Structural and functional characterization of the human and yeast TFIIIC complex

Laura Jochem

Division of Structural Biology
The Institute of Cancer Research
University of London

This thesis is submitted for the degree of Doctor of Philosophy.

July 2019

Declaration

The work described in this thesis was carried out at the Institute of Cancer Research, London, under the supervision of Dr Alessandro Vannini.

I, Laura Jochem, declare that the work presented in this thesis is my own and, where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature

Date

Abstract

TFIIIC is one of the conserved transcription factors required for RNA Polymerase (Pol) III transcription, which transcribes DNA into untranslated RNAs, such as tRNAs. However, TFIIIC's mechanistic role in Pol III regulation is poorly understood, partly due to the absence of high-resolution structures. In particular, it is unclear how TFIIIC accommodates variable spacings between its recognition motifs (A and B Box) on the DNA. To better understand TFIIIC's role in Pol III regulation, I biophysically and structurally characterized the human and yeast TFIIIC complexes.

I was able to recombinantly express and purify the human and yeast TFIIIC complexes, enabling the study of preinitiation complexes at tRNA genes in eukaryotes. Preliminary structural analysis of the human TFIIIC complex suggested a two-domain organization, as suspected previously. Contrary to the hypothesis that the B Box was indispensable for DNA binding, interaction studies suggested that the A Box element was sufficient for DNA binding.

Studies of the apo yeast TFIIIC complex via native mass spectrometry and SAXS revealed that it is a flexible complex adopting several conformations. This behavior was confirmed by cryo-EM: multi-body refinements and hierarchical masking suggested extensive flexibility between and within the two domains of TFIIIC. Despite this, it was possible to obtain an overall reconstruction of the apo-complex at 15 Å, and sub-nanometre resolution reconstructions of parts of the complex. SAXS and cryo-EM studies suggested that after DNA binding, TFIIIC retained its flexibility while extending further than apo TFIIIC complexes.

In conclusion, the presented work has demonstrated the extreme flexibility and heterogeneity of TFIIIC. It has revealed a first glimpse of apo and DNA-bound TFIIIC structures, which may facilitate solving near-atomic resolution structures in the future. This improves our understanding of Pol III transcription in eukaryotes generally and could help to understand how this regulation breaks down in cancerous cells.

Acknowledgements

First of all, I would like to thank my supervisor Dr Alessandro Vannini for all the support and freedom he has given me in the past four years to pursue this interesting project.

Secondly, I would like to thank Guiller for the amazing support, the countless hours of discussing my PhD project with me, for all the troubleshooting that we have done together, for always having an open ear for me even in the most stressful moments and for being an amazing, funny and caring colleague.

I would like to thank Ewan for teaching me how to use different EM processing software and for his amazing help throughout. I would not have been able to process my data without him and it has always been great fun working together.

Of course, I also want to thank all remaining current and former Vannini team members: Karishma, Nicolas, Jerome, Erin, Thanga, Rosie, Karolina, Andri and Helen for interesting scientific discussions and for making the ICR a place in which I enjoyed doing my PhD. In particular, I would like to thank Erin for helping me with the biGBac cloning and Helen for always being up for a chat and for looking out for everyone in the lab so well. I would also like to thank all current and former members of Sebastian's and Ed's team: Katie, Michael, Catherine, Laura, Pete, Iris, Mariola, Ruth, Yexin, Sam, Oviya, Sarah, Jennifer and Aylin for a great working environment, fantastic support and for sharing their scientific knowledge with me. I would like to thank Fabienne for her patience to screen through all my grids with me and her fantastic help. I would also like to thank Ed for sharing his large EM knowledge with me.

From the Division, I would like to thank Chris for running the IT infrastructure so smoothly and I would like to thank Nora for helping me with the crystallography side and for taking my samples to the SAXS beamline.

I would also like to thank everyone else in the Structural Biology Division that I haven't mentioned specifically yet and all my fellow PhD students who have accompanied me through the last years, in particular Kunzah, Nadia, Magda and Gabby.

I would like to extend my thanks to all the collaborators: Ian Taylor at the Francis Crick Institute for his great help to set up SEC-MALS experiments; Becky, Emma and Dan for the fantastic support during the cryo-EM data collections in Leeds; Andrew and Colin from the UCL School of Pharmacy who performed a lot of mass spectrometry analysis for me and Jack, Bruno and Prof. Perdita Barran for their great work in performing the native mass spectrometry experiments.

Last but not least, I would like to thank my brother Dominik, Nhoa and my Mum and Dad for all the love and their tireless support to encourage me not to give up even during the toughest times throughout my PhD. I would also like to thank my closest friends Ines, Yvi and Regina for always being there for me. Finally, I would like to thank Joe for making the final year to the best year of my PhD, and hopefully for many more years to follow.

Table of Contents

List of figures

List of Tables

List of Abbreviations

A	Adenine
ADNP	Activity Dependent Neuroprotector Homeobox
APS	Ammonium persulfate
Bdp1	B double prime 1
BME	Beta-mercaptoethanol
bp	Base pair
Brf1	B-related factor 1
BSA	Bovine serum albumin
C	Cytosine
CC	Closed complex
CCD	Charge coupled device
CCF	Cross correlation function
CE	Collision energy
CEN	Centromeric nucleosomes
CHD4	Chromodomain Helicase DNA Binding Protein 4
cisTEM	Computational Imaging System for Transmission Electron Microscopy
Cryo-EM	Cryogenic electron microscopy
CSD	Charge state distributions
CTF	Contrast transfer function
CV	Column volumes
Cy3	Cyanine 3
Cy5	Cyanine 5
DAMAVER	Dummy atom model averaging
DAMMIF	Dummy atom model minimization fast
DAMMIN	Dummy atom model minimization
dATP	Deoxyadenosine triphosphate

DDD	Direct detection device
Dmax	Maximum diameter
DNA	Deoxyribonucleic acid
DQE	Detective quantum efficiency
dRI	Differential refractive index
DSE	Distal Sequence Element
DTT	Dithiothreitol
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
ETC	Extra TFIIIC site
eWH	Extended winged helix
FBS	Fetal bovine serum
FEG	Field emission gun
FSC	Fourier Shell correlation
G	Guanine
GEC	Gene expression cassette
GTF	General transcription factor
GraFix	Gradient Fixation
Hepes	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
HRP	Horse radish peroxidase
ICR	Internal control region
IM	Ion mobility
IST	In-source trapping
ITC	Initially transcribing complex
KCl	Potassium chloride
kDa	Kilo Dalton
keV	Kilo electron volt
LB	Lysogeny broth

LC-MS/MS	Liquid chromatography tandem mass spectrometry
m/z	Mass to charge ratio
MES	2-(N-morpholino)ethanesulfonic acid
$MgCl_2$	Magnesium chloride
MS	Mass spectrometry
NaCl	Sodium chloride
NEB	New England Biolabs
OC	Open complex
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PBP	PSE-binding protein
PC	Principal component
PCR	Polymerase chain reaction
PDB	Protein data bank
PGC	Polygene cassette
PMSF	Phenylmethanesulfonyl fluoride
Pol	Polymerase
PSE	Proximal Sequence Element
PSF	Point spread function
PTF	PSE transcription factor
Relion	Regularised Likelihood Optimization
Rg	Radius of gyration
Rh	Hydrodynamic radius
Rh(Q)w	Hydrodynamic radius (SAXS)
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
SAXS	Small-angle X-ray scattering
SDS	Sodium dodecyl sulfate
SEC	Size exclusion
SEC-MALS	Size exclusion multi-angle light scattering

SNAPc	snRNA activating protein complex
SNR	Signal-to-noise ratio
snRNA	Small nuclear RNA
SOC	Super optimal broth with Catabolite repression
T	Thymine
TAE	Tris-acetate-EDTA
TAF	TBP associated factor
TB	Terrific Broth
TBE	Tris/Borate/EDTA
TBP	TATA binding protein
TCEP	Tris(2-carboxyethyl)phosphine
TE	Tris EDTA
TEM	Transmission electron microscope
TEMED	Tetramethylethylenediamine
TEV	Tobacco Etch Virus
TFIIIA	Transcription Factor IIIA
TFIIIB	Transcription Factor IIIB
TFIIIC	Transcription Factor IIIC
TPR	Tetratricopeptide repeat
Tris	tris(hydroxymethyl)aminomethane
tRNA	Transfer RNA
TSS	Transcription start site
UV	Ultraviolet
VPP	Volta phase plate
YFP	Yellow fluorescent protein

1. Introduction

1.1. The central dogma of molecular biology

The deoxyribonucleic acid (DNA) is the fundamental entity for all known forms of life. DNA stores the hereditary biological information that is encoded in the genes of a cell.

The very first crude precipitate of DNA was isolated in 1869 by Johann Friedrich Miescher during his studies on leucocytes (Dahm 2005). Miescher called the newly discovered compound “nuclein” since it was isolated from the nucleus (Dahm 2005). Seventy five years after the initial discovery of the DNA, Oswald T. Avery, Colin MacLeod and Maclyn McCarthy showed that the hereditary genetic information was stored inside the DNA and not inside proteins as it was initially assumed (Avery et al. 1944). Later, the composition of the DNA was deciphered. The core elements of the DNA were identified as a nitrogenous base: adenine (A), thymine (T), guanine (G) or cytosine (C), a sugar and a phosphate group (Figure 1). In 1953, James Watson and Francis Crick were able to characterize the molecular structure of DNA as a double helix (Watson & Crick 1953).

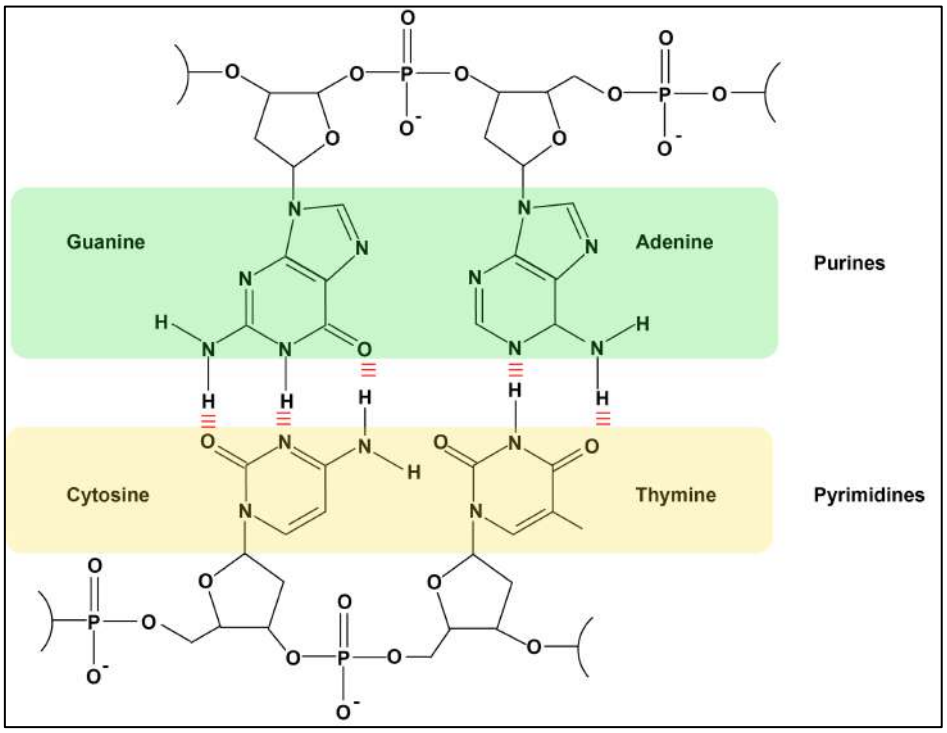


Figure 1: Schematic overview of the DNA.

The purine nucleobases Guanine (G) and Adenine (A) as well as the pyrimidine nucleobases Cytosine (C) and Thymine (T) are forming nucleosides by forming a covalent bond with β-D-Deoxyribose sugar. Nucleotides subsequently bind to phosphate groups, forming the phosphate backbone. Coupled nucleotides form the deoxyribonucleic acid (DNA). A and T are paired via two hydrogen bonds (shown in red) and G and C are paired via three hydrogen bonds. Purines are highlighted in green and pyrimidines in yellow. Adapted from (Harding et al. 2018).

For cellular expression of genetic information, the first step is to generate RNA using the DNA as a template. This process is performed by DNA-dependent RNA Polymerases (RNA Pol) and can produce different types of RNA such as messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA) and small nuclear (snRNA) (Clancy 2008). With this highly regulated process called transcription, cells ensure that proteins are correctly synthesized at the right time, in the correct place and at the desired rate. In a subsequent process, called translation, the genetic information contained in the mRNA is used by the ribosomes to synthesize proteins. Proteins are essential to sustain the life of a cell.

They are formed by a chain of amino acids which folds into secondary structures such as alpha helices and beta sheets via the formation of chemical interactions (Figure 2B). The genetic flow of information, (DNA → RNA → proteins), is usually described as the "central dogma of molecular biology" (Figure 2A). The original meaning was stated by Francis Crick: "The central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred from protein to either protein or nucleic acid" (Crick 1970).

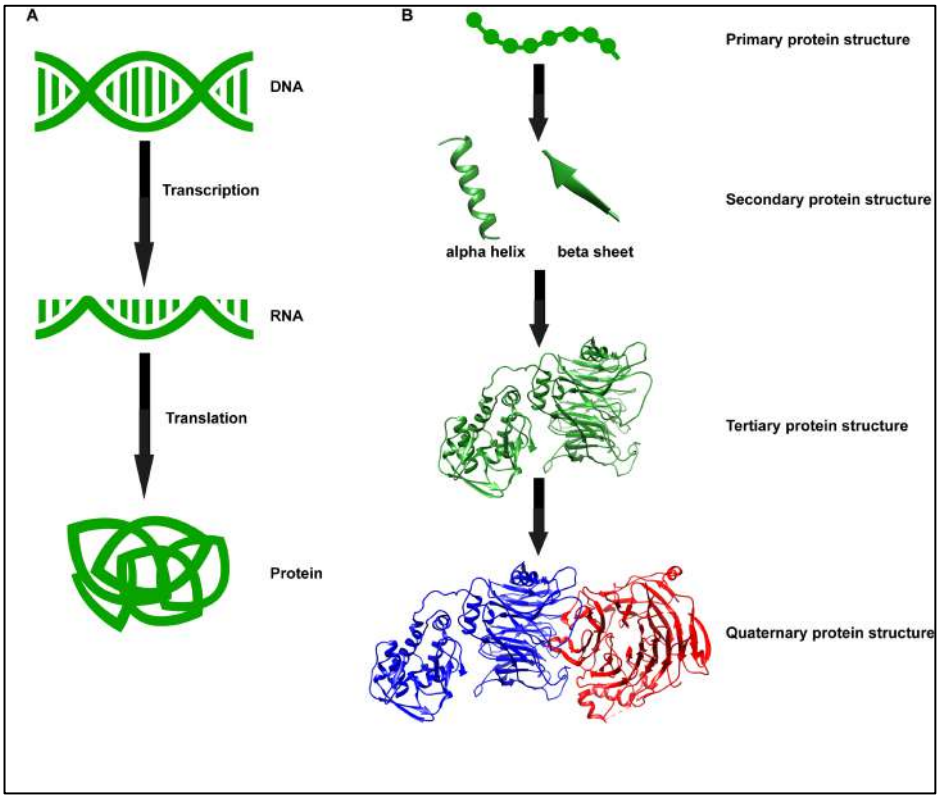


Figure 2: From DNA to proteins.

A: The central dogma of molecular biology. Double-stranded DNA is transcribed into single stranded RNA in a process called transcription. Subsequently, the RNA gets translated into proteins in a process called translation. B: Protein folding. In the primary protein structure, amino acids are connected via peptide bonds to form a polypeptide chain. The Polypeptide chain gets then compacted into regular motifs such as alpha helices, beta sheets or beta turns which are held together by hydrogen bonds. Tertiary protein structure: The secondary structure motifs fold together to form a 3D structure and give the protein its overall shape. Quaternary structure is formed when several polypeptide chains interact with each other to function as a multisubunit complex.

1.2. Transcription in the three domains of life

During transcription, DNA gets transcribed into RNA by RNA Pols in all living organisms across all domains of life: Eukarya, Bacteria and Archaea (Werner & Grohmann 2011). Eukaryotes contain three different RNA Pols, whereas bacteria and archaea contain only one to carry out transcription (Carter & Drouin 2010). Besides the three RNA Pols in Eukaryotes, plants contain plant-specific RNA Pol IV, whose catalytic activity does not overlap with RNA Pol I, II and III (Onodera et al. 2005), and plant-specific RNA Pol V (Wierzbicki et al. 2008).

All RNA Pols are catalytically active multisubunit complexes, which are derived from a common ancestor (Werner & Grohmann 2011) and thus share a conserved core consisting of five subunits that are shown in blue in Figure 3A (Best et al. 2004). The composition and number of the other subunits, however, differs between the RNA Pols (Schramm & Hernandez 2002a; Vannini & Cramer 2012). These subunits are shown in green in Figure 3A. RNA Pol I, II and III all contain an active centre within a conserved ten-subunit core (Cramer et al. 2008). An overview of the different subunits of RNA Pols in Bacteria, Eukaryotes and Archaea is shown in Figure 3 and Table 1.

The core comprises the duplex DNA-binding channel, which is formed by regions known as the jaws, and the wall that surrounds the catalytic centre containing two Mg^{2+} ions. Above the DNA-binding channel lies the clamp. The region behind the wall serves as an assembly platform for additional factors (Figure 3B, C) (Werner & Grohmann 2011).

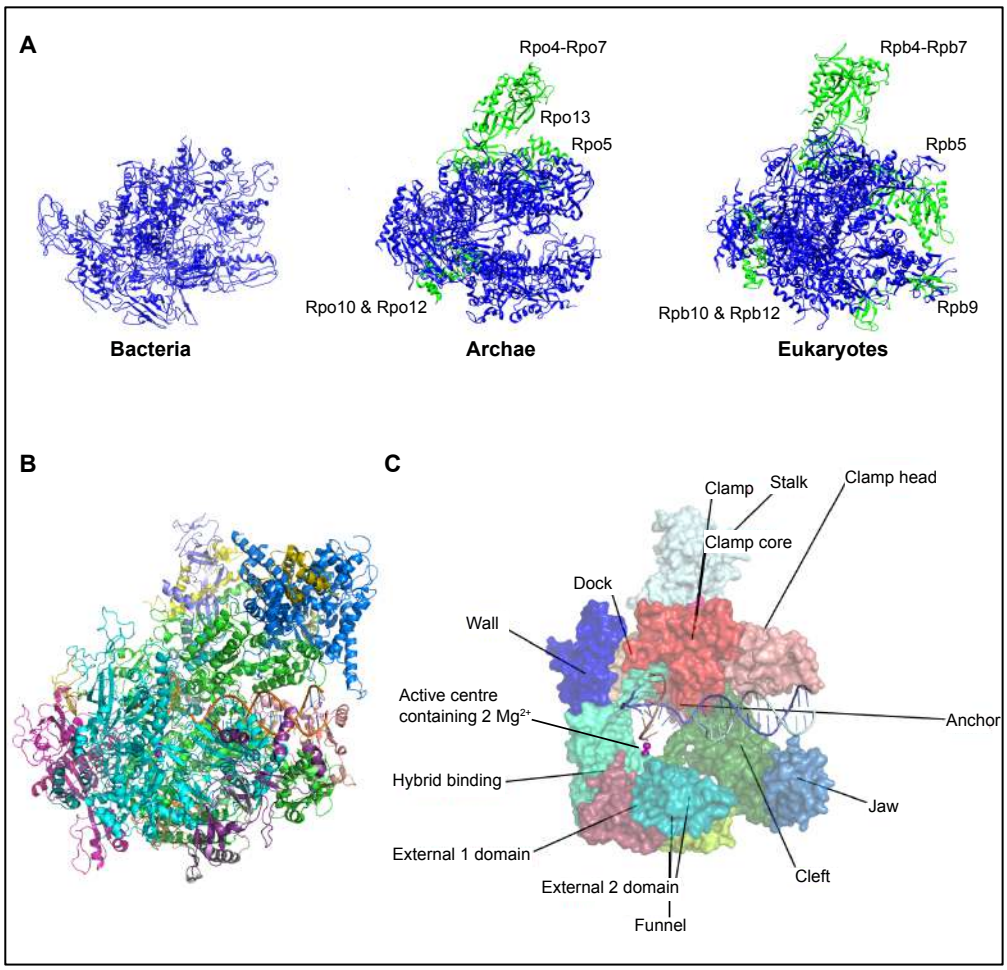


Figure 3: Comparison of the overall architecture RNA Pols across all kingdoms of life.
A: The bacterial Thermus aquaticus RNA Pol was taken from protein data bank entry (PDB) entry: 1I6V, the archaeal RNA Pol from PDB entry: 2WAQ and the eukaryotic RNA Pol from PDB entry: 3FKI. Conserved subunits are shown in blue and unique subunits to Archaea and Eukaryotes are shown in green as described in (Werner & Grohmann 2011). B: The yeast RNA polymerase III elongation complex, taken from PDB entry: 5FJ8. C: Zoom into the structure shown in B, superimposed with the RNA polymerase II elongation complex, taken from PDB entry 2E2H, containing 2 Mg^{2+} ions in the active centre. Selected domains around the active site are shown as surfaces and labelled.

Table 1: Overview of RNA Pol subunits in Bacteria, Archaea and Eukaryotes.

The conserved five-subunit core is shown in blue, the additional core subunits are shown in light green, the stalk is shown in dark green and the general transcription factors are shown in orange. Table adapted from (Vannini & Cramer 2012; Werner & Grohmann 2011).

Bacteria	Archae	Eukaryotes			Function
RNAP	RNAP	RNAP II	RNAP I	RNAP III	
5 subunits	11 subunits	12 subunits	14 subunits	17 subunits	
β subunit	Rpo1	Rpb1	A190	C160	active centre
β subunit	Rpo2	Rpb2	A135	C128	active centre
α subunit	Rpo3	Rpb3	AC40	AC40	
α subunit	Rpo11	Rpb11	AC19	AC19	
		Rpb9	A12.2 N ribbon	C11 N ribbon	RNA cleavage
		TFIIIS C ribbon	A12.2 C ribbon	C11 C ribbon	RNA cleavage
	Rpo5	Rpb5	Rpb5	Rpb5	
ω subunit	Rpo6	Rpb6	Rpb6	Rpb6	
	Rpo8	Rpb8	Rpb8	Rpb8	
	Rpo10	Rpb10	Rpb10	Rpb10	
	Rpo12	Rpb12	Rpb12	Rpb12	
	Rpo4	Rpb4	A14	C17	initiation complex formation
	Rpo7	Rpb7	A13	C25	initiation complex formation
		Tfg1	A49 N term.	C37	initiation complex formation and start site selection
		Tfg2	A34.5	C53	initiation complex formation and start site selection
	TEα	Tfa1		C82	Open complex stabilization
	TEβ	Tfa2	A49 C term.	C34	Open complex stabilization
				C31	Open complex stabilization
	TBP	TBP	TBP	TBP	DNA binding
		TAFs			
	TFB	TFIIB	Rrn7	Brf1	TBP/Pol binding, DNA opening, start site selection

In order to initiate transcription, a pre-initiation complex (PIC) is assembled. The PIC is formed by the RNA Pol together with its general transcription factors (GTFs). Each RNA Pol contains its own set of GTFs, which are able to recognize the gene promoter and to guide the RNA Pol to the promoter region. The promoter defines the starting point for transcription and enables a regulated expression of a gene. It is thought that the positioning and general function of related GTFs are conserved (Vannini & Cramer 2012). After the PIC is formed, the complex adopts several functional states during the transcription cycle as shown schematically in Figure 4. Firstly, the PIC transitions into a closed complex

(CC) bound to the closed double-stranded promoter DNA. Upon melting of the DNA, a DNA bubble is formed and the single template strand is inserted into the active centre of the RNA Pol. This forms an open complex (OC) in which the RNA Pol II scans for the initiator sequence (Zhang & Dietrich 2005). Transcribing the first RNA nucleotides leads to an initially transcribing complex (ITC). After the ITC has transcribed more than a critical length of the RNA chain, usually after adding around 8-9 RNA nucleotides, the GTFs are released (Zawel et al. 1995). The complex transitions into an elongation complex (EC), which contains a 8-9 bp of DNA-RNA hybrid (Cheung & Cramer 2012). Different functional states of the RNA Pol II PIC are reviewed in (Nogales et al. 2017).

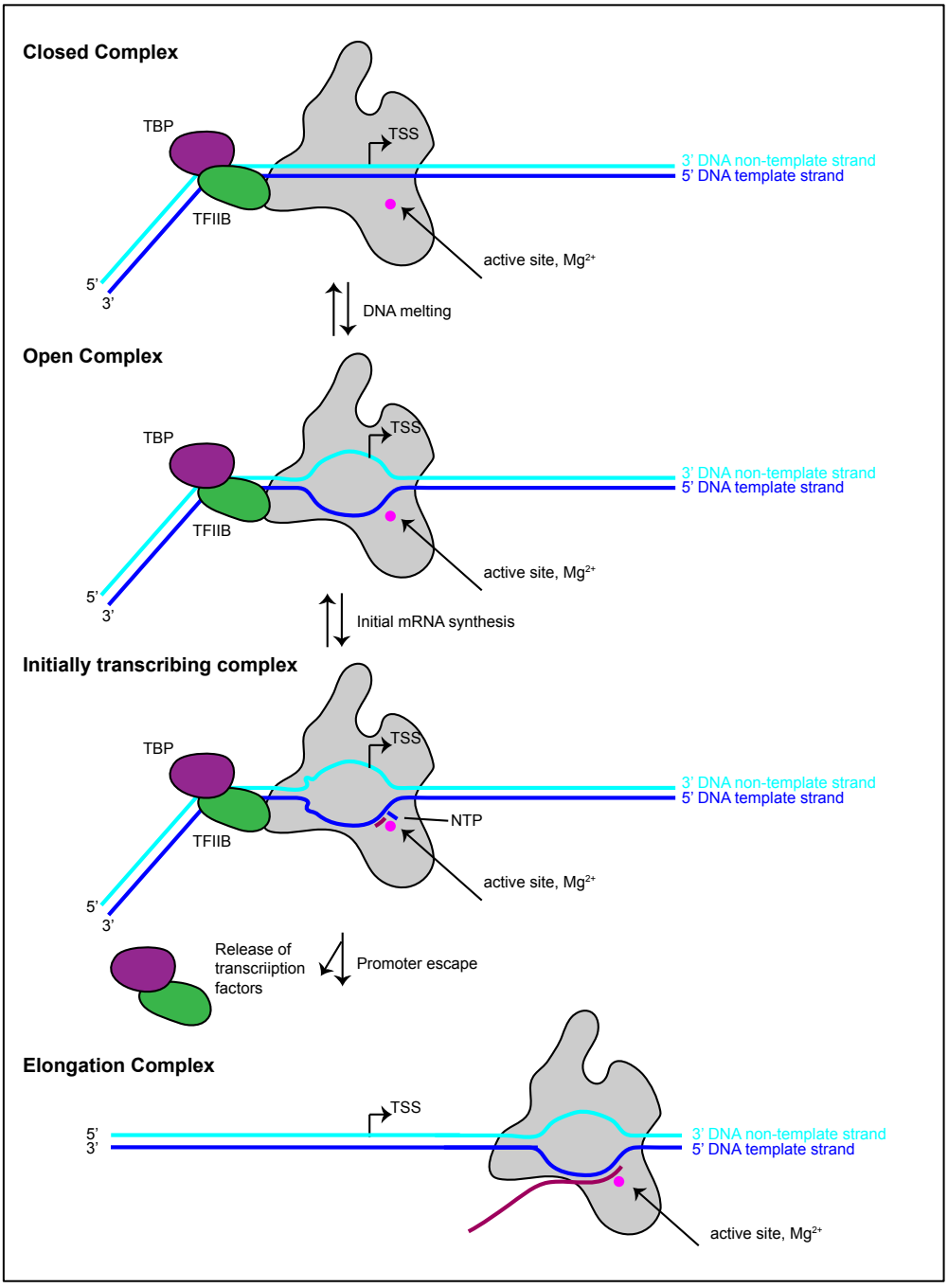


Figure 4: Schematic view of different functional states of the RNA Pol II.

Before DNA melting, the PIC, formed of GTFs and RNA Pol II, is bound to the double-stranded DNA as a closed complex. Upon melting of the DNA and positioning the DNA into the active centre of RNA Pol II, an open complex is formed. This complex transitions into an ITC by transcribing the first RNA nucleotides (shown in purple). After the first 8 nucleotides are transcribed, the GTFs are released and RNA Pol II escapes the promoter and the elongation complex is formed. Non-template strand is shown in light blue, template strand is shown in dark blue. Modified from (Cheung et al. 2011).

1.2.1. Bacterial RNA Polymerase

Studying transcription in prokaryotes can give valuable insights into the basic elements of eukaryotic transcription. From bacteria to humans, the essential core subunit composition $\alpha_2\beta\beta'\omega$ is conserved (Archambault & Friesen 1993). The core bacterial RNA Polymerase of *Thermus aquaticus* was solved by X-ray crystallography to a resolution of 3.3 Å (Zhang et al. 1999). This was the first high resolution X-ray crystal structure of any cellular RNA Pol (Jun et al. 2011). Prior to that, only low-resolution structures of bacterial and eukaryotic RNA Pols, solved by electron crystallography, were available (Darst et al. 1989). Zhang et al. revealed a "crab claw" molecular structure that contains an internal channel, bound by the Mg^{2+} co-ordinated active site at the back of the wall, with a diameter of 27 Å, large enough to accommodate double-stranded DNA molecule (Zhang et al. 1999). The Mg^{2+} is chelated by a motif that is conserved from bacterial to eukaryotic polymerases (Zhang et al. 1999) (Figure 3C). Here, the Mg^{2+} ions function as cofactors to incorporate deoxynucleotides (dNTP) during the polymerization by catalysing the phosphodiester bond formation.

Transcriptional initiation in prokaryotes requires the assembly of additional subunits, called σ, that bind to the core of the RNA Pol to form the holoenzyme (Murakami & Darst 2003). In this initial closed promoter complex, the σ factors bind upstream of the RNA Pol crab claw. The σ factors are able to recognize the Pribnow box (-10 element) and the -35 element, which directs factor spacing (Murakami et al. 2002; deHaseth et al. 1998). These DNA consensus sequences consist of six nucleotides and function as promoter recognition elements. The numbering -10 and -35 respectively corresponds to the upstream location with respect to the start site (+1). Although the binding of the σ factors to the RNA Pol is very stable (Gill et al. 1991), the σ factors are released upon entering into the

transcription elongation phase (Murakami & Darst 2003). The whole prokaryotic transcription cycle is reviewed in (Murakami & Darst 2003).

1.2.2. Archaeal RNA Polymerase

In Archaea, the transcription machinery including the general transcription factors transcription factor B (TFB), TATA binding protein (TBP), transcription factor E (TFE) and transcription factor S (TFS) can be regarded as a simpler version of the eukaryotic RNA Pol II system (Hirata & Murakami 2009). The archaeal transcription apparatus presents a mosaic of bacterial and eukaryotic characteristics, thus providing unprecedented insights into basic principles of transcription across all three domains of life (Hirata & Murakami 2009). In 2008, the X-ray crystal structure from *Sulfolobus solfataricus* (Hirata et al. 2008) provided the first atomic model of an archaeal RNA Pol (Jun et al. 2011) and in 2009, the structure of the 13 subunit RNA Pol from *Sulfolobus shibatae* was solved (Korkhin et al. 2009).

The archaeal RNA Pol requires TFB and TBP to initiate transcription (Korkhin et al. 2009). Additionally, the RNA Pol subunit Rpo13, which does not have an orthologue in eukaryotes, was identified and could function in a similar way to specific eukaryotic general transcription factors during transcription initiation (Korkhin et al. 2009).

1.2.3. Eukaryotic RNA Polymerases

In contrast to prokaryotes, the synthesis of RNA in eukaryotes is performed by three different RNA Pols which are all capable of unwinding the DNA double helix and polymerizing RNA (Cramer et al. 2000; Kassavetis & Geiduschek 2006; Arimbasseri & Maraia 2016). Each RNA

Pol is dedicated to transcribing a specific set of genes and has its own set of GTFs (Schramm & Hernandez 2002). Among other tasks, GTFs help the RNA Pols to recognize the gene promoter, which can be either located upstream or inside the target gene (Schramm & Hernandez 2002).

RNA Pol I transcribes large rRNA precursors, which make up to 60% of the nuclear transcripts transcribing from a single promoter type (Schramm & Hernandez 2002). Compared to RNA Pol II, RNA Pol I contains an additional subcomplex, A49/A34.5, whose subunits are related to the RNA Pol II GTFs TFIIF and TFIIE respectively (Geiger et al. 2010). This seems to suggest that the increased number of subunits relates to permanently recruited transcription factors to give an increase in the efficiency of transcription (Vannini & Cramer 2012b; Carter & Drouin 2010). In order to assemble an RNA Pol I PIC in yeast, Rrn3, the Core Factor (CF - consisting of Rrn6, Rrn11 and Rrn7) and the TFIIB-like factor, are required. It was shown that RNA Pol I exists as a dimer in its inactive form (Pilsl et al. 2016). However, upon binding of Rrn3, the monomeric state is favoured that can bind to the CF via Rrn3 (Blattner et al. 2011; Bedwell et al. 2012; Engel et al. 2013; Fernández-Tornero et al. 2013; Neyer et al. 2016; Tafur et al. 2016). Recently, three cryo-Electron Microscopy (cryo-EM) studies of the RNA Pol I PIC were published (Engel et al. 2017; Han et al. 2017; Sadian et al. 2017). These structures showed that a 30° bend is introduced by Rrn7 and Rrn11. Due to the low angle, the DNA runs between the wall and protrusion of the RNA Pol, forming a direct contact to the DNA. This close arrangement of the DNA with the PIC was suggested to from a “gate”, potentially facilitating rapid transcription initiation. Since the promoters are required to adopt to the distortion caused by the “gate”, Engel et al. suggested that recognition of the DNA is driven by the DNA’s biophysical properties rather than the DNA sequence, which is known to be poorly conserved in RNA Pol I promoters (Engel et al. 2017; Jochem et al. 2017).

In contrast to rRNA, the protein coding messenger RNA (mRNA) is synthesized by RNA Pol II, which is the best studied eukaryotic RNA Pol so far. In 2000, the first *S. cerevisiae* core RNA Pol II complex was solved by X-ray crystallography to a resolution of 2.8 Å (Cramer et al. 2000). RNA Pol II requires a specific set of general transcription factors: TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH to initiate transcription (Roeder 1996; Dvir et al. 1997; Kim et al. 1994).

TFIID contains TBP, which is universally required for all RNA Pols and several TBP associated factors (TAFs) (Sainsbury et al. 2015). Structures of TBP revealed that it is a saddle-shaped protein that, upon binding to the DNA in the minor grove, introduces a bend in the DNA of 90° (Chasman et al.,1993; Kim, Nikolov, & Burley, 1993a; Nikolov et al., 1992; Kim et al. 1993). This is in contrast to the 30° bend introduced by Rrn7 and Rrn11 in RNA Pol I (Engel et al. 2017; Han et al. 2017; Sadian et al. 2017). The binding mode of TBP to the DNA is highly conserved between eukaryotic RNA Pols I, II and III (Juo et al. 1996; Nikolov et al. 1996; Patikoglou et al. 1999).

The assembly of the PIC starts with the binding of TBP to the conserved TATA motif on the DNA (consensus sequence: TATAWAWR with W = T or A and R = A or G). This is located 30 bp upstream of the transcriptional start site (TSS) (Basehoar et al. 2004) and is conserved from yeast to humans (Cavallini et al. 1988). Most promoters do not contain a TATA Box. However, it was still possible to detect TBP on these promoters in yeast (Blair et al. 2012) and it was shown that TBP can bind to different variants of TATA DNA (Blair et al. 2012).

The binding of TBP to the DNA is stabilized by TFIIA, which is an auxiliary factor (Imbalzano et al. 1994). The formation of the TBP-DNA complex recruits TFIIB (Sainsbury et al. 2015). The transcription factor TFIIB contains four structurally conserved features: a N-terminal B-ribbon

region, a B-reader, B-linker and a C-terminal B-core region (Knutson & Hahn 2011; Vannini & Cramer 2012; Naidu et al. 2011). TFIIB binds to the DNA (Zhao & Herr 2002), stabilizes the TBP-DNA complex (Imbalzano et al. 1994) and is able to interact with the RNA Pol (Bushnell et al. 2004). It has structural and functional homologues in the RNA Pol I system (TAF1B) (Naidu et al. 2011) and in the RNA Pol III system (Brf1 and Brf2) (Colbert & Hahn 1992; Teichmann et al. 2000; Schramm & Hernandez 2002), which are all part of the family of TFIIB-like factors.

The core initiation complex is formed when pre-formed TFIIF-RNA Pol II binds to the TFIIB-TBP-TFIIA-DNA complex (Sainsbury et al. 2015). Therefore, TFIIB forms a bridge between RNA Pol II and TBP by binding with its C-terminal B-core region to TBP and the upstream DNA (Nikolov et al. 1995; Bagby et al. 1995) while its N-terminal B-ribbon binds to the RNA Pol II (Bushnell et al. 2004). Additional binding of TFIIE recruits TFIIH, which contains DNA kinase and helicase activities and complements the formation of a complete PIC in a closed state (Orphanides et al. 1996; Drapkin et al. 1994; Sainsbury et al. 2015).

RNA Pol III is the most complex of the eukaryotic RNA Pols. It contains 17 subunits that account for a molecular weight of approximately 0.7 MDa (Hoffmann et al. 2015) (Table 1). In addition to the conserved ten-subunit core and the two-subunit stalk, it contains an additional five subunits which are homologous to the GTFs observed in the Pol II system. Three of these additional subunits form the heterotrimer C82-C34-C31 (homologous to TFIIE) whereas the other two subunits form the C53-C37 dimer (homologous to TFIIF) (Kassavetis et al. 2010; Arimbasseri & Maraia 2015a).

RNA Pol III synthesizes short, highly structured RNAs which are not exceeding 400 bp (Schramm & Hernandez 2002) including the entire pool of tRNAs, 5S rRNA and U6 snRNA (Kassavetis & Geiduschek 2006;

Arimbasseri & Maraia 2016). Depending on the gene product that is transcribed by RNA Pol III, there are three different promoter types (Schramm & Hernandez 2002). Each of the promoter types requires a different combination of RNA Pol III GTFs to initiate transcription: TFIIIA (Engelke et al. 1980), the ubiquitously required TFIIIB, consisting of the TBP, B-related factor 1 (Brf1) and B double prime (Bdp1) (Harismendy et al. 2003; Moqtaderi & Struhl 2004; Barski et al. 2010), and TFIIIC (Lassar et al. 1983).

TFIIIB directly links the Pol III enzyme to the transcription factor assemblies at the promoter and is regarded as the minimal transcriptional machinery required to initiate Pol III transcription. This multi-subunit complex was initially described as two polypeptides with a mass of 70 kDa and 90 kDa via photo-crosslinking, which could then be separated into two distinct components via chromatography: B' and B'' (Bartholomew et al. 1991; Kassavetis et al. 1991). The subsequent transcriptional initiation step is very tightly regulated in eukaryotes (Lee et al. 2015; Fairley et al. 2012; Vannini et al. 2010) and it has been shown that a loss of regulation can lead to carcinogenesis (Gouge et al. 2017; Goodarzi et al. 2016; Dauwerse et al. 2011; Thiffault et al. 2015).

Recently, several efforts to understand the structural basis of Pol III transcription have been made. In 2015, Hoffmann et al. (Hoffmann et al. 2015) reported a 4.6 Å cryo-electron microscopy structure of the apo and a 3.9 Å structure of the transcribing *S. cerevisiae* RNA Pol III complex which revealed that the core is structurally conserved with respect to Pol I and Pol II (Hoffmann et al. 2015). This was followed in 2018 by RNA Pol III PIC structures giving unprecedented insights into the molecular mechanisms underlying transcriptional initiation in RNA Pol III (Abascal-Palacios et al. 2018; Vorländer et al. 2018). Hereafter only RNA Pol III will be discussed in further detail.

1.2.4. Transcription of RNA Polymerase III

1.2.4.1. Transcription initiation at promoter types 1, 2 and 3

RNA Pol III promoters can be divided into three different types. Type 1 and type 2 promoters are intragenic and generally do not encompass a TATA-box, while type 3 promoters are gene-external and are characterized by the presence of a TATA Box (Schramm & Hernandez 2002b) (Figure 5). An example of a type 1 promoter is the 5S promoter consisting of an intermediate element, an A Box and a C Box, which both lie within an internal control region (ICR). The A and C Box are both conserved elements on the DNA. To initiate transcription, the zinc finger protein TFIIIA binds to the ICR (Engelke et al. 1980; Sakonju et al. 1981). The formation of the TFIIIA-DNA complex serves as a specificity factor to direct TFIIIC to the 5S promoter (Lassar et al. 1983) which subsequently recruits TFIIIB and RNA Pol III (Schramm & Hernandez 2002).

The tRNA promoters can be considered as an example of type 2 promoters and contain both an A Box and a B Box. These regions encode the D and T loop of the tRNA molecule (Allison et al. 1983). *In vivo*, a faithful process of transcription initiation at type 2 promoters requires the recruitment of RNA Pol III by both TFIIIC and TFIIIB. Indeed, TFIIIC has been shown to be the first factor to bind to the promoter. When TFIIIC is bound, it serves as an assembly factor for TFIIIB, which is subsequently recruited to the region upstream of the transcription start site (TSS) (Marzouki et al. 1986; Kassavetis et al. 1990). After that, RNA Pol III is recruited and transcription can be initiated.

The type 3 promoter, only found in higher eukaryotes, contains a distal sequence element (DSE), a proximal sequence element (PSE) and a TATA Box outside of the gene (Kundu et al. 1999; Schramm & Hernandez 2002). Recognition of the PSE element is either done by the PSE-binding

protein (PBP), the PSE transcription factor (PTF) or the snRNA activating protein complex (SNAPc) (Schramm & Hernandez 2002b). The TATA Box is recognized by TBP, a component of TFIIIB. RNA Polymerase III is recruited after SNAPc, a complex that contains a TFIIB-like activity, has bound (Chong et al. 2001).

Despite the many different GTF combinations, Kassavetis et al. were able to confirm that TFIIIB was the minimal transcription factor of the RNA Polymerase III system in *S. cerevisiae* (Kassavetis et al. 1990). *In vitro*, they observed TFIIIB binding to a 5S or tRNA promoter without prior binding of TFIIIA or TFIIIC, respectively. Even after releasing TFIIIA and TFIIIC from the DNA templates, it was possible for Kassavetis et al. to initiate several rounds of RNA Pol III transcription with TFIIIB as the only transcription factor bound to the DNA (Kassavetis et al. 1990; Kassavetis et al. 1989).

Significant effort has been expended over several years to understand how TFIIIB is able to direct and regulate Pol III transcription. However, whilst this has allowed for progress in understanding Pol III transcription, the roles and mechanisms of the other GTFs remain elusive. In particular, the role and interactions of the initial factor, TFIIIC, remain unknown. Therefore, to further understand Pol III transcription, this study will focus on the function of TFIIIC in the initial steps of transcription initiation.

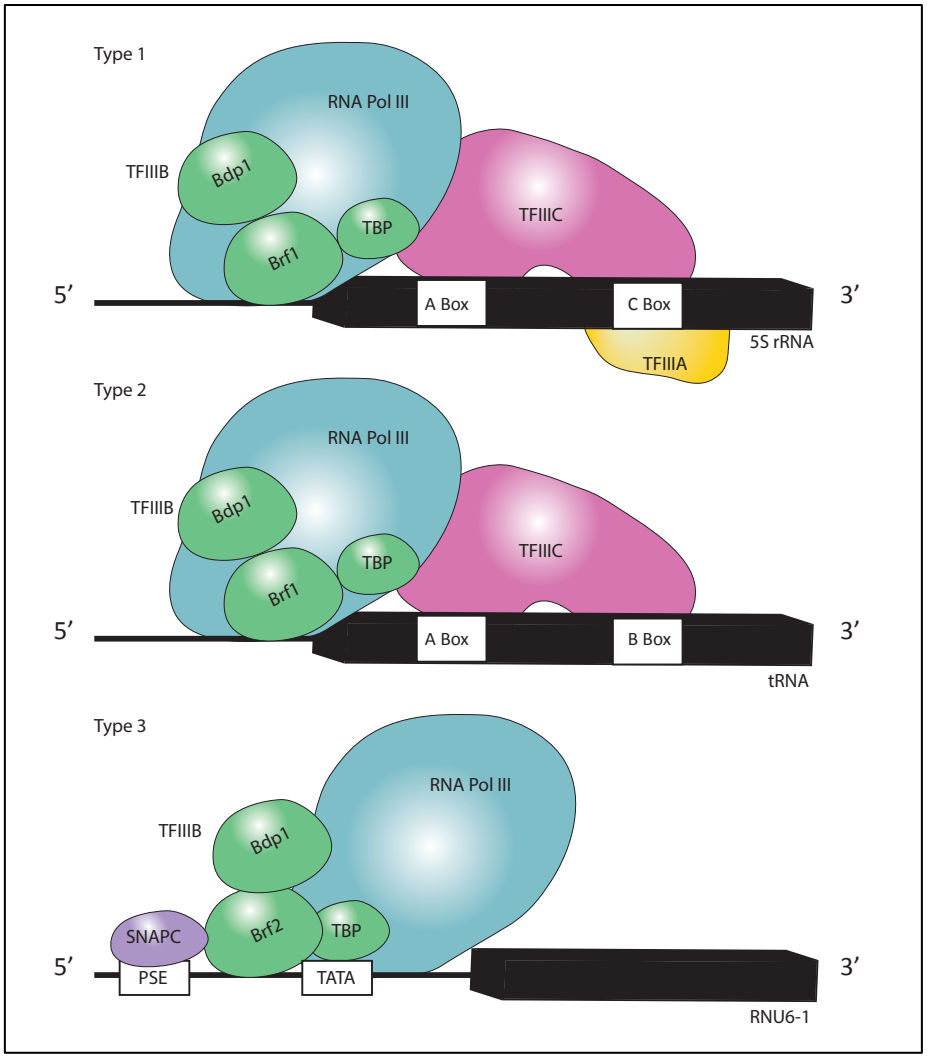


Figure 5: Schematic overview of RNA Pol III promoter types 1-3.

A 5S rRNA promoter is shown as an example of a type 1 promoter in which TFIIIA (yellow) binds to the C Box first, then TFIIIC (pink) is recruited and in a third step TFIIIB (green), consisting of Bdp1, Brf1 and TBP before RNA Pol III (blue) binds. In type 2 promoters, TFIIIC recognizes the A and the B Box and binds to the DNA first. Subsequently TFIIIB is recruited and after that RNA Pol III. In gene-external type 3 promoters, SNAPc binds and subsequently recruits TFIIIB and the RNA Pol III. Adapted from (Schramm & Hernandez 2002).

1.2.4.2. Transcriptional elongation and termination

Following initiation, Pol III must elongate the RNA transcript, a process that has yet to be studied in detail. In genome-wide studies, it was observed that RNA Pol III showed a strong signal at the A Box DNA region (Jordán-Pla et al. 2015; Turowski et al. 2016). This was interpreted as a slower clearance of RNA Pol III from the promoter site and a potential longer binding to the GTFs. Furthermore, RNA Pol III was shown to slow down transcriptional activity around the B Box DNA region (Turowski & Tollervey 2016). These non-uniform kinetics may be explained by the particular arrangement of GTFs in the intragenic promoter. Upon binding to the gene internal type 1 and type 2 promoters, TFIIIC stretches across the whole promoter until the terminator sequence (Nagarajavel et al. 2013), blocking the access to the DNA for the RNA Pol III and thus blocking polymerase passage at the A and B-boxes, the sites of TFIIIC binding.

Despite this, it is still not clear what happens to TFIIIC during transcription elongation and how it makes space for RNA Pol III transcription. One current hypothesis is that TFIIIC gets cleared away from the promoter entirely, which was supported by a lower occupancy of TFIIIC at RNA Pol III promoters than TFIIIB, (Roberts et al. 2003) (Harismendy et al. 2003). Alternatively, TFIIIC is suggested to be displaced partially. Indeed, a partial displacement of the TFIIIC subcomplex τB was hypothesized after a crosslinking mass spectrometry analysis could not detect interactions anymore between the two subcomplexes τA and τB, that constitute TFIIIC (Male et al. 2015).

In contrast to the previous two hypothesis, Ruet et al. suggested that TFIIIC remains stably bound during transcription (Ruet et al. 1984). This could be explained by τA being displaced from the A Box first whilst τB remains bound to the B Box. As the transcription continues, τA binds

again to the A Box and τB can be displaced from the B Box to continue transcription until the terminator (Turowski & Tollervey 2016). This hypothesis would support a non-simultaneous displacement of the two subcomplexes of TFIIIC, ensuring that the complex as a whole remains bound to the DNA.

Following elongation, a stretch of four to five Thymine residues in the non-template strand was shown to be sufficient for termination of the RNA Pol III transcription by genomic analysis (Braglia et al. 2005; Iben & Maraia 2012; Arimbasseri et al. 2013). *In vivo*, 7-8 Thymines are required (Turowski et al. 2016). It is thought that this termination sequence acts as a destabilizing signal due to the weak base pairing between Adenine and Thymine (Arimbasseri & Maraia 2015). Interestingly, RNA Pol III transcripts were found that extended more than 50 nucleotides beyond the canonical terminator signal of the tRNA promoter (Orioli 2016) (Turowski et al. 2016). These readthrough transcripts are directed to the exosome complex for degradation (Turowski et al. 2016; Carlsten et al. 2016).

In yeast, the heterodimer C37-C53 was shown to be specifically required for recognition of the terminator sequence together with C11, which is also involved in RNA cleavage (Landrieux et al. 2006; Chédin et al. 1998). Upon recognition of the terminator sequence, the subunits C37, C53 and C11 are able to mediate the transition of the RNA Pol III elongation complex into a metastable pre-termination complex whereby the RNA transcript is released (Arimbasseri & Maraia 2015).

RNA Pol III was shown to frequently reinitiate transcription due to recycling of the pre-assembled tDNA-TFIIIC-TFIIIB complex. This facilitated recycling model has been suggested to explain the high transcriptional rates achieved by Pol III, following an initial transcription cycle. All subsequent transcripts could be produced between 5-10 fold

faster until polymerase disengagement. Indeed, it could be demonstrated that RNA Pol III is committed to reinitiate transcription on the same gene and requires canonical termination (Dieci & Sentenac 1996).

1.2.4.3. RNA Polymerase III and cancer

RNA Pol II and III can be found in the nucleoplasm and are concentrated in discrete sites (Pombo et al. 1999). Pombo et al. could locate approx. 2000 of these sites (Pombo et al. 1999). In healthy cells, the ratio of RNA Pol I and RNA Pol III transcription is tightly regulated according to the cell's requirements. Upon transformation, these regulatory mechanisms are evaded and the cells synthesize rRNA and tRNA at a much higher rate (White 2004). Nuclear hypertrophy has long been used as a cytological feature to identify cancer cells. Indeed, it was shown later that the size of the nucleolus is proportional to the level of rRNA synthesis (Kurata et al. 1978; Altmann & Leblond 1982; Moss & Stefanovsky 1995).

In 1974, hyperactivation of RNA Pol I and RNA Pol III was shown in mice with myeloma (Schwartz et al. 1974). *In vitro*, it was possible to induce overexpression of RNA Pol I and RNA Pol III products by induction with DNA tumour viruses, RNA tumour viruses and chemical carcinogens, (Liebhaber et al. 1978; Scott et al. 1983; White et al. 1990; Wang et al. 1995; Wang et al. 1997; Gottesfeld et al. 1996; Ying et al. 1996; Larminie et al. 1999; Zhai & Comai 1999; Felton-Edkins et al. 2003) which was later confirmed *in situ* in human ovarian cancers (Winter et al. 2000).
Additionally, *in situ* hybridization experiments of breast, lung and tongue carcinomas showed an increased amount of Pol III products compared to non-transformed cells (Chen et al. 1997).

In addition to finding increased amounts of Pol III products in cancer, Haurie et al. showed that an isoform of the RNA Pol III subunit RPC7β,

called RPC7α, can transform cells and alter gene expression to favour cancer development (Haurie et al. 2010).

There are three explanations as to why RNA Pol III products are elevated in transformed cells; overexpression of RNA Pol III transcription factors, loss of repressors or a direct activation through oncogenes (White 2004). Indeed, it was shown that adenoviral SV40 transformation can overexpress five of the subunits of TFIIIC at both mRNA and protein level (Felton-Edkins et al. 2003; Larminie et al. 1999) and can increase its activity (Hoeffler & Roeder 1985; Yoshinaga et al. 1986; Hoeffler et al. 1988; White et al. 1990). This finding was confirmed in nine human ovarian carcinomas (Winter et al. 2000). After binding of TFIIIC to the promoter region, TFIIIB is recruited as described in paragraph 1.2.4.1. All components of TFIIIB have been shown to be overexpressed in transformed cells. Elevated levels of TBP were shown to contribute to cellular transformation in human colon carcinoma (Johnson et al. 2003), increased levels of Bdp1 were detected in SV40 transformed cells (Felton-Edkins et al. 2003) and increased levels of Brf1 were shown to correlate with HPV16 infected cells (Daly et al. 2005). Furthermore, Brf2, a member of the TFIIB-like core transcription factors, was shown to function as a redox-sensing module. As such, it was shown to be involved in the oxidative stress pathway in both lung and breast cancer (Gouge et al. 2017). In healthy cells, selenoproteins and seCys-tRNA are reduced during oxidative stress in a Brf2-dependent manner, leading to an increase in apoptosis. To overcome this, it was demonstrated that cancer cells are able to overexpress Brf2 to maintain redox homeostasis during oxidative stress (Gouge et al. 2017).

1.3. Transcription Factor TFIIIC

1.3.1. Domain architecture

Yeast TFIIIC is a large multisubunit complex that can be divided into two subcomplexes: τA, consisting of τ131, τ95 and τ55 and τB, consisting of τ138, τ91 and τ60 (Schramm & Hernandez 2002; Marzouki et al. 1986; Schultz et al. 1989; Ducrot et al. 2006). τA can bind via its DNA binding domain to the A Box, whereas τB binds to the B Box (Dumay-Odelot et al. 2007; Lassar et al. 1983). The consensus sequence of the A Box was defined as TRGYnnAnnnG and for the B Box , GWTCRAnnC with R = purine, Y = pyrimidine, n = nucleotide, W = weak binding (A or T) (Marck et al. 2006). The position of the A and the B Box is fixed relative to the 5' and 3' end of the tRNA gene but the spacing between the A and B Box can vary. The presence of an intron at position 37-38 of the tRNA gene, 1nt downstream of the anticodon, can for instance add between 7 and 71 nucleotides to the usual spacing of 34 nucleotides. Additionally, lengths of the tRNA flexible arm at position 47-48, can vary between 31 and 93 nucleotides creating a difference which can equal an increase in distance of 220 Å (Marck et al. 1993). Together, these insertions can significantly alter the A and B Box spacing (Nagarajavel et al. 2013; Marck et al. 2006). A flexible tether between τA and τB is thought to be able to adapt TFIIIC to differing A and B Box spacings (Kassavetis & Geiduschek 2006). Following binding site recognition, the DNA binding of τB to the B Box contributes mostly to the affinity of the TFIIIC-DNA interaction as it has been shown to bind with higher affinity than τA to the A Box (Stillman & Geiduschek 1984). Thus, the binding of τB to the B Box serves as an anchoring platform that favours the binding of τA to the A Box and the final interaction with RNA Pol III (Schultz et al. 1989). The overall architecture and composition of the different subunits seems to be conserved between human and yeast TFIIIC (Dumay-Odelot et al. 2010).

Previously, there have been multiple efforts to understand the structure and arrangement of the TFIIIC complex. In 1989, yeast TFIIIC was visualized for the first time via scanning transmission electron microscopy at low resolution. It was described as a “dumb-bell shaped” protein complex when bound to $tRNA_3^{Leu}$ genes containing A-B Box distances of 53 or 74 bp (Schultz et al. 1989). Smaller distances between the A and B Box however did not enable a separation of the two subcomplexes (Schultz et al. 1989). More recently, crosslinking mass spectrometry experiments revealed a potential subunit architecture of yeast holo-TFIIIC (Male et al. 2015). In addition, several crystal structures of domains of *S. cerevisiae* and *S. pombe* TFIIIC were solved (Male et al. 2015; Mylona et al. 2006; Taylor et al. 2013a) (Figure 6). However, despite the previous efforts, the overall structure of yeast or human TFIIIC complex remains elusive.

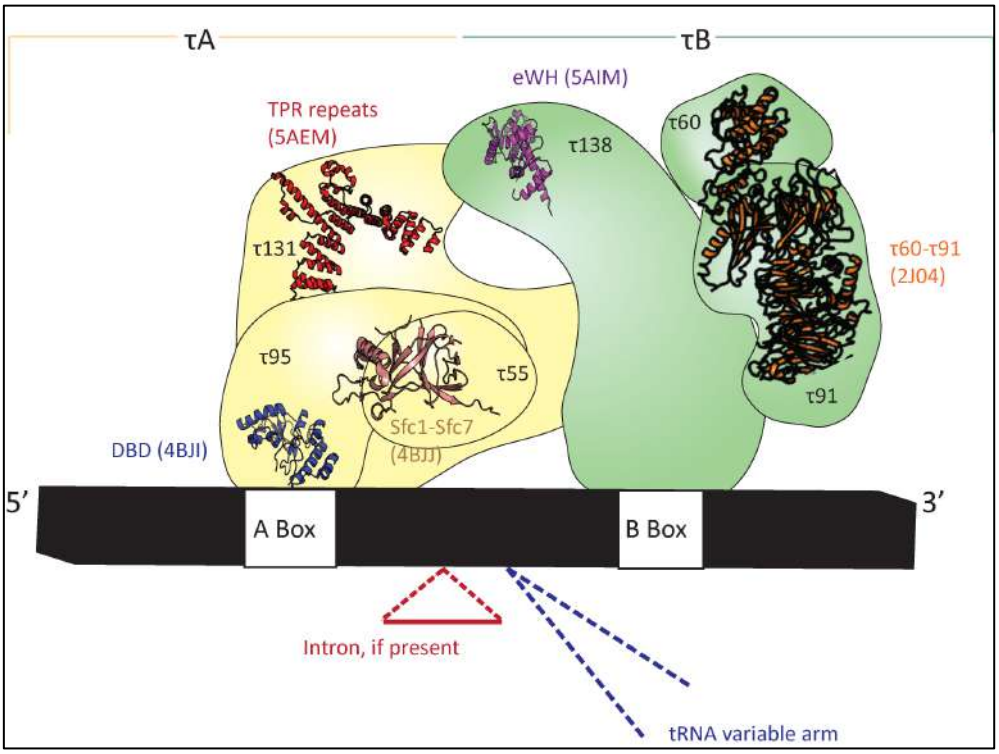


Figure 6: Schematic subunit organization of the yeast TFIIIC complex bound to the promoter type 2 DNA.

TFIIIC consists of six subunits, divided into two subcomplexes, τA (τ55, τ95, τ131) shown in yellow and τB τ138, τ91, τ60) shown in green. The distance between the A and the B Box varies depending on the presence of Introns or the tRNA variable arms. The domain architecture is based on crosslinking mass spectrometry performed by (Male et al. 2015). Crystal structures 5AIM, 5AEM (Male et al. 2015), 4BJI, 4BJJ (Taylor et al. 2013a) and 2J04 (Mylona et al. 2006) were fitted approximately.

1.3.2. Conservation of TFIIIC

TFIIIC is conserved between *S. cerevisiae*, *S. pombe* and *H. sapiens.* Functional or sequence homologues between yeast and human were found for all six subunits (L'Etoile et al. 1994; Lagna et al. 1994; Sinn et al. 1995; Hsieh et al. 1999a; Hsieh et al. 1999b; Dumay-Odelot et al. 2007) (Table 2).

The three subunits of τA contain higher sequence conservation than τB. For the latter, conservation was first found by cloning orthologues in *S. pombe* (Huang & Maraia 2001; Huang et al. 2000). τ131 possesses the highest sequence conservation, followed by τ95. The human orthologue of τ55, TFIIIC35, was identified later and shows only limited sequence conservation. Furthermore, it is missing the previously reported histidine phosphatase domain (Dumay-Odelot et al. 2007).
Upon identifying TFIIIC35, it was shown that it forms a heterodimer with TFIIIC63 in accordance with what was previously reported for the Sfc1-Sfc7 *S.pombe* heterodimer (Dumay-Odelot et al. 2007; Taylor et al. 2013a).

In τB, τ138 possesses the least conservation between yeast and human. The sequences between yeast τ138 and the human TFIIIC220 are poorly conserved, however, human TFIIIC also contains a DNA binding domain, similar to the one found in τ138 (Matsutani 2004; Shen et al. 1996). The human orthologues of τ60 and τ91, TFIIIC90 and TFIIIC110 were shown

to contain histone acetyltransferase activity (HAT) activity, which has not yet been shown for yeast subunits yet (Kundu et al. 1999).

Table 2: TFIIIC subunits in S. cerevisiae, S. pombe and H. sapiens.

Subunits marked with an asterisk were expressed with a tag for this thesis, thus increasing the molecular weight which will be discussed in the results chapter. Adapted from (Dumay-Odelot et al. 2007).

S. cerevisiae		*S. pombe*		*H. sapiens*	
subunit	molecular weight	subunit	molecular weight	subunit	molecular weight
τ138 (Tfc3)	132.1 kDa	Sfc3	153.6 kDa	TFIIIC220	238.8 kDa*
τ131 (Tfc4)	120.2 kDa	Sfc4	116.4 kDa	TFIIIC102	101.2 kDa
τ91 (Tfc6)	74.7 kDa*	Sfc6	66.2 kDa	TFIIIC110	100.6 kDa
τ95 (Tfc1)	73.6 kDa	Sfc1	52.7 kDa	TFIIIC63	59.6 kDa
τ60 (Tfc8)	67.7 kDa	Sfc9	74.8 kDa	TFIIIC90	91.9 kDa
τ55 (Tfc7)	49.1 kDa*	Sfc7	13.3 kDa	TFIIIC35	24.0 kDa*
Total	517.4 kDa		477.0 kDa		616.1 kDa

1.3.3. τ138 and TFIIIC220

τ138 (human orthologue: TFIIIC220) is part of the τB subcomplex and the largest subunit of the TFIIIC complex. It is currently the least characterized subunit of TFIIIC and has a molecular weight of 132 kDa (1,160 residues) in yeast and 238 kDa (2,109 residues) in humans (Figure 7).

Male et. al predicted the presence of several DNA binding domains, including a high mobility group (HMG)-box domain at the N-terminus of τ138 and five wing helix domains of which three are located at the N-terminus; with the others distributed across the central and C-terminal regions of the protein (Male et al. 2015; Harami et al. 2013) (Figure 7). The rest of the protein was predicted to be unstructured. Despite this, the central extended wing helix domain, that comprises residues 546-641, was crystallized (Male et al. 2015) and thus makes up only 8% of the total structure. Furthermore, Male et al. were able to identify an unstructured region from residue 641 to residue 693 that is necessary and sufficient to interact with the TPR array of τ131 (Male et al. 2015).

Photo-crosslinking experiments revealed that τ138 is able to recognize and bind to the B Box motif of the DNA through residues 108-174 (Bartholomew et al. 1990; Gabrielsen et al. 1989). Indeed, the τB - B Box DNA binding contributes mostly to the binding affinity of TFIIIC to the DNA following characterization of a weak τA -A Box DNA interaction (Marzouki et al. 1986). These findings were confirmed by a series of experiments in which either the B Box or the DNA binding site in τ138 was mutated which lead to a significant reduction in the TFIIIC-promoter affinity (Lefebvre et al. 1994; Baker et al. 1986). τ91, the second largest subunit of the τB subcomplex, was shown to cooperate with τ138 in binding to the B Box (Arrebola et al. 1998).

Furthermore, HAT activity was initially found in all subunits of human τB and was later restricted to TFIIIC220 (Kundu et al. 1999; Basu et al. 2019). It is not reported yet whether the yeast τ138 subunit contains HAT activity.

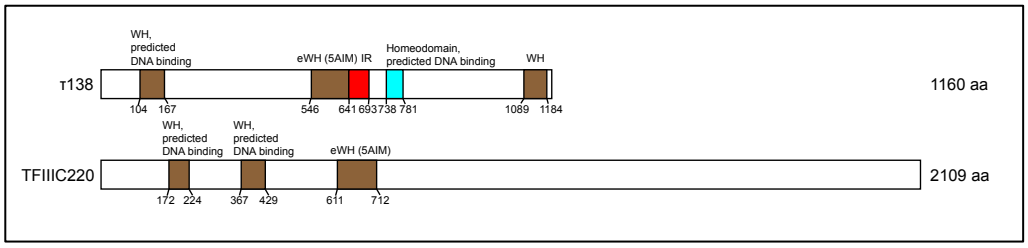


Figure 7: Domain organization of τ138 and human TFIIIC220.

τ138 is predicted to have a wing helix (WH) domain at position aa. 104-167, a homeodomain at position aa. 738-781 and a WH domain at aa. 1089-1148 according to Phyre2 result. Male et al. (Male et al. 2015) solved the structure of an eWH domain (aa.546-641) and defined an interacting region (IR) in τ138 responsible for interacting with τ131. TFIIIC220 is predicted to have a WH domains at positions aa. 172-224 and aa. 367-429. The eWH structure of 5AEM aligns with positions aa. 611-712 according to Phyre2.

1.3.4. τ131 and TFIIIC102

With 1,025 amino acids, τ131 (human orthologue: TFIIIC102) is the largest subunit of the τA subcomplex of TFIIIC and shows the highest sequence conservation (Dumay-Odelot et al. 2002). The domain architecture of τ131 according to the predictions from Marck et al. is shown in grey in Figure 8 (Marck et al. 1993). The authors observed two acidic regions at positions aa. 9-97 and aa. 310-334 as well as two basic regions at positions aa. 102-122 and aa. 599-637 which are characterized by a stretch of five consecutive basic amino acids. It is hypothesized that this region could be part of a nuclear localization signal (Marck et al. 1993). Furthermore, the presence of 11 tetratricopeptide repeat (TPR) domains was predicted. TPRs can serve as a binding interface for proteins and each domain consists of 34 amino acids that form two antiparallel alpha helixes. These repeats can be arranged in multiple copies in order to form a TPR array (Zeytuni & Zarivach 2012). According to predictions using the conserved TPR sequence from Sikorski et al. (Sikorski et al. 1991) in τ131, the TPR repeats are divided into two blocks separated by a basic helix loop helix motif (aa. 613-700). The first block contains TPR repeats 1-5 (aa. 128-297), the second one TPR repeats 6-9 (aa. 432-569) and the isolated TPR repeats 10 (aa. 875-908) and 11 (aa. 959-992) (Marck et al. 1993) (light grey, Figure 8).

Limited proteolysis revealed a stable fragment of τ131 (aa. 123-566) that was crystallized in two space groups, $P6_2$ and $P4_3$, which diffracted to 3.15Å and 3.4 Å, respectively (Male et al. 2015). The authors were able to observe 10 TPR repeats, which form two arms separated by a central region (5AEM) and identified a previously unpredicted TPR domain (TPR repeat 10) in the region where only 9 TPR repeats were predicted (Figure 8).

The amino terminal (N-terminal) TPR array of τ131 was shown to be involved in the stepwise recruitment of TFIIIB to the promoter region (Moir et al. 1997; Moir et al. 2000; Moir et al. 2002; Dumay-Odelot et al. 2002; Liao et al. 2003; Moir et al. 2004; Liao et al. 2006;). Indeed, TPR repeats 1-5 were reported to be involved in Brf1 binding (Moir et al. 2002). Upon binding to Brf1, it was hypothesized that τ131 switches from an autoinhibited form into a non-inhibited form, similar to the peroxisomal importer PEX5. In the autoinhibited form, an intramolecular interaction between the N terminus of τ131 and TPR6-9 could be formed, which in turn would be prevented via the binding of Brf1 (Liao et al. 2006). In earlier studies, it was shown that mutations in TPR2 facilitate the binding to Brf1. It was thus hypothesized that TFIIIC can adopt two different conformations. In one inhibited conformation, the TFIIIB binding site is masked and in another, the TFIIIB binding site is exposed. This conformational change was thought to be TPR mediated (Moir et al. 1997).

Besides the role of τ131 in binding to TFIIIB, it was reported that the interaction of τ131 to τ138 connects τA and τB (Rozenfeld & Thuriaux 2001). More specifically, it was confirmed later, that the TPR repeats of τ131 play a central role in linking τA and τB (Male et al. 2015). Inside τA, τ131 was also shown to interact with τ55 in a crosslinking mass spectrometry experiment (Male et al. 2015).

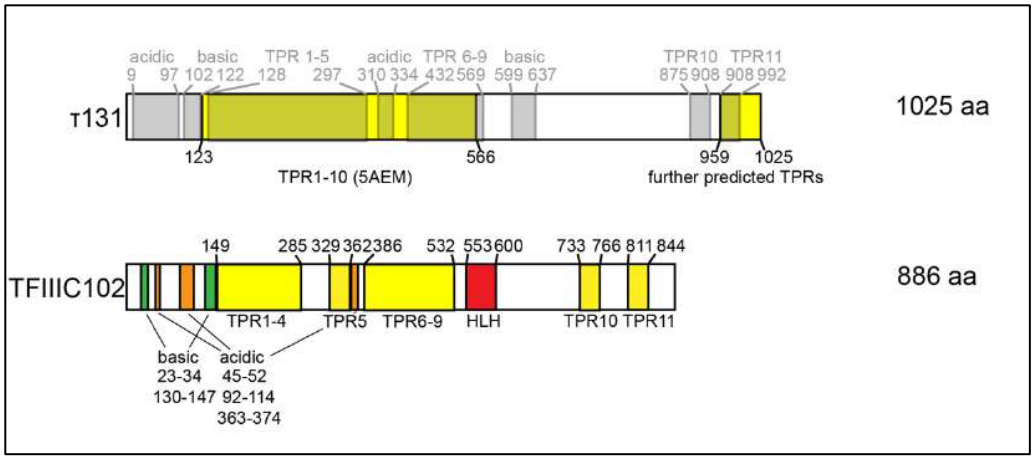


Figure 8 : Domain organization of τ131 and human TFIIIC102.

τ131 contains several TPR repeats of which TPR1-10 from position aa. 123-566 were solved (5AEM) by (Male et al. 2015). Further TPR repeats are predicted in the C-terminus of τ131 (aa. 959-1025) as derived from Uniprot. The predicted domains from Marck et al. are shown in transparent grey: acidic regions (aa. 9-97, 310-334), basic regions (aa. 102-122, 599-637), TPR repeats 1-5 (aa. 128-297), TRP 6-9 (aa. 432-569), TPR10 (aa. 875-908), TPR 11 (aa. 959-992) (Marck et al. 1993). TFIIIC102 contains two basic regions (23-34, 130-147), three acidic regions (aa. 45-52, aa. 92-114, aa. 363-374), a Helix loop Helix (HLH) motif (aa. 553-600) and 11 predicted TPR repeats: TPR 1-4 (aa. 149-285), TPR5 (aa. 329-362), TPR6-9 (aa. 386-532), TPR10 (aa. 733-766), TPR11 (aa. 811-844) (Hsieh et al. 1999a).

1.3.5. τ95 (TFIIIC63) and τ55 (TFIIIC35)

τ95 and τ55 are part of the τA subcomplex. τ95 (human orthologue: TFIIIC63) was shown to crosslink in close proximity to the A Box (Bartholomew et al. 1990) and to be specifically crosslinked to a tRNA gene (Gabrielsen et al. 1989). It was also found to be the only subunit in τA containing DNA binding properties (Ducrot et al. 2006) and able to bind both single and double-stranded DNA via an unknown mechanism (Taylor et al. 2013b).

Furthermore, τ95 was found to form a stable subcomplex with τ55 (human orthologue: TFIIIC35), a chimeric protein derived from chromosomal rearrangements (Manaud et al. 1998). The τ55 C-terminus is sufficient to ensure TFIIIC function in DNA binding assays and was shown to be

essential for cell viability (Manaud et al. 1998). The non-essential N terminus of τ55 contains a histidine phosphatase domain that is thought to be a potential link between TFIIIC and metabolic pathways via phosphorylation of an N-terminal peptide of Ugp1 (Manaud et al. 1998; Taylor, et al. 2013b). The τ55 histidine phosphatase domain (aa. 1-272) was crystallized and solved at 1.5 Å resolution (Taylor et al. 2013b). *S. cerevisiae* subunits τ95 and τ55 are homologues of *S. pombe* subunits Sfc1 and Sfc7, respectively. Sfc1 consists of a dimerization domain (aa. 1-110), a linker region (aa. 110-185), a DNA binding domain (aa. 186-396) and a tail (aa. 397-456) (Figure 9). Sfc7 has a total length of 120 residues and consists of a dimerization domain that extends from residue 1-101. The Sfc1-Sfc7 dimerization module was crystallized and the structure was solved at a resolution of 2.4 Å (Taylor et al. 2013a). The structure showed overall similarity to the N terminus of RNA Pol II specific transcription factor TFIIF, which contains the Rap30/Rap74 dimerization module (Taylor et al. 2013a). The C-terminal DNA-binding domain of Sfc1 was crystallized separately at a resolution of 1.45 Å (Taylor et al. 2013a). It also shows similarity to the C-terminal winged helix domain of TFIIF Rap30, thus indicating that TFIIIC has evolutionally incorporated a copy of an RNA Pol II specific transcription factor.

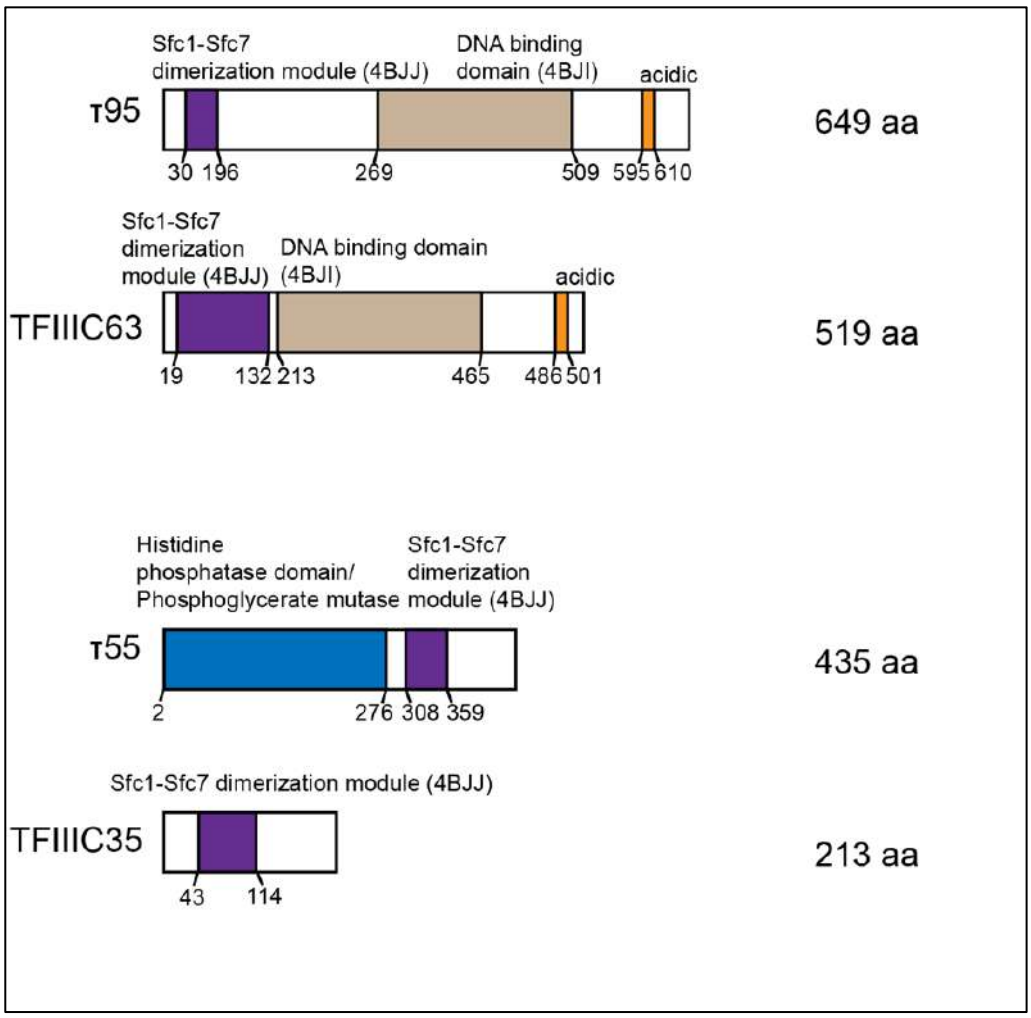


Figure 9: Domain organization of τ55 and τ95 and human TFIIIC35 and TFIIIC63.

τ95 aligns with the Sfc1-Sfc7 dimerization module (4BJJ) in position 30-196, with the DNA-binding domain (4BJI) in positions 269-509 (Phyre2) and it contains a highly acidic region (aa. 595-610) (Hsieh et al. 1999b). Human TFIIIC63 align with 4BJJ in positions 19-132 and with 4BJI in positions 213-465 (Phyre2) and an acidic region is in position 486-501 (Hsieh et al. 1999a). τ55 is predicted to have a histidine phosphatase domain (HPD) /phosphoglycerate mutase (aa. 2-276) at the N terminus and aligns to 4BJJ in the C-terminus (Phyre2). Human TFIIIC35 lacks the HPD but aligns with 4BJJ (aa. 43-114) (Phyre2).

1.3.6. τ60 (TFIIIC90) and τ91 (TFIIIC110)

τ60 and τ91(human orthologues: TFIIIC90 and TFIIIC110) are part of the τB subcomplex and were shown to both contain WD repeats (Huang et

al., 2000; Mylona et al. 2006), which are short structural motifs with a variety of functions (Figure 10). In a crosslinking mass spectrometry experiment, it was shown that τ60 only interacts with τ91 (Male et al. 2015). Previously, a 3.2 Å crystal structure of τ60 bound to a large C-terminal part of τ91 was solved (Mylona et al. 2006). The C-terminal fragment of τ91 contains a seven-bladed propeller. τ60 consists of a homologous propeller and a C-terminal domain with an alpha/beta fold. The propeller domains of each subunit mediate the formation of the heterodimer that in turn serve as a binding scaffold for τ138. Furthermore, it was shown that the C-terminal domain of τ60 interacts with TBP and thereby seems to contribute to the recruitment of TFIIIB to the promoter (Mylona et al. 2006).

