. In addition, a prominent lysine (K155) in $\alpha 5$ directly points towards the zinc site and caps it on its top, probably held in position via electrostatic interaction with an aspartic residue (D486) in helix $\alpha 2$ of p44ct C4C4 (Figure IV.42 B and C). Both K155 and D486 are highly conserved among p34 and p44 proteins from different organisms, respectively. Though the lysine is replaced by an arginine moiety in the p34 sequences of human, mouse, *S. cerevisiae* and *A. thaliana* (Figure IV.30), this conservative exchange suggests that these residues play critical roles in complex formation.

Another highly conserved residue in the p44ct C4C4 domain is F490, which was observed to engage in an unusual stacking interaction with an asparagine (N154) directly preceding K155 in helix $\alpha 5$ of p34ct vWA (Figure IV.42 C). While N154 is also present in *S. cerevisiae*, it is replaced by a histidine moiety in both human and mouse p34 proteins (Figure IV.30), the latter of which probably better accommodates the phenylalanine residue in the p44 C4C4 RING domain.

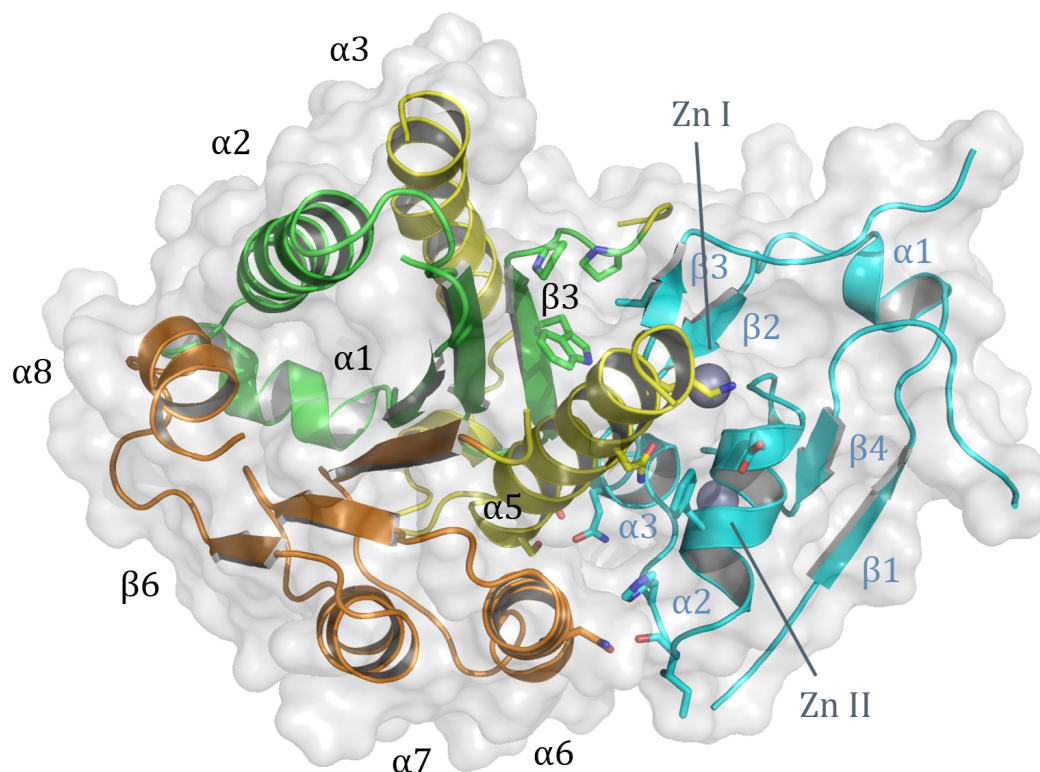


Figure IV.41: Interaction Sites in the p34ct p44ct Minimal Complex (Overview). Top view of the p34ct p44ct MC with the most prominent interacting residues shown in stick representation.

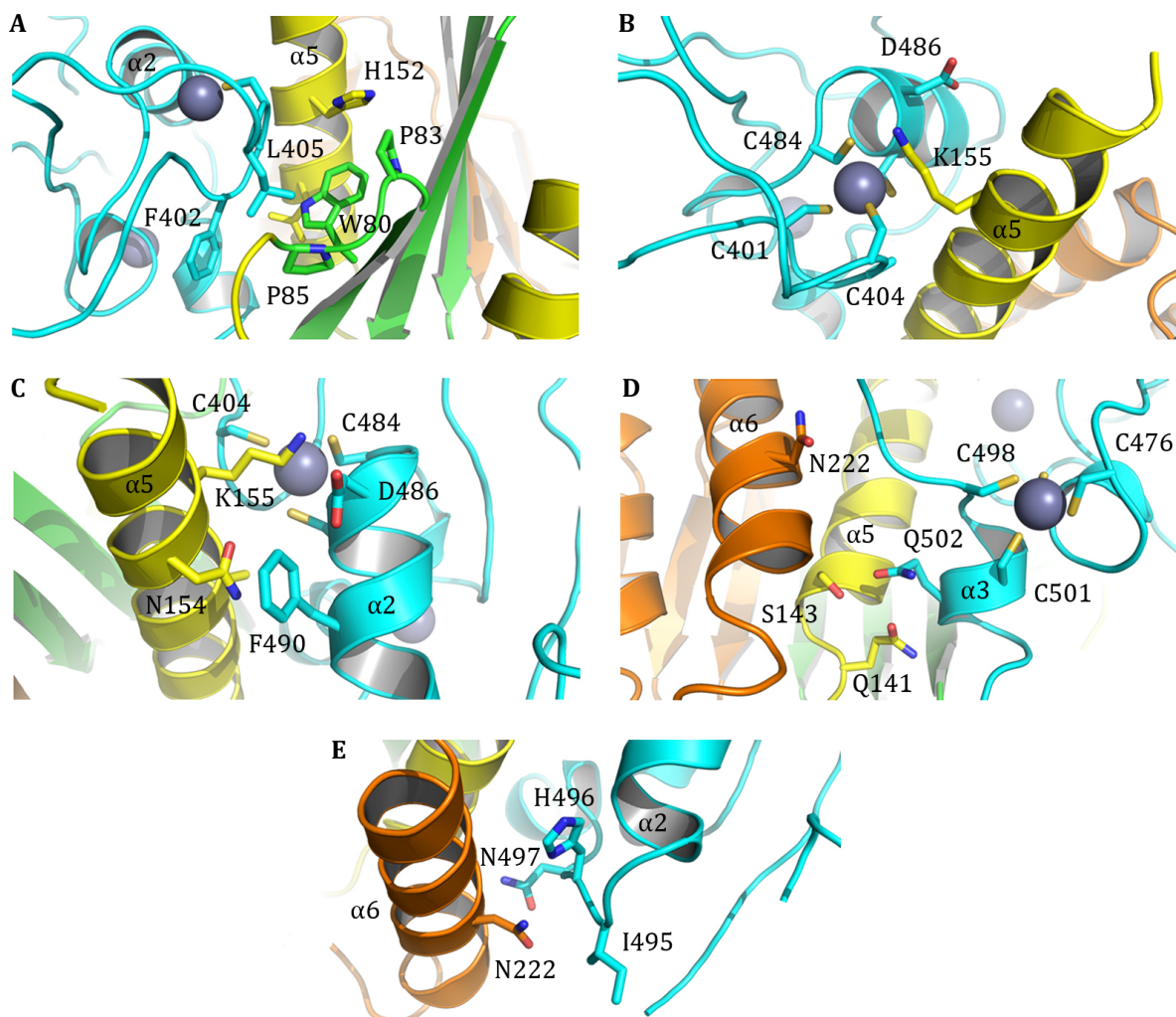


Figure IV.42: Interaction Sites in the p34ct p44ct Minimal Complex (Close Up View). Enlarged are the regions involving contacts of p34ct vWA with the $\beta 2\beta 3$ loop (A) and residues around the first zinc binding site in p44ct C4C4 (B and C) as well as interactions made via helices $\alpha 5$ and $\alpha 6$ in p34ct vWA with $\alpha 1$ (D) and the $\alpha 2\alpha 3$ loop backbone (E) in the p44ct C4C4 RING domain.

Apart from the above mentioned interactions, additional contacts between both proteins are made via the lower portion of the vWA fold, involving residues in helices $\alpha 5$ and $\alpha 6$ of p34ct as well as helix $\alpha 3$ and the $\alpha 2\alpha 3$ connecting loop in p44ct C4C4 (Figure IV.42 D and E). The latter is stabilized via a backbone interaction with an asparagine (N222) in helix $\alpha 6$ of p34ct vWA (Figure IV.42 E), while Q141 and S143 in helix $\alpha 5$ form hydrogen bonding contacts with Q502 in the C-terminal helix $\alpha 3$ of the p44ct C4C4 domain (Figure IV.42 D). The Q141 / S143 interface is especially interesting, as both residues are replaced by purely hydrophobic amino acids in human and mouse p34 proteins (Figure IV.30), likely representing an evolutionary switch in the mode of interaction at this specific site.

In summary several regions in both p34ct vWA and the p44ct C4C4 domain could be identified, harboring a number of hydrophobic as well as polar and charged residues with a high probability of being responsible for a stable complex engagement. Based on these observations, several p34ct and p44ct variants were designed to more closely investigate the critical residues involved (IV.7.2).

IV.7.2 Purification of p34ct and p44ct Variants

To assess the importance of residues located in the contact interface with respect to their contribution towards complex formation and stability, several mutated variants of both p34ct 1-277 and p44ct 368-534 were designed (Table IV.3). In case of the p34ct vWA domain 3 single (A, B, C) as well as 2 dual (D1, D2) and one multi variant (M1) were generated, of which only B, D2 and M1 could be purified (Figure IV.43). The others showed extremely low expression levels or were mostly insoluble (data not shown). Interestingly, all failed variants contained mutations to alanine, whereas highly acidic residues were introduced in the successful variants. The exception is the multi variant M1, which contained also the N222A mutation (Table IV.3). However, this variant also showed a stronger tendency to form higher oligomers. Hence, the increase in overall hydrophobicity of p34ct 1-277 by mutation of single residues to alanine presumably rendered the vWA domain more prone to aggregation.

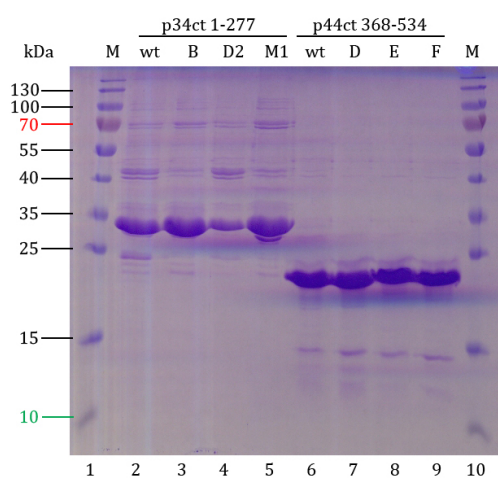


Table IV.3: Mutants of p34ct 1-277 and p44ct 368-534.

Protein	Mutated Residues	Code
p34ct 1-277	W80A	A
p34ct 1-277	K155E	B
p34ct 1-277	N222A	C
p34ct 1-277	K155E, W80A	D1
p34ct 1-277	Q141E, S143D	D2
p34ct 1-277	K155E, N222A, Q141E, S143D	M1
p44ct 368-534	L405E	D
p44ct 368-534	F490E	E
p44ct 368-534	Q502E	F

Figure IV.43: Purified Variants of p34ct 1-277 and p44ct 368-534. Of all the variants designed only p34ct 1-277 B, D2 and M1 as well as p44ct 368-534 D, E and F could be purified (Table IV.3). In terms of quantity and purity the analysis via SDS-PAGE yielded results comparable to the wild type proteins (wt).

Regarding the p44ct C4C4 RING domain all of the 3 single variants designed (D, E, F) could be purified essentially like the wild type protein, with moderate yields and good purity (Figure IV.43). The latter was often a general issue with respect to the p34ct 1-277 purification, for both wildtype and two of its variants. Here, some impurities were always present that could not be removed completely from the samples. This was mainly attributed to the far lower expression level of p34ct 1-277 adversely affecting the Ni-TED affinity chromatography step during purification (see IV.2.1). Interestingly, the p34ct 1-277 K155E variant generally showed much better expression levels and significantly cleaner purification profiles, ultimately supporting the previous hypothesis (Figure IV.43).

Taken together a total of 6 mutated variants of the p34ct vWA and p44ct C4C4 domains could be obtained with results comparable to or even better than their wild type counterparts. In the following these variants were characterized (IV.7.3) and analyzed with respect to their ability to mediate complex formation, either by combining them with their respective wild type counterpart or an appropriate mutated variant (IV.7.4).

IV.7.3 Characterization of p34ct and p44ct Variants

Prior to interaction studies each variant was analyzed for its thermal stability and proper folding in solution. While Thermofluor measurements of all three p34ct 1-277 variants yielded unfolding profiles comparable to the wild type protein, with only a marginal decrease in melting temperature for D2 and M1 (Figure IV.44 A), more significant effects were observed with p44ct 368-534. Here the F490E (E) and Q502E (F) substitutions seemed to adversely affect the overall stability of p44ct C4C4, shifting its melting temperature to values lying 3 °C and 8 °C below that of the wild type protein, respectively (Figure IV.44 B).

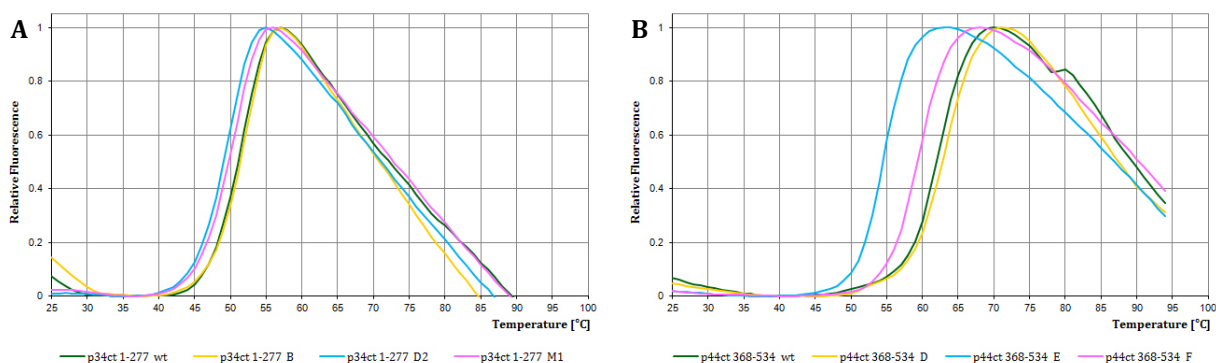


Figure IV.44: Thermofluor Analysis of p34ct and p44ct Variants. Shown are the thermal unfolding curves of purified p34ct 1-277 variants (A) in 100 mM CHES pH 9.5 and p44ct 368-534 variants (B) in reference buffer compared to the wild type proteins (wt). The reference buffer used for purification and storage consisted of 20 mM CHES pH 9.5 (A) or 20 mM Tris-HCl (B), 150 mM KCl and 1 mM TCEP.

The reduction in thermal stability for the Q502E variant could result from the disruption of a putative hydrogen bond with an asparagine located 5 residues upstream. The latter is located in a loop region between helices $\alpha 2$ and $\alpha 3$ and might be involved in fixing the position of C498 and preserving the fold of the second zinc binding site (see Figure IV.42 D). The even more significant decrease in melting temperature observed for the F490E variant could in turn be related to repulsive forces between the introduced glutamic acid residue and three negatively charged residues (D486, D488 and E493) already present in helix $\alpha 2$ of the p44ct C4C4 domain (see Figure IV.42 C). Two of these are located in close spatial proximity to F490 and are likely to be affected by the switch in electrostatic potential upon mutation.

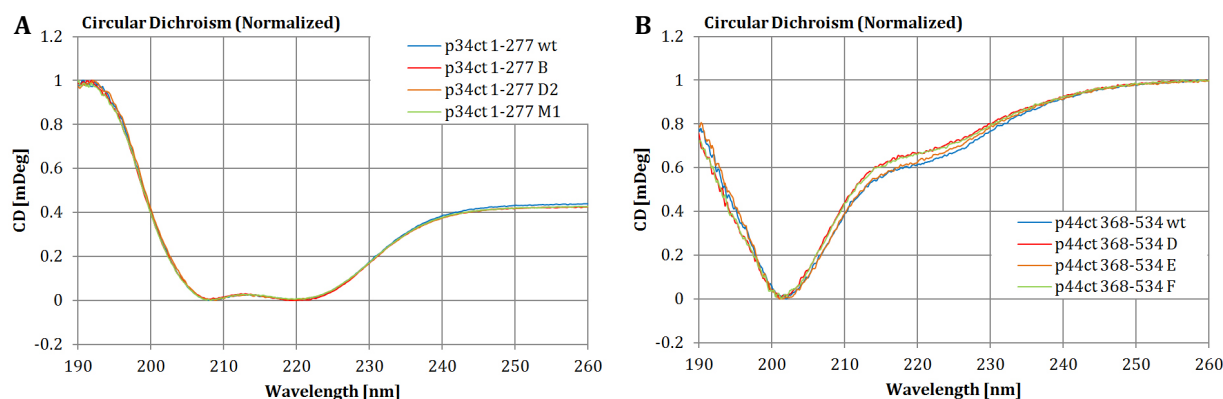


Figure IV.45: Circular Dichroism Spectra of p34ct and p44ct Variants. Shown are the CD spectra of all purified p34ct 1-277 variants (A) and p44ct 368-534 variants (B) compared to the CD spectra of their respective wild type proteins (wt). All spectra were recorded at 20 °C in 20 mM $\text{KH}_2/\text{K}_2\text{H-PO}_4$ at pH 8.1.

While the thermal stability of the p44ct C4C4 domain is clearly reduced in two of the variants, analysis of their folding in solution via circular dichroism spectroscopy (III.4.3) yields comparable results for all three variants and the wild type protein (Figure IV.45 B). Though the CD profiles for p44ct 368-534 in general are indicative of a predominant random-coil character (pronounced minimum at 200 nm) and little to no higher order folding, this finding is not at all unexpected. Due to the small size of the domain and its low amount of secondary structure elements (see Figure IV.38), the signals obtained from CD spectra for α -helices and β -strands are scarce. This is especially true for the *C. thermophilum* C4C4 domain, which contains a highly flexible 50 - 55 amino acid linker insertion, compared to its human counterpart (IV.6.2 and Figure IV.38). In contrast, the p34ct vWA domain and all of its purified variants exhibit the typical CD profile of a protein perfectly folded in solution (Figure IV.45 A). All spectra recorded are practically alike, displaying a balanced frequency of α -helices (minima at 208 nm and 222 nm) and β -strands (effective minimum at 215 nm).

IV.7.4 Determination of Critical Residues

To evaluate the importance of each of the mutated residues in p34ct vWA and p44ct C4C4 with respect to complex formation, qualitative interaction studies were performed, using either electrophoretic mobility shift assays (EMSAs) via native agarose gel-electrophoresis (NAGE, III.3.6) or semi-quantitative analytical size exclusion chromatography (A-SEC, III.3.5).

As expected, in the case of all NAGE experiments, a strong interaction could consistently be detected for the wild type proteins, resulting in a significant mobility shift of the complex formed (Figure IV.46, PQ, line 3). However, depending on pH milieu and strength of the gel, both p34ct 1-277 and p44ct 368-534 alone were severely retarded and barely entered the gel. In addition, only minor differences in the overall mobility of the two proteins were observable, which is likely due to the isoelectric points (pI) of both p34ct 1-277 (pI = 6.37) and p44ct 368-534 (pI = 7.13) lying closely together. Though considerable effort has been put in optimizing the NAGE running behavior these drawbacks made the evaluation of smaller shifts for some of the p34ct vWA and p44ct C4C4 variants more difficult.

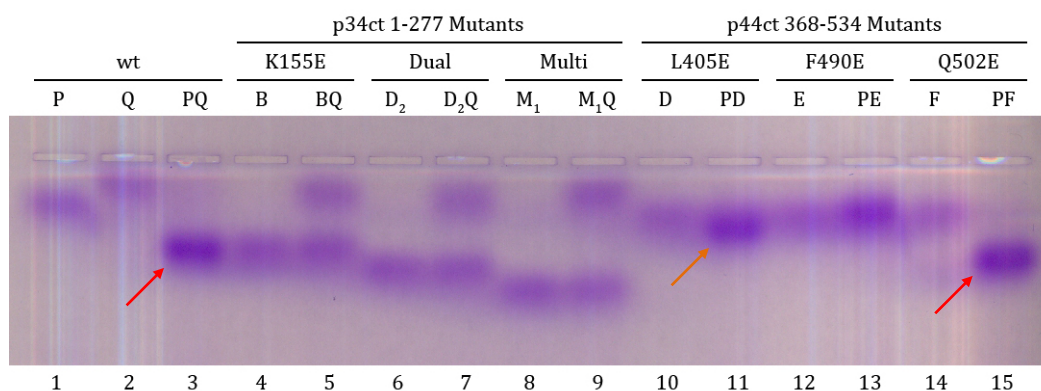


Figure IV.46: Interaction Studies using Native Agarose Gel-Electrophoresis. Wild type p34ct 1-277 vWA (P) and p44ct 368-534 C4C4 (Q) were incubated with their mutated counterpart (p34ct B, D2, M1 or p44ct D, E, F; see Table IV.3), separated via NAGE and visualized using standard Coomassie staining. Observed complex formation is indicated by the colored arrows (strong: red; weak: orange). The final concentration of each protein used in the reaction mixture was 20 μ M.

While the formation of the wild type complex perfectly depleted the control-bands for the single proteins (Figure IV.46, PQ, lanes 1-3) all tested p34ct 1-277 variants were deficient towards complex formation, displaying only the non interacting single proteins (Figure IV.46, lanes 2 and 4-9). However, when analyzing the p44ct C4C4 variants an interaction could be observed with respect to the Q502E and the L405E substitutions (Figure IV.46, lanes 1, 10-11 and 14-15). The strength of the interaction suggests that Q502E alone does not interfere with binding, while L405E at least impedes binding to some degree. No interaction was observed for the p44ct F490E mutation, which verifies the importance of this conserved residue as similarly shown for the human p44 C4C4 domain [68].

Most of the observations made using the NAGE approach are in perfect agreement with the complementary A-SEC experiments performed in parallel. Here, again a strong interaction could be visualized with the wild type control, marking a significant shift of the complex peak to earlier elution volumes (Figure IV.47 A), while both K155E and the multi mutant M1 seemed to completely disrupt complex formation (Figure IV.47 B, D). However, in contrast to the EMSA studies, the dual variant D2 indicated a weak interaction under the conditions used for the A-SEC analysis (Figure IV.47 C). Hence the Q141E and S143D dual substitution alone

was not sufficient to completely disrupt the p34ct p44ct interaction. To further investigate this binding site a combination of D2 with the Q502E mutation located in p44ct C4C4 (see Figure IV.42 D) was tested, which on its own does not interfere with complex formation (Figure IV.48 D). With both D2 and Q502E present, binding is completely abolished, emphasizing the importance of this contact interface for the formation of the p34ct p44ct minimal complex (Figure IV.48 C).

While both wild type p34ct 1-277 as well as the K155E and Q141E S143D dual variant eluted in a single and homogenous peak during A-SEC experiments, the multi variant comprising K155E, N222A, Q141E and S143D showed an equilibrium between the monomer and a higher oligomeric state of the protein (Figure IV.47 D). As mentioned above (IV.7.2), this tendency to self-associate might be related to an increase in the overall hydrophobicity of p34ct 1-277 by introduction of an alanine residue. This is supported by the finding that the N222A substitution alone rendered the p34ct vWA domain completely insoluble (IV.7.2). Interestingly, however, the double peak observable with M1 is reduced to monomers when the p44ct C4C4 wild type is combined with M1 (Figure IV.47 D). Though the reason for this remains obscure, one could speculate that either the minor increase in osmolarity or the still present but significantly reduced and undetectable affinity towards the p44ct C4C4 domain is sufficient to prevent oligomerization of the p34ct vWA domain in the presence of its natural binding partner.

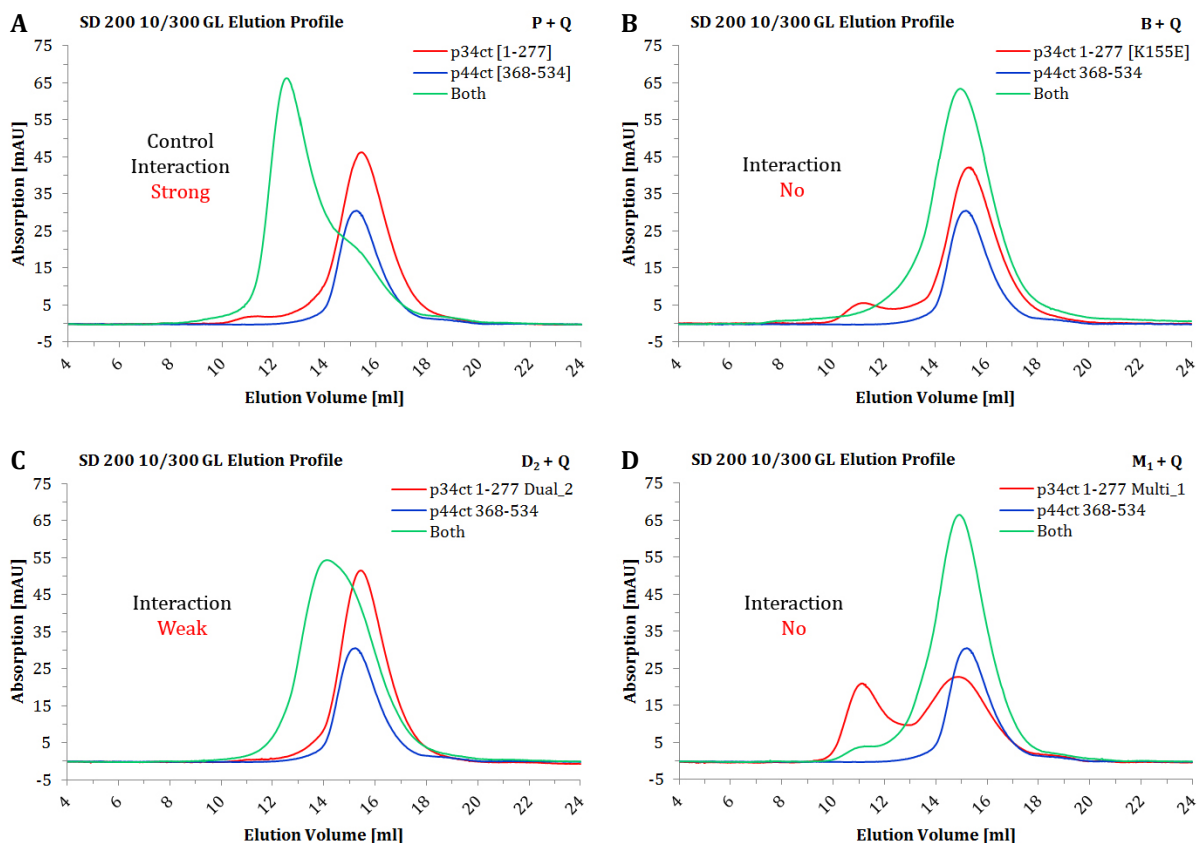


Figure IV.47: Interaction Studies with p34ct Variants and p44ct Wild Type using Analytical SEC. Shown are the elution profiles of the single proteins (red, blue) and both proteins combined (green). The wild type interaction (A) is compared to 3 different p34ct 1-277 variants (B - D, see Table IV.3). A shift of the elution profile to the left is indicative of complex formation. The separation was performed via an SD 200 10/300 GL column using final protein concentrations of 10 - 12 μ M.

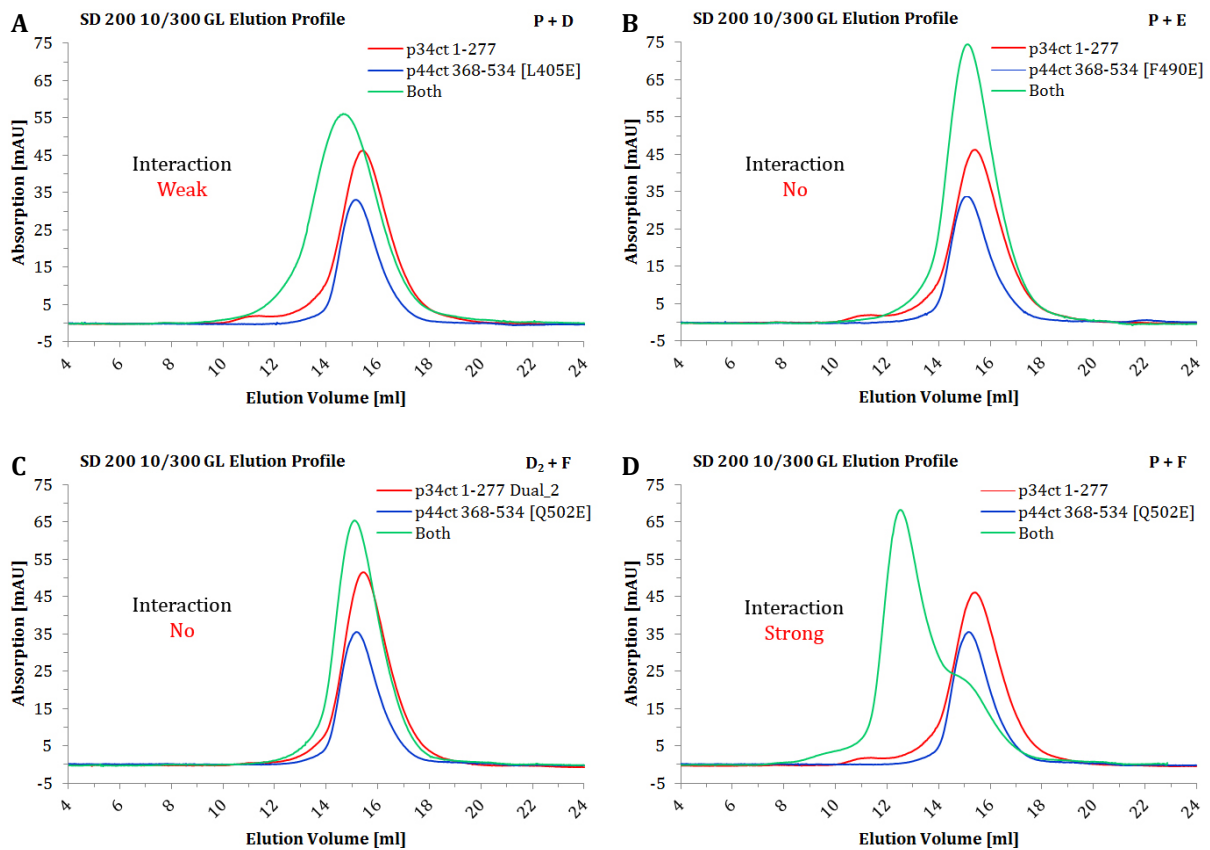


Figure IV.48: Interaction Studies with p34ct Wild Type and p44ct Variants using Analytical SEC. Shown are the elution profiles of the single proteins (red, blue) and both proteins combined (green). The wild type interaction (see Figure IV.47 A) can be compared to 3 different p44ct variants (A, B and D, see Table IV.3) and a combination of two p34ct and p44ct mutations located in close proximity to each other (C, see Table IV.3). A shift of the profile to the left is indicative of complex formation. The separation was performed via an SD 200 10/300 GL column using final protein concentrations of 10 - 12 μ M.

Though the obvious but weak interaction of D2 with p44ct 368-534 (Figure IV.47 C) was not detected during NAGE, the observed reduced affinity of p44ct L405E towards the p34ct vWA domain (Figure IV.46) could be verified using the A-SEC analysis (Figure IV.48 A). Furthermore, the F490E substitution in helix α 2 of the p44ct C4C4 domain completely abolished binding in both experiments (Figure IV.46 and Figure IV.48 B), mutually supporting the critical relevance of this single residue for complex formation.

Taken together, all of the mutated residues investigated, except for p44ct 368-534 Q502E alone, were found to influence complex formation of p34ct p44ct MC to a variable degree. The markedly reduced affinities with p34ct Q141E S143D (D2, Figure IV.47 C) and p44ct L405E (Figure IV.48 A) clearly demonstrate that both the interface at the lower portion of helix α 5 in p34ct vWA (Figure IV.42 D) as well as the hydrophobic contacts mediated via the β 2 β 3 loop in p44ct C4C4 (Figure IV.42 A) are required to stabilize the interaction. The most critical residues identified, however, were p34ct 1-277 K155E and p44ct 368-534 F490E, which completely abolished complex formation on their own when either one of these was mutated. As both these residues are located in close proximity to the first zinc binding site (Zn I) in the p44ct C4C4 domain (Figure IV.42 B and C), this area seems to constitute a central and essential interface in the p34ct p44ct minimal complex. Mutations of single residues disturbing the local and conserved fold around Zn I are tantamount to a loss of interaction, in turn emphasizing the importance of contacts made via this binding site [68]. The weaker interactions in the three flanking regions (on both sides and below Zn I) presumably serve to stabilize this central contact.

IV.7.5 Determination of Binding Affinities

With the most critical residues mediating the p34ct p44ct MC interaction resolved (IV.7.4), the final challenge was to determine binding affinities for both the wild type and the mutated variants. Initial experiments using isothermal titration calorimetry (ITC, III.3.7) were unsuccessful, due to the relatively low protein yields for p34ct 1-277, very low enthalpy signals and precipitation of the complex as a result of lower temperature and stirring in the measuring cell. However, by using the more sensitive VP-ITC instrument, slower stirring and temperatures in the range of 30 - 37 °C, consistent and reproducible binding curves could be recorded for the wild type proteins, allowing in turn to quantify the interaction.

For the p34ct p44ct wild type complex, several measurements, using different batches of purified proteins, repeatedly yielded dissociation constants (K_d) in the range of $K_d = 7.3 - 13.1$ nM, with stoichiometry (N) ratios generally indicating a 1:1 or 2:2 interaction, respectively (Figure IV.49 A). The latter were often observed to be slightly off, likely due to the faint but inherently present impurities with all p34ct 1-277 samples (IV.2.1).

Nevertheless the consistently determined K_d values in the nanomolar range are indicative of a very high affinity and tight complex formation, which is in perfect agreement with all previous observations made regarding the p34ct p44ct minimal complex (compare IV.2.3, IV.2.6, IV.6.4 and IV.7.4).