HAT activity has been reported for the human orthologues of τ60 (TFIIIC90) and τ91 (TFIIIC110) but has not been reported for the yeast orthologues yet (Hsieh et al. 1999b; Kundu et al. 1999).

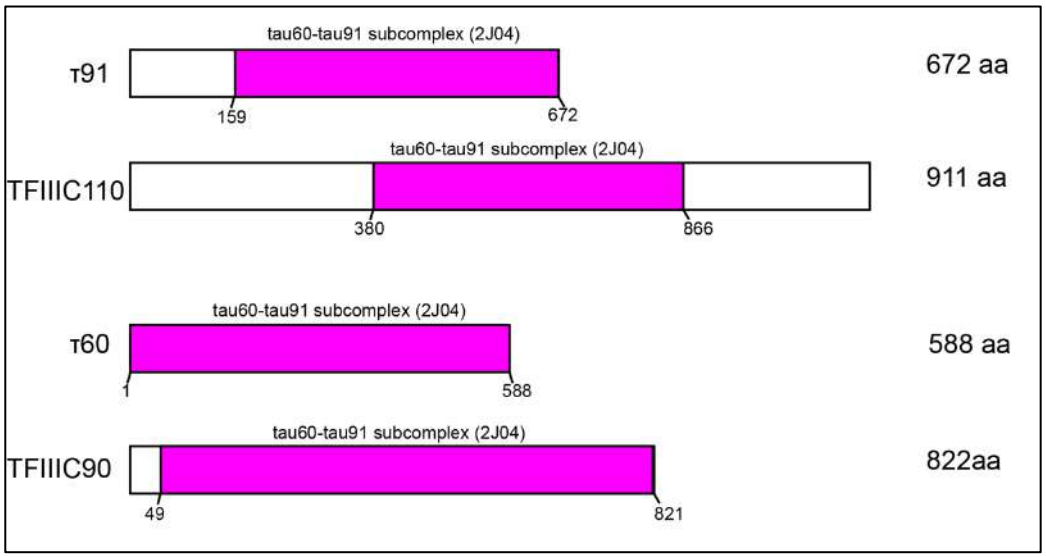


Figure 10: Domain organization of τ91 and τ60 and human TFIIIC110 and TFIIIC90.

τ91 was crystalized from 159-672 in a dimer with τ60 (1-588) (2J04). Human TFIIIC110 aligns with 2J04 in positions 380-866 and human TFIIIC90 aligns with 2J04 in positions 49-821 (Phyre2).

1.3.7. TFIIIC interaction with TFIIIB and RNA Polymerase III at type 2 promoters

As described in paragraph 1.2.4.1, TFIIIC is the first transcription factor to bind to type 2 promoters and subsequently recruits TFIIIB to the promoter, which then in turn recruits RNA Pol III in order form a full PIC. The exact molecular mechanism mediating the PIC assembly is not yet fully understood. However, a hypothesis about the different steps involved in the formation of the PIC was suggested.

Kassavetis et al. suggest that the interaction of τ131 with Brf1 involves several TPR repeats that interact with two binding sites of the N-terminus of Brf1 and one additional site on the C-terminus of Brf1 (Kassavetis & Geiduschek 2006). Both, the internal TPR repeats 6-9, as well as the N-terminal part of Brf1 vies for an interaction with the TPR binding site, located at TPRs 1-5.

The recruitment of TFIIIB to the promoter was suggested to happen in a stepwise manner (Kassavetis & Geiduschek 2006). In the first, rate-limiting step, Brf1 binds to the N-terminal arm of τ131, the TPR9 repeat (Liao et al. 2006; Moir & Willis 2004). Kassavetis and Geiduschek postulated that this subsequently leads to a series of rearrangements of the N-terminal arm (Kassavetis & Geiduschek 2006). TBP anchors Brf1 to the DNA, thereby causing an extension of τ131 towards the upstream positioned TBP-Brf1 complex. Binding of τ131 to Bdp1 most likely replaces the interaction of τ131 with Brf1 and leads to another conformational change (Liao et al. 2003; Liao et al. 2006;). The C-terminal part of Brf1 was also shown to bind TBP and Bdp1, besides binding to the TPR repeats. Most likely, not all of these binding events can take place at once at the C-terminal region of Brf1, supporting the hypothesis of a stepwise process (Liao et al., 2003; Kassavetis & Geiduschek, 2006).

Male et al. concluded from their experiments that Brf1 interacts with the N-terminal TPR repeats of τ131 first, before TBP is recruited via interactions with Brf1 and τ60 (Male et al. 2015). As a final component of TFIIIB, Bdp1 is recruited and competes for an interaction site on the TPR repeats of τ131, which are interacting with τ138. Upon binding, the τ131-τ138 interaction gets lost and τB gets displaced from the promoter region whereby TFIIIC loses most of its DNA-binding affinity. This hypothesis is in accordance with previous studies that conclude that TFIIIC is only required for the assembly of TFIIIB, but not for transcription with RNA Pol III (Kassavetis et al. 1990) and that TFIIIC is not bound to the DNA during RNA Pol III transcription (Bardeleben et al. 1994).

Several direct interactions of TFIIIC subunits τ131 as well as the human subunits TFIIIC90 and TFIIIC63 with the RNA Pol III have been reported, which could hint to the fact that TFIIIC plays an active role in recruiting RNA Pol III. τ131 was shown to bind to C53 and ABC10alpha in a two-hybrid assay (Flores et al. 1999). In the human system, interactions of TFIIIC90 were reported with HsRPC62 and HsRPC39, two subunits of RNA Pol III required for transcriptional initiation. Furthermore, TFIIIC63 was shown to interact with HsRPC62 (Hsieh et al. 1999a; Schramm & Hernandez 2002).

1.3.8. Additional roles of TFIIIC in the organization of the genome

Besides its role in transcription initiation for the RNA Pol III machinery, TFIIIC was found to bind to “extra TFIIIC sites” (ETC) throughout the genome at which neither TFIIIB nor the RNA Pol III are recruited in budding yeast. ETC sites are characterized by the presence of a B Box but lack an A Box (Moqtaderi & Struhl 2004). In *S. cerevisiae*, TFIIIC was confirmed to bind to six of the nine predicted ETC sites, as well as a newly identified tenth ETC site (Nagarajavel et al. 2013). However, Brf1 could

be located at two ETC sites, ETC9 and ETC5 in *S. cerevisiae* (Nagarajavel et al. 2013; Guffanti et al. 2006; Valenzuela et al. 2009).
Most ETC sites are located at the nuclear periphery and TFIIIC was shown to play a central role to the peripheral location since depletions in τ138 abrogated the localization in the periphery (Hiraga et al. 2012). Furthermore, TFIIIC was found to bind to centromeric nucleosomes (CEN), which could be mediated via condensin (Nagarajavel et al. 2013). In budding yeast, condensin was previously shown to be present at TFIIIC binding sites (D'Ambrosio et al. 2008; Haeusler et al. 2008). In another study, TFIIIC was found to play a role in boundary function and chromatin organization. Chromatin boundaries usually serve the function to prevent the spreading of heterochromatin, the tightly condensed form of chromatin (West et al. 2002). Chromatin boundary elements contain inverted repeats with several B Boxes. They are also located in the nuclear periphery and reported to bind TFIIIC (Noma et al. 2006). It is thought that these chromosome- organizing-clamp (COC) sites are responsible for the three dimensional genome organization in budding yeast (Noma et al. 2006) of which the mechanism is not yet known (Figure 11) (Ramsay & Vannini 2017).

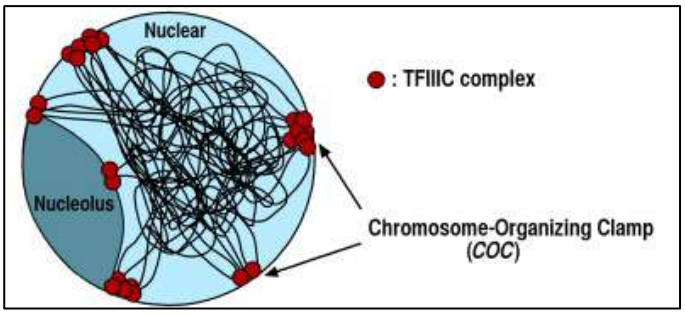


Figure 11: TFIIIC COC.
TFIIIC is recruited to inverted repeats at the boundary of fission yeast heterochromatin and COC loci and tethers them to the nuclear periphery. Through the clustering of TFIIIC, nuclear bodies are formed with the chromosome loci in the periphery which cause a higher order chromosome organization. Taken from (Noma et al. 2006)

Besides binding to ETC sites, TFIIIC was recently shown to be located at Alu elements throughout the genome (Ferrari et al. 2018). Alu elements are suggested to function in a similar way to enhancers in order to increase transcription (Su et al. 2014). The presence of TFIIIC at these sites is thought to change the 3D topology of the genome due to hyper-acetylated histones (Ferrari et al. 2018). More recent studies suggest that TFIIIC is recruited to Alu elements via interaction with the architectural protein Activity-Dependent Neuroprotector Homeobox protein (ADNP) (Ferrari et al., 2019 under review). Besides ADNP, the chromodomain helicase DNA binding protein 4 (CHD4) was also found to be associated with TFIIIC at Alu elements (Ferrari et al. 2019 under review).

1.4. The objectives of this PhD project

Besides its essential function in initiating transcription at type 1 and type 2 promoters of the RNA Polymerase III system, TFIIIC is also presumed to play an important role in the three-dimensional organization of the genome via binding to ETC sites, Alu elements and the formation of COCs. A tight regulation of RNA Pol III transcription is essential as de-regulation is associated with cancerogenesis. To date, it is unclear how TFIIIC works as structural data covering the holo-TFIIIC is lacking in all organisms. Thus, the main questions that is PhD thesis aims to study are:

1. Which is the individual architecture of the TFIIIC subcomplexes and of the holo-TFIIIC complex?
2. What is the molecular mechanism behind TFIIIC function and in particular how is TFIIIC able to adopt to variable promoter organizations?

In order to answer these questions, the six-subunit human TFIIIC complex was studied by initially dividing the complex into its two smaller subcomplexes. This was done since no structural data and only very

limited biochemical data of this complex were available. Studying the subcomplexes first was assumed to facilitate a later expression and purification of the holo-TFIIIC complex. Several DNA constructs were designed, expressed and purification protocols were optimized of different sub-assemblies before the holo-TFIIIC complex was studied.

Due to a very limited amount of human TFIIIC, the TFIIIC complex of the better studied model organism *S. cerevisiae* was also studied. The idea was to use yeast TFIIIC as a platform to later improve the results of the human TFIIIC complex. In order to characterize the yeast TFIIIC complex, DNA constructs were designed, the complex was expressed and a purification protocol was optimized. TFIIIC was then extensively biochemically and biophysically characterized before structural studies were performed.

2. Materials and Methods

2.1. Cloning

The genes used for this thesis were all purchased from Genscript and rare codon optimized for expression in insect cells or *Escherichia coli* (*E. coli*) respectively. In general, each cloning comprised the following steps which will be explained in more details further below: Firstly, the expression vector was linearized by digestion with two different restriction enzymes to ensure directionality during the cloning. Secondly, the gene of interest was amplified from a cloning vector using a Polymerase Chain Reaction (PCR), with primers containing overhangs which were compatible to the digested plasmid vector. Thirdly, the gene of interest was ligated with the digested vector. The construct was subsequently transformed into Stellar™ competent cells and single colonies were picked after overnight incubation. Each of the colonies was grown in 5 ml Lysogeny Broth (LB) medium overnight. The DNA was isolated and purified using the GeneJet Plasmid Mini Prep Kit (Thermo Fisher Scientific). As a final step the DNA construct was analytically digested with an appropriate restriction enzyme and analyzed on an agarose gel before sending it to Source Bioscience for sequencing (Figure 12).

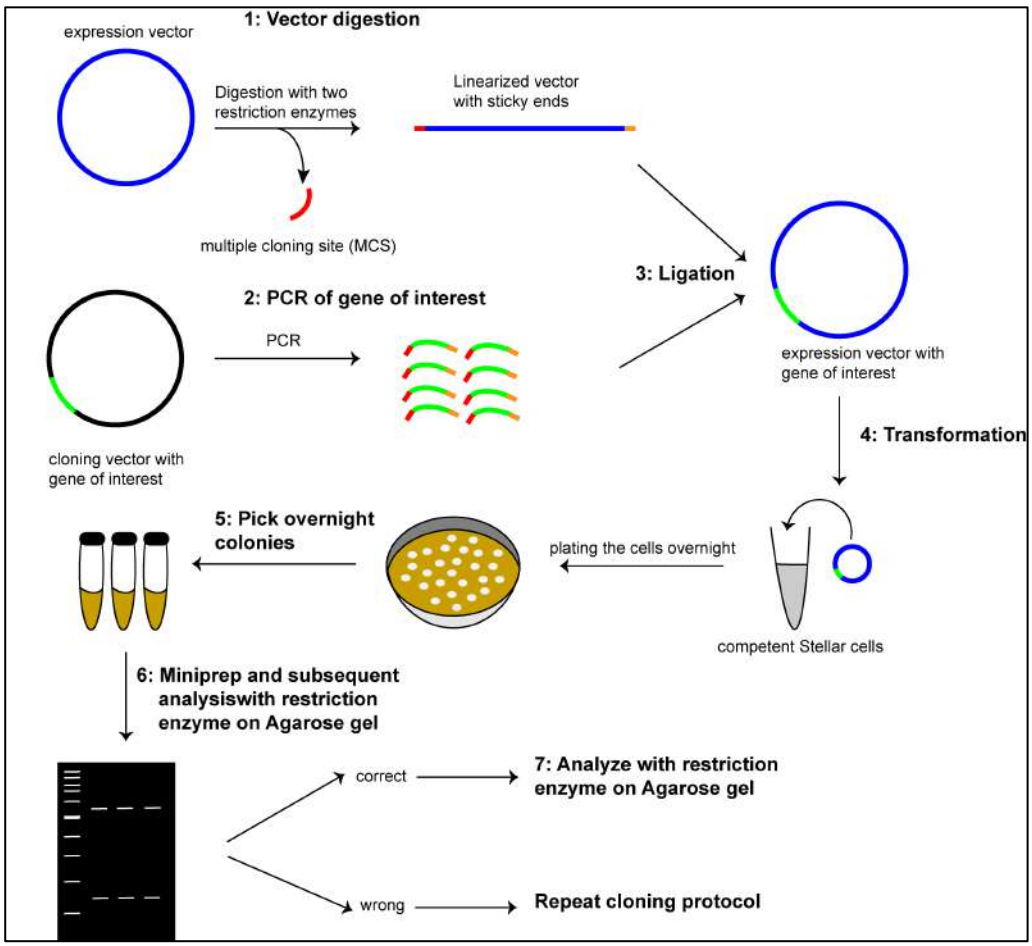


Figure 12: Main steps involved in cloning.

1) The expression vector is digested with two restriction enzymes. 2) The gene of interest is amplified via a PCR reaction with compatible overhangs to the digested vector. 3) Insert and vector are ligated and 4) transformed into Stellar competent cells. The cells are plated on a LB-Agar plate overnight. 5) Single colonies are picked on the next day, 6) the DNA is purified by using a Miniprep Kit and the constructs are then analyzed on an Agarose gel. 7) If the correct pattern is observed, the construct is sent to sequencing, if not, the protocol has to be repeated.

2.1.1. Plasmids and DNA constructs

All plasmids used for cloning and expression are shown in Table 3. The different DNA constructs used for expression in *E.coli* were cloned using the In-Fusion Cloning (Table 4). MultiBac constructs (Table 5) were cloned using the MultiBac Cloning system and biGBac constructs (Table

6) using the Gibson Assembly respectively for expression in insect cells which can also be found below. The different cloning techniques that were used are described in paragraphs 2.1.4, 2.1.5 and 2.1.6.

Table 3: Overview of vectors used for DNA cloning.

Name	Antibiotic resistance	Tag and cleavage site	Vector size (bp)	Expression system
pACEBac1	Gentamicin		2,904	Insect cells
pOPINE	Ampicillin	C-His_6	5,534	*E. coli*
pIDK	Kanamycin		2,077	*E. coli*
pUC57	Ampicillin		2,710	*E. coli*
pOPINF	Ampicillin	N-His_6, 3C	6,218	*E. coli*
pOPINJ	Ampicillin	N-His_6, GST, 3C	5,457	*E. coli*
pLIB	Ampicillin, Gentamicin		4,970	Insect cells
pBIG1a	Ampicillin, Spectinomycin, Gentamicin		5,915	Insect cells
pBIG1b	Ampicillin, Spectinomycin, Gentamicin		5,915	Insect cells
pBIG1c	Ampicillin, Spectinomycin, Gentamicin		5,915	Insect cells
pBIG2ab	Chloramphenicol Ampicillin, Gentamicin		5,668	Insect cells
pBIG2abc	Chloramphenicol Ampicillin, Gentamicin		5,669	Insect cells

Table 4: DNA constructs for expression in E.coli.

*FL = full length. * cloned by Dr Karishma Satia, ** cloned by Dr Nicolas Guthertz, *** cloned by Dr Jerome Gouge.*

construct	organism	vector	tag and cleavage site
TBP core (158-339)*	human	pOPINF	C-His_6, 3C cleavage
Brf1 C-terminus (339-550)/ TBP core (158-339)**	human	pIDK	N-His_6 on Brf1, 3C cleavage
Bdp1 (130-484)***	human	pOPINE	N-His_6, 3C cleavage
TFIIIA, FL	human	pOPINJ	N-His_6, GST, 3C

Table 5: MultiBac DNA constructs for expression in insect cells.

*FL = full length. * cloned by Dr Karishma Satia, ** cloned by Dr Nicolas Guthertz.*

construct	organism	vector	tag and protease
His-τA, FL*	human	pACEBac1	N-His_6 on TFIIIC35, TEV
TFIIIC220, FL	human	pACEBac1	
TFIIIC110, FL	human	pACEBac1	C-Strep
TFIIIC90, FL	human	pUC57	
τB-110Strep, FL*	human	pACEBac1	C-Strep on TFIIIC110
τB-90Strep, FL	human	pACEBac1	C-Strep on TFIIIC90

Brf1-His, FL**	human	pACEBac1	C-His_6
Brf1ΔZR (78-677) **	human	PACEBac1	N-His-3C
HVNL-Bcore-CC-C (73-677) term **	human	pACEBac1	C-His_6
HVNL-Bcore-CC-C (73-677) term **	human	pACEBac1	

Table 6: BiGBac DNA constructs for expression in insect cells.

*FL = full length. * cloned by Dr Erin Cutts. TEV = TEV protease, 3C = 3C protease.*

#	Construct	Organism	Vector	Tag and cleavage site
1	His-τA, FL *	human	pBIG1a	N-His_6 on TFIIIC35, TEV
2	τB-Strep, FL *	human	pBIG1b	C-twin-Strep on TFIIIC220, TEV
3	holo-His-TFIIIC-Strep, FL *	human	pBIG2ab	C-twin-Strep on TFIIIC220 and N-His_6 on TFIIIC35, TEV
4	Brf1 *	human	pLIB	C-twin Strep, TEV
5	TFIIIB-Strep, FL *	human	pBIG1c	C-twin Strep on Brf1, TEV
6	His-τA- TFIIIB-Strep, FL *	human	pBIG2abc	N-His_6 on TFIIIC35, C-twin-Strep on Brf1, TEV
7	TFIIIC220-FLAG, FL	human	pBIG1c	C-3xFLAG, 3C
8	TFIIIC90, FL	human	pLIB	-
9	TFIIIC110, FL	human	pLIB	-
8	TFIIIC90-TFIIIC110, FL (cloned from 8+9)	human,	pBIG1b	-

9	holo-His-TFIIIC-FLAG, FL (cloned from 1+7+8)	human	pBIG2abc	N-His_6 on TFIIIC35, C-3xFLAG on TFIIIC220
10	TFIIIC63, FL	human	pLIB	-
11	His-TFIIIC35, FL	human	pLIB	N-His_6 on TFIIIC35, TEV
12	TFIIIC102, FL	human	pLIB	-
13	TFIIIC102-TFIIIC220-FLAG, FL (cloned from 7+12)	human	pBIG1a	C-3xFLAG on TFIIIC220, 3C
14	His-TFIIIC35-TFIIIC63, FL (cloned from 11+10)	human	pBIG1b	N-His_6 on TFIIIC35
15	TFIIIC4-FLAG/His, FL (cloned from 13+14)	human	pBIG2ab	C-3xFLAG on TFIIIC220, N-His_6 on TFIIIC35
16	τ55, FL	yeast	pLIB	-
17	τ60, FL	yeast	pLIB	-
18	τ91, FL	yeast	pLIB	N-twin Strep
19	τ95, FL	yeast	pLIB	-
20	τ131, FL	yeast	pLIB	-
21	τ138, FL	yeast	pLIB	-
22	His-τ55, FL (cloned from 16)	yeast	pLIB	N-His_6

23	His-τA, FL (cloned from 22+19+20)	yeast	pBIG1a	N-His_6 on τ55
24	τB-Strep, FL *	yeast	pBIG1b	N-twin Strep on τ91
25	holo-TFIIIC-Strep, FL *	yeast	pBIG2ab	N-twin Strep on τ91
26	138-FLAG, FL	yeast	pLIB	C-3xFLAG, 3C
27	τ131-τ95, FL (cloned from 19+20)	yeast	pBIG1b	
28	His-τ55 -τ138-FLAG- (cloned from 26+22), FL	yeast	pBIG1a	C-3xFLAG on τ138, 3C, N-His_6 on τ55
29	His-TFIIIC4-FLAG, FL (cloned from 27+28)	yeast	pBIG2ab	C-3xFLAG on τ138, 3C, N-His_6 on τ55
30	τB_FLAG, FL (cloned from 26+17+18)	yeast	pBIG1b	C-3xFLAG on τ138, 3C
31	TFIIIA FL	human	pLIB	N-His_6

Table 7: Sequences of cleavage sites and protein tags.

The DNA sequence of the tags is shown at the top and the protein sequence is depicted below. Cleavage sites are shown in red and the tags are shown in green.

C-terminal TEV - Twin Strep tag

GAAAACCTGTACTTCCAGTCATGGAGCCATCCGCAATTTGAAAAAGGTGGCGGATCC
GGC
E N L Y F Q S W S H P Q F E K G G G S G

GGAGGTAGCGGCGGAGGTTCTTGGTCTCACCCTCAGTTCGAGAAGTAA
G G S G G G S W S H P Q F E K *

N-terminal His tag - TEV

ATGGCTCACCATCACCATCACCATCACCATCACCATGAAAACCTGTATTTCCAAGGC
M A H H H H H H H H H H E N L Y F Q G

C-terminal 3C - 3x FLAG tag

CTGGAAGTTCTGTTTCAGGGCCCGGACTACAAGGACCACGACGGTGACTACAAGGA
CCACGACATCGACTACAAGGACGACGACGACAAGTGA
M L E V L F Q G P D Y K D H D G D Y K D H D I D Y K D D D D K

2.1.2. Polymerase Chain Reaction

The DNA fragments used for cloning were amplified in a PCR using the Phusion® High-Fidelity DNA Polymerase, (New England Biolabs). For each PCR reaction, forward and reverse primers were designed that annealed at around 60°C with the 3' and 5' end of the gene of interest. Each PCR reaction contained 1x Phusion GC buffer (NEB), 200 μM deoxyadenosine triphosphate (dATP), 0.5 μM of the forward and 0.5 μM of the reverse primer as well as 1 μl of template DNA and nuclease free water to top up the reaction to 50 μl. During the initial denaturation, the template DNA melts and becomes single-stranded before the PCR cycles start. In each cycle, a denaturation step is performed, followed by an annealing step in which the primers anneal with the single stranded template DNA. The Phusion Polymerase has its optimal activity at 72°C in which it synthesizes the complementary DNA to the DNA template in 5'-3' direction. The cycles are repeated for 30 to 35 times to exponentially

increase the amount of DNA generated. After the cycles are finished, a final elongation step follows at 72°C before the reaction goes over into a final hold at 4°C. The schematic PCR protocol is shown in Figure 13.

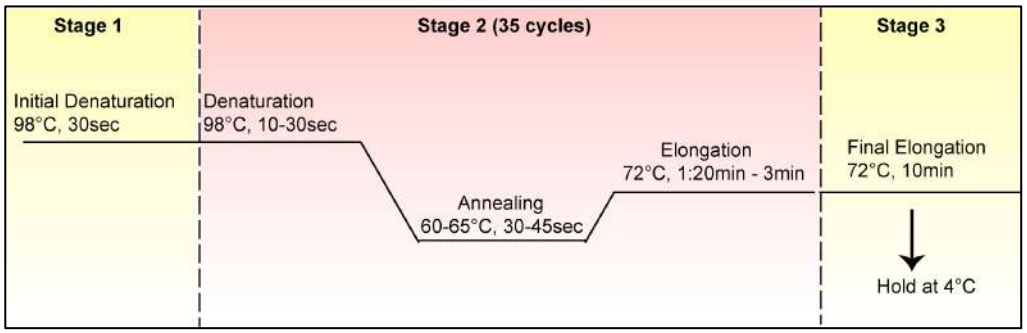


Figure 13: PCR reaction. Schematic overview of the PCR cycle.
sec = seconds, min = minutes.

2.1.3. Transformation into Stellar™ competent cells

Between 0.5 µl and 2.5 µl of a DNA construct were mixed and incubated with 20 µl of Stellar™ competent cells on ice for 30 minutes. After that, the reaction mix with the competent cells was heat-shocked at 42°C for 45 seconds and immediately placed on ice again for another 2 to 3 minutes. 500 µl of prewarmed Super Optimal Broth (SOC) media (Table 11) were added and the cells were grown at 37°C in a shaking incubator for 1h. 100 µl were then plated on a LB-Agar plate containing the appropriate antibiotic as a selection marker and incubated overnight at 37°C. A successful transformation led to single colonies the next morning. Single colonies were picked with a tip and grown separately in 5 ml LB media with the appropriate antibiotic overnight. On the next day, the DNA of each colony was purified using the GeneJET Plasmid Miniprep Kit from Thermo Fisher according to the manufacturer's instructions.

2.1.4. Cloning using the In-Fusion Kit (Clontech)

The In-Fusion Cloning Kit (Clontech) was used to clone constructs for expression in *E.coli*. The protocol was performed according to the manufacturer's instructions.
For the PCR, two primers were designed that contained a 15 bp overhang on each side which was complementary to the vector. For the In-Fusion cloning reaction, the molar ratio of the insert over the plasmid was calculated using the online Clontech Calculator (http://bioinfo.clontech.com/infusion/molarRatio.do). 1x In-Fusion HD Enzyme Premix was added, the reaction was topped up to 10 µl and incubated at 50°C for 15 minutes before it was placed on ice. 2.5 µl of the reaction mix were transformed into Stellar cells as described in paragraph 2.1.3.

2.1.5. Cloning using the MultiBac expression system

The MultiBac expression system, which was used for Baculovirus expression in insect cells, has been developed in Imre Berger's lab (Berger et al. 2004). The gene of interest was thereby cloned into the multiple cloning site (MCS) of a pACEBac1 plasmid digested with HindIII and BamHI and purified as described in paragraph 2.1.8. using the In-Fusion HD Cloning Kit according to paragraph 2.1.4. The MCS of the pACEBac1 was flanked by Tn7 sequences under the control of a polyhedrin promoter and a SV40 terminator sequence. Furthermore, it contained a Gentamycin antibiotic resistance. In order to clone several genes of interest together, which were on different pACEBac1 plasmids, the AlwNI, AvrII, EcoRV and SpeI restriction enzyme system was used. Thereby, the pACEBac1 construct with the larger DNA insert served as the "plasmid" fragment and the pACEBac1 with the smaller DNA insert as an "insert" fragment. The "plasmid" was digested with AlwNI and AvrII and the "insert" with AlwNI, EcoRV and SpeI (Table 8). The digested vectors

were purified according to 2.1.8 and the concentration was measured. For the ligation, a 5 to 7 times molar excess of the “insert” was mixed with the “plasmid” and was ligated using the T4 DNA Ligase (NEB) for 2 to 3h at room temperature according to the NEB online protocol. The ligated construct was then transformed into Stellar™ cells (refer to section 2.1.3).

Table 8: Digestions required for the MultiBac cloning.

The reactions were either incubated 3h at 37°C or overnight at room temperature.

	pACEBac1 digest	AlwNI/AvrII	AlwNI/EcoRV/ SpeI
Template	30 µl	30 µl	30 µl
BamHI (NEB)	2 µl	-	-
HindIII (NEB)	2 µl	-	-
AlwNI (NEB)	-	2 µl	1.5 µl
AvrII (NEB)	-	2 µl	-
EcoRV (NEB)	-	-	1.5 µl
SpeI (NEB)	-	-	1.5 µl
CutSmart Buffer (NEB)	5 µl	5 µl	5 µl
Water	11 µl	11 µl	10.5 µl

2.1.6. Cloning using the biGBac expression system

An optimized “mix and match” cloning strategy for the expression of large multisubunit complexes in insect cells was published from Weissmann et al. in 2016 (Weissmann et al. 2016). Using this approach, all genes of interest were amplified in a PCR reaction and cloned into a BamHI/HindIII digested pLIB vector using the In Fusion Cloning (2.1.4). In a pLIB vector, the gene of interest is under the control of a polyhedron promoter and a SV40 terminator sequence. As shown in Figure 14A, the gene expression

cassette (GEC) in the pLIB could be amplified in a PCR reaction using pre-defined primers which generate the desired linker sequences (α, β, γ, δ, ε, ω) that was necessary for the next assembly step. The primer sequences CasI_for/rev etc. can be found in (Weissmann et al. 2016). In the next assembly step, shown in Figure 14B, a maximum of up to five different GECs, each containing its own polyhedron promoter and terminator, could be cloned into a pBIG1 vector using the Gibson Assembly® Cloning Kit (NEB) to form a polygene cassette (PGC). In order to do so, the pBIG1 vector (pBIG1a-e) was digested with SwaI (Table 9) and gel purified. The pBIG1 vectors are designed to have complementary overlaps to the PCR generated overlaps of the GEC after they were digested with SwaI. For the Gibson Assembly, which was performed for 15 minutes at 50°C, a 3 to 7 molar excess of the PCR amplified GECs was mixed with 100 ng of the linearized pBIG1 vector, mixed with 5µl of 2x Gibson Assembly Master Mix and topped up to 10 µl with water per reaction. During the assembly reaction (shown in Figure 14D), an exonuclease digested the 5' ends, the fragments annealed, a DNA polymerase closed the gaps and a DNA ligase sealed the nicks. Successful integration of all GECs could be seen with an analytical SwaI digest and subsequent analysis of the digestion pattern on an agarose gel.

In the next assembly step, shown in Figure 14C, the PGC from each pBIG1 construct could be released by digesting the construct with PmeI (Table 9). Thereby each PGC has specifically designed overhangs called A-F. For specific overhang sequences, please refer to (Weissmann et al. 2016). Digestion of the pBIG2 vector with PmeI, which was performed according to Table 9, led to a linearized vector with an A and a F overhang sequence, complementary to the overhang sequences from the PGCs. The PGCs were combined into the pBIG2 vector using another Gibson Assembly.

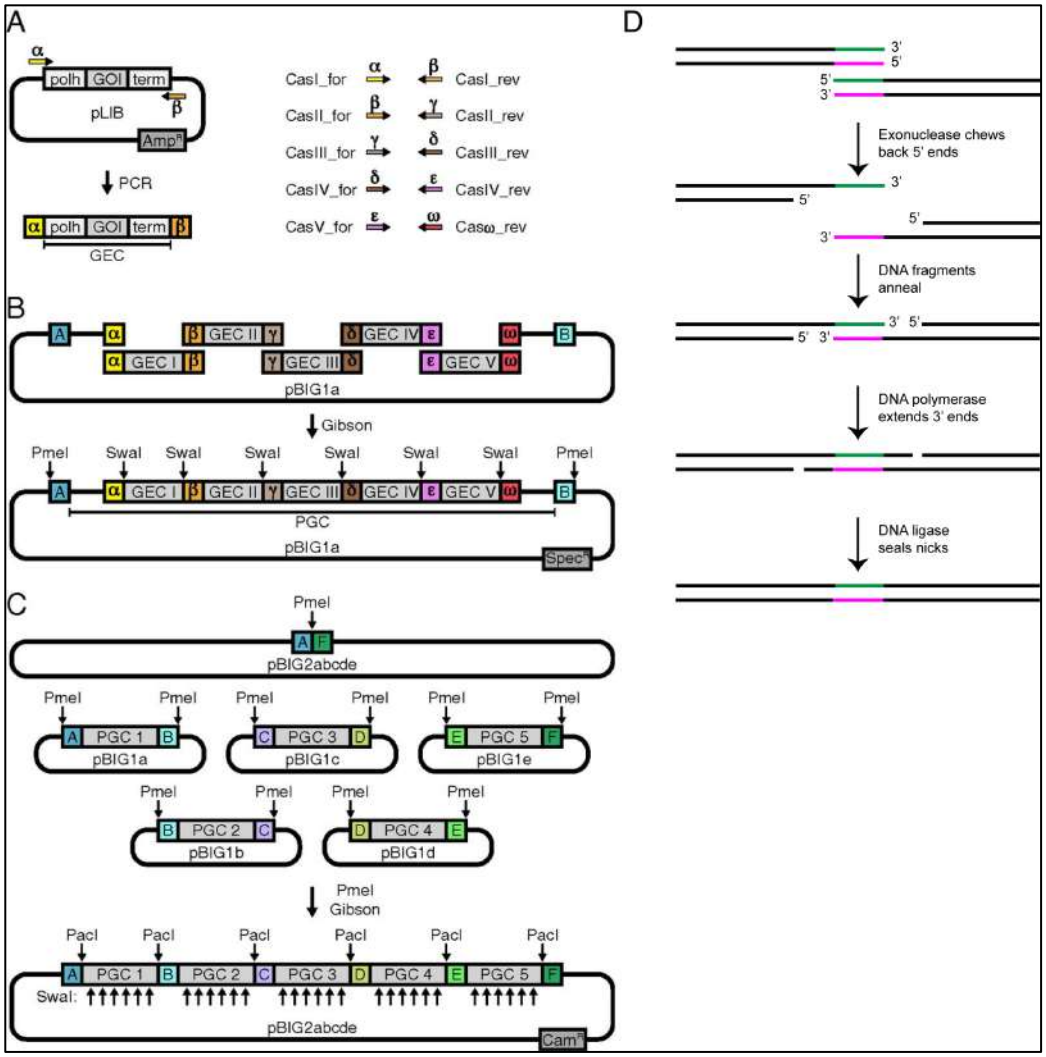


Figure 14: Schematic overview of the biGBac cloning.

A: The gene expression cassette (GEC) is amplified with pre-defined primers via a PCR reaction from the pLIB. B: Stage 1 of the assembly where a SwaI digested pBIG1 vector is assembled via Gibson assembly with the GECs. C: Step 2 of the assembly. The pBIG2 vector is digested with PmeI and assembled with the polygene cassettes (PGC) that were released via PmeI digestion. D: Gibson assembly as shown in the manufacturer's instructions. Adapted from (Weissmann et al. 2016).

Table 9: DNA digestion reactions required for the biGBac cloning strategy.

	pLIB digest	SwaI digest	PmeI digest
DNA	30 µl	30 µl	30 µl
BamHI (NEB)	2 µl	-	-
HindIII (NEB)	2 µl	-	-
SwaI (NEB)	-	2 µl	-
PmeI (NEB)	-	-	2 µl
CutSmart Buffer (NEB)	5 µl	-	5 µl
NEB Buffer 3.1	-	5 µl	-
Water	11 µl	13 µl	13 µl
Reaction conditions	37°C, 3h	room temperature, overnight	37°C, 3h

2.1.7. Bacmid generation using DH10EmBacY cells

Bacmids for protein expression in insect cells could be generated from a pACEBac, a pLIB or a pBIG construct after the genes of interest were successfully assembled as described in 2.1.5 and 2.1.6.

In order to generate a bacmid, the construct was transformed into DH10EmBacY cells. These cells contain a Bacmid with a Kanamycin resistance and a LacZ alpha peptide sequence as well as a helper plasmid encoding for the transposase and containing a Tetracyclin resistance. Via transposition, the gene of interest was inserted in the Tn7 attachment site within the LacZ alpha peptide sequence in the Bacmid. This abolishes the expression of the lacZα peptide, so that no functional β-Galactosidase can be formed. Active β-Galactosidase is able to cleave X-Gal, which is an analogue of Lactose, in order to form 5-bromo-4-

chloro-indoxyl. This substrate can dimerize and oxidize to form a blue colored pigment called 5-5'-dibromo-4,4'-dichloro-indigo. After the transposition, the DH10EmBacY cells containing a Bacmid, were plated onto LB-Agar plates containing selective antibiotics and X-Gal. A correct construct could then be identified as a white colony whereas blue colonies indicated the presence of active β-Galactosidase.

Protocol:

For the transposition, 2.5 µl of the successfully assembled DNA construct were transformed into 75 µl DH10EmBacY cells. The DNA-cell mixture was incubated on ice for 30 minutes, heat shocked at 42°C for 45 seconds and immediately stored on ice for another 2 to 3 minutes. 500 µl of SOC media were added, the cells were grown at 37°C in a shacking incubator for 6 h to enable the transposition and then selectively grown on a Bacmid Transposition plate containing X Gal, Gentamycin, Kanamycin and Tetracylin.

A white colony was picked and grown overnight in 5 ml of LB medium containing Tetracyclin, Kanamycin and Gentamycin in a shaking incubator at 37°C. The cells were pelleted and the supernatant was discarded. The following steps were all performed on ice: the pellet was resuspended in 300 µl resuspension buffer (taken from GeneJet Plasmid Miniprep Kit, Thermo Fisher Scientific) and transferred into a 2 ml Eppendorf tube. 100 µl of lysis buffer were added (taken from GeneJet Plasmid Miniprep Kit, Thermo Fisher Scientific) and 300 µl sodium acetate at pH 5.5, mixed and incubated on ice for 10 min. After that, the samples were centrifuged for 10 minutes at 17,000 g, the supernatant was taken and mixed with 800 µl of Isopropanol. It was incubated on ice for another 10 minutes, centrifuged for 10 minutes at 17,000 g and the supernatant was removed. 500 µl Ethanol were added to wash the pellet, it was centrifuged again, the supernatant was removed and the pellet was dried at room temperature. The dried pellet was afterwards resuspended

in 40 µl of ultra-filtered water and the Bacmid could be used for transfection in Sf9 cells.

2.1.8. DNA purification from an agarose gel

In order to purify PCR products or to identify the correct digestion pattern of DNA cloning constructs after incubation with respective restriction enzymes, the samples were mixed with the purple gel loading dye (NEB) and loaded on a 0.8% agarose 1xTris-acetate-EDTA gel (TAE), containing 40 mM tris base, 20 mM acetic acid and 1 mM EDTA at pH 8. The gel contained ethidium bromide which can intercalate into the DNA and be visualized with UV light. The gel was run at 100 V for 40 minutes. The DNA band of interest was cut out of the gel and extracted from the TAE gel using the Zymoclean Gel DNA recovery Kit from Cambridge Bioscience according to the manufacturer's instructions.

2.2. Protein expression

All proteins expressed for this thesis, were either expressed in insect cells or in *E. coli* as shown in Table 10.

Table 10: E. coli strains and insect cells used for cloning and protein expression.
**will be explained in 2.2.2.*

E. coli	Purpose	Insect cells	Purpose
Stellar (Clontech)	Molecular Cloning	Sf9 cells (Thermo Fisher Scientifc)	virus generation and amplification in stages: P1 and P2 *
BL21 RIL (Stratagene)	Protein expression	High5 cells (Thermo Fisher Scientifc)	Protein expression

Rosetta™ (DE3) pLysS (Novagen)	Protein expression		

2.2.1. Protein expression in *E.coli*

The recombinant plasmid was transformed into competent BL21 RIL or Rosetta cells for protein expression according to section 2.1.3. Several colonies from the LB Agar plate were picked on the next day and grown overnight in 100 ml LB media with the respective selective antibiotics in a shaking incubator at 37°C. 20 ml of the overnight culture were added into 1 L of TB medium, also containing the respective antibiotics. For each protein expression, 12 L of *E.coli* cells were grown in a shaking incubator at 37°C until an OD_{600} between 0.6 and 1.0 was reached. The cultures were shifted to 4°C for 1h. Subsequently, induction followed by addition of 0.5 mM IPTG at 18°C overnight. The cells were harvested via centrifugation for 30 minutes at 3,993 g and the pellet was stored at -80°C.

Table 11: Media composition for expression in E. coli.

The components marked with an asterisk were added on the day of the growing as described in section 2.2.1.

Lysogeny Broth (LB) media	Terrific Broth (TB) media	Super Optimal broth with Catabolite repression (SOC) media
1% w/v tryptone	1.2% w/v tryptone	20mM glucose
0.5% w/v yeast extract	2.4% w/v yeast extract	2% w/v tryptone
170 mM NaCl	0.4% w/v glycerol	0.5% w/v yeast extract
	17 mM KH_2PO_4	10 mM NaCl
	72 mM K_2HPO_4	2.5 mM KCl
		10 mM $MgCl_2$

2.2.2. Protein expression in insect cells

Expressing eukaryotic proteins in insect cells has the advantage that proteins are expressed at a very high level. Furthermore, the recombinant proteins get post-translationally modified through similar pathways than to the ones present in mammalian cells. This allows the production of a functionally very similar protein(-complex) compared to the native eukaryotic protein. Insect cell lines Sf9 and High5 were used. They were cultured in Lonza Insect Xpress media in a shaking incubator at 27°C. High5 cells were passaged three times a week and Sf9 cells two times a week by the Baculovirus facility manager Ruth Knight.

1) Transfection of Sf9 cells

The generated Bacmid, as described in section 2.1.7, was used to generate a P1 virus via transfection of Sf9 cells. Therefore, 2 ml of Sf9 cells at a density of $5x10^5$/ml were transferred into each well of a six-well plate. To allow the Sf9 cells to adhere to the bottom of the wells, the six well plate was incubated for 30 minutes without shaking.
In the meantime, 10 µl of the Bacmid were mixed with 100 µl of the growing media (mix A) and 10 µl of Cellfectin™ II Reagent (Thermo Fisher Scientific) were mixed with 100 µl of growing media (mix B).
The supernatant of the Sf9 cells was removed and replaced with 2 ml of fresh growing media. Mix A and B were combined and added into one well. The six well plate was incubated at 27°C for three days. After that, the transfection could be analyzed using a light microscope. The supernatant of each well was harvested and supplemented with 5% FBS and called “P1 virus”.

2) Virus amplification: P1 → P2

In order to amplify the virus concentration, the P1 virus was added into 25 ml Sf9 cells at a density of $5x10^5$/ml and grown at 27°C in Lonza Insect Xpress media. The efficiency of the transfection could be monitored via cell viability and fluorescence since the EMBacY cells produced yellow fluorescent protein (YFP) which could be used as a marker to identify transfected cells. After 3 to 5 days, 80-90% of the Sf9 cells showed fluorescence. The cells were centrifuged and the 25 ml supernatant containing the virus (called P2 virus) was mixed with 5% fetal bovine serum (FBS) and stored at 4°C.

3) Protein expression

For protein expression, 2 ml of the P2 virus were added into 500 ml of High5 cells at a density of $5x10^5$/ml and grown for 3-4 days at 27°C in a shaking incubator. Cells were harvested for 5 minutes at 233 g and the pellet was frozen at -20°C for short term storage and -80°C for long term storage.

2.3. Protein purification

The purification protocols used for each of the proteins and protein complexes that were purified during the course of the PhD are explained in the following sections. The general purification procedure comprised three steps (Figure 15): an affinity purification step in which the affinity tag of the protein was specifically bound to a resin, an ion exchange step that enabled purifying proteins based on their charge and thereby removing further contaminants such as DNA or proteins with a different charge and a size exclusion step that fractionated the sample according to the size, preventing the presence of aggregates inside the protein prep. All cell pellets, used for the protein purifications explained in the following, were lysed using the Vibra-Cell Ultrasonic Liquid Processors (Sonics).

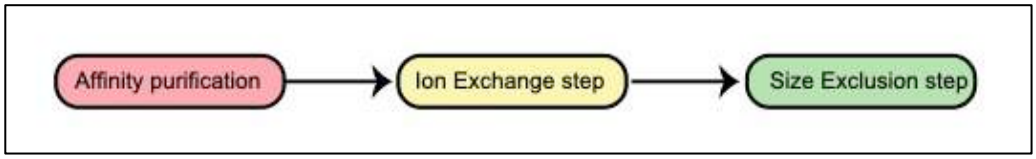


Figure 15:Purification steps.

2.3.1. Purification protocols for human proteins expressed in E.coli

2.3.1.1. Purification protocol for human TBP (158-339)

The TBP core construct (Table 4) was expressed in 6 L of Rosetta pLysS cells which were grown in TB medium and induced with 0.5 mM IPTG at an OD_{600} of 1.5 at 18°C. The pellet was stored at -80°C before use.
TBP was purified in a three-step process via a HisTrap HP column, a HiTrap Heparin HP column and a HiLoad 16/600 Superdex 200pg gel filtration column using a protocol which was already established in the lab. On the day of the purification, the pellet was resuspended in 250 ml lysis buffer (375 mM KCl, 50 mM Hepes pH 7.8, 10% glycerol, 2 mM $MgCl_2$, 10 mM Imidazole), supplemented with 10 mM beta-mercaptoethanol (BME), a tip of DNase I (2-5 mg) (Sigma Aldrich), 1 EDTA free Protease Inhibitor tablet (Thermo Fisher Scientific) and 1 ml self-made protease inhibitor (Table 12). The fully resuspended pellet was sonicated in three cycles of 5 minutes each at 4°C at 60% (each cycle: 3 seconds on, 5 seconds off) and centrifuged at 37,565 g for 30 minutes. The filtered supernatant was supplemented with 15 mM Imidazole and loaded on a 5 ml His Trap HP column, washed with 5 column volumes (CV) of lysis buffer and eluted in 3 CV elution buffer (375 mM KCl, 50 mM Hepes pH 7.8, 10% glycerol, 2 mM $MgCl_2$, 300 mM Imidazole, 10 mM BME). The elution fractions were loaded onto a HiTrap Heparin HP column pre-equilibrated with 25% Buffer B. (Buffer A: 50 mM Hepes pH 7.8, 10% glycerol, 10 mM BME, Buffer B: 50 mM Hepes pH 7.8, 10% glycerol, 2 M KCl, 10 mM BME). The sample was eluted in a gradient from 25 to 75% of Buffer B over a course of 10 CV. TBP eluted at ~ 1 M KCl. The His tag of TBP was cleaved overnight at 4°C using 3C protease in a ratio of protease:protein 1:100 (w/w). On the next day, the cleaved protein was supplemented with 50 mM Imidazole, re-loaded on a His Trap HP column and the flow through collected. It was concentrated using a 10 kDa molecular weight cut-off (MWCO) VivaSpin concentrator (Generon) and

loaded on a HiLoad 16/600 Superdex 200pg gel filtration column (running buffer: 25 mM Hepes pH 7.8, 5% glycerol, 200 mM KCl, 2 mM Dithiothreitol (DTT)). TBP was concentrated again, flash-frozen in liquid nitrogen and stored at -80°C.

2.3.1.2. Purification protocol for human Brf1 C-terminus (339-550)

The Brf1 C terminus/ TBP core construct (Table 4) was expressed in 12 L BL21 RIL *E. coli* cells which were grown at 37°C in TB medium and induced with 1 mM IPTG at an OD_{600} of 1.8 at 18°C overnight. The pellet was stored at -80°C before use. During the purification process, TBP was lost and only Brf1 could be purified so that this purification is referred to as Brf1 C-terminus. This method was used to avoid proteolytic cleavage of the Brf1 C-terminus, as observed when Brf1 C-terminus was expressed in isolation. The purification protocol was optimized by Nicolas Guthertz, a former PhD student, in the lab.

All steps were performed at 4°C. The pellet was resuspended in lysis buffer (50 mM Hepes pH 7.8, 10% glycerol, 500 mM KCl, 2 mM $MgCl_2$), supplemented with 10 mM BME, a tip of DNase I (Sigma Aldrich), 1 EDTA free Protease Inhibitor tablet (Thermo Fisher Scientific) and 1 ml self-made protease inhibitor (Table 12). The pellet was sonicated for three cycles with 3 seconds off and 5 seconds on (60% amplitude) and centrifuged for 30 minutes at 37,565 g. The protein was purified via a His-Trap HP, adding 40 mM Imidazole for the washing step and eluted with lysis buffer containing 300 mM Imidazole. The eluted fractions were diluted with Buffer A (50 mM Hepes pH 7.8, 10% glycerol, 10 mM BME) to a KCl concentration of 375 mM and subsequently loaded on a HiTrap Heparin HP column. A gradient from 15 to 50% of Buffer B (50m M Hepes pH 7.8, 10% glycerol, 2 M KCl, 10 mM BME) over 10 CV was performed and the recombinant protein eluted at ~ 500 mM KCl. The His tag of the protein was cleaved overnight using 3C protease in a protease:protein ratio of 1:50. On the next day, the cleaved protein was supplemented with

50 mM Imidazole, re-loaded on a His Trap HP column and the flow through collected. It was concentrated using a 10 kDa MWCO VivaSpin concentrator (Generon) and loaded on a Superdex 200 16-600 gel filtration column (running buffer: 25 mM Hepes pH 7.8, 5% glycerol, 200 mM KCl, 2 mM DTT). Brf1 C-terminus was concentrated again, flash-frozen in liquid nitrogen and stored at -80°C.

2.3.1.3. Purification protocol for human Bdp1 (130-484)

The expression and purification protocol for Bdp1 (130-484) was obtained from Dr. Jerome Gouge. Bdp1 (Table 4) was expressed in 12 L BL21-RIL *E. coli* cells which were grown at 37°C in TB medium and induced with 0.5 mM IPTG at an OD_{600} of 1.5 at 18°C overnight. The pellet was stored at -80°C before use.

The pellet was resuspended in lysis buffer (50 mM Hepes pH 7.8, 10% glycerol, 500 mM KCl, 2 mM $MgCl_2$,), supplemented with 10 mM BME, a tip of DNase I (Sigma Aldrich), 1 EDTA free Protease Inhibitor tablet, (Thermo Fisher Scientific), 4 µl Benzonase and 1 ml self-made protease inhibitor (Table 12). The pellet was sonicated for three cycles with 3 seconds off and 5 seconds on (60% amplitude) and centrifuged for 30 minutes at 37,565 g. The protein was purified via a His-Trap HP, adding 5 mM Imidazole for the loading step and 40 mM Imidazole for the washing step and eluted with 300 mM Imidazole. The eluted fractions were diluted with Buffer A (50 mM Hepes pH 7.8, 10% glycerol, 10 mM BME) to a KCl concentration of 200 mM and subsequently loaded on a HiTrap Heparin HP column. A gradient from 20 to 50% of Buffer B (50 mM Hepes pH 7.8, 10% glycerol, 2 M KCl, 10 mM BME) over 10 CV was performed. The recombinant protein eluted at ~500 mM KCl.

The Bdp1 (130-484) construct contained a cleavable His tag which was cleaved overnight using 3C protease in a protease:protein ratio of 1:50. On the next day, the cleaved protein was supplemented with 40mM Imidazole, re-loaded on a His Trap HP column and the flow through

collected. It was concentrated using a 10 kDa MWCO VivaSpin concentrator (Generon) and loaded on a Superdex 200pg 16/600 gel filtration column (running buffer: 50 mM Hepes pH 7.8, 10% glycerol, 500 mM Ammonium Acetate, 2 mM DTT). After the gel filtration step, Bdp1 was concentrated again, flash-frozen in liquid nitrogen and stored at -80°C.

2.3.2. Purification protocols for human protein constructs expressed in insect cells

2.3.2.1. Purification protocol of Multibac expressed human constructs His-τA

It was possible to optimize a three-step purification protocol for τA (Table 5). τA was expressed for three days in High5 cells using the Baculovirus expression system. The pellets were stored at -80°C before being used. All steps were performed at 4°C. The pellet was resuspended in 125 ml lysis buffer (500 mM NaCl, 20 mM Hepes pH 7.9, 10 mM Imidazole, 10 mM BME), supplemented with DNase I (Sigma Aldrich), 2 EDTA free Protease Inhibitor tablets (Thermo Fisher Scientific) and 1 ml self-made protease inhibitor (Table 12). The resuspended pellet was sonicated at 20% amplitude for 1.5 minutes (15 sec on, 59 sec off) before centrifugation for 40 minutes at 37,565 g. The sample was subsequently loaded onto a His Trap HP column, washed with 5 CV lysis buffer and eluted in 3 CV using lysis buffer supplemented with 300 mM Imidazole. The elution fractions were diluted to 200mM NaCl using Buffer A (20 mM Hepes pH 7.9, 10% glycerol, 10 mM BME) and loaded onto a HiTrap Heparin HP column. Protein elution was done during a gradient from 10% to 50% of Buffer B (2M NaCl, 20mM Hepes pH 7.9, 10% glycerol, 10mM BME). Although τA contained a TEV cleavable His Tag, no cleavage was performed since it led to a high loss in protein yield. Instead, the sample was concentrated and loaded onto a HiLoad 16/600 Superdex 200pg

(running buffer: 500 mM NaCl, 10% glycerol, 20 mM Hepes pH 7.9, 1 mM DTT). Analytical gel filtrations were performed by injecting 500 µl of the sample onto a Superose 6 Increase 10/300 GL column. After the gel filtration step, τA was concentrated, flash-frozen in liquid nitrogen and stored at -80°C.

2.3.2.2. Purification protocol of Multibac expressed human constructs τB-110Strep and τB-90Strep

Both τB constructs (Table 5) were expressed for three days in High5 cells at 27°C using the Baculovirus expression system. The pellets were stored at -80°C before being used. For both constructs, τB was mainly found in the insoluble fraction and thus a protocol to recover a small yield of the protein from the cell pellet via refolding with 8 M Urea was established. The cell pellet was resuspended in 100 ml phosphate-buffered saline (PBS) containing 2 tablets of EDTA free Protease Inhibitor tablets (Thermo Fisher Scientific) and sonicated at 20% amplitude for 1.5 minutes (15 sec on, 59 sec off) before centrifugation for 40 minutes at 37,565 g.

Since the protein was found in the pellet instead of the supernatant, the pellet was re-suspended again in 100 ml PBS with 1% Triton X-100, stirred for 2h at 4°C and centrifuged for 15 minutes at 12,000 g. This procedure was repeated twice before the Urea unfolding. Then, 50 ml of 8 M Urea with 10 mM CAPS pH 10.5 were added to the washed cell pellet and stirred for 1h at room temperature. The solution was transferred into snake skin (Thermo Fisher Scientific) for dialysis (dialysis buffer: 50 mM Tris pH 7.5, 0.8 M NaCl, 1 mM $MnCl_2$, 0.02% Tween, 10% glycerol, 10 mM BME) overnight at 4°C. The dialyzed sample was diluted to 200 mM with Buffer A (50 mM Tris pH 7.5, 0 M NaCl, 1 mM $MnCl_2$, 0.02% Tween, 10% glycerol, 10 mM BME) and loaded on a HiTrap Heparin HP column. The column was washed with 10 CV of 10% Buffer B (50 mM Tris pH 7.5,

2 M NaCl, 1 mM $MnCl_2$, 0.02% Tween, 10% glycerol, 10 mM BME) and the protein was eluted in a 10 to 100% gradient of Buffer B.
For analytical purposes, 500 µl of the sample were loaded onto a Superose 6 Increase 10/300 GL column (running buffer: 500 mM NaCl, 50 mM Tris pH 7.9, 10% glycerol, 1 mM $MnCl_2$, 1 mM DTT).

2.3.2.3. Purification protocol of biGBac expressed human holo-His-TFIIIC-Strep

Holo-His-TFIIIC-Strep (Table 6, #3) was expressed for three days in High5 cells using the Baculovirus expression system. The pellet was stored at -80°C before being used.
All steps were performed at 4°C. The pellet was resuspended in 150 ml lysis buffer (500 mM NaCl, 20 mM Hepes pH 8, 1 mM $MgCl_2$,10% glycerol, 10 mM BME), supplemented with DNase I (Sigma Aldrich), 3 EDTA free Protease Inhibitor tablets (Thermo Fisher Scientific), 4 µl Benzonase and 2 ml self-made protease inhibitor (Table 12). The resuspended pellet was sonicated at 20% amplitude for 45 seconds (5 sec on, 10 sec off) before centrifugation for 30 minutes at 37,565 g. The sample was subsequently loaded onto a StrepTrap HP at a constant flow rate of 1.5ml/min, washed with 5 CV lysis buffer and eluted in 3 CV lysis buffer supplemented with 0.05% (w/v) D-Desthiobiotin. The elution fractions were diluted to 200 mM NaCl using Buffer A (20 mM Hepes pH 8, 10% glycerol, 10 mM BME) and loaded onto a 1 ml HiTrap Heparin HP column. To prevent dilution of the sample, the protein was eluted over a sharp gradient from 10 to 100% of Buffer B (2 M NaCl, 20 mM Hepes pH 7.9, 10% glycerol, 10 mM BME) in 8 CV. The sample was concentrated and loaded onto a Superose 6 Increase 10/300 GL (running buffer: 150 mM NaCl, 5% glycerol, 20 mM Hepes pH 8, 2 mM DTT). For structural studies, TFIIIC could neither be concentrated nor be frozen after the gel filtration step. This was mainly due to the low yield and the thus resulting low protein concentration as

well as the complex not being stable enough to be frozen and thawed. For biochemical experiments, TFIIIC could be concentrated and flash-frozen in liquid nitrogen before being stored at -80°C.

2.3.2.4. Purification protocol of biGBac expressed human holo-His-TFIIIC-FLAG and human TFIIIC4-FLAG

Both constructs (Table 6, #9 and #15) were expressed for three days in High5 cells using the Baculovirus expression system. The pellets were stored at -80°C before they were used and they were all purified using the same protocol.

All steps were performed at 4°C. The pellet was resuspended in 150 ml lysis buffer (500 mM NaCl, 20 mM Hepes pH 8, 1 mM $MgCl_2$,10% glycerol, 5 mM BME), supplemented with DNase I (Sigma Aldrich), 2 EDTA free Protease Inhibitor tablets (Thermo Fisher Scientific), 4 μl Benzonase and 2 ml self-made protease inhibitor (Table 12). The resuspended pellet was sonicated at 20% amplitude for 45 seconds (5 sec on, 10 sec off) before centrifugation for 30 minutes at 37,565 g. The supernatant was subsequently incubated with 4 ml Anti-DYKDDDDK G1 Affinity resin (Genscript) on a rolling plate at 4°C for 3h. Elution was done in a three-step process using the lysis buffer supplemented with 200 μg/ml DYKDDDDK peptide (B2 Scientific). Firstly, the beads were equilibrated with 4 ml elution buffer for 1 minute (E1), then 10 ml of elution buffer were added and incubated with the beads for 15 minutes (E2) and a final elution was done adding 4 ml of elution buffer for another 5 minutes (E3). Since the FLAG elution already contained highly pure material, the elution fractions were pooled, concentrated to 1 mg/ml and flash-frozen in liquid nitrogen for storage at -80°C. This sample could be used for biochemical experiments. In order to further increase the purity for structural studies, the elution fractions were diluted to 150 mM using Buffer A (20 mM Hepes pH 8, 10% glycerol, 10 mM BME) and loaded onto a 1 ml HiTrap Heparin

HP column. Using a 1 ml instead of a 5 ml column, allowed to concentrate the sample. TFIIIC was eluted over a gradient from 7.5 to 100% of Buffer B (2 M NaCl, 20 mM Hepes pH 7.9, 10% glycerol, 10 mM BME) in 10 CV. The sample was concentrated, flash frozen and stored at -80°C for biochemical studies or loaded analytically onto a Superose 6 Increase 10/300 GL column (running buffer: 200 mM NaCl, 5% glycerol, 20 mM Hepes pH 8, 2 mM DTT). After the gel filtration step, TFIIIC had to be used immediately and could not be stored.

2.3.3. Purification protocols for S. cerevisiae TFIIIC constructs

2.3.3.1. Purification protocol of biGBac expressed S. cerevisiae holo-TFIIIC-Strep and τB-Strep constructs

Holo-TFIIIC and τB-Strep (Table 6, #24 and #25) were expressed for three days in High5 cells using the Baculovirus expression system. The pellets were stored at -80°C before being used. During the purification, τ138 of the τB-Strep sample became insoluble so that instead of τB, only τ60/τ91 could be purified using this construct.

All steps were performed at 4°C. The pellet was resuspended in 125 ml lysis buffer (500 mM NaCl, 20 mM Hepes pH 8, 1 mM $MgCl_2$, 10% glycerol, 10 mM BME), supplemented with DNase I (Sigma Aldrich), 2 EDTA free Protease Inhibitor tablets (Thermo Fisher Scientific), 4 µl Benzonase and 2 ml self-made protease inhibitor (Table 12). The resuspended pellet was sonicated at 20% amplitude for 45 seconds (5 sec on, 10 sec off) before centrifugation for 30 minutes at 37,565 g. The sample was subsequently loaded onto a Strep Trap HP at a flow rate of 2.0 ml/min, washed with 5 CV lysis buffer and eluted in 3 CV lysis buffer supplemented with 0.05% (w/v) g D-Desthiobiotin. The elution fractions were diluted to 150 mM NaCl using Buffer A (20 mM Hepes pH 8, 10% glycerol, 10 mM BME). TFIIIC was loaded onto a 5 ml HiTrap Heparin HP column whereas τ60/τ91 was loaded onto a 1 ml HiTrap Heparin HP

column. The protein complexes were eluted over a sharp gradient from 7.5 to 100% of Buffer B (2 M NaCl, 20 mM Hepes pH 7.9, 10% glycerol, 10 mM BME) in 10 CV.
After this step, TFIIIC was concentrated and loaded onto a Superose 6 16/70 column (running buffer: 200 mM NaCl, 2.5% glycerol, 20 mM Hepes pH 8, 1 mM DTT), concentrated and flash-frozen in liquid nitrogen for storage at -80°C.
τ60/τ91 was concentrated and flash-frozen after the Hi Trap Heparin purification for biochemical experiments. For analytical purposes, τ60/τ91 was loaded onto a Superose 6 Increase 10/300 GL column (running buffer: 200 mM NaCl, 2.5% glycerol, 20 mM Hepes pH 8, 1 mM DTT).

2.3.3.2. Purification protocol of the biGBac expressed S. cerevisiae His-τA construct

The yeast construct of His- τA (Table 6, #23) was expressed for three days in High5 cells before the pellet was harvested.
All steps were performed at 4°C. The pellet was resuspended in 125 ml lysis buffer (500 mM NaCl, 20 mM Hepes pH 8, 10% glycerol, 1 mM $MgCl_2$, 10 mM Imidazole, 10 mM BME), supplemented with DNase I (Sigma Aldrich), 2 EDTA free Protease Inhibitor tablets (Thermo Fisher Scientific) and 1 ml self-made protease inhibitor (Table 12). The resuspended pellet was sonicated at 20% amplitude for 45 seconds (5 sec on, 10 sec off) before centrifugation for 40 minutes at 37,565 g. The sample was subsequently loaded onto a His Trap HP column, washed with 5 CV lysis buffer and eluted in 3 CV lysis buffer supplemented with 300 mM Imidazole. The elution fractions were diluted to 200 mM NaCl using Buffer A (20 mM Hepes pH 8, 10% glycerol, 10 mM BME) and loaded onto a 1 ml HiTrap Heparin HP column. The protein complex was eluted over a gradient from 10 to 50% of Buffer B (2 M NaCl, 20 mM Hepes pH 8, 10% glycerol, 10 mM BME). After the Heparin step, τA was

pooled, concentrated and loaded onto a Superose 6 Increase 10/300 GL column. The elution peak was concentrated again, flash-frozen in liquid nitrogen and stored at -80°C.

2.3.3.3. Purification protocol of biGBac expressed S. cerevisiae τB-FLAG and His-TFIIIC4-FLAG constructs

Both constructs (Table 6, #29 and #30) were expressed and purified as described for the human holo-TFIIIC FLAG complex further above.

Table 12: Composition of self-made protein inhibitor (250x).

Benzamidine	3.3 g
Phenylmethylsulfonyl fluoride (PMSF)	1.7 g
Ethanol	40 ml

2.4. Biochemical and biophysical analysis

2.4.1. Protein concentration determination

After or during protein purifications, the absorption at 280 nm was measured using a Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific) in order to calculate the protein concentration. Using the Beer-Lambert Law (Equation 1), the protein concentration could be calculated using the protein extinction coefficient determined by the online ProtParam tool (https://web.expasy.org/protparam/). The absorption at 260 nm was also measured in order to analyze the amount of DNA contamination. A_{260}/A_{280} ratio of 0.57 corresponded to no DNA contamination whereas higher ratios indicated DNA contamination

(https://www.biotek.com/resources/docs/PowerWave200_Nucleic_Acid_Purity_Assessment.pdf).

$$A = \varepsilon * c * x$$

$$c = \frac{A}{\varepsilon * x}$$

Equation 1: Beer-Lambert Law.

A = absorbance, ε = extinction coefficient, c = concentration, $x =$ path length = 1.

2.4.2. SDS-PAGE analysis

Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) analysis is an electrophoretic separation technique. The protein sample was mixed with a gel loading dye (Table 13) in a 1:6 ratio and boiled at 95°C for 3 minutes. This step was necessary to denature all proteins. SDS charges the linearized protein chains negatively so that all proteins migrate from the cathode to the anode which allows a separation purely by mass inside the gel.

The samples were loaded in commercial 4-12% NuPAGE SDS-PAGE (Life Technologies) and ran in 1x NuPAGE MES buffer (Invitrogen). A protein ladder (P7712S, NEB) was used to identify the molecular weights. The gels were run at 200 V for 40 minutes and stained with Instant Blue (Expedeon).

Table 13: Composition of the 6x SDS Loading dye.

6x SDS Loading Dye
400 mM DTT
8% SDS
0.4% bromphenol blue
40% glycerol
200 mM Tris-HCl pH 6.8

2.4.3. Silver Stain Analysis

Silver Staining was performed to detect proteins that were present at a very low yield and not detectable via staining an SDS-PAGE with Instant Blue (refer to section 2.4.2). Therefore, the Pierce™ Silver Stain Kit was used according to the manufacturer's instructions. The ultra-sensitive Silver Stain Kit allowed to detect proteins at 0.25 ng.

2.4.4. LC-MS/MS mass spectrometry

In order to unambiguously identify the protein bands in the SDS-PAGE, the gel bands were sent for liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis.
The band of interest was cut and either given to the Proteomics Facility of the Institute of Cancer Research for a tryptic digest and a subsequent LC-MS/MS analysis or digested using the In-Gel Tryptic Digestion Kit (Thermo Fisher Scientific) and sent to the UCL School of Pharmacy for an LC-MS/MS analysis. For the mass spectrometry analysis, the trypsin-digested peptides were ionized by electrospray ionization and separated by the Velos Orbitrap mass spectrometer (Thermo Fisher Scientific). Comparing the resulting spectra to a library gave the results which could be read out using the Scaffold software.

2.4.5. EMSAs

All electrophoretic mobility shift assays (EMSAs) were run at 4°C in 1x Tris-borate EDTA (TBE) Buffer either using self-made 4% Acrylamide native pages or commercial Novex™ TBE Gels 4-12% (Invitrogen).

Self-made native PAGEs were pre-run for 10 minutes, 1.5 µl of each sample were loaded in each well and ran at a constant amperage of 40 mA.

Commercial native pages were pre-run for 1.5h at 130 V, then 10 – 12 µl of each sample was loaded per well. The EMSA was run for 1.5h at 130 V.

Table 14: Composition of self-made 5% Agarose native pages.

The protocol shown below is sufficient to prepare 4 gels.

4% Acrylamide (19:1)	2.0 ml
0.5x TBE (10x stock: 1x 89 mM Tris, Boric acid 89 mM, EDTA 2mM, pH 8.3)	1.0 ml
5% glycerol	1.0 ml
Water	15.5 ml
TEMED	30 µl
APS	300 µl

Table 15: DNA Oligos used for EMSA experiments.

TATA Box shown in purple, A Box in orange, B Box in green, Terminator sequence in red. bp = base pairs.

Human DNA constructs

TRNAM2_ABox_BBox CATGATAAATAAGAAGAGTAGAAGCGTGTTTTCCGTT AGCAGAGTGGCGCAGCGGAAGCGTGCTGGGCCCAT AACCCAGAGGTCGATGGATCTAA	95 bp
TRNAM2_ABox CATGATAAATAAGAAGAGTAGAAGCGTGTTTTCCGTT AGCAGAGTGGCGCAGCGGAAGCGT	61 bp
HVG2_ABox_BBox AGAAAATATAATCCGGAGATAGCCTCTGGGCTGGCT TTAGCTCAGCGGTTACTTCGAGTACATTGTAACCACC TCTCTGGGTGGTTCGAGACCCGCGG	98 bp
HVG2_ABox AGAAAATATAATCCGGAGATAGCCTCTGGGCTGGCT TTAGCTCAGCGGTTACT	53 bp
TRNAM1_TSS_till_term GGCAGAACAGCAGAGTGGCGCAGCGGAAGCGTGCT GGGCCCATAACCCAGAGGTCGATGGATCGAAACCAT CCTCTGCTAGGTCCTTTTTTTTT	94 bp
TRNAM2_deleted_ABox TCCGTTAGCAGAGAGCGTGCTGGGCCCATAACCCAG AGGTCGATGGATCTAA	52 bp
(Cy5)-TRNAM2_BBox_24bp_upstream AGCGTGCTGGGCCCATAACCCAGAGGTCGATGGATC TAA	39 bp
Cy3-TRNAM2_ Abox CCGTTAGCAGAGTGGCGCAGCGGAAGCGTGCTGGG C	36 bp

yeast DNA constructs	
tD(GUC)K_noTATA_short CCGTGATAGTTTAATGGTCAGAATGGGCGCTTGTCG CGTGCCAGATCGGGGTTCAATTCCCCGTCGCGGAGA TTTTTTATTTT	83 bp
tD(GUC)K _Ultramer CTGTAGAGGTACCAATTCGACGAAATATAAAAATTTA AAACTAAGAGAAAAAATCCAACAAATAACGTTCCGTG ATAGTTTAATGGTCAGAATGGGCGCTTGTCGCGTGC CAGATCGGGGTTCAATTCCCCGTCGCGGAGATTTTTT ATTTTGCTTT	157 bp
tD(GUC)K_noTATA_short_A_Box CCGTGATAGTTTAATGGTCAGAATGGGCGCTTGTCG CGTGCCAGATCGGG	50 bp
tD(GUC)K_noTATA_short_B_Box TCAGAATGGGCGCTTGTCGCGTGCCAGATCGGGGTT CAATTCCCCGTCGCGGAGATTTTTTATTTT	66 bp

Table 16: Interaction buffers.

Each reaction was topped up with the respective interaction buffer to a total volume of 20 µl.

Interaction buffer A	150 mM NaCl, 20 mM Tris pH 7.5, 1mM $MgCl_2$, 10% glycerol, 2 mM DTT
Interaction buffer B	500 mM NaCl, 20 mM Tris pH 7.5, 1mM $MgCl_2$, 10% glycerol, 2 mM DTT
Interaction buffer C	700mM NaCl final, 20 mM Hepes pH 8, 10% glycerol
Interaction buffer D	150 mM NaCl, 20 mM Hepes pH 8, 5% glycerol, 1mM DTT

2.4.6. Pulldowns

The principle of pulldown experiments is that the protein of interest, called the bait protein, contains an affinity tag which can be immobilized via binding to an appropriate affinity resin. Potential interaction partners, called prey proteins, that do not contain an affinity tag, are incubated with the affinity resin and the bait protein. After a set incubation time, the resin is washed a few times with the interaction buffer and the presence of the prey proteins is detected in an SDS-PAGE after washing a few times with the interaction buffer.

For the τA-TFIIIB pulldown, 0.13 nmol of bait protein were incubated with 0.40 nmol of prey proteins and 50 µl of affinity resin. The samples were topped up with low salt interaction buffer (150 mM NaCl, 20 mM Hepes pH 8, 10% glycerol, 20 mM Imidazole, 10 mM BME) or high salt interaction buffer (500 mM NaCl, 20 mM Hepes pH 8, 10% glycerol, 10 mM BME) to 350 µl and incubated at 4°C for 3h. Then, the resins were washed three times with 10 CV of the respective interaction buffer. The supernatant was removed, 6x SDS Loading Dye added and the beads were directly analyzed via an SDS-PAGE.

For the TFIIIC-ADNP/CHD4 pulldown, High5 cells were co-infected with the TFIIIC and the ADNP Baculovirus, or the CHD4 Baculovirus respectively, and expressed for three days at 27 °C in a shaking incubator. The cells were pelleted at 233g for 5 minutes, resuspended (lysis buffer: 500mM NaCl, 20mM Hepes pH 8, 1mM $MgCl_2$, 10% glycerol, 5mM BME) and incubated with 50 µL of HA resin or 50 µL of Anti-DYKDDDDK (FLAG) resin for 3h at 4°C on a rolling plate. The beads were subsequently washed three times with 20 CV of lysis buffer and mixed with 6x SDS Loading dye before being analyzed via an SDS-PAGE.

2.4.7. Analytical gel filtration chromatography

Analytical gel filtration chromatographies were performed to study either protein-protein or protein-DNA interactions. Furthermore, running profiles of purified protein complexes could be analyzed.
A gel chromatography column consists of a stationary phase that is tightly packed with polymer beads. A constant flow of running buffer is applied and the sample travels through the stationary phase after the injection. Depending on their size and shape, the injected protein complexes have specific retention properties in the packed resin. Small proteins and complexes are subjected to higher delays inside the bead/holes whereas large proteins and aggregates are unable to enter into any of the holes and elute earlier from the resin. In all performed analytical gel filtration experiments in this thesis, 250 μg of the protein of interest were used. It was mixed at a specific molar ratio with either its interacting partner proteins or DNA and then loaded onto a Superose 6 Increase 10/300 GL column (GE Healthcare). The fractions that eluted from the column were analyzed via SDS-PAGE.

2.4.8. SEC-MALS

Size exclusion chromatography coupled with multi-angle static laser light scattering (SEC-MALS) is used to determine the absolute molar mass M_w and the root mean square (rms) radius, also called the radius of gyration R_g of a sample in solution. Furthermore, most setups are coupled to dynamic light scattering (DLS). In contrast to static light scattering, in which the average scattered intensity is integrated over time, in dynamic light scattering, the fluctuations of scattered photos are measured in very short time intervals. This is related to the Brownian movement and gives an indication of the particle shape depending on its diffusion characteristics. From DLS, the hydrodynamic radius R_H of the sample

,which takes the viscosity and the temperature into account, can be derived. The R_H is the radius obtained from a particle rotating in all directions including the hydration layer.

The intensity of how much light is scattered by a sample depends on the mass, the concentration and the refractive index (Folta-Stogniew 2006). This relationship enables to calculate the molar mass from the detected light scattering. In a SEC-MALS setup, the scattered light is detected by an array of detectors at discrete angles. This enables the measurement of the light scattering depending on the angle and the intensity. Furthermore, measuring at different angles ensures that the light scattering is detected at high precision.

The system is usually also coupled to a refractive index (dRI) detector which is sensitive to the sample concentration. The refractive index is a measure of the amount of light getting bent or refracted by entering from one medium into another (Folta-Stogniew 2006) which can be described by the Snell's law. The dRI detector is important because an aggregate that exists independently of the concentration would only give a small dRI signal but a strong light scattering signal at the same time.

In a SEC-MALS setup (Figure 16), the sample of interest is initially separated by a size exclusion column and then passes through a linear polarized laser at a constant flow. The light of the laser interacts with all media, even if they are transparent, as well as all molecules, which leads to a scattering of the light. Thus, the system needs to be carefully calibrated before the measurement to correct for the scattering caused by the running buffer. Each of the 18 detectors in the array measures the scattered light and its intensity at a discrete angle (θ) and extrapolates it back to zero degrees.

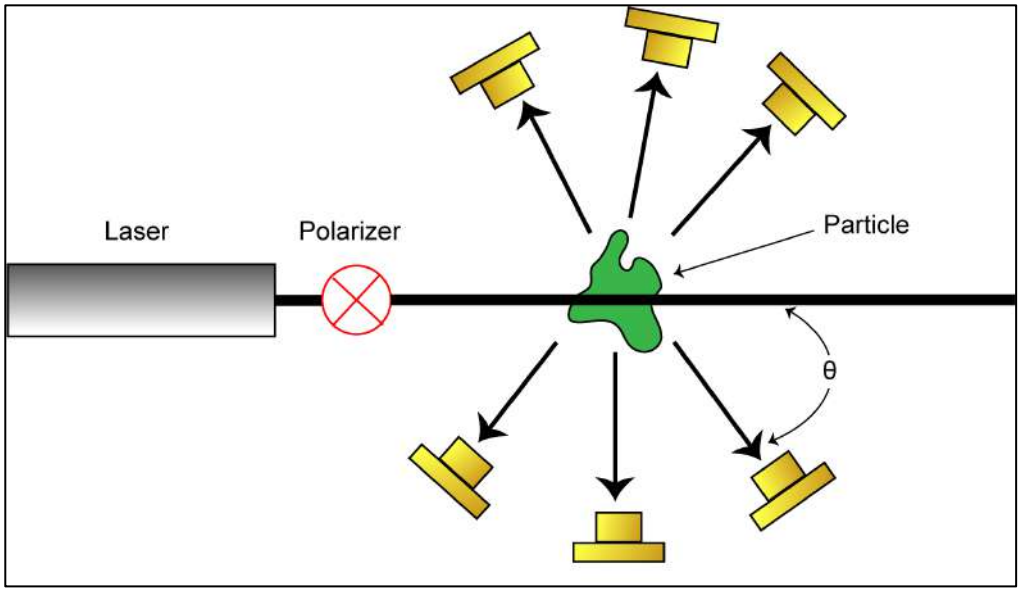


Figure 16: SEC-MALS setup.

The incident beam is polarized and scattered by the particle. Detectors at discrete angles detect the scattered beam.

For the DLS measurement, the R_H can be derived from the Stokes Einstein equation shown in Equation 2 by solving the equation for R_H. The Stokes Einstein equation describes the Brownian motion of a spherical particle containing the radius r through a fluid with the viscosity η at a constant temperature T (Minton 2016).

$$D_T = \frac{kT}{6\pi\eta R_H}$$

$$R_H = \frac{kT}{6\pi\eta D_T}$$

Equation 2: Stokes Einstein equation.

K = Boltzmann constant, T = temperature, r = spherical particle radius, η = viscosity, D_T = translational diffusion coefficient.

Experimental setup: The SEC-MALS analysis of apo yeast TFIIIC was performed at the Francis Crick Institute in cooperation with Dr. Ian Taylor.