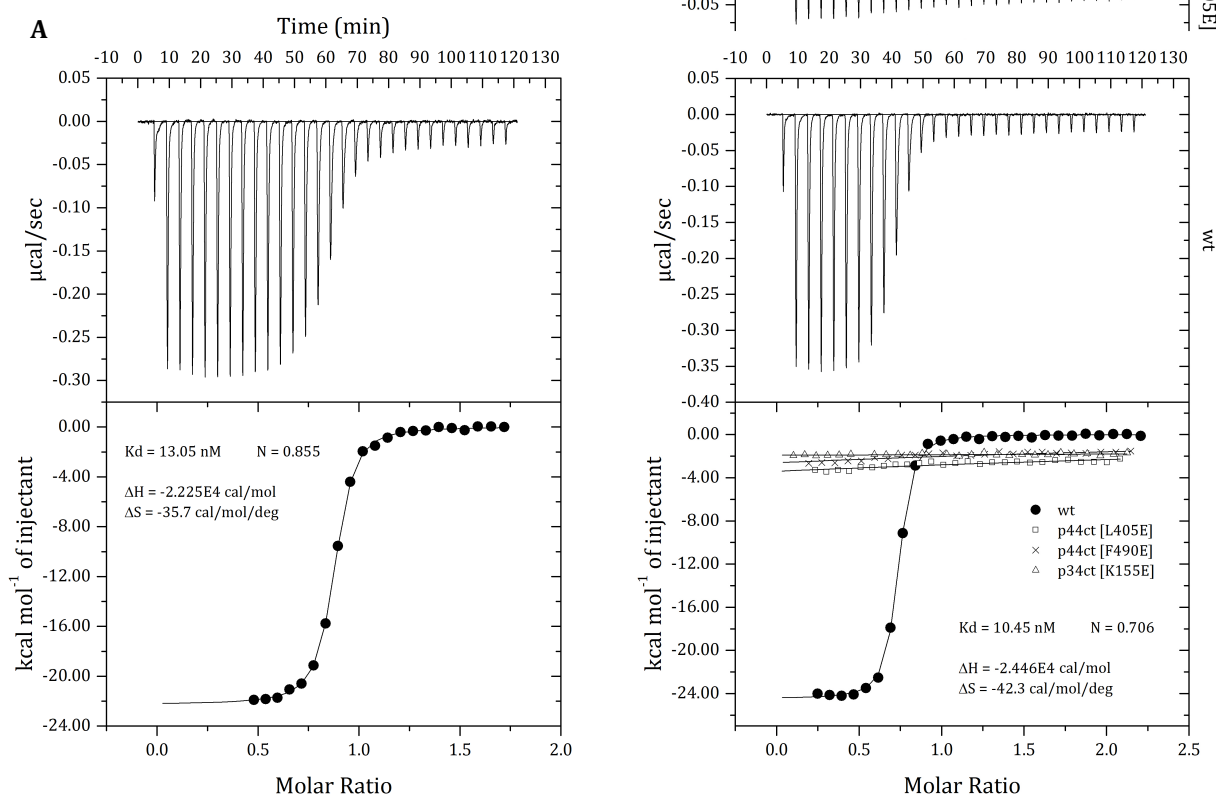


Figure IV.49: Interaction Studies with p34ct and p44ct Proteins using ITC Measurements. Shown are the isothermal titration curves (top: raw enthalpy, bottom: fit of data) for the interaction of wild type p34ct 1-277 and p44ct 368-534 (A). In addition, measurements with one p34ct 1-277 (K155E) and two p44ct 368-534 mutants (F490E, L405E) are compared to another wild type interaction (B). For clarity, the enthalpy plots of each mutant are shown at an equal scale above the graph for the wild type control (B). In each experiment 40 - 48 μ M of p44ct 368-534 were added incrementally to 4 - 5 μ M of p34ct 1-277.

In contrast, identical experiments with p34ct 1-277 K155E as well as p44ct 368-534 F490E and L405E yielded practically no detectable titration curves, indicative of a significantly reduced affinity with these variants (Figure IV.49 B). The effect was observed to be more pronounced going from L405E via F490E to K155E, the latter of which showed no detectable enthalpy signature at all, except for the residual background noise, yielding faint signals of equal intensity throughout the course of the experiment (Figure IV.49 B). However, no distinct binding affinity could be calculated for any of the mutated variants tested at the protein concentrations used to carry out the measurements.

Taken together the ITC binding studies nicely support the observations made with respect to the EMSA and A-SEC analysis (IV.7.4). While the wild type complex exhibits a tight interaction with high affinity, yielding K_d values around 10 nM (Figure IV.49 A), the critical variants p34ct 1-277 K155E and p44ct 368-534 F490E completely abolish binding, with their dissociation constants not being determinable (Figure IV.49 B). Similarly p44ct 368-534 L405E shows a significantly reduced affinity towards p34ct 1-277. Yet the disruptive effect is not as pronounced as with p34ct 1-277 K155E and the underlying albeit lower K_d values might be determinable upon a significant increase in the concentrations used during the experiment.

V. Conclusion and Outlook

V.1 The TFIIH Subunits p34 and p44 from *C. thermophilum*

In the present work the TFIIH subunits p34 and p44 were cloned from the genomic DNA of the eukaryotic thermophile *C. thermophilum* (ct), a remarkable fungus, which can thrive at temperatures of up to 60 °C. Both proteins as well as several shortened variants were successfully expressed in *E. coli* and subsequently purified to homogeneity without observable degradation or solubility issues (IV.2.1 to IV.2.3), whereas most of the human TFIIH subunits tested were either insoluble in *E. coli* and insect cells (IV.1.1) or prone to degradation (IV.1.2). The purified p34ct and p44ct variants were highly thermostable (IV.2.5) and found to be monomers in solution (IV.2.6), except for the p44ct full-length protein, which was detected to form higher oligomers during purification (IV.2.2). The latter could be attributed to the central zinc finger domain of the protein, as both the N-terminal and C-terminal domains alone did not exhibit this behavior (IV.2.2).

In addition to the improved properties in solution, as compared to their mesophilic homologues, p34ct also readily yielded protein crystals, which, after extensive optimization, were suitable for x-ray crystallographic analysis (IV.5). In contrast, the full-length p44ct subunit of TFIIH generally eluded crystallization, while shorter variants including only the N-terminal region of the protein or its vWA domain, respectively, resulted in crystalline aggregates and a few needle or platelet clusters (IV.4). The latter, though still resisting optimization attempts, seemed to improve with progressive shortening of the construct borders. Hence, it is not unlikely that further refinement of the p44ct N-terminal domain boundaries will result in reproducible protein crystals that could allow to determine the structure of the p44ct vWA domain as well.

Taken together the p34 and p44 proteins from *C. thermophilum* nicely demonstrate the applicability of this novel model organism to provide for highly stable eukaryotic TFIIH subunits that are suitable for biochemical as well as x-ray crystallographic studies. Further experiments, also employing other TFIIH core and possibly CAK subunits from *C. thermophilum*, will certainly advance our understanding of the molecular structure and function of the TFIIH complex in the context of transcription, DNA repair and beyond.

V.2 The p34ct vWA Domain

Structural analysis of the protein crystals grown from samples of the p34ct subunit of TFIIH revealed the presence of a novel *von Willebrand Factor A* (vWA) like domain at the N-terminus of the protein, which has previously not been reported. The p34ct vWA structure assumes a compact fold, comprising a 6-stranded β -sheet, which is surrounded by 3 α -helices on either side. Two additional, shorter α -helices are located at the C-terminal end of the central β -sheet and are usually replaced by loops in many other vWA domain proteins. These differences also result in the disruption of a putative MIDAS motif, which would otherwise enable the protein to capture a divalent metal ion, as seen, for example, in the vWA domains of several integrins.

The vWA domain in general is typically involved in mediating protein-protein interactions and the p34ct vWA structure itself was found to be highly similar to the *von Willebrand Factor A1* domain and the vWA domain of Rpn10, both of which are known to contact several binding partners (IV.5.6 and IV.5.7). The latter is a subunit of the 26S proteasome regulatory particle and a good example of a vWA domain engaged in multiple protein-protein interactions, while concomitantly serving additional functions, which are mediated via regions in its C-terminal domain. Interestingly, also the p44 subunit of TFIIH contains a vWA like domain at its N-terminus, which is closely related to that of Rpn10, based on phylogenetic studies [188,207]. Hence, it is conceivable that the roles of p34 and p44 within TFIIH are comparable to the structural and accessory functions of Rpn10 in the 26S proteasome regulatory particle (IV.5.7).

In support of this notion, the p34 and p44 subunits of TFIIH are highly similar with respect to their primary domain architecture. Both proteins contain an N-terminal vWA like domain, which is followed by a C4-type zinc finger motif. In addition, p44 contains a C-terminal extension, comprising a C4C4 RING domain, which is known to form a tight complex with the p34 vWA domain, directly linking both proteins within TFIIH [42]. While the p44 subunit is a positive regulator of the XPD helicase, its interaction with p34 might result in the formation of a regulatory triad, which closely monitors and directs the helicase activity of XPD with respect to other sensory inputs. The latter could be mediated through additional and putatively transient protein-protein interactions of p34 and p44 with other TFIIH subunits or associated proteins, either encompassing their vWA domains or zinc finger motifs, respectively.

In terms of TFIIH integrity the vWA domains of p34 and p44 might serve as a general scaffold and stabilize the interaction of several subunits whose affinity for each other would normally be too low. As similarly seen with the *von Willebrand Factor A1* / GP-Ib- α / Botrocetin complex (IV.5.7), the vWA domain could potentially accommodate two binding partners and thereby serve as a mediator. By binding both proteins and bringing them in close proximity, additional contacts might be established, which would further stabilize the assembly.

V.3 The p34ct p44ct Minimal Complex

In addition to the p34ct vWA domain also a structure of the p34ct p44ct minimal complex was determined at a resolution of 3.8 Å (IV.6). The structure comprises the vWA domain of p34ct and the C-terminal C4C4 RING domain of p44ct and allowed for the first time to examine the binding interface between both proteins and delineate their tight mode of interaction (IV.7).

The p44ct C4C4 RING domain was found to contact p34ct at a mainly hydrophobic and slightly basic region, involving α -helices α_5 , α_6 and the extension of β -strand β_3 in the p34ct vWA domain as well as α -helix α_2 and the $\beta_2\beta_3$ loop in the p44ct C4C4 domain (Figure IV.37). The latter inserts itself into a largely hydrophobic pocket in the vWA fold, tightly locking both proteins and bringing the first zinc binding site in p44ct C4C4 in close proximity to helix α_5 in the p34ct vWA domain, effectively shielding the zinc ion from the surrounding solvent. Several residues in the p34ct vWA domain, like K155, which points towards the p44ct C4C4 RING domain and caps the first zinc binding site, seem to further enhance this effect.

A more detailed analysis of the interaction interface then made it possible to decipher the critical residues involved in binding, which were subsequently mutated and tested for their significance in mediating the formation of a tight p34 p44 complex. In this regard, both the p34ct K155E and p44ct F490E mutations were found to completely abolish binding during analytical size exclusion chromatography and electrophoretic mobility shift assays, while p44ct L405E led to a reduced affinity (IV.7.4). These amino acids are located in the interface formed by the p34ct vWA domain and the first p44ct zinc binding site, including the $\beta_2\beta_3$ loop, in the p44ct C4C4 RING domain, indicative of this interface being essential for complex formation.

Additional accessory contacts, which seem to stabilize the primary interface, were found in the lower portion of the vWA fold, involving residues in helices α_5 and α_6 in p34ct vWA as well as α -helix α_3 and the $\alpha_2\alpha_3$ connecting loop in the p44ct C4C4 RING domain (IV.7.1). While certain residues in these regions evidently weakened the p34 p44 interaction when they were mutated, the combination of a p34ct Q141E S143D dual variant and the p44ct Q502E substitution abolished complex formation (IV.7.4).

To quantify the p34 p44 interaction, binding affinities were determined via isothermal titration calorimetry. The measurements reproducibly yielded dissociation constants in the range of 10 - 13 nM for the wild type complex, clearly demonstrating the strength of this particular vWA and C4C4 RING domain mediated protein-protein contact (IV.7.5). In addition, the disruptive effects observed with the p34ct K155E as well as the p44ct F490E and L405E variants could be reproduced and verified in the course of these experiments, even though no exact binding affinities for these variants could be determined under the conditions tested (IV.7.5).

V.4 Implications from the p34ct p44ct Minimal Complex

Apart from forming a tight structural assembly, the structure of the p34ct p44ct minimal complex revealed no additional clues as to the actual purpose of the p34 p44 interaction within the scope of the TFIID complex. Compared to the structure of the p34ct vWA domain alone (IV.5.4) the p34ct p44ct minimal complex did not result in any conformational changes in the vWA fold, at least not in those regions that can be observed in the calculated electron density maps. However, considering the spacious linker regions in the p34ct vWA and the p44ct C4C4 RING domains, which are disordered in both structures, it cannot be excluded that these regions serve as mediator or signal transducer with respect to the impact the p34 p44 interaction might have on other TFIID subunits. In addition, subtle differences in the overall fold of the p44ct C4C4 RING domain compared to its human counterpart (IV.6.2), could mark structural changes induced upon p34ct vWA binding, though the low resolution of the collected data hampered a more detailed investigation in this direction.

In this regard it is interesting to note that the p34ct p44ct minimal complex assumed a hetero-tetrameric assembly in solution, while under the same conditions both the p34ct vWA and p44ct C4C4 RING domain alone were found to be monomers (IV.2.6). However, even though the true oligomeric state of each of the TFIID subunits still has to be determined and a hetero-tetrameric p34 p44 complex *in vivo* is a distinct possibility, the physiological relevance of such a higher order complex is unclear. Indeed, based on a thorough analysis of the p34ct p44ct MC crystal packing, the hetero-tetramer could also be a result of an artificial dimerization of the p34ct vWA domain (IV.6.4). Upon binding of the p44ct C4C4 RING domain, the shielding of a large portion of the p34ct vWA fold might result in a shift of the equilibrium towards a p34ct vWA dimer formation, which in turn seems to be driven exclusively by hydrophobic effects.

Though putatively artificial in nature, the p34ct vWA dimerization resulting in a p34ct p44ct hetero-tetramer inevitably raises the question if the vWA domain is occupied by another binding partner, in addition to the p44ct C4C4 RING domain, in the context of the TFIID complex *in vivo*. Together with the findings made with respect to other vWA folds, especially those of the structurally highly similar *von Willebrand Factor A1* domain and the 26S proteasomal regulatory subunit Rpn10 (IV.5.7), an engagement of the p34ct vWA domain in multiple interactions is not unlikely. It will be very interesting to further investigate the p34 p44 complex with respect to additional contacts with other TFIID subunits that might play a role in regulating the intricate enzymatic function of TFIID in the context of transcription, DNA repair and other cellular processes.

In this regard the dissection of the TFIID interaction network, by purification of other TFIID subunits and performing combinatorial interaction studies *in vitro*, will help to verify already known associations and discover additional binding partners, which have so far eluded detection. Using the p34 p44 minimal complex as a starting point further assemblies can then be identified, including additional subunits or interacting domains thereof, which could be similarly stable and suitable for crystallographic analysis. With respect to improving the overall quality of potential protein crystals, regions that are not involved in the interaction, or which are highly flexible and protruding outwards from an otherwise compact fold, could be removed. In this way the p34ct p44ct minimal complex could be re-designed to allow for a better crystal packing and hence higher resolution data, in turn providing a better starting model for further studies.

V.5 The Role of the p34ct and p44ct Zinc Finger Domains

Apart from the vWA like fold at the N-terminus of p34ct and p44ct and the C-terminal C4C4 RING domain exclusive to p44, both proteins contain an additional C-terminal (p34) or central (p44) zinc finger motif. Zinc finger motifs are relatively small protein folds, which coordinate a zinc ion and often occur in clusters of several repeats. They were initially identified as DNA binding regions in the transcription factor III A of the African clawed frog *Xenopus laevis* [216], but have subsequently also been shown to mediate RNA binding as well as protein-protein interactions [217–219]. Zinc finger domains usually coordinate metal ions via a combination of four cysteine and/or histidine residues, forming finger-like protrusions that can make tandem contacts with their target molecule. Based on the nature and arrangement of their zinc-coordinating residues, they can be grouped into more than 14 classes, each serving a multitude of different biological functions, which in turn emphasizes the versatility of the zinc finger fold [220].

Both the C-terminal zinc finger in p34 as well as the central zinc binding domain in p44 belong to the C4-type of zinc finger motifs, which coordinate one zinc ion via a highly conserved sequence of four cysteine residues in a C-x-x-C ... C-x-x-C arrangement. While the closely related C4C4 RING domain of p44 has long been known to tightly interact with the p34 subunit of TFIIH, the precise role of the zinc finger domains in each of these binding partners still remains elusive. For p44 it has been suggested that the zinc finger interacts with the p62 subunit of TFIIH, based on the finding that TFIIH lacked p62 when cysteine residues in the central zinc finger motif of p44 were mutated to alanines [208]. In contrast, nothing at all is known with respect to the C-terminal zinc finger domain in p34.

Given the DNA and RNA binding properties of many zinc finger proteins and the involvement of TFIIH in both transcription and nucleotide excision repair, it can be hypothesized that the zinc finger motifs in p34 and p44 are involved in mediating the interaction of TFIIH with its DNA substrate. In support of this notion, full-length p44ct showed a weak but positive DNA binding capability for double stranded DNA, while both its N-terminal and C-terminal domains alone did not (Figure IV.15). In contrast, the fact that neither full-length p34ct nor p34ct 1-277 were able to bind to the same DNA substrate would rule against a general DNA binding property inferred by the p34 zinc finger. However, this does not necessarily mean that the protein would not serve a similar function within TFIIH *in vivo*. As many zinc finger proteins with DNA binding properties are known to contact their DNA substrate via tandem contacts made by multiple zinc finger motifs [220], it is conceivable that p34 and p44 alone provide only parts of the overall DNA binding capability within TFIIH and that both proteins together could have a synergistic effect, in turn increasing the affinity towards the DNA substrate. More studies will be required to determine the exact role of the p34 and p44 zinc finger domains and their influence on TFIIH regulation and function.

V.6 The Importance of TFIIH Research

To unravel the molecular structure and function of the TFIIH complex has been a challenge since its discovery more than 20 years ago. Originally identified as a general transcription factor, TFIIH has soon been found to assume additional roles during nucleotide excision repair and cell cycle control, emphasizing the complexity and dynamics of this 10-subunit assembly [27]. All these processes are essential for each and every organism and provide for two central characteristics of cellular life: survival and replication. Hence, what we can learn about TFIIH will greatly expand our knowledge with respect to life in its essence.

In addition to these fundamental aspects, one can also use the knowledge gained for the benefit of mankind, especially in the context of severe diseases that affect our species. In this regard, single point mutations in the TFIIH subunits XPB, XPD and p8 are known to be causative for xeroderma pigmentosum (XP), Cockayne syndrome (CS) and trichothiodystrophy (TTD), three genetic disorders that severely compromise their host and usually lead to an early death [30–32]. Due to the complexity of these diseases, in turn reflecting the intricate entanglement of TFIIH in various processes, there is, to date, no viable cure. A better understanding of the underlying mechanisms with respect to TFIIH function, will surely help to fathom the cellular cause of these and similar afflictions and may provide the basis for an improved treatment in the long run.

Though a lot has been learned on a genetic and biochemical level, so far the structural information on TFIIH, especially its 7-subunit core, is still limited (I.3.2). In this regard, the structures of the p34ct vWA domain and the p34ct p44ct minimal complex (MC) presented in this work, appreciably contribute towards a detailed and complete structural characterization of the TFIIH complex. In fact, together with an even smaller minimal complex structure of the yeast homologue of p8 (Tfb5) bound to the C-terminal domain of p52 (Tfb2) [47,48], the p34ct p44ct MC represents the second structure only, that reveals the interaction of two TFIIH subunits at an atomic level. Hence, we are barely scratching the surface with respect to a complete understanding of the molecular structure, subunit interplay and regulation within the TFIIH complex, which would provide us both with a mastered research challenge and an invaluable asset.

VI. Literature References

- [1] **Hoeijmakers, J. H.**, Genome maintenance mechanisms for preventing cancer, *Nature* 411(6835), 366–374, 2001.
- [2] **Hoeijmakers, J. H.**, DNA repair mechanisms, *Maturitas* 38(1), 17-22; discussion 22-3, 2001.
- [3] **Schärer, O. D.**, Chemistry and biology of DNA repair, *Angewandte Chemie (International ed. in English)* 42(26), 2946–2974, 2003.
- [4] **Wang, F.**, Yang, W., Structural insight into translesion synthesis by DNA Pol II, *Cell* 139(7), 1279–1289, 2009.
- [5] **Sancar, A.**, Mechanisms of DNA excision repair, *Science (New York, N.Y.)* 266(5193), 1954–1956, 1994.
- [6] **Kisker, C.**, Kuper, J., van Houten, B., Prokaryotic nucleotide excision repair, *Cold Spring Harbor perspectives in biology* 5(3), a012591, 2013.
- [7] **Gillet, L. C. J.**, Schärer, O. D., Molecular mechanisms of mammalian global genome nucleotide excision repair, *Chemical reviews* 106(2), 253–276, 2006.
- [8] **Hanawalt, P. C.**, Spivak, G., Transcription-coupled DNA repair: two decades of progress and surprises, *Nature reviews. Molecular cell biology* 9(12), 958–970, 2008.
- [9] **Araújo, S. J.**, Tirode, F., Coin, F., Pospiech, H., Syväoja, J. E., Stucki, M., Hübscher, U., Egly, J. M., Wood, R. D., Nucleotide excision repair of DNA with recombinant human proteins: definition of the minimal set of factors, active forms of TFIIH, and modulation by CAK, *Genes & development* 14(3), 349–359, 2000.
- [10] **Sugasawa, K.**, Okuda, Y., Saijo, M., Nishi, R., Matsuda, N., Chu, G., Mori, T., Iwai, S., Tanaka, K., Tanaka, K., Hanaoka, F., UV-induced ubiquitylation of XPC protein mediated by UV-DDB-ubiquitin ligase complex, *Cell* 121(3), 387–400, 2005.
- [11] **Yeh, J. I.**, Levine, A. S., Du, S., Chinte, U., Ghodke, H., Wang, H., Shi, H., Hsieh, C. L., Conway, J. F., van Houten, B., Rapić-Otrin, V., Damaged DNA induced UV-damaged DNA-binding protein (UV-DDB) dimerization and its roles in chromatinized DNA repair, *Proceedings of the National Academy of Sciences of the United States of America* 109(41), E2737-46, 2012.
- [12] **Sugasawa, K.**, Akagi, J.-i., Nishi, R., Iwai, S., Hanaoka, F., Two-step recognition of DNA damage for mammalian nucleotide excision repair: Directional binding of the XPC complex and DNA strand scanning, *Molecular cell* 36(4), 642–653, 2009.
- [13] **Clement, F. C.**, Camenisch, U., Fei, J., Kaczmarek, N., Mathieu, N., Naegeli, H., Dynamic two-stage mechanism of versatile DNA damage recognition by xeroderma pigmentosum group C protein, *Mutation research* 685(1-2), 21–28, 2010.
- [14] **Schärer, O. D.**, Nucleotide Excision Repair in Eukaryotes, *Cold Spring Harbor perspectives in biology* 5(10), 2013.
- [15] **Mathieu, N.**, Kaczmarek, N., Naegeli, H., Strand- and site-specific DNA lesion demarcation by the xeroderma pigmentosum group D helicase, *Proceedings of the National Academy of Sciences of the United States of America* 107(41), 17545–17550, 2010.

-
- [16] **Costa, R. M. A.**, Chiganças, V., Galhardo, R. d. S., Carvalho, H., Menck, C. F. M., The eukaryotic nucleotide excision repair pathway, *Biochimie* 85(11), 1083–1099, 2003.
- [17] **Hoeijmakers, J. H.**, Bootsma, D., DNA repair. Incisions for excision, *Nature* 371(6499), 654–655, 1994.
- [18] **Schaeffer, L.**, Roy, R., Humbert, S., Moncollin, V., Vermeulen, W., Hoeijmakers, J. H., Chambon, P., Egly, J. M., DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor, *Science (New York, N.Y.)* 260(5104), 58–63, 1993.
- [19] **Feaver, W. J.**, Gileadi, O., Kornberg, R. D., Purification and characterization of yeast RNA polymerase II transcription factor b, *The Journal of biological chemistry* 266(28), 19000–19005, 1991.
- [20] **Mattick, J. S.**, Makunin, I. V., Non-coding RNA, *Human molecular genetics* 15 Spec No 1, R17-29, 2006.
- [21] **Cramer, P.**, Armache, K.-J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G. E., Dengl, S., Geiger, S. R., Jasiak, A. J., Jawhari, A., Jennebach, S., Kamenski, T., Kettenberger, H., Kuhn, C.-D., Lehmann, E., Leike, K., Sydow, J. F., Vannini, A., Structure of eukaryotic RNA polymerases, *Annual review of biophysics* 37, 337–352, 2008.
- [22] **Grünberg, S.**, Hahn, S., Structural insights into transcription initiation by RNA polymerase II, *Trends in biochemical sciences* 38(12), 603–611, 2013.
- [23] **Kornberg, R. D.**, The molecular basis of eukaryotic transcription, *Proceedings of the National Academy of Sciences of the United States of America* 104(32), 12955–12961, 2007.
- [24] **Murakami, K.**, Elmlund, H., Kalisman, N., Bushnell, D. A., Adams, C. M., Azubel, M., Elmlund, D., Levi-Kalisman, Y., Liu, X., Gibbons, B. J., Levitt, M., Kornberg, R. D., Architecture of an RNA polymerase II transcription pre-initiation complex, *Science (New York, N.Y.)* 342(6159), 1238724, 2013.
- [25] **Chen, J.**, Suter, B., Xpd, a structural bridge and a functional link, *Cell cycle (Georgetown, Tex.)* 2(6), 503–506, 2003.
- [26] **Busso, D.**, Keriél, A., Sandrock, B., Poterszman, A., Gileadi, O., Egly, J. M., Distinct regions of MAT1 regulate cdk7 kinase and TFIIH transcription activities, *The Journal of biological chemistry* 275(30), 22815–22823, 2000.
- [27] **Egly, J.-M.**, Coin, F., A history of TFIIH: Two decades of molecular biology on a pivotal transcription/repair factor, *DNA repair*, 2011.
- [28] **Murakami, K.**, Gibbons, B. J., Davis, R. E., Nagai, S., Liu, X., Robinson, Philip J J, Wu, T., Kaplan, C. D., Kornberg, R. D., Tfb6, a previously unidentified subunit of the general transcription factor TFIIH, facilitates dissociation of Ssl2 helicase after transcription initiation, *Proceedings of the National Academy of Sciences of the United States of America* 109(13), 4816–4821, 2012.
- [29] **Mourgues, S.**, Gautier, V., Lagarou, A., Bordier, C., Mourcet, A., Slingerland, J., Kaddoum, L., Coin, F., Vermeulen, W., Gonzales de Peredo, Anne, Monsarrat, B., Mari, P.-O., Giglia-Mari, G., ELL, a novel TFIIH partner, is involved in transcription restart after DNA repair, *Proceedings of the National Academy of Sciences of the United States of America* 110(44), 17927–17932, 2013.
- [30] **Kraemer, K. H.**, Sander, M., Bohr, V. A., New areas of focus at workshop on human diseases involving DNA repair deficiency and premature aging, *Mechanisms of ageing and development* 128(2), 229–235, 2007.
-

-
- [31] **Hashimoto, S.**, Egly, J. M., Trichothiodystrophy view from the molecular basis of DNA repair/transcription factor TFIIH, *Human molecular genetics* 18(R2), R224-30, 2009.
- [32] **Andressoo, J. O.**, Hoeijmakers, J H J, Transcription-coupled repair and premature ageing, *Mutation research* 577(1-2), 179-194, 2005.
- [33] **Oksenysh, V.**, Coin, F., The long unwinding road: XPB and XPD helicases in damaged DNA opening, *Cell cycle (Georgetown, Tex.)* 9(1), 90-96, 2010.
- [34] **Fuss, J. O.**, Tainer, J. A., XPB and XPD helicases in TFIIH orchestrate DNA duplex opening and damage verification to coordinate repair with transcription and cell cycle via CAK kinase, *DNA repair*, 2011.
- [35] **Sung, P.**, Bailly, V., Weber, C., Thompson, L. H., Prakash, L., Prakash, S., Human xeroderma pigmentosum group D gene encodes a DNA helicase, *Nature* 365(6449), 852-855, 1993.
- [36] **Coin, F.**, Marinoni, J. C., Rodolfo, C., Fribourg, S., Pedrini, A. M., Egly, J. M., Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH, *Nature genetics* 20(2), 184-188, 1998.
- [37] **Rossignol, M.**, Kolb-Cheynel, I., Egly, J. M., Substrate specificity of the cdk-activating kinase (CAK) is altered upon association with TFIIH, *The EMBO journal* 16(7), 1628-1637, 1997.
- [38] **Coin, F.**, Proietti Santis, L. de, Nardo, T., Zlobinskaya, O., Stefanini, M., Egly, J.-M., p8/TTD-A as a repair-specific TFIIH subunit, *Molecular cell* 21(2), 215-226, 2006.
- [39] **Jawhari, A.**, Lainé, J.-P., Dubaele, S., Lamour, V., Poterszman, A., Coin, F., Moras, D., Egly, J.-M., p52 Mediates XPB function within the transcription/repair factor TFIIH, *The Journal of biological chemistry* 277(35), 31761-31767, 2002.
- [40] **Coin, F.**, Oksenysh, V., Egly, J.-M., Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair, *Molecular cell* 26(2), 245-256, 2007.
- [41] **Iyer, N.**, Reagan, M. S., Wu, K. J., Canagarajah, B., Friedberg, E. C., Interactions involving the human RNA polymerase II transcription/nucleotide excision repair complex TFIIH, the nucleotide excision repair protein XPG, and Cockayne syndrome group B (CSB) protein, *Biochemistry* 35(7), 2157-2167, 1996.
- [42] **Fribourg, S.**, Romier, C., Werten, S., Gangloff, Y. G., Poterszman, A., Moras, D., Dissecting the interaction network of multiprotein complexes by pairwise coexpression of subunits in *E. coli*, *Journal of molecular biology* 306(2), 363-373, 2001.
- [43] **Sandrock, B.**, Egly, J. M., A yeast four-hybrid system identifies Cdk-activating kinase as a regulator of the XPD helicase, a subunit of transcription factor IIH, *The Journal of biological chemistry* 276(38), 35328-35333, 2001.
- [44] **Adamczewski, J. P.**, Rossignol, M., Tassan, J. P., Nigg, E. A., Moncollin, V., Egly, J. M., MAT1, cdk7 and cyclin H form a kinase complex which is UV light-sensitive upon association with TFIIH, *The EMBO journal* 15(8), 1877-1884, 1996.
- [45] **Tirode, F.**, Busso, D., Coin, F., Egly, J. M., Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7, *Molecular cell* 3(1), 87-95, 1999.
- [46] **Bede, F.**, Linard, B., Brochet, X., Ripp, R., Thompson, J. D., Moras, D., Lecompte, O., Poch, O., Functional insights into the core-TFIIH from a comparative survey, *Genomics* 101(3), 178-186, 2013.
-

- [47] **Kainov, D. E.**, Vitorino, M., Cavarelli, J., Poterszman, A., Egly, J.-M., Structural basis for group A trichothiodystrophy, *Nature structural & molecular biology* 15(9), 980–984, 2008.
- [48] **Kainov, D. E.**, Cura, V., Vitorino, M., Nierengarten, H., Poussin, P., Kieffer, B., Cavarelli, J., Poterszman, A., Structure determination of the minimal complex between Tfb5 and Tfb2, two subunits of the yeast transcription/DNA-repair factor TFIIH: a retrospective study, *Acta crystallographica. Section D, Biological crystallography* 66(Pt 7), 745–755, 2010.
- [49] **Humbert, S.**, van Vuuren, H., Lutz, Y., Hoeijmakers, J. H., Egly, J. M., Moncollin, V., p44 and p34 subunits of the BTF2/TFIIH transcription factor have homologies with SSL1, a yeast protein involved in DNA repair, *The EMBO journal* 13(10), 2393–2398, 1994.
- [50] **Takagi, Y.**, Masuda, C. A., Chang, W.-H., Komori, H., Wang, D., Hunter, T., Joazeiro, C. A. P., Kornberg, R. D., Ubiquitin ligase activity of TFIIH and the transcriptional response to DNA damage, *Molecular cell* 18(2), 237–243, 2005.
- [51] **Schultz, P.**, Fribourg, S., Poterszman, A., Mallouh, V., Moras, D., Egly, J. M., Molecular structure of human TFIIH, *Cell* 102(5), 599–607, 2000.
- [52] **Chang, W. H.**, Kornberg, R. D., Electron crystal structure of the transcription factor and DNA repair complex, core TFIIH, *Cell* 102(5), 609–613, 2000.
- [53] **Takagi, Y.**, Komori, H., Chang, W.-H., Hudmon, A., Erdjument-Bromage, H., Tempst, P., Kornberg, R. D., Revised subunit structure of yeast transcription factor IIH (TFIIH) and reconciliation with human TFIIH, *The Journal of biological chemistry* 278(45), 43897–43900, 2003.
- [54] **Gibbons, B. J.**, Brignole, E. J., Azubel, M., Murakami, K., Voss, N. R., Bushnell, D. A., Asturias, F. J., Kornberg, R. D., Subunit architecture of general transcription factor TFIIH, *Proceedings of the National Academy of Sciences of the United States of America* 109(6), 1949–1954, 2012.
- [55] **Andersen, G.**, Busso, D., Poterszman, A., Hwang, J. R., Wurtz, J. M., Ripp, R., Thierry, J. C., Egly, J. M., Moras, D., The structure of cyclin H: common mode of kinase activation and specific features, *The EMBO journal* 16(5), 958–967, 1997.
- [56] **Lolli, G.**, Lowe, E. D., Brown, N. R., Johnson, L. N., The crystal structure of human CDK7 and its protein recognition properties, *Structure (London, England : 1993)* 12(11), 2067–2079, 2004.
- [57] **Gervais, V.**, Busso, D., Wasielewski, E., Poterszman, A., Egly, J. M., Thierry, J. C., Kieffer, B., Solution structure of the N-terminal domain of the human TFIIH MAT1 subunit: new insights into the RING finger family, *The Journal of biological chemistry* 276(10), 7457–7464, 2001.
- [58] **Fan, L.**, Fuss, J. O., Cheng, Q. J., Arvai, A. S., Hammel, M., Roberts, V. A., Cooper, P. K., Tainer, J. A., XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations, *Cell* 133(5), 789–800, 2008.
- [59] **Liu, H.**, Rudolf, J., Johnson, K. A., McMahon, S. A., Oke, M., Carter, L., McRobbie, A.-M., Brown, S. E., Naismith, J. H., White, M. F., Structure of the DNA repair helicase XPD, *Cell* 133(5), 801–812, 2008.
- [60] **Wolski, S. C.**, Kuper, J., Hänzelmann, P., Truglio, J. J., Croteau, D. L., van Houten, B., Kisker, C., Crystal structure of the FeS cluster-containing nucleotide excision repair helicase XPD, *PLoS biology* 6(6), e149, 2008.

- [61] **Fan, L.**, Arvai, A. S., Cooper, P. K., Iwai, S., Hanaoka, F., Tainer, J. A., Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair, *Molecular cell* 22(1), 27–37, 2006.
- [62] **Kelman, Z.**, White, M. F., Archaeal DNA replication and repair, *Current opinion in microbiology* 8(6), 669–676, 2005.
- [63] **Rouillon, C.**, White, M. F., The evolution and mechanisms of nucleotide excision repair proteins, *Research in microbiology*, 2010.
- [64] **Hilario, E.**, Li, Y., Nobumori, Y., Liu, X., Fan, L., Structure of the C-terminal half of human XPB helicase and the impact of the disease-causing mutation XP11BE, *Acta crystallographica. Section D, Biological crystallography* 69(Pt 2), 237–246, 2013.
- [65] **Vitorino, M.**, Coin, F., Zlobinskaya, O., Atkinson, R. A., Moras, D., Egly, J.-M., Poterszman, A., Kieffer, B., Solution structure and self-association properties of the p8 TFIIH subunit responsible for trichothiodystrophy, *Journal of molecular biology* 368(2), 473–480, 2007.
- [66] **Gervais, V.**, Lamour, V., Jawhari, A., Frindel, F., Wasielewski, E., Dubaele, S., Egly, J.-M., Thierry, J.-C., Kieffer, B., Poterszman, A., TFIIH contains a PH domain involved in DNA nucleotide excision repair, *Nature structural & molecular biology* 11(7), 616–622, 2004.
- [67] **Fribourg, S.**, Kellenberger, E., Rogniaux, H., Poterszman, A., van Dorsseleer, A., Thierry, J. C., Egly, J. M., Moras, D., Kieffer, B., Structural characterization of the cysteine-rich domain of TFIIH p44 subunit, *The Journal of biological chemistry* 275(41), 31963–31971, 2000.
- [68] **Kellenberger, E.**, Dominguez, C., Fribourg, S., Wasielewski, E., Moras, D., Poterszman, A., Boelens, R., Kieffer, B., Solution structure of the C-terminal domain of TFIIH P44 subunit reveals a novel type of C4C4 ring domain involved in protein-protein interactions, *The Journal of biological chemistry* 280(21), 20785–20792, 2005.
- [69] **Feaver, W. J.**, Svejstrup, J. Q., Bardwell, L., Bardwell, A. J., Buratowski, S., Gulyas, K. D., Donahue, T. F., Friedberg, E. C., Kornberg, R. D., Dual roles of a multiprotein complex from *S. cerevisiae* in transcription and DNA repair, *Cell* 75(7), 1379–1387, 1993.
- [70] **Xie, Z.**, Liu, S., Zhang, Y., Wang, Z., Roles of Rad23 protein in yeast nucleotide excision repair, *Nucleic acids research* 32(20), 5981–5990, 2004.
- [71] **Araki, M.**, Masutani, C., Takemura, M., Uchida, A., Sugasawa, K., Kondoh, J., Ohkuma, Y., Hanaoka, F., Centrosome protein centrin 2/caltractin 1 is part of the xeroderma pigmentosum group C complex that initiates global genome nucleotide excision repair, *The Journal of biological chemistry* 276(22), 18665–18672, 2001.
- [72] **Bernardes de Jesus, B. M.**, Bjørås, M., Coin, F., Egly, J. M., Dissection of the molecular defects caused by pathogenic mutations in the DNA repair factor XPC, *Molecular and cellular biology* 28(23), 7225–7235, 2008.
- [73] **Coin, F.**, Bergmann, E., Tremeau-Bravard, A., Egly, J. M., Mutations in XPB and XPD helicases found in xeroderma pigmentosum patients impair the transcription function of TFIIH, *The EMBO journal* 18(5), 1357–1366, 1999.