A Superose 6 Increase 10/300 GL was used as the initial size exclusion column. 20 µl of sample were injected in the sample loop via an Hamilton syringe and loaded onto the column (running buffer: 200 mM NaCl, 20 mM Hepes pH 8, 0.5 mM TCEP, 3 mM NaN_3) at a constant flow of 0.5 ml/min. Before loading the sample, TFIIIC had to be dialyzed overnight to remove DTT and glycerol to prevent a strong background signal (dialysis buffer: 200 mM NaCl, 20 mM Hepes pH 8, 0.5 mM TCEP). Light scattering was measured using a DAWN Heleos-II light scattering instrument (Wyatt) and a differential refractive index was measured using an Optilab rEX instrument (Wyatt). DLS was coupled to the SEC-MALS machine. The analysis was done using the ASTRA Software from Wyatt.

The SEC-MALS analysis of crosslinked yeast TFIIIC was performed in-house using the same setup as described above. The human τA SEC-MALS analysis was also performed in-house but using a TSK G4000SWXL 300x7.8 mm column from Tosoh bioscience (running buffer: 500 mM NaCl, 10% glycerol, 20 mM Hepes pH 7.9, 1 mM DTT).

2.4.9. SEC-SAXS

Using small angle X-ray scattering (SAXS), macromolecular flexibility, shape, size and volume of a protein in solution as well as a low resolution (approx. 2 nm) reconstruction of a molecular envelope can be assessed (Rambo & Tainer 2011; Rambo & Tainer 2013; Jacques & Trewhella 2010). The *ab initio* rigid body modeling enables the assessment of the oligomeric state of proteins and protein complexes(Mertens & Svergun 2010). SAXS is traditionally used to obtain structural information about non-crystalline samples by using X-ray scattering (Feigin & Svergun 1987). Thereby, the protein solution is exposed to a focused and collimated X-ray beam (Figure 17A).
In SEC-SAXS experiment, the sample is separated via a size exclusion chromatography column before it is exposed to the X-ray beam. The

sample elutes continuously from the size exclusion column. Then, the total scattering intensity is plotted against the time. This generates a graph that is similar to a SEC UV measurement and is called a signal plot or SAXS chromatograph.

The scattering intensity of the dissolved monodisperse macromolecules is detected as a function of the scattering angle, at a detector that is placed at 2θ (angle between incident and scattering beam) from the incident beam (Mertens & Svergun 2010). The scattering of the particles in solution is dependent on the concentration and the scattering power of a protein, called the excess scattering length density. Thereby, each atom, surrounded by an electron cloud, serves as a scattering centre for the X-rays that propagate through space as a planar wave. At these scattering centers, the electromagnetic radiation gets scattered because in each of the electrons, polarization is induced which causes additional secondary smaller waves. Compared to the transmitted radiation, these waves are scattered (Jacques & Trewhella 2010; Mertens & Svergun 2010).

Measurement of the scattering intensity of the buffer is required and needs to be subtracted from the protein solution scattering since the total scattering (ρ) is a sum of all scattering events, including the scattering of the buffer and of the sample. The difference in excess scattering length density, also called contrast, $\Delta\rho(r)$ arises thus from the difference of the electron clouds between the sample and the solvent. Since the difference in electron density between the solvent and the protein is very small, a SAXS setup needs to be optimized carefully (Mertens & Svergun 2010; Putnam et al. 2007).

Scattering in solution occurs on randomly oriented molecules which ultimately limits the resolution of SAXS. After subtracting the solvent scattering from the protein-solvent scattering, the scattering intensity I(s) is directly related to the scattering of a single particle that is averaged over all orientations (Mertens & Svergun 2010). Besides that, the scattering is elastic which means that the oscillating electrons emit light of the same

wavelength. Furthermore, the scattering is coherent, leading to patterns of interference. These patterns carry information about the internal protein structure that allows to study the structure of the molecule.
The scattering is detected at two distinct positions: at low angles to detect the low resolution information and at higher angle scattering that contains more information about the structure of the protein (Jacques & Trewhella 2010; Mertens & Svergun 2010).
The scattering objects in SAXS are small compared to the wavelength which makes the scattering event isotropic. The scattering profile gets radially averaged and as a readout, a one-dimensional scatter curve is obtained (Mertens & Svergun 2010) (Figure 17B). The scattering vector (q) is usually converted to the momentum transfer as given (Liu & Zwart 2012):

$$q = \frac{4\pi \sin(\theta)}{\lambda}$$

Equation 3: Momentum Transfer (q) of the scattering vector.
2θ = angle between incident and scattered beam. λ = wavelength (Jacques & Trewhella 2010; Mertens & Svergun 2010)

In the scattering curve in Figure 17B, the logarithm of the scattering intensity $\log_{10}[I(q)]$ is plotted against the scattering vector q (unit: 1/Å) at different spatial frequencies and is obtained from Equation 3. The higher the value of q is, the higher is the resolution of the transferred information. It can be observed that the scattering curve becomes broader for higher resolutions, which is due to the increasing difficulty to distinguish the protein signal from the surrounding solvent (Rambo & Tainer 2011).
The Guinier analysis allows to calculate the radius of gyration R_g of a monodisperse solution of macromolecules. Thereby the linear region of the one-dimensional scattering curve, called the Guinier region, is extrapolated to the y intercept or the 0 angle called I(0) and fitted onto a straight line in a linear Guinier plot. The Guinier region is the first, lowest

resolution part of the curve. The gradient of the line is equal to Equation 4:

$$m = -\frac{R_g^2}{3} * q^2$$

Equation 4: Gradient of a Guinier plot.

R_g = radius of gyration, q = scattering vector, m = gradient.

The radius of gyration is the longest axis between the centre of mass and the end of the particle. R_g is thus larger for elongated proteins compared to compact proteins. Standards such as Bolvine Albumin Serum (BSA) are used to calibrate the measurement and the molecular mass can be estimated following the Guinier analysis (Mertens & Svergun 2010).

Information about the folding state can be obtained from the Kratky plot. In this plot, $q^2I(q)$ is plotted against q (Mertens & Svergun 2010). Compact, globular proteins tend to have a Gaussian shaped curve and very flexible or unfolded proteins tend to have a plateau at low q values. A protein containing both, compact regions and flexible regions will have a mixture of both curves (Mertens & Svergun 2010; Rambo & Tainer 2013).

Another piece of information that can be extracted from the one-dimensional scattering curve is the pair-distance distribution function P(r) which is a Fourier Transform of the 1D scatter curve. It enables to obtain more information about the shape of a protein by measuring the probability of finding a given intramolecular distance r P(r) by plotting it against the real-space intramolecular distance (Putnam et al. 2007 ;Mertens & Svergun 2010). This plot shows a real-space profile of the protein shape from which the maximum dimension, D_{max}, can be read from the x intercept. At this point, the probability of finding larger distances falls to zero. D_{max} can be used as a measure for the protein size.

It is important that P(r), which is closely related to the intensity profile I(q), has been estimated correctly for a later reconstruction of a 3D model (Liu & Zwart 2012).

Besides characterizing the protein with the analysis mentioned above, a low-resolution *ab initio* three-dimensional model can be calculated from the one-dimensional scatter. A popular method to calculate a bead-based model is the Dummy Atom Model Minimization (DAMMIN) (Svergun 1999). Thereby beads inside a constraint volume by D_{max} are randomly assigned as solvent or protein and a model is created in an iterative process which compares the computed scattering of the bead model at each stage to the true experimental profile (Figure 17C) (Mertens & Svergun 2010). In 2009, DAMMIN was optimized to operate slightly faster, creating DAMMIF in which F is referred to fast (Franke & Svergun 2009; Mertens & Svergun 2010). The speed increase comes from the fact that the random start or “seed” models must home all the protein beads connected. If this is not the case, then the seed is discarded from refinement in contrast to DAMMIN,

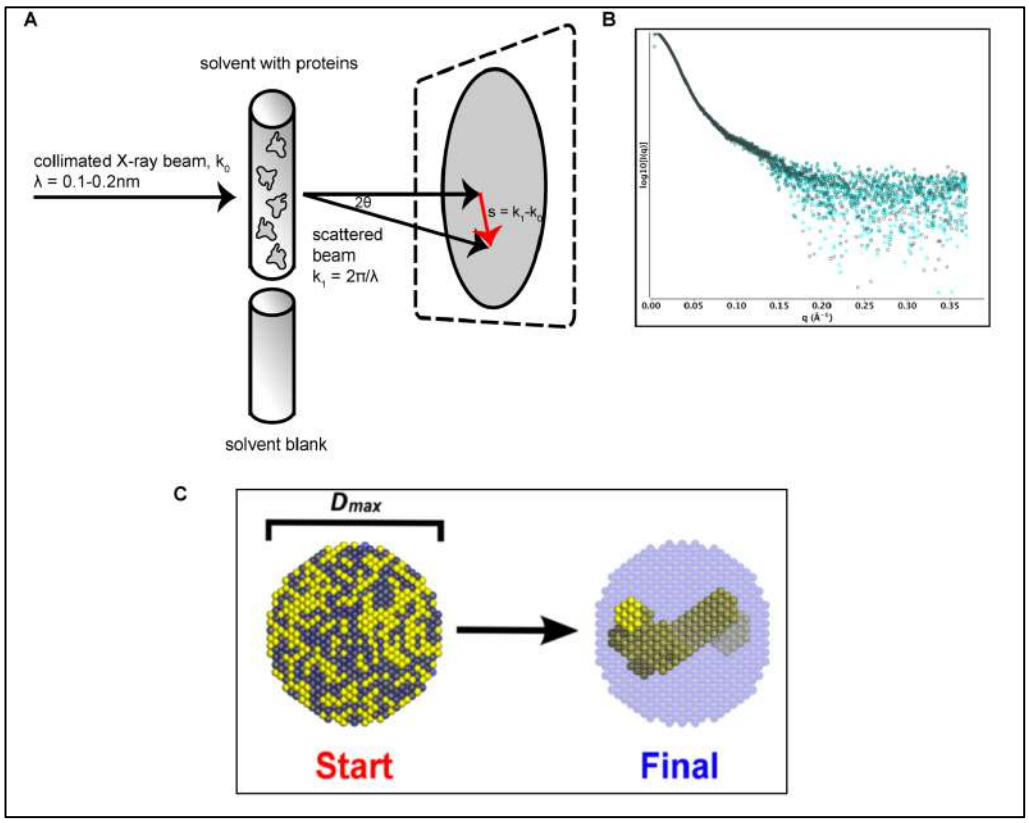


Figure 17: SAXS analysis.

A) Experimental setup. The focused and collimated X-ray beam (wave vector k0) gets exposed onto the solution containing the sample whereby the incident beam gets scattered (wave vector k1). 2θ is the angle between the incident and the scattered beam. λ = wavelength, s = scattering vector. B) Intensity plot of subtracted and merged SAXS frames from the leading fraction of 5 mg/ml yeast TFIIIC (Diamond Lightsource, Didcot) in 150 mM NaCl, 20mM Hepes pH 8, 56 mM Tris pH 8, 1 mM DTT. Black represents averaged buffer frames subtracted from averaged sampled frames. Cyan represents median buffer frames subtracted from averaged sampled frames. C) Ab initio modeling with DAMMIN. Spherical search volume containing particle beads (yellow) and solvent beads (blue) which are randomly assigned. D_{max} = maximum diameter, obtained experimentally. Adapted from (Mertens & Svergun 2010).

For the SAXS experiment conducted in this thesis, samples were analyzed at Beamline B21 at the Diamond light source national facility. Apo yeast TFIIIC (5.02 mg/ml) and yeast TFIIIC bound to the DNA (5.02 mg/ml, TFIIIC:DNA ratio 2:1) were tested in 150 mM NaCl, 20 mM Hepes pH 8, 56 mM Tris pH 8, 1 mM DTT). Analysis of the data was done using the JAVA based application Scatter, developed by Robert Rambo.

DAMMIF modelling was performed using the EMBL Hamburg online server.

2.4.10. Native mass spectrometry

The stoichiometry and interactions of protein complexes can be studied in the gas-phase using native mass spectrometry. The biological sample of interest is thereby sprayed from a non-denaturing solvent via electrospray ionization (Leney & Heck 2016). The term “native” relates to the fact that the native state of the protein is preserved prior to ionizing the sample. The detection of the sample however takes place in the vacuum of the mass analyzer (Leney & Heck 2016). Using this technique, molecular weight measurements are performed in gentle conditions to preserve the native structure inside the mass spectrometer (Zhou et al. 2012).

The two native mass spectrometry experiments explained in section 2.4.10.1 and 2.4.10.2 were conducted by Jack Roberts and Dr Bruno Bellina from Prof. Perdita Barran’s team at the Michael Barber Centre for Collaborative Mass Spectrometry at the University of Manchester. The nanoelectrospray ionisiation mass spectrometry experiment was performed using the Thermo Scientific Q Exactive UHMR Hybrid Quadrupole-Orbitrap mass spectrometer (Figure 18) and the ion mobility experiments were performed at Waters on a Cyclic IMS with the help of Jakub Ujma and at the University of Manchester using the Waters Synapt G2, that had been modified to possess a RF confining linear field drift cell.

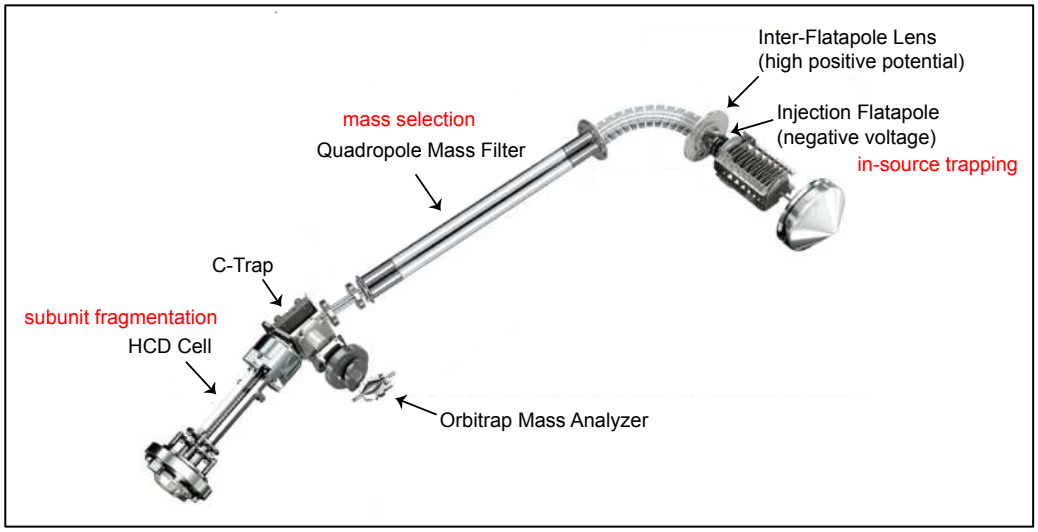


Figure 18: Schematic view of the Thermo Scientific Q Exactive UHMR Hybrid Quadrupole-Orbitrap mass spectrometer.

Adapted from the manufacturer's instructions.

2.4.10.1. Nanoelectrospray ionisation mass spectrometry (nano-ESI-MS)

Using this approach, apo TFIIIC was analyzed in two parts. Prior to ionization, 100 mM ammonium acetate at pH 7 was used as a non-denaturing solvent for TFIIIC. It is a very routine solvent/concentration/pH to use in native mass spectrometry since it mimics the physiological conditions of a cell.

In the first part, the in-source trapping (IST) voltage was increased from 0 until 300V in increments of 50V to detect if and how TFIIIC dissociates. At each of the voltage steps, a spectrum was taken and deconvoluted using the UniDec software (University of Oxford, UK), to assign masses to the precursor and fragment ions.

Subsequently, specific charge states of a mass were selected and the dissociation pathway of each of them was followed using a tandem mass spectrometry (MS/MS) analysis. The noncovalently linked

macromolecular complex is thereby introduced into the gas phase via electrospray ionization. A specific state can then be chosen which is directly dissociated into subcomplexes before being analyzed (VanAernum et al. 2019). The ion activation for the tandem mass spectrometry analysis was performed via collision-induced dissociation (CID) in which an inert gas is used as a collision target for the generated ions. Thereby, fragmentation of the sample into highly charged monomers and (n-1)-mers occurs. The stoichiometry and stability of the complex in the gas-phase can be analyzed (VanAernum et al. 2019).

In the performed experiments, a specific charge state was selected using the Quadrupole Mass Filter. The selected ions subsequently entered the HCD cell where a collision energy was provided to induce dissociation via CID, with the inert buffer gas (N_2). The fragment ions then entered the Orbitrap for mass analysis. The collision energy (CE) was increased in increments of 50 V from 0 to 300 V, to see how the dissociation of each of the selected charge states of the complex was happening.

2.4.10.2. Cyclic ion mobility analysis

For this approach, cyclic ion mobility (IM) was used in combination with a tandem mass spectrometry analysis. The advantage of this technique is, that conformational information of the gas-phase ions can be obtained which cannot be studied solely by using mass spectrometry. (Zhou et al. 2012).
In the experimental setup, the protein complexes get dissociated in the trap cell and are subsequently separated according to their size, shape and charge in the IM cell containing a neutral buffer gas. The separation of the ions occurs due to a series of electrical pulses and friction (Zhou et al. 2012). Thereby, the drift time of the ions is registered which is proportional to the size and charge state and serves as a physical

property reflecting the shape of the ions. This data can be used to calculate collisional cross sections (CCS) which serve as constraints for the molecular architecture (Zhou et al. 2012). After the IM step, the ions are subjected to a tandem mass spectrometry analysis and are analyzed in a mass analyzer to measure the mass to charge ratio.
The data was analyzed using MassLynx V4.1 software (Waters, UK) and Driftscope v2.9 (Waters, UK).

2.4.11. Western Blots

A Western Blot can selectively identify proteins. In this technique, an SDS-PAGE analysis (2.4.2) is performed to separate the proteins according to their molecular weight. Instead of staining the SDS-PAGE, the gel is blotted onto a membrane using a constant voltage. The membrane can selectively be incubated with antibodies that recognize and bind to specific protein epitopes. Antibodies can either be primary antibodies or horseradish peroxidase (HRP)-coupled antibodies. Primary antibodies recognize directly the epitope but need to be bound to a secondary antibody, that is coupled to a fluorophore, in order to be detectable by an Imaging system. On the other hand, coupled antibodies are covalently bound to HRP which can be detected using a Western Blot Kit and be imaged via exposure to an X-ray film.

Western Blot protocol:
The SDS-PAGE was run as described in to 2.4.2 and the proteins were transferred to a membrane using a XCell II BlotModule (Invitrogen) at 4°C for 1h at 100 V (or overnight at 10 V) in 1x Transfer Buffer (10% Methanol, 25 mM Tris Base, 192 mM Glycine) in a sandwich containing (from bottom to top): a sponge, two Whatman papers, the SDS-PAGE, the membrane (Amersham Protran Premium Blotting membrane, nitrocellulose 0.45, VWR), two Whatman papers and a sponge all soaked in cold Transfer Buffer. The membrane was blocked overnight at 4°C in 5% skimmed milk

(Sigma) in TBS-T (TBS containing 0.05% Tween) or for 1h at room temperature, then washed three times with TBS-T and incubated with the first antibody (Table 17) for either 2h at room temperature or overnight at 4°C in a 1:1000 dilution in 5% skimmed milk in TBS-T. For coupled primary antibodies, the membrane was then washed twice with PBS with 0.01% Tween and twice with PBS before staining it with the Pierce ECL Western Blotting Substrate (Thermo Fisher Scientific) according to the manufacturer's instructions. The Western Blot was read out at the Curix 60 X-Ray film processor (AFGA Healthcare).

Table 17: Antibody used for Western Blots.

Antibody name	Description	Supplier
StrepMAB-Classic	conjugated to horseradish peroxidase	Iba GmbH

2.5. Electron Microscopy

2.5.1. Historical and theoretical background

The first light microscope was invented in the 17th century by Antonj van Leeuwenhoek who was using a single magnifying lens to observe biological samples and by Robert Hooke who used a compound three lens system similar to modern light microscopes (Wollman et al. 2015). It took until the 19th century to develop modern light microscopes and to understand the theory behind the imaging.

In 1873, Ernst Abbe described the resolution limitation of a light microscope as half the wavelength of visible light (400 – 700 nm) (Equation 5) (Abbe 1873) . The term $n\ sin\theta$ is the numerical aperture (NA) of a microscope, indicating the ability of an optical system to focus light, thereby describing the resolution limitation. This means in a simplified

case, assuming the NA is 1, that visible light with a wavelength of 400 nm would lead to a maximum resolution of 200 nm.

$$d = \frac{\lambda}{2n\,sin\theta} = \frac{\lambda}{2NA}$$

Equation 5: Abbe diffraction limit.

d = minimum resolvable distance, λ = wavelength, n = refractive index of medium, θ = half angle of object (refer to Figure 19), NA = Numerical aperture.

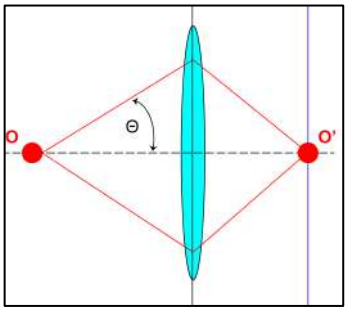


Figure 19: Numerical aperture (NA).

Image formation of object point "O". The outer light rays are maximum angle of light rays which can enter the lens and get diffracted. They can be used to determine the half angle θ of the object site which in turn can be used to determine the numerical aperture.

In order to overcome the resolution limitation caused by the wavelength of visible light, Ernst Ruska and Max Knoll developed the first transmission electron microscope (TEM) in 1931 which used electrons instead of photons from visible light as a source (Freundlich 1963). The use of electrons as a wave source was possible due to Louis-Victor de Broglie's discovery of the wave nature of electrons in 1924. He was able to explain that matter can have the same diffraction pattern as a beam of light. He defined the De Broglie relationship (Equation 6) in which the wavelength (λ) is associated with a particle containing the momentum (ρ) and the Planck constant (h) at $6.626*10^{-34}$ J.

$$\lambda = \frac{h}{\rho}$$

Equation 6: De Broglie relationship*. λ = wavelength, ρ = momentum, h = Planck constant.*

Since the momentum ρ is product of the mass (m) and the velocity (v) of a particle, the formula can be written as:

$$\lambda = \frac{h}{mv}$$

The velocity of electrons is determined by the accelerating voltage eV where e is the electric charge ($1.6*10^{-17}$ coulombs) and V the voltage, so that the velocity of electrons can be calculated with the following formula (Equation 7):

$$eV = \frac{1}{2}mv^2 \rightarrow \quad v = \sqrt{\frac{2eV}{m}}$$

Equation 7: Velocity of electrons.

e = electric charge, V = voltage, m = mass, v = velocity.

The wavelength of electrons at a given accelerating voltage can thus be calculated with Equation 8:

$$\lambda = \frac{h}{\sqrt{2meV}}$$

Equation 8: Wavelength of electrons at a given voltage.

λ = wavelength, h = Planck constant, m = mass, e = electric charge, V = voltage.

The effective wavelength of electrons is however slightly different due to relativistic effects on the electrons inside the microscope. Also, due to the objective lens system, the resolution is practically limited to approx. 0.8 Å for radiation insensitive samples (Ziegler et al. 2002).

2.5.2. Transmission electron microscopy

A transmission electron microscope (TEM) works under the same basic principles as an optical microscope in order to generate an image. In an optical microscope, an enlarged virtual image of the sample is generated through a series of lenses that diffract visible light. In a TEM, electrons are deflected instead of visible light through a system of electromagnetic lenses in an electromagnetic field. The TEM consists of an electron source, three electromagnetic lenses, each consisting of a coil of wire through which a current flows, and an image detecting system which nowadays usually is a charge- coupled device (CCD) camera or a direct detection device (DDD) (Orlova & Saibil 2011) (Figure 20).

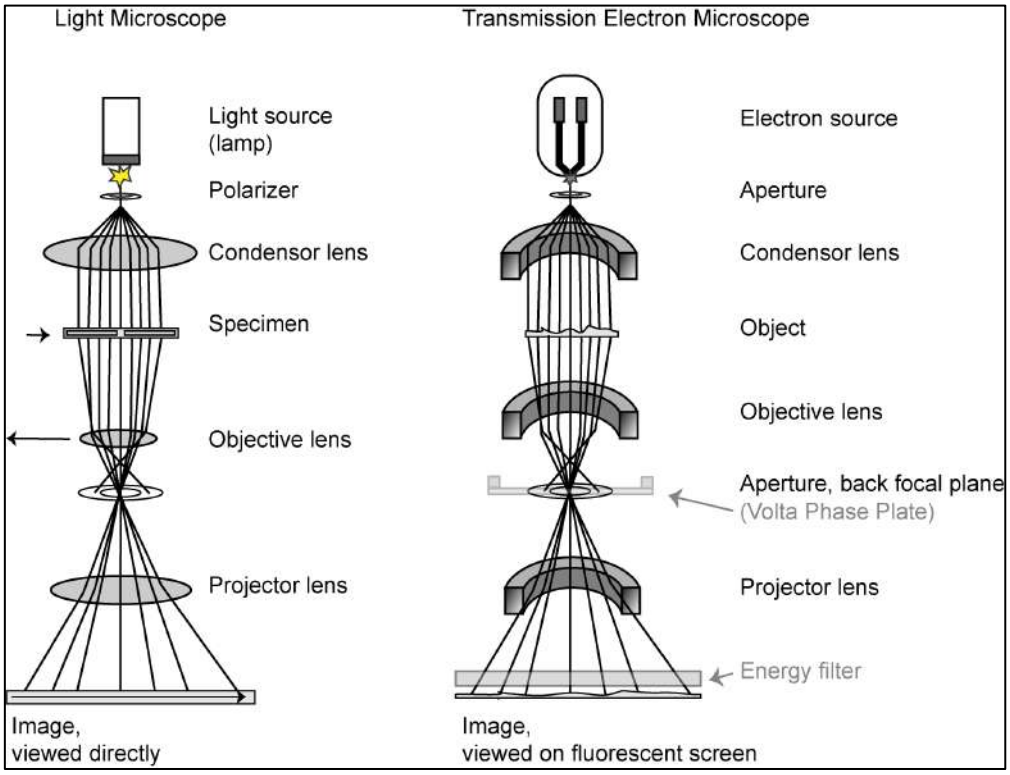


Figure 20: Schematic overview light microscope and TEM.
Adapted from (Orlova & Saibil 2011).

The electrons are emitted from the electron source which can be a tungsten filament at 2,000 - 3,000°C, a LaB_6 crystal or a field emission gun (FEG) with a tungsten filament as a cathode. The latter one produces a smaller, more coherent and 500x brighter electron beam (Spence 2013, p.275). The emitted electrons are accelerated by an anode at voltages of 100 – 300 keV and reach the condenser lens. The diverging electron beam is thereby converted into parallel electron beams which illuminate the sample, which is located in the middle of the objective lens in the magnetic field. In the back focal plane, an objective aperture is placed to increase the image contrast by preventing electrons, that were scattered at very high angles, to be used for image formation. The objective lens is capable of magnifying 20-50x, making it the most important lens of the electron microscope. Its lens aberrations also play an important role in imaging. After the objective lens, the image is further magnified by the projector lens before the electrons reach the detector (Orlova & Saibil 2011). A CCD detector converts the analogue optical signal in form of photon energy (light) into an electric signal by converting it into a small electric charge in a photo sensor. A scintillator converts the incoming electrons into photons to prevent irreversible damages to the photosensitive wells in the CCD. To gain a better signal-to-noise ratio (SNR), direct detection devices (DDD) such as the Falcon III EC from the FEI Company, the DE camera from Direct Electron or the K2 camera from Gatan Inc. are used. In contrast to the CCD cameras, the incoming electron signal is not converted into a light signal by a phosphor scintillator. This leads to a narrower distribution of the charges generated by the primary electron which in turn leads to a smaller point spread function (Wu et al. 2016). Thereby the detective quantum efficiency (DQE) of the detectors improves significantly. The DQE measures how the detector affects the SNR ratio. An improved DQE means that a lower dose is required to distinguish signal from noise (Clough et al. 2014). Over all spatial frequencies, a perfect detector would possess a DQE of 1 which

however is not practically achievable due to detector imperfections (Clough et al. 2014).

DDD allow to detect electrons at 200 or 300 keV at much better quality than film (McMullan et al. 2014) .These detectors record high frame rate movies of the exposure, made up of stacks of many frames, rather than images. This allows to correct for beam-induced specimen movement and image blurring (McMullan et al. 2016). DDD camera models can work in integration mode in which the different charges caused by the electrons on the sensor are integrated (Wu et al. 2016). In addition to that, the Gatan K2 camera and the Falcon III EC camera have been reported to be working in counting mode (Mooney et al. 2011; Song et al. 2019). Instead of integrating the signal of the primary electrons, each primary electron is counted and the center of the event is allocated to a quarter of a pixel. Single electrons can be counted and identified because a very high internal frame rate of 400 frames per second is used. The advantage of counting individual electrons is that the point spread function (PSF) is reduced to a single pixel and thus the DQE of the detector is significantly improved (Wu et al. 2016).

The lens system used in the electron microscope is not perfect and has some defects similar to the defects that occur in optical light microscopes such as spherical and chromatic aberration, curvature of the field, astigmatism and coma (Kohl & Reimer 2008; Orlova & Saibil 2011). Spherical aberration leads to a stronger refraction of rays further away from the optical axis which can be corrected for during image processing. A lens has chromatic aberration if it is able to focus rays with longer wavelengths better than rays with shorter wavelengths. This partly leads to an image formation in the plane closer to the object causing colored halos. Introducing an energy filter before the electron detector can remove these electrons during illumitation (Figure 20). A lens is astigmatic if it refracts beams stronger in one direction than in the perpendicular

direction leading to two different focal points. Astigmatism can be caused by an asymmetric magnet field. It can be corrected using stigmator coils.

2.5.3. Image formation and phase contrast

Depending on the sample composition and the electron energy, the interaction of the electrons with the specimen can vary (Hanszen 1971). Some of the interacting electrons are scattered without energy loss (elastic scattering) whereas other electrons are losing energy during the scattering (inelastic scattering) (Figure 21C). Inelastic scattering can occur through interaction with an outer shell electron or it can lead to an energy transfer onto the sample if the incident electron hits an inner shell electron. The energy transfer leads to the emission of X-Rays, chemical bond rearrangements, free radicals or secondary electron scattering. All of these processes lead to a change in the specimen structure, known as radiation damage. Due to these effects, the resolution is limited. Prolonged exposure of the sample to an electron beam can have similar radiation damage effects compared to an atomic explosion (Glaeser &. Taylor 1978). Other electrons however pass the sample without any interactions or get deflected by the electrostatic field generated by the nucleus. Alternatively, electrons can collide or almost collide with the atomic nuclei which leads to backscattering or high angle deflection (Orlova & Saibil 2011) (Figure 21).

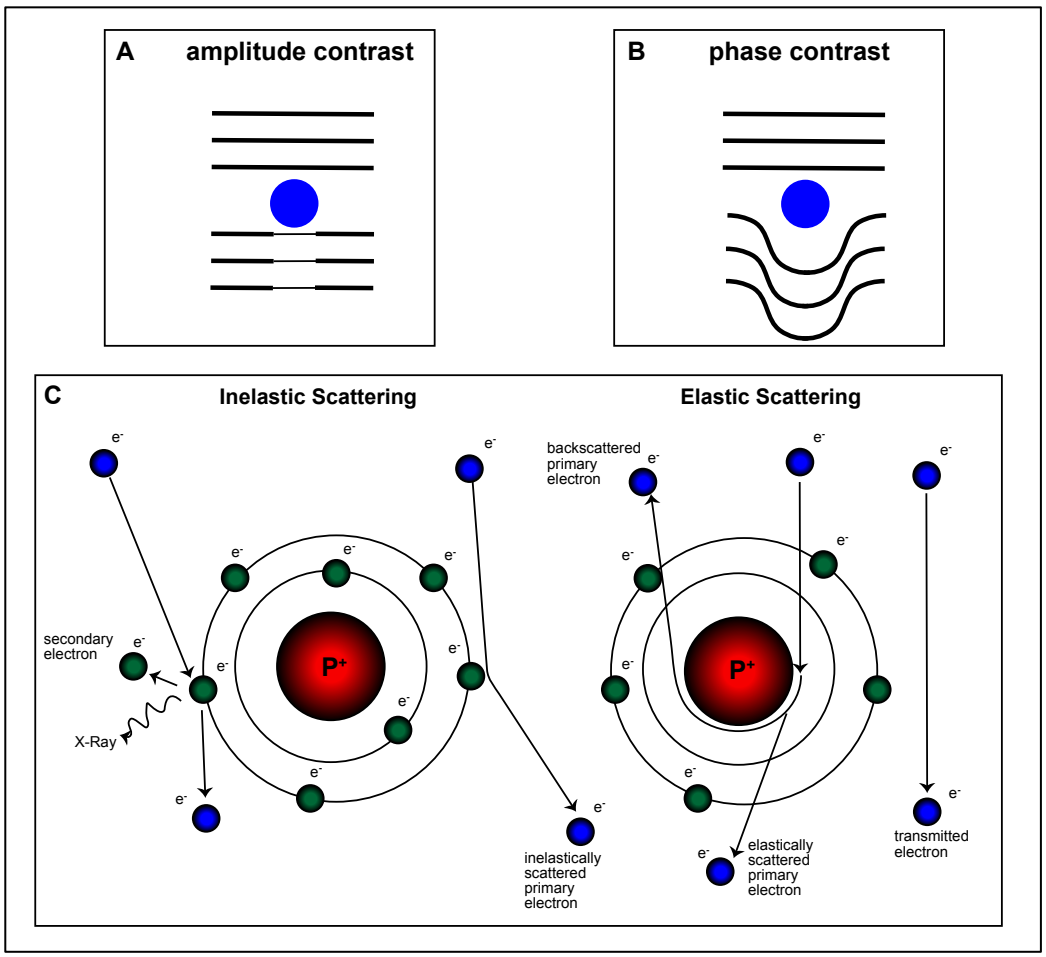


Figure 21: Contrast and scattering.

A: amplitude contrast. The particle absorbs part of the incoming beam. B: phase contrast: The particle changes the phase of the incoming beam. C: Inelastic and elastic scattering of the incident electrons with the specimen during illumination.

The image contrast in electron microscopy is an interplay between amplitude and phase contrast (Figure 21A, B). Amplitude contrast is formed by electrons which are partly absorbed by the specimen whereas phase contrast occurs when the specimen changes the phase of the incident beam. Biological specimen however contain light atoms such as H, O, N or C which deflect rather than absorb electrons via inelastic scattering and thus generate a very small amplitude contrast. When the incident beam interacts with the specimen, the exit wave of the beam gets changed due to this interaction. It can also be explained as a location

dependent phase shift. During this elastic scattering, the electrons get scattered at different angles depending on the distance of the scattering centers. This in turn extends the path lengths. The objective lens focusses both, the unscattered and scattered beams after interacting with the sample, leading to different path lengths which are phase shifted with respect to each other producing the phase contrast (Orlova & Saibil 2011).

The phase contrast can be visualized by combining the wavelength of the unscattered wave with the wavelength of the scattered wave to generate amplitude contrast. Thereby, the interference can either be constructive or destructive. This thus leads to an increase or decrease of the amplitude of the resulting wave which in turn leads to a positive or negative phase contrast.

Positive and negative contrast are alternating in a way that it creates a sinusoidal function, called the contrast transfer function (CTF) (Orlova & Saibil 2011) (Figure 22B). The effect of the CTF can be represented by a power spectrum or Fourier Transform of a micrograph obtained from a TEM. It shows concentric, alternating dark and light thon rings whereby each light ring represents a contrast transfer maximum and a dark ring indicates a spatial frequency without a signal (Figure 22A). Every time the sinusoidal function crosses the x axis, no information can be transferred. Due to the imperfections of the microscope, envelope functions cause the amplitude of the CTF to decay for higher spatial frequencies. These frequencies describe fine details in the image, whereas low spatial frequencies describe the overall shape. Defocusing the microscope changes the envelope function in a way that more low spatial frequency details are detected in exchange for losing high spatial frequency details (Cheng 2015).

During the data collection, one way to generate phase contrast is to take micrographs out of focus, called the defocus phase contrast (Danev et al. 2017). Indeed, a range of different defocus steps is collected to ensure that information at all spatial frequencies is collected (Figure 22C). The

lost information is restored by correction of the CTF via “phase flipping”. It inverts the negative amplitudes of the CTF into positive ones and restores the phase information (Figure 22D) (Orlova & Saibil 2011). In order to correct for amplitudes and phases of the CTF, a Wiener Filter is used. Thereby, the Fourier Transform of the image is multiplied by the CTF to correct for the phases. This value is subsequently divided by the CTF squared + 1/(SNR) to correct for the amplitudes. The term 1/(SNR) is added to the denominator in order to avoid dividing by values close to zero (Orlova & Saibil 2011).

Furthermore, imperfections of the microscope also cause a small dot to appear blurry rather than as a sharp dot when imaged in real space. The relationship between the dot and the blurred image is described as the PSF. A Fourier Transform of the PSF leads to the Contrast Transfer Function (CTF) (Orlova & Saibil 2011). The Fourier Transform of a spatial function is presented in reciprocal space.

Since the resulting phase shifts are very poor for biological specimen, the image contrast can be increased by defocusing, spherical aberration and substances that scatter more electrons such as uranyl acetate, (Erickson & Klug 1970; Orlova & Saibil 2011; Danev et al. 2017). Besides that, phase plates such as the Volta Phase Plate (VPP) can be added in the microscope (Danev & Baumeister 2016) (Figure 20). The VPP consist of a thin amorphous (10 nm) carbon film which is heated to 250°C. A negative Volta potential is generated when the beam hits the thin carbon film. The cause of this phenomenon has not been entirely understood yet (Danev et al. 2014). The Volta potential interacts directly with the incident electron beam and allows for a phase shift between the scattered and unscattered wave in the diffraction plane. Theoretically, the ideal phase shift is reported to be $\pi/2$, as the so-called Zernike phase contrast (Danev & Nagayama 2001) but has been found to give satisfactory results for a

variety of shifts $\frac{\pi}{4} \sim \frac{3\pi}{4}$ which equals 22.5 - 135° shifts (Danev & Nagayama 2011). This allows the acquisition of the image almost in focus whilst still retaining good image contrast (Danev et al. 2017). Over time, the shift in phase advances. To prevent the VPP potential to build up too high, the position on the VPP gets periodically adjusted, which is typically every 25 – 50 micrographs (Danev et al. 2017). Thereby, the zero point of the CTF shifts, thus making it possible to collect micrographs with only one defocus value.

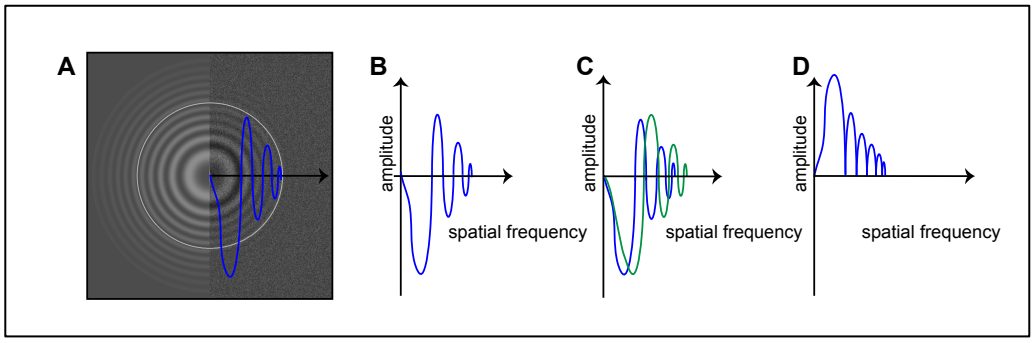


Figure 22: CTF curves.

A: Thon rings and respective CTF function. B: Uncorrected CTF curve from single defocus value. C: Uncorrected CTF curves from two different defocus values. D: Phase-flipped corrected CTF curve from single defocus value. Adapted from (Orlova & Saibil 2011).

2.5.4. Sample preparation

Imaging in an electron microscope requires a stable vacuum since electrons are able to interact with any particles in the air. It is thus important to stabilize the sample before inserting it into the vacuum. Currently there are two main techniques used for sample preparation: negative staining and cryo preparation of the sample. For negative staining, a heavy salt such as uranyl acetate is applied to the sample, staining the whole grid except the proteins (Brenner & Horne 1959). Thereby the negative imprint of the protein and not the protein itself is imaged after the heavy stain has been dried. This technique is easy to

use but it traps the sample in a non-physiological state due to flattening of the particle and dehydration (Orlova & Saibil 2011). Furthermore, the resolution is limited to about 15-20 Å due to the grain size of the heavy salt and the generation of artifacts.

Imaging a sample in the vacuum under native conditions is possible by freezing it under cryo conditions. A rapid freezing technique was developed in which the grid with the applied sample is rapidly plunge frozen in liquid ethane, cooled by liquid nitrogen, before being transferred and stored in liquid nitrogen (Adrian et al. 1984; Dubochet et al. 1988). Freezing the sample so quickly prevents the formation of ice crystals and traps the sample in a vitrified state. This allows imaging of the sample at high resolutions.

2.5.5. Single particle reconstruction workflow

After the sample has been imaged in a TEM and the images have been recorded, the processing of the data starts. Nowadays a variety of software is available to process the data such as Relion 3.0, EMAN2, cisTEM, cryoSPARC or IMAGIC.

Firstly, the recorded movies from the direct detection device are motion corrected by aligning and weighting the different movie frames. The motion corrected images are then CTF corrected by phase flipping as described in section 2.5.3. To restore the information in a more advanced way, amplitude distortions and phases can be corrected for by applying a Wiener Filter (Orlova & Saibil 2011).

After these initial steps, the individual particles can be picked using a pre-defined box size either manually or by picking algorithms such as the Gaussian picking in Relion 3.0 (Zivanov et al. 2018) or cisTEM (Grant et al. 2018). The individual particles are then extracted as a particle stack and aligned in order to increase the low SNR of individual images. Alignment is done by the cross-correlation function (CCF). It measures the consistency or dependency between two values and can thus

compare image similarity (Orlova & Saibil 2011). Images are subsequently normalized and compared to reference images that arise from peaks in the CCF function (Orlova & Saibil 2011).

When a 3D protein sample is imaged, the micrograph provides a snapshot of the individual particles which can have various orientations and conformations. Each of these views forms a separate 2D class in which particles with similar conformation and similar orientation are bundled to increase the SNR. This step is called reference free class averaging and represents a two-dimensional image of the protein structure (Orlova & Saibil 2011).

The classification of particles can either be performed by multivariate statistical analysis (MSA) as implemented in IMAGIC and EMAN (van Heel & Frank 1981) or by Maximum Likelihood Methods, as implemented in Relion and cyroSPARC (Sigworth 1998). Rather than maximizing CCF values between particles and references, maximum likelihood seeks to maximize the likelihood that the model represents the data by refining the class averages and 3D models. Particles are thus not immediately assigned to one class but during the comparison of the class averages and model projections, probabilities are calculated for all orientations of each particle, whilst applying optimized masking and filtering. The contribution a particle makes to a class gets weighted according to its probability. This method proved to be a robust method that is less likely to re-enforce noise (Sigworth et al. 2010).

2.5.6. Angle Assignment and 3D reconstruction

In order to reconstruct a three-dimensional model from the generated 2D classes, the relative orientation of the different classes needs to be known. Historically, this was done by using the common lines theorem. The theorem states that a pair of two 2D Fourier Transforms of the same particle in different orientations will have a common line between them in reciprocal space. As soon as at least three common lines between the

two Fourier Transforms are found, the orientation is fixed in 3D and angles can be assigned (Crowther et al. 1970). The assigned angles are called Euler angles and describe the orientation of a three-dimensional object in a fixed common centre (Orlova & Saibil 2011).

After angles are assigned, a 3D structure of the particle can be reconstructed. This can either be done in real space or in reciprocal space. In the latter one, the 3D model is calculated from 2D Fourier Transforms. Each of the 2D Transforms is a forward projection of the same 3D model at different angles. The central sections of the Fourier Transforms are built together into three dimensional space and the density of the 3D map in real space is calculated by an inverse Fourier Transform (Rosier & Klug 1968).

Modern processing software packages solely use the projection-matching approach. An initial model is generated either from a similar structure or can be generated de novo from EM images (Nogales & Scheres 2015). Throughout the processing, the 3D reconstruction is then refined in an iterative process; the orientation of each particle in the dataset is assigned by comparing it to generated 2D projections of the initial 3D reference volume. This determines the optimal orientation of each particle. Using these optimized angles, a more detailed reference volume, compared to the initial reference volume, is generated. In an iterative process, projections of this new reference map are compared again to the particles to further improve the angle assignment and to refine the 3D volume. Projection-matching can also be applied to several 3D reference volumes in one dataset, allowing to classify particles three-dimensionally in a process called 3D classification. Using this approach, specific conformations of a mixed dataset can be separated (Scheres et al. 2007). Thus, 3D classification is able to deal with discontinuous heterogeneity of a dataset. However, the limiting factor of 3D classifications is that the number of classes has to be determined before and that low populated states might not be found (Sigworth 2016).

2.5.7. Resolution estimation and postprocessing

After several rounds of 3D refinement, the resolution of the 3D reconstruction can be calculated using the Fourier Shell Correlation (FSC). It compares how consistent two models of the same dataset, containing half the particles each, are in spatial frequency using Fourier shells. These two models are not masked to prevent overfitting when calculating the FSC curves (Scheres 2016). Usually, 3D maps are consistent at low spatial frequencies that can be used for the overall shape of the molecule but decay at higher frequencies which represent finer details of the structure. The resolution can be estimated by reading the FSC value at a defined criterion. The most commonly used criterion is the FSC value of 0.143. It is used in the "gold standard" refinement in which two maps are refined individually with half of the particles using projection matching and are combined in the end (Rosenthal & Henderson 2003). This however gives a single resolution estimation for the whole map and does not account for local differences in resolution, in particular in larger, flexible macromolecular complexes.

Not using masks for the FSC usually results in an underestimation of the resolution. Thus, in the final postprocessing step, a mask around the 3D reconstruction is required. By using the mask, the noise is reduced and the FSC values will increase. This however also leads to convolution effects that can be corrected for by using phase-randomization in which the phases of the Fourier components above a resolution threshold are randomized (Chen et al. 2013). In Relion, phase-randomization is implemented together with B-factor sharpening during postprocessing. Whilst sharpening, the map is multiplied with a negative B factor. This is necessary to restore high frequency information that is damped throughout the image formation and processing. The sharpened map is then multiplied with the corrected FSC curve in order to obtain high-resolution information (Scheres 2016).

2.5.8. Particle subtraction

During single particle processing, unwanted densities inside particles can be subtracted from the rest of the particle. Therefore, all particles need to be aligned to a common 3D map first. A mask around the subunits of interest is designed in Chimera which is then applied to the whole 3D reconstruction, resulting in a 3D map containing a reduced density without the densities of interest. This reduced map is projected to the orientation of the particle stack, thereby generating a masked particle stack. In a subsequent step, the masked density is subtracted, resulting in particles only containing the densities of interest (Ilca et al. 2015).

2.5.9. Multi-body Refinement

A big challenge in analyzing the structure of macromolecular complexes via cryo-EM is continuous intrinsic flexibility inside a complex that cannot be classified by a conventional 3D classification. Relion 3.0 implemented a new tool, called multi-body refinement, which allows the analysis of continuous movement inside a complex. Traditionally it would lead to a blurry averaged reconstruction of that region (Nakane et al. 2018). Multi-body refinement allows to define up to three rigid bodies inside the complex which are generated by a consensus refinement. During a consensus refinement, all particles are aligned to the same 3D map in order to ensure all particles are in the correct orientation. The bodies are defined by customized masks in Chimera and individually refined via a focus refinement and the use of iteratively subtracted particles. The relative orientation of the bodies can be assigned by a principal component analysis that is used over all particles in the data set. A change of the orientation of the bodies towards each other is visualized in a volume series. Besides being able to visualize the flexibility of mobile regions inside a complex, it is possible to individually post-process the maps to gain a higher resolution map (Nakane et al. 2018).

2.6. Negative stain electron microscopy

Negative stain EM was used in this thesis to screen for adequate protein fractions and to assess their quality.

2.6.1. Grid preparation protocol

For the preparation of negative stain grids, copper Quantifoil grids (R1.2/1.3) coated with a carbon film (Quantifoil Microtool GmbH) were glow-discharged with the easiGlow Glow Discharge Cleaning System (PELCO) for 60 seconds at 15 mA. Thereby the grid surface becomes negatively charged and hydrophilic which allows an even spreading of the sample. After that, 3 µl of the sample were pipetted onto the grid and incubated for 1 minute. The excess liquid was removed with a filter paper and the grid was washed once in a drop of water. The water was soaked away with filter paper and the grid was tapped into a droplet of 2% uranyl acetate stain. The excess stain was removed with a filter paper and the grid was stained for an additional 30 seconds in 2% uranyl acetate. After that, the excess stain was removed whilst carefully ensuring that a thin layer of stain remained on the grid. The grid was air dried before it was placed into the electron microscope.

2.6.2. Negative stain screening and data collection

Negative stain grids were screened at the Tecnai T12 with a LaB_6 filament at an accelerating voltage of 120 keV with a 4K x 4K CCD F415 detector (Tietz). A magnification of 50,000x and an electron dose of 100e^-/Å^2 was used for image collection.

Large negative stain data collections were performed by Fabienne Beuron at the Tecnai F20 (FEI, ThermoFisher Scientific, USA) with a field emission gun and an accelerator voltage of 200 keV. The electrons were

detected with a 4k X 4k CCD F415 detector (Tietz) at a magnification of 50,000x and an electron dose of 100e-/Å^2. The resulting pixel size was 1.73Å. The automated collection was done with EM-Menu (TVIPS GmbH).

2.6.3. Sample preparation for negative stain data collections

2.6.3.1. Negative stain data collection of the human apo TFIIIC complex

Human His-TFIIIC-FLAG was purified according to its purification protocol described in section 2.3.2.4. After the gel filtration analysis (running buffer: 200 mM NaCl, 20 mM Hepes pH 8, 5% glycerol, 2 mM DTT) fraction D4 was chosen to be analyzed via negative stain.

2.6.3.2. Negative stain data collection of yeast TFIIIC bound to the DNA

400 μg (774 pmol) of yeast TFIIIC were incubated with a four times molar excess (3.1 nmol) of DNA (tD(GUC)K_noTATA_short) for 30 minutes at room temperature and topped up with interaction buffer to 500 μl in total (interaction and running buffer: 200mM NaCl, 20 mM Hepes, 2.5% glycerol, 2mM DTT). The excess of DNA was chosen to ensure that all TFIIIC complexes were bound to the DNA. By loading the TFIIIC-DNA complex onto a Superose 6 Increase 10/300 GL column at room temperature, the free DNA could be separated from the TFIIIC-DNA complex.

2.7. Cryo-electron microscopy

As mentioned previously, cryo-EM images the vitrified sample in its native state and thus overcomes resolution limitations imposed by negative stain EM.

2.7.1. Sample preparation

2.7.1.1. Yeast TFIIIC bound to the DNA

225 µg (434 pmol, 48 µl) of yeast TFIIIC were incubated with 217 pmol (3.1 µl) of DNA tD(GUC)K_noTATA_short (Table 15) at room temperature for 30 minutes. In the meantime, a fresh 2.3% Glutaraldehyde solution was prepared. After the incubation, 450 µl of interaction buffer (150 mM NaCl, 20 mM Hepes pH 8, 1 mM DTT) were added to dilute the sample in order to prevent aggregation during the crosslinking. The sample was immediately crosslinked by adding 6.62 µL of 2.3% Glutaraldehyde (0.03% final concentration) for 3 minutes at 37ºC and quenched by adding 50 mM Tris pH 8 (30 µl). The sample was centrifuged for 10 minutes and loaded onto a Superose 6 Increase 10/300 GL column with a constant flow rate of 0.3 ml/min. The interaction buffer was used as running buffer. Fractions were collected every 300 µl and fraction D6 was used for EM sample preparation of LJ012, LJ013, LJ014 and LJ015.

2.7.1.2. Apo yeast TFIIIC

225 µg (434 pmol, 39 µl) of yeast TFIIIC were mixed with 460 µl interaction buffer (150 mM NaCl, 20 mM Hepes pH 8, 1 mM DTT) and a fresh 2.3% Glutaraldehyde solution was prepared. The sample was immediately crosslinked by adding 6.62 µl of 2.3% Glutaraldehyde (0.03% final concentration) and quenched by adding 50 mM Tris pH 8 (30 µl). The

sample was centrifuged for 10 minutes and loaded onto a Superose 6 Increase 10/300 GL column with a constant flow rate of 0.3 ml/min and the interaction buffer used as running buffer. Fractions were collected every 300 µl and fraction D11 was used for EM sample preparation of LJ017.

2.7.2. Grid preparation protocol

2.7.2.1. Yeast TFIIIC bound to the DNA grid preparation

This protocol was used to produce grid samples LJ012, LJ013, LJ014 and LJ015. Quantifoil Copper R2/2 400 mesh (Quantifoil Microtool GmbH) grids with a freshly floated thin carbon layer were glow discharged for 20 seconds at 15 mA at the easi Glow Glow Discharge Cleaning System (PELCO). The grid was mounted inside a Vitroblot Mark IV (FEI) and 3 µl of the sample were pipetted onto the grid, incubated for 30 seconds, blotted for 3 seconds at a force of 5 and after a drain time of 0.5 seconds immediately plunge frozen in liquid ethane before being transferred into liquid nitrogen for storage.

2.7.2.2. Apo yeast TFIIIC unsupported grid preparation

For the grid preparation of sample LJ017, Gold Ultrafoil R2/2 200 mesh (Electron Microscopy Sciences) grids were glow discharged for 20 seconds at 15 mA at the easi Glow Glow Discharge Cleaning System (PELCO). The grid was mounted inside a Vitroblot Mark IV (FEI) and the sample was pipetted onto the grid in a total of three times according to the following scheme: Initially, 4 µl of the sample were pipetted onto the grid, after 30 seconds, 3 µl were removed from the grid and 3 µl were added from the sample, incubated for another 30 seconds, again 3 µl

were removed from the grid, 3 µl were added and finally 2 µl were removed, leaving a 2 µl thick wet film on the grid. The grid was incubated for another 30 seconds, blotted for 3 seconds at a force of 5 and after a drain time of 0.5 seconds immediately plunge frozen in liquid ethane before being transferred into liquid nitrogen for storage.

2.7.3. Cryo-EM Data collections

2.7.3.1. In-house cryo data collections

Yeast TFIIIC bound to the DNA (LJ012) was imaged in an in-house cryo data collection by Fabienne Beuron at the Tecnai F20 (FEI, ThermoFisher Scientific, USA) with a field emission gun and an accelerator voltage of 200kV. The electrons were detected with a 4k X 4k CCD F416 detector (Tietz) at a magnification of 50,000x and an electron dose of 37.8 e-/Å^2. The resulting pixel size was 1.73 Å and 430 micrographs were collected. The automated collection was done with EM-Menu (TVIPS GmbH).
582 micrographs of apo yeast TFIIIC (LJ017) were also imaged via an in-house cryo data collection using the same settings as above.

2.7.3.2. eBIC data collection (yeast TFIIIC + DNA)

The yeast TFIIIC bound to the DNA sample (LJ012) was imaged by Dr Kyle Dent at a 48-hour data collection in eBIC at Krios 3 at Diamond Light Source national facility. The FEI Titan Krios microscope was operated at 300 ekV connected to the Gatan Quantum K2 Summit detector with 3838 x 3710 pixels in integration mode. The pixel size was of 1.043 Å^2 , the defocus range was set from -1.7 to -3.5 µm in steps of 0.3 µm and a dose of electrons 80 e$^-$/Å^2 per second was used. Per hole, 5 shots were taken. The 3216 movies were motion corrected, aligned and dose weighted “on

the fly" using the Scipion pipeline MotionCor2 (la Rosa-Trevín et al. 2016) (Zheng et al. 2017).

2.7.3.3. 72h Leeds data collections (yeast TFIIIC bound to the DNA)

For the data collections at the Astbury Centre for Structural Biology in Leeds, yeast TFIIIC bound to the DNA was used (LJ014). The data was collected in parallel for 72h each on the FEI Titan Krios 1 and the Titan Krios 2.

The Titan Krios 1 was operated at 300 keV and equipped with a Falcon III EC camera (FEI) that was collecting in integration mode. The magnification of 130,000x lead to a pixel size of 1.065 $Å^2$,the defocus range was set from -1.6 to -3.1 µm in steps of 0.3 µm and a dose of 77.5 $e^-/Å^2$ per second was used. Per hole, 5 shots were taken. The 8547 movies were motion corrected, aligned and dose weighted "on the fly" using MotionCor2.

The Titan Krios 2 was operated at 300 keV and equipped with the Gatan energy filtered BioQuantum967 (K2 summit camera) and to a volta-potential phase plate (Danev et al. 2017), also operated in integration mode. The pixel size was of 1.065 $Å^2$ and the defocus was set to -0.5 µm to ensure that no over focus micrographs were recorded. The phase shift of the volta-potential plate increases over time with increasing accumulated dose (Danev & Baumeister 2016). In order to prevent the phase shift exceed much over $\frac{\pi}{2}$ (90º), the electron beam was moved periodically to a new position on the phase plate every 50 micrographs. A dose of 64 $e^-/Å^2$ per second was used. Per hole, 4 shots were taken. The 3167 movies were motion corrected, aligned and dose weighted "on the fly" using the MotionCor2 software inside the Relion 3.0 package.

2.7.3.4. 72h Leeds data collection of apo yeast TFIIIC

Data of the yeast apo TFIIIC (LJ017) was collected for 72h at the Astbury Centre for Structural Biology in Leeds. The Titan Krios 1 was operated at 300 keV and equipped with a Falcon III EC camera (FEI) operating in counting mode. The magnification was 75,000x which lead to a pixel size of 1.065 $Å^2$.The defocus was set to a range from -1.6 to -3.6 µm in increments of 0.4 µm. A total dose of 49 $e^-/Å^2$ per second was used. Per hole, 5 shots were taken. The 2682 movies were motion corrected, aligned and dose weighted “on the fly” using the MotionCor2 software integrated in Relion 3.0.

3. Results and Discussion

3.1. Human TFIIIC studies

This chapter describes the reconstitution and biochemical characterization of the human recombinant holo-TFIIIC complex. Furthermore, a low-resolution 3D reconstruction that was obtained from an in-house negative stain EM data collection will be discussed.
The human TFIIIC complex was initially divided into its two subcomplexes, τA and τB, to reduce the size of the complexes significantly. The aim was to characterize them individually, to understand their individual DNA-binding affinity better and to study their behavior in isolation before the holo-TFIIIC complex was reconstituted.

3.1.1. Human τA

To better understand the role of τA inside the TFIIIC complex, the main questions addressed in this section were to understand how τA behaves in isolation and whether it is capable to bind to the DNA without the presence of τB. Additionally the interactions of τA with other proteins were studied.

3.1.1.1. Expression and purification of human τA

τA, consisting of TFIIIC102, TFIIIC63 and TFIIIC35, was expressed to very high levels in High5 cells for three days. The whole subcomplex was purified via the N-terminal His_6 tag on subunit TFIIIC35 using a His Trap HP column. The His Trap HP was followed by a Hi Trap Heparin to remove DNA contamination and a HiLoad Superdex 200 16/600 pg gel filtration column as described in section 2.3.2. Although a TEV cleavage

site was inserted downstream of the His tag, it was not possible to cleave the His tag after the affinity step. A possible explanation for this could be that the cleavage site was shielded inside the protein complex, so that it was inaccessible for the TEV protease. After the purification, an additional analytical gel filtration, using a Superose 6 10/300 column was performed to better study the running profile of τA (Figure 23A). The remaining τA was pooled, concentrated to 1.6 mg/ml (total yield of 3.2 mg), flash frozen and stored at -80°C.

τA eluted in three distinct peaks from the Superose 6 10/300 column, with peak 1 eluting at 13 ml, peak 2 at 15 ml and peak 3 at 16.5 ml. All three protein bands, TFIIIC102, TFIIIC63 and TFIIIC35, could be detected in fractions from all peaks in the SDS-PAGE (Figure 23B). It was noticed, that the stoichiometry of the proteins changed throughout the peak. In peak 1 and 2 the subunits of τA appeared relatively stoichiometric, whereas in peak 3 TFIIIC102 was significantly reduced in comparison to TFIIIC63 and TFIIIC35.

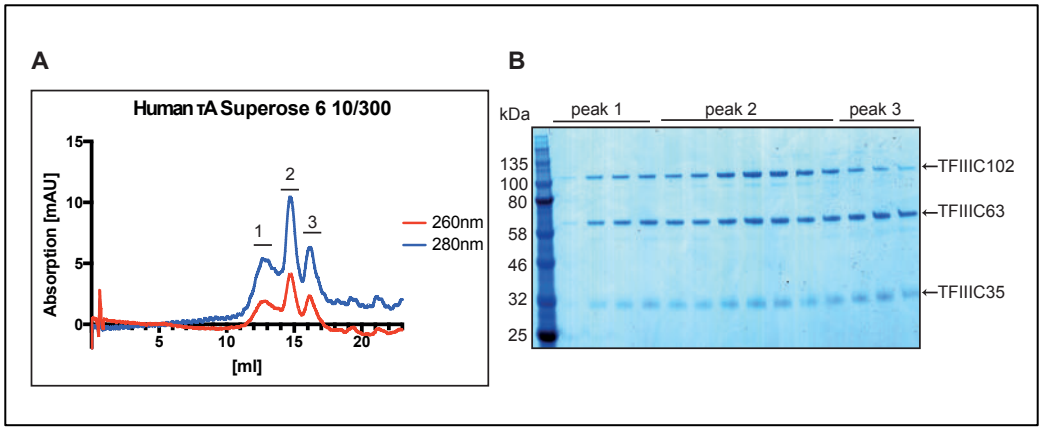


Figure 23: τA gel filtration elution profile after the purification.

A: τA chromatogram after elution from a Superose 6 10/300 column. τA elutes in three peaks. B: SDS-PAGE showing the three subunits of τA.

Comparing the elution volumes to a gel filtration standard (Bio-Rad) that was run on the same column in the same buffer (refer to section 5.1), it was observed that at 13 ml thioglobulin (670 kDa) and at 16.5 ml gamma

Globulin (155 kDa) eluted. The molecular weight of τA including the His tag accounted for 187.1 kDa. This led to the hypothesis that peak 2 probably contained stoichiometric τA whereas peak 3 could contain a mixture of stoichiometric τA and the TFIIIC35-TFIIIC63 subcomplex. Inside peak 1, an oligomeric state of τA could potentially be present, which was studied further in a SEC-MALS experiment described in section 3.1.1.3.

3.1.1.2. DNA binding studies of τA

τA was reported to be capable of binding to the A Box of the DNA (Gabrielsen et al. 1989; Bartholomew et al. 1990). It was unclear though whether τA was able to bind to the A Box in the absence of τB. Therefore, EMSAs were performed with four different DNA oligos that contained a different A Box sequence in the internal promoter region: TRNAM2 (A Box sequence: TGGCGCAGCGGA) and HVG2 (A Box sequence: TAGCTCAGCGG). From both constructs, short DNA oligos containing only the A Box and longer DNA constructs containing the A and the B Box were used (Table 15). The EMSAs were performed with self-made native PAGEs in 0.5x TBE buffer.

1.2 μM τA was incubated in a 1:1 ratio with the DNA for 30 minutes at room temperature using interaction buffer A (Table 16) before the samples were loaded. After staining the EMSA with ethidium bromide, no DNA binding of τA could be identified (data not shown). Increasing the ratio of τA to DNA up to 5:1 did not lead to a shift either (data not shown). Instead, protein aggregation or the formation of oligomeric states due to concentration effects inside the well could be observed that prevented the protein complex from entering the gel. Reducing the salt in the interaction buffer or using TGEM as a running buffer instead of TBE buffer did not lead to any DNA binding shift either (data not shown). The lack of DNA

binding could be due to τA's requirement for additional factors such as τB or TFIIIB to form a stable and strong DNA binding. This would be in accordance with previous papers in which it was reported that the yeast τA-DNA-binding affinity was weaker than the τB-DNA-binding (Stillman & Geiduschek 1984). Furthermore, it is important to note that only a limited amount of salt and running conditions could be tested during these studies. Potentially the binding could thus not have been captured in the tested setup. Additionally, the intrinsic setup of an EMSA experiment requires applying a voltage which could potentially disrupt the binding of a protein-DNA complex.

3.1.1.3. SEC-MALS analysis of τA

Next, τA was further characterized biophysically to understand whether it was present as a single heterotrimeric subcomplex or whether oligomers were formed. A SEC-MALS analysis was performed in order to better understand why τA could not enter into the native PAGE anymore when it was present in higher quantities. Due to the limitation of τA supply, the highest concentrations that could be analyzed were 1 mg/ml and 2 mg/ml respectively.

The SEC-MALS analysis revealed that at neither 1 mg/ml nor at 2 mg/ml, the expected molecular weight of τA (187.2 kDa) was measured (Figure 24). Both chromatograms, in which the light scattering and dRI signal were plotted (Figure 24A,D), showed a high background signal. This was potentially due to the amount of glycerol and the high salt present in the sample and running buffer. However, it was possible to identify two peaks that had a signal in the refractive index detector (dRI) and a signal in the light scattering, which are marked with light grey boxes (Figure 24A, D). For these peaks, the estimated molecular mass was plotted against the detected refractive index (n) (Figure 24B, E) and the light scattering signal

(Figure 24C, F). It was observed that at a concentration of 1 mg/ml, the detected molecular weight ranged from 600 kDa to 1 MDa and was thus significantly higher than the expected molecular weight. The fractions of these peaks were collected and loaded onto an SDS-PAGE that confirmed the presence of τA in the measured fractions. Increasing the concentration of τA from 1 mg/ml to 2 mg/ml led to an increase in the molecular weight from 750 kDa to 1.25 MDa. Both SEC-MALS measurements were repeated twice in the same setup, using the TSK G4000SWXL 300x7.8mm column (Tosoh bioscience).

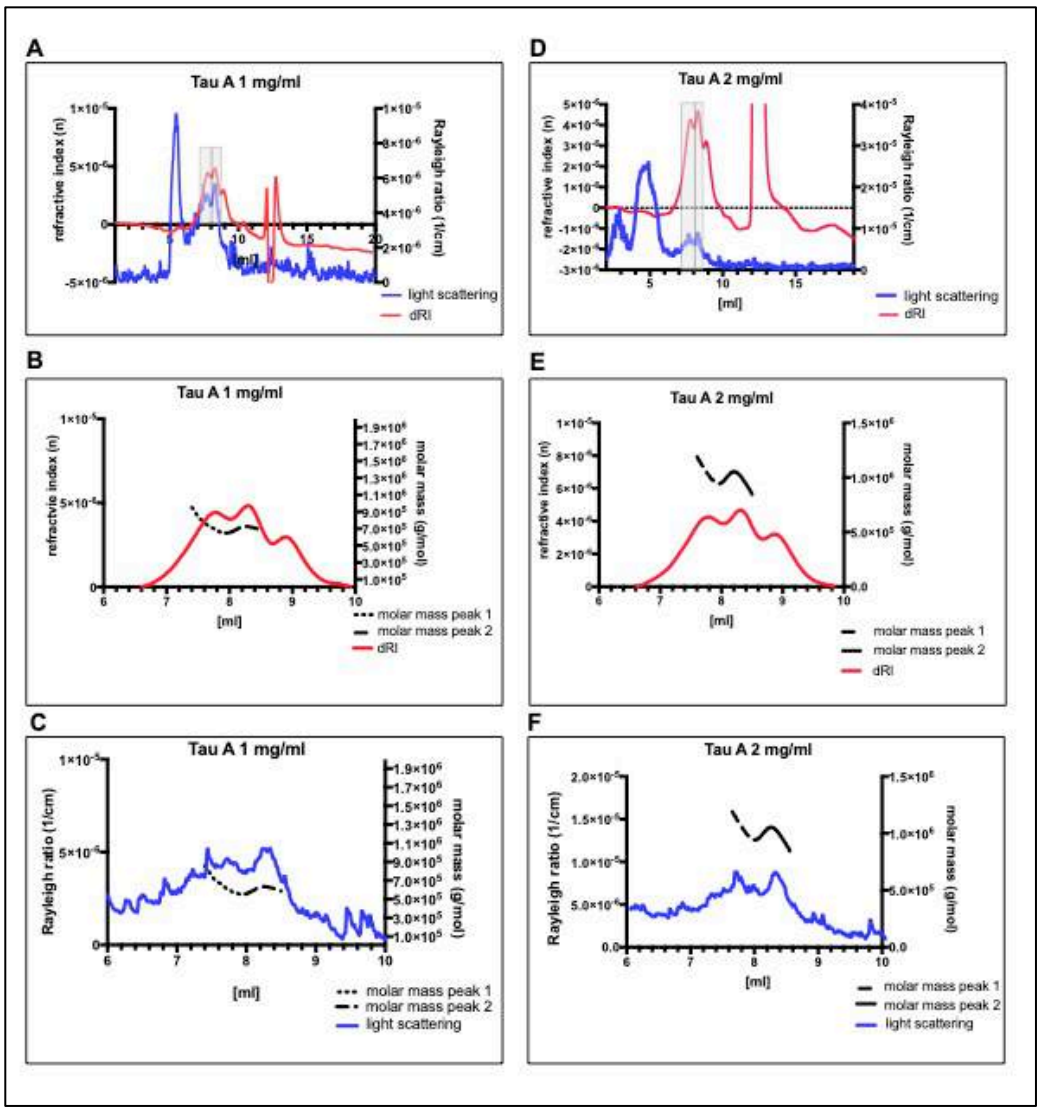


Figure 24: SEC-MALS results of τA.

A: Light scattering and dRI spectrum of τA at 1 mg/ml. The grey boxes indicate the two peaks where light scattering and the refractive index detector detected a signal. B: Signal plot showing the refractive index plotted against the elution volume and the molar mass of τA at 1 mg/ml. C: Signal plot showing the light scattering intensity plotted against the elution volume and the molar mass of τA at 1 mg/ml. D: Light scattering and dRI spectrum of τA at 2 mg/ml. E: Signal plot showing the refractive index plotted against the elution volume and the molar mass of τA at 2 mg/ml. The grey boxes indicate the two peaks where light scattering and the refractive index detector detected a signal. F: Signal plot showing the light scattering intensity plotted against the elution volume and the molar mass of τA at 2 mg/ml.

Due to technical issues it was not possible to conduct further SEC-MALS analysis. However, it would be beneficial to repeat the analysis using a

Superose 6 Increase 6 10/300 GL column to obtain a direct readout of the molecular weight of the peaks from the gel filtration column elution profile after the purification of τA.

The observed results led to the conclusion that τA was potentially forming multimeric states of four up to six τA molecules that were interacting with each other. These oligomers could have formed due to the high concentration of τA. Indeed, the detected molecular weight at 2 mg/ml was higher than the molecular weight detected at 1 mg/ml. Due to the limited sensitivity of the detector, no reliable readout could be obtained from a measurement at a lower concentration of 0.5 mg/ml to analyze whether the measured molecular weight at this concentration was lower than the one measured at 1 mg/ml. Furthermore, the lack of τB, as the natural binding partner of τA, could have caused τA to interact with itself. In the experiments described, a running buffer containing 500 mM NaCl and 10% glycerol was used. For confirmation, it would be important to repeat the SEC-MALS analysis using a lower salt buffer and less glycerol to reduce the background signal.

3.1.1.4. Negative stain analysis of τA

Following the SEC-MALS analysis, negative stain EM was used to further analyze τA and to investigate whether multimeric complexes were being formed. The peak fraction of peak 2 (Figure 23) was therefore stained with uranyl acetate and screened at the FEI T12 electron microscope.
The micrographs showed that the sample was heterogenous. Exemplary particles that could potentially represent τA are marked in green (Figure 25). Some particles appeared very elongated, others more globular. On average, the diameter was estimated to be approximately 15 -19 nm. The observed size of the particles could not confirm the hypothesis that τA forms multimers. However, it needs to be noted that the concentration

used for the preparation of the grid was 50 μg/ml and thus 20 – 40x lower than the concentration measured at the SEC-MALS. Oligomer formation could thus still be happening at higher concentrations. Due to the heterogeneity on the grid it was decided to not proceed with a data collection.

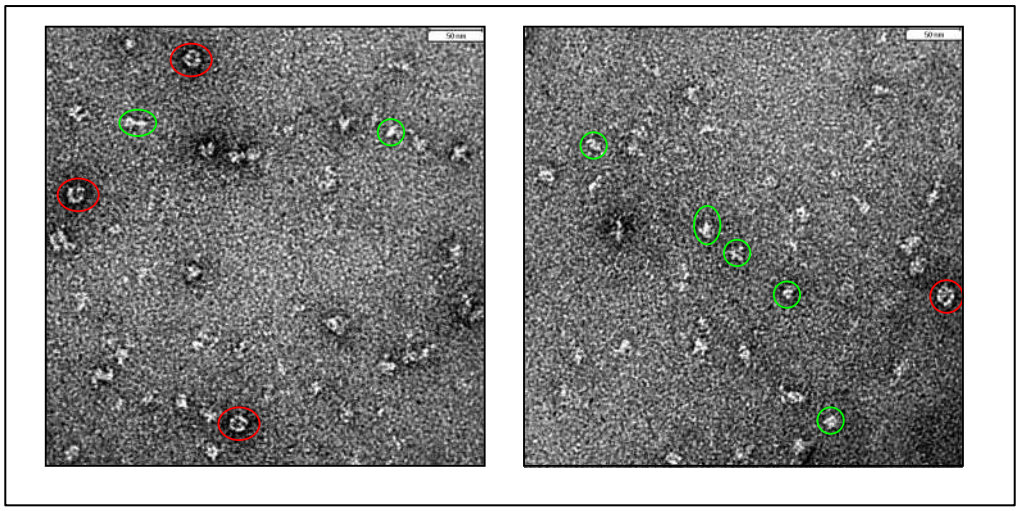


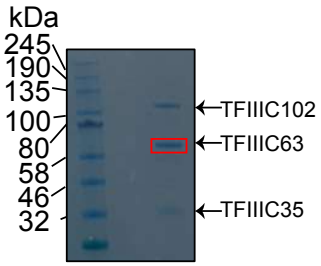
Figure 25: Negative stain micrographs of human τA.

Negative stain micrographs of τA taken at the FEI T12 microscope. Potential chaperones are marked in red and exemplary τA particles in green.

Another interesting observation on the micrographs was the presence of chaperones, which are marked in red (Figure 25). To validate their presence, the band running at the position where TFIIIC63 ran in the SDS-PAGE (Figure 26A, marked in red) was analyzed via LC-MS/MS. This band was analyzed because individual subunits of chaperones have an approximate size of 60 kDa and thus run at the same position as TFIIIC63 in the SDS-PAGE. The results of the mass spectrometry analysis (Figure 26B) revealed that in addition to TF3C5 and TF3C3, individual T-complex subunits that form the TCP1 complex, also known as the TCP1 ring complex (TRiC), are present. These results indicated that τA could contain unfolded or flexible regions that were exposed to the solvent in the absence of τB and were thus bound tightly to chaperones. This could also be an explanation for the higher molecular weight measured with SEC-MALS, since the TRiC complex accounts for a

molecular weight of almost 1 MDa. Attempts to remove the chaperones during the purification were unsuccessful so it was decided to focus on the expression of binding partners for τA instead of studying the subcomplex further in isolation.

A



B

identified proteins	protein score
TF3C5_HUMAN (General transcripation factor 3C polypeptide 5)	344
TF3C3_HUMAN (General transcripation factor 3C polypeptide 5)	16
TCPG_DROME (T complex protein 1 subunit gamma)	9
TCPH_BOVIN (T complex protein 1 subunit eta)	9
TCPA_DROME (T complex protein 1 subunit alpha)	7
TCPW_HUMAN (T complex protein 1 subunit zeta -2)	5
TCPE_SCHPO (T complex protein 1 subunit epsilon)	3
TCPQ_CAEBR (T complex protein 1 subunit theta)	3
TCPD_AEDTR (T complex protein 1 subunit delta)	2
TCPB_BOVIN (T complex protein 1 subunit beta)	2
TCPB_CAEEL (T complex protein 1 subunit beta)	2

Figure 26: LC-MS/MS analysis of τA.

A: SDS-PAGE indicating the band that was cut and analyzed via LC-MS/MS. B: Results of the inhouse LC-MS/MS analysis.

3.1.1.5. Purification of τA binding partners

The detection of chaperones in the τA sample led to the hypothesis that τA was lacking a binding partner. In order to stabilize τA, the idea was to purify reported binding partners; the three components of TFIIIB: TBP, Brf1 and Bdp1, as well as TFIIIA were identified as binding partners in addition to the natural binding partner τB. The latter will be discussed in chapter 3.1.2.

The TBP core (158-339), a C-terminal construct of Brf1 lacking the human specific extension (389-550) and Bdp1 (130-484) were expressed individually in *E. coli* and purified as described in section 2.3.1.1, 2.3.1.2

and 2.3.1.3. The gels of the purification are shown in Figure 27. All proteins could be purified in high yields.

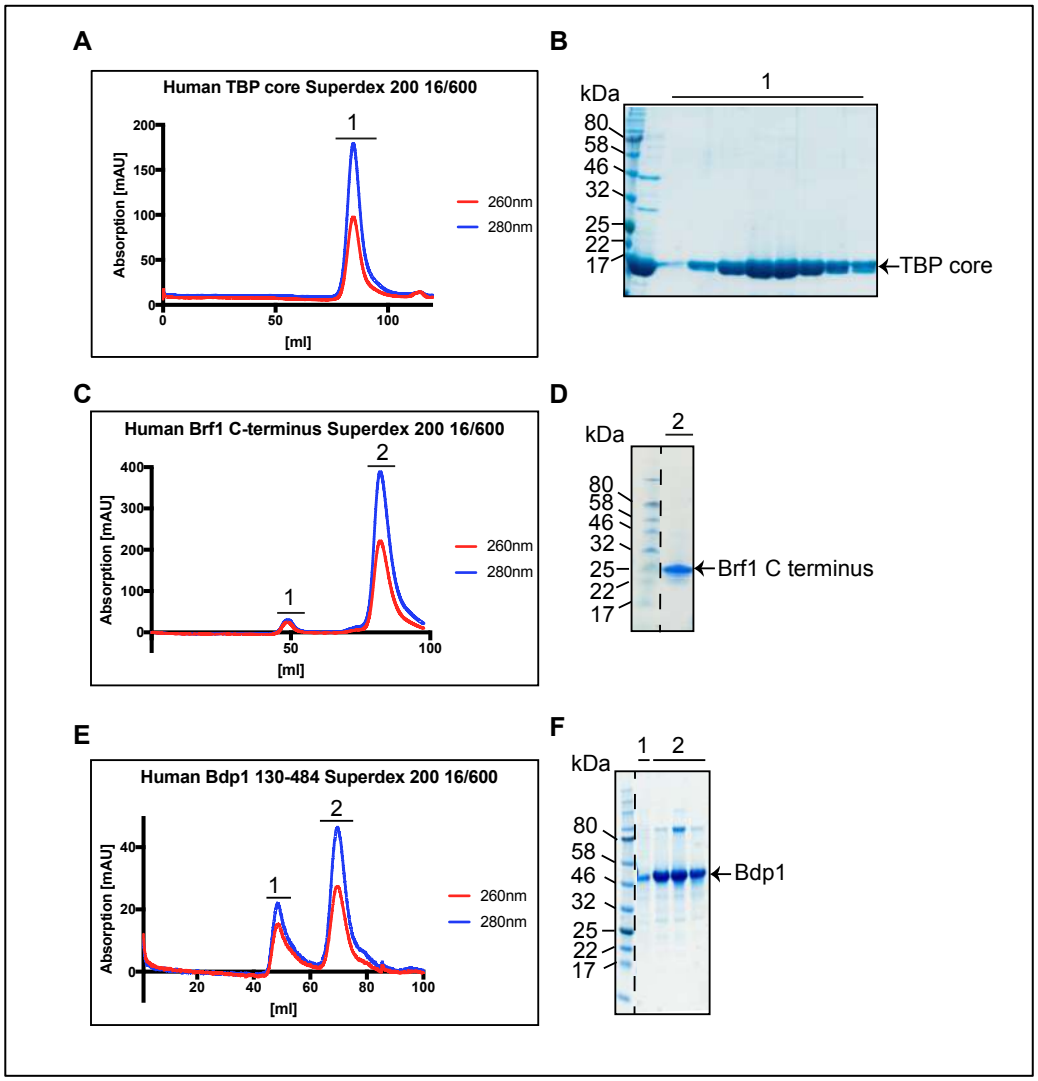


Figure 27: Purification of τA binding partners.

A: Chromatogram of the TBP core eluting from a Superdex 200 16/600, B: SDS-PAGE showing the peak fractions. C: Chromatogram of the C-terminus of Brf1 eluting from a Superdex 200 16/600. D: SDS-PAGE that shows the peak fraction. The dashed line indicates that the gel was cut. E: Chromatogram of Bdp1 130-484 eluting from a Superdex 200 16/600. F: SDS-PAGE with fractions from both peaks. The dashed line indicates that the gel was cut.

Previously, it has been reported that the N terminus of Brf1 was important for the interaction between TFIIIB and τA. Based on previous work in the lab, a variety of longer Brf1 constructs (Table 5) were designed but it was not possible to purify them in isolation.

Co-expressing the Brf1 constructs with TFIIIC components in High5 cells did not make it possible either to purify full-length Brf1 (data not shown). Hence, the only available human Brf1 construct that could be used for further studies was the C-terminal Brf1 region.

TFIIIA, a member of the zinc finger protein family, was identified as another potential binding partner. It functions as an affinity factor in type 1 promoters, which encode for 5S RNA (Wolffe & Brown 1988). For the expression of TFIIIA, two constructs were generated. The TFIIIA pLIB construct was used for expression in High5 cells (Table 6) and the TFIIIIA pOPINJ construct (Table 4) was used for expression in BL12 RIL *E. coli* cells. As shown in Figure 28, no TFIIIA expression could be observed in High5 cells. The His pulldown in lanes 4 and 8 did not show any TFIIIA either. In *E. coli*, TFIIIA could be expressed at good yields. However, it was not possible to remove DNA contamination during the purification process and was thus decided not to investigate TFIIIA further as a potential binding partner of τA (data not shown).

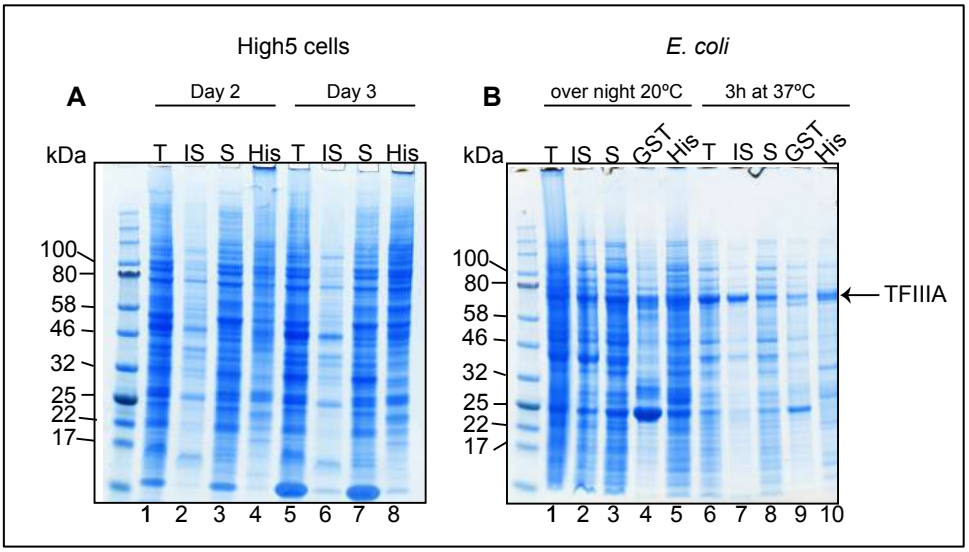


Figure 28: TFIIIA expression test.

A: Expression in insect cells (High5 cells). The cell lysis (T), the insoluble fraction (IS), the soluble fraction (S) and the fraction bound to the Ni NTA beads (His) are analyzed on the SDS Page after 2 and 3 days. In none of the fractions TFIIIA could be detected. B: Expression in E. coli (BL21 RIL cells). The cell lysis, the soluble, the insoluble, the fraction bound to the GST beads (GST) and the Ni NTA resin was analyzed on a SDS PAGE. In both induction conditions at 20°C overnight and 3h at 37°C expression of TFIIIA could be detected which is shown with a black arrow.

3.1.1.6. τA-TFIIIB interaction studies

The interaction of τA with the above-mentioned purified components of TFIIIB, the TBP core (158-339), Bdp1 (130-484) and the C-terminal fragment of Brf1 (389-550) was initially tested *in vitro* in a His-tag pulldown experiment taking advantage of the N-terminal His tag on TFIIIC35 of τA. None of the TFIIIB components contained an affinity tag after the purification. τA was incubated with a threefold molar excess of the potential binding partners using 50 µl of Ni NTA resin in both low (150 mM NaCl) and high salt (500 mM NaCl) conditions, incubated for 2h at 4°C, washed three times with the interaction buffer before analysis via SDS-PAGE.

Figure 29 shows the results of the His-Pulldown in low salt (Figure 29A) and high salt (Figure 29B). In Figure 29A, a weakly stained band of Bdp1 bound to the Ni NTA resin could be observed (lane 6, blue asterisk). This could arise from an unspecific interaction or it could be that the TEV protease did not cleave the His tag during the Bdp1 purification. The previously reported interaction between τA and Brf1 could not be confirmed (Figure 29A, lane 12). However, it needs to be taken into consideration that the Brf1 construct only contained the C-terminus of the protein and was lacking the N terminus that was reportedly important for the interaction (Kassavetis et al. 2001). Therefore, the small C-terminal part of Brf1 may not be sufficient to establish a stable interaction with τA. Adding TBP together with Brf1 to τA, did not improve the binding either (Figure 29A, lane 6). Interestingly, TBP, without the other components of TFIIIB, showed a strong interaction with τA (lane 13, red asterisk) which has not been reported before. To exclude nonspecific interactions, the His-Pulldown was repeated in high salt to increase the ionic strength (Figure 29B). In these conditions, the TBP band was very faint (Figure 29B, lane 13) and could not be easily detected in the SDS-PAGE.

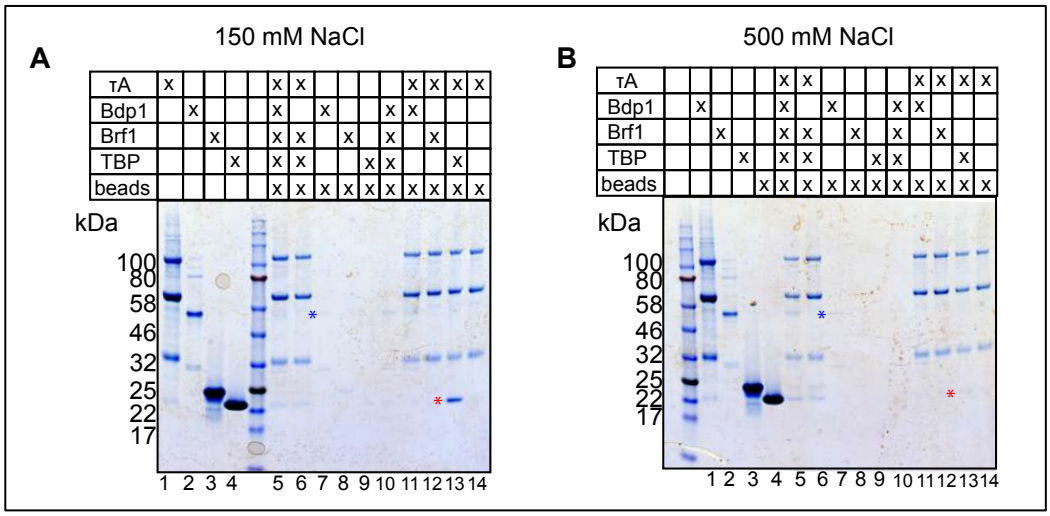


Figure 29: His-Pulldown between τA and TFIIIB components.

A: SDS-PAGE of the His Pulldown in low salt buffer (150 mM) stained with Coomassie Blue. B: SDS-PAGE of the His Pulldown in high salt (500 mM) stained with Coomassie Blue.

Staining the SDS-PAGE using Silver Stain (Figure 30) however revealed that TBP was detectable (lane 13, red asterisk). This demonstrated that high salt conditions can significantly reduce the τA-TBP interaction, but not disrupt it completely. Indeed, in the literature, most TFIIIB interactions were observed at low salt conditions *in vitro* (Kassavetis et al. 2001). Testing this His-Pulldown in lower salt conditions will thus be a promising experiment for future studies.

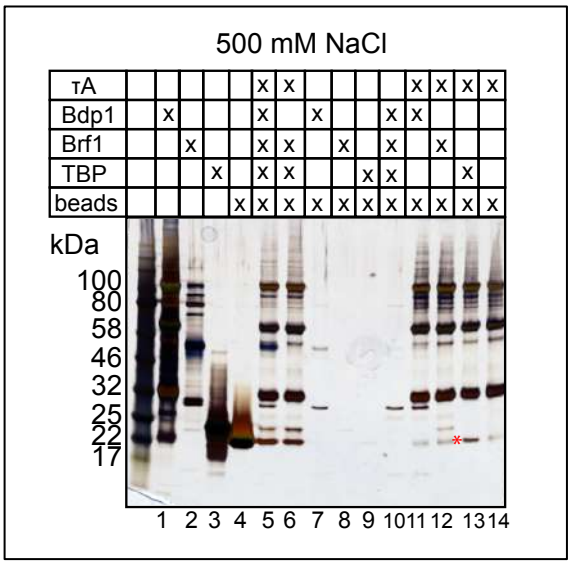


Figure 30: Silver Stain of the high salt His-Pulldown SDS-PAGE.
Lanes 1-4 served as running controls: proteins were loaded without Ni NTA beads. The blue asterisk marks Bdp1 band and red asterisk marks TBP band.

Studying the interactions further via an EMSA for example, was not possible because in the tested EMSA setups τA did not show a DNA binding shift whereas TBP did bind to the DNA in isolation. Incubating both proteins together before running the EMSA, did not lead to a super-shift (data not shown).

At this point, it was decided to investigate the expression and purification of τB as the natural binding partner of τA. This would allow reconstitution of the holo-TFIIIC complex, which could be used for further interaction studies with TFIIIB and to stabilize τA.

3.1.2. Human τB

In order to better understand the role of the τB complex inside TFIIIC, the main aim in this section was to express, purify and study τB in isolation. In particular, analyzing the stability of τB would allow us to understand whether τB could potentially be present without τA inside the cell.

3.1.2.1. Expression and purification of human τB

After τA was successfully purified, the purification of τB was required in order to reconstruct the whole TFIIIC complex and perform individual DNA binding studies of the two subcomplexes.
τB was expressed using the Multibac expression in High5 cells and contained the subunits TFIIIC220, TFIIIC110 and TFIIIC90. It could be observed that the expression of τB was low in comparison to the expression levels of τA. Furthermore, an expression test showed that τB was mostly insoluble (data not shown). A range of salt concentrations and buffers were screened to optimize the buffer conditions but this could not overcome the issue of insolubility.

The initial τB construct was expressed using the MultiBac system and contained a C-terminal Strep tag on TFIIIC110 (Table 5). At this position, the Strep Tag would not mediate the binding of τB to the Strep Trap HP column or the Strep resin. Our hypothesis was that the tag was at a position in the protein complex where it was either shielded or folded away inside a protein domain. Thus, the idea was to change the Strep-tag position to the C-terminus of TFIIIC90 using the In-Fusion Cloning Kit (2.1.4). Moving the tag to TFIIIC90 in the new construct (Table 5) made the tag more accessible, however the binding affinity to the Strep Trap HP column was still very low (data not shown).

The low binding affinity of the soluble protein subcomplex, together with the issue of insolubility, prompted us to test a Urea unfolding approach on the cell pellet (2.3.2) in order to recover the insoluble protein. The 8M urea containing buffer was exchanged by dialysis overnight and the sample was subsequently loaded onto a Heparin HP column. The three subunits of τB eluted together from the Heparin HP and were loaded onto a Superose 6 10/300 gel filtration column (Figure 31A). It was observed that the majority of the protein complex eluted in the void volume of the column (peak 1), which is at approximately 8 ml. Peak 2 contained stoichiometric τB of which all subunits were detectable in the SDS-PAGE (Figure 31B). However, this peak showed a high absorption at 260 nm which could have been caused by DNA contamination that was not removed during the purification. Peak 3 seemed to contain TFIIIC220, TFIIIC110 and TFIIIC90 but the stoichiometry of the complex was disrupted. It was thus concluded that purifying τB on its own was not the preferred strategy to with which proceed.

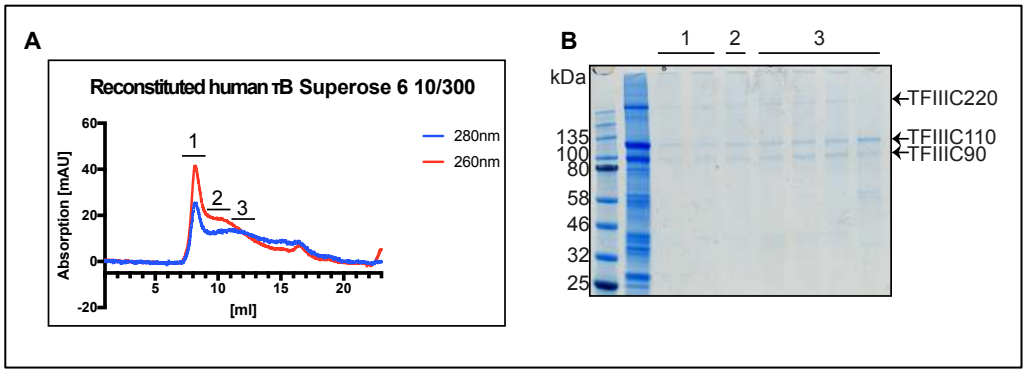


Figure 31: τB gel filtration running profile after the purification.

A: τB chromatogram after a Superose 6 10/300. τB eluted in three peaks. B: SDS-PAGE indicating the three bands of τB.

3.1.2.2. Purification of the human τB biGBac construct

During the cloning and purification attempts described above, Weissmann et al. published a paper describing the biGBac Cloning system, which was

reported to facilitate the assembly of multi-subunit constructs and to increase the level of protein expression (Weissmann et al. 2016). In addition to the low expression levels of τB using the MultiBac expression system, it was also noted that the baculovirus, generated with the MultiBac cloning system, was unstable over time. It was therefore decided to assemble a biGBac construct of τB in which the Strep tag was cloned to the C-terminus of TFIIIC220. This approach was used to increase the binding affinity of the tag that previously showed a low binding affinity at the C-terminus of TFIIIC90 and TFIIIC110. This could have been caused due to the Strep tag at those subunits being shielded or, alternatively, by degradation of the Strep tag. A Western Blot of the purified human τB and human TFIIIC samples, that were tagged with a Strep tag at TFIIIC90, was performed using an anti-Strep antibody to investigate this (Figure 32). The antibody was able to bind to both, TFIIIC and τB. This suggested that the Strep tag was expressed and accessible for the monoclonal anti-Strep antibody. However, it could not be excluded that the Strep tag at TFIIIC90 was potentially not accessible for binding to the Strep resin or a Strep Trap column due to steric hindrance. These results encouraged us to proceed with a construct containing the Strep tag at TFIIIC220.

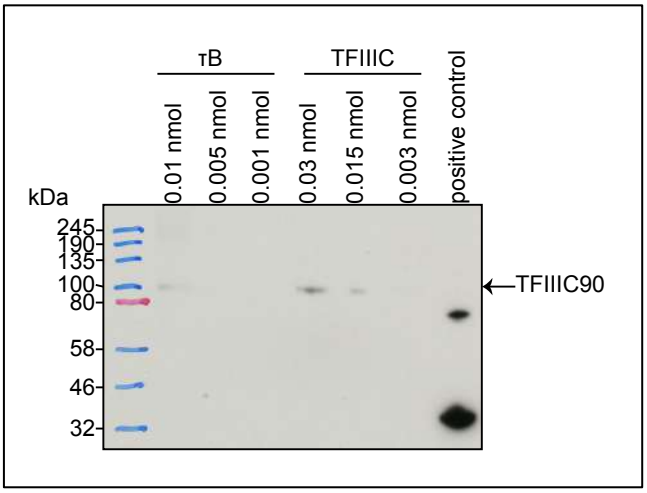


Figure 32: Western Blot of human τB-TFIIIC90-Strep and human TFIIIC-TFIIIC90-Strep.
Both samples were loaded in a concentration series together with 0.06 nmol of a positive control (PA28β-Strep, provided by Jennifer Ching, a fellow PhD student). The membrane was stained with StrepMAB-classic antibody (IBA GmbH).

The biGBac τB-Strep construct (Table 6) (Dr Erin Cutts) was used to generate a baculovirus that could potentially allow us to purify this τB construct from the soluble fraction. In contrast to the baculovirus which was cloned using the MultiBac system, it was possible to purify a small amount of τB using the biGBac baculovirus construct (42 μg, 882 nmol) (Figure 33). It was unclear whether this was due to the new cloning system or the new position of the Strep tag at TFIIIC220 but it was possible to store a small amount of τB after the Strep affinity step. The yield of τB after the Superose 6 Increase 10/300 GL column was however too low, meaning that the position of the peaks could not be determined precisely. Additionally, degradation was observed on the SDS-PAGE, which led to the decision to not proceeded further with purifications of the isolated human τB complex.

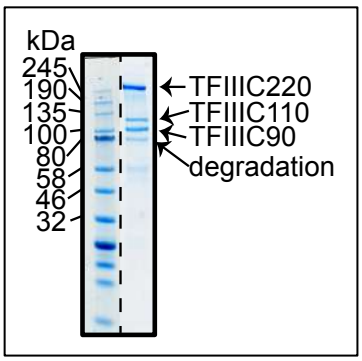


Figure 33: BiGBac human τB purification.

SDS-PAGE indicating the different subunits of human τB. The dashed line indicates that the gel was cut.

3.1.2.3. DNA binding studies of τB

The functionality of the purified biGBac τB subcomplex (Figure 33) was studied by testing its DNA-binding affinity. Therefore, 1 μmol of TRNAM2_ABox_BBox DNA was incubated with 1 μmol of τB using interaction buffer A (Table 16). τA was titrated in increasing amounts, using 2, 3, 4 and 5 μmol. The reaction was incubated for 30 minutes at room temperature before it was run on a 0.5x TBE acrylamide native PAGE that was stained with ethidium bromide (Figure 34).

In the EMSA, it was confirmed that τB was able to bind to the DNA (Figure 34, lane 5). The τB-DNA shift was clearly visible in all samples containing τB. In addition to the DNA binding, a significant amount of free DNA and some aggregation in the wells could be detected. This could indicate that the buffer or running conditions were not optimized or that τB aggregated during the freezing and thawing cycle after the purification.

Titrating τA over τB did not cause a shift of the τB -DNA band which indicated that there was no τA-τB-DNA interaction forming. Instead, only τB was capable to form a stable DNA interaction. As observed previously,

τA was not able to bind to the DNA on its own in the tested EMSA setup (Figure 34, lane 6).

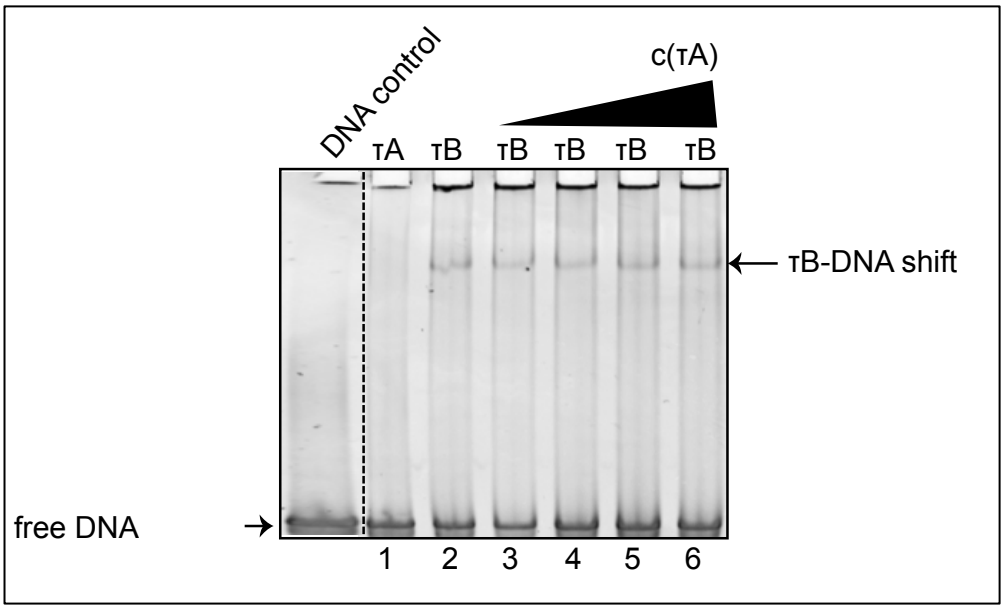


Figure 34: EMSA of human τB.

1 μmol of DNA was incubated with 1 μmol of τB and increasing amounts of τA (2, 3, 4, 5 μmol), incubated for 30 minutes at room temperature and loaded onto a 0.5xTBE acrylamide native PAGE stained with ethidium bromide.

In summary, τB was expressed at very low levels, it was very challenging to purify in isolation and it aggregated during the EMSA experiments. Taken together with the presence of chaperones after the purification of τA and its inability to bind to the A Box of the DNA, these results hinted to the fact that human τA and τB did not tolerate expression and purification in isolation very well. Hence, reconstitution of the holo-TFIIIC complex was attempted.

3.1.3. τA-τB interaction studies

The previous experiments suggested that τA and τB potentially function better together as a unit rather than as individual subcomplexes. This

could also mean that inside the cell, TFIIIC is mostly present as one complex. In order to reconstitute holo-TFIIIC *in vitro*, an analytical gel filtration of τA and τB, both independently expressed and purified, was performed. However, the performed experiments suggested that the two protein subcomplexes were only co-eluting instead of interacting, which would have caused a shift in the elution volume (data not shown). Therefore, the next aim was to reconstitute the holo-TFIIIC complex already at the expression level.

3.1.3.1. Purification of τA and τB co-infections

Initially, a co-purification of τA and τB pellets was tested. Both subcomplexes were expressed independently in High5 cells for three days, however no satisfactory result could be achieved using this approach. Only τA could be recovered whereas τB was mostly insoluble (data not shown).

Thus, High5 cells were co-infected with τA and τB viruses at different ratios. Purifying the co-infection of τA and τB gave a small yield of stoichiometric holo-TFIIIC complex. The sample was purified via a Hi Trap Heparin HP column, a His Trap HP affinity column and finally concentrated and loaded onto a Superose 6 Increase 10/300 GL column. The main peak eluted relatively close to the void volume (9 ml) (**Figure 35**A) and was shown to contain stoichiometric TFIIIC by analyzing the sample on an SDS PAGE (Figure 35B). This peak had a 260:280 ratio of 2:1, which implied that TFIIIC was either aggregated or contaminated with DNA. The following smaller peaks between 12 ml and 15 ml contained non-stoichiometric TFIIIC (data not shown). The presence of all six subunits of the reconstituted TFIIIC complex was confirmed via inhouse LC-MS/MS analysis, although TFIIIC35 was only detected at trace levels. Since the protein complex eluted very close to the void volume of the column and showed a high absorbance at 260 nm, we did not proceed further with this sample.

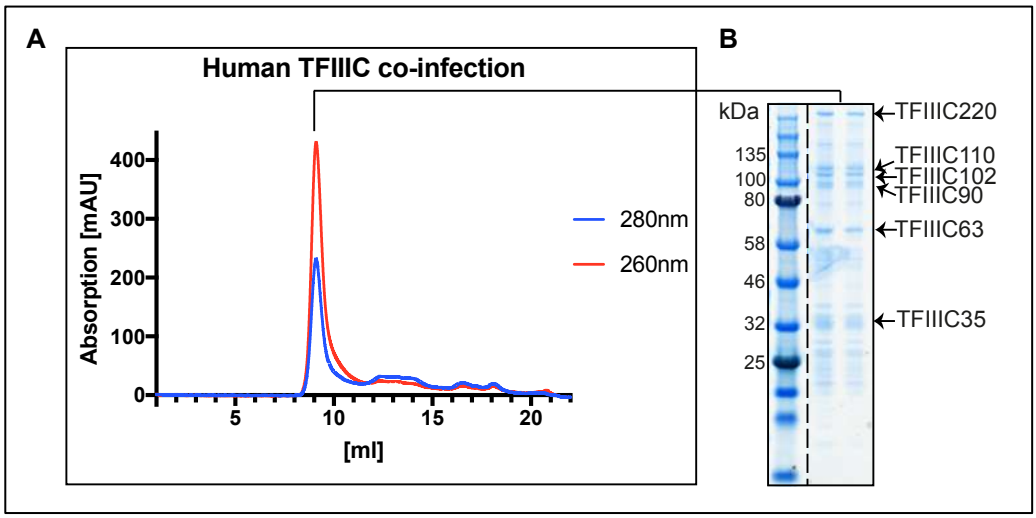


Figure 35: Gel filtration chromatogram of a co-infection of τA and τB.

A: Chromatogram after the Superose 6 Increase 10/300 GL elution. B: SDS-PAGE of human TFIIIC. The different subunits are shown. The dashed line indicates that the gel was cut.

3.1.4. Human holo-TFIIIC

By purifying the holo-TFIIIC complex, two main questions were aimed to address; firstly, how expressing holo-TFIIIC would change the DNA binding affinity and whether a binding to the A Box could be recovered and secondly, it would allow to structurally study TFIIIC to better understand the complex architecture.

3.1.4.1. Purification of the human holo-TFIIIC-Strep complex

As an alternative strategy to reconstitute the complex *in vitro*, the six TFIIIC subunits were cloned into a single DNA vector, which would ensure that all cells contain both, τA and τB subcomplexes. Previous attempts to assemble all six subunits into one pACEBac construct using the MultiBac cloning system were unsuccessful. This could potentially be overcome by switching to the biGBac cloning system. It was thought that all six subunits

could be cloned into a pBIG2ab vector with a His tag at the N terminus of TFIIIC35, as part of the τA subcomplex, and a Strep tag at the C-terminus of TFIIIC220, as part of the τB subcomplex. The presence of all six subunits of the holo His-TFIIIC-Strep construct (Dr Erin Cutts) was confirmed via control digest with SwaI and DNA sequencing.

After generation of a bacmid, the construct was expressed for three days in High5 cells. A purification protocol was established for TFIIIC consisting of three steps: a Strep-Trap HP column, a 1 ml Heparin column and a gel filtration column as described in section 2.3.2. TFIIIC had to be purified via the affinity tag on τB, since τA subunits were expressed at higher levels. Using the His tag on TFIIIC35 thus led to a non-stoichiometric TFIIIC complex. Using this protocol, it was possible to obtain a small amount of stoichiometric TFIIIC, however the expression and purification of this TFIIIC construct had major challenges. First, the expression in High5 cells was not scalable, making 500 ml culture of High5 cells the maximum volume in which TFIIIC could be expressed. Furthermore, a significant amount of TFIIIC did not bind to the Strep Trap column. The complex was found in the flow-through fraction (data not shown) but was able to bind to the Strep Trap column by re-loading it onto a new Strep Trap column. It was thought that the lack of binding was not due to a shielding of the Strep tag, as observed in previous constructs, but rather due to steric hinderance whilst binding to the column. Incubating TFIIIC with Strep-Tactin® Sepharose (IBA Lifesciences) for a few hours at 4 °C was not able to overcome this problem either. A chromatogram of holo-TFIIIC elution from a Superose 6 Increase 10/300 GL column is shown in Figure 36A.

The elution of TFIIIC occurred in three distinct peaks (Figure 36A). The first peak eluted at the column void volume. This suggested that it probably contained aggregated or higher oligomeric states of the protein

complexes. Peak 2, eluting at 12 ml, contained a stoichiometric TFIIIC complex that was detected in the SDS-PAGE (Figure 36B). The fractions of this peak were pooled and concentrated. This generated a total yield of 81 µg (130 pmol) from a 500 ml pellet of High5 cells. Peak 3, eluting at 13.5 ml, showed a non-stoichiometric TFIIIC complex in which a four-subunit core complex consisting of TFIIIC220, TFIIIC102, TFIIIC63 and TFIIIC35 was present. The TFIIIC110/TFIIIC90 subcomplex potentially dissociated from the holo-TFIIIC complex or was expressed at a lower level compared to the other subunits. A dissociation of the TFIIIC90/TFIIIC110 subcomplex from holo-TFIIIC could have been possible since it was reported to form a stable subcomplex (Mylona et al. 2006) and neither TFIIIC90 nor TFIIIC110 contained an affinity tag for the purification.

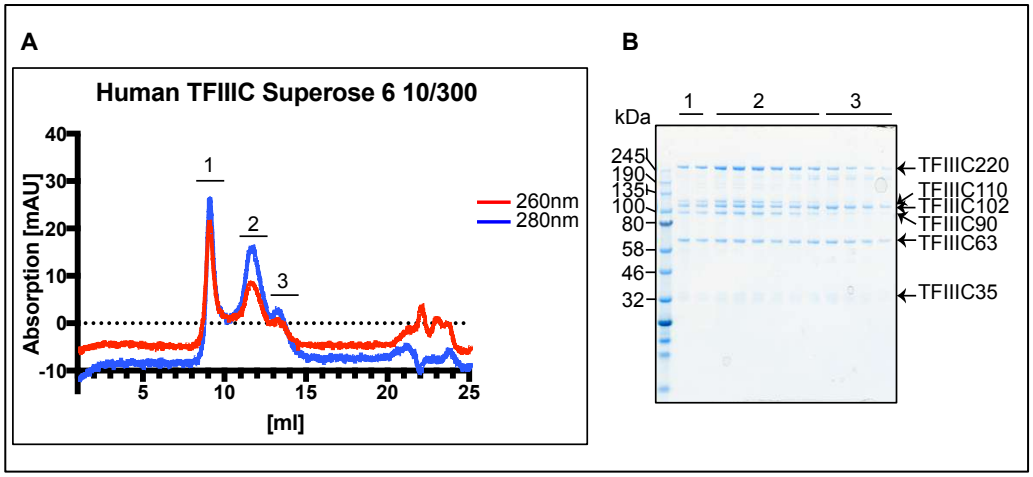


Figure 36: Human TFIIIC-Strep.

A: Gel filtration chromatogram after the three-step purification. B: SDS-PAGE indicating the six subunits: TFIIIC220, TFIIIC110, TFIIIC102, TFIIIC90, TFIIIC63 and TFIIIC35.

All six subunits of TFIIIC could be detected in the SDS-PAGE and were confirmed via an in-house LC-MS/MS analysis (Table 18). The smallest subunit, TF3C6 (TFIIIC35) was detected at lower levels compared to the other subunits which is due to the smaller size. On the SDS-PAGE,

TFIIIC35 appeared to run in a double band which could be due to different phosphorylation states. Indeed, a phosphorylated serine was identified in a mass spectrometry analysis.

Table 18: In-house LC-MS/MS analysis of holo-TFIIIC-Strep.

identified proteins	protein score
TF3C1_HUMAN (General transcription factor 3C polypeptide 1)	206
TF3C3_HUMAN (General transcription factor 3C polypeptide 3)	89
TF3C5_HUMAN (General transcription factor 3C polypeptide 5)	80
TF3C4_HUMAN (General transcription factor 3C polypeptide 4)	85
TF3C2_HUMAN (General transcription factor 3C polypeptide 2)	55
TF3C6_HUMAN (General transcription factor 3C polypeptide 6)	6

3.1.4.2. Purification of the human holo-TFIIIC-FLAG complex

Due to the low yield and low affinity of the Strep tag of human His-TFIIIC-Strep to the Strep Trap column, it was thought that trying a different affinity tag may help. It was important to choose a tag that provided high affinity and purity, similar to a Strep tag. A FLAG tag possesses hydrophilic characteristics. This allows the tag to be located at the protein surface, which makes it very easily accessible (Zhao et al. 2013). Using this tag, the risk of poor accessibility could be overcome. It was therefore decided to insert a 3xFLAG tag (Table 7) to the C-terminus of TFIIIC220. Using the FLAG tagged TFIIIC220, the six-subunit human TFIIIC construct was assembled using the biGBac cloning system. The pre-assembled His-τA construct (Dr Erin Cutts) was assembled with a TFIIIC90-TFIIIC110 construct and the TFIIIC220-FLAG construct. The presence of all six

subunits was confirmed via an analytical SwaI digestion (Figure 39) and DNA sequencing.

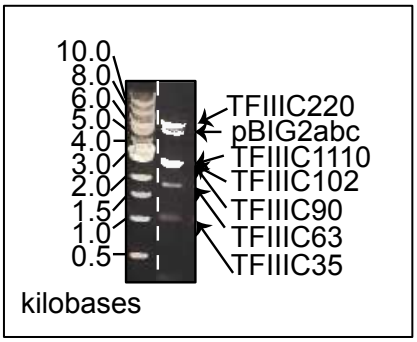


Figure 37: SwaI digest of holo-TFIIIC-FLAG.

After the SwaI digest, the DNA construct was loaded onto a 0.8% agarose gel stained with ethidium bromide. The dashed line indicates that the gel was cut.

The TFIIIC-FLAG construct was expressed in High5 cells for 3 days and purified via Anti-DYKDDDDK beads (Genscript). This was followed by a 1 ml Heparin column to remove DNA contamination and a subsequent concentration of the sample before it was loaded onto a gel filtration column using a Superose 6 Increase 10/300 GL.

As previously observed, TFIIIC eluted in three distinct peaks (Figure 38A) in which all six subunits could be identified in the SDS-PAGE (Figure 38B).

The spike observed after peak 3 occurred due to a re-start of the system and did not indicate a protein or DNA containing peak. As seen for the TFIIIC-Strep construct (Figure 36), the composition of the peaks indicated that the stoichiometric holo-TFIIIC complex eluted in peak 2 at 12 ml whereas a non-stoichiometric TFIIIC containing a four-subunit complex and a small amount of holo-TFIIIC eluted at 13 - 13.5 ml. TFIIIC35 ran as a double band again in the SDS-PAGE. The total yield of this purification after the gel filtration step was comparable to the yield obtained with the Strep-TFIIIC construct, however the amount of protein complex that was bound to the Anti-DYKDDDDK resin was significantly higher compared to the amount of protein that was bound to the Strep Trap column. Indeed,

it was of high purity and subsequently stored for further biochemical experiments immediately after the Anti-DYKDDDDK elution step.

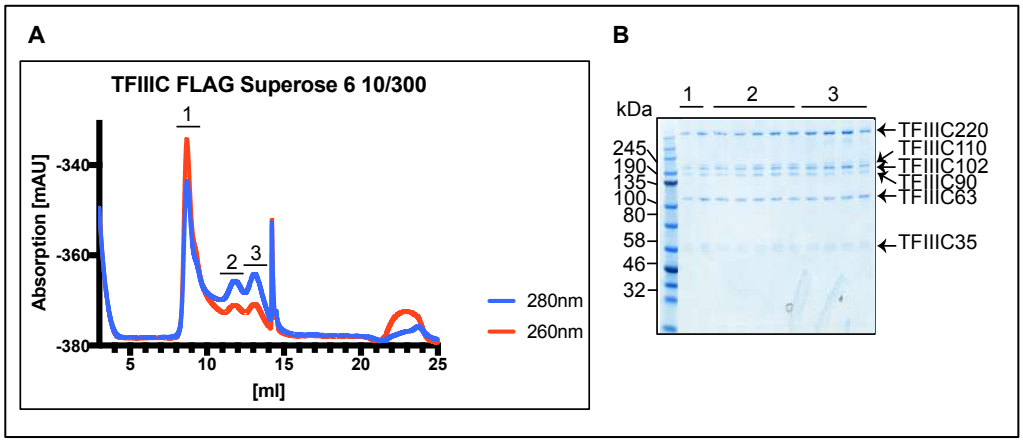


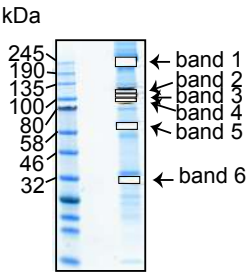
Figure 38: Holo-TFIIIC-FLAG purification.

A: Chromatogram of holo-TFIIIC-FLAG after elution from a Superose 6 Increase 10/300 GL column. B: SDS-PAGE showing different fractions from the three peaks of the chromatogram in A.

To identify the protein bands present in the SDS-PAGE, the composition of each was analyzed via LC-MS/MS at the UCL School of Pharmacy (Figure 39). Therefore, individual bands were cut out, digested with the tryptic in-gel digestion kit and sent to UCL. In each of the bands, several proteins could be identified. This was due to the high sensitivity of the mass spectrometer, cross-contamination whilst cutting out the individual bands and protein degradation. As expected, in band 1 mostly TF3C1 (TFIIIC220) was identified. In band 2, degradation products of TF3C1 were identified together with TF3C3 (TFIIIC102) and TF3C2 (TFIIIC110). It could be assumed that this band corresponded mainly to TFIIIC110 since slightly below, in band 3, TF3C3 (TFIIIC102) could be identified again together with TF3C5 (TFIIIC63). In band 4, TF3C4 (TFIIIC90) was identified with the highest protein score. Additionally, degradation and contamination products from other protein subunits could be detected. Band 5 was identified as TF3C5 (TFIIIC63). Finally, it could be assumed that band 6 accounted for TF3C6 (TFIIIC35) since it was the only band in

which this subunit was detected. As TFIIIC35 is the smallest subunit, degradation products from the other, larger proteins could be identified at the same position.

A



B

band	identified proteins	protein score
1	TF3C1_HUMAN (TFIIIC220)	3245
	TF3C3_HUMAN (TFIIIC102)	55
2	TF3C1_HUMAN (TFIIIC220)	823
	TF3C3_HUMAN (TFIIIC102)	471
	TF3C2_HUMAN (TFIIIC110)	373
	TF3C5_HUMAN (TFIIIC63)	51
3	TF3C3_HUMAN (TFIIIC102)	2575
	TF3C5_HUMAN (TFIIIC63)	167

band	identified proteins	protein score
4	TF3C4_HUMAN (TFIIIC90)	1935
	TF3C3_HUMAN (TFIIIC102)	580
	TF3C1_HUMAN (TFIIIC220)	151
	TF3C5_HUMAN (TFIIIC63)	72
5	TF3C5_HUMAN (TFIIIC63)	1592
	TF3C3_HUMAN (TFIIIC102)	101
6	TF3C3_HUMAN (TFIIIC102)	115
	TF3C6_HUMAN (TFIIIC35)	97
	TF3C1_HUMAN (TFIIIC220)	77
	TF3C5_HUMAN (TFIIIC63)	48

Figure 39: LC-MS/MS analysis of holo-TFIIIC-FLAG purification.

A: SDS-PAGE indicating the different bands that were cut out, digested and sent for LC-MS/MS analysis. B: LC-MS/MS analysis from the UCL School of Pharmacy.

3.1.4.3. DNA binding studies of TFIIIC

After the purification of holo-TFIIIC complex, its ability to bind the DNA *in vitro* was tested. Therefore, 1 μmol TFIIIC was incubated in a 1:1 ratio with the HVG2 and the TRNAM2 promoter for 30 minutes at room temperature using interaction buffer B (Table 16) to maintain a high ionic

strength. It was then loaded onto a 0.5xTBE 4% acrylamide Native PAGE that was incubated with ethidium bromide. In Figure 40, it could be seen that despite some aggregation present in the wells, TFIIIC was able to bind to both promoters (Figure 40, lanes 2 and 4). Indeed, TFIIIC mediated two DNA binding events, indicated by the two TFIIIC-DNA binding shifts (Figure 40, lanes 2 and 4). Based on these results, the difference between the two TFIIIC-DNA-binding events was not clear. One possibility could be the existence of different TFIIIIC conformations or different phosphorylation states. Alternatively, the presence of two bands could represent two DNA-binding states by a B-Box-only interaction as reported for the binding of TFIIIC to ETC sites (Nagarajavel et al. 2013) and a simultaneous binding to the A and B Box es, as expected on RNA Polymerase III transcribed genes.

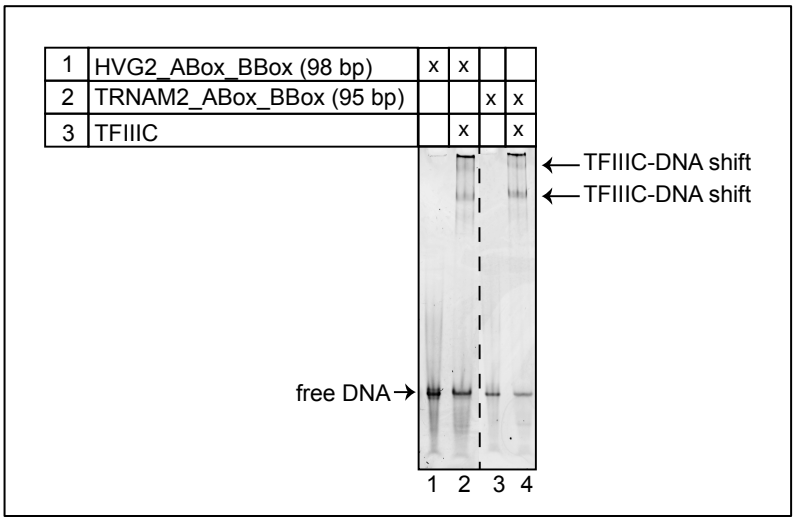


Figure 40: Holo-TFIIIC DNA-binding study.

Two different promoters were tested on a 0.5x TBE 4% acrylamide native PAGE stained with ethidium bromide.

To investigate on the latter hypothesis, the EMSA was repeated using a 52 bp long construct, that had a deleted A Box (TRNAM2_deleted_ABox) and a very short 39 bp construct that only contained the B Box and 24 bp upstream of the B Box but no A Box (TRNAM2_BBox_24bp_upstream) (Table 15, Figure 41A). As a control, the TRNAM2 promoter binding

performed in Figure 40 was used. It was thought that the two shorter constructs would cause a single DNA-binding shift whereas the two shifts caused by differential binding to the A and the B Box. However, instead of a single TFIIIC-DNA shift, all three constructs showed two DNA-binding events (Figure 41B). These observations showed that the two DNA-binding events of TFIIIC were caused independently of the A and B Box presence. As expected, it could be demonstrated that the presence of a B Box on the DNA was sufficient for the formation of a stable TFIIIC-DNA complex (lane 10), supporting the previous finding that the majority of the TFIIIC-DNA-binding affinity was mediated via binding to the B Box (Stillman & Geiduschek 1984). Furthermore, it was shown that TFIIIC can bind to extra TFIIIC sites (ETC) sites throughout the genome that exclusively contain a B Box (Nagarajavel et al. 2013).

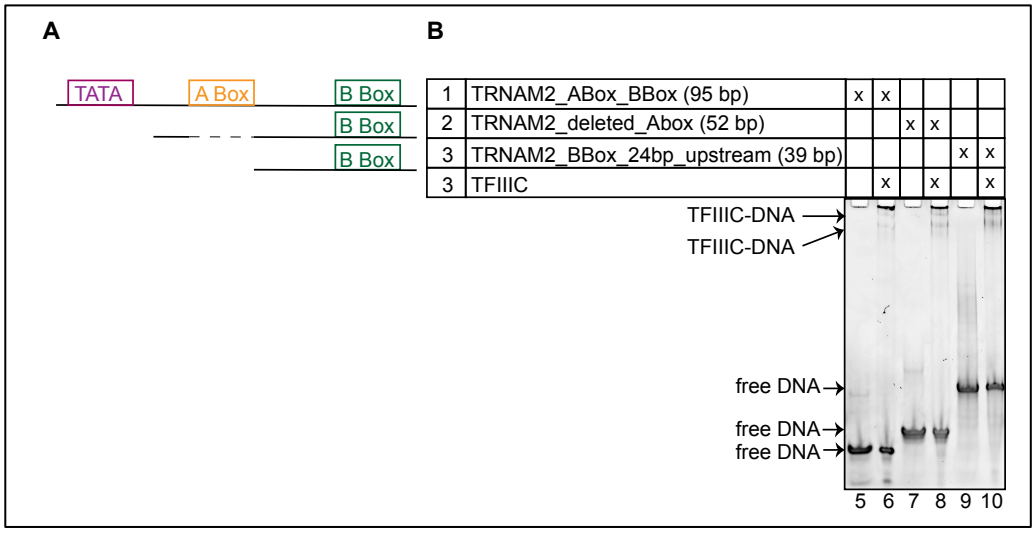


Figure 41: Differential TFIIIC-DNA-binding screen.

A: Schematic overview of the respective DNA constructs. B: DNA constructs with deleted A and B Boxes were used on a 0.5x TBE 4% acrylamide native PAGE stained with ethidium bromide.

To fully exclude differential binding to the A and B Box and to study whether TFIIIC would be able to bind to a DNA sequence only containing an A Box and lacking a B Box; two DNA constructs were designed that contained exclusively an A Box or a B Box. From the TRNAM2 promoter,

the 39 bp construct containing the B Box, shown in Figure 40B, was labelled with Cy5 (Cy5-TRNAM2_BBox_24bp_upstream). Additionally, a 36 bp construct solely containing an A Box was labelled with Cy3 (Cy3-TRNAM2_ Abox) (Figure 42A). 1 μmol of TFIIIC was incubated with these DNA constructs in at a 1:1 ratio using interaction buffer B (Table 16). The DNA oligos were labelled with Cy5 and Cy3, therefore no incubation with ethidium bromide was required to visualize the DNA. The native PAGE was detected at the excitation wavelength of Cy3 (554 nm) and Cy 5 (649 nm). The DNA-binding was studied using a 0.5xTBE 4% acrylamide native PAGE (Figure 42). The native PAGE (Figure 42B) was detected at 554 nm, thus visualizing the DNA containing the A Box.

Lanes 2 and 10-12 gave no signal since they only contained the Cy5 labelled DNA. Interestingly, a double DNA-binding event could be observed in lanes 5,6,8 and 9. The two bands appeared similar to the previously observed double-DNA-binding in Figure 40. Titrating increasing amounts of TFIIIC over the A Box DNA, increased the intensity of the TFIIIC-DNA shift (lanes 7 to 9). Binding of TFIIIC to a DNA containing exclusively an A Box was observed previously for yeast TFIIIC (Schultz et al. 1989). It was, however, unexpected for the human TFIIIC complex, since τA, which was reported to mediate the A Box binding (Bartholomew et al. 1990), was unable to bind to the DNA in previous EMSAs. These findings led to two hypotheses: either τA required the presence of τB to form a stable DNA binding or τB itself was able to bind to the A Box as well as to the B Box. Most likely, τA and τB are only fully functional when reconstituted as holo-TFIIIC. This hypothesis was supported by the fact that isolated τA contained chaperones and that purifying τB in isolation led to insolubility problems. Due to the limited amount of isolated τB, it was not possible to follow up on the hypothesis that τB could be able to also bind to the A Box DNA.

The native PAGE in Figure 42C was excited at 649 nm. TFIIIC-DNA-binding could be detected in lanes 5,6,11 and 12, confirming the prior observations that the B Box DNA was sufficient to mediate a TFIIIC-DNA-binding.

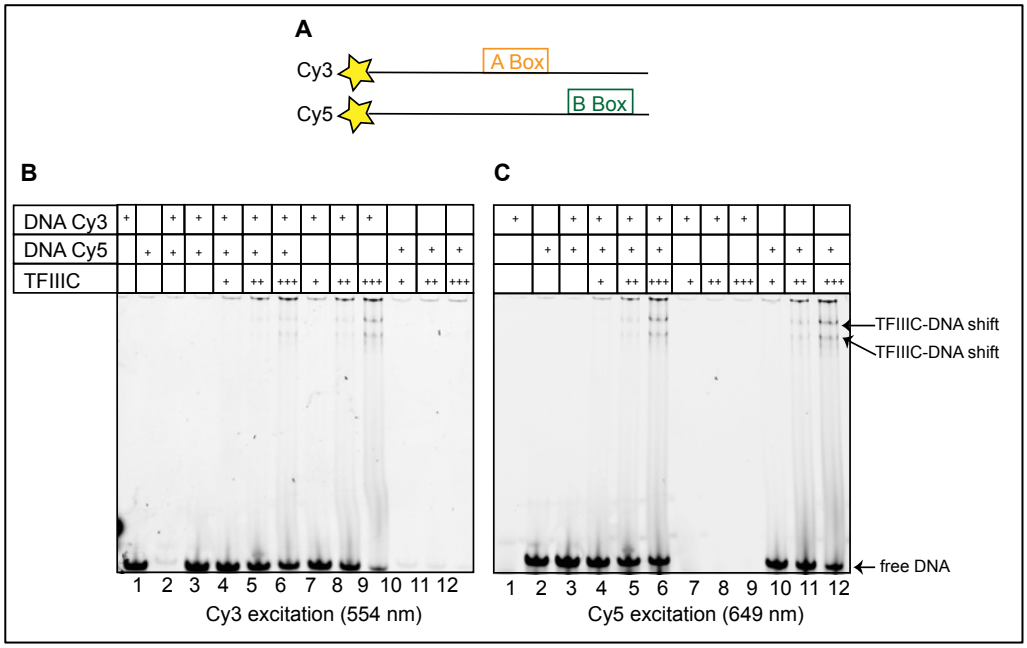


Figure 42: Human TFIIIC differential A and B Box DNA binding screen.

A: Schematic representation of the Cy3 and Cy5 labelled DNA construct. B: 4% acrylamide 0.5x TBE native PAGE detected at 554 nm. C: 4% acrylamide 0.5x TBE native PAGE detected at 649 nm. The TFIIIC-DNA shifts and the free DNA is indicated with arrows.

At both wavelengths, no significant difference in the DNA binding between the constructs containing either the A Box or the B Box sequence could be detected. Some aggregation could be observed in the wells, so that the samples were also loaded onto a 2% agarose 1xTBE native PAGE (Figure 43A, B). The gel detection at both excitation wavelengths confirmed the previous results (Figure 42). Due to the wider pores and lower resolution capacity of the 2% agarose gel, the sample could enter into the gel, but the DNA binding could not be resolved into the two individual binding events as detected in 4% acrylamide gels.

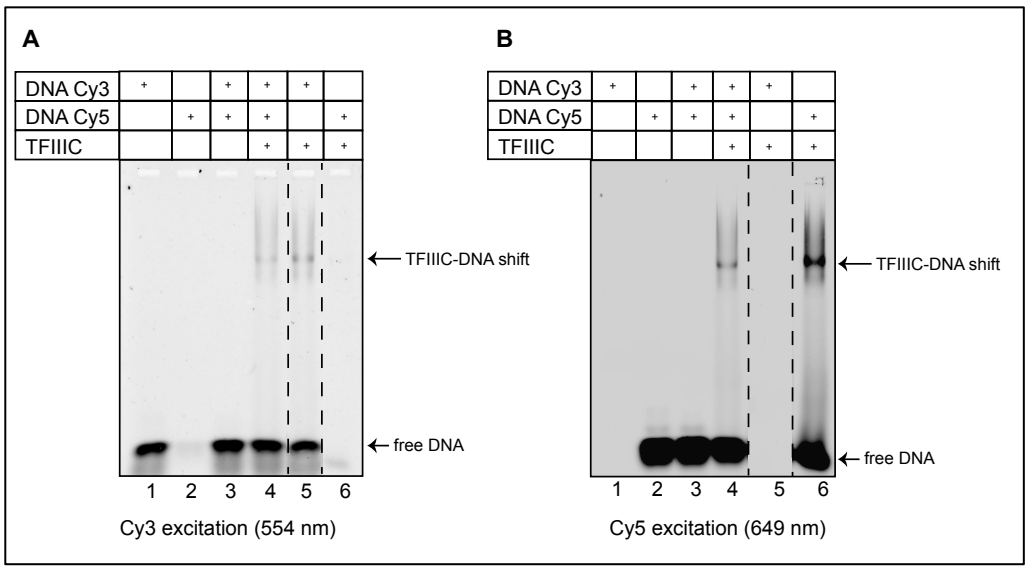


Figure 43: Differential A and B Box DNA binding screen.

A: 2% agarose 1x TBE native PAGE detected at 554 nm. B: 2% agarose 1x TBE native PAGE detected at 649 nm. The dashed line indicates that the gel was cut to show only the relevant lanes. The TFIIIC-DNA shift and the free DNA is indicated with an arrow.

In summary, the DNA binding studies of the human TFIIIC complex showed that two independent DNA-binding events took place, these were independent of the presence of the A and B Box DNA sequence. Furthermore, it was possible to demonstrate that either an A Box sequence or the B Box sequence on their own were sufficient for a stable TFIIIC-DNA-binding. In the performed experiments, binding to the A Box was only possible for holo-TFIIIC and not for τA. It remains unclear what caused the occurrence of two DNA binding events; this will be subject of future studies.

3.1.4.4. Structural analysis of the human TFIIIC complex using negative stain EM

In the previous experiments, it could be demonstrated that TFIIIC was functional and capable of binding to the DNA. The next step was to characterize apo TFIIIC structurally via negative stain EM. During the process of screening negative stain grids of TFIIIC, it became apparent that TFIIIC had to be prepared fresh and could not be frozen and stored since it showed a high degree of aggregation after thawing. Hence, to collect negative stain micrographs, TFIIIC had to be prepared and studied on the same day.

Several fractions of peak 2 (Figure 39A) were screened initially at the FEI T12 microscope before a data collection at the FEI F20 microscope was conducted The chosen fraction was the peak fraction in which all six subunits could be identified in the SDS-PAGE (Figure 44A). A representative micrograph of the data collection (76 images) is shown in Figure 44B.

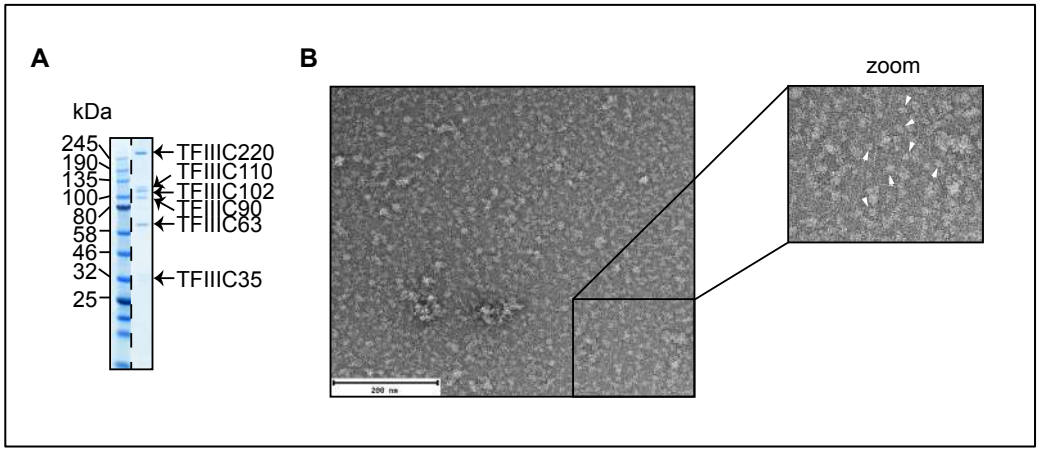


Figure 44: Human TFIIIC negative stain analysis.

A: SDS-PAGE indicating the different subunits of human TFIIIC-FLAG that was used for the grid preparation. B: Negative stain EM micrograph taken at the FEI F20 microscope.

The concentration of TFIIIC on the grid was slightly too highly concentrated but using a combination of manual picking and swarm picking with EMAN2 enabled the selection of 12,130 individual particles.

These were imported into Relion 2.1, subjected to 2D classification and used as references for an auto-picking that identified 15,268 particles. After two rounds of 2D classification and subsequent class selection, 4,210 particles were chosen. Six of the selected 2D classes are shown in Figure 45A which were produced by masking tightly around the particles, with a mask diameter of 210 Å. From the 4,210 particles, an initial *ab initio* model was generated that was further refined and post-processed to a resolution of 30.4 Å (Figure 45B).

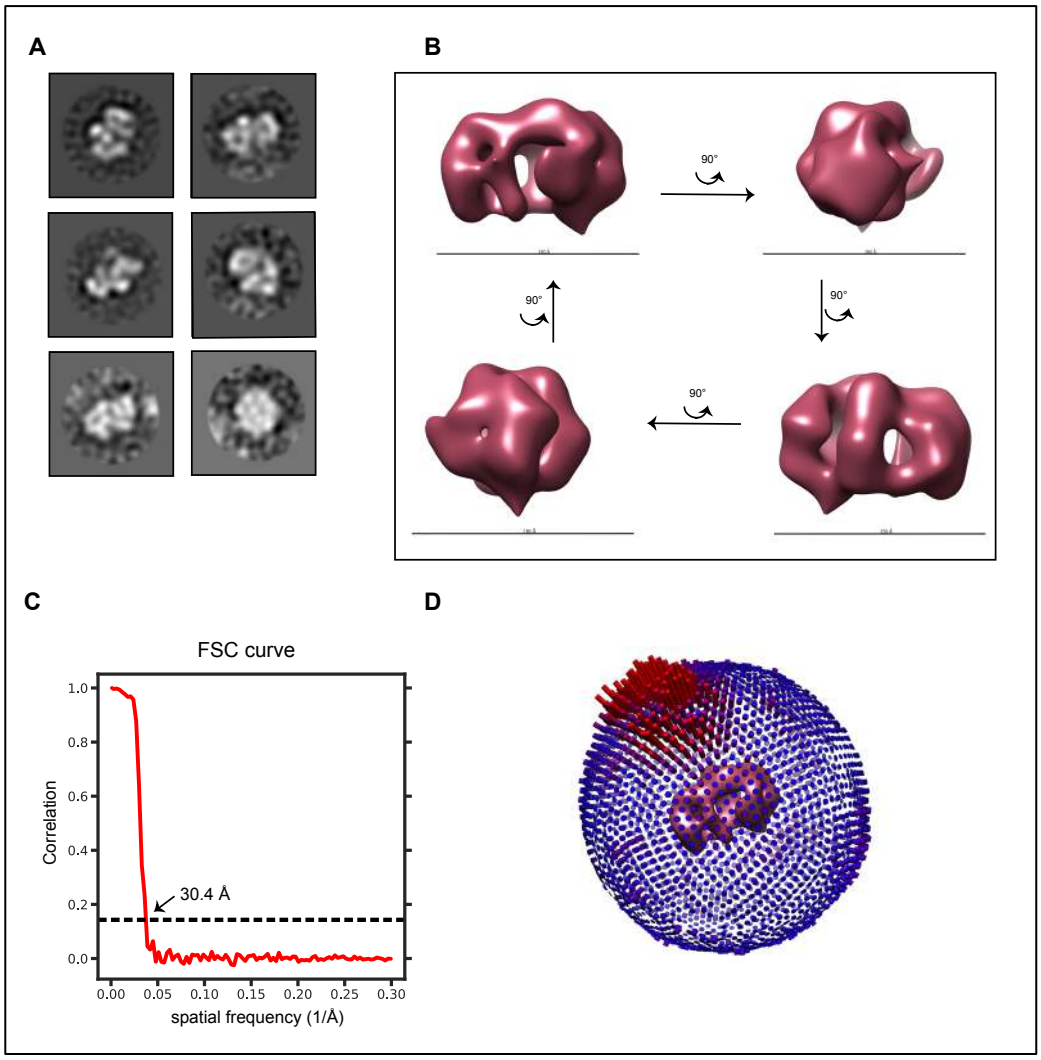


Figure 45: Human TFIIIC negative stain reconstruction.

A: Exemplary 2D classes. B: 3D reconstruction of apo human TFIIIC at 30.4 Å. C: Fourier-Shell Correlation, taken from Relion 3.0. D: Angular distribution plot taken from Relion 3.0.

To date, no structure of the human TFIIIC complex was available, which made the assessment of the obtained EM map challenging. Yeast TFIIIC was reported to adopt a dumb-bell shape as observed by scanning transmission electron microscopy (Schultz et al. 1989) and some isolated crystal structures have been solved but no high resolution reconstruction of a holo-TFIIIC complex was available.

The resulting 3D reconstruction had a diameter of approximately 100 Å. Due to the low number of particles, it was not possible to refine the structure to a higher resolution. Crystal structures of the yeast TFIIIC complex could also not be fitted unambiguously due to the low resolution. However, it was interesting to observe that the apo TFIIIC complex was elongated and showed two main domains, as previously reported (Schultz et al. 1989), which were connected via a third domain.

Limitations to the resolution arose due to heterogeneity on the grid. This could have been caused by TFIIIC partially disassembling on the grid due to the uranyl acetate stain. Another explanation for the heterogeneity could be the presence of a four-subunit complex of TFIIIC as reported previously (Figure 38). Furthermore, the small number of particles limited the 3D classification, which would have been necessary to deal better with the heterogeneity. Thirdly, apo TFIIIC potentially possessed a certain degree of conformational flexibility when not bound to the DNA.

To address the issue of flexibility, the effect of binding TFIIIC to the TRNAM2 promoter (as shown in section 3.1.4.3) was explored. The TFIIIC-DNA complex was purified via a Superose 6 Increase 10/300 GL column and analyzed via negative stain EM. The 3D reconstruction obtained, however, was at a similar resolution and did look very comparable to the apo TFIIIC model. In order to decrease the heterogeneity due to the disassembly of TFIIIC on the grid, crosslinking

with Glutaraldehyde was tested. However, instead of stabilizing TFIIIC, the protein complex aggregated and no grid could be stored.

The low yield obtained after TFIIIC purification limited the number of conditions that could be tested for structural analysis. However, it was possible to obtain a preliminary low-resolution reconstruction, which can be used as a reference for further structural studies. A validation of the generated 3D density will first be possible upon obtaining higher-resolution structures. Crystal screenings of apo and DNA bound TFIIIC, as well as τA, were set up using JCSG+ (Molecular Dimensions), PACT (Molecular Dimensions), PEG/Ion (Hampton Research) and Salt Rx (Hampton Research) in concentrations between 25mg/ml and 9 mg/ml, but no crystals were obtained.

3.1.5. Human four-subunit TFIIIC complex

The low yield that was obtained from human holo-TFIIIC purifications limited the amount of conditions that could be tested for structural analysis. To overcome this problem, a four-subunit TFIIIC complex, consisting of τA and TFIIIC220 was generated, as a dissociation of TFIIIC90 and TFIIIC110 could be observed at previous holo-TFIIIC purifications. After establishing that the individual subcomplexes τA and τB were not stable, the main question that was addressed in this section was whether TFIIIC may be present in other sub-assemblies such as a four-subunit TFIIIC complex and TFIIIC90/TFIIIC110 and what role the TFIIIC90/TFIIIC110 heterodimer could play with regards to the full TFIIIC complex.

3.1.5.1. Expression and purification of the human four-subunit TFIIIC-FLAG complex

The observation that a four-subunit sub-complex, consisting of TFIIIC220-FLAG, TFIIIC102, TFIIIC63 and TFIIIC35, was present in both purification attempts of the holo-TFIIIC constructs led to the decision to further clone another construct with only these four-subunits. It was thought that this would facilitate the structural analysis of TFIIIC by decreasing the heterogeneity.

The four-subunit TFIIIC was cloned using the biGBac cloning system. The pBIG1b containing His-TFIIIC35 and TFIIIC63 was assembled together with the pBIG1a containing TFIIIC102 and 3xFLAG tagged TFIIIC220. The presence of all subunits in the resulting construct was confirmed via an analytical SwaI digest (Figure 46A) in addition to DNA sequencing.

The four-subunit TFIIIC was expressed for three days in insect cells and purified via three steps, starting with an incubation with anti-DYKDDDDK resin as described in section 2.3.2. The protein complex eluted from the anti-DYKDDDDK resin and was pure enough to be stored for biochemical purposes. However, for structural studies, a Hi Trap Heparin HP column and a gel filtration column (Superose 6 Increase 10/300 GL) were performed. A chromatogram of the final step is shown in Figure 46B. We observed that the complex eluted in two distinct peaks as indicated in the chromatogram. Peak 1 eluted at 10 – 11 ml and showed a 1:1 ratio of absorption at 260 and 280 nm. It was unclear whether this peak contained protein aggregates or an oligomeric form of the four-subunit complex as it eluted earlier than predicted for a protein complex of 429.8 kDa; also eluting earlier than holo-TFIIIC. Peak 2 was detected as a broader peak that eluted between 13 and 16 ml. The absorbance at 260 nm in this peak was higher than the absorbance at 280 nm which indicated that, despite the Heparin ion exchange step, DNA contamination was still present. The peak marked with an asterisk was caused due to stopping and re-starting the ÅKTA.

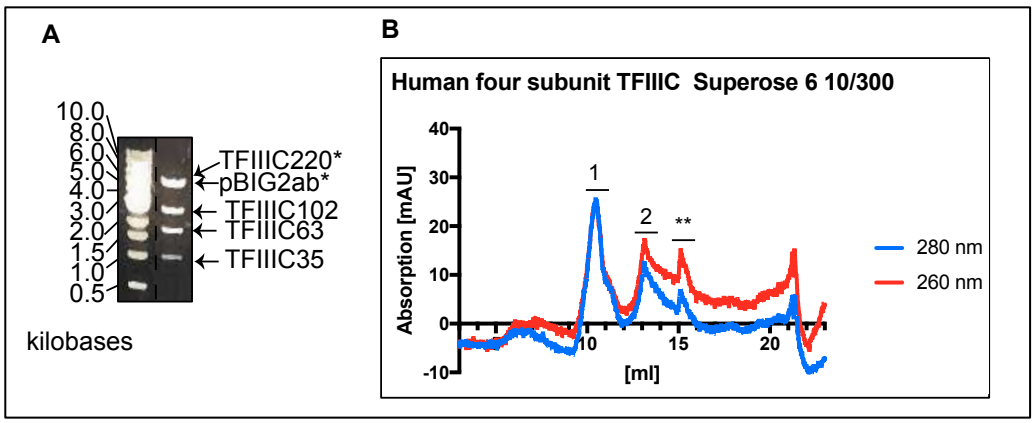
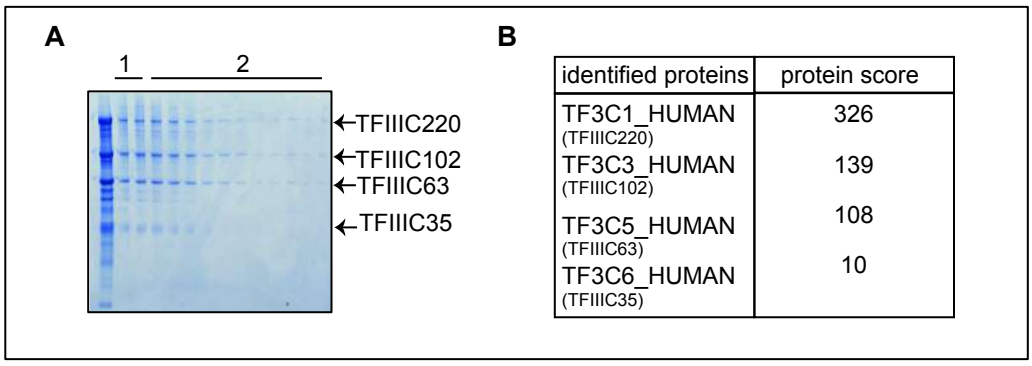


Figure 46: Four-subunit human TFIIIC.

A: 0.8% agarose DNA gel of the SwaI digestion pattern of the four-subunit TFIIIC construct. The bands marked with an Asterisk, pBIG2ab (5668 bp) and TFIIIC220-FLAG (6435 bp), are not separated under this running condition. B: Chromatogram of four-subunit TFIIIC-FLAG after elution from a Superose 6 10/300 Increase. The two asterisks indicate that this peak was due to the system stopping and restarting.

The different elution fractions were analyzed on an SDS-PAGE (Figure 47A) and four bands could be identified, present in a stoichiometric ratio. In addition to the protein bands, a significant number of degradation products could be seen in the SDS-PAGE. One of the peak fractions was analyzed via inhouse LC-MS/MS analysis and confirmed the presence of all expected subunits (Figure 47B).

The yield after the gel filtration purification was too low meaning no protein complex could be stored after this step.



identified proteins	protein score
TF3C1_HUMAN (TFIIIC220)	326
TF3C3_HUMAN (TFIIIC102)	139
TF3C5_HUMAN (TFIIIC63)	108
TF3C6_HUMAN (TFIIIC35)	10

Figure 47: LC-MS-MS analysis of the four-subunit human TFIIIC complex.

A: SDS-PAGE showing different fractions from the peaks in ***Figure 46****B. B: Identified proteins from the inhouse LC-MS/MS analysis.*

3.1.5.2. DNA binding studies

The four-subunit TFIIIC complex was subsequently tested for its ability to stably bind the DNA. Therefore, the four-subunit TFIIIC complex was incubated with the TRNAM1 promoter containing both an A Box and B Box and the Terminator sequence, labelled with Cy5. An excess of TFIIIC complex was titrated over the DNA (7.5 fold and 10-fold) and incubated with the DNA for 30 minutes at room temperature (interaction buffer: 500 mM NaCl, 20 mM Tris pH 7.5, 1 mM $MgCl_2$, 10% glycerol, 2 mM DTT). The samples were loaded onto a 1x TBE 2% agarose gel (Figure 48). The native PAGE was detected at the Cy5 excitation wavelength of 649 nm. We observed that the four-subunit TFIIIC complex was capable of binding to the DNA. Indeed, a double DNA binding could be observed although the upper band appeared less intense than the lower TFIIIC-DNA band.

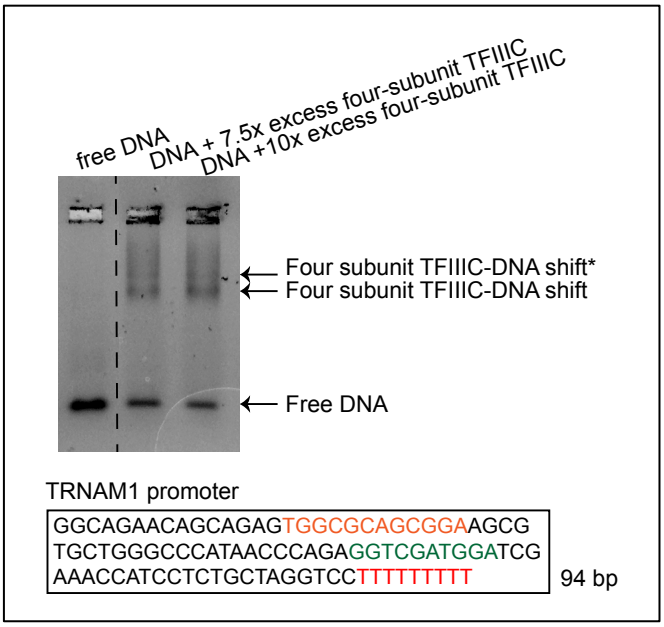


Figure 48: Four-subunit TFIIIC complex EMSA.

The four-subunit TFIIIC complex was titrated over the DNA in a 7.5x and a 10x excess. The 1x TBE 2% agarose gel was detected at 649 nm. The dashed line indicates that the gel was cut.

From these results, it was decided to study the four-subunit complex structurally by negative stain EM.

3.1.5.3. Structural analysis using negative stain EM

Due to the limited yield, the DNA binding step was integrated into the purification protocol to avoid losing material by loading the DNA-bound sample again onto a gel filtration column. Thus, the four-subunit complex was purified as described in section 3.1.5.1 via Anti-DYKDDDDK beads and a Heparin column. TFIIIC was incubated for 30 minutes with the TRNAM1 promoter, that was found to be a suitable DNA oligo in the previous EMSA, and then loaded onto the Superose 6 Increase 10/300 GL column (Figure 49A, B). The different elution fractions were then analyzed by SDS-PAGE (Figure 49C) and on a 1x TBE 2% agarose gel stained with ethidium bromide to analyze the TFIIIC-DNA-binding (Figure 49D). The EMSA suggested that fractions D6-D11, that eluted in the peak at 13 ml (Figure 49B), contained the TFIIIC-DNA complex, whereas fractions E2 and E3 contained free DNA.

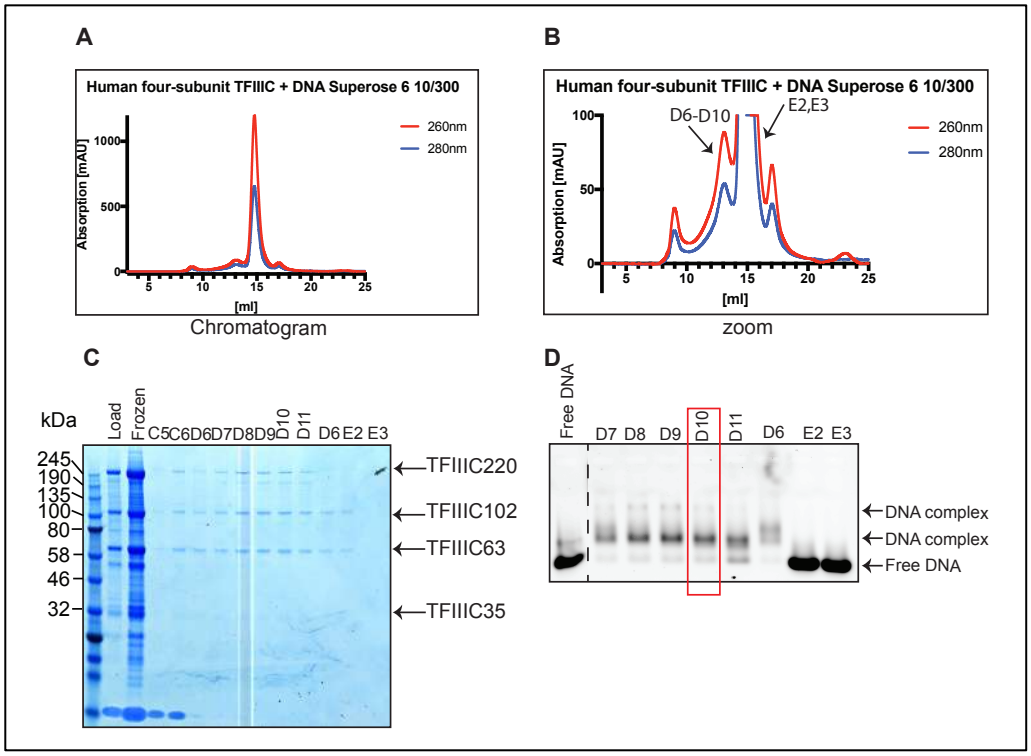


Figure 49: Human four-subunit TFIIIC-DNA complex formation for structural analysis.
A: Chromatogram of the four-subunit TFIIIC complex bound to the DNA. B: zoom into the peak of the chromatogram in A. C: SDS-PAGE, stained with Coomassie Blue, analyzing the fractions that eluted from the Superose 6 Increase 10/300 GL column. D: EMSA of the fractions on a 1x TBE 2% agarose gel stained with ethidium bromide. The dashed line indicates that the gel was cut.

The fractions containing the DNA bound complex were used to prepare negative stain grids and screened at the FEI T12 microscope (Figure 50). Fractions D6 and D7 appeared to contain some filamentous contamination (marked with red arrows); it was unclear what these filaments were. Fractions D8 and D9 showed positive staining, meaning no thorough analysis could be conducted. In fraction D11, a variety of different sizes of particles could be seen which meant the picking of the correct target was challenging. The most promising fraction was fraction D10. The particles were monodisperse, although there was a great variability among the size and shape of the particles. Furthermore, chaperones (marked with blue arrows) could be seen in the micrograph, indicating that the complex was potentially contaminated during the

purification to protect unfolded regions. This observation could mean that the preferred state of TFIIIC would still be in form of the holo-TFIIIC complex and that TFIIIC90 and TFIIIC110 indeed play a more important role than expected.