-
- [74] **Wood, R. D.**, DNA damage recognition during nucleotide excision repair in mammalian cells, *Biochimie* 81(1-2), 39–44, 1999.
- [75] **Bootsma, D.**, Hoeijmakers, J. H., DNA repair. Engagement with transcription, *Nature* 363(6425), 114–115, 1993.
- [76] **Oksenych, V.**, Jesus, B. B. de, Zhovmer, A., Egly, J.-M., Coin, F., Molecular insights into the recruitment of TFIIH to sites of DNA damage, *The EMBO journal*, 2009.
- [77] **Dubaele, S.**, Proietti Santis, L. de, Bienstock, R. J., Keriel, A., Stefanini, M., van Houten, B., Egly, J.-M., Basal transcription defect discriminates between xeroderma pigmentosum and trichothiodystrophy in XPD patients, *Molecular cell* 11(6), 1635–1646, 2003.
- [78] **Abdulrahman, W.**, Iltis, I., Radu, L., Braun, C., Maglott-Roth, A., Giraudon, C., Egly, J.-M., Poterszman, A., ARCH domain of XPD, an anchoring platform for CAK that conditions TFIIH DNA repair and transcription activities, *Proceedings of the National Academy of Sciences of the United States of America* 110(8), E633-42, 2013.
- [79] **Compe, E.**, Egly, J.-M., TFIIH: when transcription met DNA repair, *Nature reviews. Molecular cell biology* 13(6), 343–354, 2012.
- [80] **Coin, F.**, Oksenych, V., Mocquet, V., Groh, S., Blattner, C., Egly, J. M., Nucleotide excision repair driven by the dissociation of CAK from TFIIH, *Molecular cell* 31(1), 9–20, 2008.
- [81] **Vermeulen, W.**, Bergmann, E., Auriol, J., Rademakers, S., Frit, P., Appeldoorn, E., Hoeijmakers, J. H., Egly, J. M., Sublimiting concentration of TFIIH transcription/DNA repair factor causes TTD-A trichothiodystrophy disorder, *Nature genetics* 26(3), 307–313, 2000.
- [82] **Giglia-Mari, G.**, Coin, F., Ranish, J. A., Hoogstraten, D., Theil, A., Wijgers, N., Jaspers, Nicolaas G J, Raams, A., Argentini, M., van der Spek, P J, Botta, E., Stefanini, M., Egly, J.-M., Aebersold, R., Hoeijmakers, Jan H J, Vermeulen, W., A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A, *Nature genetics* 36(7), 714–719, 2004.
- [83] **Dvir, A.**, Conaway, J. W., Conaway, R. C., Mechanism of transcription initiation and promoter escape by RNA polymerase II, *Current opinion in genetics & development* 11(2), 209–214, 2001.
- [84] **Assfalg, R.**, Lebedev, A., Gonzalez, O. G., Schelling, A., Koch, S., Iben, S., TFIIH is an elongation factor of RNA polymerase I, *Nucleic acids research* 40(2), 650–659, 2012.
- [85] **Iben, S.**, Tschochner, H., Bier, M., Hoogstraten, D., Hozák, P., Egly, J. M., Grummt, I., TFIIH plays an essential role in RNA polymerase I transcription, *Cell* 109(3), 297–306, 2002.
- [86] **Barski, A.**, Chepelev, I., Liko, D., Cuddapah, S., Fleming, A. B., Birch, J., Cui, K., White, R. J., Zhao, K., Pol II and its associated epigenetic marks are present at Pol III-transcribed noncoding RNA genes, *Nature structural & molecular biology* 17(5), 629–634, 2010.
- [87] **Oler, A. J.**, Alla, R. K., Roberts, D. N., Wong, A., Hollenhorst, P. C., Chandler, K. J., Cassidy, P. A., Nelson, C. A., Hagedorn, C. H., Graves, B. J., Cairns, B. R., Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors, *Nature structural & molecular biology* 17(5), 620–628, 2010.

-
- [88] **Holstege, F. C.**, van der Vliet, P C, Timmers, H. T., Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH, *The EMBO journal* 15(7), 1666–1677, 1996.
- [89] **Moreland, R. J.**, Tirode, F., Yan, Q., Conaway, J. W., Egly, J. M., Conaway, R. C., A role for the TFIIF XPB DNA helicase in promoter escape by RNA polymerase II, *The Journal of biological chemistry* 274(32), 22127–22130, 1999.
- [90] **Dvir, A.**, Conaway, R. C., Conaway, J. W., Promoter escape by RNA polymerase II. A role for an ATP cofactor in suppression of arrest by polymerase at promoter-proximal sites, *The Journal of biological chemistry* 271(38), 23352–23356, 1996.
- [91] **Douziech, M.**, Coin, F., Chipoulet, J. M., Arai, Y., Ohkuma, Y., Egly, J. M., Coulombe, B., Mechanism of promoter melting by the xeroderma pigmentosum complementation group B helicase of transcription factor IIH revealed by protein-DNA photo-cross-linking, *Molecular and cellular biology* 20(21), 8168–8177, 2000.
- [92] **Akhtar, M. S.**, Heidemann, M., Tietjen, J. R., Zhang, D. W., Chapman, R. D., Eick, D., Ansari, A. Z., TFIIF kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II, *Molecular cell* 34(3), 387–393, 2009.
- [93] **Larochelle, S.**, Pandur, J., Fisher, R. P., Salz, H. K., Suter, B., Cdk7 is essential for mitosis and for in vivo Cdk-activating kinase activity, *Genes & development* 12(3), 370–381, 1998.
- [94] **Fesquet, D.**, Morin, N., Doree, M., Devault, A., Is Cdk7/cyclin H/MAT1 the genuine cdk activating kinase in cycling *Xenopus* egg extracts?, *Oncogene* 15(11), 1303–1307, 1997.
- [95] **Wu, L.**, Chen, P., Hwang, J. J., Barsky, L. W., Weinberg, K. I., Jong, A., Starnes, V. A., RNA antisense abrogation of MAT1 induces G1 phase arrest and triggers apoptosis in aortic smooth muscle cells, *The Journal of biological chemistry* 274(9), 5564–5572, 1999.
- [96] **Fisher, R. P.**, Secrets of a double agent: CDK7 in cell-cycle control and transcription, *Journal of cell science* 118(Pt 22), 5171–5180, 2005.
- [97] **Ito, S.**, Tan, L. J., Andoh, D., Narita, T., Seki, M., Hirano, Y., Narita, K., Kuraoka, I., Hiraoka, Y., Tanaka, K., MMXD, a TFIIF-independent XPD-MMS19 protein complex involved in chromosome segregation, *Molecular cell* 39(4), 632–640, 2010.
- [98] **Frit, P.**, Kwon, K., Coin, F., Auriol, J., Dubaele, S., Salles, B., Egly, J. M., Transcriptional activators stimulate DNA repair, *Molecular cell* 10(6), 1391–1401, 2002.
- [99] **Fong, Y. W.**, Inouye, C., Yamaguchi, T., Cattoglio, C., Grubisic, I., Tjian, R., A DNA repair complex functions as an Oct4/Sox2 coactivator in embryonic stem cells, *Cell* 147(1), 120–131, 2011.
- [100] **Schmitz, K.-M.**, Schmitt, N., Hoffmann-Rohrer, U., Schäfer, A., Grummt, I., Mayer, C., TAF12 recruits Gadd45a and the nucleotide excision repair complex to the promoter of rRNA genes leading to active DNA demethylation, *Molecular cell* 33(3), 344–353, 2009.
- [101] **Le May, N.**, Mota-Fernandes, D., Vélez-Cruz, R., Iltis, I., Biard, D., Egly, J. M., NER factors are recruited to active promoters and facilitate chromatin modification for transcription in the absence of exogenous genotoxic attack, *Molecular cell* 38(1), 54–66, 2010.
-

- [102] **Lehmann, A. R.**, The xeroderma pigmentosum group D (XPD) gene: one gene, two functions, three diseases, *Genes & development* 15(1), 15–23, 2001.
- [103] **Aguilar-Fuentes, J.**, Fregoso, M., Herrera, M., Reynaud, E., Braun, C., Egly, J. M., Zurita, M., p8/TTDA overexpression enhances UV-irradiation resistance and suppresses TFIIH mutations in a *Drosophila* trichothiodystrophy model, *PLoS genetics* 4(11), e1000253, 2008.
- [104] **Castro, J.**, Merino, C., Zurita, M., Molecular characterization and developmental expression of the TFIIH factor p62 gene from *Drosophila melanogaster*: effects on the UV light sensitivity of a p62 mutant fly, *DNA repair* 1(5), 359–368, 2002.
- [105] **Fregoso, M.**, Lainé, J.-P., Aguilar-Fuentes, J., Mocquet, V., Reynaud, E., Coin, F., Egly, J.-M., Zurita, M., DNA repair and transcriptional deficiencies caused by mutations in the *Drosophila* p52 subunit of TFIIH generate developmental defects and chromosome fragility, *Molecular and cellular biology* 27(10), 3640–3650, 2007.
- [106] **Manuguerra, M.**, Saletta, F., Karagas, M. R., Berwick, M., Veglia, F., Vineis, P., Matullo, G., XRCC3 and XPD/ERCC2 single nucleotide polymorphisms and the risk of cancer: a HuGE review, *American journal of epidemiology* 164(4), 297–302, 2006.
- [107] **Zhang, J.**, Gu, S.-Y., Zhang, P., Jia, Z., Chang, J.-H., ERCC2 Lys751Gln polymorphism is associated with lung cancer among Caucasians, *European journal of cancer (Oxford, England : 1990)* 46(13), 2479–2484, 2010.
- [108] **Kraemer, K. H.**, Levy, D. D., Parris, C. N., Gozukara, E. M., Moriwaki, S., Adelberg, S., Seidman, M. M., Xeroderma pigmentosum and related disorders: examining the linkage between defective DNA repair and cancer, *The Journal of investigative dermatology* 103(5 Suppl), 96S-101S, 1994.
- [109] **Cleaver, J. E.**, Cancer in xeroderma pigmentosum and related disorders of DNA repair, *Nature reviews. Cancer* 5(7), 564–573, 2005.
- [110] **Boer, J. de**, Hoeijmakers, J. H., Nucleotide excision repair and human syndromes, *Carcinogenesis* 21(3), 453–460, 2000.
- [111] **Nance, M. A.**, Berry, S. A., Cockayne syndrome: review of 140 cases, *American journal of medical genetics* 42(1), 68–84, 1992.
- [112] **Itin, P. H.**, Sarasin, A., Pittelkow, M. R., Trichothiodystrophy: update on the sulfur-deficient brittle hair syndromes, *Journal of the American Academy of Dermatology* 44(6), 891-920; quiz 921-4, 2001.
- [113] **Schärer, O. D.**, Hot topics in DNA repair: the molecular basis for different disease states caused by mutations in TFIIH and XPG, *DNA repair* 7(2), 339–344, 2008.
- [114] **Compe, E.**, Malerba, M., Soler, L., Marescaux, J., Borrelli, E., Egly, J.-M., Neurological defects in trichothiodystrophy reveal a coactivator function of TFIIH, *Nature neuroscience* 10(11), 1414–1422, 2007.
- [115] **Ito, S.**, Kuraoka, I., Chymkowitch, P., Compe, E., Takedachi, A., Ishigami, C., Coin, F., Egly, J.-M., Tanaka, K., XPG stabilizes TFIIH, allowing transactivation of nuclear receptors: implications for Cockayne syndrome in XP-G/CS patients, *Molecular cell* 26(2), 231–243, 2007.
- [116] **Bogliolo, M.**, Schuster, B., Stoepker, C., Derkunt, B., Su, Y., Raams, A., Trujillo, J. P., Minguillón, J., Ramírez, M. J., Pujol, R., Casado, J. A., Baños, R., Rio, P., Knies, K., Zúñiga, S., Benítez, J., Bueren, J. A.,

- Jaspers, Nicolaas G J, Schärer, O. D., de Winter, Johan P, Schindler, D., Surrallés, J., Mutations in ERCC4, encoding the DNA-repair endonuclease XPF, cause Fanconi anemia, *American journal of human genetics* 92(5), 800–806, 2013.
- [117] **Hickey, D. A.**, Singer, Gregory A C, Genomic and proteomic adaptations to growth at high temperature, *Genome biology* 5(10), 117, 2004.
- [118] **Amlacher, S.**, Sarges, P., Flemming, D., van Noort, V., Kunze, R., Devos, D. P., Arumugam, M., Bork, P., Hurt, E., Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile, *Cell* 146(2), 277–289, 2011.
- [119] **La Touche, C. J.**, A Chætomium-like Thermophile Fungus, *Nature <London>* 161(4087), 320, 1948.
- [120] **Inoue, H.**, Nojima, H., Okayama, H., High efficiency transformation of *Escherichia coli* with plasmids, *Gene* 96(1), 23–28, 1990.
- [121] ADXV. Version 1.9.8, Andrew Arvai, The Scripps Research Institute.
- [122] **Winn, M. D.**, Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, Andrew G W, McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., Wilson, K. S., Overview of the CCP4 suite and current developments, *Acta crystallographica. Section D, Biological crystallography* 67(Pt 4), 235–242, 2011.
- [123] **Emsley, P.**, Cowtan, K., Coot: model-building tools for molecular graphics, *Acta crystallographica. Section D, Biological crystallography* 60(Pt 12 Pt 1), 2126–2132, 2004.
- [124] **Battye, T. G.**, Kontogiannis, L., Johnson, O., Powell, H. R., Leslie, Andrew G. W., iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM, *Acta crystallographica. Section D, Biological crystallography* 67(Pt 4), 271–281, 2011.
- [125] **Jones, T. A.**, Zou, J. Y., Cowan, S. W., Kjeldgaard, M., Improved methods for building protein models in electron density maps and the location of errors in these models, *Acta crystallographica. Section A, Foundations of crystallography* 47 (Pt 2), 110–119, 1991.
- [126] **McCoy, A. J.**, Solving structures of protein complexes by molecular replacement with Phaser, *Acta crystallographica. Section D, Biological crystallography* 63(Pt 1), 32–41, 2007.
- [127] **Adams, P. D.**, Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta crystallographica. Section D, Biological crystallography* 66(Pt 2), 213–221, 2010.
- [128] The PyMOL Molecular Graphics System. Version 1.6.0, Schrödinger, LLC.
- [129] **Evans, P.**, Scaling and assessment of data quality, *Acta crystallographica. Section D, Biological crystallography* 62(Pt 1), 72–82, 2006.
- [130] **Sheldrick, G. M.**, Experimental phasing with SHELXC/D/E: combining chain tracing with density modification, *Acta crystallographica. Section D, Biological crystallography* 66(Pt 4), 479–485, 2010.
- [131] **Kabsch, W.**, XDS, *Acta crystallographica. Section D, Biological crystallography* 66(Pt 2), 125–132, 2010.
- [132] **Altschul, S. F.**, Gish, W., Miller, W., Myers, E. W., Lipman, D. J., Basic local alignment search tool, *Journal of molecular biology* 215(3), 403–410, 1990.

-
- [133] **Joosten, R. P.**, te Beek, Tim A H, Krieger, E., Hekkelman, M. L., Hooft, Rob W W, Schneider, R., Sander, C., Vriend, G., A series of PDB related databases for everyday needs, *Nucleic acids research* 39(Database issue), D411-9, 2011.
- [134] GENTle. Version 1.9.4, Magnus Manske.
- [135] **Waterhouse, A. M.**, Procter, J. B., Martin, David M A, Clamp, M., Barton, G. J., Jalview Version 2--a multiple sequence alignment editor and analysis workbench, *Bioinformatics (Oxford, England)* 25(9), 1189–1191, 2009.
- [136] **Huang, X.**, Miller Webb, A time-efficient, linear-space local similarity algorithm, *Advances in Applied Mathematics* 12(3), 1991.
- [137] **Edgar, R. C.**, MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic acids research* 32(5), 1792–1797, 2004.
- [138] OligoAnalyzer. Version 3.1, IDT Integrated DNA Technologies Inc.
- [139] **Kelley, L. A.**, Sternberg, Michael J E, Protein structure prediction on the Web: a case study using the Phyre server, *Nature protocols* 4(3), 363–371, 2009.
- [140] **Gasteiger, E.**, Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., Bairoch, A., Protein Identification and Analysis Tools on the ExPASy Server, In: John M. Walker (ed): *The Proteomics Protocols Handbook*, Humana Press, , 571–607, 2005.
- [141] **Holm, L.**, Rosenström, P., Dali server: conservation mapping in 3D, *Nucleic acids research* 38(Web Server issue), W545-9, 2010.
- [142] **Artimo, P.**, Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., Castro, E. de, Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., Stockinger, H., ExPASy: SIB bioinformatics resource portal, *Nucleic acids research* 40(Web Server issue), W597-603, 2012.
- [143] **Berman, H. M.**, Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E., The Protein Data Bank, *Nucleic acids research* 28(1), 235–242, 2000.
- [144] **Berman, H.**, Henrick, K., Nakamura, H., Announcing the worldwide Protein Data Bank, *Nature structural biology* 10(12), 980, 2003.
- [145] **Letunic, I.**, Doerks, T., Bork, P., SMART 7: recent updates to the protein domain annotation resource, *Nucleic acids research* 40(Database issue), D302-5, 2012.
- [146] **Schultz, J.**, Milpetz, F., Bork, P., Ponting, C. P., SMART, a simple modular architecture research tool: identification of signaling domains, *Proceedings of the National Academy of Sciences of the United States of America* 95(11), 5857–5864, 1998.
- [147] **The Uniprot Consortium**, Activities at the Universal Protein Resource (UniProt), *Nucleic acids research* 42(1), D191-D198, 2014.
- [148] **Mullis, K. B.**, Faloona, F. A., Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction, *Methods in enzymology* 155, 335–350, 1987.
- [149] **Saiki, R. K.**, Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., Erlich, H. A., Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase, *Science (New York, N.Y.)* 239(4839), 487–491, 1988.