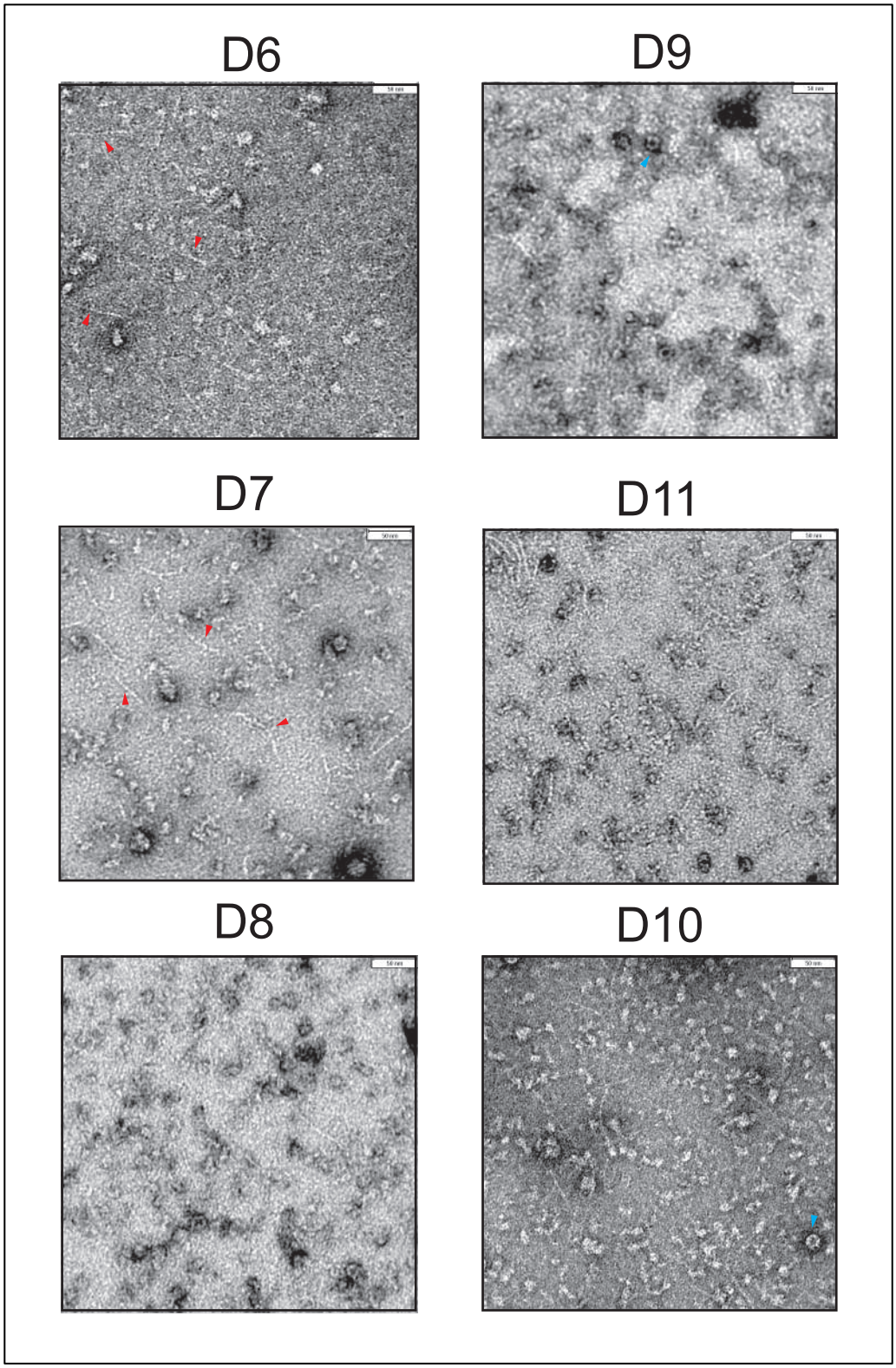


Figure 50: Negative stain analysis of the four-subunit TFIIIC-DNA complex.

The micrographs of the different fractions in ***Figure 49*** *were taken at the FEI T12 microscope. Exemplary filaments are marked with red arrows and chaperones are marked with blue arrows.*

No data collection using the grid containing fraction D10 was conducted as the grid appeared unsuitable. Regardless, the performed experiment demonstrated that the four-subunit complex was able to form a stable DNA complex even after a gel filtration column. Furthermore, it was possible to see that the complex formed individual particles and did not aggregate fully on the grid. However, the presence of chaperones might be a sign that the holo-TFIIIC complex is the better target to be studied.

3.1.6. Probing TFIIIC interactions with CHD4 and ADNP

As afore mentioned, TFIIIC not only plays an important role in transcriptional initiation at RNA Polymerase III promoters but also in genome organization. In a recent study carried out by (Ferrari et al., 2019 under review), it was shown that TFIIIC plays a direct role in loop forming through association with SMC complexes at Alu elements. In order to understand better how TFIIIC gets recruited to Alu elements throughout the genome, interactions with CHD4 and ADNP were tested, that were found to interact with TFIIIC in a RNA Seq experiment. In order to do so, HA tagged ADNP and CHD4 Baculovirus were produced (Dr Erin Cutts). With these viruses, individual expression tests and co-infections with Flag tagged TFIIIC were tested in High5 cells.

TFIIIC and ADNP or TFIIIC and CHD4 were co-expressed for three days in High5 cells. The harvested cell pellets were incubated with HA resin or Anti-DYKDDDDK (FLAG) resin, washed three times and analyzed via an SDS-PAGE (Figure 51). The fractions that were bound to the FLAG beads are shown in green and fractions bound to the HA beads are shown in blue.

We observed that TFIIIC bound to the FLAG beads, however some nonspecific binding to the HA resin was also detected. CHD4 bound to the HA beads, as expected, but also nonspecifically to the FLAG beads (lanes 3 and 4). ADNP only bound to the HA resin and did not bind nonspecifically to the FLAG beads (lanes 5 and 6). An interaction between TFIIIC and ADNP could be observed by binding the co-infection to the FLAG beads (lane 7, Figure 51). However, in order to confirm this interaction without doubt, further experiments with the purified proteins rather than cell extracts would be required. An interaction between TFIIIC and CHD4 could not be confirmed in this assay due to nonspecific binding and because CHD4 ran at the same position as TFIIIC220.

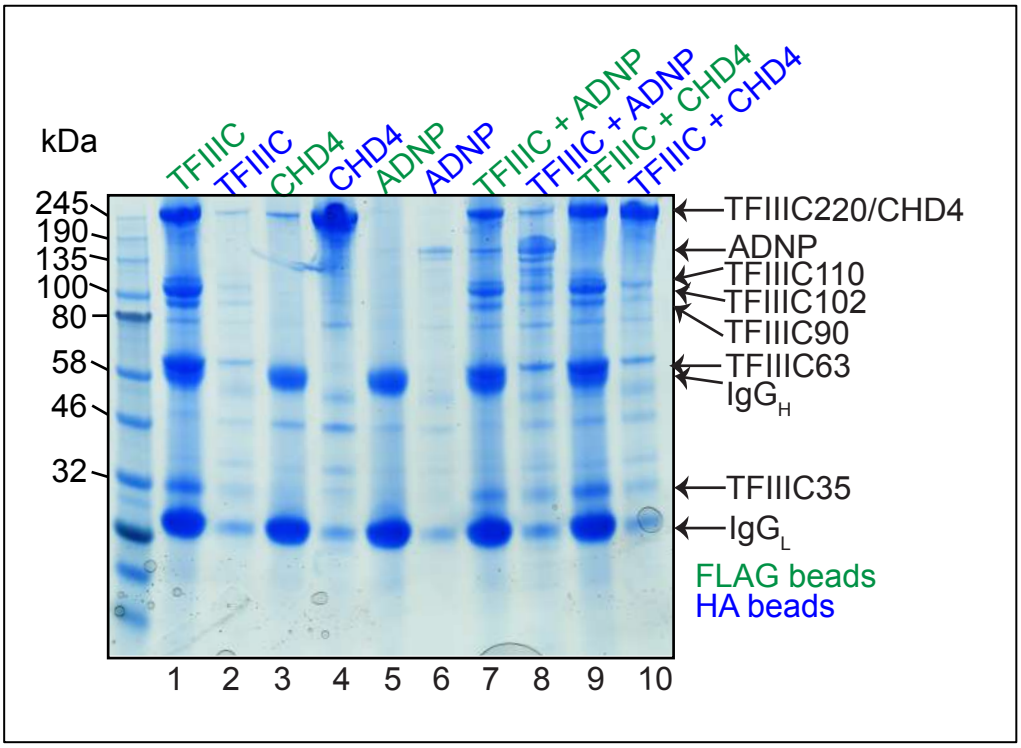


Figure 51: TFIIIC, ADNP and CHD4 co-infection expression test.

The SDS-PAGE was stained with Coomassie Blue. The fractions bound to the FLAG beads are shown in green, fractions bound to the HA beads are shown in blue. The position of the different subunits is indicated.

For future experiments, it would be interesting to test the co-purification of ADNP and TFIIIC. Co-purifying TFIIIC with an interaction partner could provide additional stability to TFIIIC itself.

3.1.7. Summary of human TFIIIC characterization

In summary, I was able to successfully clone a construct of holo-TFIIIC containing a FLAG tag and I managed to find an optimized position for the affinity tag on τB. I was able to establish a preliminary purification protocol for τB and to optimize purification protocols for τA and holo-TFIIIC. I could show that holo-TFIIIC can only be purified as a stoichiometric complex by purifying it via its affinity tag on τB since τA subunits were expressed to higher levels.

Furthermore, I could demonstrate that it was necessary to reconstitute the holo-TFIIIC complex to obtain a robust complex that can form a stable DNA bound complex. I was able to confirm that the isolated τB subcomplex bound to the B Box DNA but I was not able to see the binding of τA to the A Box DNA in the tested EMSA conditions. It is important to note at this point, that τA could have formed a DNA-bound complex, which was subsequently disrupted by applying a voltage to run the EMSA. It would thus be important to study the τA-DNA binding reaction using different techniques such as fluorescence polarization.

Additionally, I studied the binding of TFIIIC to DNA constructs that exclusively contained the A or the B Box and I was able to show that the presence of either of the recognition motifs was sufficient for the formation of a stable TFIIIC-DNA complex. It could be observed that the TFIIIC-DNA complex caused two shifts in the EMSA. This could have been due to the phosphorylation status or due to different confirmations of TFIIIC. This will be subject to further studies in the future.

I analyzed the binding of τA to TFIIIB, its natural binding partner by pulldowns. I was able to detect an interaction between TBP and τA which could be characterized further in the future. I could show that the C-terminal region of Brf1 was not sufficient to form an interaction between τA and Brf1. Optimizing an expression and purification protocol for a longer Brf1 construct that contains the N-terminal region will probably enable the detection of a τA-Brf1 interaction *in vitro*.

The activity of the TFIIIC complex was also tested by HAT Assays (Kundu et al. 1999; Basu et al. 2019). The assay was performed using a colorimetic and a fluorimetric assay kit as well as a Western Blot with antibodies detecting acetylated Histone residues. Unfortunately, due to the low sample concentration or a readout below the detection level of the assay, it was not possible to confirm the HAT activity of the purified TFIIIC complex.

Regarding the structural studies, I successfully obtained a preliminary 3D reconstruction of apo TFIIIC that provides a good starting point for further structural studies using electron microscopy. The generated 3D map can however first be validated by obtaining higher resolution maps.

3.2. *S. cerevisiae* TFIIIC studies

The expression and purification of the human TFIIIC complex proved challenging, which made an extensive biochemical and structural characterization difficult. A new approach using the functionally conserved *S. cerevisiae* TFIIIC complex, provided an alternative opportunity to further investigate the TFIIIC complex. This would allow us to understand the molecular mechanism of TFIIIC better and to understand the differences between DNA-bound and apo TFIIIC by analyzing structural data.

In this chapter, the cloning, expression and purification of different yeast TFIIIC constructs and their ability to bind to the DNA will be discussed. Furthermore, the reconstitution of the holo-TFIIIC complex and its biochemical and biophysical characterization will be addressed. Several cryo-EM data collections gave unprecedented insights into the structural organization and the behavior of the complex. EM reconstructions of the τ60/τ91 domain with a resolution of 7 - 8 Å and an overall reconstruction of the holo-TFIIIC complex at 15 Å could be obtained. Multi-body refinements were performed to illustrate the flexibility and the different conformational states of the TFIIIC complex, which was supported by native mass spectrometry and SEC-SAXS data.

3.2.1. Cloning, expression and purification of TFIIIC

Form the previous studies of the human TFIIIC complex, we concluded that a better strategy for the yeast TFIIIC complex would be to focus mainly on the holo-TFIIIC complex. The subcomplexes of TFIIIC were still expressed and purified in order to characterize which contribution each of the subcomplexes has towards the DNA binding affinity and whether alternative sub-assemblies of the TFIIIC complex were stable. Comparing

these results to the human TFIIIC complex could give valuable information about how conserved the function of TFIIIC is.

3.2.1.1. τA subcomplex

Following the human TFIIIC constructs, yeast TFIIIC was cloned with an N-terminal His-tag on its smallest subunit τ55. The three subunits of τA, His-τ55, τ95 and τ131 were assembled into a pBIG1a plasmid. The presence of all subunits in the vector was confirmed via analytical SwaI digest (Figure 52A) and DNA sequencing. From this construct, a baculovirus was generated to express τA in High5 cells. After three days, the cells were pelleted and a purification protocol was established to purify τA in three steps via a Ni-NTA agarose batch affinity step, a Heparin ion exchange step and a final gel filtration step using a Superose 6 Increase 10/300 GL column. The chromatogram (Figure 52B) showed that the majority of τA eluted in a peak at 15 ml (peak 3) and a small amount of the protein complex ran in the void volume (peak 1). All three subunits could be detected when the fractions were analyzed via SDS-PAGE (Figure 52C). The total yield from a 500 ml High5 cell culture was 1.31 mg, frozen in aliquots at a concentration of 2.62 mg/ml (10.8 μM).

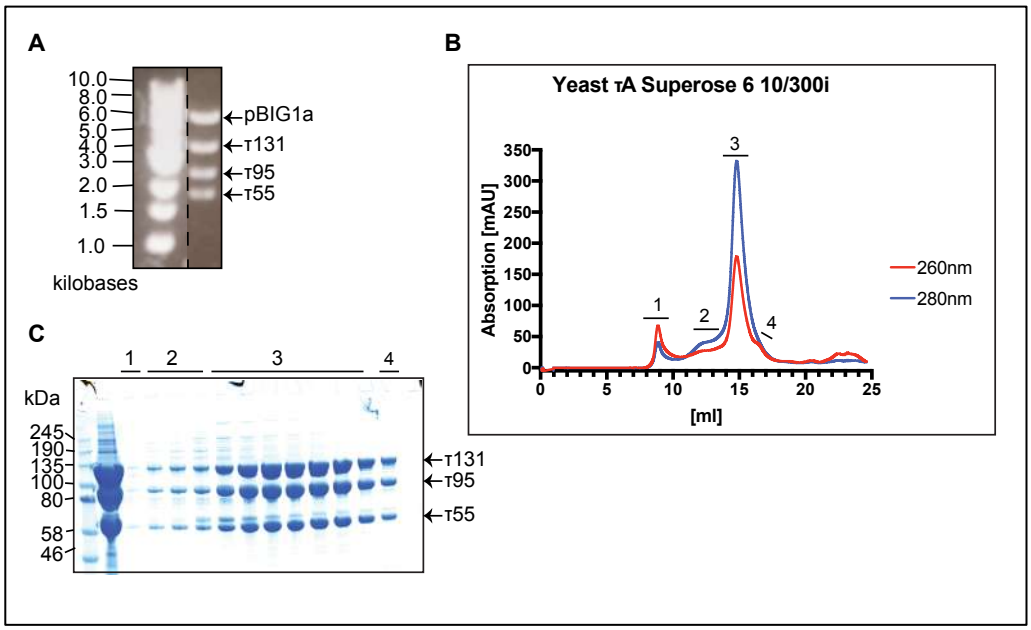


Figure 52: Yeast τA cloning and purification.

A: 1x TAE 0.8 % agarose gel showing an analytical SwaI digestion of τA cloned into a pBIG1a plasmid. B: Gel elution profile from a Superose 6 Increase 10/300 GL column. C: SDS-PAGE showing the different fractions that eluted from the Superose 6 Increase 10/300 GL column shown in B.

3.2.1.2. τB subcomplex

For the expression and purification of τB, the pBIG1b construct was used to produce a baculovirus (Dr Erin Cutts, Table 6, #24). This construct had a twin Strep-tag at the C-terminus of τ91. The protein complex was expressed for three days in High5 cells, the cells were then pelleted and the purification started with a Strep Trap HP column. After the affinity step, it became apparent that instead of τB, only the heterodimer τ60/τ91 was pulled down. This reduced complex was further purified via a Hi Trap Heparin HP (Figure 53A). The total yield from a 500 ml infection of High5 cells was 1.7 mg, which were stored in aliquots at a concentration of 2.1 mg/ml (14.93 μM). To determine whether τ138 was not expressed or whether it became insoluble during the cell lysis step, the insoluble fraction was analyzed on an SDS-PAGE (Figure 53B). According to the molecular weight of the bands in the insoluble fraction, it was presumed that τ138 was expressed, as indicated by the asterisk (Figure 53B), but became insoluble. The two asterisks could indicate a degradation product of τ138.

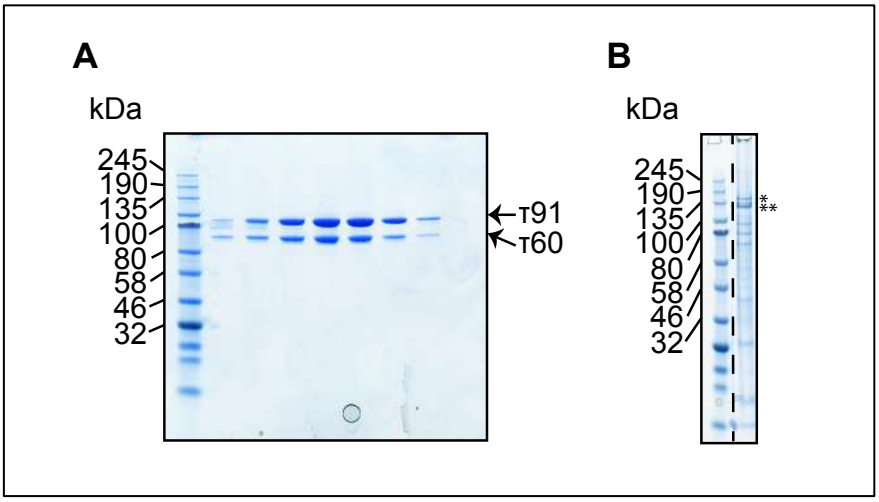


Figure 53: τB purification.

A: SDS-PAGE of τ60/τ91 after elution from the Heparin column. B: SDS-PAGE showing the insoluble fraction of τB

The purified heterodimer of τ60/τ91 was later injected onto a Superose 6 Increase 10/300 GL column to analyze its elution profile, and to confirm that no aggregation was present in the sample (Figure 54). It could be observed that τ60/τ91 eluted at approximately 16 - 17 ml and that the heterodimer had a A260:A280 ratio of 0.5. This suggested that no DNA contamination was present, which supports the fact that τ138 contains the DNA-binding domain of τA.

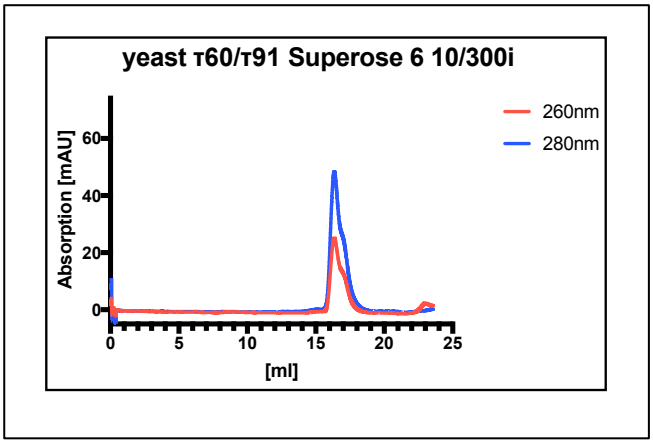


Figure 54: τ60/τ91 Superose 6 Increase 10/300 GL elution profile.

After successfully purifying τ60/τ91, the aim to purify the whole τB subcomplex still remained. To do this, we introduced an affinity tag on τ138, the subunit that became insoluble during the previous purification. Since a Strep-tag was already present at τ91, an N-terminal 3xFLAG tagged τ138 construct was designed. The subunits (FLAG-tagged τ138, Strep-tagged τ91 and τ60) were assembled in a pBIG1b plasmid using the Gibson Assembly. Their presence was confirmed via an analytical SwaI digest (Figure 55A) and DNA sequencing. The new τB construct was purified using Anti-DYKDDDDK resin (Genscript) to bind to the 3xFLAG tag. Additionally, a control pulldown was performed with Strep resin (Figure 55B). However, it was observed that the majority of τ138 remained insoluble and that only a small amount of τB was bound to the Anti-DYKDDDDK resin. The τ91 subunit, however, bound with high affinity to the Strep resin, which pulled down the τ60/τ91 complex as previously observed.

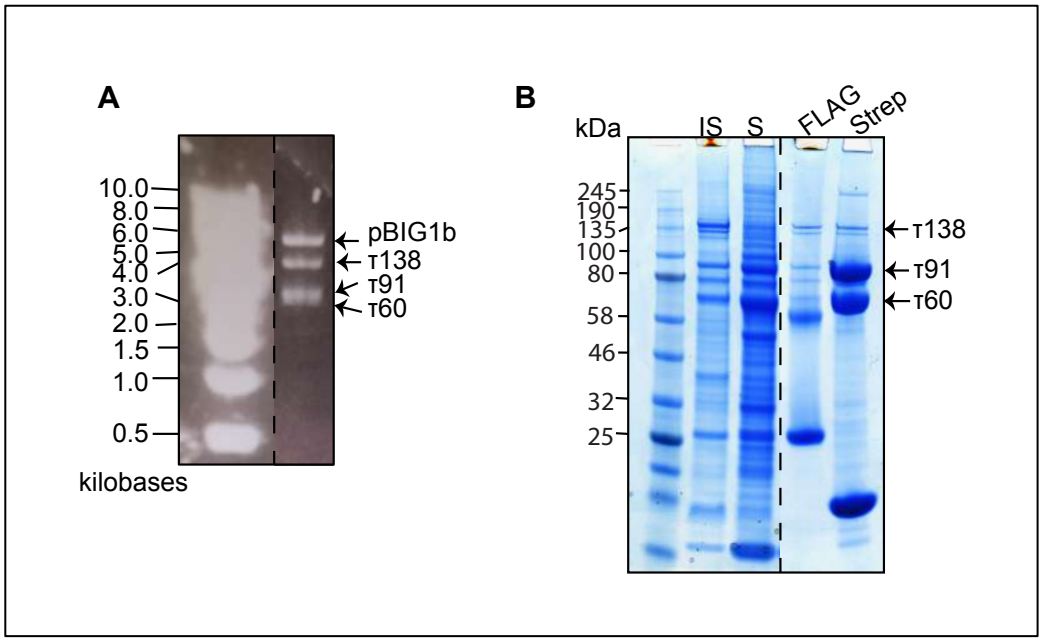


Figure 55: τB-FLAG cloning and pulldown.

A: SwaI digest of the FLAG tagged τB construct on a 1xTAE 0.8% agarose gel. B: SDS-PAGE showing the insoluble (IS) and soluble (S) fraction as well as the pulldown from the Anti-DKYDDDK resin (FLAG) and the Strep resin (Strep). The dashed line indicates that the SDS-PAGE was cut.

In summary, these findings indicated that τ60 and τ91 formed a stable subcomplex that could be purified in high yields, obtaining 1.7 mg from a 500 ml culture of High5 cells. Furthermore, it was shown that τ138 required additional binding partners to remain soluble and bound to the τ60/τ91 dimer. This hypothesis was tested in section 3.2.1.3.

3.2.1.3. Four-subunit TFIIIC complex

As the next step in the assembly of the yeast TFIIIC complex, we tested the purification of a four-subunit complex, as previously observed with the human TFIIIC complex. A successful formation of the four-subunit TFIIIC complex could hint to an architectural conservation between human and yeast TFIIIC and was thus tested in the following.

Initially, a baculovirus was generated from the FLAG-tagged τ138 construct was used for a co-infection with τA. The complex was expressed in High5 cells for three days and it was purified via Anti DYKDDDDK resin and a 1 ml Hi Trap Heparin column (Figure 56). The yield was relatively low at 189 μg and stored in aliquots of 0.189 mg/ml (504 nM).

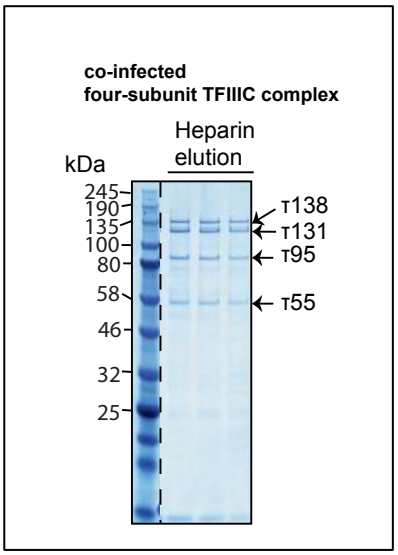


Figure 56: SDS-PAGE of the HiTrap Heparin elution from co-infected four-subunit TFIIIC complex.

As it was possible to purify a four-subunit complex in the yeast system, the next aim was to generate a virus that contained all four-subunits in one plasmid. Therefore, FLAG tagged τ138 was assembled with His-tagged τ55 into a pBIG1a plasmid and τ131 and τ95 were assembled into a pBIG1b plasmid. Both constructs were combined into a pBIG2ab using the Gibson Assembly to generate this construct. The presence of the four-subunits was confirmed via DNA sequencing and an analytical SwaI digest (Figure 57A). During the purification of this construct, however, it was clear that a large amount of the protein complex was insoluble (Figure 57B, lane IS) and that part of τ138 got degraded (Figure 57B, marked with an asterisk). The four-subunit complex was subsequently loaded onto a Superose 6 Increase 10/300 GL column (Figure 57C). Two peaks were identified in the chromatogram and were further analyzed via SDS PAGE (Figure 57D). It was not possible to identify the content of peak 1 using an SDS PAGE analysis but it was observed that the four-subunit complex eluted between 14 and 15 ml in peak 2. However, the yield was too low, so that it was decided to not further invest time into optimizing the purification of the four-subunit complex but rather to proceed with a reconstitution of the holo-TFIIIC complex.

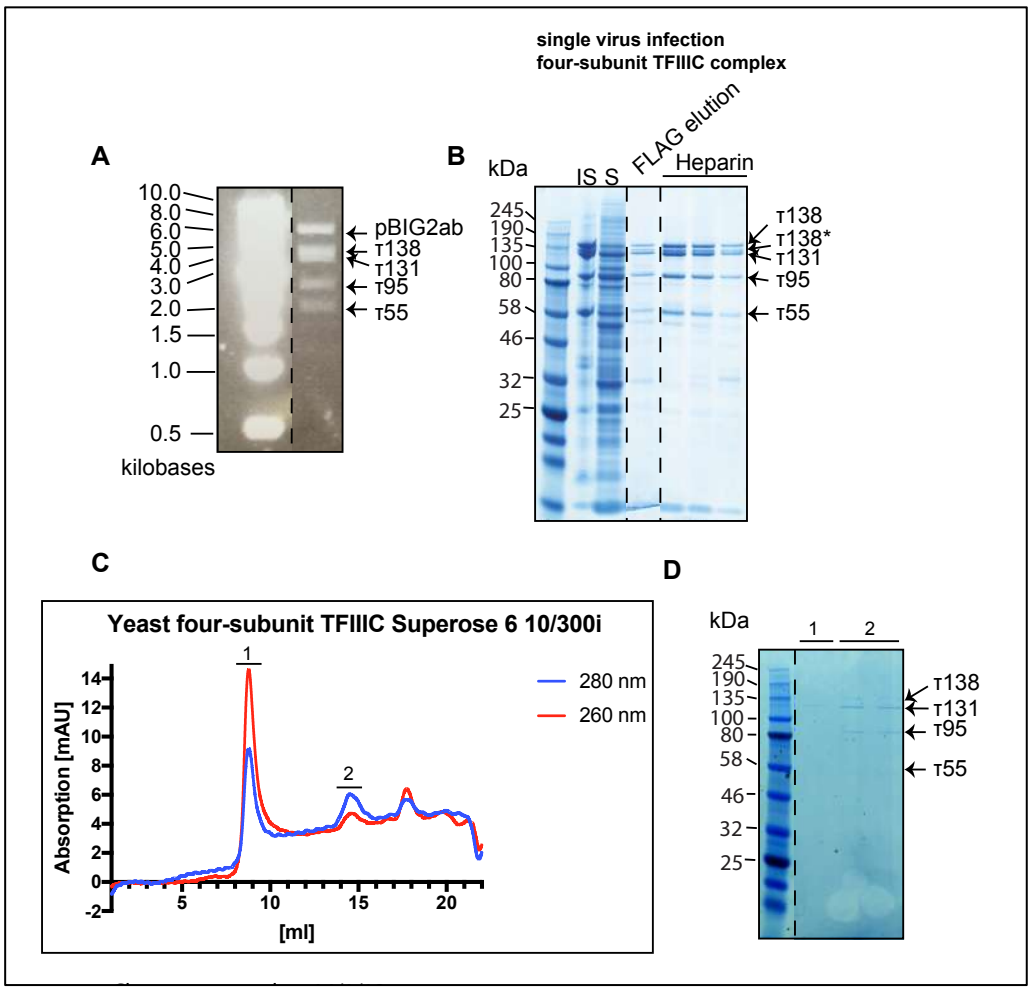


Figure 57: Yeast four-subunit TFIIIC cloning into one baculovirus and purification.

A: Analytical SwaI digest of the four-subunit yeast TFIIIC construct on a 1x TAE 0.8% agarose gel. B: SDS-PAGE of the yeast four-subunit TFIIIC complex that was cloned into one virus after the FLAG elution and the Hi Trap Heparin HP 1 ml elution. IS = insoluble, S = soluble. The asterisk indicates a degradation product of τ138. C: Superose 6 Increase 10/300 GL elution profile of the four-subunit TFIIIC complex, purified after infection with a single virus. D: SDS-PAGE of the fractions from peak 1 and 2 of the chromatogram shown in C.

Purifying sub-assemblies of TFIIIC, as described in this section, provided a better understanding how the yeast TFIIIC complex works and were subsequently used to reconstitute the holo-TFIIIC complex. Furthermore, these sub-assemblies enabled the study of the DNA-binding properties of TFIIIC in the absence of specific subunits.

3.2.1.4. Holo-TFIIIC complex

The following experiments were designed to obtain a holo-TFIIIC complex. TFIIIC was first reconstituted *in vitro* from its purified component parts and then via a co-infection with baculoviruses from both subcomplexes before it was purified as a holo-TFIIIC complex from a single baculovirus infection.

3.2.1.4.1. Reconstitution of holo-TFIIIC from purified subcomplexes

Initially, it was tested whether it was possible to reconstitute a holo-TFIIIC complex from the τ60/τ91 subcomplex, which had been previously purified (section 3.2.1.2) and the four-subunit TFIIIC complex *in vitro*. Therefore, 0.1 nmol Strep-tagged τ60/τ91 heterodimer was incubated with 30 µl of Strep resin using interaction buffer C (section 2.4.5). A titration of 0.1 and 0.2 nmol of the FLAG tagged four-subunit TFIIIC complex was added and incubated for 30 minutes at room temperature. After this, the beads were washed three times with 30 CV of interaction buffer C, mixed with 6x SDS Loading Buffer and analyzed on an SDS-PAGE (Figure 58).
It could be observed that a holo-TFIIIC complex was formed already in a 1:1 ratio. It became more stoichiometric with an increasing concentration of the four-subunit TFIIIC complex.

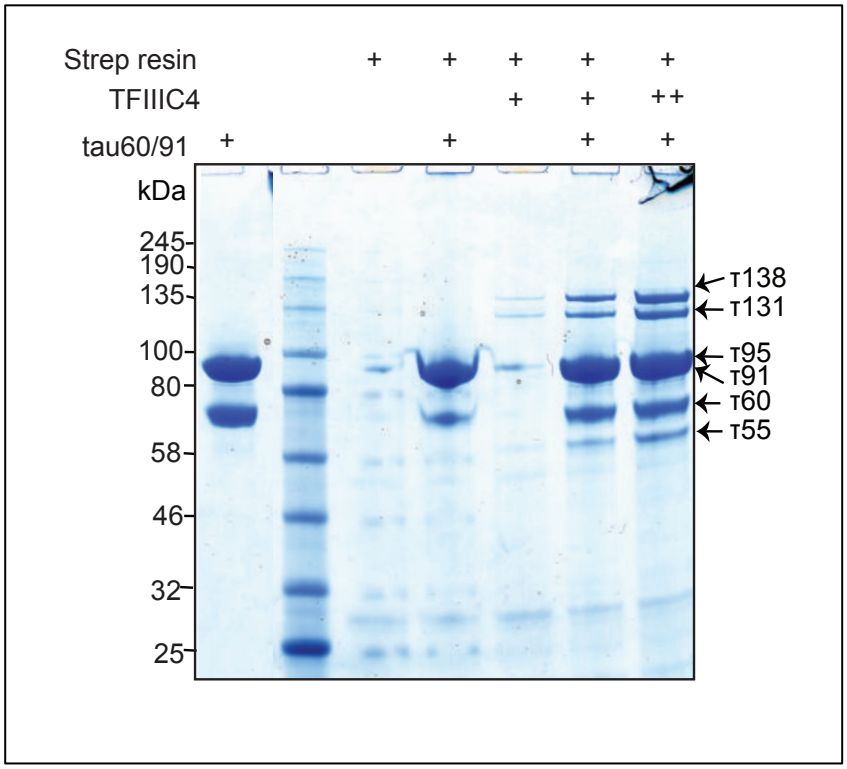


Figure 58: Reconstitution of holo-TFIIIC from purified subcomplexes.

Strep Pulldown of the FLAG tagged four-subunit TFIIIC complex with a titration of Strep-tagged τ60/τ91.

3.2.1.4.2. Co-infections of subcomplex baculoviruses

To further investigate the interactions between τA and τB, it was tested whether it was possible to express a holo-TFIIIC complex by co-infecting High5 cells with a Baculovirus containing τA and a baculorvirus containing τB. It was thought this may solubilize and improve stabililty of τ138, which was insoluble when τB was expressed in isolation. High5 cells were infected in a 1:1 ratio with τA and τB virus for three days before the cells were pelleted. The cells were lysed using the lysis buffer from section 2.3.2.4 and split in half. Half was incubated with Ni-NTA resin that could be bound by τ55 (τA) and the other was incubated with Strep resin, that could be bound by τ91 (τB). After a 2h of incubation, the beads were

washed three times and analyzed via an SDS-PAGE (Figure 59). The running controls of τA and the purified τ60/τ91 complex were loaded in parallel to the soluble fraction, the insoluble fraction and the fraction bound to the respective beads. Whereas previously alone τB led to insolubility of τ138, this subunit was recovered by co-infecting τB and τA. Pulling the complex down via His beads led to a non-stoichiometric complex that showed an over-expression of τA subunits, as previously observed for the human TFIIIC complex. Pulling the co-infected complex down with Strep beads led to a stoichiometric six subunit holo-TFIIIC complex.

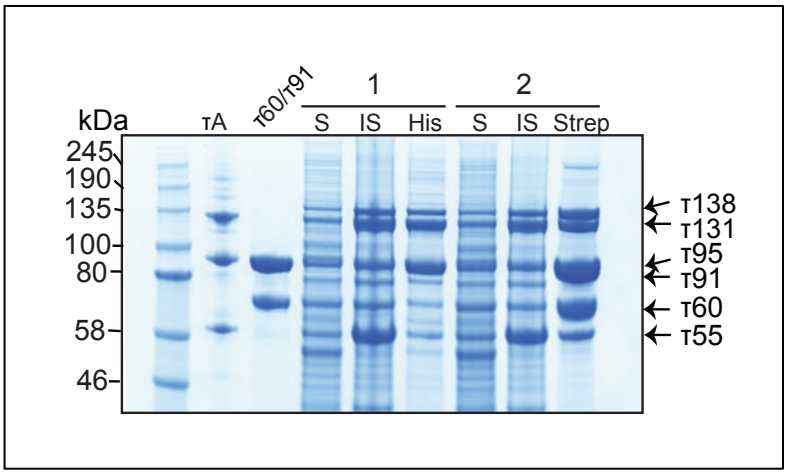


Figure 59: Holo-TFIIIC reconstitution using a co-infection.

A: Pulldown of τA and τB co-infection of High5 cells. S = soluble, IS = insoluble, His = Ni-NTA resin, Strep = Strep resin.

These results suggested that binding of τA to τ138 was required for it to remain soluble. Indeed, the interactions of τ138 with τA seemed significantly more important for the stability of the protein than its interaction with τ60/τ91. However, τ60/τ91 still appeared to play an important role in stabilizing the holo-TFIIIC complex as no degradation of τ138 was observed in the SDS-PAGE of the holo-TFIIIC compared to the degradation observed when only the four-subunit complex was expressed (Figure 57C). The stable heterodimer of τ60/τ91 was able to bind the four-

subunit TFIIIC complex to form a holo-TFIIIC complex. Furthermore, it suggested that the interaction between τA and τB was mainly driven by the interaction between τ138 and τA. Indeed, τ138 required τA to be soluble and could not be purified without the presence of τA. τ60/τ91 could play an important role in stabilizing the holo-TFIIIC complex. Without the presence of τ60/τ91, the formation of a four-subunit complex was possible but showed higher insolubility. From this it was hypothesized that separating TFIIIC into its two subcomplexes τA and τB needs to be reconsidered as other sub-assemblies of the complex are more stable than isolated τB. As it was not possible to purify the human or the yeast τB subcomplex in isolation, it may imply that inside the cell, τB is only ever present together with τA and that these complexes can't be regarded as two individual entities but rather only as a whole complex in which both subcomplexes function together. It was thus decided to proceed with reconstitution of the full TFIIIC complex from a single baculovirus infection.

3.2.1.4.3. Holo-TFIIIC purification from one baculovirus

To express the holo-TFIIIC complex, all six subunits with a twin Strep-tag at τ91 were cloned into a pBIG2ab vector (Dr Erin Cutts), using the Gibson Assembly. From this construct (Table 6 #25), a baculovirus was generated via transposition. TFIIIC was expressed for three days in 500 ml of High5 cells. The cells were harvested and a purification protocol was established. Initially, the complex was purified via a Strep Trap HP column as reported before. The Strep-tag enabled pulling down large amounts of high purity material (Figure 60A). In contrast to the human TFIIIC complex, the complex bound with high specificity and affinity to the Strep Trap HP column. The Strep-tag at the TFIIIC110 C-terminus, which is the functional homologue of the yeast τ91 subunit, was unable to bind to the

Strep column and the Strep tag at the C-terminus of TFIIIC90 could recover some protein bound to the Strep resin. This suggested that the folding of the TFIIIC110 and the τ91 subunits is different, indicating that the C-terminus of TFIIIC110 in humans was not as accessible as the C-terminus of τ91 in yeast.

After the affinity purification, the sample was loaded onto a Hi Trap Heparin column, to remove the DNA contamination, and subsequently loaded onto a custom made Superose 6 16 70 column to separate aggregates from the protein complex. It was then injected onto a Superose 6 Increase 10/300 GL column to perform an analytical comparison of the running profiles of this and previous TFIIIC subcomplexes. TFIIIC eluted in a single peak without any significant aggregation (Figure 60B). All subunits of the complex could be detected in the SDS-PAGE in a stoichiometric ratio, assuming that τ95 and τ91 ran at the same position in the gel. After the final gel filtration chromatography, an average of 6 -20 mg of the yeast TFIIIC complex was obtained in high purity from a 500 ml culture of High5 cells. The difference in yield was most likely due to variability of expression in High5 cells.

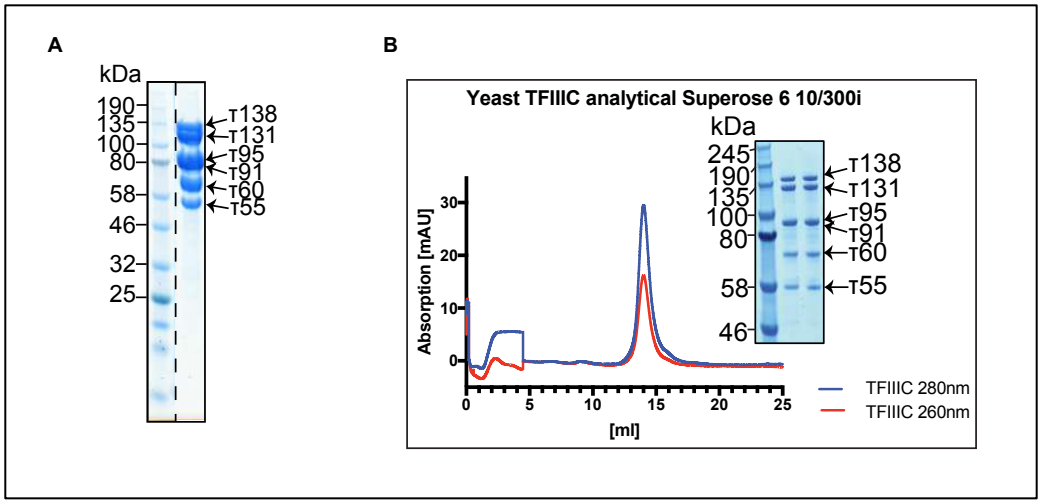


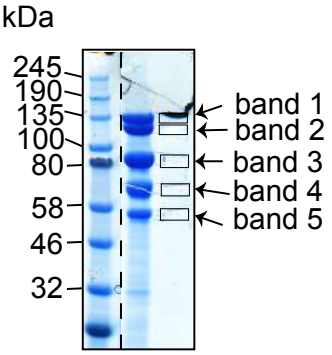
Figure 60: Yeast holo-TFIIIC reconstitution.

A: SDS-PAGE of the Strep pulldown of yeast TFIIIC. B: Gel elution profile after a three-step purification showed a single, Gaussian-shaped peak at 14 ml.

Before proceeding further with this complex, the individual bands were cut and submitted for LC-MS/MS analysis at the UCL School of Pharmacy. This aimed to identify the presence of all TFIIIC subunits, including the co-localization of τ95 and τ91(Figure 61A).

The LC-MS/MS analysis (Figure 61B) revealed that band 1 contained TFC3 (τ138), band 2 contained some degradation from TFC3 *(*τ138) and TFC4 (τ131) as well as AGE2, an ADP ribosylation protein. Band 3 was shown to contain TFC1 (τ95) and TFC6 (τ91), thus confirming that both subunits were present in this band in the SDS-PAGE. Band 4 contained TFC8 (τ60) and band 5 contained TFC7 (τ55) (Figure 61B). Thus, the presence of all six subunits as well as their position in the SDS-PAGE could be confirmed.

A



B

band	identified proteins	protein score
1	TF3C3_YEAST	505
2	TFC4_YEAST TFC3_YEAST AGE2_YEAST	435 155 32
3	TFC1_YEAST TFC6_YEAST	144 50
4	TFC8_YEAST	120
5	TFC7_YEAST GIR2_YEAST	67 32

Figure 61: Yeast holo-TFIIIC LC-MS/MS analysis.

A: SDS-PAGE indicating the position of the different bands that were cut and submitted for LC-MS/MS analysis. B: Results of the LC-MS/MS analysis of the individual bands performed by the UCL School of Pharmacy.

In summary, we were able to express and purify the yeast holo-TFIIIC complex in high purity and with a large yield after a three-step purification. This demonstrates a fundamental difference to the human TFIIIC

complex. Indeed, compared to the human complex, yeast TFIIIC was expressed to higher levels, bound with high affinity to the Strep column and could be frozen and thawed without any reduction in the quality of the sample. The purified protein assemblies were characterized further and described in the next sections.

3.2.2. DNA binding studies of TFIIIC

The following experiments were designed to study the individual DNA-binding ability of τA, the other sub-assemblies and the holo-TFIIIC complex. Testing the DNA binding ability was one way to assess whether the purified TFIIIC complexes were functional *in vitro*. The main questions addressed in this section were to establish what contribution each of these subcomplexes makes towards the DNA binding and whether two DNA binding events could be observed, as previously detected for the human TFIIIC complex. This could potentially hint to a conserved DNA-binding behavior.

3.2.2.1. DNA binding studies of yeast τA, τB and the four-subunit TFIIIC

The ability of TFIIIC to bind to the DNA was tested in EMSAs. First, τA was titrated on 8 pmol of DNA (tD(GUC)K_noTATA_short_A_Box) (Table 15) and incubated for 30 minutes at room temperature using interaction buffer D. As a control, holo-TFIIIC was titrated on the same DNA construct using the same conditions. All fractions were loaded onto a commercial 4-12% TBE Native PAGE. In Figure 62 lanes 5-7, we observed that binding of τA to the DNA led to a double DNA-binding shift. A ratio of 2.5 molar excess of τA over the DNA was almost enough to ensure that no free DNA was present anymore (lane 7). At a molar excess of 5x τA over the DNA, the lower DNA shift disappeared and only the higher band was

observed. The lower DNA-binding shift appeared not to represent a stable τA-DNA-binding since it did not show a sharp shift. This suggested that the τA-DNA-binding is an on-off event rather than a stable binding event.

In the controls, where TFIIIC was incubated with the DNA (Figure 63, lanes 2-4), a similar binding pattern could be observed. This may imply that maybe τB dissociated from τA once bound to the DNA as TFIIIC is not able to form a stable DNA bound complex where only an A Box was present. It could also be observed that TFIIIC formed two DNA-binding events as seen previously for the human TFIIIC complex when binding to the DNA. A molar excess of 5x TFIIIC over the DNA also led to the disappearance of the lower DNA shift (Figure 63, lane 4). The native page originally stained with ethidium bromide was later stained with Coomassie Blue to detect the proteins. The patterns were identical, which meant that each of the DNA-binding shifts in fact contained not only DNA but also protein.

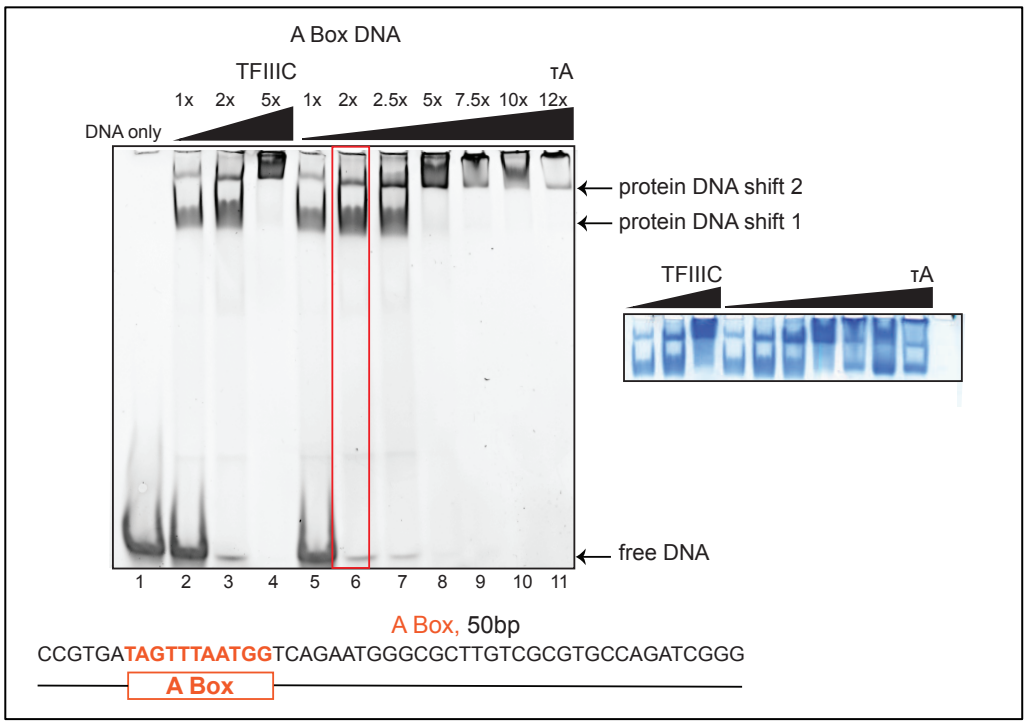


Figure 62 EMSA τA and TFIIIC with an A Box DNA construct.

A: Left: Native 4-12% TBE PAGE stained with ethidium bromide. The molar excess of protein over the DNA is indicated. The DNA construct used is shown below the EMSA as a DNA sequence and additionally as a schematic. Right: The same native PAGE stained with Coomassie Blue.

Next, τ60/τ91 were titrated on 8 pmol of DNA containing only the B Box (tD(GUC)K_noTATA_short_B_Box) (Table 15) and incubated for 30 minutes at room temperature using the same binding conditions as in the previous experiment. As a control, τA and holo-TFIIIC were incubated with the DNA construct. The samples were loaded onto a 4-12% TBE Native PAGE and stained with ethidium bromide (Figure 63, left) as well as Coomassie Blue (Figure 63, right).

Titrating TFIIIC over the DNA led to a double binding event to the DNA similar to that observed in the τA binding experiment (Figure 62) but the lower band had sharper features (Figure 63, lanes 2-3). This might imply that the binding to the B Box DNA was more stable than the binding to the A Box DNA. τ60/τ91 was able to bind nonspecifically to the DNA

construct. This could be due to the ability of τ91 to only assist τ138 to bind to the B Box DNA as previously reported (Arrebola et al. 1998). In the absence of the τ138 subunit, τB is not expected to bind to the B Box in a stable manner. τA was not able to bind to the DNA construct containing only a B Box (Figure 63, red rectangle). This suggested that the binding of τA to the A Box, shown in Figure 62, was a specific binding and did not require the presence of the B Box. As τA was not able to bind to the B Box DNA construct and also holo-TFIIIC was able to bind this DNA construct on its own, suggested that either τB could specifically bind to the B Box DNA or that τA, in combination with τB, was able to contribute towards the binding to the B Box.

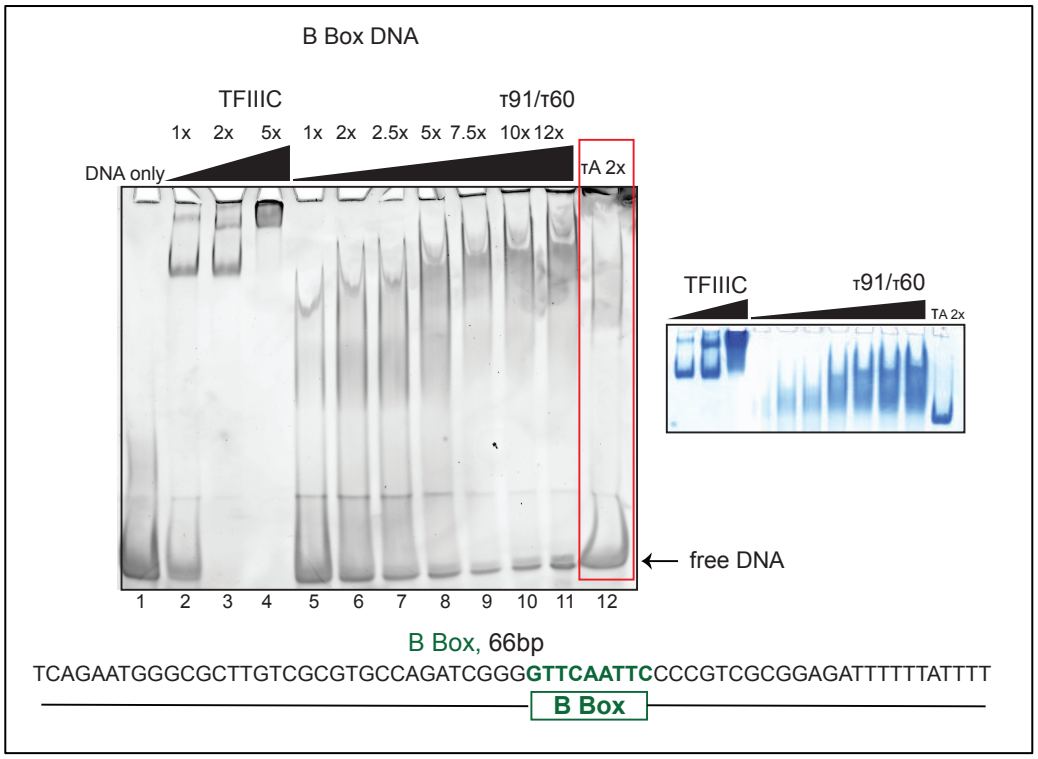


Figure 63: EMSA τB and TFIIIC with a B Box DNA construct.

Left: Native 4-12% TBE PAGE stained with ethidium bromide. The molar excess of protein over the DNA is indicated. The DNA construct used is shown below the EMSA as a DNA sequence and additionally as a schematic. Right: The same native PAGE stained with Coomassie Blue. The 2x molar excess of τA over the DNA for the A Box and the B Box DNA construct is marked in red.

In summary these experiments suggested that the binding of TFIIIC via the A Box was less stable than binding the complex to the DNA via the B Box. The τ60/τ91 heterodimer of τB was not able to form a stable complex bound to the DNA in the absence of τ138 in the tested conditions. A two-fold molar excess of τA over the A Box DNA led to two DNA-binding events, whereas a two-fold molar ratio of τA over the B Box did not give rise to any binding event (Figure 62 and Figure 63 marked with red rectangles). For future studies, if a successful optimization of the τB purification is achieved, the binding of this complex to the A Box would be of interest. Testing lower salt conditions for the DNA-binding reaction and trying other techniques such as fluorescence polarization might be beneficial in the future to be able to exclude a binding of τ60/τ91 to the DNA.

3.2.2.2. DNA binding studies of the yeast holo-TFIIIC complex

After studying the binding of the subcomplexes with the DNA individually, the formation of the TFIIIC-DNA complex was next investigated. Therefore, TFIIIC was initially titrated over 500 nmol of Aspartate tRNA (tD(GUC)K_Ultramer) (Table 15) with molar ratios of 1:1, 1:2, 1:4 and 1:8. The long 157 bp DNA construct, containing the TATA Box and the Terminator sequence (Figure 64A) was chosen to enable later assemblies of the full PIC containing the RNA Polymerase III, TFIIIB and TFIIIC. Interaction buffer A, containing 500 mM NaCl, was used initially to prevent nonspecific interactions. TFIIIC was incubated with the DNA for 30 minutes at room temperature and then analyzed via a 1x TBE 2% agarose gel (Figure 64B) and a 4% acrylamide 0.5x TBE native PAGE (Figure 64C), both stained with ethidium bromide.

In the EMSAs (Figure 64), we observed that the yeast TFIIIC complex was able to bind to the DNA. Free DNA was no longer present at a 1:4

molar ratio (DNA:TFIIIC). Increasing concentrations of TFIIIC, caused a further shift of the DNA-binding complex, most likely due to electrostatic effects. Whereas in the agarose EMSA (Figure 64B), higher concentrations of yeast TFIIIC still formed a single DNA band shift, the higher resolution of the 4% acrylamide EMSA (Figure 64C) allowed visualization of several binding events, as detected previously for the DNA binding of human TFIIIC. As previously observed, a higher concentration of TFIIIC led to aggregation inside the well, which could imply the formation of oligomers/aggregates.

The ability of TFIIIC to bind to the DNA was not disrupted at 500 mM NaCl. Indeed, preparing the interaction buffer A (section 2.4.5) with 1 M NaCl or 500 mM Ammonium sulfate instead, did not disrupt the binding of TFIIIC to the DNA either (data not shown). Higher amounts of salt led to increased aggregation and did not prevent the formation of several DNA-binding events.

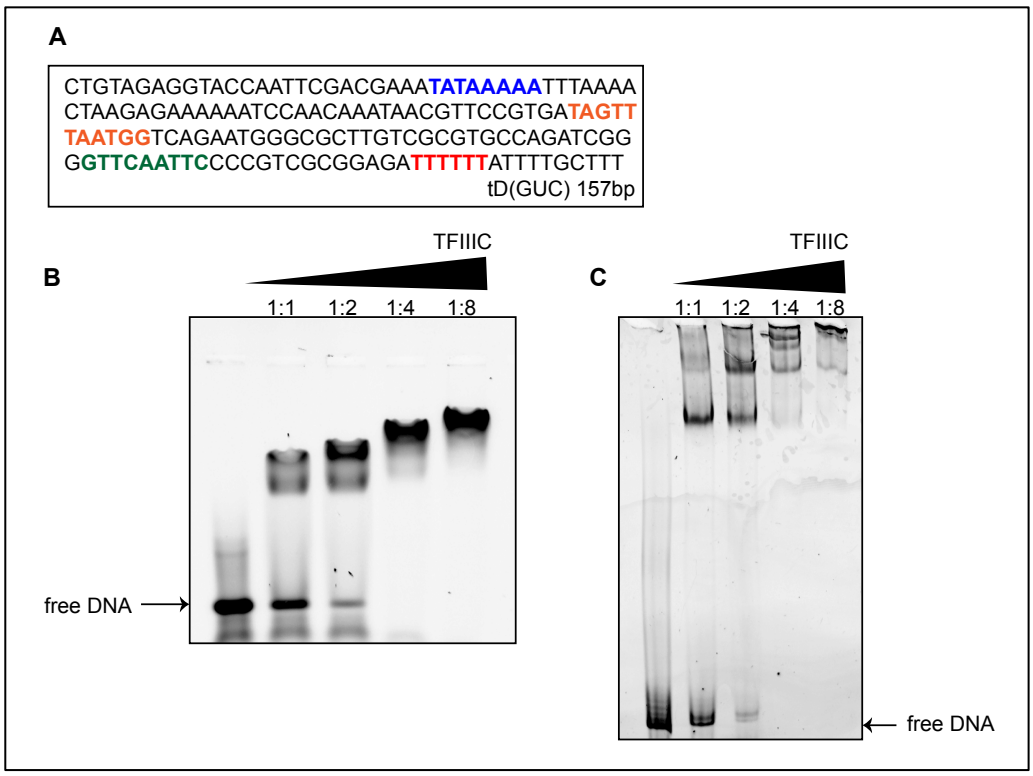


Figure 64: TFIIIC-DNA-binding studies.

A: tD(GUC) DNA sequence. In blue: TATA Box, in orange: A Box, in green: B Box, in red: terminator sequence. B: 1x TBE 2% agarose native PAGE stained with ethidium bromide. TFIIIC was titrated in ratios from 1:1, 1:2, 1:4 and 1:8 over the DNA oligo shown in A. C: 4% acrylamide 0.5x TBE native PAGE stained with ethidium bromide. TFIIIC was titrated in the same ratios over the DNA as in A.

By using high salt in the interaction buffer, it was excluded that the TFIIIC-DNA complex formation was nonspecific. However, high salt buffers can be problematic during structural analysis. In prospect of future electron microscopy structural studies, it was thus tested whether the TFIIIC-DNA complex was also formed in low salt using interaction buffer D with 150 mM NaCl. Furthermore, a shorter version of the tD(GUC)K DNA construct was tested that did not contain the TATA Box and started 6 bp upstream of the A Box (tD(GUC)K_noTATA_short) (Table 15) (Figure 65A). Using a short DNA construct with only 83 bp would facilitate the structural biology analysis.

We first assessed, whether TFIIIC was able to bind to the shorter DNA construct in the low salt buffer. Therefore, 500 nmol of the DNA construct was incubated with 1.75 μm of TFIIIC for 30 minutes at room temperature and analyzed on a 2% agarose 1x TBE gel stained with ethidium bromide (Figure 65B). This confirmed that the DNA construct was suitable for TFIIIC binding studies and that the interaction buffer was appropriate.

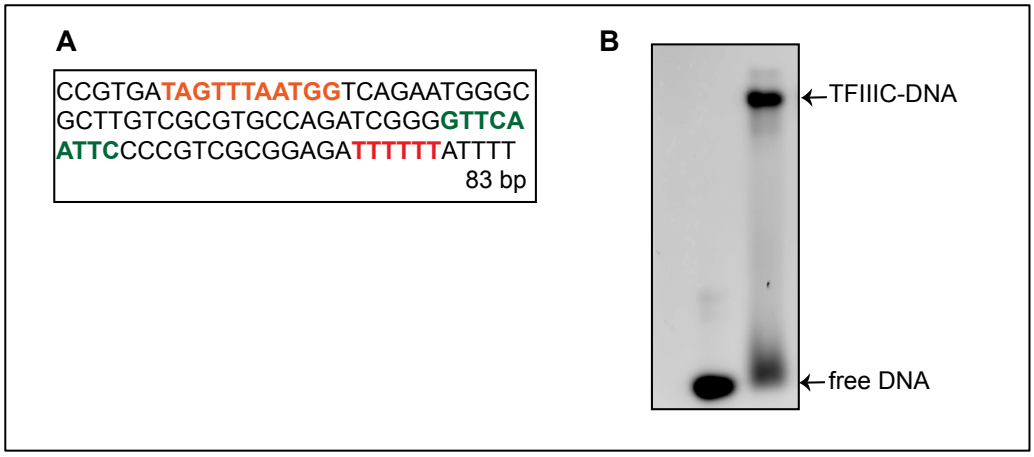


Figure 65: TFIIIC-DNA-binding to short tD(GUC)K DNA construct (83 bp).
A: tD(GUC)K DNA sequence with the A Box (orange), B Box (green) and the terminator (red). B: Control 2% agarose 1x TBE native PAGE of TFIIIC binding to this DNA construct.

As TFIIIC was able to bind to this shorter DNA construct using a low salt buffer, the effect the DNA binding had on TFIIIC regarding its DNA-bound conformation was next investigated. Therefore, a 3.5 molar excess of TFIIIC (225 μg, 434 pmol) were mixed with 123 pmol of the short tD(GUC)K construct in interaction buffer D, containing only 5 % glycerol compared to the 10 % glycerol that was used previously (Figure 65). The reaction was incubated for 30 minutes at room temperature. The sample was subsequently injected onto a Superose 6 Increase 10/300 GL column and run at room temperature in order to establish whether the TFIIIC-DNA complex showed a different elution pattern compared to apo TFIIIC.

The TFIIIC-DNA sample eluted in two distinct peaks, one at 12 ml and a second at 14 ml (Figure 66A, green chromatogram). As a control, apo

TFIIIC (434 pmol) was injected onto the same column (Figure 66A, blue chromatogram), as an additional control, 123 pmol DNA (Figure 66A, red chromatogram) was run. Apo TFIIIC eluted, as previously observed, in a single peak at 14 ml and the free DNA eluted later, at approximately 15 ml. To confirm whether the DNA binding caused TFIIIC to elute earlier in the gel filtration column, the peak fractions were analyzed on an SDS-PAGE (Figure 66B). The fractions at the volume where peak 1 eluted did not show any TFIIIC subunits in the apo TFIIIC chromatogram. In the TFIIIC/DNA chromatogram, TFIIIC could be detected at this position in the gel filtration column. At peak 2, TFIIIC subunits could be detected in both cases since 14 ml is the position where apo TFIIIC elutes. This most likely represented an excess of apo TFIIIC in the TFIIIC-DNA experiment. These observations suggested that the DNA-binding may have caused either an extension of the TFIIIC complex that shifted the elution peak from 14 ml to 12 ml or that the length of the bound DNA itself could have caused the shift. Analyzing the different fractions of peak 1 and peak 2 on a 4-12% TBE native PAGE, stained with ethidium bromide and Coomassie Blue (Figure 66C), showed that the TFIIIC-DNA complex mainly eluted in peak 1.

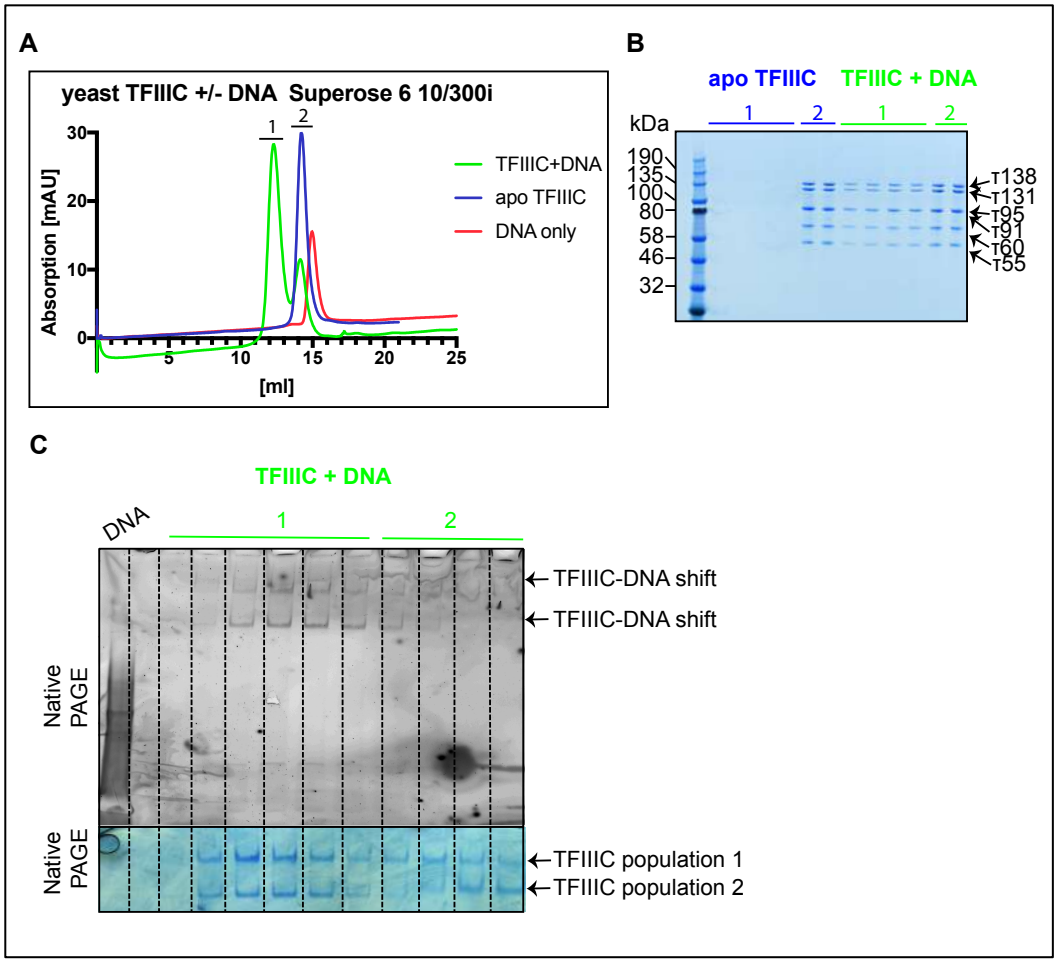


Figure 66: TFIIIC DNA interaction studies using a Superose 6 Increase 10/300 GL column.

A: Superose 6 Increase 10/300 GL chromatogram of TFIIIC+DNA (green), apo TFIIIC (blue) and the DNA (red) absorption at 280 nm. B: SDS-PAGE of fractions from peaks 1 and 2 from A. C: Native-PAGE showing fractions from peak 1 and peak 2 from A. Top: Native PAGE stained with ethidium bromide. Bottom: Stained with Coomassie Blue.

The fractions in peak 1 showed a signal when stained with ethidium bromide which confirmed the presence of DNA. Furthermore, the Coomassie staining also confirmed the presence of protein where the DNA bands were detected. Peak 2, which was assumed to be the apo TFIIIC peak, only showed a weak signal for the DNA bound complex when stained with ethidium bromide which was most likely due to a slight overlap of peak 1 into peak 2. When stained with Coomassie Blue, the native PAGE of peak 2 showed two bands which indicated the presence

of protein inside this peak. The resolution of the native PAGE (Figure 66C) was higher than the agarose native PAGE (Figure 65B) and showed two distinct DNA-binding events. The two TFIIIC-DNA shifts may be caused by a differential phosphorylation status, the formation of oligomers or by two different conformations of TFIIIC. The origin of the two DNA bands was studied further and is discussed in the following sections.

In summary, it was observed that the TFIIIC-DNA complex showed a different elution pattern compared to the apo TFIIIC complex. This could be due to an extension of TFIIIC upon binding to the DNA, as reported previously (Kassavetis et al. 1989). It was also detected that two DNA-binding events of TFIIIC took place, which could be a conserved behavior between yeast and human TFIIIC. Further investigation will be required to clarify why two binding events were observed.

3.2.3. Biochemical and biophysical characterization of the yeast TFIIIC complex

The purification of a large amount of pure yeast TFIIIC complex allowed extensive characterization of this fairly uncharacterized complex both biochemically and biophysically by using SEC-MALS, SEC-SAXS and native mass spectrometry analysis. The main questions addressed in this section were to understand how TFIIIC behaves in solution, whether TFIIIC was present as a monomer and in one specific conformation or whether it was subject to movements. The hypothesis that TFIIIC was a flexible complex that could potentially adopt several conformations was previously reported (Schultz et al. 1989; Liao et al. 2006; Marck et al. 2006; Nagarajavel et al. 2013). Besides that, we wanted to analyze how the different subunits are connected with each other to understand why a four-subunit complex was observed before.

3.2.3.1. SEC-MALS analysis of apo TFIIIC

After confirming the ability of TFIIIC to bind to the DNA and before moving into structural studies, it was assessed whether TFIIIC was a monodisperse protein sample. Therefore, SEC-MALS experiments in a concentration series of TFIIIC of 0.5 mg/ml, 1 mg/ml ,2 mg/ml and 4 mg/ml were performed together with Dr. Ian Taylor at the Francis Crick Institute.

The results of the SEC-MALS analysis revealed that TFIIIC was a monodisperse complex. With decreasing concentration, the molecular weight and the hydrodynamic radius (Rh(Q)w) also decreased which was likely due to technical variability (Figure 67A, B). The results at 0.5 mg/ml were beyond the lower concentration limit of the SEC-MALS and thus needed to be treated with caution. Besides the main peak, a smaller peak could be detected, (peak 2). This peak appeared to contain sub-assemblies of TFIIIC. However, it needs to be noted at this point that the ratio between peak 2 and peak 1 did not change depending on the concentration. A lower concentration did not lead to an increase in sub-assemblies. This suggested that the complex did not fall apart but rather that a second sub-population was purified together with holo-TFIIIC. This population could be separated out by gel filtration to avoid issues arising during structural studies.
More importantly, the SEC-MALS analysis also demonstrated that TFIIIC was present as a monomer in solution since the measured molecular weight of 473 - 509 kDa accounted for a monomeric TFIIIC.

A

concentration	Mw peak 1	Rh(Q)w peak 1	Mw peak 2	Rh(Q)w peak 2
4 mg/ml	509 kDa	8.25	208 kDa	3.77
2 mg/ml	499 kDa	8.01	208 kDa	2.42
1 mg/ml	473 kDa	8.09	187 kDa	1.98
0.5 mg/ml	447 kDa	6.45	135 kDa	2.29

B

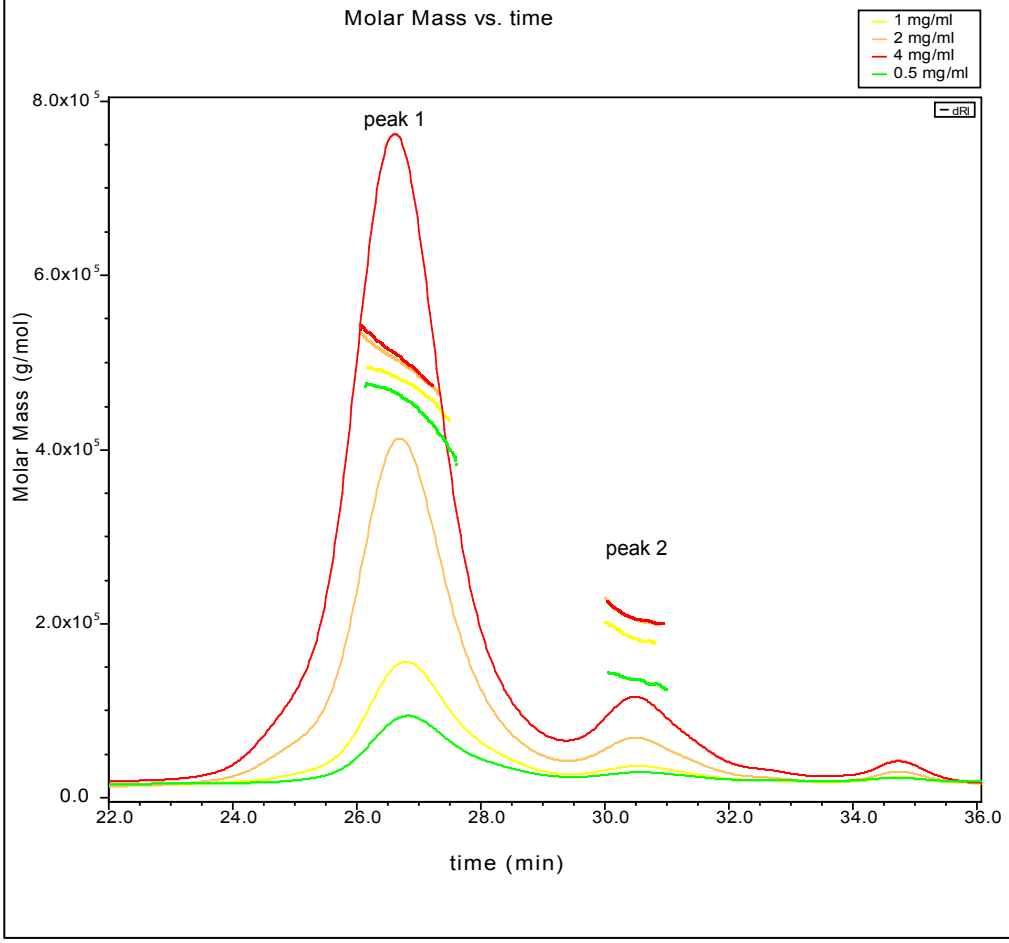


Figure 67: SEC-MALS of apo TFIIIC.

A: Table indicating the measured molecular weight (kDa) and the hydrodynamic radius measured by the integrated DLS per peak. B: Easy graph of the TFIIIC concentration series.

Following the SEC-MALS experiment, which revealed that TFIIIC was present as a monomer, it was tested whether the two bands, in which apo TFIIIC and DNA bound TFIIIC observed on native PAGE were a

concentration dependent effect. Therefore, 9 pmol of TFIIIC was loaded onto each well of a 4-12% TBE native PAGE in a concentration series from 5.75 mg/ml to 0.25 mg/ml and stained with Coomassie Blue.

In this experiment (Figure 68), we observed that decreasing the concentration of TFIIIC did not have an effect on the two different bands present in the native PAGE. If the top band was a concentration dependent oligomer, it should have decreased with the dilution. This therefore suggested that the two bands of TFIIIC were not a concentration dependent oligomerization product.

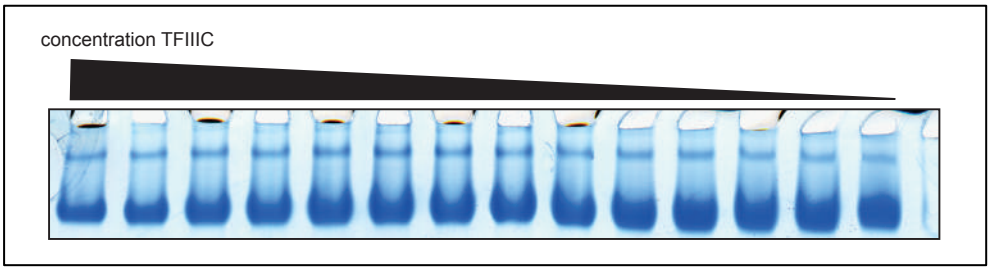


Figure 68: Concentration dependence of yeast TFIIIC.
9 pmol each of TFIIIC were loaded onto a 4-12% TBE native PAGE in a concentration series. The gel was stained with Coomassie Blue.

Together with the SEC-MALS results, this suggested that either phosphorylation or different conformations of the complex were responsible for the two different bands.

3.2.3.2. SEC-SAXS analysis of apo and DNA bound TFIIIC

After the SEC-MALS experiment was used to describe the TFIIIC complex hydrodynamically, a SEC-SAXS analysis of apo TFIIIC (LJ001) and DNA bound TFIIIC (LJ002) was used to analyze whether a conformational difference could be observed inside each sample. The two samples were

taken to the SAXS beamline by Dr. Nora Cronin at the Diamond Light Source Facility. The samples were run on a 4.3 ml Shodex HPLC column KW403 and the data was analyzed using the Scatter software developed by Robert Rambo.

Plotting the radius of gyration (Rg) into the SEC-SAXS signal plot (Figure 69, top) revealed that there were two populations present in the apo TFIIIC sample since the R_g plateaued at two distinct R_g values. These populations were selected and analyzed independently.
The scatter profile was calculated for each of the two populations in the scatter intensity plot where the log of the intensity of the X-Ray scatter log[I(q)} was represented against the scattering vector q ($Å^{-1}$) in reciprocal space (Figure 69A, Ai). The scattering intensity at low q values was strong and dropped quickly for higher q values, introducing high noise. The low q range is characteristic for the overall size and dimension of the particle and is called the Guinier region. The ln(I) of this region was plotted against the squared scattering vector q^2 ($Å^{-2}$), which led to a straight line in both cases. This indicated that both populations were monodisperse. The Guinier plots reported a R_g of 75.8 Å for the leading population and a R_g of 71.4 Å for the trailing population, indicating that the leading population was larger and potentially more elongated compared to the other population (Figure 69B, Bi).

The gyration radius versus molecular weight plot for globular proteins (Smilgies & Folta-Stogniew 2015), suggests a R_g of 55 - 65 Å for a protein complex with a molecular weight of 520 kDa. Observing R_g values of. 70 Å and larger could thus indicate that both populations of apo TFIIIC appear to be extended rather than fully globular.

The dimensionless Kratky plot analyzed the shape of the protein complex further (Figure 69C, Ci). The Kratky plot is sensitive to the compactness and morphology of the particle so that folded and unfolded states can be

distinguished. It is scaled using the previously obtained radius of gyration from the Guinier analysis. For monodisperse, globular proteins the top of the peak of the graph is expected to go through the Guinier-Kratky point ($\sqrt{3}$, 1.1) of the plot and back down to the baseline. An unfolded protein on the contrary, would show a graph that is raising over time and not dropping back down to the baseline. The course of the Kratky plots obtained from the SEC-SAXS analysis indicated that both TFIIIC populations were indeed folded. The shift of the top of the peak away from the Guinier-Kratky point could indicate an extended state, as previously hypothesized (Durand et al. 2010). Indeed, it could be noted that the Kratky plot of the leading TFIIIC population was further away from the Guinier-Kratly point than the trailing TFIIIC population for which a smaller R_g was obtained.

The flexibility of the two apo TFIIIC populations was next analyzed using the Porod's law, which can be applied to the high q values of the scattering curve (Figure 69D, Di). Thereby, information regarding the surface per volume can be obtained. The first plateau region was obtained for q^{-4} in the Porod Debye plot, obtaining a Porod slope of 3.8 for the leading population and 3.7 for the trailing population. A plateau at q^{-4} indicates that a particle has a uniform density and globular overall shape. With the Porod slope between 3 to 4 in both populations, it indicated that TFIIIC did not have a perfectly uniform density (Boldon et al. 2015).

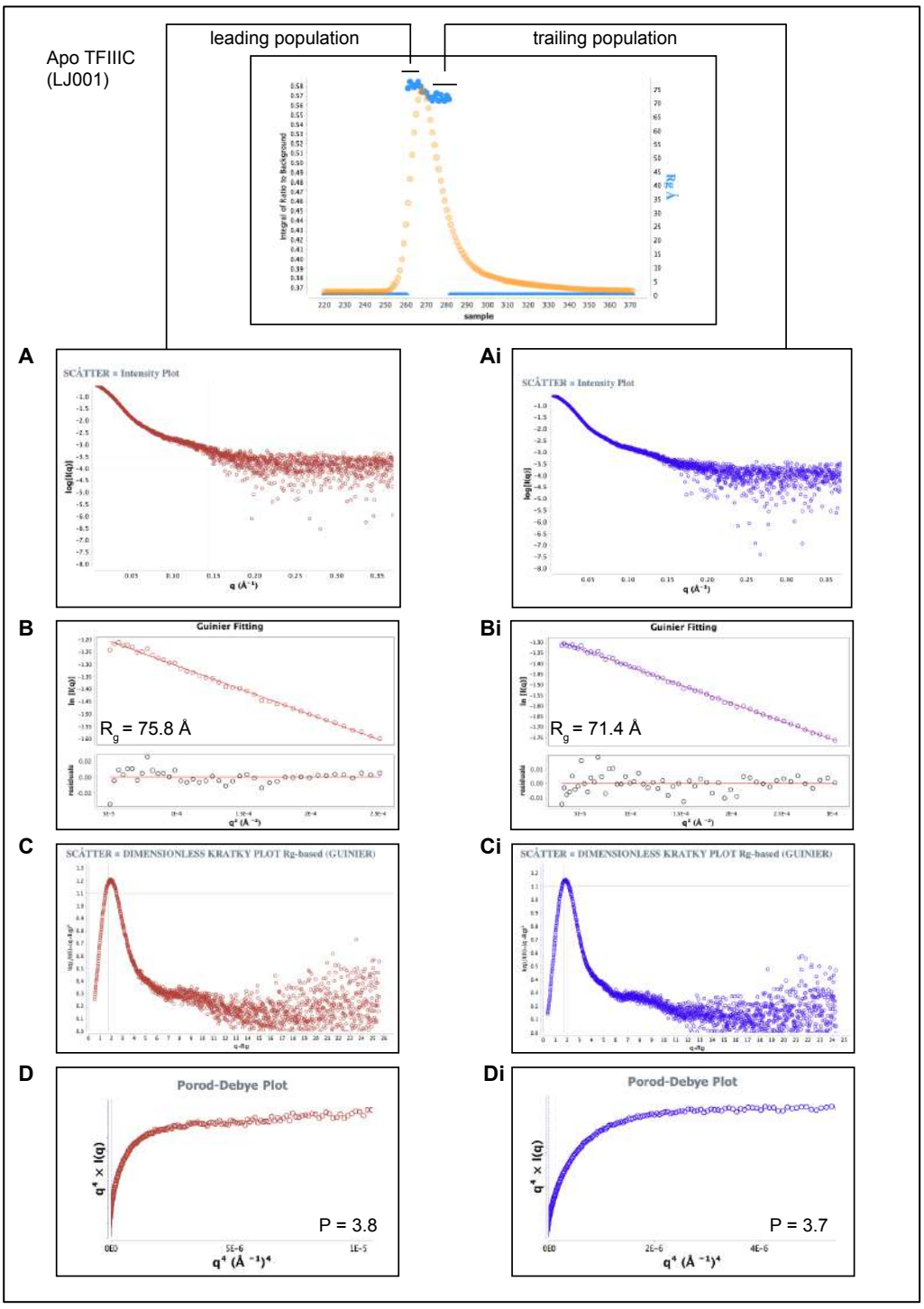


Figure 69: LJ001 SEC-SAXS analysis.

Top: SEC-SAXS signal plot that indicated the presence of two distinct populations that were analyzed independently. A: Scatter Intensity plot (Log[I(q)] vs scattering vector q ($Å^{-1}$)). B: Guinier plot (ln[I(q)] vs. q^2 ($Å^{-2}$)) C: Dimensionless Kratky plot $(I(q)/I(0)*(q*R_g)^2$ vs. qR_g) indicating a folded protein complex. D: Porod-Debye plot (q^4 x I(q) vs. q^4 $(Å^{-1})^4$).

In addition to information regarding the folding state, the radius of gyration and the compactness of a protein complex, further information could be extracted from the SAXS experiments by performing a Fourier Transform of the scattering profile. This allows to determine a range of intramolecular distances and their relative proportions in a pair distance distribution function P(r). Furthermore, the maximal dimension D_{max} of a protein can be estimated using the P(r) function (Mertens & Svergun 2010).

The P(r) distribution was calculated for each population independently using the Scatter Software (Figure 70A, D) obtaining a D_{max} of 320 Å for the leading population and 243.5 Å for the trailing population. Both P(r) distributions showed a Gaussian peak with an extended tail at larger dimensions. The tail was more pronounced for the leading population, suggesting that it contained a more extended domain attached to a globular protein complex compared to the trailing population. An inverse Fourier Transform of the P(r) showed a scattering profile (Figure 70B, E), which was similar to the experimentally obtained scattering profile.

As a final step, the SAXS data was subjected to an *ab initio* model building software, called DAMMIF (Dummy Atom Model Minimization Fast) (Franke & Svergun 2009) integrated into the Scatter Software. Using 10 iterations, a molmap of the final averaged DAMAVER bead model was generated at a resolution of 15 Å and is shown in Figure 70C, F.
Both models showed a triangular-like shape. The model of the leading population was significantly larger than the model of the trailing population and adopted a more extended conformation.

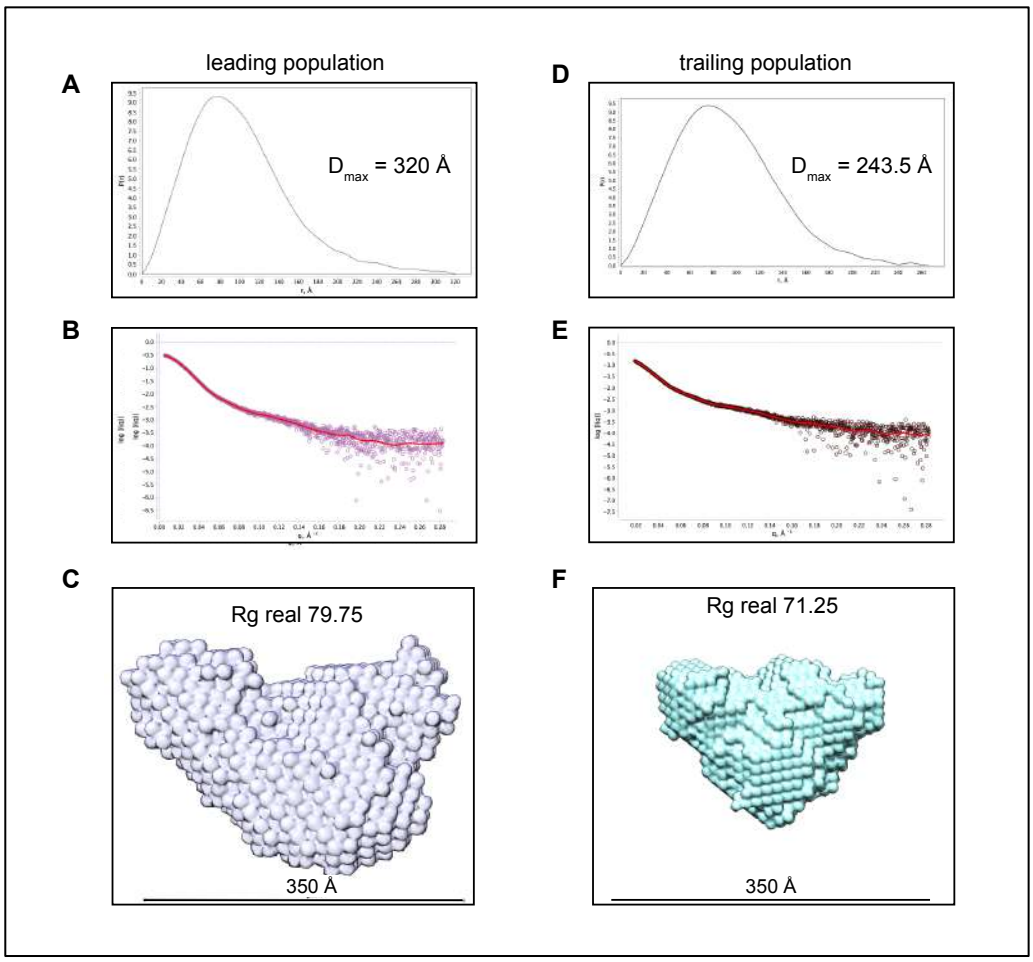


Figure 70: SAXS structural determination of apo TFIIIC.

A: P(r) distribution function of the leading population of apo TFIIIC, D: P(r) distribution function of the trailing population of apo TFIIIC (P(r) vs. r (Å)). B: Inverse Fourier Transform of the P(r) distribution of the leading apo TFIIIC, E: Inverse Fourier Transform of the P(r) distribution of the trailing apo TFIIIC population log [I(q)] vs. q ($Å^{-1}$)).C: DAMAVER molmap of the leading apo TFIIIC population. F: DAMAVER molmap of the trailing apo TFIIIC population.

The same SEC-SAXS analysis was performed with the DNA bound TFIIIC sample, called LJ002. The signal plot revealed that in the DNA bound TFIIIC sample also two distinct populations were present (Figure 71, top row). The two populations were again analyzed separately. In the region of low q values in the Scatter Intensity plots (Figure 71A, Ai), the Guinier region was identified by plotting the ln[I(q)] versus the squared scattering vector q^2 ($Å^{-2}$) in Figure 71B,Bi. The obtained R_g reported 100.8 Å for the leading population and 79.5 Å for the lagging TFIIIC population. The

leading population that most likely contained the DNA bound TFIIIC complex was thus significantly larger compared to the trailing population and suggested thus to be apo TFIIIC. This hypothesis was based on previous observations that suggested that the DNA-binding caused TFIIIC to elute in an earlier elution peak. Furthermore, the obtained R_g of the trailing population was in a similar size range compared to the leading population of apo TFIIIC whereas the leading population of DNA bound TFIIIC showed a significantly higher R_g value.

Analyzing the folding state with the Kratky plot revealed that the trailing population was completely folded, as observed for the apo TFIIIC population (Figure 71Ci) whereas the Kratky plot of the leading population showed a little shoulder instead of a direct progression back to the baseline (Figure 71C). A shoulder in a Kratky plot often indicates the presence of a multidomain protein. Thus, it was concluded that the DNA-binding forced TFIIIC into a conformation in which the globular protein shape was distributed into multidomains.

Both populations plateaued at the q^{-4} plot (Figure 71D, Di), obtaining Porod slopes of 3.7 for the leading population and 3.5 for the trailing population. Neither of these two populations thus seemed to contain a uniform density distribution.

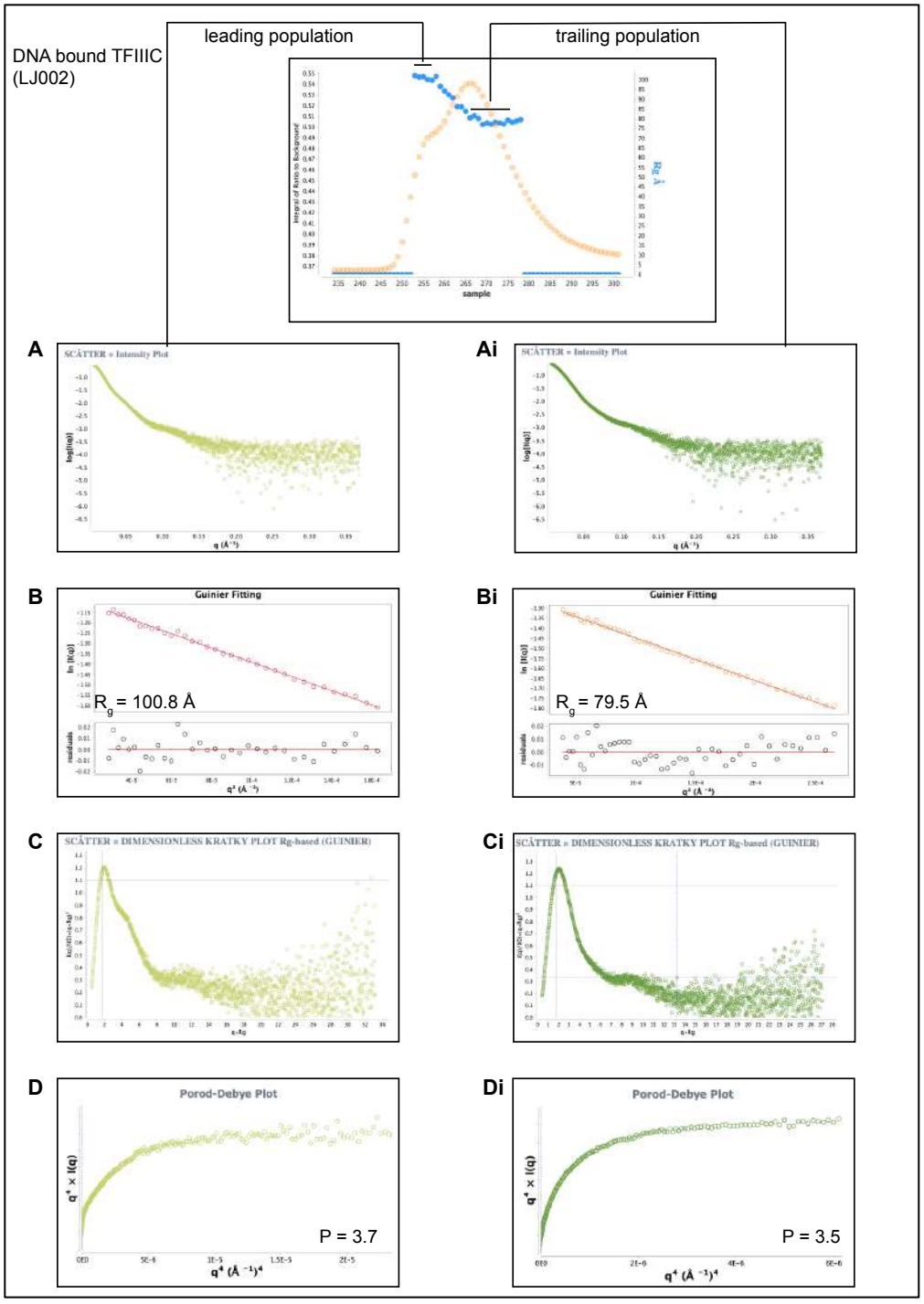


Figure 71: LJ002 SEC-SAXS analysis.

Top: SEC-SAXS signal plot that indicated the presence of two distinct populations that were analyzed independently. A: Scatter Intensity plot (Log[I(q)] vs scattering vector q ($Å^{-1}$)). B: Guinier plot (ln[I(q)] vs. q^2 ($Å^{-2}$)) C: Dimensionless Kratky plot ($I(q)/I(0)*(q*R_g)^2$ vs. qR_g) indicating a folded protein complex for the trailing population and a multidomain protein complex for the leading TFIIIC-DNA population. D: Porod-Debye plot (q^4 x I(q) vs. q^4 ($Å^{-1}$)^4).

Next, the P(r) distribution functions were analyzed of both populations (Figure 72A, D) whereby a D_{max} of 373 Å was obtained for the leading population and 331 Å for the trailing population. The D_{max} value of the trailing population was only 11 Å larger than the D_{max} obtained for the leading peak of the apo TFIIIC sample and suggested that three main conformational states of TFIIIC were evident in the SEC-SAXS experiments: the TFIIIC-DNA complex that had a D_{max} of 373 Å, an extended apo TFIIIC complex with a D_{max} in the range of 320 – 331 Å and a more compact apo TFIIIC complex with a D_{max} of 243.5 Å.

The inverse Fourier Transform of both P(r) distributions (Figure 72B, E) resembled the experimental data. Lastly, an *ab initio* model building using DAMMIF was performed for both populations using 10 iterations. This led to two averaged DAMAVER bead models of which molmaps at a resolution of 15 Å were generated (Figure 72C,F).

The map obtained for the trailing population showed similarities to the leading population of apo TFIIIC with two main domains on the right and left hand-side of the structure. The molmap of the DNA bound TFIIIC complex appeared to be enlarged and contained extra densities which could be due to a more open conformation or the incorporation of the DNA.

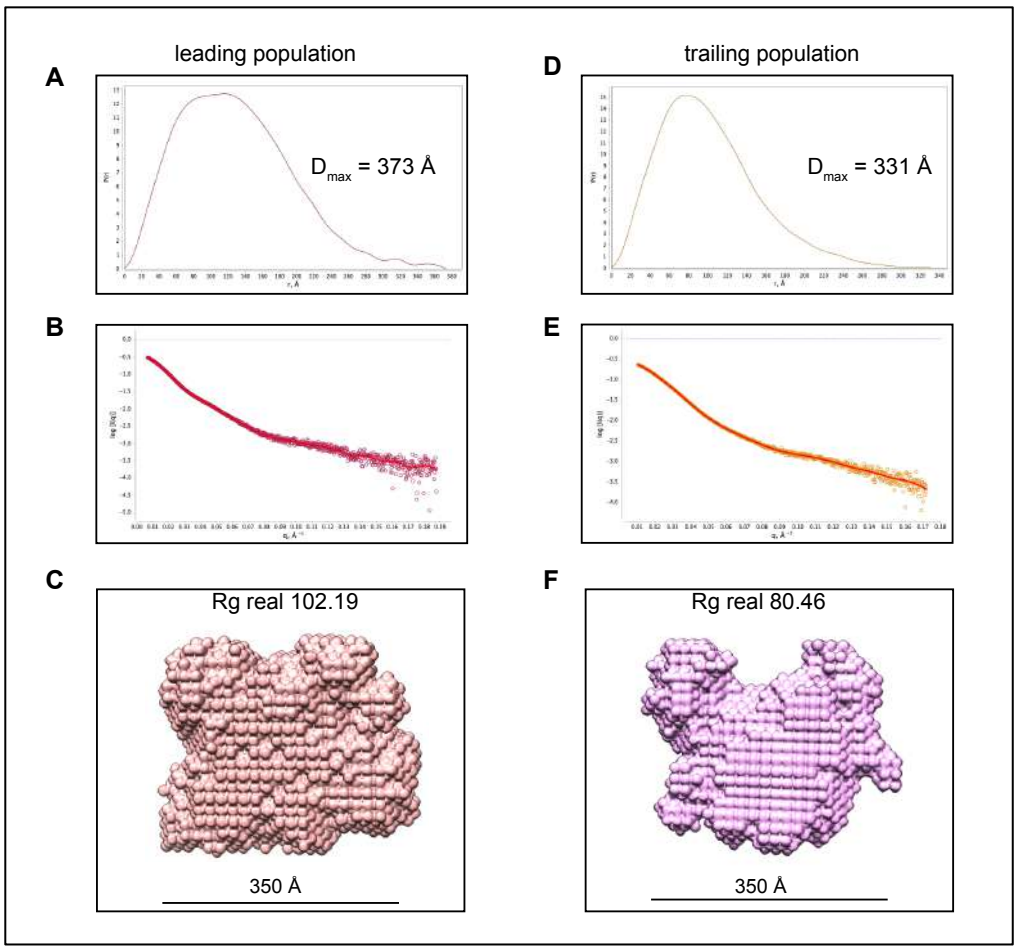


Figure 72: SAXS structural determination of DNA bound TFIIIC.

A: P(r) distribution function of the leading population of DNA bound TFIIIC, D: P(r) distribution function of the trailing population of TFIIIC (P(r) vs. r (Å)). B: Inverse Fourier Transform of the P(r) distribution of the leading DNA bound TFIIIC, E: Inverse Fourier Transform of the P(r) distribution of the trailing TFIIIC population log [I(q)] vs. q ($Å^{-1}$)).C: DAMAVER molmap of the leading DNA bound TFIIIC population. F: DAMAVER molmap of the trailing TFIIIC population.

In summary, the SAXS experiments suggested that apo TFIIIC was present in two conformations, a more elongated and a more compact form. Furthermore, the DNA bound TFIIIC sample contained apo TFIIIC as a second population which appeared to have adopted the more elongated conformation of apo TFIIIC rather than the compact conformation. It was also confirmed that TFIIIC was a properly folded protein complex and not aggregated or unfolded. The different conformations of TFIIIC could either be due to an auto-inhibited form as suggested in (Liao et al. 2006) or could be similar to the conformations

observed for TFIID, in which a canonical and a rearranged state were observed (Louder et al. 2016). In order to clarify on this, the observed different conformations of yeast TFIIIC further were studied further using a native mass spectrometry analysis was performed in collaboration with Prof. Perdita Barran's team at Manchester University.

3.2.3.3. Native Mass Spectrometry analysis of the apo TFIIIC complex

The apo yeast TFIIIC complex was sent for native mass spectrometry and ion mobility analysis. It was hoped these experiments would indicate the presence of different conformations, which subunits were more prone to dissociate from the complex and whether other sub-assemblies of TFIIIC could be observed.

3.2.3.3.1. Tandem LC-MS/MS analysis with increasing IST voltage

First, the effect of an increasing in-source trapping voltage (IST) onto the TFIIIC complex was studied in order to see which products could be detected. The IST was incrementally increased from 0V to 250 – 300V in total. Even before any IST voltage was applied at all, different populations of TFIIIC could be detected (Figure 73). As previously hypothesized, two conformations of TFIIIC, one elongated and one compact, could be identified because of their different charge to mass ratio. Interestingly, isolated τ60 could also be detected before the IST voltage was applied. This observation could explain the two different bands that were observed in native PAGEs for apo and DNA bound TFIIIC. A compact form of TFIIIC would be able to travel faster through a native PAGE compared to an elongated form, thus resulting in two different bands. However, it was not clear what drove the transition between those two conformations or whether there was a continuous transition.

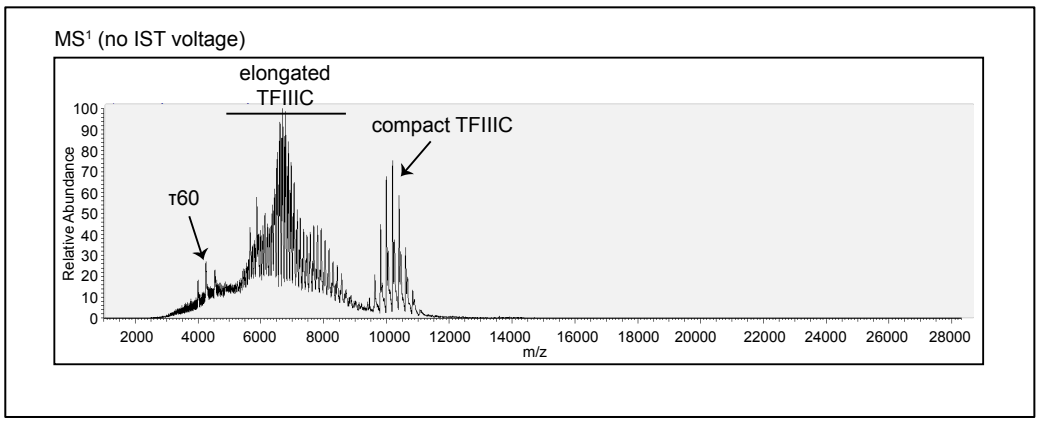
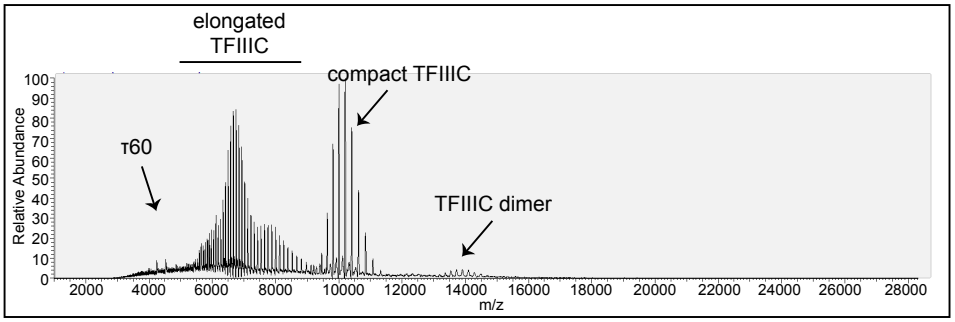


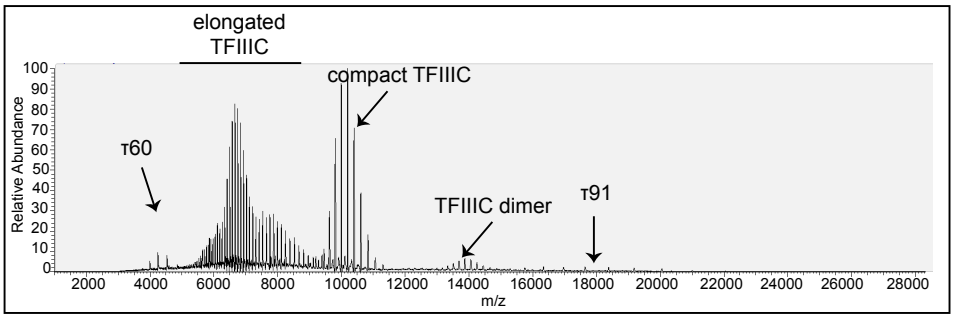
Figure 73: Native mass spectrometry analysis 0V IST.

Next, the IST voltage was progressively increased in steps of 50 V (Figure 74). With increasing IST voltage, the compact form of TFIIIC became more prominent compared to the elongated TFIIIC conformation. The TFIIIC dimers could be detected until a voltage of 200 V, but were not detected at higher voltages. Furthermore, besides being able to detect isolated τ60, increasing the IST voltage led to the presence of the heterodimer τ60/τ91 as well as the isolated subunit τ91.

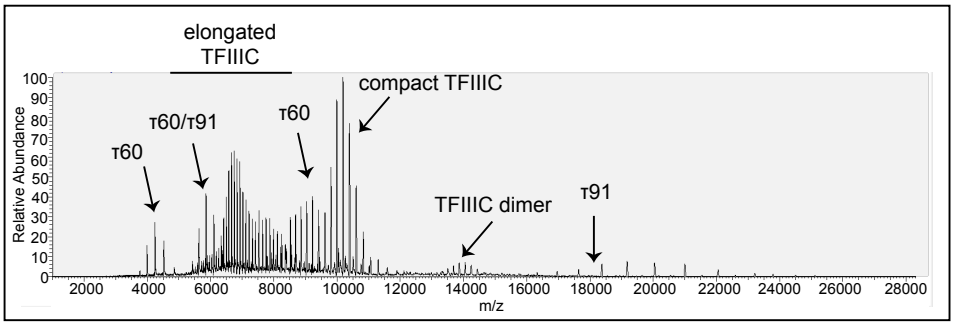
MS² (50V IST voltage)



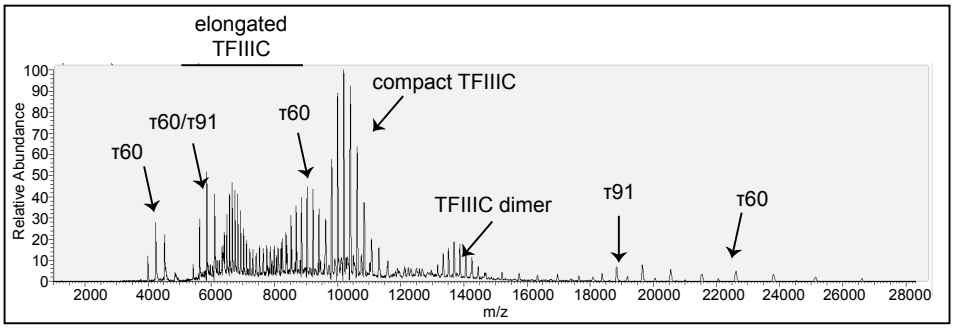
MS² (100V IST voltage)



MS² (150V IST voltage)



MS² (200V IST voltage)



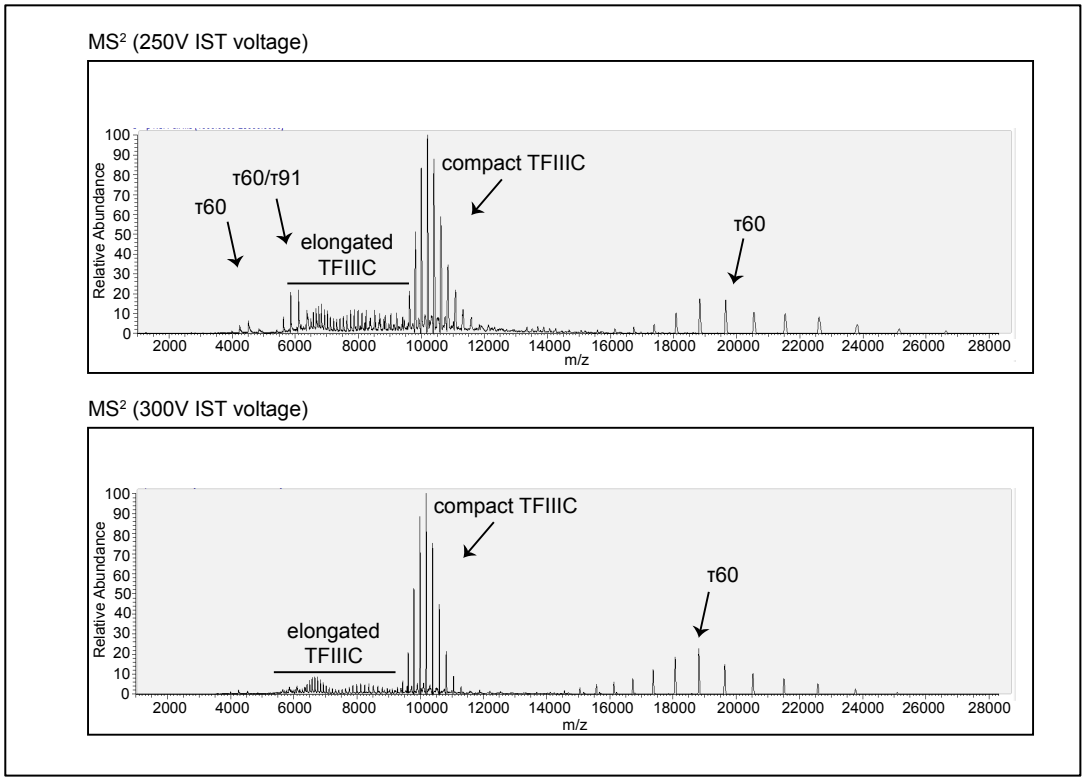


Figure 74: Native mass spectrometry analysis of apo TFIIIC with an incremental increase of the IST voltage from 50V to 300V in steps of 50V.

To study the elongated and the compact form of TFIIIC independently from each other, one charge state of the elongated TFIIIC conformation (6,692 m/z) and one charge state of the compact TFIIIC conformation (10,194 m/z) (Figure 75) were chosen.

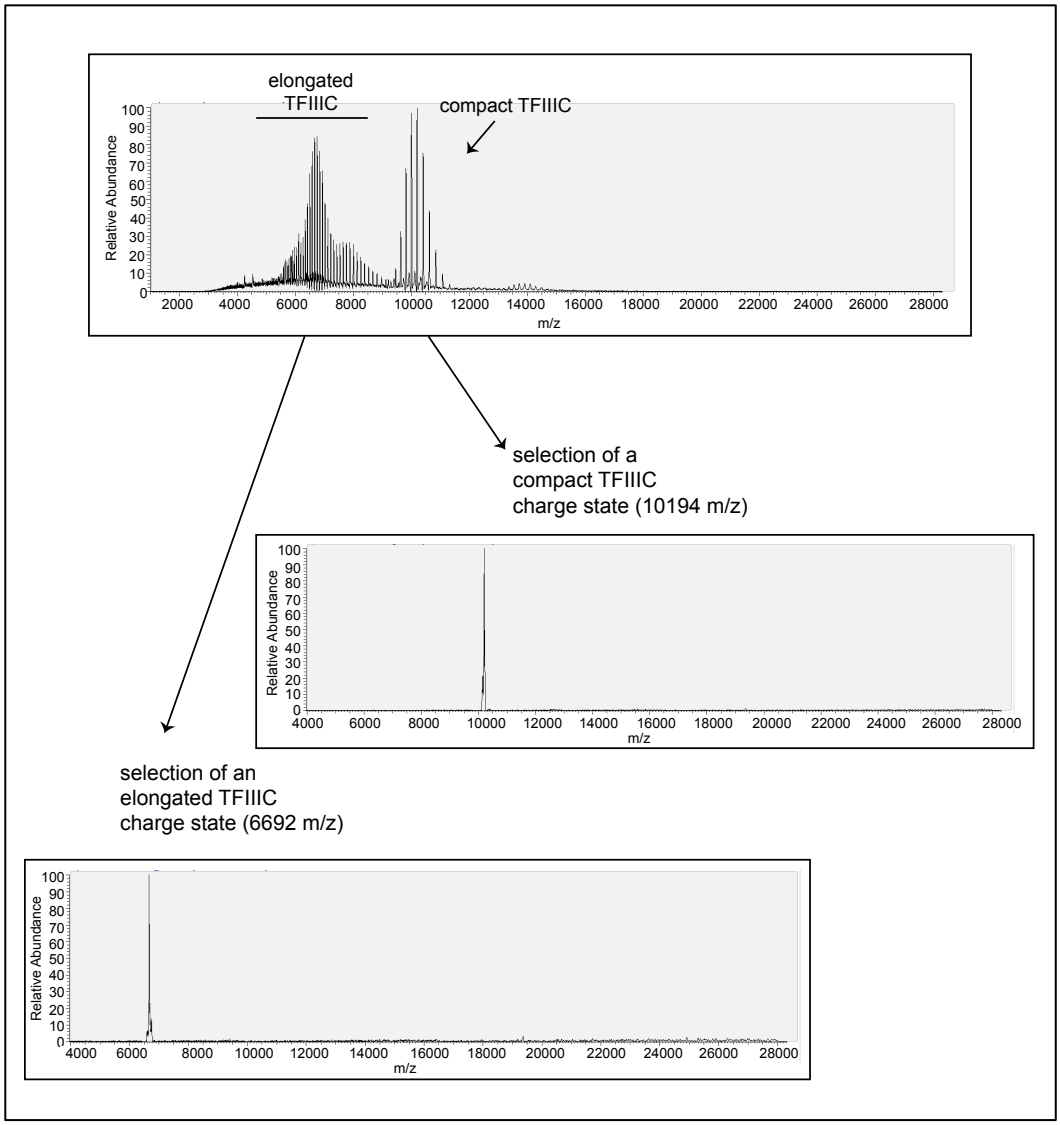


Figure 75: Different charge states of the elongated and compact TFIIIC complex analyzed by native mass spectrometry.

One charge state of the elongated and one charge state of the compact form of TFIIIC was chosen.

By analyzing exclusively the elongated conformation of TFIIIC (Figure 76), it was observed that τ60 got expelled at 150 CE. By further increasing the voltage up to 200 CE, τ91 was also expelled from TFIIIC. At 250 V, the amount of expelled τ60 and τ91 increased further.

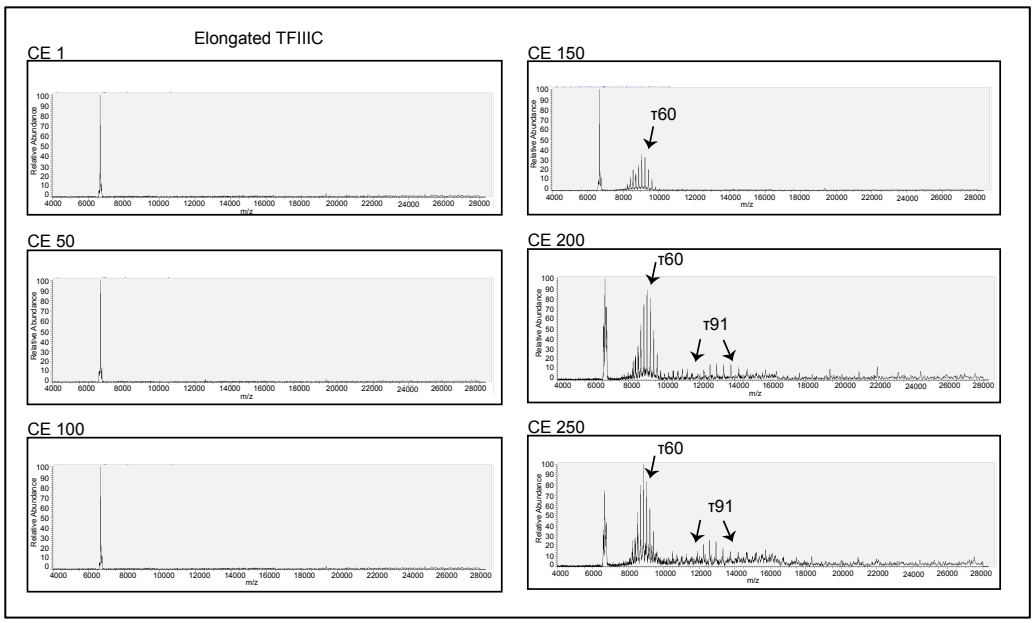


Figure 76: Native mass spectrometry analysis of the elongated TFIIIC complex.
The voltage was increased in increments of 50V.

After analyzing the elongated TFIIIC complex in detail, the compact conformation of TFIIIC was chosen (Figure 77). The compact TFIIIC conformation showed the loss of τ60 at 200 V which was at a slightly higher voltage than previously observed for the elongated form of TFIIIC where τ60 got expelled at 150 V. Furthermore, even with increasing the voltage up to 300 V, no τ91 was detected.

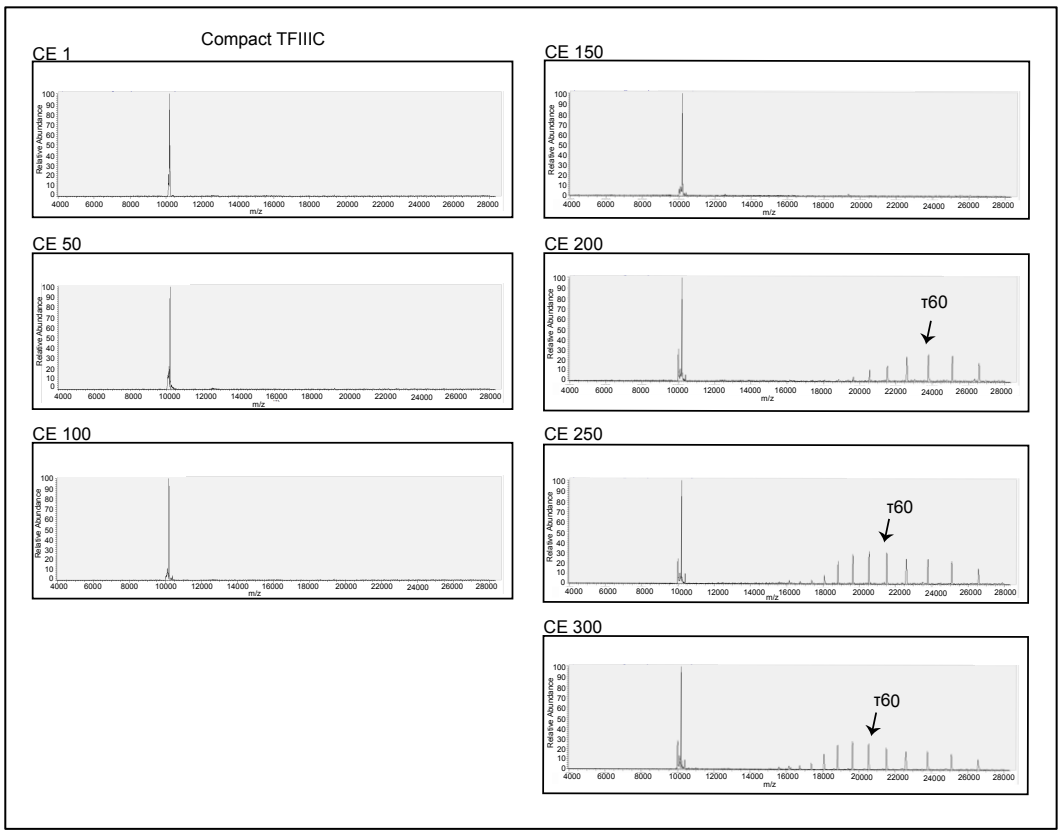


Figure 77: Native mass spectrometry analysis of the compact TFIIIC complex.
The voltage was increased in increments of 50V.

This could indicate that in its elongated form, the τ60/τ91 is less tightly bound to the rest of the TFIIIC complex, making it more likely for this subcomplex to detach. In particular τ91 appears to be protected in the compact form since it could not be detected.

Furthermore, τ60 and τ91 are not only stable as a heterodimer but also as individual subunits outside the TFIIIC complex. In particular the presence of τ60 without detection of τ91 would confirm this hypothesis.

The results from this experiment could reveal that apo TFIIIC is present in both, a compact and an elongated conformation. The presence of the two conformations could therefore explain the two bands observed in a native PAGE for apo TFIIIC as well as for DNA bound TFIIIC. It still remains elusive, however, what triggers the conformational state of TFIIIC.

The overall observed wide charge state distributions (CSD) (Figure 73) and the dominant high charge states in this experiment suggested that some disorder and/or dynamics inside the TFIIIC complex were present. Thus, it was decided to perform a cyclic ion mobility mass spectrometry analysis.

3.2.3.3.2. Cyclic ion mobility mass spectrometry analysis of apo TFIIIC

Next, an ion mobility analysis of apo TFIIIC was performed to characterize the different TFIIIC conformations further and to confirm the results obtained in the previous experiment using a different mass spectrometer. In this approach, apo TFIIIC was first separated according to its size, shape and charge in the ion mobility cell before it was subjected to a tandem LC-MS/MS analysis.

The cyclic ion mobility analysis was able to identify five populated states of TFIIIC, ranging from a compact form to a highly extended conformation (Figure 78).

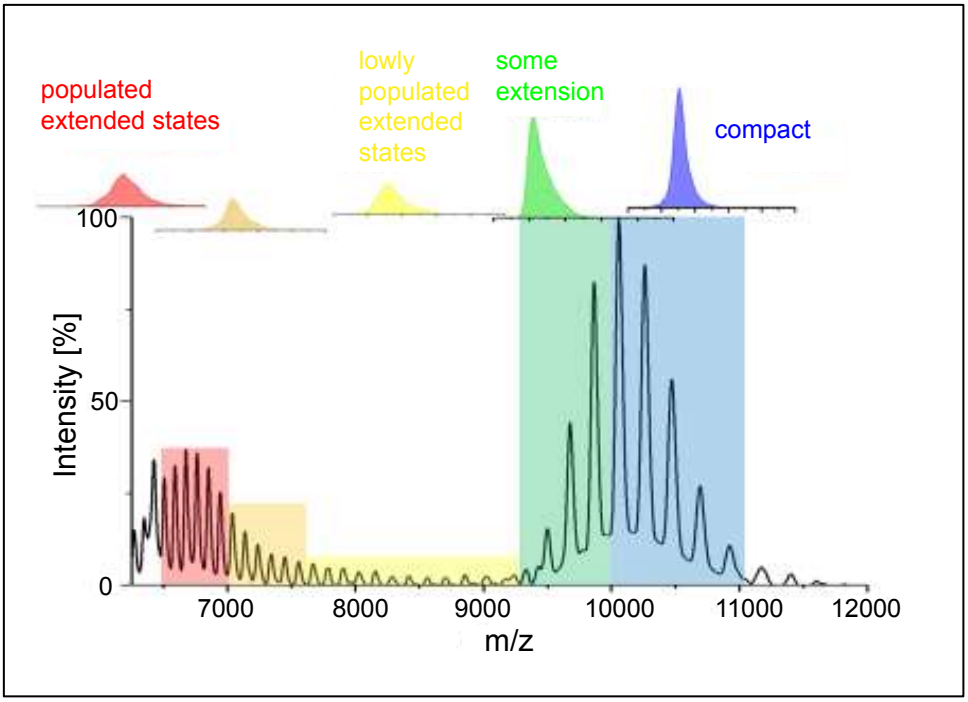


Figure 78: Cyclic ion mobility analysis of apo TFIIIC.

The native mass spectrometry experiments demonstrated that TFIIIC is present in different conformations. However, whether the identified states of TFIIIC are distinct states or whether a constant transition between these states was occurring was not fully elucidated and will be subject to future studies. Furthermore, it would be interesting to find a way to separate the different conformations and analyze them independently for posttranslational modifications. This could help to establish what drives the formation of different conformations. Additionally, better analyzing the structure of them individually would enable a better understanding of their biological relevance.

The results from the SEC/MALS, the SEC-SAXS and the native mass spectrometry experiments suggested that TFIIIC was mainly present as a monomer in solution, adapting several conformations. The TFIIIC dimers, as observed using native mass spectrometry, were most likely not detectable using SEC-SAXS or SEC-MALLS because of the low amount of dimers compared to monomers. Besides a compact and an elongated conformation, a variety of partly extended conformations were observed that could hint to the fact that possibly TFIIIC was transitioning between these different states. In a model suggested in Figure 79, representing the two main conformational states, TFIIIC could be more tightly folded together in the compact conformation. Thereby, τ91 could stay closely attached to the rest of the complex whereas τ60 may only be connected loosely via τ91. The compact conformation could potentially present an inhibited form of TFIIIC, in which the interaction sites for Brf1 in τ131 are shielded at the cost of reducing the tight binding to τ60, similar to the previously suggested model (Liao et al. 2006). In the elongated conformation, TFIIIC could potentially open up, thereby extending the overall diameter, as also overserved in the SEC-SAXS analysis and thus loose interaction sites with τ91. This elongation could allow for

interactions with binding partners whilst the interaction to τ91 was weakened. This may cause τ60 and τ91 to dissociate.

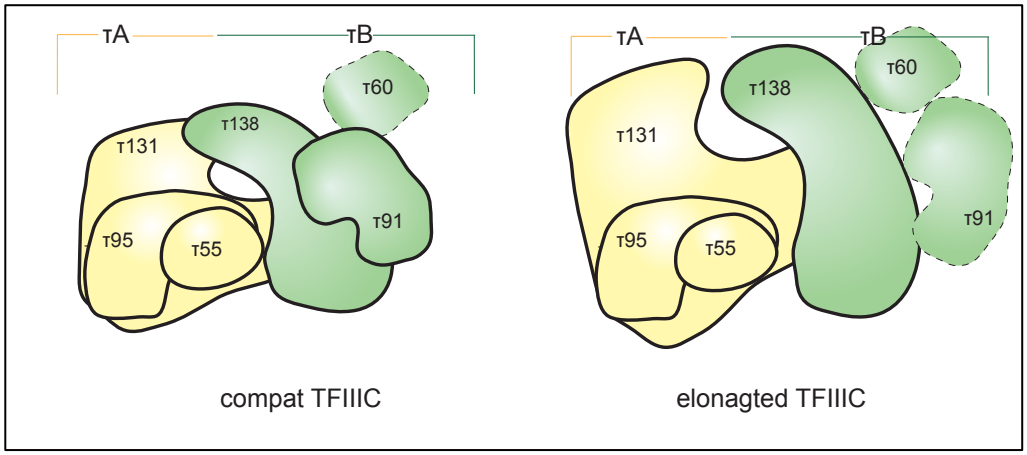


Figure 79: Schematic view of compact and elongated TFIIIC.

Structural studies were performed next with the main aim to clarify on the different conformations by analyzing the structure of them.

3.2.4. Structural characterization of the TFIIIC-DNA complex

After biochemically characterizing yeast TFIIIC, structural studies were performed. Binding TFIIIC to the DNA before analyzing it at the electron microscope was thought to stabilize TFIIIC and reduce the heterogeneity. Thus, TFIIIC was first analyzed bound to the DNA in negative stain to get an initial idea about the overall shape before proceeding with cryo-EM.

3.2.4.1. Negative stain EM of the DNA bound TFIIIC complex

For the negative stain analysis, yeast TFIIIC was incubated with a 4x molar excess of DNA (tD(GUC)K_noTATA_short) for 30 minutes at room temperature. This excess of DNA was chosen to ensure that all TFIIIC complexes were bound to the DNA. By loading the TFIIIC-DNA complex onto a Superose 6 Increase 10/300 GL column at room temperature, the

free DNA was separated from the TFIIIC-DNA complex as shown in (Figure 80A). Due to the technical setup, it was only possible to measure the absorption at 280 nm. Loading the fractions onto an SDS-PAGE and a native 1x TBE 2% agarose gel (Figure 80B), confirmed that the TFIIIC-DNA complex eluted in the first peak and that the free DNA eluted in the second peak. Fraction D6, which is indicated with the red rectangle, was chosen to be analyzed via negative stain EM since all subunits of TFIIIC could be detected in the SDS PAGE and the formation of a TFIIIC-DNA complex could be confirmed in the EMSA.

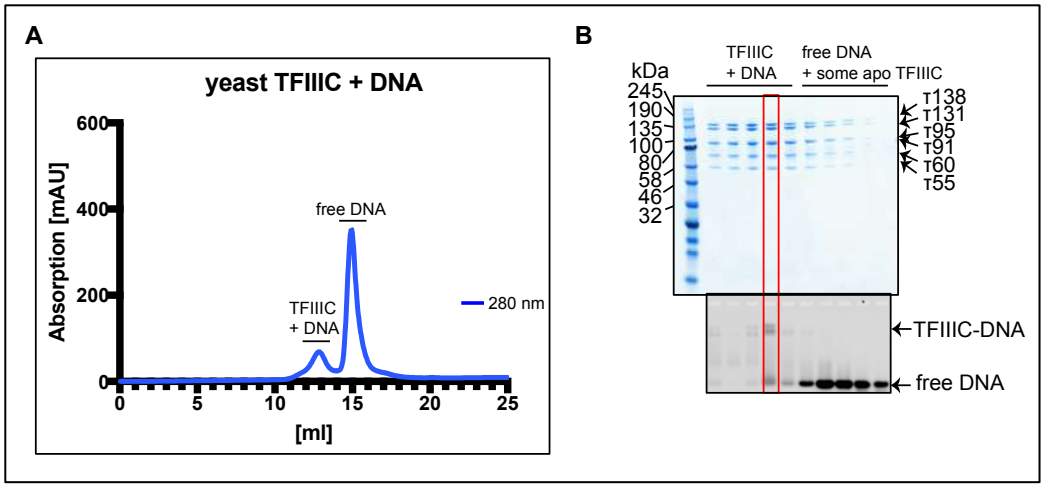


Figure 80: Formation of the TFIIIC-DNA complex for negative stain analysis.

A: Gel filtration elution profile from a Superose 6 Increase 10/300 GL column of the TFIIIC-DNA complex with two main peaks. B: Top: SDS-PAGE of the different peak fractions, bottom: Native 1xTBE 2% agarose EMSA stained with ethidium bromide to detect the formation of the DNA bound complex.

Dr Fabienne Beuron collected 224 micrographs at the FEI TF20 microscope (Figure 81A). Initially, the micrographs were imported into EMAN2 and particles were picked using the swarm pick option. Using this picking approach, it was possible to combine automated picking with manual picking to have a high degree of control about the particle selection. The initial particle subset (16,428 particles) was subjected to 2D classification. Five of these classes, containing 5,854 particles, were chosen as references for automated picking in Relion 2.1. These

micrographs were imported into Relion 2.1 and an auto pick was started with the provided references which produced 105,496 particles. After three rounds of 2D classification using a mask diameter of 190 Å, 30,299 particles were selected (Figure 81B). The obtained 2D classes appeared homogenous and showed a similar shape containing three main domains, one in the middle and one at each side, forming an “M” shape. Using these particles, an initial model was built and a 3D classification into 5 different classes was performed. The most promising class contained 6,159 particles. It was refined and post-processed to a resolution of 25.6 Å (Figure 81C).

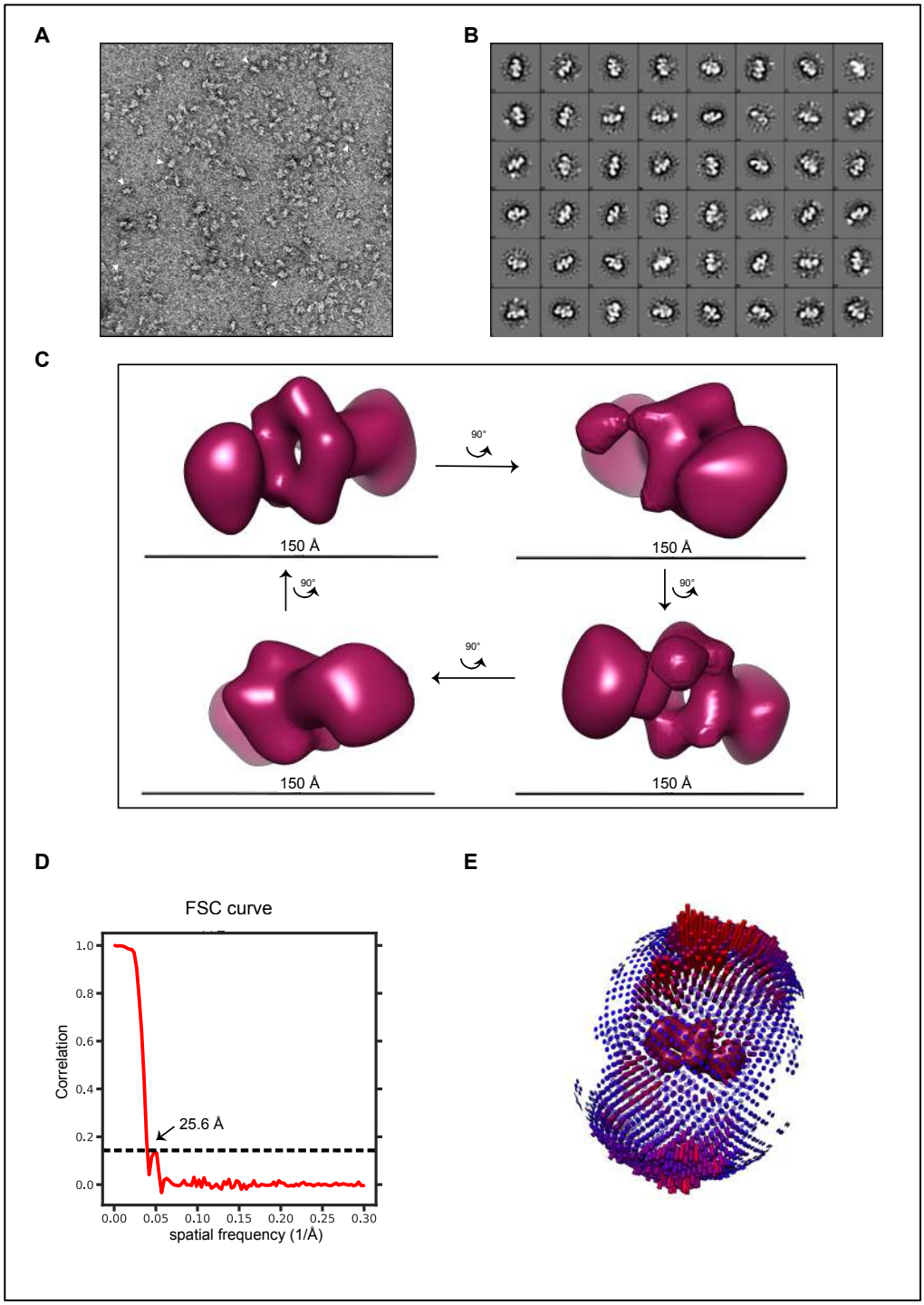


Figure 81: Negative stain analysis of the TFIIIC-DNA complex.

A: Negative stain micrograph taken at the FEI TF20. Exemplary TFIIIC particles are indicated with white arrows. B: 2D classes obtained from Relion 2.1 using a mask of 190 Å. C: Refined and post-processed 3D map of most promising 3D classes at a resolution of 25.6 Å. D: Fourier Shell Correlation. E: Angular distribution plot.

Due to this resolution, it was not possible to detect the DNA or to allocate domains according to previously solved crystal structures. It was interesting to see that at low resolution, the map looked symmetric with two globular domains at each end and a domain with almost five-fold symmetry in the middle. According to previous observations using scanning transmission electron microscopy, a dumb-shell shape was expected for yeast TFIIIC (Schultz et al. 1989). However, the obtained map appeared too small to accommodate all crystal structures. An approximation of the molecular weight using the Chimera Area and Volume measure and a converting factor of 1.21 $Å^3$ per Dalton, led to a molecular weight of 220 kDa. This was only half of the expected molecular weight of 521.2 kDa. These observations suggested that potentially TFIIIC fell apart on the grid. Thus, only one subcomplex, either τA or τB, may have been solved or that a mixture of the two subcomplexes was present, which might look similar at low resolution. Since no structure of a holo-TFIIIC complex was available to compare to the obtained EM reconstruction, it was not possible to verify the map at this point.

3.2.4.2. Glutaraldehyde titration

With the hypothesis that TFIIIC fell apart on the grid, the next step was to titrate a crosslinker over the TFIIIC-DNA complex in order to create intra- and inter-subunit crosslinks that might reduce the observed dissociation. To further optimize the yeast TFIIIC sample for cryo-EM, it was decided to decrease the amount of DNA used in the binding reaction. This would ensure that, despite the separation on the gel filtration column, no free DNA could leak into the fraction to be analyzed at the electron microscope and therefore no artifacts would be produced. BS3, Formaldehyde and Glutaraldehyde were tested as crosslinkers. Formaldehyde dissociated the DNA from TFIIIC-DNA entirely (data not shown) and crosslinked BS3 particles did not look as homogenous as Glutaraldehyde crosslinked

particles on negative stain grids (data not shown). Thus, Glutaraldehyde was chosen as a crosslinker.

225 μg of TFIIIC (432 pmol) of TFIIIC were incubated with 216 pmol of the tD(GUC)K_noTATA_short DNA construct for 30 minutes at room temperature. A small-scale Glutaraldehyde titration experiment showing that a final concentration of 0.01% Glutaraldehyde was sufficient to crosslink the complex was performed (data not shown). The sample was then crosslinked with a final concentration of 0.01%, 0.03% and 0.05% Glutaraldehyde. Each experiment was quenched with 50 mM Tris pH 8 and loaded onto a Superose 6 Increase 10/300 GL column using the running buffer from section 2.7.1.2 and the peak fractions were analyzed at the FEI T12 electron microscope.

The elution profiles of each gel filtration run were overlaid to compare whether the increased concentration of crosslinker influenced the elution pattern (Figure 82A). It was observed that the first peak area was reduced and the second peak area increased with Glutaraldehyde concentration. Free DNA was detected in the second peak (data not shown), which implies that Glutaraldehyde interfered with the TFIIIC-DNA-binding. However, in all chromatograms, the absorption at 260 nm was higher than the absorption at 280 nm indicating that DNA-binding was present. An EMSA of the peak 1 fractions using a 2% 1x TBE agarose gel, confirmed that a TFIIIC-DNA complex was formed (Figure 82I). As previously reported, two TFIIIC-DNA-binding events could be detected. We also observed that free DNA was present in the samples and that the signal of the TFIIIC-DNA complex decreased with increasing Glutaraldehyde concentration. The fraction loaded onto the EMSA of each Glutaraldehyde titration experiment, was analyzed via negative stain EM. Compared to un-crosslinked TFIIIC+DNA (Figure 82B), the observed particles increased in size at 0.01% Glutaraldehyde and further with 0.03% Glutaraldehyde crosslinking (Figure 82D, F). Indeed, the particles

in Figure 82F were more than twice the size than the un-crosslinked sample. Increasing Glutaraldehyde concentration further to 0.05% (Figure 82H), led to aggregation and was thus not considered as a suitable condition.

It was concluded that at a concentration of 0.01%, TFIIIC was only partially crosslinked, thus leading to a heterogenous sample containing smaller particles. At a concentration 0.03%, all TFIIIC appeared fully crosslinked since a further increase in crosslinker concentration led to aggregation, most likely caused by intramolecular crosslinks.

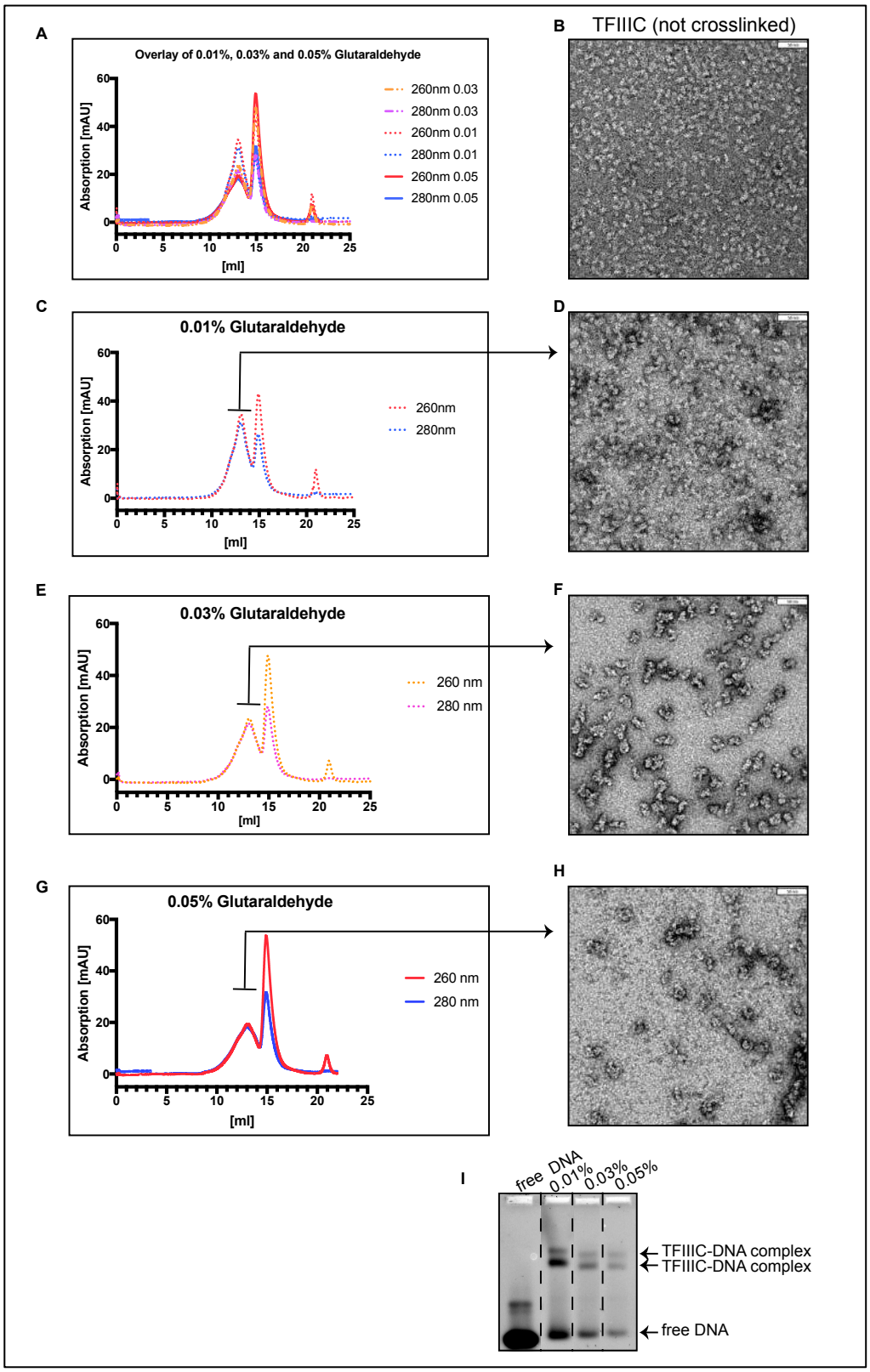

A
Overlay of 0.01%, 0.03% and 0.05% Glutaraldehyde
Absorption [mAU]
[ml]
260nm 0.03
280nm 0.03
260nm 0.01
280nm 0.01
260nm 0.05
280nm 0.05
B
TFIIIC (not crosslinked)
C
0.01% Glutaraldehyde
260nm
280nm
D
E
0.03% Glutaraldehyde
260 nm
280 nm
F
G
0.05% Glutaraldehyde
260 nm
280 nm
H
I
free DNA
0.01%
0.03%
0.05%
TFIIIC-DNA complex
TFIIIC-DNA complex
free DNA

Figure 82: TFIIIC-DNA Glutaraldehyde titration.

A: Overlay of gel filtration elution profiles of the TFIIIC-DNA complex crosslinked at a final concentration of Glutaraldehyde of 0.01%, 0.03% and 0.05%. B: Negative stain micrograph of the un-crosslinked TFIIIC-DNA complex. C: Superose 6 Increase 10/300 GL gel filtration elution profile of the TFIIIC-DNA complex crosslinked with 0.01% Glutaraldehyde with its respective negative stain micrograph shown in D. E: Superose 6 Increase 10/300 GL gel filtration elution profile of the TFIIIC-DNA complex crosslinked with 0.03% Glutaraldehyde with its respective negative stain micrograph shown in F. G: Superose 6 Increase 10/300 GL gel filtration elution profile of the TFIIIC-DNA complex crosslinked with 0.05% Glutaraldehyde with its respective negative stain micrograph shown in H. I: Native 1xTBE 2% agarose PAGE stained with ethidium bromide on which the three peak fractions of the first peak were analyzed.

It was therefore decided to continue with the cryo-EM experiments using a Glutaraldehyde concentration of 0.03%. As the particles were now significantly larger, this confirmed the previous hypothesis that un-crosslinked TFIIIC fell apart into its two subcomplexes during the preparation of the negative stain grid. Using Glutaraldehyde as a crosslinker seemed to compromise the DNA-binding slightly. Since TFIIIC otherwise fell apart on the grid, it was however not possible to perform further structural analysis without crosslinking.

To attempt to decrease the heterogeneity on the grid, Gradient Fixation (GraFix) was tested. GraFix combines a glycerol density gradient with a Glutaraldehyde crosslinking gradient. By ultra-centrifugation, the protein sample becomes gradually crosslinked and different crosslinked species are separated according to their weight into different glycerol layers. Using GraFix, however, did not increase the homogeneity on the grid and it was thus decided not to use gradient fixation instead of the standard crosslinking protocol followed by a gel filtration.

An initial negative stain analysis of TFIIIC-DNA, crosslinked with 0.03% Glutaraldehyde, was performed. 187 micrographs were collected by Dr Fabienne Beuron at the FEI TF20 in order to assess the quality of the sample. A promising and larger EM map compared to the non-crosslinked TFIIIC-DNA sample was obtained at a resolution of 30 Å (Figure 83).

Comparing the newly obtained structure (Figure 83A) to the previously obtained uncrosslinked TFIIIC-DNA complex (Figure 83B), it was concluded that the larger size of Figure 83A would allow the crystal structure of the TPR repeats to be accomodated, which comprises less than 10 % of the total structure (Figure 83C). However, the structure in Figure 83B did not appear to be big enough and this led to the decision to proceed with the preparation of crosslinked TFIIIC-DNA cryo grids.

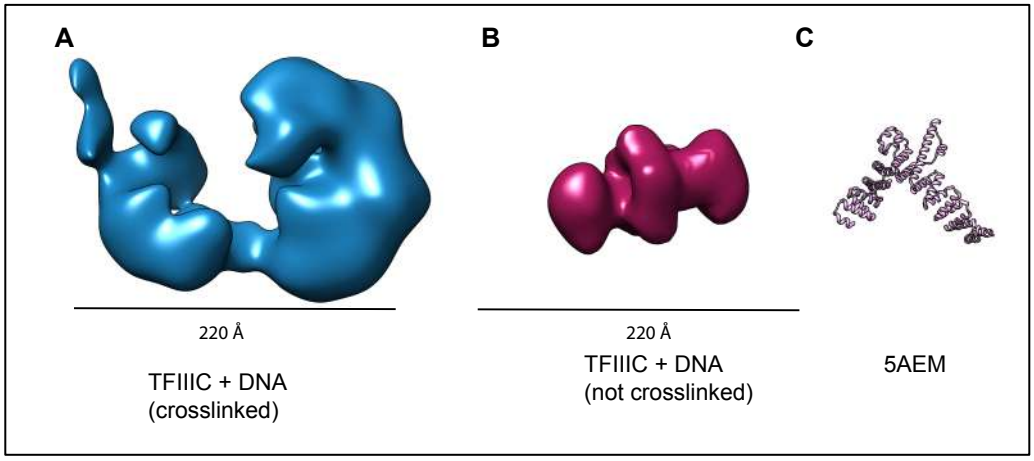


Figure 83: Comparison of the crosslinked and non-crosslinked TFIIIC-DNA complex.

3.2.4.3. Cryo-electron microscopy studies of DNA bound yeast TFIIIC

For the preparation of cryo grids, yeast TFIIIC was bound to the tD(GUC)K_noTATA_short DNA (Table 15), crosslinked with 0.03% Glutaraldehyde, quenched with 50 mM Tris pH 8 and loaded on a Superose 6 Increase 10/300 GL column (Figure 84A). The SDS-PAGE (Figure 84B, bottom) confirmed that TFIIIC was crosslinked.

The fractions corresponding to the first peak were initially screened via negative staining. Fraction D6 was chosen for the subsequent cryo-EM experiments since individual particles were clearly detectable in the negative stain (Figure 84C) and the EMSA (1x TBE 2% agarose native PAGE) (Figure 84B, top) indicated that this fraction only contained the top

population of the two TFIIIC-DNA complexes described before. Having only one DNA-binding event was thought to reduce heterogeneity and to facilitate processing compared to analyzing a TFIIIC-DNA sample with two different binding conformations.

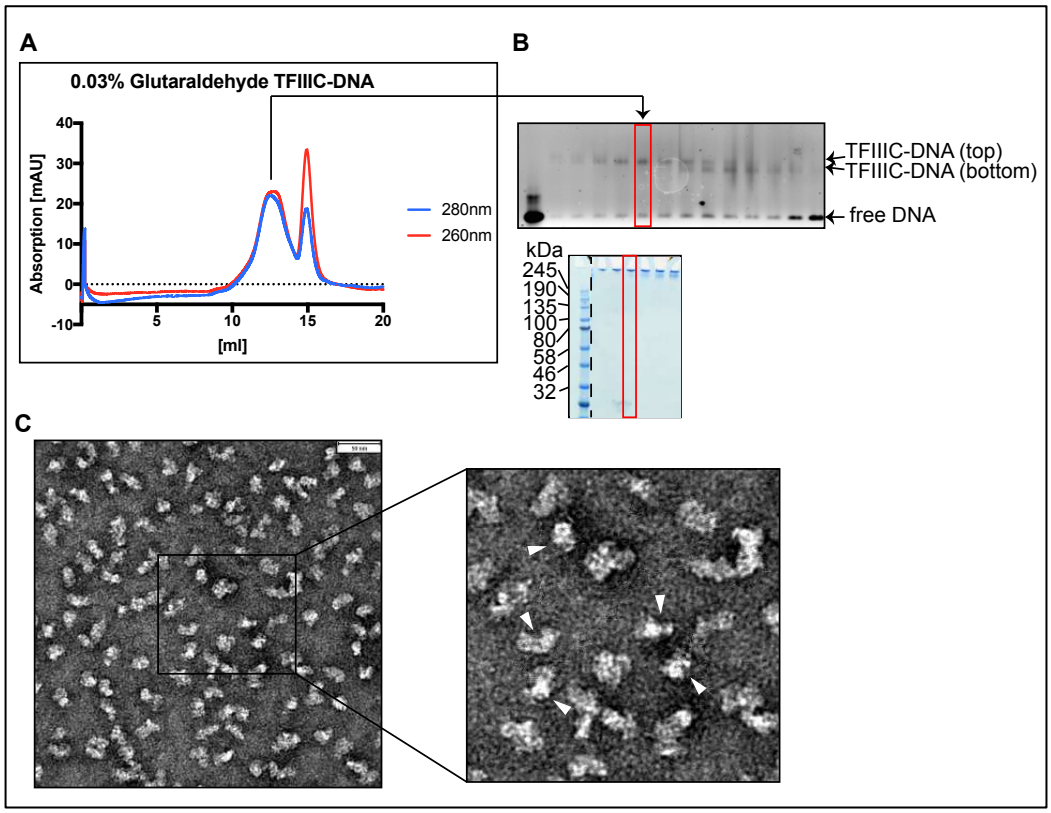


Figure 84: Yeast TFIIIC-DNA cryo grid preparation.

A: Elution profile of the TFIIIC-DNA complex, crosslinked with 0.03% Glutaraldehyde from a Superose 6 Increase 10/300 GL column. B: EMSA and SDS-PAGE. The first peak contained the TFIIIC-DNA complex as confirmed by the native 1xTBE 2% agarose PAGE stained with ethidium bromide and the SDS-PAGE confirming the presence of crosslinked protein. C: Negative stain micrograph of the peak fraction taken at the FEI T12 with a zoom that shows TFIIIC-DNA particles marked by white arrows.

Initially, the cryo grids (LJ012) were screened inhouse, at the FEI TF20 before being taken to Electron Biology Imaging Centre (eBIC, DL, Didcot) for a 48h data collection (Figure 85A). There, 3,216 micrographs were collected from which 128,811 particles could be picked in Relion 2.1. After several rounds of 2D and 3D classifications, it was not possible to generate an EM reconstruction at a resolution higher than 33 Å. To

improve the resolution, particle picking with Gautomatch instead of the autopick method in Relion was tested. Gautomatch has a better picking algorithm that prevents the selection of carbon artefacts and thus could improve the alignment of the particles and increase the final resolution. However, in our case, following this method did not increase the resolution of the obtained 3D reconstructions (Figure 85B).

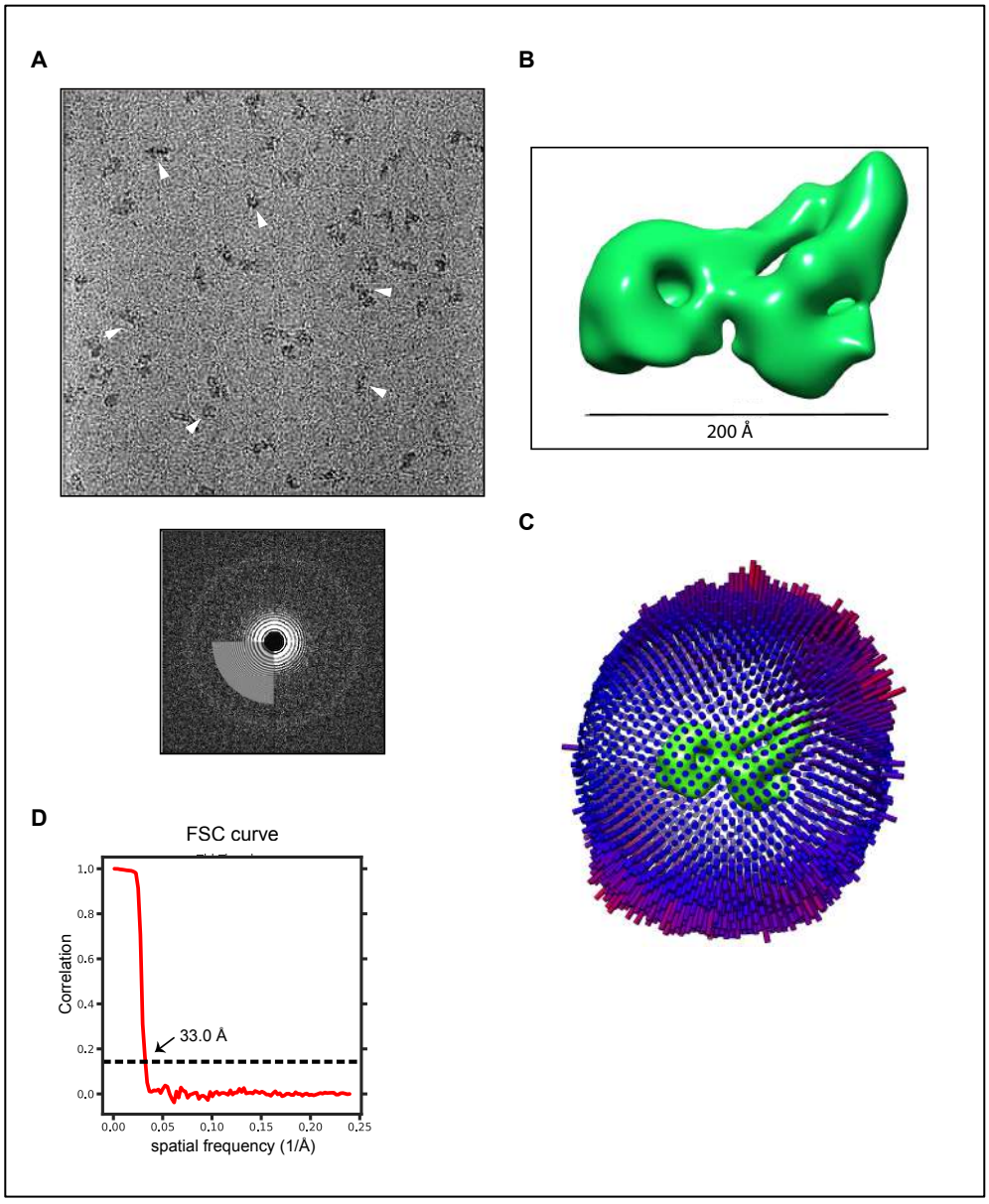


Figure 85: Data collection of the TFIIIC-DNA complex at eBIC.

A: Micrograph was taken at the Krios 1 at eBIC using a K2 camera. Exemplary particles are marked with white arrows. B: TFIIIC-DNA 3D reconstruction obtained from this dataset at 33 Å. C: Angular distribution plot. D: Fourier Shell Correlation.

It was hypothesized that this would be caused by the poor alignment of the particles. Particle misalignment could could be caused by potential pseudo-symmetry of τA and τB (Figure 81C). Relion 2.1 was not able to estimate the accuracy of the translations and rotations of most 2D classes.
Another reason could be the flexibility of the complex. It was not clear if TFIIIC was able to adopt a stable DNA-binding conformation on its own in the absence of TFIIIB and the RNA Pol III. Additionally, some heterogeneity was present on the grid and the carbon coating caused a high background signal, which reduced the SNR.

The heterogeneity and the background noise from the carbon posed a significant challenge. However, the pseudo-symmetry and the flexibility of the complex could be addressed. To overcome the pseudo-symmetry problem, it was thought that collecting more data would increase the number of particles per class and therefore increase the SNR. Thus, the prepared cryo grids were taken for two more data collections at the Astbury Centre for Structural Biology in Leeds. One data collection was run for 72h using a Falcon III EC camera in integrating mode to collect a large number of micrographs (Figure 86A) whereas the second data collection was done for 72h using a Volta Phase Plate (VPP) to enhance the contrast (Figure 86B). Due to the enhanced contrast from the VPP, a variety of particle sizes was visible on the grid detected exclusively using the VPP. Fairly small particles could be caused by the blotting conditions. This confirmed that some heterogeneity was present in the sample. Both datasets were processed independently.