- [150] **Siwek, W.**, Czapinska, H., Bochtler, M., Bujnicki, J. M., Skowronek, K., Crystal structure and mechanism of action of the N6-methyladenine-dependent type IIM restriction endonuclease R.DpnI, *Nucleic acids research* 40(15), 7563–7572, 2012.
- [151] **Roberts, R. J.**, Restriction endonucleases, *CRC critical reviews in biochemistry* 4(2), 123–164, 1976.
- [152] **Pingoud, A.**, Jeltsch, A., Structure and function of type II restriction endonucleases, *Nucleic acids research* 29(18), 3705–3727, 2001.
- [153] **Sambrook, J.**, Russell, D. W., *Molecular Cloning: A Laboratory Manual (3rd Edition)*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2001.
- [154] **Li, M. Z.**, Elledge, S. J., Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC, *Nature methods* 4(3), 251–256, 2007.
- [155] **Lorenz, M. G.**, Wackernagel, W., Bacterial gene transfer by natural genetic transformation in the environment, *Microbiological reviews* 58(3), 563–602, 1994.
- [156] **Chen, I.**, Dubnau, D., DNA uptake during bacterial transformation, *Nature reviews. Microbiology* 2(3), 241–249, 2004.
- [157] **Laemmli, U. K.**, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature* 227(5259), 680–685, 1970.
- [158] **Pierce, M. M.**, Raman, C. S., Nall, B. T., Isothermal titration calorimetry of protein-protein interactions, *Methods (San Diego, Calif.)* 19(2), 213–221, 1999.
- [159] **Pfeiffer, H. G.**, Liebhafsky, H. A., The origins of Beer's law, *Journal of Chemical Education* 28(3), 123–125, 1951.
- [160] **Pantoliano, M. W.**, Petrella, E. C., Kwasnoski, J. D., Lobanov, V. S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B. A., Lane, P., Salemme, F. R., High-density miniaturized thermal shift assays as a general strategy for drug discovery, *Journal of biomolecular screening* 6(6), 429–440, 2001.
- [161] **Matulis, D.**, Kranz, J. K., Salemme, F. R., Todd, M. J., Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor, *Biochemistry* 44(13), 5258–5266, 2005.
- [162] **Niesen, F. H.**, Berglund, H., Vedadi, M., The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability, *Nature protocols* 2(9), 2212–2221, 2007.
- [163] **Ericsson, U. B.**, Hallberg, B. M., Detitta, G. T., Dekker, N., Nordlund, P., ThermoFluor-based high-throughput stability optimization of proteins for structural studies, *Analytical biochemistry* 357(2), 289–298, 2006.
- [164] **Greenfield, N. J.**, Using circular dichroism spectra to estimate protein secondary structure, *Nature protocols* 1(6), 2876–2890, 2006.
- [165] **Kelly, S. M.**, Jess, T. J., Price, N. C., How to study proteins by circular dichroism, *Biochimica et biophysica acta* 1751(2), 119–139, 2005.
- [166] **Wen, J.**, Arakawa, T., Philo, J. S., Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions, *Analytical biochemistry* 240(2), 155–166, 1996.

- [167] **Zimm, B. H.**, The Scattering of Light and the Radial Distribution Function of High Polymer Solutions, *The Journal of Chemical Physics* 16(12), 1093, 1948.
- [168] **Karas, M.**, Glückmann, M., Schäfer, J., Ionization in matrix-assisted laser desorption/ionization, *Journal of Mass Spectrometry* 35(1), 1–12, 2000.
- [169] **Wollnik, H.**, Time-of-flight mass analyzers, *Mass Spectrometry Reviews* 12(2), 89–114, 1993.
- [170] **Bragg W.L.**, The Diffraction of Short Electromagnetic Waves by a Crystal, *Proceedings of the Cambridge Philosophical Society* 17(1), 43–57, 1913.
- [171] **Rhodes, G.**, *Crystallography Made Crystal Clear (Third Edition)*, Academic Press, Elsevier, 2006.
- [172] **Cowtan, K.**, *Book of Fourier*, The University of York,
<http://www.yesbl.york.ac.uk/~cowtan/fourier/fourier.html>.
- [173] **Taylor, G.**, The phase problem, *Acta crystallographica. Section D, Biological crystallography* 59(Pt 11), 1881–1890, 2003.
- [174] **Dauter, Z.**, Dauter, M., Dodson, E., Jolly SAD, *Acta crystallographica. Section D, Biological crystallography* 58(Pt 3), 494–506, 2002.
- [175] **Rossmann, M. G.**, Blow, D. M., The Detection of Sub-Units within the Crystallographic Asymmetric Unit, *Acta crystallographica. Section D, Biological crystallography* 15, 24–31, 1962.
- [176] **Li, L.**, Ismagilov, R. F., Protein crystallization using microfluidic technologies based on valves, droplets, and SlipChip, *Annual review of biophysics* 39, 139–158, 2010.
- [177] **Chayen, N. E.**, Comparative studies of protein crystallization by vapour-diffusion and microbatch techniques, *Acta crystallographica. Section D, Biological crystallography* 54(Pt 1), 8–15, 1998.
- [178] **Murshudov, G. N.**, Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., Vagin, A. A., REFMAC5 for the refinement of macromolecular crystal structures, *Acta crystallographica. Section D, Biological crystallography* 67(Pt 4), 355–367, 2011.
- [179] **Adachi, N.**, Matsumoto, M., Hasegawa, S., Yamamoto, T., Horikoshi, M., Analysis of TFIID subunit through isolation of the gene from *Schizosaccharomyces pombe* corresponding to that of *Saccharomyces cerevisiae* SSL1, reveals the presence of conserved structural motifs, *Yeast (Chichester, England)* 15(3), 255–262, 1999.
- [180] **Feaver, W. J.**, Huang, W., Gileadi, O., Myers, L., Gustafsson, C. M., Kornberg, R. D., Friedberg, E. C., Subunit interactions in yeast transcription/repair factor TFIID. Requirement for Tfb3 subunit in nucleotide excision repair, *The Journal of biological chemistry* 275(8), 5941–5946, 2000.
- [181] **Schultz, S. C.**, Shields, G. C., Steitz, T. A., Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees, *Science (New York, N.Y.)* 253(5023), 1001–1007, 1991.
- [182] **Marianayagam, N. J.**, Sunde, M., Matthews, J. M., The power of two: protein dimerization in biology, *Trends in biochemical sciences* 29(11), 618–625, 2004.
- [183] **Grant, R. A.**, Filman, D. J., Finkel, S. E., Kolter, R., Hogle, J. M., The crystal structure of Dps, a ferritin homolog that binds and protects DNA, *Nature structural biology* 5(4), 294–303, 1998.
- [184] **Sadler, J. E.**, Biochemistry and genetics of von Willebrand factor, *Annual review of biochemistry* 67, 395–424, 1998.

- [185] **Sadler, J. E.**, Matsushita, T., Dong, Z., Tuley, E. A., Westfield, L. A., Molecular mechanism and classification of von Willebrand disease, *Thrombosis and haemostasis* 74(1), 161–166, 1995.
- [186] **Krissinel, E.**, Henrick, K., Inference of macromolecular assemblies from crystalline state, *Journal of molecular biology* 372(3), 774–797, 2007.
- [187] **Schmitt, D. R.**, Kuper, J., Elias, A., Kisker, C., The structure of the TFIIH p34 subunit reveals a von Willebrand Factor A like fold, Manuscript submitted for publication.
- [188] **Whittaker, C. A.**, Hynes, R. O., Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere, *Molecular biology of the cell* 13(10), 3369–3387, 2002.
- [189] **Springer, T. A.**, Complement and the multifaceted functions of VWA and integrin I domains, *Structure (London, England : 1993)* 14(11), 1611–1616, 2006.
- [190] **Ponting, C. P.**, Aravind, L., Schultz, J., Bork, P., Koonin, E. V., Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer, *Journal of molecular biology* 289(4), 729–745, 1999.
- [191] **Emsley, J.**, Cruz, M., Handin, R., Liddington, R., Crystal structure of the von Willebrand Factor A1 domain and implications for the binding of platelet glycoprotein Ib, *The Journal of biological chemistry* 273(17), 10396–10401, 1998.
- [192] **Dumas, J. J.**, Kumar, R., McDonagh, T., Sullivan, F., Stahl, M. L., Somers, W. S., Mosyak, L., Crystal structure of the wild-type von Willebrand factor A1-glycoprotein Ibalph complex reveals conformation differences with a complex bearing von Willebrand disease mutations, *The Journal of biological chemistry* 279(22), 23327–23334, 2004.
- [193] **Celikel, R.**, Varughese, K. I., Madhusudan, Yoshioka, A., Ware, J., Ruggeri, Z. M., Crystal structure of the von Willebrand factor A1 domain in complex with the function blocking NMC-4 Fab, *Nature structural biology* 5(3), 189–194, 1998.
- [194] **Fukuda, K.**, Doggett, T., Laurenzi, I. J., Liddington, R. C., Diacovo, T. G., The snake venom protein botrocetin acts as a biological brace to promote dysfunctional platelet aggregation, *Nature structural & molecular biology* 12(2), 152–159, 2005.
- [195] **Maita, N.**, Nishio, K., Nishimoto, E., Matsui, T., Shikamoto, Y., Morita, T., Sadler, J. E., Mizuno, H., Crystal structure of von Willebrand factor A1 domain complexed with snake venom, bitiscetin: insight into glycoprotein Ibalph binding mechanism induced by snake venom proteins, *The Journal of biological chemistry* 278(39), 37777–37781, 2003.
- [196] **Huang, R.-H.**, Fremont, D. H., Diener, J. L., Schaub, R. G., Sadler, J. E., A structural explanation for the antithrombotic activity of ARC1172, a DNA aptamer that binds von Willebrand factor domain A1, *Structure (London, England : 1993)* 17(11), 1476–1484, 2009.
- [197] **Kainov, D. E.**, Selth, L. A., Svejstrup, J. Q., Egly, J.-M., Poterzsmann, A., Interacting partners of the Tfb2 subunit from yeast TFIIH, *DNA repair* 9(1), 33–39, 2010.
- [198] **Nymalm, Y.**, Puranen, J. S., Nyholm, Thomas K M, Käpylä, J., Kidron, H., Pentikäinen, O. T., Airene, T. T., Heino, J., Slotte, J. P., Johnson, M. S., Salminen, T. A., Jararhagin-derived RKKH peptides induce

- structural changes in alpha1I domain of human integrin alpha1beta1, *The Journal of biological chemistry* 279(9), 7962–7970, 2004.
- [199] **Riedinger, C.**, Boehringer, J., Trempe, J.-F., Lowe, E. D., Brown, N. R., Gehring, K., Noble, Martin E M, Gordon, C., Endicott, J. A., Structure of Rpn10 and its interactions with polyubiquitin chains and the proteasome subunit Rpn12, *The Journal of biological chemistry* 285(44), 33992–34003, 2010.
- [200] **Walker, J. R.**, Corpina, R. A., Goldberg, J., Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair, *Nature* 412(6847), 607–614, 2001.
- [201] **Rao, S. T.**, Rossmann, M. G., Comparison of super-secondary structures in proteins, *Journal of molecular biology* 76(2), 241–256, 1973.
- [202] **Luo, B.-H.**, Carman, C. V., Springer, T. A., Structural basis of integrin regulation and signaling, *Annual review of immunology* 25, 619–647, 2007.
- [203] **Huizinga, E. G.**, Tsuji, S., Romijn, Roland A P, Schiphorst, M. E., de Groot, Philip G, Sixma, J. J., Gros, P., Structures of glycoprotein Ibalpha and its complex with von Willebrand factor A1 domain, *Science (New York, N.Y.)* 297(5584), 1176–1179, 2002.
- [204] **Lander, G. C.**, Estrin, E., Matyskiela, M. E., Bashore, C., Nogales, E., Martin, A., Complete subunit architecture of the proteasome regulatory particle, *Nature* 482(7384), 186–191, 2012.
- [205] **Śledź, P.**, Unverdorben, P., Beck, F., Pfeifer, G., Schweitzer, A., Förster, F., Baumeister, W., Structure of the 26S proteasome with ATP-γS bound provides insights into the mechanism of nucleotide-dependent substrate translocation, *Proceedings of the National Academy of Sciences of the United States of America* 110(18), 7264–7269, 2013.
- [206] **Beck, F.**, Unverdorben, P., Bohn, S., Schweitzer, A., Pfeifer, G., Sakata, E., Nickell, S., Plitzko, J. M., Villa, E., Baumeister, W., Förster, F., Near-atomic resolution structural model of the yeast 26S proteasome, *Proceedings of the National Academy of Sciences of the United States of America* 109(37), 14870–14875, 2012.
- [207] **Aravind, L.**, Ponting, C. P., Homologues of 26S proteasome subunits are regulators of transcription and translation, *Protein science : a publication of the Protein Society* 7(5), 1250–1254, 1998.
- [208] **Tremeau-Bravard, A.**, Perez, C., Egly, J. M., A role of the C-terminal part of p44 in the promoter escape activity of transcription factor IIH, *The Journal of biological chemistry* 276(29), 27693–27697, 2001.
- [209] **Saurin, A. J.**, Borden, K. L., Boddy, M. N., Freemont, P. S., Does this have a familiar RING?, *Trends in biochemical sciences* 21(6), 208–214, 1996.
- [210] **Freemont, P. S.**, RING for destruction?, *Current biology : CB* 10(2), R84-7, 2000.
- [211] **Lorick, K. L.**, Jensen, J. P., Fang, S., Ong, A. M., Hatakeyama, S., Weissman, A. M., RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination, *Proceedings of the National Academy of Sciences of the United States of America* 96(20), 11364–11369, 1999.
- [212] **VanDemark, A. P.**, Hill, C. P., Structural basis of ubiquitylation, *Current opinion in structural biology* 12(6), 822–830, 2002.
- [213] **Albert, T. K.**, Hanzawa, H., Legtenberg, Yvonne I A, de Ruwe, Marjolein J, van den Heuvel, Fiona A J, Collart, M. A., Boelens, R., Timmers, H Th Marc, Identification of a ubiquitin-protein ligase subunit within the CCR4-NOT transcription repressor complex, *The EMBO journal* 21(3), 355–364, 2002.

- [214] **Duda, D. M.**, Olszewski, J. L., Tron, A. E., Hammel, M., Lambert, L. J., Waddell, M. B., Mittag, T., DeCaprio, J. A., Schulman, B. A., Structure of a glomulin-RBX1-CUL1 complex: inhibition of a RING E3 ligase through masking of its E2-binding surface, *Molecular cell* 47(3), 371–382, 2012.
- [215] **Zurita, M.**, Merino, C., The transcriptional complexity of the TFIID complex, *Trends in genetics : TIG* 19(10), 578–584, 2003.
- [216] **Miller, J.**, McLachlan, A. D., Klug, A., Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes, *The EMBO journal* 4(6), 1609–1614, 1985.
- [217] **Hall, T. M. T.**, Multiple modes of RNA recognition by zinc finger proteins, *Current opinion in structural biology* 15(3), 367–373, 2005.
- [218] **Brown, R. S.**, Zinc finger proteins: getting a grip on RNA, *Current opinion in structural biology* 15(1), 94–98, 2005.
- [219] **Gamsjaeger, R.**, Liew, C. K., Loughlin, F. E., Crossley, M., Mackay, J. P., Sticky fingers: zinc-fingers as protein-recognition motifs, *Trends in biochemical sciences* 32(2), 63–70, 2007.
- [220] **Matthews, J. M.**, Sunde, M., Zinc fingers - Folds for many occasions, *IUBMB life* 54(6), 351–355, 2002.

Protein Data Bank References

Data files contained in the Protein Data Bank (PDB) archive at www.rcsb.org (<ftp://ftp.wwpdb.org>) are free of all copyright restrictions and made fully and freely available for both non-commercial and commercial use.

The following table provides an alphabetical summary of all PDB structures used and discussed in the present doctoral thesis, including the authors of the respective data and their year of publication.