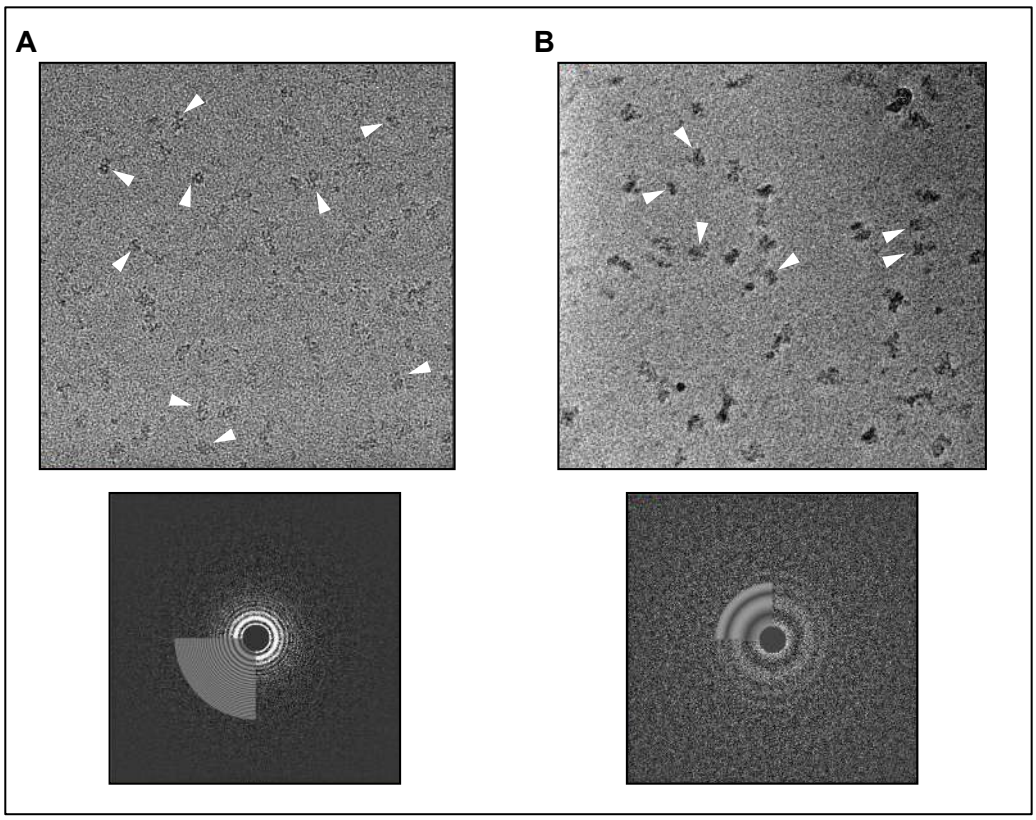


Figure 86: Cryo-EM micrographs of the TFIIIC-DNA complex and its respective thon rings. *A: Micrograph was taken at the Krios 1 at the Astbury Centre for Structural Biology using a Falcon III EC camera. B: Micrograph was taken at the Krios 2 using at Volta Phase Plate and a K2 camera at the Astbury Centre for Structural Biology.*

3.2.4.3.1. Data collection using a Falcon III EC camera in integration mode

8,547 micrographs were collected at the Krios1 using the Falcon III EC camera in integration mode.

The CTF estimation was performed using Gctf and an autopick was performed in Relion 2.1 selecting 656,196 particles. Due to the large box size of 512 pixels (1.065 Å per pixel), the data had to be binned by a factor of 2 and split into 7 different subsets due to computational difficulties. 2D classifications were gradually merged together and the dataset finally contained 161,599 particles. The SNR in each 2D class was still relatively

low (Figure 87A). This meant, the generated 3D classes (Figure 87B) appeared very noisy and the particles could not be classified properly. Instead, the particles were evenly distributed into the classes, leading to 15,000 up to 18,000 particles per class.

Due to the poor-quality particle picking in Relion, a significant number of carbon artefacts were selected which decreased the SNR in each 2D class significantly (Figure 87A and B).

Thus, the newly developed software cisTEM with an optimized picking algorithm was used. By setting the picking parameters very conservatively, 248,565 particles were picked. These particles were imported into Relion and 2D classified using a tighter mask of 400 Å (Figure 87C). The obtained 2D classes had a significantly increased SNR compared to the classes shown in Figure 87A. They showed a clear overall shape in which some internal features started to become visible.

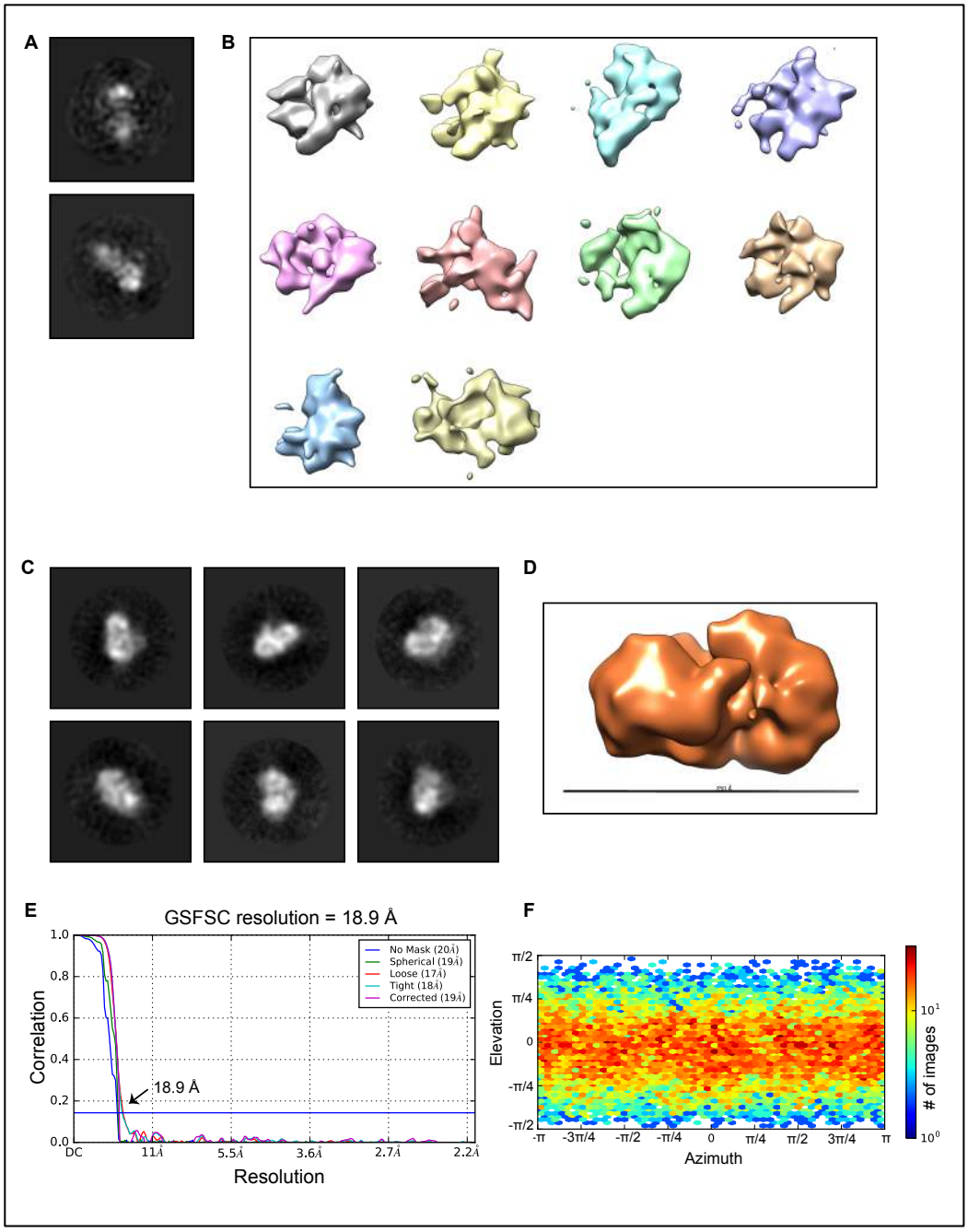


Figure 87: Krios 1 (72h) data set of the yeast TFIIIC-DNA complex (LJ012) taken at the Astbury Centre for Structural Biology in Leeds.

A: Initial 2D classes using Relion 2.1. with a mask of 420 Å. B: Initial 3D classes obtained from Relion 2.1. C: Relion 2D classification with a mask of 400 Å using the particles from the cisTEM automated picking. D: 3D reference map of the TFIIIC-DNA complex at 18.9 Å obtained from the imported cisTEM particles in cryoSPARC v2. E: Resolution estimation using the gold standard Fourier Shell Correlation, taken from cryoSPARC v2. F: Heat map showing angular distribution plot, taken from cryoSPARC v2. The colour indicates the number of particles per view (few – more = blue to red).

The particle set obtained from cisTEM, was imported into cryoSPARC v2 to generate an *ab initio* 3D model. Through several rounds of 2D classifications, 3D Hetero Refinements and Refinements, a reconstruction at 18.9 Å was obtained from 24,972 particles (Figure 87D). This generated *ab initio* 3D map, which was at significantly higher resolution than the initial 3D maps obtained in Relion, was imported into Relion 2.1. There, the 3D map was refined and post-processed with the whole imported particle set from cisTEM. Finally, an EM reconstruction at 11.85 Å was obtained (Figure 88).

The EM map of TFIIIC bound to the DNA contained a dumb-bell-like shape with two main domains being present at each side; this could be the two subcomplexes τA and τB. Also, it was notable that the two domains had a similar half-moon-like shape that were related by a rotation axis. This could have caused the pseudo-symmetry problem described previously. At this resolution, it was not possible to unambiguously place the crystal structures into the map. However, the atomic model of τ60/τ91 fit well into one of the lobes (Figure 88A) which led to the hypothesis that this might be the τB subcomplex region.

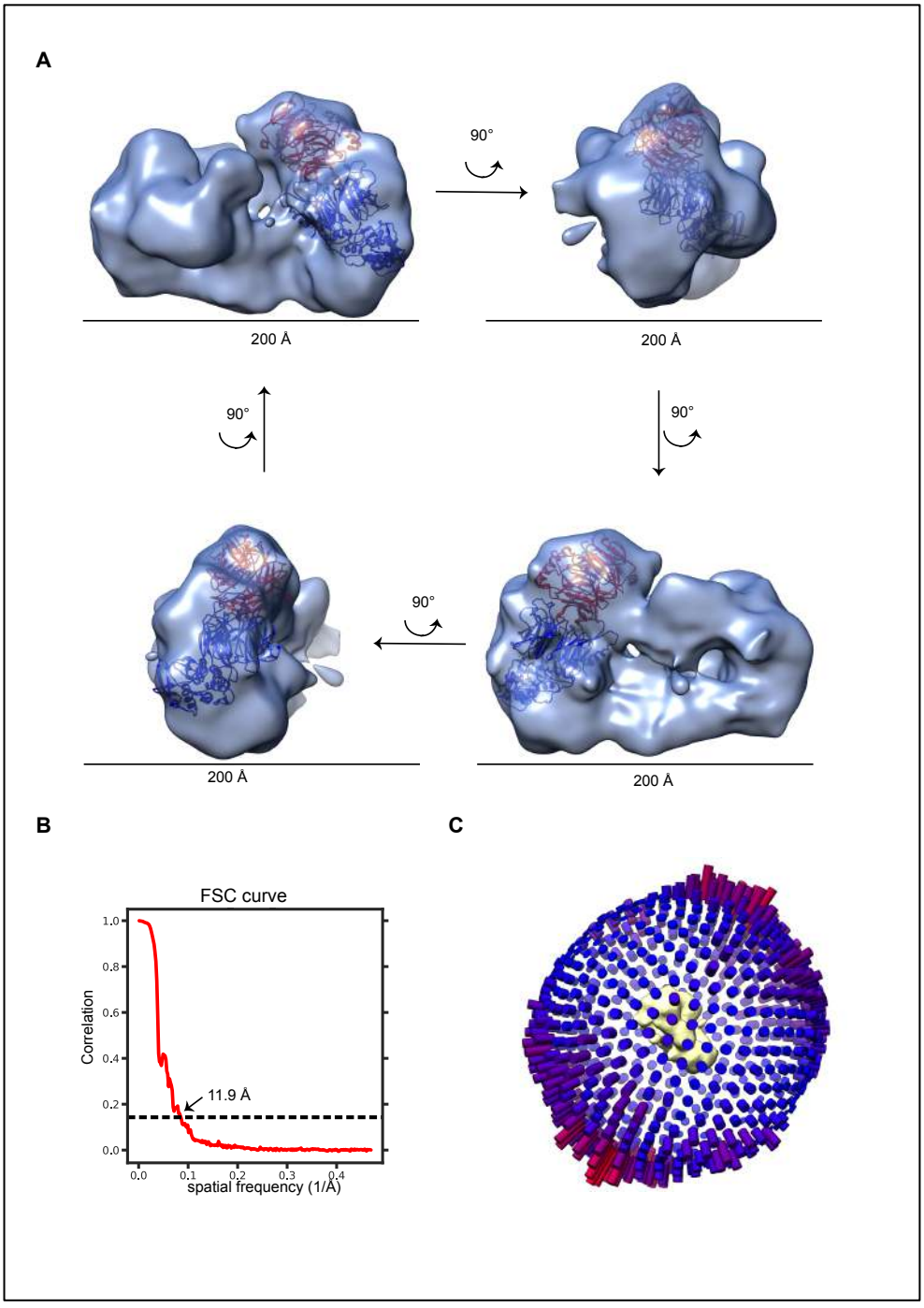


Figure 88: Refined and post-processed 3D map of the TFIIIC-DNA complex (LJ012) collected at the Krios 1 at the Astbury Centre for Structural Biology in Leeds.

A: The resolution was estimated to 11.9 Å. The atomic model of τ60/τ91(PDB entry: 2J04) was fitted with τ60 shown in blue and τ91 shown in red. B: Fourier Shell Correlation. C: Angular distribution plot.

Further improvement of the resolution will be necessary to unambiguously identify the domains in the structure and to understand how the DNA is bound by TFIIIC.

3.2.4.3.2. Data collection of the TFIIIC-DNA complex using a volta phase plate (VPP)

3,167 micrographs were collected at the Krios 2 at the Astbury Centre for Structural Biology in Leeds using a VPP. The VPP induced phase shifts between 22.5 and 101° to enhance the SNR (Figure 89A). In total, 159,644 particles were picked using the autopick in Relion 2.1, then 40,619 particles were selected after four rounds of 2D classification using a mask diameter of 412 Å. The classification showed classes with a clear overall shape and a good SNR, however the details of internal structures were very poor (Figure 89B). A 3D classification did not generate a high-resolution 3D map. Performing particle subtraction and focused 3D classifications with local masks did not improve the resolution significantly either.

Relion 2.1 was not able to generate a high-resolution structure. Therefore, the particles were imported into cyroSPARC v2. After several rounds of 2D classifications, several domains were clearly visible in the classes (Figure 89C). However, the structures were still blurry and did not reveal a great amount of internal detail. The highest resolution 3D map that was generated was at 21 Å and is shown in Figure 89D. Two distinct modules were visible that had a half-moon-like shape as previously observed.

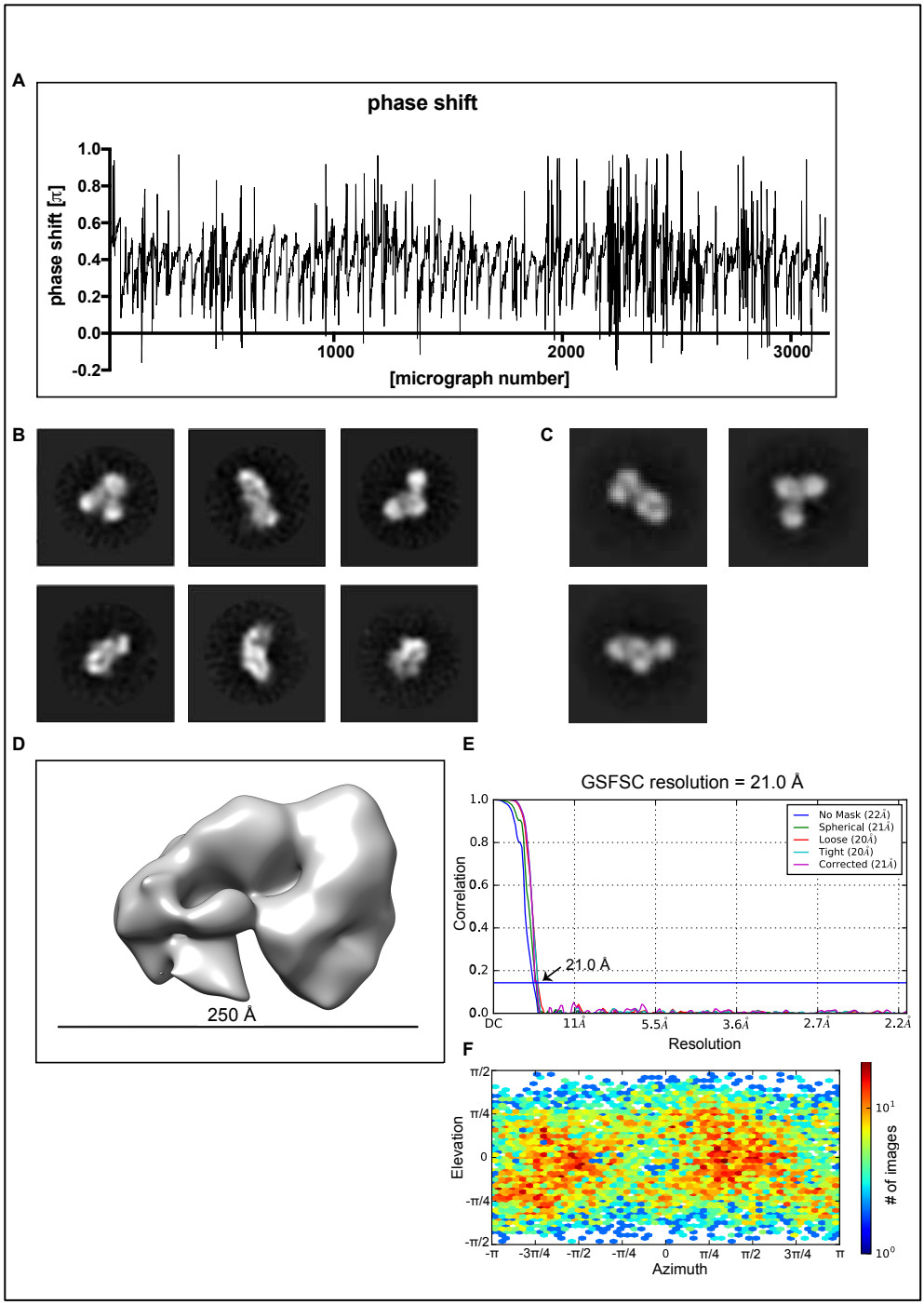


Figure 89: VPP yeast TFIIIC bound to the DNA.

A: Plotted phase shifts during the data collection induced by the VPP. B: Initial 2D classes using Relion 2.1. C: 2D classes obtained from cryoSPARC v2. D: 3D map obtained from TFIIIC bound to the DNA from cryoSPARC v2 obtained at a resolution of 21 Å. E: Resolution estimation using the gold standard Fourier Shell Correlation, taken from cryoSPARC v2. F: Heat map showing angular distribution plot, taken from cryoSPARC v2. The colour indicates the number of particles per view (few – more = blue to red).

Since the processing in Relion 2.1 and cryoSPARC v2 did not give satisfactory results, it was decided to manually sort the micrographs in cisTEM, grouping them into good and bad quality micrographs according to their particle distribution and ice thickness. A mixture of automated and manual picking was performed on the group of good quality micrographs. The particles were used for further 2D classifications and 3D classification in cisTEM and Relion 2.1 but no improved result was obtained (data not shown).

Summarizing the results from this section, the processing of TFIIIC bound to the DNA proved to be very challenging. The alignment of particles worked slightly better using a VPP or by increasing the number of particles compared to previous data collections. The main problems, which were the degree of heterogeneity and flexibility of the complex, still prevented the particles to align properly. Using a combined approach of different software such as Relion 2.1, cryoSPARC v2, cisTEM and Gautomatch was the only way to improve the resolution slightly. Although the individual particles are easily detectable and were picked correctly, the available software packages were still not able to process the data to high resolution. Furthermore, one question that still remained was the potential presence of apo-TFIIIC dimers, as observed in the native mass spectrometry analysis performed by Prof Perdita Barran's team. These dimers could be mixed together with TFIIIC complexes bound to the DNA and TFIIIC complexes that lost the DNA-binding after the crosslinking or during the grid preparation. A mixture of all these different populations, together with the fact that TFIIIC is a flexible complex could have been the reason for the poor particle alignment. Furthermore, the quality of the low-resolution structures could unfortunately not be compared to other holo-TFIIIC structures since no structures were available.
Unsupported UltrAUFoil 2/2 grids of yeast TFIIIC bound to the DNA were prepared in an attempt to reduce the background caused by the carbon layer. Due to time restraints and microscope access, no data was

collected from these grids. However, preparing these grids enabled to establish a protocol of re-applying the sample to the grid which significantly increased the concentration of the protein complex on the grid without concentrating it beforehand. This grid preparation protocol as described in section 2.7.2.2, was used for the preparation of the apo TFIIIC complex.

3.2.5. Structural analysis of the apo TFIIIC complex

As aforementioned, the structural analysis of TFIIIC bound to the DNA proved to be quite challenging. Thus, it was decided to simplify the system by structurally studying apo TFIIIC.

3.2.5.1. Negative stain analysis of the large and small apo TFIIIC fractions

In an initial experiment, apo TFIIIC was crosslinked with 0.03% Glutaraldehyde, quenched and loaded onto a Superose 6 Increase 10/300 GL column.
Compared to the running profile of uncrosslinked apo TFIIIC (Figure 60B) that showed a single peak at 14ml, it was observed that crosslinked apo TFIIIC led to two distinct peaks in the chromatogram (Figure 90A). The first peak at 12 ml was at the same position where DNA bound TFIIIC sample elutes and thus gave an indication that either an elongated TFIIIC conformation or an oligomeric form of TFIIIC was stabilized with the crosslinking. Analyzing the fractions from peak 1 and 2 on an 1xTBE 2% agarose native PAGE revealed that indeed two populations of TFIIIC could be observed (Figure 90B). The population of the second peak ran much further into the native PAGE than the population from the first peak, indicating a difference in size and/or charge. Compared to the uncrosslinked TFIIIC, the crosslinked sample ran further into the gel.

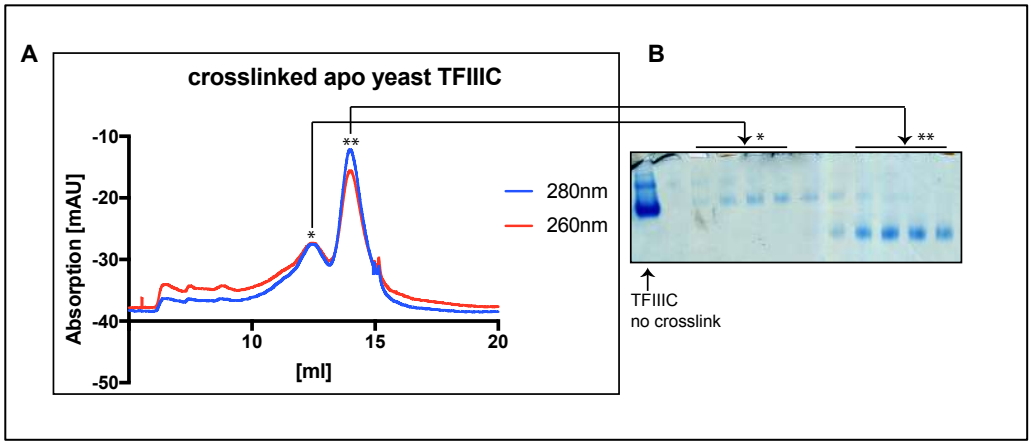


Figure 90: Gel filtration analysis of the crosslinked apo TFIIIC complex.

A: Chromatogram from a Superose 6 Increase 10/300 GL column. B: 2% agarose 1x TBE Native PAGE stained with Coomassie Blue.

Negative stain grids were prepared of the peak fractions from peak 1 and peak 2 and compared to uncrosslinked TFIIIC (Figure 91). A micrograph of the uncrosslinked TFIIIC showed that the complex appeared to have fallen apart into different subcomplexes (Figure 91A). This confirmed earlier observations about the disassembly of TFIIIC. The TFIIIC sample from peak 1 (Figure 91B) and peak 2 (Figure 91C) showed significant differences in the size of the particles. The particles from the first peak could be TFIIIC dimers, which would be in accordance with the results from the native mass spectrometry analysis described in section 3.2.3.3. Crosslinking TFIIIC could have stabilized the TFIIIC-dimers, which are otherwise potentially subject to a constant on-and-off reaction and thus not detectable without crosslinking on a gel filtration column. The particles from peak 2 appeared larger than the uncrosslinked sample and approximately half the size of the TFIIIC complexes from peak 1. It was thus hypothesized that monomeric apo TFIIIC eluted in peak 2.

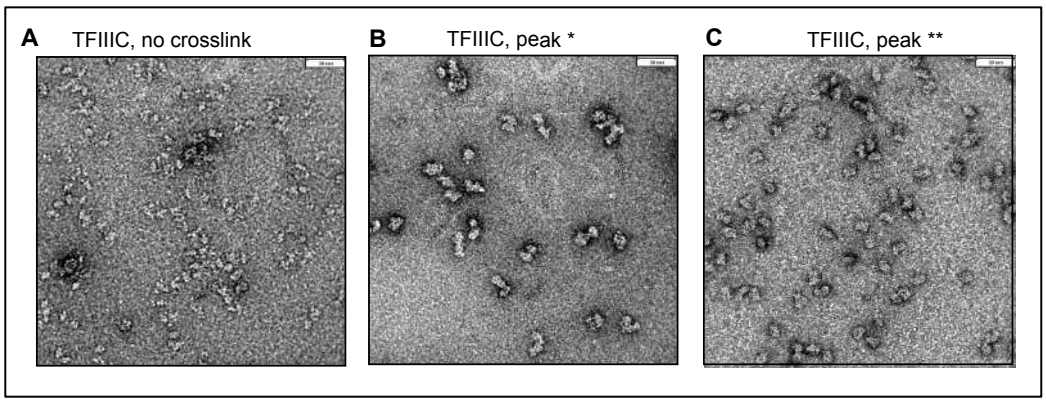


Figure 91: Negative stain micrographs of crosslinked and non-crosslinked apo TFIIIC.
A: Micrograph of non-crosslinked apo TFIIIC. B: Micrograph of crosslinked TFIIIC from the first peak, marked with one asterisk. C: Micrograph of crosslinked TFIIIC from the second peak, marked with two asterisks. All micrographs were taken at the FEI T12 microscope.

To better understand the different TFIIIC populations, a negative stain analysis of each of the peaks was performed. 303 micrographs were collected at the FEI TF20 from the first peak and 157 micrographs from the second peak. The two data sets were analyzed independently using Relion 3.0. 15,513 particles were picked from the first peak. These particles were 2D classified twice using a mask diameter of 450 Å. Some of the 2D classes from the selected subset of 12,693 particles are shown in Figure 92A. With these particles, an initial model and a refined and post-processed map were generated (Figure 92B, shown in yellow).
39,800 particles were obtained from the data collection of the second peak. After four rounds of 2D classifications with a mask diameter of 400 Å, 24,130 particles were obtained (Figure 92C), which were used to build an initial model and a refined, post-processed map (Figure 92D, shown in grey).
The two maps showed significant difference in size and shape. The yellow map showed two distinct domains each approximately comprising the size of the grey reconstruction. This observation suggested that the yellow map could be TFIIIC dimers. However, at low resolution, it was not possible to detect symmetry which would have confirmed the hypothesis about the formation of dimers. Thus, a SEC-MALS experiment was

performed for further clarification. It was performed to unambiguously determine the molecular weight of the two populations of crosslinked apo TFIIIC and to unambiguously identify the position of the monomeric TFIIIC complex.

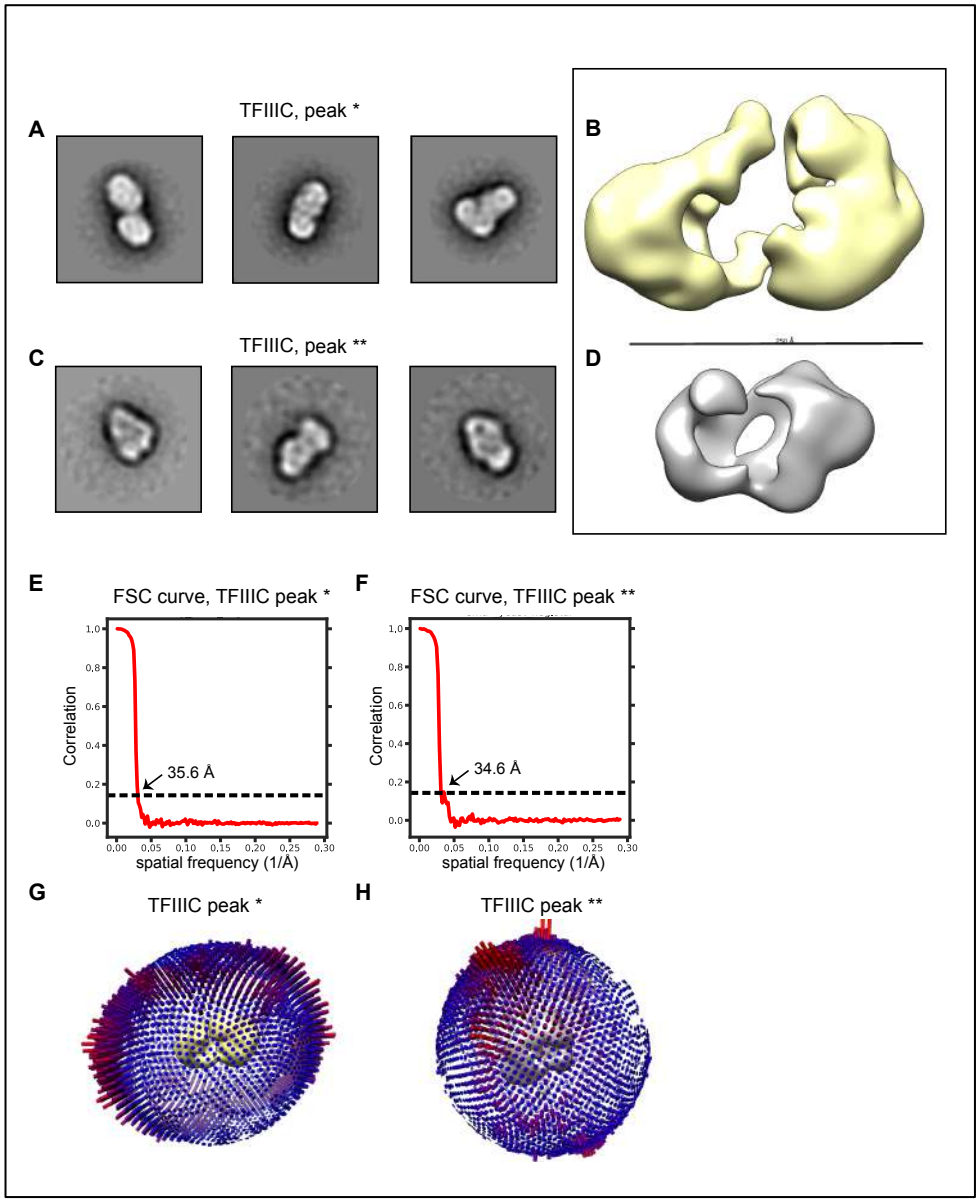


Figure 92: Negative stain analysis of crosslinked apo TFIIIC.

A: 2D classes of the apo TFIIIC complex eluting in peak 1 using a mask diameter of 450 Å.B: Obtained 3D map for the TFIIIC complex eluting in peak 1 shown in yellow. C: 2D classes of the apo TFIIIC complex eluting in peak 2 using a mask diameter of 400 Å. D: Obtained 3D map for the TFIIIC complex eluting in peak 2 shown in grey. E, F: Fourier Shell Correlations of both peaks. F, G: Angular distribution plots of both peaks.

3.2.5.2. SEC-MALS analysis of the crosslinked apo TFIIIC complex

Crosslinked apo TFIIIC was analyzed using the in-house SEC-MALS setup. Two peaks could be identified in the chromatogram in which a signal of the refractive index (dRI) and the light scattering was observed (Figure 93, top row). The first peak eluted at 12 ml, whereas the second peak eluted at 13.5 ml. From both peaks, the molecular weight was measured and plotted against both the dRI signals (Figure 93, middle row) and the light scattering signal (Figure 93, bottom row). The molecular weight of the crosslinked apo TFIIIC was calculated to be 1.18 MDa in the first peak and 661 kDa in the second peak. Factoring in the internal error of the SEC-MALS equipment, it could be concluded that the second peak contained monomeric apo TFIIIC. The first peak had an estimated molecular weight twice that observed in the second peak, which could arise from the stabilization of apo TFIIIC dimers through crosslinking.

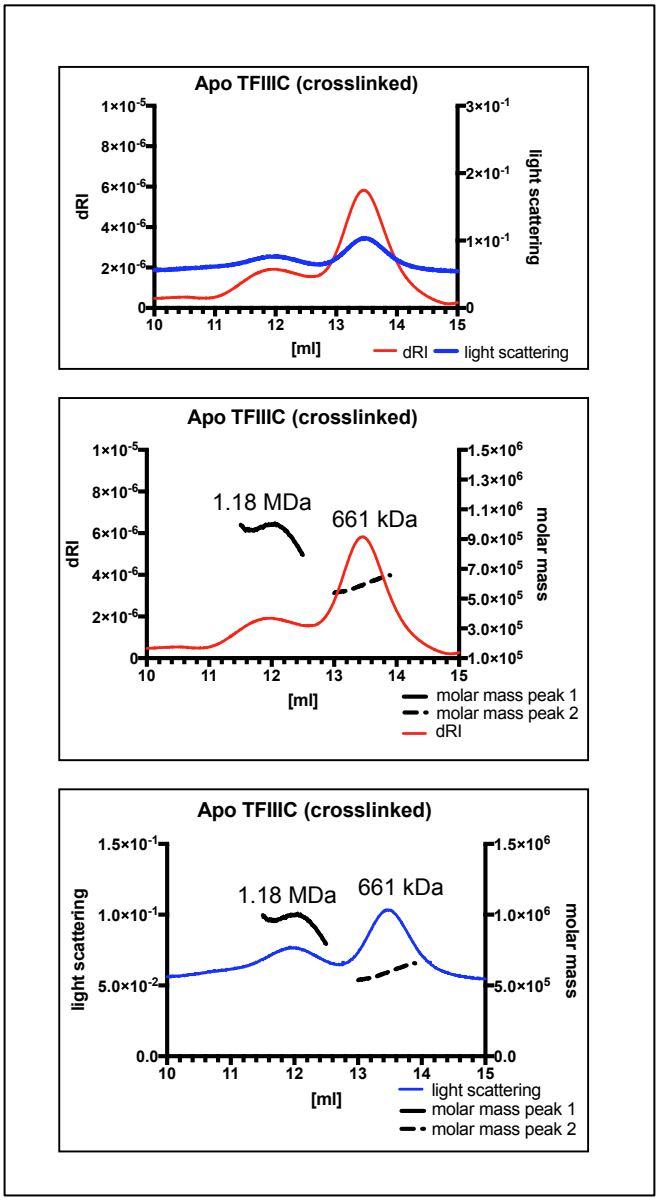


Figure 93: In-house SEC-MALS analysis of crosslinked apo TFIIIC.

Top row: The refractive index (dRI) and light scattering is plotted against the elution volume for crosslinked apo TFIIIC. Middle row: The dRI is plotted against the molar mass and the elution volume for crosslinked apo TFIIIC. Bottom row: The light scattering is plotted against the molar mass and the elution volume for crosslinked apo TFIIIC.

These results suggested that in the absence of a crosslinking reaction, no dimers were observed, as seen previously in section 3.2.3.1. This could imply that the formation of TFIIIC dimers is an on-off reaction in solution

which is stabilized by crosslinking the complex. Alternatively, it could be a crosslinking artefact.

Labelling TFIIIC with an anti-Strep antibody that bound to τ91, was used to identify the location of τ91 via negative stain EM. Thereby, detection of two bound antibodies per particle was also attempted to confirm the presence of dimers. However, after loading the antibody-TFIIIC complex onto a Superose 6 Increase 10/300 GL column, heterogeneity remained on the grid, which meant the analysis was not possible. Loading the crosslinked apo TFIIIC complex onto a Superose 6 Increase 10/300 GL column however enabled the separation of the two different TFIIIC populations. Since the second peak appeared to contain monomeric TFIIIC, we proceeded to prepare cryo EM grids using this sample.

3.2.5.3. SEC-MALS analysis of the crosslinked and non-crosslinked TFIIIC-DNA complex

Before proceeding into cryo-EM with the crosslinked apo TFIIIC sample, we were prompted to investigate whether we could identify dimers in the crosslinked DNA-bound-TFIIIC complex via SEC-MALS. In particular, it was of interest to analyze whether the peak eluting at 12 ml, that was shown to contain TFIIIC-DNA complexes, also contained TFIIIC dimers, as observed for the crosslinked apo TFIIIC complex.

The crosslinked DNA bound TFIIIC sample showed two peaks as previously observed (Figure 94). The first peak had a molecular weight of 1.26 MDa and the second peak had a molecular weight of 555 kDa. The molecular weight estimations suggested that monomeric apo TFIIIC eluted in the second peak, as previously suggested. Furthermore, these results were in accordance with the observed molecular weight of un-crosslinked apo TFIIIC using SEC-MALS setup at the Francis Crick Institute. According to the estimated molecular weight, the first peak of the crosslinked DNA-bound sample could correspond to dimers or TFIIIC-

DNA complexes with an overestimated molecular weight due to the presence of the nucleic acids.

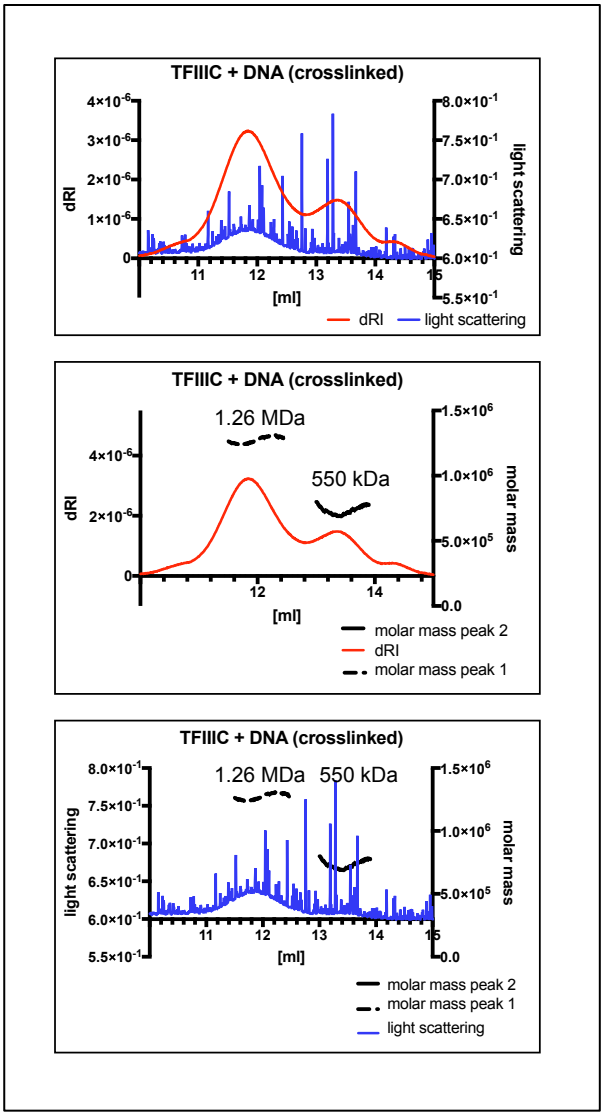


Figure 94: In-house SEC-MALS analysis of crosslinked TFIIIC bound to the DNA.
Top row: The refractive index (dRI) and light scattering is plotted against the elution volume for crosslinked TFIIIC bound to the DNA. Middle row: The dRI is plotted against the molar mass and the elution volume for crosslinked, DNA bound TFIIIC. Bottom row: The light scattering is plotted against the molar mass and the elution volume for crosslinked DNA bound TFIIIC.

To determine if the over-estimation of the molecular weight was caused by the presence of the DNA or by the stabilization of dimers from the

crosslinker, a SEC-MALS analysis was performed of the uncrosslinked TFIIIC-DNA sample. One main peak at 12.5 could be identified and a shoulder at 13.5 ml (Figure 95). For each peak, the molar mass was plotted against the dRI and the light scattering signal to obtain the molecular weight. For the first peak a molecular weight of 680 kDa was obtained and for the second peak a molecular weight of 583 kDa was measured.

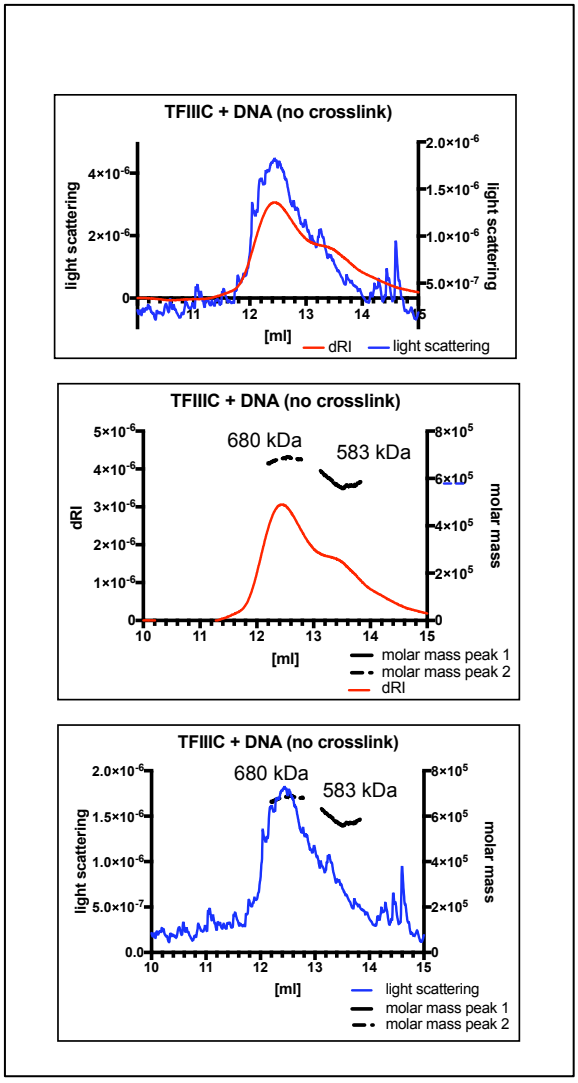


Figure 95: In-house SEC-MALS analysis of un-crosslinked TFIIIC bound to the DNA.
Top row: The refractive index (dRI) and light scattering is plotted against the elution volume for crosslinked TFIIIC bound to the DNA. Middle row: The dRI is plotted against the molar mass and the elution volume for crosslinked DNA bond TFIIIC. Bottom row: The light scattering is plotted against the molar mass and the elution volume for crosslinked DNA bound TFIIIC.

These experiments suggested that it was not the binding to the DNA, but rather the crosslinking reaction that stabilized the TFIIIC dimer conformation. Taken together, the 2D classes of the two peaks observed for crosslinked apo TFIIIC and the significant difference in size of the two crosslinked apo TFIIIC EM maps, it strongly suggested that the complex eluting in the first peak, most likely corresponded to a higher oligomeric state (dimer) of TFIIIC.

A dimeric form of apo TFIIIC eluting at 12 ml, suggested an explanation for the challenges that arose during the processing of the DNA bound TFIIIC complex. Binding of the DNA shifted the TFIIIC elution peak from its original position at 14 ml to approximately 12 ml, where the potentially dimeric form of crosslinked apo TFIIIC eluted. Most likely, the fraction that was used for the preparation of DNA bound TFIIIC thus contained a mixture of crosslinked TFIIIC-DNA complexes, crosslinked dimeric TFIIIC bound and not bound to the DNA, crosslinked TFIIIC that lost the DNA-binding during the blotting process as well as some degradation products and artifacts. Together with the fact that TFIIIC was reportedly a flexible complex, this could have been the cause for the challenging data processing of the DNA bound TFIIIC sample. Pursuing cryo-EM of the monomeric crosslinked apo TFIIIC would thus provide way to overcome the previous difficulties.

3.2.5.4. Cryo-electron microscopy studies of the apo yeast TFIIIC

From the previous trouble-shooting, decreasing the background noise caused by the carbon layer to obtain a higher SNR facilitated the particle alignment. Thus, unsupported gold grids were prepared according to section 2.7.2.2. The grids (LJ017) were taken to the Astbury Centre for Structural Biology and data was collected over 72h. The micrographs

(Figure 96) showed globular particles that could be picked individually. Imaging at higher defocus facilitated the recognition of the particles (Figure 96A) but TFIIIC was still visible even at close to focus parameters (Figure 96B). 2,682 micrographs were taken with a Falcon III EC camera in counting mode over the 72h.

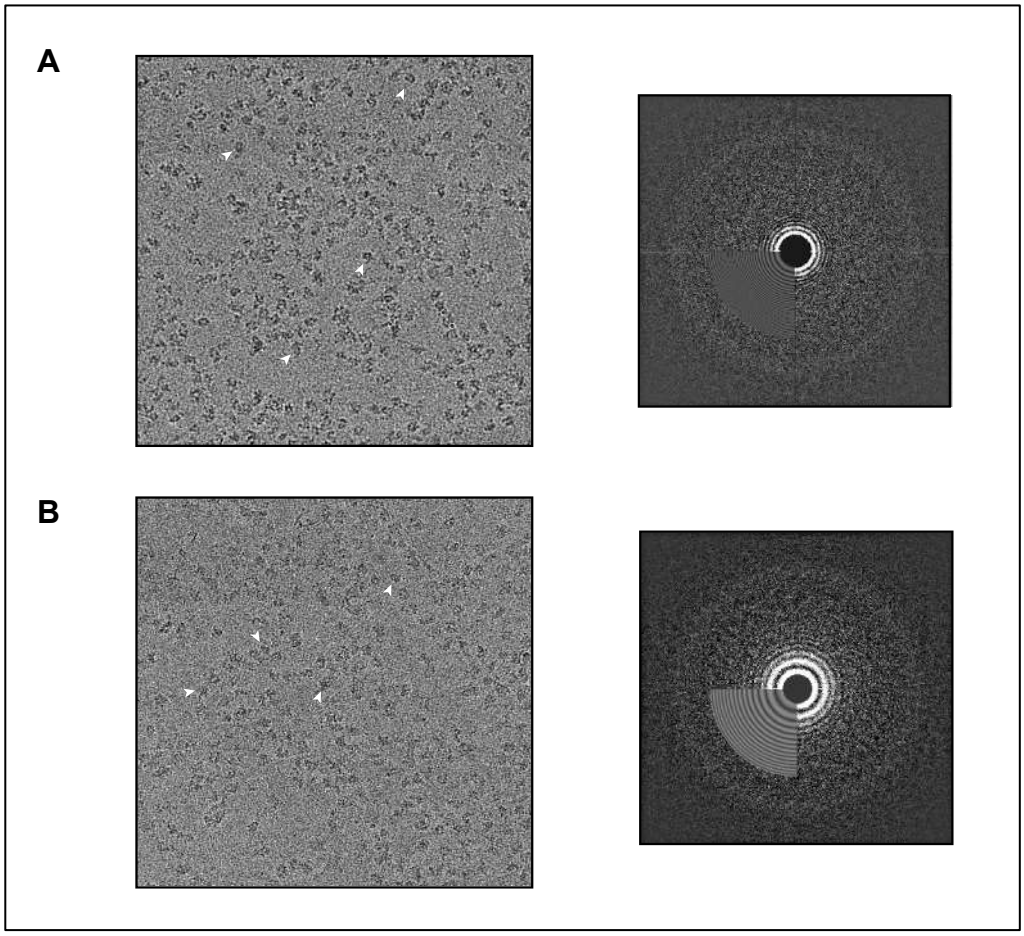


Figure 96: Apo TFIIIC cryo-EM micrographs taken at the Krios 1 in counting mode at the Astbury Centre for Structural Biology in Leeds.

A: Defocused micrograph with its respective thon rings (defocus value of 3.43 μM). B: Less defocused micrograph with its respective thon rings (defocus value of 1.90 μM). TFIIIC particles are shown with white arrows.

The micrographs were manually sorted according to their ice thickness in cisTEM and 314 micrographs were discarded. From the remaining 2,368 micrographs, 199,385 particles were picked in cisTEM using a conservative approach and imported into Relion 3.0. After an initial 2D

classification in Relion 3.0 with a mask of 324 Å, 144,156 particles were selected. These particles were classified into the 2D classes shown in Figure 97. The obtained SNR of the classes was fairly strong and some internal domain features became visible but no secondary structure details were identified.

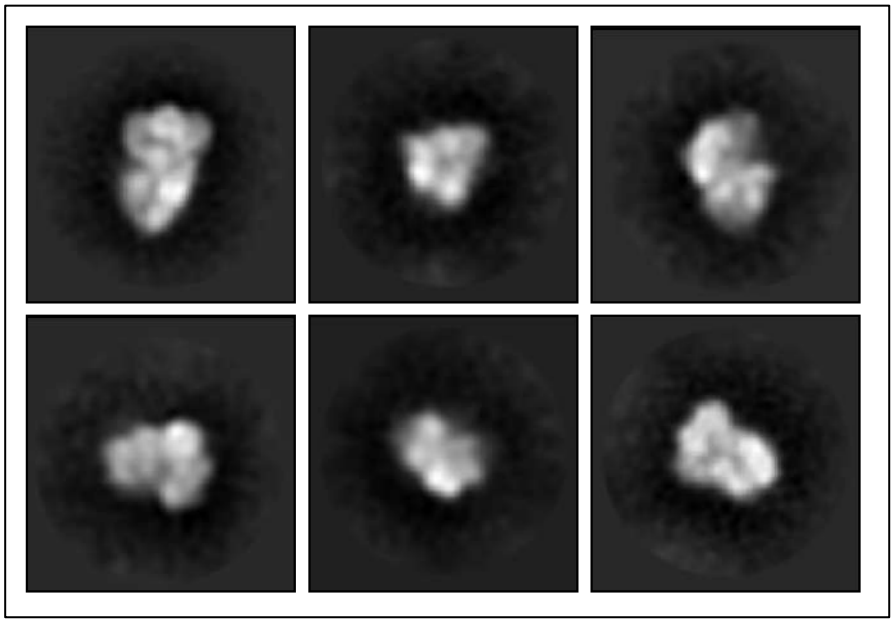


Figure 97: Initial 2D classes of apo yeast TFIIIC.
The 2D classes were obtained using Relion 3.0 with a mask diameter of 324 Å.

To increase the resolution, a hierarchical masking process was started. Four initial 3D classes were built from the particle set and a mask around the half-moon shaped domain of Class 2 (Figure 98A), which was thought to be τB, was created. Using this mask, the particle set was 3D classified again into four classes (Figure 98B). The detail inside the τB subcomplex increased significantly compared to the previous structures. To further tighten the mask, the τ60/ τ91 crystal structure (2J04) was fitted into the map; into the top domain of Class 1 (Figure 98B). The particles of Class 1 and Class 2 were pooled (106,218 particles) and the previously generated tighter mask around τ60/ τ91 was applied. Classifying the τ60/ τ91 domain into four classes led to structures of approximately 10 –

11 Å resolution (Figure 98C). This was a significant increase compared to the structures obtained previously, however no secondary structure details were visible to fully unambiguously place the crystal structure.

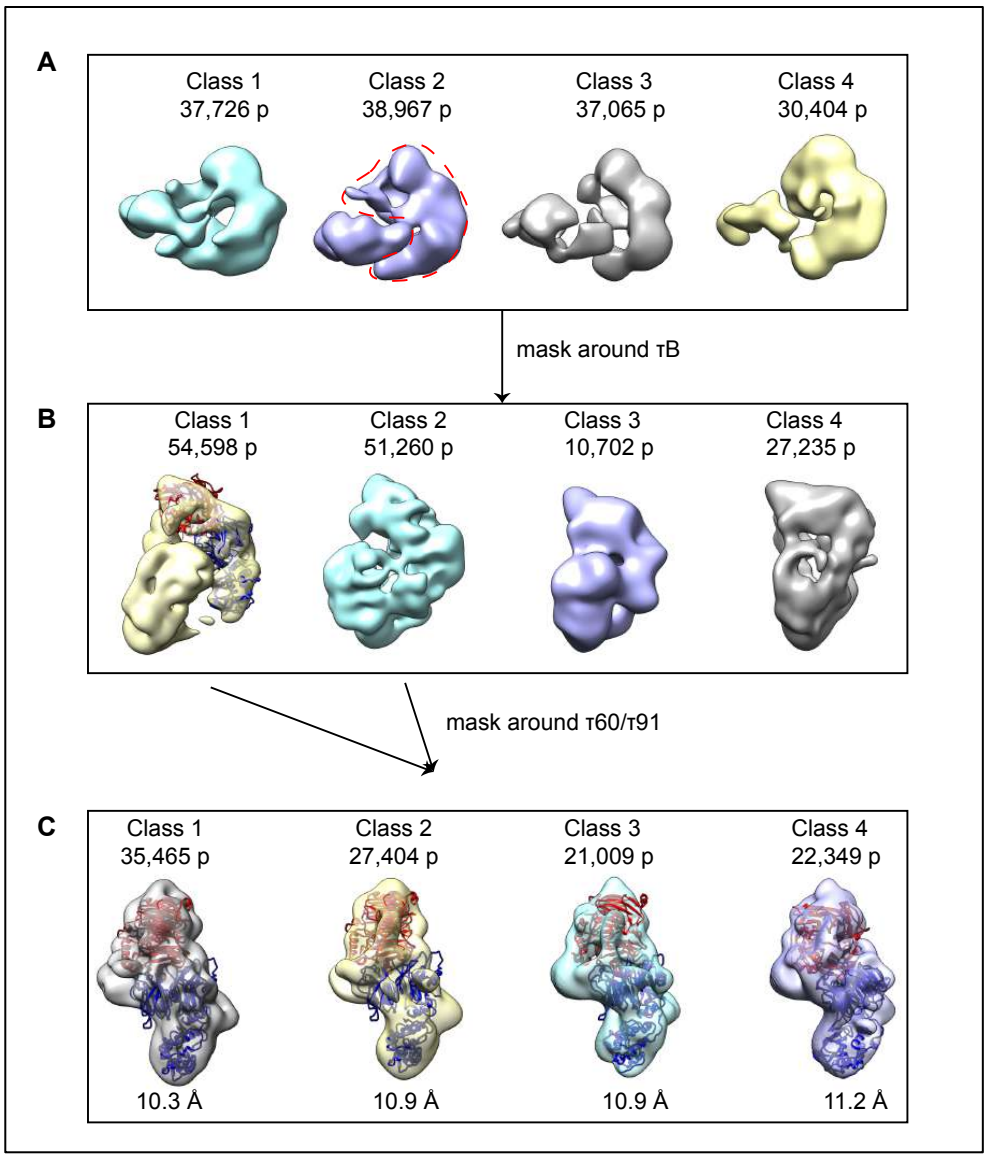


Figure 98: Hierarchical masking of apo TFIIIC.

A: 144,162 particles were initially classified into four 3D classes. From the most populated class (class 2), a mask around the suspected area of *τB was generated (marked with a red dashed line). B: All particles were classified again using Class 2 as a reference and masking around τB. The particles in resulting classes 1 and 2 were pooled, the crystal structure pdb: 2J04 (τ60 shown in blue, τ91 shown in red) was fitted into Class 1 and a mask was generated around that. C: Using Class 1 from C as a reference and masking around 2J04, resulted in four different 3D classes that were all at a resolution between 10 and 11 Å.*

The Fourier Shell Correlation and an angular distribution plot for Class 2 (Figure98C) is shown exemplary in Figure 99.

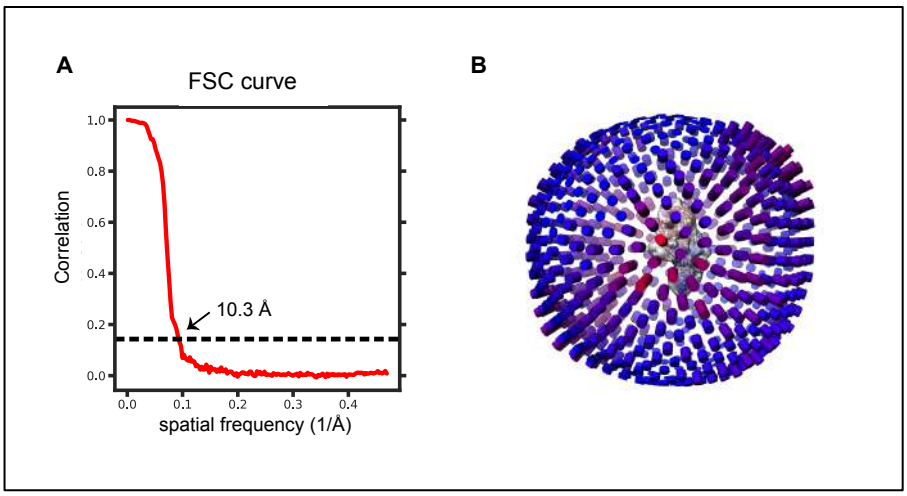


Figure 99: FSC curve and angular distribution of masked region.

A: Fourier Shell Correlation of Class 2, Figure98C. B: Angular distribution plot.

At this point, we hypothesized that increasing the resolution of a sub-part of the apo TFIIIC complex would potentially lead to a loss of alignment in the rest of the TFIIIC particle. Thus, 2D classes without alignment of the whole TFIIIC complex were generated from Class 1 in Figure 100. It was interesting to observe that only the masked area of the particles showed structural information whereas the rest of the protein complex was featureless. This suggested that aligning around one area in TFIIIC led to mis-alignment of the rest of the TFIIIC complex. This could have been caused by conformational heterogeneity or a strong flexibility of the complex. Furthermore, particle misalignment and potential reference bias caused by an arbitrary mask, could have caused the 2D classes shown in Figure 100.

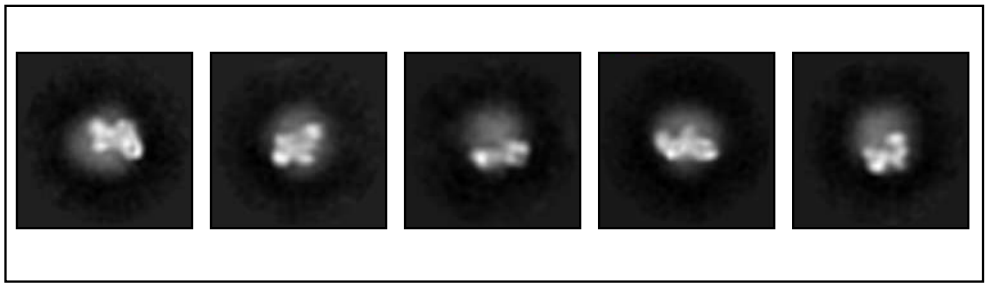


Figure 100: 2D classes without alignment of the apo TFIIIC complex.

The 106,218 particles from Figure 98A were additionally sub-classified into four 3D classes without using any mask. The most promising 3D class, containing 29,460 particles, was refined and post-processed to a resolution of 15.4 Å, which was the highest resolution obtained so far for the apo TFIIIC complex (Figure 101).

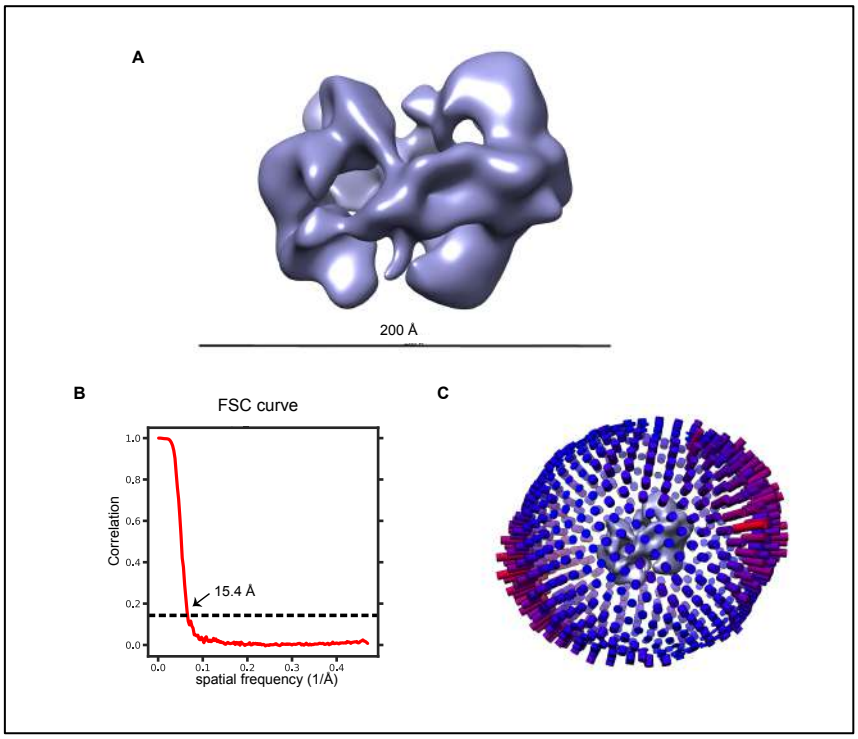


Figure 101: 3D map obtained for holo-TFIIIC.

A: 29,460 particles were used to reconstruct the map at a resolution of 15.4 Å using Relion 3.0. B: Fourier Shell Correlation. C: Angular distribution plot.

In order to test whether TFIIIC was subjected to strong flexibility, it was decided to try a multi-body refinement, which was implemented as a new feature into Relion 3.0. This would give the opportunity to locally increase the resolution and to study the movement of the complex. Therefore, a consensus 3D map of apo TFIIIC was generated to align all particles according to this map (Figure 102A). The map showed two main domains which were similar to those obtained during the sub-classification (Figure 101). It was thus possible to divide the complex into two subcomplexes as indicated in Figure 102A and a mask around τA and a mask around τB were defined. By masking around the different subcomplexes, they were defined as two rigid bodies so that the movement of the subcomplexes with respect to each other could be studied using the 3D multi-body refinement option in Relion 3.0. Furthermore, after the 3D multi-body refinement, each rigid body could be post-processed individually, taking the movement into consideration, which could increase the resolution. The reconstruction of τA is shown in Figure 102B and reconstruction of τB is shown in Figure 102C. In the reconstruction of τA, it was not possible to place any crystal structure yet. Interestingly, by specifically masking around τA, the density became much more defined and contained more features compared to the τA densities obtained from holo-TFIIIC (Figure 98, top row). In the reconstruction of τB, the crystal structure of τ60/τ91 (2J04) could be placed. Most likely, the rest of the density could be part of τ138 and the N terminus of τ91. From each of the reconstructed subcomplexes, 2D classes without alignment were generated to visualize how well the particles were aligned during the multi-body refinement (Figure 102D, E). Thereby the previously applied mask was removed and the whole particle was visible. The masked area of the 2D classes showed a significant amount of detail and secondary structure information which could not be observed previously. Outside of the masked area, only a featureless density was visible, accounting for the other half of the

complex. This was observed in Figure 98D when the mask around τ60/τ91 was removed. It indicated that by masking and aligning around one half of the complex, no detail about the other subcomplex was visible. Most likely, τB, or at least the part of τB containing τ60/τ91, was slightly less flexible than τA and was thus driving the alignment (Figure 98A). Thereby some of the density of τA got averaged out and details were lost when the complex was aligned as a whole.

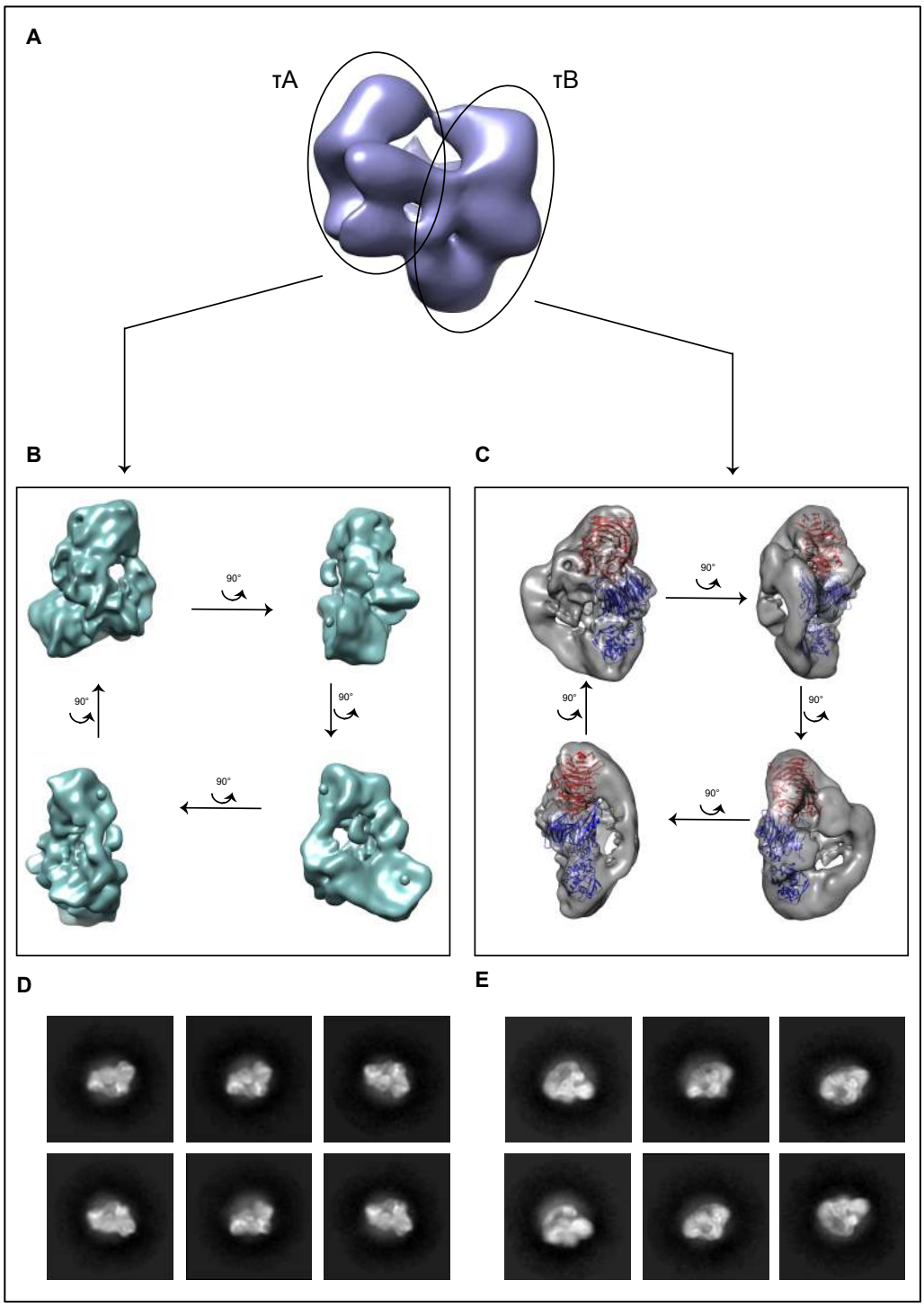


Figure 102: Apo TFIIIC consensus refinement and individual τA and τB refinement.

A: Consensus refinement of apo TFIIIC using 106,218 particles which was used to generate a mask around τA and a mask around τB. B: Refinement of the τA subcomplex. D: 2D classes generated without alignment from the masked refinement of τA. C: Refinement of the τB subcomplex. D: 2D classes generated without alignment from the masked refinement of τB.

The Fourier Shell Correlation and the angular distribution plots for τA and τB are shown in Figure 103.

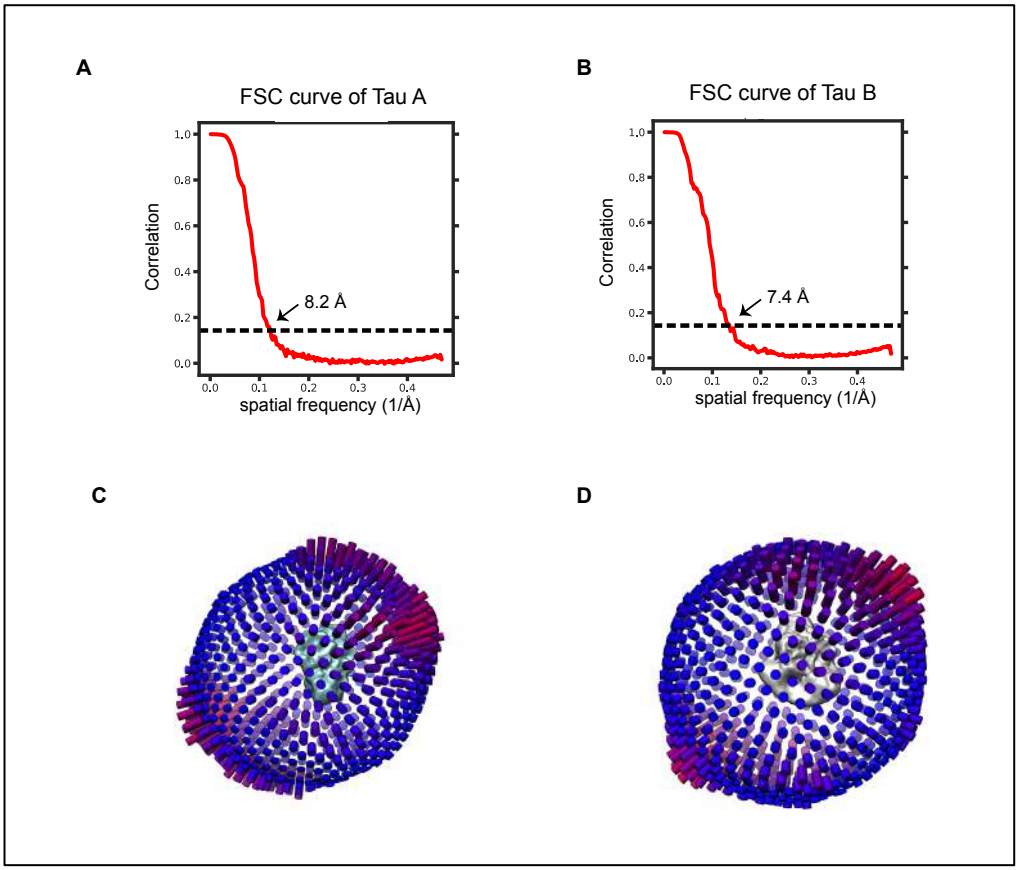


Figure 103: FSC curves and angular distribution plots of τA and τB refinements.
A, B: Fourier Shell Correlations. C, D: Angular distribution plots.

For each 3D multi-body refinement, the movement along the six principal components (PC) that contributed the most to the variance in the data (Figure 104) was analyzed. In the resulting volume series, the movements were visualized and are represented with arrows. It could be observed that both subcomplexes show movement with respect to each other. It is not known yet whether the observed movement was due to continuous movement or rather due to several different conformations.

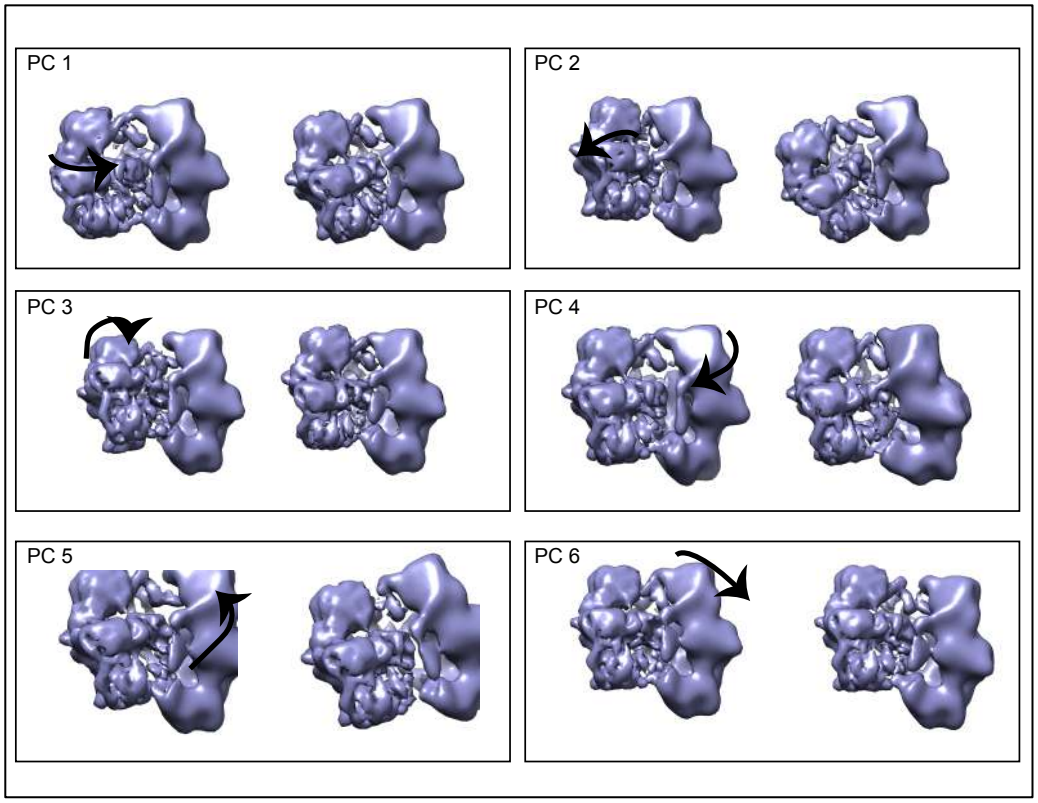


Figure 104: 3D multi-body refinement of apo TFIIIC using Relion 3.0.

The movement along the six principal components was studied. PC = principal component.

It would be interesting to analyze the movement inside each of the subcomplexes separately. As from this data, it could not be unambiguously confirmed that movement inside each subcomplex is happening. However, there is strong evidence suggesting movement inside τA, in particular around the TPR domain of TFIIIC102 in order to interact with the components of TFIIIB (Kassavetis & Geiduschek 2006; Male et al. 2015; Moir et al. 2000; Kassavetis et al. 1992). Furthermore, the area where the crystal structure of τ60/τ91 (2J04) was fitted, could not be solved to higher resolution, suggests movement is present there too. In the atomic model 2J04, two τ60/τ91 subcomplexes were crystallized per asymmetric unit of the crystal. Overlaying both showed flexibility inside the crystal structure (Figure 105).

It was observed that there was a clear movement of the C-terminal domain of τ60 of approx. 4Å. The crystallized part of τ91 (165-672) appeared to be much more stable. However, the first 165 amino acids were not crystallized, which implies that they might be part of a flexible region.

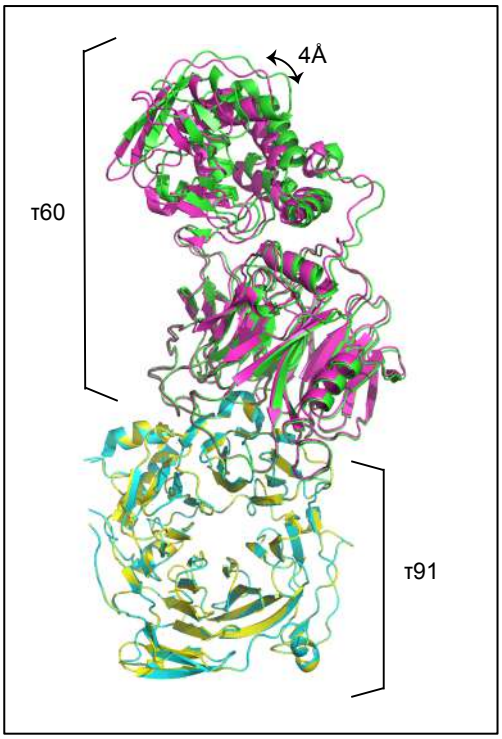


Figure 105: 2J04 overlay of the τ60/τ91 heterodimer of the asymmetric unit of the crystal.

These different states, together with the flexibility of the TFIIIC complex, were most likely the cause of the alignment issues that prevented the solving of a high-resolution structure. The best processing strategy for the cryo-EM data of this complex thus seemed to be hierarchical masking. However, even using tight masks did not lead to high resolution. Possible explanations for this include that the masks could not be placed unambiguously, that it was potentially cut off between domains, that the the particles were poorly aligned for the generation of the consensus map and the flexibility itself of the complex. The results from the 3D multi-body refinement gave evidence for the flexibility of the TFIIIC complex. The

cryo-EM analyzes concluded the studies of the yeast TFIIIC complex for this thesis.

3.2.6. Comparison of structures obtained from *S. cerevisiae* TFIIIC

Throughout the PhD, several cryo-EM data collections were performed of apo and DNA bound TFIIIC. Figure 106 shows a comparison of the different reconstructions that were obtained. In Figure 106A, the reconstructions of the TFIIIC-DNA complex, obtained from the data set collected at eBIC (shown in green), the large data collection in Leeds using the Falcon III EC camera (shown in blue) and using the volta phase plate (shown in grey) are displayed. All reconstructions show a bilobed architecture, in which the lobe on the left-hand side is more compact compared to the lobe on the right-hand side. As shown in the overlay of the three structures in B, the overall dimensions of the three maps are similar. The density of the domains however is not identical. This is most likely due to the low resolution and due to the observed flexibility of TFIIIC. Obtaining similar sized reconstructions from all data sets, that were processed independently from each other, suggested that the maps are good starting models to generate higher resolution reconstructions of the TFIIIC-DNA complex in the future.

Comparing the apo TFIIIC reconstruction (purple model) with the highest resolution TFIIIC-DNA complex reconstruction (blue model) in Figure 106C, seemed to suggest that the complex as a whole extends upon binding to the DNA, which was observed also during the SEC-SAXS experiments (section 2.4.9). The whole complex seems to open up and the two main lobes appear to be moving away from each other. A similar process could possibly also occur when TFIIIC transitions between a compact apo conformation and a more elongated conformation. The obtained D_{max} values and the molecular envelopes obtained from the SEC-SAXS experiments appear larger than in the cryo-EM

reconstructions (comparison not shown). This could be due to an overestimation of the size during the SEC-SAXS analysis due to movements and flexible regions of TFIIIC.