PDB ID	Authors	Year
1AUQ	Emsley, J., Cruz, M., Handin, R., Liddington, R.	1998
1PT6	Nymalm, Y., Puranen, J. S., Nyholm, Thomas K M, Käpylä, J., Kidron, H., Pentikäinen, O. T., Airene, T. T., Heino, J., Slotte, J. P., Johnson, M. S., Salminen, T. A.	2004
1SQ0	Dumas, J. J., Kumar, R., McDonagh, T., Sullivan, F., Stahl, M. L., Somers, W. S., Mosyak, L.	2004
1U0N	Fukuda, K., Doggett, T., Laurenzi, I. J., Liddington, R. C., Diacovo, T. G.	2005
1Z60	Kellenberger, E., Dominguez, C., Fribourg, S., Wasielewski, E., Moras, D., Poterszman, A., Boelens, R., Kieffer, B.	2005
2X5N	Riedinger, C., Boehringer, J., Trempe, J.-F., Lowe, E. D., Brown, N. R., Gehring, K., Noble, Martin E M, Gordon, C., Endicott, J. A.	2010
4C0V	Unverdorben, P., Beck, F., Sledź, P., Schweitzer, A., Pfeifer, G., Plitzko, J. M., Baumeister, W., Förster, F.	2014

VII. Figures and Tables

VII.1 List of Figures

Figure I.1: DNA Damage and Repair Mechanisms.....	1
Figure I.2: Mechanism of Eukaryotic Nucleotide Excision Repair.....	3
Figure I.3: Composition of the TFIIH Complex and Structural Knowledge on its Subunits.....	6
Figure II.1: Phase Diagram and Methods of Protein Crystallization.....	43
Figure IV.1: Domain Structure of p44h Variants.....	47
Figure IV.2: Variants of p44h.....	47
Figure IV.3: Domain Structure of p34h Variants and the p34h p44h MC.....	48
Figure IV.4: The p34h p44h Minimal Complex.....	48
Figure IV.5: Analysis of p34h p44h MC Crystals and X-Ray Diffraction.....	49
Figure IV.6: Domain Structure of p34ct Variants.....	51
Figure IV.7: Variants of p34ct.....	52
Figure IV.8: SEC Elution Profiles of p34ct Samples.....	52
Figure IV.9: Variants of p44ct.....	53
Figure IV.10: Domain Structure of p44ct Variants.....	54
Figure IV.11: SEC Elution Profiles of p44ct Samples.....	54
Figure IV.12: Domain Structure of p34ct p44ct MC.....	55
Figure IV.13: Interaction Studies with p34ct and p44ct.....	55
Figure IV.14: The p34ct p44ct Minimal Complex.....	56
Figure IV.15: DNA Binding Studies with p34ct and p44ct Variants.....	57
Figure IV.16: Thermofluor Analysis of p34ct and p44ct Full-Length Proteins.....	59
Figure IV.17: Thermofluor Analysis of Shortened p34ct and p44ct Variants.....	60
Figure IV.18: Thermofluor Analysis of p34ct p44ct MC.....	61
Figure IV.19: Molar Mass Analysis of p34ct and p44ct Variants.....	62
Figure IV.20: Molar Mass Analysis of p34ct p44ct MC.....	63
Figure IV.21: Crystals and overall Structure of DPS.....	64
Figure IV.22: Samples of p44ct obtained from different Purification Strategies.....	65
Figure IV.23: Crystals of the p44ct 101-285 vWA Domain.....	66
Figure IV.24: Analysis of p34ct Crystals and X-Ray Diffraction.....	67
Figure IV.25: Electron Density of p34ct Crystals and Preliminary Model.....	68
Figure IV.26: Crystals of p34ct 1-277 and X-Ray Diffraction.....	69
Figure IV.27: Electron Density of p34ct 1-277.....	70
Figure IV.28: Structure of the p34ct vWA like Domain.....	71
Figure IV.29: Topology of the p34ct vWA Domain.....	72
Figure IV.30: Multiple Sequence Alignment of p34 Proteins.....	73
Figure IV.31: Location of the p34ct C-terminal Domain.....	73
Figure IV.32: Crystal Packing of the p34ct vWA Domain.....	74
Figure IV.33: Electrostatic Surface Potential of the p34ct vWA Domain.....	75

Figure IV.34: Superposition of p34ct vWA with other vWA Domains.....	77
Figure IV.35: Superposition of p34ct vWA with Complex Structures of vWA Domain Homologues.....	78
Figure IV.36: Crystals of p34ct p44ct MC and X-Ray Diffraction.	81
Figure IV.37: Structure of the p34ct p44ct Minimal Complex.....	82
Figure IV.38: Comparison of p44ct C4C4 to p44h C4C4.	83
Figure IV.39: Multiple Sequence Alignment of RING Domains.	84
Figure IV.40: Crystal Packing of the p34ct p44ct Minimal Complex.	86
Figure IV.41: Interaction Sites in the p34ct p44ct Minimal Complex (Overview).	87
Figure IV.42: Interaction Sites in the p34ct p44ct Minimal Complex (Close Up View).....	88
Figure IV.43: Purified Variants of p34ct 1-277 and p44ct 368-534.....	89
Figure IV.44: Thermofluor Analysis of p34ct and p44ct Variants.....	90
Figure IV.45: Circular Dichroism Spectra of p34ct and p44ct Variants.....	90
Figure IV.46: Interaction Studies using Native Agarose Gel-Electrophoresis.....	92
Figure IV.47: Interaction Studies with p34ct Variants and p44ct Wild Type using Analytical SEC.....	93
Figure IV.48: Interaction Studies with p34ct Wild Type and p44ct Variants using Analytical SEC.....	94
Figure IV.49: Interaction Studies with p34ct and p44ct Proteins using ITC Measurements.....	95

VII.2 List of Tables

Table I.1: Composition and Function of the TFIID Complex in Human and Yeast.....	5
Table III.1: Setup for Standard PCR Reaction.....	24
Table III.2: Program for Standard PCR Reaction.....	24
Table III.3: Setup for Standard DNA Double Digestion.....	25
Table III.4: Setup for SLIC Cloning (Stage 1).....	26
Table III.5: Setup for SLIC Cloning (Stage 2).....	26
Table III.6: Setup for Standard Colony-PCR Reaction.....	27
Table III.7: Program for Standard Colony-PCR Reaction.....	27
Table III.8: Setup for DNA Hybridization.....	28
Table III.9: Setup for EMSA and NAGE Interaction Studies.....	36
Table III.10: Parameters for ITC Measurements.....	37
Table III.11: Setup for Thermofluor Experiments.....	38
Table IV.1: Thermofluor Melting Temperatures.....	61
Table IV.2: Results from Molar Mass Analysis of p34ct and p44ct Variants.....	63
Table IV.3: Mutants of p34ct 1-277 and p44ct 368-534.....	89

VIII. Abbreviations

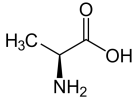
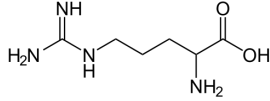
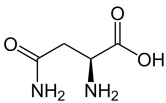
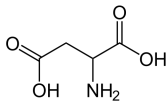
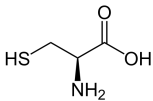
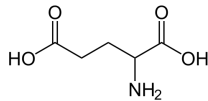
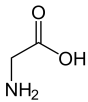
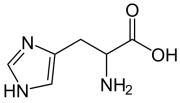
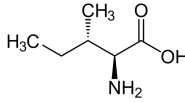
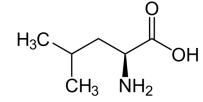
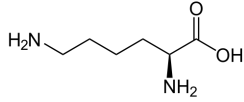
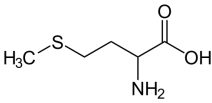
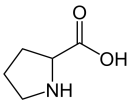
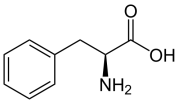
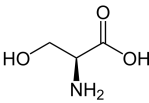
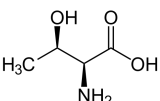
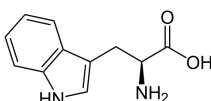
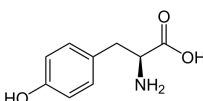
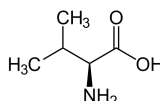
Abbreviation	Detailed Description	Synonym / Type
3D	3-dimensional	-
aa	Amino acid	-
Amp	Ampicillin	Antibiotic
ATP	Adenosine triphosphate	-
BESSY	Berliner Elektronen-Speicherring Synchrotron	Synchrotron in Berlin (D)
CAK	Cyclin activating kinase	-
Cam	Chloramphenicol	Antibiotic
CAPS	3-(Cyclohexylamino)-1-propanesulfonic acid	Buffer
CD	Circular dichroism	-
CHES	2-(Cyclohexylamino)ethanesulfonic acid	Buffer
CSS	Complex significance score	-
ct	<i>Chaetomium thermophilum</i>	-
CTD	C-terminal domain	-
CV	Column volume	-
Da	Dalton	Molecular mass in g/mol
ddH ₂ O	Double distilled millipore water	-
DNA	Deoxyribonucleic acid	-
DNA Pol	DNA Polymerase	-
dNTP	Deoxyribonucleotide triphosphate	-
DPS	DNA binding protein from starved cells	-
dsDNA	Double stranded DNA	-
DTT	1,4-Dithiotreitol	Reducing agent
ϵ	Extinction coefficient	-
<i>E. coli</i>	<i>Escherichia coli</i>	-
EC	Enzyme classification	-
EM	Electron microscopy	-
EMSA	Electrophoretic mobility shift assay	-
ESRF	European synchrotron radiation facility	Synchrotron in Grenoble (F)
EtBr	Ethidium bromide	-
<i>et al.</i>	et alii (m) / et aliae (f) / et alia (n)	and others
FCS	Fetal calf serum	-
FPLC	Fast protein liquid chromatography	-
g	Gram / acceleration of gravity (9.81 m/s ²)	-
Gen	Gentamicin	-
GG-NER	Global genome nucleotide excision repair	-
h	Hours	-
His	Histidine	Amino acid

hpi	Hours post induction/infection	-
IEX	Ion exchange chromatography	-
IPTG	Isopropyl- β -D-thiogalactopyranoside	-
ITC	Isothermal titration calorimetry	-
k	Kilo	One thousand, 10^3
Kan	Kanamycin	Antibiotic
KCl	Potassium chloride	-
λ	Wavelength	-
L	Liters	-
LB	Lysogeny-Broth / Luria-Bertani Broth	Culture medium
μ	Micro	One part in a million, 10^{-6}
m	Milli	One part in a thousand 10^{-3}
M	Molarity (mol/l)	Molar concentration in mol/l
MALS	Multi angle light scattering	-
MCS	Multiple cloning site	-
min	Minutes	-
MOI	Multiplicity of infection	-
MW	Molecular weight	Molecular mass in Dalton
MWCO	Molecular weight cut off	-
n	Nano	One part in a billion, 10^{-9}
NaCl	Sodium chloride	-
NAGE	Native agarose gel electrophoresis	-
NER	Nucleotide excision repair	-
Ni-MAC	Nickel metal affinity chromatography	-
Ni-TED	Nickel tris-carboxymethyl ethylene diamine	-
nm	Nanometer	One billionth meter, 10^{-9} m
NMR	Nuclear magnetic resonance	-
nt	Nucleotide / nucleotides	-
OD ₆₀₀	Optical density at 600 nm	-
PA	Polyacrylamide	-
PAGE	Polyacrylamide gel electrophoresis	-
PBS	Phosphate buffered saline	-
PCR	Polymerase chain reaction	-
PEG	Polyethylene glycol	-
pH	pH value	Strength of acid / base, $-\lg [H_3O^+]$
pI	Isoelectric point	pH for neutral electrical charge
Pol	Polymerase	-
r.m.s.	Root mean square	-
r.m.s.d.	Root mean square deviation	-
RB	Running buffer	-
RING	Really Interesting New Gene	Type of zinc finger domain

RNA	Ribonucleic acid	-
RT	Room temperature	-
SB	Sample buffer	-
SDS	Sodium dodecyl sulfate	-
SEC	Size exclusion chromatography	Gelfiltration
<i>Sf</i>	<i>Spodoptera frugiperda</i>	Insect cells
ssDNA	Single stranded DNA	-
TAE	Tris acetic acid	Buffer
TC-NER	Transcription coupled nucleotide excision repair	-
TCEP	Tris-(2-carboxyethyl)-phosphine	Reducing agent
TE	Tris EDTA	Buffer
TEMED	Tetramethyl ethylene diamine	-
Tet	Tetracyclin	-
TEV	Tobacco etch virus	-
TF	Transcription factor	-
TFIIH	General transcription factor II H	-
T _m	Melting temperature	-
Tn7	Transposon / transposable element 7	-
TRIS	Tris(hydroxymethyl)-aminomethane	-
V	Volt	-
var.	variable	-
vWA	von Willebrand Factor A Domain	-
vWF	von Willebrand Factor	-
X-Gal	Bromo-chloro-indolyl β-D-galactopyranoside	Dye for blue/white screening

IX. Appendix

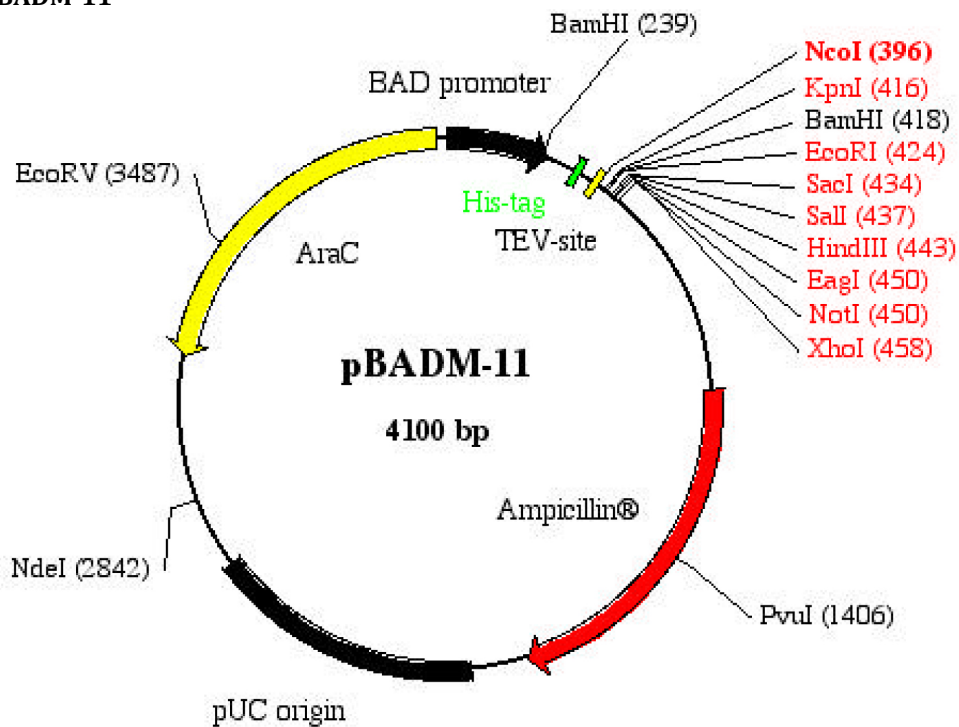
IX.1 Table of Amino Acids

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

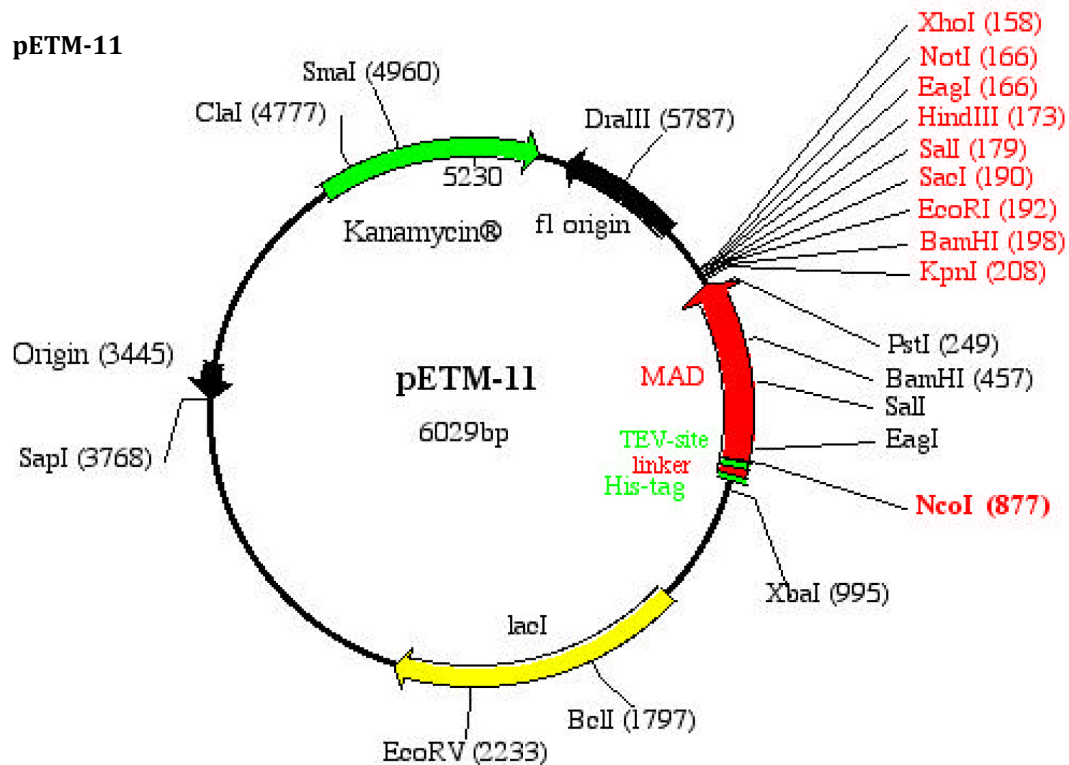
All amino acid chemical structure images have been adapted from www.wikipedia.org.

IX.2 Detailed Maps of the Vectors used for Protein Expression

IX.2.1 pBADM-11

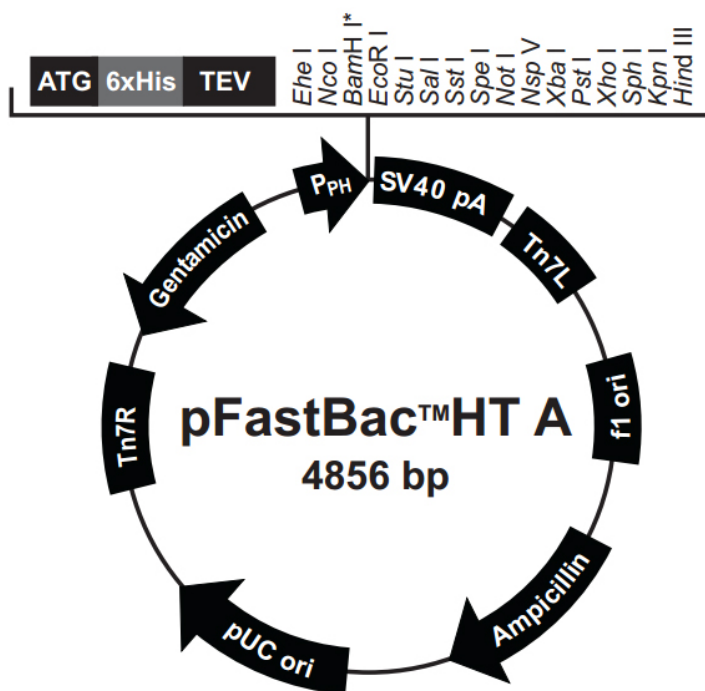


IX.2.2 pETM-11



Both images were retrieved from the vector documentation provided by EMBL on www.embl.de.

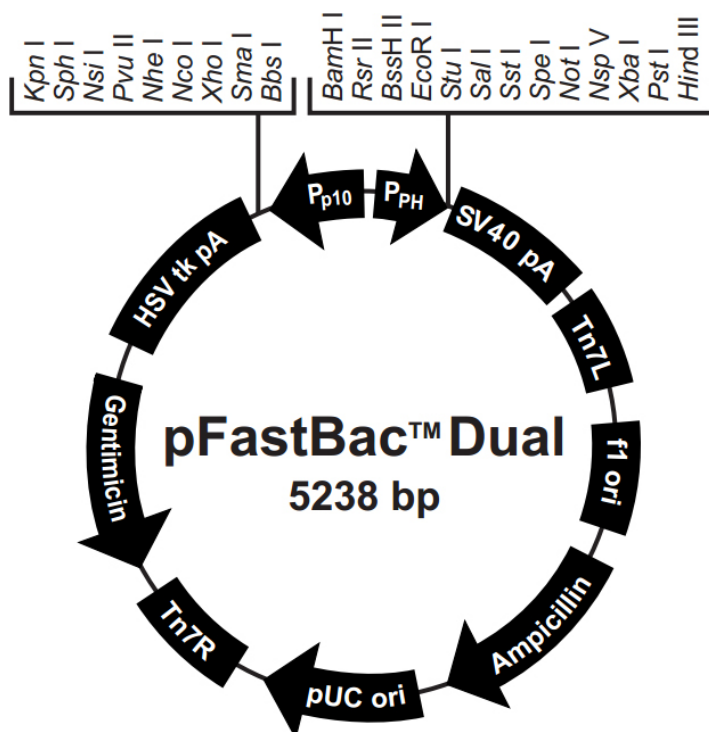
IX.2.3 pFastBac HTa

**Comments for pFastBac™ HT A**
4856 nucleotides

f1 origin: bases 2-457
 Ampicillin resistance gene: bases 589-1449
 pUC origin: bases 1594-2267
 Tn7R: bases 2511-2735
 Gentamicin resistance gene: bases 2802-3335
 Polyhedrin promoter (P_{PH}): bases 3904-4032
 Initiation ATG: bases 4050-4052
 6xHis tag: bases 4062-4079
 TEV recognition site: bases 4101-4121
 Multiple cloning site: bases 4119-4222
 SV40 polyadenylation signal: bases 4240-4480
 Tn7L: bases 4509-4674

*Frameshift occurs at the
BamH I site in each vector

IX.2.4 pFastBac Dual

**Comments for pFastBac™ Dual**
5238 nucleotides

f1 origin: bases 102-557
 Ampicillin resistance gene: bases 689-1549
 pUC origin: bases 1694-2367
 Tn7R: bases 2611-2835
 Gentamicin resistance gene: bases 2902-3435
 HSV tk polyadenylation signal: bases 3992-4274
 Multiple cloning site: bases 4274-4337
 p10 promoter (P_{p10}): bases 4338-4459
 Polyhedrin promoter (P_{PH}): bases 4478-4606
 Multiple cloning site: bases 4606-4704
 SV40 polyadenylation signal: bases 4722-4962
 Tn7L: bases 4991-5156

Both images were retrieved from the Bac-to-Bac® Baculovirus Expression System User Manual
 provided by Invitrogen on www.lifetechnologies.com.

IX.3 Detailed Information on p34ct and p44ct Constructs

IX.3.1 Constructs of p34ct

The following constructs of p34ct were used in the present study. The values for molecular weight (MW), the extinction coefficient (ϵ) and the theoretical isoelectric point (pI) were derived from sequence analysis using the ProtParam tool on the ExPASy server [140]. All sizes include the His₆-Tag, if present.

Construct	MW [Da]	ϵ [M ⁻¹ cm ⁻¹]	Size [AA]	pI	Tag
p34ct pBADM-11	46,767.0	28,420	455	6.94	N-Term His ₆
p34ct pETM-11 CH ₆	44,463.4	25,440	435	6.29	C-Term His ₆
p34ct [1-387] pBADM-11	42,563.1	28,420	413	6.29	N-Term His ₆
p34ct [1-368] pBADM-11	40,711.0	26,930	394	6.41	N-Term His ₆
p34ct [1-277] pBADM-11	31,941.0	26,930	303	6.37	N-Term His ₆
p34ct pETM-11 CH ₆	44,463.4	25,440	435	6.29	C-Term His ₆

IX.3.2 Constructs of p44ct

The following constructs of p44ct were used in the present study. The values for molecular weight (MW), the extinction coefficient (ϵ) and the theoretical isoelectric point (pI) were derived from sequence analysis using the ProtParam tool on the ExPASy server [140]. All sizes include the His₆-Tag, if present.

Construct	MW [Da]	ϵ [M ⁻¹ cm ⁻¹]	Size [AA]	pI	Tag
p44ct pBADM-11	61,144.9	45,380	560	6.53	N-Term His ₆
p44ct [55-534] pBADM-11	55,521.8	43,890	507	6.36	N-Term His ₆
p44ct [77-534] pBADM-11	53,218.5	32,890	487	6.90	N-Term His ₆
p44ct [1-285] pBADM-11	34,974.3	29,910	311	5.86	N-Term His ₆
p44ct [55-285] pBADM-11	29,351.2	28,420	258	5.54	N-Term His ₆
p44ct [75-285] pBADM-11	27,047.9	17,420	238	5.98	N-Term His ₆
p44ct [101-285] pBADM-11	23,985.2	17,420	212	5.78	N-Term His ₆
p44ct [368-534] pETM-11 NH ₆	21,084.6	16,960	194	7.18	N-Term His ₆
p44ct [368-534] pETM-11 CH ₆	18,643.9	13,980	173	8.14	C-Term His ₅
p44ct [368-534] pETM-11 NoTag	17,958.2	13,980	168	8.13	-

IX.4 Secondary Structure Prediction of p34ct using the Phyre² Algorithm

The secondary structure prediction was derived using the Phyre² server (<http://www.sbg.bio.ic.ac.uk/>) with p34ct full-length including its N-terminal hexa-histidine-tag as input sequence.



Regions marked in color represent structural elements visible in the p34ct structure solved (Figure IV.28).

Residues of the putative noncontiguous MIDAS motif are boxed in grey, while the borders of all p34ct variants used in the present study are indicated by brackets according to the following color code:

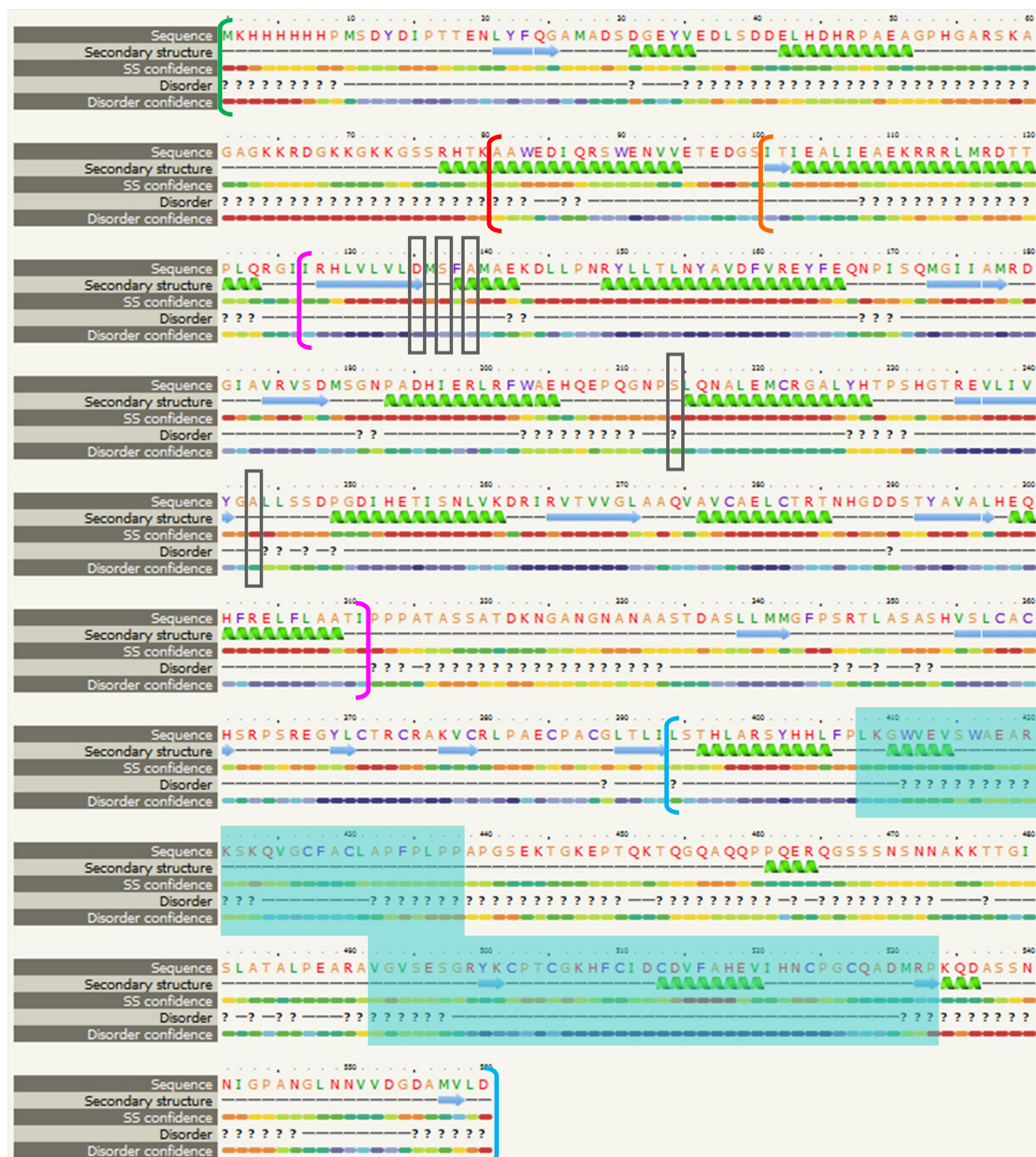
[p34ct 1-387]

[p34ct 1-368]

[p34ct 1-277]

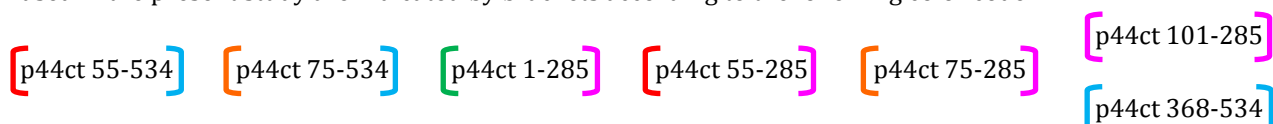
IX.5 Secondary Structure Prediction of p44ct using the Phyre² Algorithm

The secondary structure prediction was derived using the Phyre² server (<http://www.sbg.bio.ic.ac.uk/>) with p44ct full-length including its N-terminal hexa-histidine-tag as input sequence.



Regions marked in cyan represent structural elements of the p44ct 368-534 C4C4 zinc binding domain visible in the structure of the p34ct p44ct minimal complex (Figure IV.37).

Residues of the putative noncontiguous MIDAS motif are boxed in grey, while the borders of all p44ct variants used in the present study are indicated by brackets according to the following color code:



IX.6 Data Collection and Refinement Statistics for p34ct

Data Collection	p34ct	p34ct KI	p34ct 1-277
Space Group	F 4 ₁ 3 2	F 4 ₁ 3 2	F 4 ₁ 3 2
Cell Dimensions			
a, b, c [Å]	257.31, 257.31, 257.31	257.07, 257.07, 257.07	257.11, 257.11, 257.11
α , β , γ [°]	90, 90, 90	90, 90, 90	90, 90, 90
Monomers per AU	1.0	1.0	1.0
Resolution [Å]	3.8	4.2	2.7
Wavelength [Å]	0.8726	1.6000	0.9199
Observed Reflection	77,505	468,424	1,432,226
Unique Reflections	7,694	10,090	20,598
Completeness [%]	99.8 (100.0)	99.9 (100.0)	100.0 (100.0)
Redundancy	10.1	46.4	69.5
I / σ I	16.2 (1.4)	18.7 (2.1)	22.8 (1.7)
R _{merged} -F [%]	14.2 (142.1)	14.5 (78.8)	
Rp _{im} [%]			1.9 (51.0)
Completeness [%]	99.8 (100.0)	99.9 (100.0)	100.0 (100.0)
Redundancy	10.1	46.4	69.5
SigA _{no}		1.288	
Software	XDS	XDS	iMOSFLM, SCALA
Refinement	p34ct	p34ct KI	p34ct 1-277
Software	Phenix		Phenix
Resolution	3.8		2.7
Number of Atoms	2,058		3,190
R _{work} (R _{free})	38.2 (40.0)		22.2 (23.5)
Wilson B-Factor [Å ²]	142.5		92.8
R.m.s. Deviations			
Bond Lengths [Å]	0.004		0.005
Bond Angles [°]	0.924		0.835
Chirality [°]	0.062		0.028
Planarity [°]	0.006		0.004
Dihedral	16.793		12.624
Ramachandran Plot			
Favored [%]	85.39		98.52
Allowed [%]	9.74		1.48
Outliers [%]	4.87		0.00
Rotamer Outliers [%]	23.70		0.00

IX.7 Data Collection and Refinement Statistics for p34ct p44ct MC

Data Collection	p34ct p44ct MC
Space Group	P 6 ₃ 2 2
Cell Dimensions	
a, b, c [Å]	138.21, 138.21, 95.23
α , β , γ [°]	90, 90, 120
Monomers per AU	1.0
Resolution [Å]	3.7
Wavelength [Å]	0.8726
Observed Reflections	49,937
Unique Reflections	6,140
Completeness [%]	99.7 (99.0)
Redundancy	8.1
I / σ I	9.57 (1.17)
Rmeas [%]	21.3 (168.5)
Rpim [%]	7.9 (60.5)
Software	XDS, SCALA
Refinement	p34ct p44ct MC
Software	Phenix
Resolution	3.7
Number of Atoms	4,164
R _{work} (R _{free})	27.8 (32.4)
Mean B-Factor [Å ²]	138.3
R.m.s. Deviations	
Bond Lengths [Å]	0.011
Bond Angles [°]	1.297
Chirality [°]	0.102
Planarity [°]	0.004
Dihedral	14.294
Ramachandran Plot	
Favored [%]	92.45
Allowed [%]	7.17
Outliers [%]	0.38
Rotamer Outliers [%]	2.25

Acknowledgment

First, I would like to express my sincere gratitude to Prof. Caroline Kisker, for giving me the opportunity to work in her lab during both my diploma and doctoral studies. The challenging subject and excellent research environment, together with her constant support, her kindness and admirable devotion, are things I will always remember as a positive experience.

Next to my primary supervisor, special thanks go to all the other members of my supervisory committee, Prof. Thomas Müller, Prof. Hanspeter Naegeli and Dr. Robert Hock, for their guidance throughout my doctoral studies as well as their countless suggestions on how to improve my work.

Many thanks also to Dr. Jochen Kuper and Prof. Hermann Schindelin, who helped me to learn a great deal, especially with respect to the data collection and computational part of x-ray crystallography.

In addition, several people in the lab deserve a big “thank you” for all their contribution to my work, including their ideas and technical assistance, which often made life a lot easier. Especially, I would like to mention Agnes Elias and Wolfgang Kölmel, my recent co-workers on the TFIIH project, as well as Bodo Sander and Carolyn Delto, who helped me to learn and perform the ITC experiments. Special thanks also go to Gudrun Michels, for her help with the purification of the most recent batch of p34ct and p44ct mutants.

Furthermore, I would like to thank all the people who spent their hours at a synchrotron to collect x-ray diffraction data, including Christing Schäfer, Dr. Maria Hirschbeck, Dr. Jochen Kuper, Dr. Heide Marie Roth and Dr. Marcus Resch. Also Florian Rohleder, Sebastian Kaiser and Dr. Johannes Schiebel deserve my thanks, for their help in maintaining a relaxed and steady atmosphere in the office and for many fruitful discussions.

Apart from my lab colleagues, I am very grateful to Elisabeth Schönwetter, Imke Spöring and Olga Fust, who contributed to the TFIIH project as part of their bachelor thesis or practical course, respectively. In addition, I would like to thank Dr. Joop van den Heuvel and Prof. Thomas Müller for their advice with respect to insect cell culture and their willingness to share several samples of their *Sf21* and *Sf9* insect cells.

Next, I kindly thank Ed Hurt, from the University of Heidelberg, for the provision of *C. thermophilum* DNA and Agnes Elias for the initial cloning the p34ct and p44ct full-length constructs. Furthermore, the synchrotron staff of both ID 23.2 and ID 29 of the ESRF in Grenoble as well as BL14.1 of the BESSY II in Berlin deserve many thanks for their excellent beamline support.

I also thank Prof. Clemens Steegborn, from the University of Bayreuth, who made it possible for me to already start in his lab, before I was completely finished with my doctoral thesis.

Finally, I would like to express my heartfelt gratitude to my parents, Hans-Peter and Elfriede, as well as my beloved wife Veronika. I will never forget your constant support and encouragement throughout the years.

Without you, and the elk, this endeavour would have never been possible. Thank you so much!

A last big “thank you” goes to Veronika and her brother Wolfram for proofreading the draft of this thesis ...

... and to Douglas Adams for his „Hitchhiker’s Guide to the Galaxy” and his inspiring words to live by ...

“Don’t panic!”

Affidavit

Affidavit

(Eidesstattliche Erklärung)

I hereby declare that my thesis entitled

**“Structural Characterization of the TFIID Subunits
p34 and p44 from *C. 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 a similar form.

Weidenberg, 23.05.2014

Place

Date

Signature