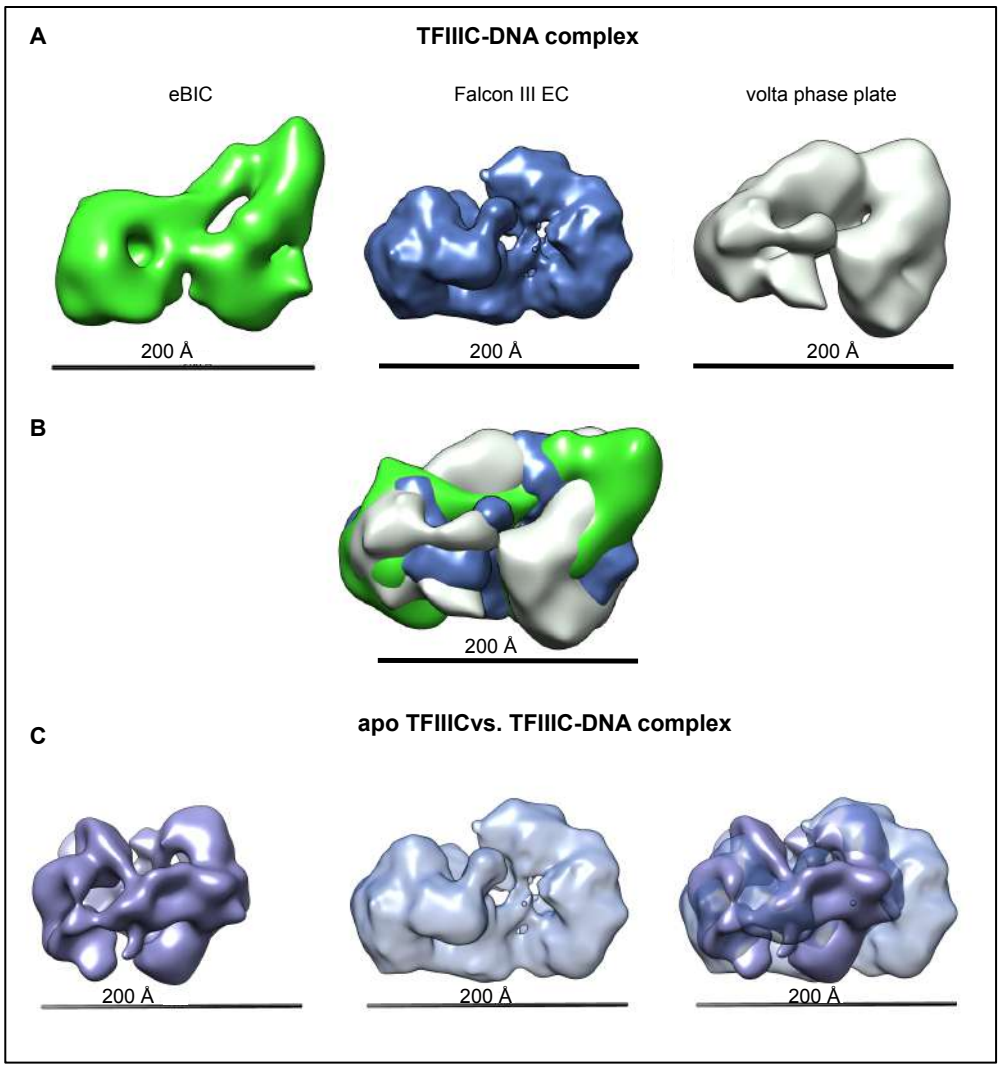


Figure 106: Comparsion of S.cerevisiae TFIIIC reconstructions.

A: Different reconstructions obtained from the data collection in eBIC (green), the Falcon III EC data collection in Leeds (blue) and the data collection using the volta phase plate (grey) of the TFIIIC-DNA complex. B: Overlay of the three TFIIIC-DNA complexes. C: Comparison of apo TFIIIC (purple) with the TFIIIC-DNA complex (blue) obtained using the Falcon III EC camera.

In summary, this comparison suggests that the observations that TFIIIC extends upon binding to the DNA are consistent. The apo TFIIIC complex appears more compact, as observed also in the SEC-SAXS analysis.

3.2.7. Summary of *S. cerevisiae* TFIIIC characterization

Using the biGBac cloning system, I was able to successfully express and purify the yeast holo-TFIIIC complex in high yields and to high purity. I was able to characterize the complex extensively biochemically and biophysically by using EMSAs, SEC-MALS, SEC-SAXS as well as native mass spectrometry experiments that were performed in collaboration with Prof Perdita Barran's team. Furthermore, I was able to obtain the first cryo-EM reconstruction of the apo TFIIIC complex at 15 Å, as well as to obtain reconstructions of the subcomplexes τA and τB to a resolution of 7 – 8 Å. Low resolution reconstructions of the DNA bound TFIIIC complex were also obtained using cryo-EM.

Using this integrated structural biology approach, we were able to identify several compact and extended conformations of TFIIIC as well as the presence of potential dimers. These observations allowed us to understand why we were not able to obtain a higher resolution of the TFIIIC complex using a limited number of particles and the current cryo-EM setup.

Furthermore, I was able to show that τB could not be purified in isolation with the tested expression and purification protocols. The experiments suggested that τ138 required τA in order to stay soluble. Thereby, I was also able to show that a four-subunit TFIIIC complex, consisting of τ138 and τA could be purified. Additionally, I also studied the individual DNA binding of τA, τ60/τ91 and TFIIIC. We could observe that an A Box was sufficient for the formation of a τA-DNA complex that however did not show a single, clear DNA-binding, as expected due to its lower DNA-binding affinity (Stillman & Geiduschek 1984). Most of the binding affinity of TFIIIC was reported to be mediated via the B Box (Stillman & Geiduschek 1984), thus we also studied the binding of TFIIIC and τ60/τ91 to a B Box DNA construct. We were able to see the formation of a stable

TFIIIC-DNA complex, whereas τ60/τ91 was only able to bind to the DNA unspecifically.

4. Concluding remarks and outlook

TFIIIC is a key transcription factor required for transcription initiation in the RNA Polymerase III system. It is a large, multisubunit complex that is comprised of two subcomplexes, τA and τB (Schramm & Hernandez 2002). Functional or sequence homologues were found between *S. pombe, S. cerevisiae* and *H. sapiens*, thus emphasizing the conservation of the TFIIIC complex (Hsieh et al.1999a; Hsieh et al. 1999b; L'Etoile et al. 1994; Lagna et al. 1994; Sinn et al. 1995; Dumay-Odelot et al. 2007). Besides observations at low-resolution using scanning transmission electron microscopy (Schultz et al. 1989), a crosslinking mass spectrometry analysis suggesting an overall organization of the six different subunits (Male et al. 2015) and crystal structures of isolated parts of the complex (Male et al. 2015; Mylona et al. 2006; Taylor et al. 2013a), the overall structure of the human and the yeast TFIIIC complex remains elusive until now. This thesis reports structural, biochemical and biophysical studies of the yeast and the human TFIIIC complex.

4.1. Human TFIIIC

It was possible to optimize an expression and purification protocol of the whole human TFIIIC complex and to obtain a first overall reconstruction of the human apo TFIIIC complex at 30 Å resolution via negative stain. Furthermore, the binding to different DNA constructs that contained only an A Box and only a B Box was tested which demonstrated that the presence of an A Box was sufficient for the formation of a DNA-bound complex. It could also be shown that τB was mostly insoluble when expressed in isolation, whereas it was possible to express τA in isolation. However, the presence of chaperones in the τA sample suggested that

τA requires a stable binding partner or that potentially the buffer conditions could be optimized. Expressing a four-subunit complex, consisting of τA and TFIIIC220, seemed to overcome part of this problem. However, it was not possible to obtain any preliminary structural reconstruction of this complex yet due to low yields and time restrictions.

Regarding the main objectives of this PhD, it was not possible to entirely understand the molecular mechanism behind the human TFIIIC function. A preliminary reconstruction of the holo-TFIIIC complex can be used as a starting point for further structural studies, however the individual architecture of the subcomplexes remains elusive still. In order to answer these questions in the future, it is important to increase the yield of TFIIIC. Most likely, this can be done by optimizing the purification conditions rather than the expression conditions since the bottleneck of the purification proved to be the binding to the affinity tag. It could be shown that a stoichiometric TFIIIC complex could only be obtained by binding TFIIIC via its affinity tag on τB. In contrast, using an affinity tag on τA led to an over-representation of the τA subunits. The position of the tag has already been tested at the C-termini of the τB subunits TFIIIC110, TFIIIC90 and TFIIIC220. TFIIIC220 proved to be the most suitable subunit to be tagged at the C-terminus. However, it would be interesting to see whether a tag at the N-terminus at any of these subunits would enhance the binding reaction.
Additionally, it could also be worth testing more buffer conditions to stabilize TFIIIC. Obtaining a higher yield of TFIIIC would allow to test different DNA-binding conditions, buffers and grid preparation protocols for the preparation of negative stain as well as eventually cryo-EM grids without constantly purifying TFIIIC.

Regarding the structural studies, the binding of TFIIIC to specific antibodies or nanobodies (Pleiner et al. 2015) to label domains could be

tested. This might be an important experiment to do since no structural data about the human TFIIIC complex is available yet. This would allow to assign domains of the reconstruction to specific subunits. Besides the structural studies, an increased amount of TFIIIC would also allow us to optimize the HAT Assay conditions to obtain a reliable readout.
It would also be interesting to send human TFIIIC for a native mass spectrometry analysis to compare its spectrum to the one obtained from the yeast TFIIIC complex and to see which subunits get ejected from the human complex. In particular, it would be interesting to investigate whether human TFIIIC is also present in different compact and elongated conformations via native mass spectrometry or SAXS experiments.

On the condition that it is possible to optimize the four-subunit TFIIIC purification and to be able to express and purify the TFIIIC90/TFIIIC110 heterodimer, it will be interesting to study what influence the subcomplex has on the four-subunit complex. Whether it enhances the DNA-binding or potentially has an inhibiting role in binding to the DNA. Regarding the DNA-binding, it will also be of interest to study the influence of longer linkers between the A and the B Box on the conformation of TFIIIC since most variability in gene length arises from the presence of introns. It was observed before that the region of the DNA where TFIIIC binds, is protected from MNase cleavage (Nagarajavel et al. 2013). Increasing the length of introns by 10bp however showed some vulnerability in the intronic region towards the MNase cleavage. The authors thus suggested that TFIIIC binds in a dumb-bell like manner to the DNA, loops out the DNA between the A and the B Box but does not expose it to the outside. An increasing length of introns however leads to exposure of this loop to the outside which makes it prone to MNase cleavage (Nagarajavel et al. 2013). Indeed, looped DNA could only be observed by artificially increasing the length of the gene between the A and the B Box (Schultz et al. 1989). It would be interesting to see whether it is possible to confirm the formation of DNA loops by TFIIIC using structural analysis or by

performing a FRET analysis in which two DNA bases could be tagged and to investigate what influence the length has on the conformation.

4.2. Yeast TFIIIC

Since working solely with the human TFIIIC complex could not clarify on the main objectives of this PhD, the functionally conserved *S. cerevisiae* TFIIIC complex was studied. Yeast is a well-characterized model organism and could serve as a platform to understand the molecular mechanism and structure of the human TFIIIC better. In contrast to the human TFIIIC complex, it was possible to express and purify the whole yeast TFIIIC complex in high purity and high yield. It was also possible to purify the subcomplexes τA, τ60/τ91 and a four-subunit TFIIIC complex, consisting of τA and τ138. Furthermore, it was possible to reconstitute TFIIIC via a co-infection and via reconstituting it from its purified component parts, the four-subunit complex and τ60/τ91.

Each of the purified complexes were used to study the DNA-binding individually. It was possible to observe that τA specifically bound to the poorly conserved A Box DNA construct but was not able to bind to the B Box DNA. TFIIIC seemed to be able to form a DNA bound complex with either the A Box DNA or the B Box DNA or a DNA construct containing both, the A and the B Box. There still remain some doubts whether it was only τA binding to the A Box DNA which might have caused τB to dissociate as suggested before (Male et al. 2015) or whether a full TFIIIC-DNA complex formed upon binding to the A Box DNA construct. This question could be clarified by performing a Western Blot of the EMSA with a monoclonal Strep antibody that binds to τ91 (τB). A signal at the same position where the DNA-bound complex was detected would thus confirm the presence of τB subunits.

Analyzing yeast TFIIIC via native mass spectrometry analysis, performed by Prof Perdita Barran's lab in Manchester, disclosed that TFIIIC is present in five different conformations, reaching from a compact form to a highly extended form. It could also be observed that τ60 and τ91 are more prone to getting ejected from the elongated complex conformation. Two populations, a more compact and a more extended conformation of TFIIIC could also be observed in the SEC-SAXS experiment and would explain the two protein and DNA-binding bands observed in native PAGEs. Of particular interest would be to investigate what triggers each conformation of TFIIIC and what their biological meaning is. So far, it was not possible to separate the different conformations of complexes as observed in native mass spectrometry. An important further experiment would be to look at posttranslational modifications of each of the complex conformations individually. Thereby, potential mutation sites could be identified which could be mutated as a potential trigger to force TFIIIC into one specific conformation.

Regarding the structural analysis of the yeast TFIIIC complex, it was possible to obtain a reconstruction of the apo yeast TFIIIC complex at 15.4 Å and to refine some part of the structure, in particular the domain in which the atomic model of τ60/τ91 could be placed, to a resolution of 10 – 11 Å. Most likely, it was not possible to obtain a higher resolution due to the flexibility of the complex. Masking around certain domains in TFIIIC led to a higher resolution inside the mask. Generating 2D classes without alignment from the orientation of the particles from the masked refinement however showed that the rest of the particle outside of the mask was not aligned but only represented as a big featureless density. This observation was complemented with a multi-body refinement in which the movement along the principal component axis between τA and τB was visualized. Of particular interest would be to analyze the movement inside each of the subcomplexes and to see whether this flexibility could be decreased by binding TFIIIC together with its binding partners TFIIIB or

TFIIIA to the DNA. Besides potentially reducing the flexibility, it would also give insights about what happens on the molecular level upon the binding reaction. Furthermore, labelling apo or DNA-bound TFIIIC with specific nanobodies as reported in (Pleiner et al. 2015), would allow to unambiguously assign domains of the complex.

Besides studying apo TFIIIC structurally, a low-resolution reconstruction of the DNA-bound TFIIIC complex could be obtained. It was observed that without crosslinking, TFIIIC fell apart on the grid, which made it inevitable to use the crosslinker. However, the crosslinking reaction favoured or stabilized the formation of TFIIIC dimers that could be separated from crosslinked apo TFIIIC by size exclusion chromatography but not from the DNA-bound TFIIIC complex. Together with the hypothesis that some TFIIIC-DNA complexes lost the DNA-binding during the blotting process, there was potentially a mixture of all these complexes present on the grid. With a limited number of micrographs and thus particles, it was not possible to classify the different conformations. An interesting approach would be to collect data of the DNA-bound complex again at the Falcon III EC camera and to merge datasets in an attempt to significantly increase the number of particles. This would lead however to some computational difficulties since the boxsize of the DNA-bound TFIIIC complex was very large and already difficult to process in their individual datasets. Potentially it could also be an idea to optimize a grid preparation protocol for unsupported copper or gold grids, similar to the one used for apo TFIIIC, to reduce the background noise and to obtain a better SNR. This could also show whether the resolution of the TFIIIC-DNA complex could be increased just by reducing the noise. Exploring graphene as a support film instead of carbon, that was used in this thesis for the preparation of TFIIIC-DNA grids, would also be an option. Another idea would be to bind TFIIIC to Biotin labelled DNA which is bound to Streptavidin beads. Instead of loading this complex onto a gel filtration column, the DNA could be cleaved from the beads by a DNA restriction

enzyme and subsequently analyzed via cryo-EM. Using this approach, apo TFIIIC dimers could be eliminated from the grid.

Regarding the main objectives of this thesis, the performed studies allowed to characterize the architecture of holo-TFIIIC and as well as the subcomplexes at low-resolution. It became clear that TFIIIC is mostly present as a monomer in solution but was also found to be present as dimers which could potentially be a concentration effect or could have a biological relevance, similar to RNA Pol I dimerization (Milkereit et al. 1997). Furthermore, this thesis could shed light onto the molecular mechanism of yeast TFIIIC. TFIIIC was shown to extend upon binding to the DNA which might also allow TFIIIC to adapt to different promoter organizations. Furthermore, these observations resemble the canonical and rearranged conformation of TFIID, in which only the rearranged form was able to bind to the promoter DNA (Louder et al. 2016). The presence of compact and extended conformations of apo TFIIIC could represent a possibility to auto-inhibit itself from interactions with binding partners as suggested in (Liao et al. 2006). The different confirmations of TFIIIC were shown to have an impact on the binding of τ60 and τ91. This could therefore suggest that these subunits are potentially less essential for TFIIIC function, which would have to be confirmed in future studies. Following up on Figure 79, the model shown in Figure 107 could be proposed about the molecular mechanism of TFIIIC. Initially, Tau A and Tau B are closely connected, shielding τ131 interactions sites for binding partners. τ91 is stably attached to the complex, whereas τ60 may only be loosely connected to TFIIIC. Upon extension, Tau A and Tau B move away from each other, open up a gap between them which uncovers τ131 interaction sites and rearranges the binding to τ60 and τ91. In the further extended TFIIIC conformation, both subunits, τ60 and τ91 are prone to dissociate from the complex. In the fully extended apo TFIIIC conformation, Tau A and Tau B are even further away from each other,

reducing the interaction sites between them. Upon binding to the DNA, Tau A and Tau B are furthest away from each other to allow for an extension that almost covers the whole gene promoter.

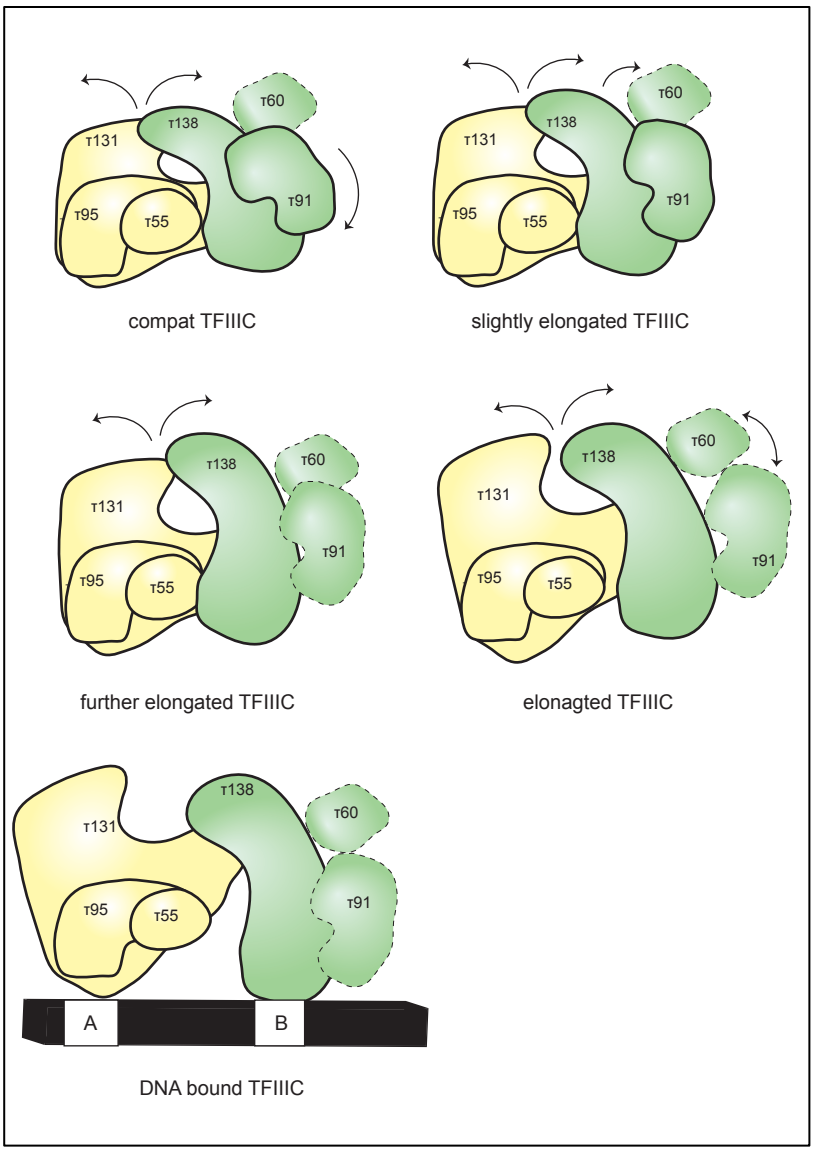


Figure 107: Cartoon showing proposed molecular mechanism of TFIIIC.

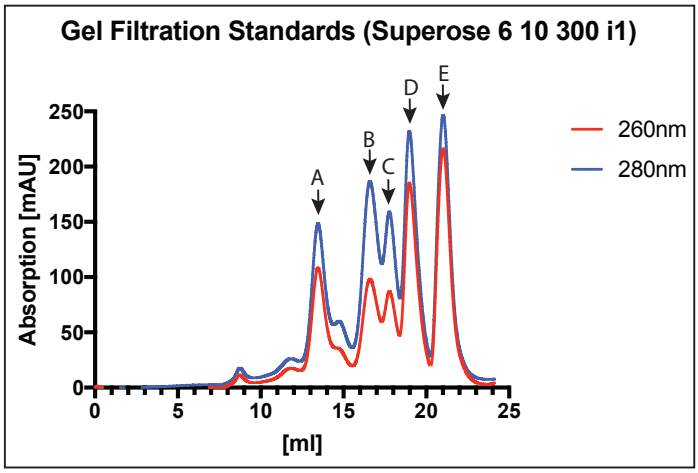
In the compact conformation, Tau A and Tau B are interacting closely. Through transition into the elongated conformation, τ131 and τ138 move further away from each other to open up binding sites on τ131. This rearranges the interaction to τ60 and τ91 which weakens the interactions of both subunits with the rest of the complex. A further extension of TFIIIC is caused by binding to the DNA. Tau A is shown in yellow and Tau B is shown in green. The arrows indicate the direction of the movement. The dashed line indicates that these subunits are prone to dissociate from the complex.

In summary, this thesis provides unprecedented insights into the biochemical behaviour of the yeast TFIIIC complex, its different conformational states and preliminary reconstructions of the apo and the DNA bound TFIIIC complex. Also, an extensive basis for further studying the human TFIIIC complex biochemically and structurally was developed.

5. Appendix

5.1. Gel filtration standards run on a Superose 6 Increase 10/300 GL column

100 µl of the gel filtration standards (Bio-Rad) were loaded onto a Superose 6 Increase 10/300 GL column using the following running buffer: 150 mM NaCl, 20 Hepes pH 8, 1 mM DTT. The elution profile is shown in Figure 108 .



A: Thioglobulin (670kDa)
B: Gamma Globulin (155kDa)
C: Ovalbumin (44kDa)
D: Myoglobin (17kDa)
E: Vitamin B12 (1.35kDa)

Figure 108: Gel filtration standards on a Superose 6 Increase 10/300 GL column.

6. References

Abbe, E., 1873. Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. *Archiv für mikroskopische Anatomie*, 9(1), pp.413–418.

Adrian, M., Dubochet, J., Lepault, J. & McDowall, A.W., 1984. Cryo-electron microscopy of viruses. *Nature*, 308(5954), pp.32–36.

Allison, D.S., Goh, S.H. & Hall, B.D., 1983. The promoter sequence of a yeast tRNAtyr gene. *Cell*, 34(2), pp.655–664.

Altmann, G.G. & Leblond, C.P., 1982. Changes in the size and structure of the nucleolus of columnar cells during their migration from crypt base to villus top in rat jejunum. *Journal of cell science*, 56, pp.83–99.

Archambault, J. & Friesen, J.D., 1993. Genetics of eukaryotic RNA polymerases I, II, and III. *Microbiological reviews*, 57(3), pp.703–724.

Arimbasseri, A.G., Rijal, K. & Maraia, R.J., 2013. Transcription termination by the eukaryotic RNA polymerase III. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1829(3-4), pp.318–330.

Arimbasseri, A.G. & Maraia, R.J., 2015. Mechanism of Transcription Termination by RNA Polymerase III Utilizes a Non-template Strand Sequence-Specific Signal Element. *Molecular cell*, 58(6), pp.1124–1132.

Arimbasseri, A.G. & Maraia, R.J., 2016. RNA Polymerase III Advances: Structural and tRNA Functional Views. *Trends in biochemical sciences*, 41(6), pp.546-559.

Arrebola, R., Manaud, N., Rozenfeld, S., Marsolier, M.C., Lefebvre, O., Carles, C., Thuriaux, P., Consea, C. & Sentenac, A., 1998. Tau91, an essential subunit of yeast transcription factor IIIC, cooperates with tau138 in DNA binding. *Molecular and cellular biology*, 18(1), pp.1–9.

Avery, O.T., Macleod, C.M. & McCarty, M., 1944. Studies of the chemical nature of the substance inducing transformation of Pneumococcal types: Induction of transformation by a deoxyribonucleic acid fraction isolated from Pneumococcus Type III. *The Journal of experimental medicine*, 79(2), pp.137–158.

Bagby, S., Kim, S., Maldonado, E., Tong, K., Reinberg, D. & Ikura, M., 1995. Solution structure of the c-terminal core domain of human TFIIB: Similarity to cyclin A and interaction with TATA-binding protein. *Cell*, 82(5), pp.857–867.

Baker, R.E., Gabrielsen, O. & Hall, B.D., 1986. Effects of tRNATyr point mutations on the binding of yeast RNA polymerase III transcription factor C. *The Journal of biological chemistry*, 261(12), pp.5275–5282.

Bardeleben, C., Kassavetis, G.A. & Geiduschek, E.P., 1994. Encounters of Saccharomyces cerevisiae RNA polymerase III with its transcription factors during RNA chain elongation. *Journal of molecular biology*, 235(4), pp.1193–1205.

Barski, A., Chepelev, I., Liko, D., Cuddapah, S., Fleming, A.B., Birch, J., Cui, K., White, R.J. & Zhao, K., 2010. Pol II and its associated epigenetic marks are present at Pol III-transcribed noncoding RNA genes. *Nature structural & molecular biology*, 17(5), pp.629–634.

Bartholomew, B., Kassavetis, G.A., Braun, B.R. & Geiduschek, E.P., 1990. The subunit structure of Saccharomyces cerevisiae transcription factor IIIC probed with a novel photocrosslinking reagent. *The EMBO journal*, 9(7), pp.2197–2205.

Bartholomew, B., Kassavetis, G.A. & Geiduschek, E.P., 1991. Two components of Saccharomyces cerevisiae transcription factor IIIB (TFIIIB) are stereospecifically located upstream of a tRNA gene and interact with the second-largest subunit of TFIIIC. *Molecular and cellular biology*, 11(10), pp.5181–5189.

Basehoar, A.D., Zanton, S.J. & Pugh, B.F., 2004. Identification and Distinct Regulation of Yeast TATA Box-Containing Genes. *Cell*, 116(5), pp.699–709.

Basu, M.,Boopathi, R., Das, S. & Kundu, T.K., 2019. The Largest Subunit of Human TFIIIC Complex, TFIIIC220, a Lysine Acetyltransferase Targets Histone H3K18. *bioRxiv*, p.513127.

Bedwell, G.J., Appling, F.D., Anderson, S.J. & Schneider, D.A., 2012. Efficient transcription by RNA polymerase I using recombinant core factor. *Gene*, 492(1), pp.94–99.

Berger, I., Fitzgerald, D.J. & Richmond, T.J., 2004. Baculovirus expression system for heterologous multiprotein complexes. *Nature biotechnology*, 22(12), pp.1583–1587.

Best, A.A., Morrison, H.G., McArthur, A.G., Sogin, M.L. & Olsen, G.J., 2004. Evolution of eukaryotic transcription: insights from the genome of Giardia lamblia. *Genome research*, 14(8), pp.1537–1547.

Blair, R.H., Goodrich, J.A. & Kugel, J.F., 2012. Single-Molecule Fluorescence Resonance Energy Transfer Shows Uniformity in TATA Binding Protein-Induced DNA Bending and Heterogeneity in Bending Kinetics. *Biochemistry*, 51(38), pp.7444–7455.

Blattner, C., Jennebach, S., Herzog, F., Mayer, A., Cheung, A.C., Witte, G., Lorenzen, K., Hopfner, KP, Heck, AJ, Aebersold, R. & Cramer, P., 2011. Molecular basis of Rrn3-regulated RNA polymerase I initiation and cell growth. *Genes & development*, 25(19), pp.2093–2105.

Boldon, L., Laliberte, F. & Liu, L., 2015. Review of the fundamental theories behind small angle X-ray scattering, molecular dynamics simulations, and relevant integrated application. *Nano Reviews*, 6(1), p.25661.

Braglia, P., Percudani, R. & Dieci, G., 2005. Sequence context effects on oligo(dT) termination signal recognition by Saccharomyces cerevisiae RNA polymerase III. *The Journal of biological chemistry*, 280(20), pp.19551–19562.

Brenner, S. & Horne, R.W., 1959. A negative staining method for high resolution electron microscopy of viruses. *Biochimica et biophysica acta*, 34, pp.103–110.

Bushnell, D.A., Westover, K.D., Davis, R.E. & Kornberg, R.D., 2004. Structural Basis of Transcription: An RNA Polymerase II-TFIIB Cocrystal at 4.5 Angstroms. *Science*, 303(5660), pp.983–988.

Carlsten, J.O.P., Zhu, X., Lopez, M.D., Samuelsson, T. & Gustafsson, C.M., 2016. Loss of the Mediator subunit Med20 affects transcription of tRNA and other non-coding RNA genes in fission yeast. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1859(2), pp.339–347.

Carter, R. & Drouin, G., 2010. The increase in the number of subunits in eukaryotic RNA polymerase III relative to RNA polymerase II is due to the permanent recruitment of general transcription factors. *Molecular biology and evolution*, 27(5), pp.1035–1043.

Cavallini, B., Heut, J., Plassat, J.L., Sentenac, A., Egly, J.M. & Chambon, P., 1988. A yeast activity can substitute for the HeLa cell TATA box factor. *Nature*, 334(6177), pp.77–80.

Chen, S., McMullan, G., Faruqi, A.R., Murshudow, G.N., Short, J.M., Scheres, S.H. & Henderson, R., 2013. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy*, 135, pp.24–35.

Cheng, Y., 2015. Single-Particle Cryo-EM at Crystallographic Resolution. *Cell*, 161(3), pp.450–457.

Cheung, A.C.M., Sainsbury, S. & Cramer, P., 2011. Structural basis of initial RNA polymerase II transcription. *The EMBO journal*, 30(23), pp.4755–4763.

Cheung, A.C.M. & Cramer, P., 2012. A movie of RNA polymerase II transcription. *Cell*, 149(7), pp.1431–1437.

Chédin, S., Riva, M., Schultz, P., Sentenac, A. & Carles, C., 1998. The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIS-like subunit and is important for transcription termination. *Genes & development*, 12(24), pp.3857–3871.

Chong, S.S., Hu, P. & Hernandez, N., 2001. Reconstitution of transcription from the human U6 small nuclear RNA promoter with eight recombinant polypeptides and a partially purified RNA polymerase III complex. *The Journal of biological chemistry*, 276(23), pp.20727–20734.

Clancy, S., 2008. RNA functions. *Nature Education*,1(1):102.

Clough, R.N., Moldovan, G. & Kirkland, A.I., 2014. Direct Detectors for Electron Microscopy. *Journal of Physics: Conference Series*, 522(1), p.012046.

Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R. & Kornberg, R.D., 2000. Architecture of RNA polymerase II and implications for the transcription mechanism. *Science*, 288(5466), pp.640–649.

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., 2008. Structure of eukaryotic RNA polymerases. *Annual review of biophysics*, 37, pp.337–352.

Crick, F., 1970. Central Dogma of Molecular Biology. *Nature*, 227(5258), pp.561–563.

Crowther, R.A., Amos, L.A., Finch, J.T., DeRosier, D.J. & Klug, A., 1970. Three Dimensional Reconstructions of Spherical Viruses by Fourier Synthesis from Electron Micrographs. *Nature*, 226(5244), pp.421–425.

D'Ambrosio, C., Schmidt, C.K., Katou, Y., Kelly, G., Itoh, T., Shirahige, K. & Uhlmann, F., 2008. Identification of cis-acting sites for condensin loading onto budding yeast chromosomes. *Genes & development*, 22(16), pp.2215–2227.

Dahm, R., 2005. Friedrich Miescher and the discovery of DNA. *Developmental Biology*, 278, pp. 274 -288.

Daly, N.L., Arvanitis, D.A., Fairley, J.A., Gomez-Roman, N., Morton, J.P., Graham, S.V., Spandidos, D.A. & White, R.J., 2005. Deregulation of RNA polymerase III transcription in cervical epithelium in response to high-risk human papillomavirus. *Oncogene*, 24(5), pp.880–888.

Danev, R. & Nagayama, K., 2001. Transmission electron microscopy with Zernike phase plate. *Ultramicroscopy*, 88(4), pp.243–252.

Danev, R. & Nagayama, K., 2011. Optimizing the phase shift and the cut-on periodicity of phase plates for TEM. *Ultramicroscopy*, 111(8), pp.1305–1315.

Danev, R., Buijsse, B., Khoshouei, M., Plitzko, J. & Baumeister, W., 2014. Volta potential phase plate for in-focus phase contrast transmission electron microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 111(44), pp.15635–15640.

Danev, R. & Baumeister, W., 2016. Cryo-EM single particle analysis with the Volta phase plate. *eLife*, 5:e13046.

Danev, R., Tegunov, D. & Baumeister, W., 2017. Using the Volta phase plate with defocus for cryo-EM single particle analysis. *eLife*, 6:e230006.

Darst, S.A., Kubalek, E.W. & Kornberg, R.D., 1989. Three-dimensional structure of Escherichia coli RNA polymerase holoenzyme determined by electron crystallography. *Nature*, 340(6236), pp.730–732.

Dauwerse, J.G. et al., 2011. Mutations in genes encoding subunits of RNA polymerases I and III cause Treacher Collins syndrome. *Nature genetics*, 43(1), pp.20–22.

deHaseth, P.L., Zupancic, M.L. & Record, M.T., 1998. RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. *Journal of bacteriology*, 180(12), pp.3019–3025.

Dieci, G. & Sentenac, A., 1996. Facilitated Recycling Pathway for RNA Polymerase III. *Cell*, 84(2), pp.245–252.

Drapkin, R., Reardon, J.T., Ansari, A., Huang, J.C., Zawel, L., Ahn, K., Sancar, A.& Reinberg, D., 1994. Dual role of TFIIH in DNA excision repair and in transcription by RNA polymerase II. *Nature*, 368(6473), pp.769–772.

Dubochet, J., Adrian, M., Chang, J.-J.& Homo, J.C.,1988. Cryo-electron microscopy of vitrified specimens. *Quarterly Reviews of Biophysics*, 21(2), pp.129–228.

Ducrot, C., Lefebvre, O., Landrieux, E., Guirouilh-Barbat, J., Sentenac, A. & Acker, J., 2006. Reconstitution of the yeast RNA polymerase III transcription system with all recombinant factors. *The Journal of biological chemistry*, 281(17), pp.11685–11692.

Dumay-Odelot, H., Acker, J., Arrebola, R., Sentenac, A. & Marck, C., 2002. Multiple roles of the tau 131 subunit of yeast transcription factor IIIC (TFIIIC) in TFIIIB assembly. *Molecular and cellular biology*, 22(1), pp.298–308.

Dumay-Odelot, H., Marck, C., Durrieu-Gaillard, S., Lefebvre, O., Jourdain, S., Prochazkova, M., Pflieger, A. & Teichmann, M., 2007. Identification, molecular cloning, and characterization of the sixth subunit of human transcription factor TFIIIC. *The Journal of biological chemistry*, 282(23), pp.17179–17189.

Dumay-Odelot, H., Durrieu-Gaillard, S., DaSilva, D., Roeder, R.D. & Teichmann, M., 2010. Cell growth- and differentiation-dependent regulation of RNA polymerase III transcription. *Cell cycle*, 9(18), pp.3687–3699.

Durand, D., Vivès, C., Cannella, D., Pérez, J., Pebay-Peyroula, E., Vachette, P. & Fieschi, F., 2010. NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers. *Journal of structural biology*, 169(1), pp.45–53.

Dvir, A., Conaway, R.C. & Conaway, J.W., 1997. A role for TFIIH in controlling the activity of early RNA polymerase II elongation complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 94(17), pp.9006–9010.

Engel, C., Sainsbury, S., Cheung, A.C., Kostrewa, D. & Cramer, P., 2013. RNA polymerase I structure and transcription regulation. *Nature*, 502(7473), pp.650–655.

Engel, C., Gubbey, T., Neyer, S., Sainsbury, S., Oberthuer, C., Baejen, C., Bernecky, C. & Cramer, P., 2017. Structural Basis of RNA Polymerase I Transcription Initiation. *Cell*, 169(1), pp.120–131.e22.

Engelke, D.R., Ng, S.Y., Shastry, B.S. & Roeder, R.G., 1980. Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. *Cell*, 19(3), pp.717–728.

Erickson, H.P. & Klug, A., 1970. The Fourier Transform of an Electron Micrograph: Effects of Defocussing and Aberrations, and Implications for the Use of Underfocus Contrast Enhancement. *Berichte der Bunsengesellschaft für physikalische Chemie*, 74(11), pp.1129–1137.

Fairley, J.A., Mitchell, L.E., Berg, T., Kenneth, N.S., vonSchubert, C., Silljé, H.H., Medema, R.H., Nigg, E.A. & White, R.J., 2012. Direct regulation of tRNA and 5S rRNA gene transcription by Polo-like kinase 1. *Molecular cell*, 45(4), pp.541–552.

Feigin, L.A. & Svergun, D.I., 1987. *Structure Analysis by Small-Angle X-Ray and Neutron Scattering* G. W. Taylor, ed., Boston, MA: Springer US.

Felton-Edkins, Z.A., Kenneth, N.S., Brown, T.R., Daly, N.L., Gomez-Roman, N., Grandori, C., Eisenman, R.N. & White, R.J., 2003. Direct regulation of RNA polymerase III transcription by RB, p53 and c-Myc. *Cell cycle*, 2(3), pp.181–184.

Fernández-Tornero, C., Moreno-Morcillo, M., Rashid, U.J., Taylor, N.M., Ruiz, F.M., Gruene, T., Legrand, P., Steuerwald, U.& Mueller, C.W., 2013. Crystal structure of the 14-subunit RNA polymerase I. *Nature*, 502(7473), pp.644–649.

Ferrari, R., deLlobet-Cucalon, L.I., DiVona, C., Vidal, E., Lioutas, A., LeDilly, F., Quilez-Oliete, J., Dieci, G., Teichmann, M., de-la-Luna, S. & Beato, M., 2018. TFIIIC dynamically binds Alu elements to control gene expression through chromatin looping. *bioRxiv*, p.455733.

Ferrari, R., deLlobet-Cucalon, L.I., DiVona, C., LeDilly, F., Vidal, E., Lioutas, A., Quilez-Oliete, J., Jochem, L., Cutts, E., Dieci, G., Vannini, A., Teichmann, M., de-la-Luna, S. & Beato, M., 2019. TFIIIC binding to Alu elements controls gene expression via chromatin looping and histone acetylation. *Molecular Cell,* under review.

Flores, A., Briand, J.-F., Gadal, O., Andrau, J.-C., Rubbi, L., VanMullem, V., Boschiero, C., Goussot, M., Marck, C., Calres, C., Thuriaux, P., Sentenac, A. & Werner, M., 1999. A protein-protein interaction map of yeast RNA polymerase III. *Proceedings of the National Academy of Sciences of the United States of America*, 96(14), pp.7815–7820.

Folta-Stogniew, E., 2006. Oligomeric states of proteins determined by size-exclusion chromatography coupled with light scattering, absorbance, and refractive index detectors. *Methods in molecular biology*, 328, pp.97–112.

Franke, D. & Svergun, D.I., 2009. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of applied crystallography*, 42(Pt 2), pp.342–346.

Freundlich, M.M., 1963. Origin of the Electron Microscope. *Science*, 142(3589), pp.185–188.

Gabrielsen, O.S., Marzouki, N., Ruet, A., Sentenac, A. & Fromageot, P., 1989. Two polypeptide chains in yeast transcription factor tau interact with DNA. *The Journal of biological chemistry*, 264(13), pp.7505–7511.

Geiger, S.R., Lorenzen, K., Scheieck, A., Hanecker, P., Kostrewa, D. Heck, A.J.& Cramer, P., 2010. RNA polymerase I contains a TFIIF-related DNA-binding subcomplex. *Molecular cell*, 39(4), pp.583–594.

Gill, S.C., Weitzel, S.E. & Hippel, von, P.H., 1991. Escherichia coli sigma 70 and NusA proteins. I. Binding interactions with core RNA polymerase in solution and within the transcription complex. *Journal of molecular biology*, 220(2), pp.307–324.

Glaeser, R.M. & Taylor, K.A., 1978. Radiation damage relative to transmission electron microscopy of biological specimens at low temperature: a review. *Journal of Microscopy*, 112(1), pp.127–138.

Goodarzi, H., Nguyen, H.C.B., Zhang, S., Dill, B.D., Molina, H. & Tavazoie, S.F., 2016. Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression. *Cell*, 165(6), pp.1416–1427.

Gottesfeld, J.M., Johnson, D.L. & Nyborg, J.K., 1996. Transcriptional activation of RNA polymerase III-dependent genes by the human T-cell leukemia virus type 1 tax protein. *Molecular and cellular biology*, 16(4), pp.1777–1785.

Gouge, J., Guthertz, N., Kramm, K., Dergai, O., Abascal-Palacios, G., Satia, K., Cousin, P., Hernandez, N., Grohmann, D. & Vannini, A., 2017. Molecular mechanisms of Bdp1 in TFIIIB assembly and RNA polymerase III transcription initiation. *Nature communications*, 8(1), p.130.

Grant, T., Rohou, A. & Grigorieff, N., 2018. cisTEM, user-friendly software for single-particle image processing. *eLife,* 7:e35383.

Guffanti, E., Percudani, R., Harismendy, O., Soutourina, J., Werner, M.m Iacovella, M.G., Negri, R. & Dieci, G., 2006. Nucleosome depletion activates poised RNA polymerase III at unconventional transcription sites in Saccharomyces cerevisiae. *The Journal of biological chemistry*, 281(39), pp.29155–29164.

Haeusler, R.A., Pratt-Hyatt, M., Good, P.D., Gipson, T.A. & Engelke, D.R., 2008. Clustering of yeast tRNA genes is mediated by specific association of condensin with tRNA gene transcription complexes. *Genes & development*, 22(16), pp.2204–2214.

Han, Y., Yan, C., Nguyen, T.H.D., Jackobel, A.J., Ivanov, I., Knutson, B.A. & He, Y., 2017. Structural mechanism of ATP-independent transcription initiation by RNA polymerase I. *eLife*, 6:e27414.

Hanszen, K.J., 1971. The optical transfer theory of the electron microscope: fundamental, principles and applications. *Advanced Optical and Electron Microscopy*, 4, pp.1-84.

Harami, G.M., Gyimesi, M. & Kovács, M., 2013. From keys to bulldozers: expanding roles for winged helix domains in nucleic-acid-binding proteins. *Trends in biochemical sciences*, 38(7), pp.364–371.

Harding, S.E., Channell, G. & Phillips-Jones, M.K., 2018. The discovery of hydrogen bonds in DNA and a re-evaluation of the 1948 Creeth two-chain model for its structure. *Biochemical Society transactions*, 46(5), pp.1171–1182.

Harismendy, O., Gendrel, C.-G., Soularue, P., Gidrol, X., Sentenac, A., Werner, M. & Lefebvre, O., 2003. Genome-wide location of yeast RNA polymerase III transcription machinery. *The EMBO journal*, 22(18), pp.4738–4747.

Haurie, V., Durrieu-Gaillard, S., Dumay-Odélot, H., DaSilva, D., Rey, C., Prochazkova, M., Roeder, R.G., Besser, D. & Teichmann, M., 2010. Two isoforms of human RNA polymerase III with specific functions in cell growth and transformation. *Proceedings of the National Academy of Sciences of the United States of America*, 107(9), pp.4176–4181.

Hiraga, S.-I., Botsios, S., Donze, D. & Donaldson, A.D., 2012. TFIIIC localizes budding yeast ETC sites to the nuclear periphery. *Molecular biology of the cell*, 23(14), pp.2741–2754.

Hirata, A. & Murakami, K.S., 2009. Archaeal RNA polymerase. *Current opinion in structural biology*, 19(6), pp.724–731.

Hirata, A., Klein, B.J. & Murakami, K.S., 2008. The X-ray crystal structure of RNA polymerase from Archaea. *Nature*, 451(7180), pp.851–854.

Hoeffler, W.K. & Roeder, R.G., 1985. Enhancement of RNA polymerase III transcription by the E1A gene product of adenovirus. *Cell*, 41(3), pp.955–963.

Hoeffler, W.K., Kovelman, R. & Roeder, R.G., 1988. Activation of transcription factor IIIC by the adenovirus E1A protein. *Cell*, 53(6), pp.907–920.

Hoffmann, N.A., Jakobi, A.J., Moreno-Morcillo, M., Glatt, S., Kosinski, J., Hagen, W.J., Sachse, C. & Mueller, C.W., 2015. Molecular structures of unbound and transcribing RNA polymerase III. *Nature*, 528(7581), pp.231–236.

Hsieh, Y.J., Wang, Z., Kovelman, R. & Roeder, R.G., 1999a. Cloning and characterization of two evolutionarily conserved subunits (TFIIIC102 and TFIIIC63) of human TFIIIC and their involvement in functional interactions with TFIIIB and RNA polymerase III. *Molecular and cellular biology*, 19(7), pp.4944–4952.

Hsieh, Y.J., Kundu, T.K., Wang, Z., Kovelman, R. & Roeder, R.G., 1999b. The TFIIIC90 subunit of TFIIIC interacts with multiple components of the RNA polymerase III machinery and contains a histone-specific acetyltransferase activity. *Molecular and cellular biology*, 19(11), pp.7697–7704.

Huang, Y. & Maraia, R.J., 2001. Comparison of the RNA polymerase III transcription machinery in Schizosaccharomyces pombe, Saccharomyces cerevisiae and human. *Nucleic acids research*, 29(13), pp.2675–2690.

Huang, Y., Hamada, M. & Maraia, R.J., 2000. Isolation and cloning of four subunits of a fission yeast TFIIIC complex that includes an ortholog of the human regulatory protein TFIIICbeta. *The Journal of biological chemistry*, 275(40), pp.31480–31487.

Iben, J.R. & Maraia, R.J., 2012. tRNAomics: tRNA gene copy number variation and codon use provide bioinformatic evidence of a new anticodon:codon wobble pair in a eukaryote. *RNA*, 18(7), pp.1358–1372.

Ilca, S.L., Kotecha, A., Sun, X., Poranen, M.M., Stuart, D.I. & Huiskonen, J.T., 2015. Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. *Nature communications*, 6(1), pp.8843–8.

Imbalzano, A.N., Zaret, K.S. & Kingston, R.E., 1994. Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA. *The Journal of biological chemistry*, 269(11), pp.8280–8286.

Jacques, D.A. & Trewhella, J., 2010. Small-angle scattering for structural biology—Expanding the frontier while avoiding the pitfalls. *Protein science: a publication of the Protein Society*, 19(4), pp.642–657.

Jochem, L., Ramsay, E.P. & Vannini, A., 2017. RNA polymerase I, bending the rules? *The EMBO journal*, 36(18), pp.2664–2666.

Johnson, S.A., Dubbeau, L., Kawalek, M., Dervan, A., Schoenthal, A.H., Dang, C.V. & Johnson, D.L., 2003. Increased expression of TATA-binding protein, the central transcription factor, can contribute to oncogenesis. *Molecular and cellular biology*, 23(9), pp.3043–3051.

Jordán-Pla, A., Gupta, I., de Miguiel-Jiménez, L., Steinmetz, L.M., Chávez, S., Pelechano, V. & Pérez-Ortín, J.E., 2015. Chromatin-dependent regulation of RNA polymerases II and III activity throughout the transcription cycle. *Nucleic acids research*, 43(2), pp.787–802.

Jun, S.-H., Reichlen, M.J., Tajiri, M. & Murakami, K.S., 2011. Archaeal RNA polymerase and transcription regulation. *Critical reviews in biochemistry and molecular biology*, 46(1), pp.27–40.

Juo, Z.S., Chui, T.K., Leiberman, P.M., Baikalov, I., Berk, A.J. & Dickerson, R.E., 1996. How Proteins Recognize the TATA Box. *Journal of molecular biology*, 261(2), pp.239–254.

Kassavetis, G.A., Riggs, D.L., Negri, R., Nguyen, L.H. & Geiduschek, E.P., 1989. Transcription factor IIIB generates extended DNA interactions in RNA polymerase III transcription complexes on tRNA genes. *Molecular and cellular biology*, 9(6), pp.2551–2566.

Kassavetis, G.A., Braun, B.R., Nguyen, L.H. & Geiduschek, E.P., 1990. S-Cerevisiae TFIIIB Is the Transcription Initiation-Factor Proper of RNA Polymerase-III, While TFIIIA and TFIIIC Are Assembly Factors. *Cell*, 60(2), pp.235–245.

Kassavetis, G.A., Bartholomew, B., Blanco, J.A., Johnson, T.E. & Geiduschek, E.P., 1991. Two essential components of the Saccharomyces cerevisiae transcription factor TFIIIB: transcription and DNA-binding properties. *Proceedings of the National Academy of Sciences of the United States of America*, 88(16), pp.7308–7312.

Kassavetis, G.A., Joazeiro, C.A., Pisano, M., Geiduschek, E.P., Colbert, T., Hahn, S. & Blanco, J.A., 1992. The role of the TATA-binding protein in the assembly and function of the multisubunit yeast RNA polymerase III transcription factor, TFIIIB. *Cell*, 71(6), pp.1055–1064.

Kassavetis, G.A., Letts, G.A. & Geiduschek, E.P., 2001. The RNA polymerase III transcription initiation factor TFIIIB participates in two steps of promoter opening. *The EMBO journal*, 20(11), pp.2823–2834.

Kassavetis, G.A. & Geiduschek, E.P., 2006. Transcription factor TFIIIB and transcription by RNA polymerase III. *Biochemical Society transactions*, 34(Pt 6), pp.1082–1087.

Kassavetis, G.A., Prakash, P. & Shim, E., 2010. The C53/C37 subcomplex of RNA polymerase III lies near the active site and participates in promoter opening. *The Journal of biological chemistry*, 285(4), pp.2695–2706.

Kim, Y., Geiger, J.H., Hahn, S. & Sigler, P.B., 1993. Crystal structure of a yeast TBP/TATA-box complex. *Nature*, 365(6446), pp.512–520.

Kim, Y.J., Broerklund, S., Li, Y., Sayre, M.H. & Kornberg, R.D., 1994. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell*, 77(4), pp.599–608.

Knutson, B.A. & Hahn, S., 2011. Yeast Rrn7 and Human TAF1B Are TFIIB-Related RNA Polymerase I General Transcription Factors. *Science*, 333(6049), pp.1637–1640.

Kohl, H. & Reimer, L., 2008. *Transmission Electron Microscopy: Physics of Image Formation*, New York, NY: Springer.

Korkhin, Y., Unligil, U.M., Littlefield, O., Nelson, P.J., Stuart, D.I., Sigler, P.B., Bell, S.D. & Abrescia, N.G.A., 2009. Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure. *PLoS biology*, 7(5), e1000102.

Kundu, T.K., Wang, Z. & Roeder, R.G., 1999. Human TFIIIC relieves chromatin-mediated repression of RNA polymerase III transcription and contains an intrinsic histone acetyltransferase activity. *Molecular and cellular biology*, 19(2), pp.1605–1615.

Kurata, S., Koga, K. & Sakaguchi, B., 1978. Nucleolar size in parallel with ribosomal RNA synthesis at diapause termination in the eggs of Bombyx mori. *Chromosoma*, 68(4), pp.313–317.

L'Etoile, N.D., Noelle, D., Fahnestock, M.L., Shen, Y., Aebersold, R. & Berk, A.J., et al., 1994. Human transcription factor IIIC box B binding subunit. *Proceedings of the National Academy of Sciences of the United States of America*, 91(5), pp.1652–1656.

la Rosa-Trevín, De, J.M., Quintana, A., Del Cano, L., Zaldívar, A., Foche, I., Gutiérrez, J., Gómez-Blanco, J., Burguet-Castell, J., Cuenca-Alba, J., Abrishami, V., Vargas, J., Otón, J., Sharov, G., Vilas, JL., Navas, J., Conesa, P., Kazemi, M., Marabini, R., Sorzano, C.O. & Carazo, J.M., 2016. Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *Journal of structural biology*, 195(1), pp.93–99.

Lagna, G., Kovelman, R., Sukegawa, J. & Roeder, R.G., 1994. Cloning and characterization of an evolutionarily divergent DNA-binding subunit of mammalian TFIIIC. *Molecular and cellular biology*, 14(5), pp.3053–3064.

Landrieux, E., Alic, N., Ducrot, C., Acker, J., Riva, M. & Carles, C., 2006. A subcomplex of RNA polymerase III subunits involved in transcription termination and reinitiation. *The EMBO journal*, 25(1), pp.118–128.

Larminie, C.G., Sutcliffe, J.E., Tosh, K., Winter, A.G., Felton-Edkins, Z.A. & White, R.J., 1999. Activation of RNA polymerase III transcription in cells transformed by simian virus 40. *Molecular and cellular biology*, 19(7), pp.4927–4934.

Lassar, A.B., Martin, P.L. & Roeder, R.G., 1983. Transcription of class III genes: formation of preinitiation complexes. *Science*, 222(4625), pp.740–748.

Lee, J., Moir, R.D. & Willis, I.M., 2015. Differential Phosphorylation of RNA Polymerase III and the Initiation Factor TFIIIB in Saccharomyces cerevisiae. *PloS one*, 10(5), e0127225.

Lefebvre, O., Rueth, J. & Sentenac, A., 1994. A mutation in the largest subunit of yeast TFIIIC affects tRNA and 5 S RNA synthesis. Identification of two classes of suppressors. *The Journal of biological chemistry*, 269(37), pp.23374–23381.

Leney, A.C. & Heck, A.J.R., 2016. Native Mass Spectrometry: What is in the Name? *Journal of The American Society for Mass Spectrometry*, 28(1), pp.5–13.

Liao, Y., Willis, I.M. & Moir, R.D., 2003. The Brf1 and Bdp1 subunits of transcription factor TFIIIB bind to overlapping sites in the tetratricopeptide repeats of Tfc4. *The Journal of biological chemistry*, 278(45), pp.44467–44474.

Liao, Y., Moir, R.D. & Willis, I.M., 2006. Interactions of Brf1 peptides with the tetratricopeptide repeat-containing subunit of TFIIIC inhibit and promote preinitiation complex assembly. *Molecular and cellular biology*, 26(16), pp.5946–5956.

Liebhaber, S.A., Wolf, S. & Schlessinger, D., 1978. Differences in rRNA metabolism of primary and SV40-transformed human fibroblasts. *Cell*, 13(1), pp.121–127.

Liu, H. & Zwart, P.H., 2012. Determining pair distance distribution function from SAXS data using parametric functionals. *Journal of structural biology*, 180(1), pp.226–234.

Louder, R.K., He, Y., López-Blanco, J.R., Fang, J., Chacón, P. & Nogales, E., 2016. Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature*, 531(7596), pp.604-609.

Male, G., von Appen, A., Glatt, S., Taylor, N.M., Cristovao, M., Groetsch, H., Beck, M. & Mueller, C.W., 2015. Architecture of TFIIIC and its role in RNA polymerase III pre-initiation complex assembly. *Nature communications*, 6, p. 7387.

Manaud, N., Arrebola, R., Buffin-Meyer, B., Lefebvre, O., Voss, H., Riva, M., Conesa, C. & Sentenac, A., 1998. A chimeric subunit of yeast transcription factor IIIC forms a subcomplex with tau95. *Molecular and cellular biology*, 18(6), pp.3191–3200.

Marck, C., Lefebvre, O., Carles, C., Riva, M., Chaussivert, N., Ruet, A. & Sentenac, A., 1993. The TFIIIB-assembling subunit of yeast transcription factor TFIIIC has both tetratricopeptide repeats and basic helix-loop-helix motifs. *Proceedings of the National Academy of Sciences of the United States of America*, 90(9), pp.4027–4031.

Marck, C., Kachouri-Lafond, R., Lafontaine, I., Westhof, E., Dujon, B. & Grosjean, H., 2006. The RNA polymerase III-dependent family of genes in hemiascomycetes: comparative RNomics, decoding strategies, transcription and evolutionary implications. *Nucleic acids research*, 34(6), pp.1816–1835.

Marzouki, N., Camier, S., Ruet, A., Moenne, A. & Sentenac, A., 1986. Selective proteolysis defines two DNA binding domains in yeast transcription factor tau. *Nature*, 323(6084), pp.176–178.

Matsutani, S., 2004. Similarities in transcription factor IIIC subunits that bind to the posterior regions of internal promoters for RNA polymerase III. *BMC evolutionary biology*, 4:26.

McMullan, G., McMullan, G., Faruqi, A.R., Clare, D. & Henderson, R., 2014. Comparison of optimal performance at 300 keV of three direct electron detectors for use in low dose electron microscopy. *Ultramicroscopy*,147 pp.156–163.

McMullan, G., Faruqi, A.R. & Henderson, R., 2016. Direct Electron Detectors. *Methods in enzymology*, 579, pp.1–17.

Mertens, H.D.T. & Svergun, D.I., 2010. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *Journal of structural biology*, 172(1), pp.128–141.

Milkereit, P., Schultz, P. & Tschochner, H., 1997. Resolution of RNA polymerase I into dimers and monomers and their function in transcription. *Biological chemistry*, 378(12), pp.1433–1443.

Minton, A.P., 2016. Recent applications of light scattering measurement in the biological and biopharmaceutical sciences. *Analytical biochemistry*, 501, pp.4–22.

Moir, R.D., Sethy-Coraci, I., Puglia, K., Librizzi, M.D. & Willis, I.M., 1997. A tetratricopeptide repeat mutation in yeast transcription factor IIIC131 (TFIIIC131) facilitates recruitment of TFIIB-related factor TFIIIB70. *Molecular and cellular biology*, 17(12), pp.7119–7125.

Moir, R.D., Puglia, K.V. & Willis, I.M., 2000. Interactions between the tetratricopeptide repeats-containing transcription factor TFIIIC131 and its Ligand, TFIIIB70 evidence for a conformational change in the complex. *The Journal of biological chemistry*, 275(34), pp.26591–26598.

Moir, R.D., Puglia, K.V. & Willis, I.M., 2002. A gain-of-function mutation in the second tetratricopeptide repeat of TFIIIC131 relieves autoinhibition of Brf1 binding. *Molecular and cellular biology*, 22(17), pp.6131–6141.

Moir, R.D. & Willis, I.M., 2004. Tetratricopeptide repeats of TFC4 and a limiting step in the assembly of the initiation factor TFIIIB. *Proteins in Eukaryotic Transcription*, 67, pp.93–121.

Mooney, P., Contarato, D., Denes, P., Gubbens, A., Lee, B., Lent, M. & Agard, D., 2011. A High-Speed Electron-Counting Direct Detection Camera for TEM. *Microscopy and Microanalysis*, 17(S2), pp.1004–1005.

Moqtaderi, Z. & Struhl, K., 2004. Genome-wide occupancy profile of the RNA polymerase III machinery in Saccharomyces cerevisiae reveals loci with incomplete transcription complexes. *Molecular and cellular biology*, 24(10), pp.4118–4127.

Moss, T. & Stefanovsky, V.Y., 1995. Promotion and regulation of ribosomal transcription in eukaryotes by RNA polymerase I. *Progress in nucleic acid research and molecular biology*, 50, pp.25–66.

Murakami, K.S. & Darst, S.A., 2003. Bacterial RNA polymerases: the wholo story. *Current opinion in structural biology*, 13(1), pp.31–39.

Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O. & Darst, S.A., 2002. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science*, 296(5571), pp.1285–1290.

Mylona, A., Fernández-Tornero, C., Legrand, P., Haupt, M., Sentenac, A., Acker, J. & Mueller, C.W., 2006. Structure of the tau60/Delta tau91 subcomplex of yeast transcription factor IIIC: insights into preinitiation complex assembly. *Molecular cell*, 24(2), pp.221–232.

Nagarajavel, V., Iben, J.R., Howard, B.H., Maraia, R.J. & Clark, D.J., 2013. Global “bootprinting” reveals the elastic architecture of the yeast TFIIIB-TFIIIC transcription complex in vivo. *Nucleic acids research*, 41(17), pp.8135–8143.

Naidu, S., Friedrich, J.K., Russell, J. & Zomerdijk, J.C., 2011. TAF1B Is a TFIIB-Like Component of the Basal Transcription Machinery for RNA Polymerase I. *Science*, 333(6049), pp.1640–1642.

Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S.H.W., 2018. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife*, 7e:36961.

Neyer, S., Kurz, M., Geiss, C., Hantsche, M., Hodirnau, V.V., Seybert, A., Engel, C., Scheffer, M.P., Cramer, P. & Frangakis, A.S., 2016. Structure of RNA polymerase I transcribing ribosomal DNA genes. *Nature*, 540(7634), pp.607-610.

Nikolov, D.B., Chen, H., Halay, E.D., Usheva, A.A., Hisatake, K., Lee, D.K.,Roeder, R.G. & Burley, S.K., 1995. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature*, 377(6545), pp.119–128.

Nikolov, D.B., Chen, H., Halay, E.D., Hoffmann, A., Roeder, R.G. & Burley, S.K., 1996. Crystal structure of a human TATA box-binding protein/TATA element complex. *Proceedings of the National Academy of Sciences of the United States of America*, 93(10), pp.4862–4867.

Nogales, E. & Scheres, S.H.W., 2015. Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. *Molecular cell*, 58(4), pp.677–689.

Nogales, E., Louder, R.K. & He, Y., 2017. Structural Insights into the Eukaryotic Transcription Initiation Machinery. *Annual review of biophysics*, 46, pp.59–83.

Noma, K.-I., Cam, H.P., Maraia, R.J. & Grewal, S.I., 2006. A role for TFIIIC transcription factor complex in genome organization. *Cell*, 125(5), pp.859–872.

Onodera, Y., Haag, J.R., Ream, T., Costa Nunes, P., Pontes, O. & Pikaard, C.S., 2005. Plant Nuclear RNA Polymerase IV Mediates siRNA and DNA Methylation-Dependent Heterochromatin Formation. *Cell*, 120(5), pp.613–622.

Orioli, A., Pascali, C., Quartararo, J., Diebel, K.W., Praz, V., Romascano, D., Percudani, R., van Dyk, L.F., Hernandez, N., Teichmann, M. & Dieci, G., 2011. Widespread occurrence of non-canonical transcription termination by human RNA polymerase III. *Nucleic Acids Research,* 39 (13), pp.5499-512.

Orlova, E.V. & Saibil, H.R., 2011. Structural analysis of macromolecular assemblies by electron microscopy. *Chemical reviews*, 111(12), pp.7710–7748.

Orphanides, G., Lagrange, T. & Reinberg, D., 1996. The general transcription factors of RNA polymerase II. *Genes & development*, 10(21), pp.2657–2683.

Patikoglou, G.A., Kim, J.L., Sun, L., Yang, S.H., Kodadek, T. & Burley, S.K., 1999. TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes & development*, 13(24), pp.3217–3230.

Pilsl, M. et al., 2016. Structure of the initiation-competent RNA polymerase I and its implication for transcription. *Nature communications*, 7, p.12126.

Pleiner, T., Bates, M., Trakhanov, S., Chung-Tien, L., Schliep, J.E., Chug, H., Boehning, M., Stark, H., Urlaub, H. & Goerlich, D., 2015. Nanobodies: site-specific labeling for super-resolution imaging, rapid epitope-mapping and native protein complex isolation. *eLife*, 4e:11349.

Pombo, A., Jackson, D.A., Hollinshead, M., Wang, Z., Roeder, R.G. & Cook, P.R., 1999. Regional specialization in human nuclei: visualization of discrete sites of transcription by RNA polymerase III. *The EMBO journal*, 18(8), pp.2241–2253.

Putnam, C.D., Hammel, M., Hura, G.L. & Tainer, J.A., 2007. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly Reviews of Biophysics*, 40(3), pp.191–285.

Rambo, R.P. & Tainer, J.A., 2011. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. T. Sosnick & S. A. Woodson, eds. *Biopolymers*, 95(8), pp.559–571.

Rambo, R.P. & Tainer, J.A., 2013. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature*, 496(7446), pp.477–481.

Ramsay, E.P. & Vannini, A., 2017. Structural rearrangements of the RNA polymerase III machinery during tRNA transcription initiation. *BBA - Gene Regulatory Mechanisms*.1861(4), pp. 285-294.

Roberts, D.N., Stewart, A.J., Huff, J.T. & Cairns, B.R., 2003. The RNA polymerase III transcriptome revealed by genome-wide localization and activity–occupancy relationships. *Proceedings of the National Academy of Sciences of the United States of America*, 100(25), pp.14695–14700.

Roeder, R.G., 1996. The role of general initiation factors in transcription by RNA polymerase II. *Trends in biochemical sciences*, 21(9), pp.327–335.

Rosenthal, P.B. & Henderson, R., 2003. Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. *Journal of molecular biology*, 333(4), pp.721–745.

Rosier, D.J.D. & Klug, A., 1968. Reconstruction of Three Dimensional Structures from Electron Micrographs. *Nature*, 217(5124), pp.130–134.

Rozenfeld, S. & Thuriaux, P., 2001. Genetic interactions within TFIIIC, the promoter-binding factor of yeast RNA polymerase III. *Molecular Genetics and Genomics*, 265(4), pp.705–710.

Ruet, A., Camier, S., Smagowicz, W., Sentenac, A. & Fromageot, P., 1984. Isolation of a class C transcription factor which forms a stable complex with tRNA genes. *The EMBO journal*, 3(2), pp.343–350.

Sadian, Y., Tafur, L., Kosinski, J., Jakobi, A.J., Wetzel, R., Buczak, K., Hagen, W.J., Beck, M., Schase, C. & Mueller, C.W., 2017. Structural insights into transcription initiation by yeast RNA polymerase I. *The EMBO journal*, 36(18), pp.2698–2709.

Sainsbury, S., Bernecky, C. & Cramer, P., 2015. Structural basis of transcription initiation by RNA polymerase II. *Nature reviews. Molecular cell biology*, 16(3), pp.129–143.

Sakonju, S., Brown, D.D., Engelke, D., Ng, S.Y., Shastry, B.S. & Roeder, R.G., 1981. The binding of a transcription factor to deletion mutants of a 5S ribosomal RNA gene. *Cell*, 23(3), pp.665–669.

Scheres, S.H.W., Gao, H., Valle, M., Herman, G.T., Eggermont, P.P., Frank, J. & Carazo, J.M., 2007. Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nature Methods*, 4(1), pp.27–29.

Scheres, S.H.W., 2016. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Methods in enzymology*, 579, pp.125–157.

Schramm, L. & Hernandez, N., 2002. Recruitment of RNA polymerase III to its target promoters. *Genes & development*, 16(20), pp.2593–2620.

Schultz, P., Marzouki, N., Marck, C., Ruet, A., Oudet, P. & Sentenac, A., 1989. The two DNA-binding domains of yeast transcription factor tau as observed by scanning transmission electron microscopy. *The EMBO journal*, 8(12), pp.3815–3824.

Schwartz, L.B., Sklar, V.E., Jaehning, J.A., Weinmann, R. & Roeder, R.G., 1974. Isolation and partial characterization of the multiple forms of deoxyribonucleic acid-dependent ribonucleic acid polymerase in the mouse myeloma, MOPC 315. *The Journal of biological chemistry*, 249(18), pp.5889–5897.

Scott, M.R., Westphal, K.H. & Rigby, P.W., 1983. Activation of mouse genes in transformed cells. *Cell*, 34(2), pp.557–567.

Shen, Y., Igo, M., Yalamanchili, P., Berk, A.J. & Dasgupta, A., 1996. DNA binding domain and subunit interactions of transcription factor IIIC revealed by dissection with poliovirus 3C protease. *Molecular and cellular biology*, 16(8), pp.4163–4171.

Sigworth, F.J., 1998. A Maximum-Likelihood Approach to Single-Particle Image Refinement. *Journal of structural biology*, 122(3), pp.328–339.

Sigworth, F.J., Doerschuk, P.C., Carazo, J.M. & Scheres, S.H.W., 2010. An introduction to maximum-likelihood methods in cryo-EM. *Methods in enzymology*, 482, pp.263–294.

Sigworth, F.J., 2016. Principles of cryo-EM single-particle image processing. *Microscopy*, 65(1), pp.57–67.

Sikorski, R.S., Michaud, W.A., Wootton, J.C., Boguski, M.S., Connelly, C. & Hieter, P., 1991. TPR Proteins as Essential Components of the Yeast Cell Cycle. *Cold Spring Harbor symposia on quantitative biology*, 56(0), pp.663–673.

Sinn, E., Wang, Z., Kovelman, R. & Roeder, R.G., 1995. Cloning and characterization of a TFIIIC2 subunit (TFIIIC beta) whose presence correlates with activation of RNA polymerase III-mediated transcription by adenovirus E1A expression and serum factors. *Genes & development*, 9(6), pp.675–685.

Smilgies, D.M. & Folta-Stogniew, 2015. Molecular weight–gyration radius relation of globular proteins: a comparison of light scattering, small-angle X-ray scattering and structure-based data. *Journal of applied crystallography*, 48(5), pp.1604–1606.

Song, B., Lenhart, J., Flegler, V.J., Makbul, C., Rasmussen, T. & Boettcher, B. 2019. Capabilities of the Falcon III detector for single-particle structure determination. *Ultramicroscopy*, 203, pp.145–154.

Spence, J.C.H., 2013. *High-Resolution Electron Microscopy*, 4rd edn, Oxford University Press, New York.

Stillman, D.J. & Geiduschek, E.P., 1984. Differential binding of a S. cerevisiae RNA polymerase III transcription factor to two promoter segments of a tRNA gene. *The EMBO journal*, 3(4), pp.847–853.

Su, M., Han, D., Boyd-Kirkup, J., Yu, X. & Han, J.J., 2014. Evolution of Alu elements toward enhancers. *Cell reports*, 7(2), pp.376–385.

Svergun, D.I., 1999. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical journal*, 76(6), pp.2879–2886.

Tafur, L., Sadian, Y., Hoffmann, N.A., Jakobi, A.J., Wetzel, R., Hagen, W.J.H., Sachse, C. & Mueller, C.W., 2016. Molecular Structures of Transcribing RNA Polymerase I. *Molecular cell*, 64(6), pp.1135–1143.

Taylor, N.M.I., Baudin, F., von Scheven, G. & Mueller, C.W., 2013a. RNA polymerase III-specific general transcription factor IIIC contains a heterodimer resembling TFIIF Rap30/Rap74. *Nucleic acids research*, 41(19), pp.9183–9196.

Taylor, N.M.I., Glatt, S., Hennrich, M.L., von Scheven, G., Groetsch, H., Fernández-Tornero, C., Rybin, V., Gavian, A.C., Kolb, P. & Mueller, C.W., 2013b. Structural and functional characterization of a phosphatase domain within yeast general transcription factor IIIC. *The Journal of biological chemistry*, 288(21), pp.15110–15120.

Thiffault, I., Wolg, N.I., Forget, D., Guerrero, K., Tran, L.T., Choquet, K., Lavallée-Adam, M., Poitras, C., Brais, B., Yoon, G., Sztriha, L., Webster, R.I., Timmann, D., van de Warrenburg, B.P., Seeger, J., Zimmermann, A., Máté, A., Goizet, C., Fung, E., van der Knaap, M.S., Fribourg, S., Vanderver, A., Simons, C., Taft, R.J., Yates, J.R., Coulombe, B. & Bernhard, G., 2015. Recessive mutations in POLR1C cause a leukodystrophy by impairing biogenesis of RNA polymerase III. *Nature communications*, 6, p.7623.

Turowski, T.W. & Tollervey, D., 2016. Transcription by RNA polymerase III: insights into mechanism and regulation. *Biochemical Society transactions*, 44(5), pp.1367–1375.

Turowski, T.W., Leśniewska, E., Delan-Forino, C., Sayou, C., Boguta, M. & Tollervey, D., 2016. Global analysis of transcriptionally engaged yeast RNA polymerase III reveals extended tRNA transcripts. *Genome research*, 26(7), pp.933–944.

Valenzuela, L., Dhillon, N. & Kamakaka, R.T., 2009. Transcription independent insulation at TFIIIC-dependent insulators. *Genetics*, 183(1), pp.131–148.

van Heel, M. & Frank, J., 1981. Use of multivariates statistics in analysing the images of biological macromolecules. *Ultramicroscopy*, 6(1), pp.187–194.

VanAernum, Z.L., Gilbert, J.D., Belov, M.E., Makarov, A.A., Horning, S.R. & Wysocki, V.H., 2019. Surface-Induced Dissociation of Noncovalent Protein Complexes in an Extended Mass Range Orbitrap Mass Spectrometer. *Analytical Chemistry*, 91(5), pp.3611–3618.

Vannini, A., Ringel, R., Kusser, A.G., Berninghausen, O., Kassevetis, G.A. & Cramer, P., 2010. Molecular basis of RNA polymerase III transcription repression by Maf1. *Cell*, 143(1), pp.59–70.

Vannini, A. & Cramer, P., 2012. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Molecular cell*, 45(4), pp.439–446.

Wang, H.D., Yuh, C.H., Dang, C.V. & Johnson, D.L., 1995. The hepatitis B virus X protein increases the cellular level of TATA-binding protein, which mediates transactivation of RNA polymerase III genes. *Molecular and cellular biology*, 15(12), pp.6720–6728.

Wang, H.D., Trivedi, A. & Johnson, D.L., 1997. Hepatitis B virus X protein induces RNA polymerase III-dependent gene transcription and increases cellular TATA-binding protein by activating the Ras signaling pathway. *Molecular and cellular biology*, 17(12), pp.6838–6846.

Waston, J.D. & Crick, F.H., 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356), pp.737–738.

Weissmann, F., Petzold, G., VanderLinden, R., Huis In't Veld, P.J., Brown, N.G., Lampert, F., Westermann, S., Stark, H., Schulman, B.A. & Peters, J.M., 2016. biGBac enables rapid gene assembly for the expression of large multisubunit protein complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 113(19), pp.E2564–9.

Werner, F. & Grohmann, D., 2011. Evolution of multisubunit RNA polymerases in the three domains of life. *Nature reviews. Microbiology*, 9(2), pp.85–98.

West, A.G., Gaszner, M. & Felsenfeld, G., 2002. Insulators: many functions, many mechanisms. *Genes & development*, 16(3), pp.271–288.

White, R.J., Stott, D. & Rigby, P.W., 1990. Regulation of RNA polymerase III transcription in response to Simian virus 40 transformation. *The EMBO journal*, 9(11), pp.3713–3721.

White, R.J., 2004. RNA polymerase III transcription and cancer. *Oncogene*, 23(18), pp.3208–3216.

Wierzbicki, A.T., Haag, J.R. & Pikaard, C.S., 2008. Noncoding Transcription by RNA Polymerase Pol IVb/Pol V Mediates Transcriptional Silencing of Overlapping and Adjacent Genes. *Cell*, 135(4), pp.635–648.

Winter, A.G., Sourvinos, G., Allison, S.J., Tosh, K., Scott, P.H., Spandidos, D.A. & White, R.J., 2000. RNA polymerase III transcription factor TFIIIC2 is overexpressed in ovarian tumors. *Proceedings of the National Academy of Sciences of the United States of America*, 97(23), pp.12619–12624.

Wolffe, A.P. & Brown, D.D., 1988. Developmental regulation of two 5S ribosomal RNA genes. *Science*, 241(4873), pp.1626–1632.

Wollman, A.J.M., Nudd, R., Hedlund, E.G. & Leake, M.C., 2015. From Animaculum to single molecules: 300 years of the light microscope. *Open Biology*, 5(4):150019.

Wu, S., Armache, J.-P. & Cheng, Y., 2016. Single-particle cryo-EM data acquisition by using direct electron detection camera. *Microscopy*, 65(1), pp.35–41.

Ying, C., Gregg, D.W. & Gorski, J., 1996. Estrogen-induced changes in rRNA accumulation and RNA polymerase I activity in the rat pituitary: correlation with pituitary tumor susceptibility. *Molecular and cellular endocrinology*, 118(1-2), pp.207–213.

Yoshinaga, S., Dean, N., Han, M. & Berk, A.J., 1986. Adenovirus stimulation of transcription by RNA polymerase III: evidence for an E1A-dependent increase in transcription factor IIIC concentration. *The EMBO journal*, 5(2), pp.343–354.

Zawel, L., Kumar, K.P. & Reinberg, D., 1995. Recycling of the general transcription factors during RNA polymerase II transcription. *Genes & development*, 9(12), pp.1479–1490.

Zeytuni, N. & Zarivach, R., 2012. Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. *Structure*, 20(3), pp.397–405.

Zhai, W. & Comai, L., 1999. A kinase activity associated with simian virus 40 large T antigen phosphorylates upstream binding factor (UBF) and promotes formation of a stable initiation complex between UBF and SL1. *Molecular and cellular biology*, 19(4), pp.2791–2802.

Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K. & Darst, S.A., 1999. Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. *Cell*, 98(6), pp.811–824.

Zhang, Z. & Dietrich, F.S., 2005. Identification and characterization of upstream open reading frames (uORF) in the 5' untranslated regions (UTR) of genes in Saccharomyces cerevisiae. *Current genetics*, 48(2), pp.77–87.

Zhao, X. & Herr, W., 2002. A Regulated Two-Step Mechanism of TBP Binding to DNA: A Solvent-Exposed Surface of TBP Inhibits TATA Box Recognition. *Cell*, 108(5), pp.615–627.

Zhao, X., Li, G. & Liang, S., 2013. Several Affinity Tags Commonly Used in Chromatographic Purification. *Journal of Analytical Methods in Chemistry*, 2013(1), pp.1–8.

Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y. & Agard, D.A., 2017. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods*, 14(4), pp.331–332.

Zhou, M., Dagan, S. & Wysocki, V.H., 2012. Protein Subunits Released by Surface Collisions of Noncovalent Complexes: Nativelike Compact Structures Revealed by Ion Mobility Mass Spectrometry. *Angewandte Chemie*, 51(18), pp.4336–4339.

Ziegler, A., Kisielowski, C. & Ritchie, R.O., 2002. Imaging of the crystal structure of silicon nitride at 0.8 Ångström resolution. *Acta Materialia*, 50(3), pp.565–574.

Zivanov, J., Nakane, T., Forsberg, B., Kimanius, D., Hagen, W.J.H., Lindahl, E. & Scheres, S.H.W., 2018. RELION-3: new tools for automated high-resolution cryo-EM structure determination. *eLife, 7:e42166.*