Dissertation

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

M. Biochem. Gary Male

Born in: Jersey, Channel Islands

Oral examination: 30th September 2014

Insights into the Architecture of Transcription Factor IIIC

Referees: Dr. Carsten Sachse

Prof. Dr. Bernd Bukau

Acknowledgements

I firstly want to thank my supervisor, Christoph Müller, for giving me the opportunity to work towards my PhD thesis at EMBL on a challenging project, and for giving me the freedom to pursue my own ideas and project directions.

I owe a lot to the mentoring and support of Sebastian Glatt, a brilliant Staff Scientist in our lab, who has supported and challenged me throughout my PhD.

I thank Michele Cristovao in our lab for support and discussions as she studies my *in vitro* findings *in vivo*. I am grateful to Martin Beck and Alexander von Appen at EMBL for an excellent collaboration with the cross-linking studies. I thank Vladimir Rybin for the ITC experiments and for valuable discussions. I acknowledge Nicholas Taylor for his efforts in solving the original τ131 structure. I thank Imre Berger and his group at EMBL Grenoble for their work and support in the MultiBac project.

I thank Florence Baudin for help with the EMSA experiments and for useful discussions.

I am grateful to Angelika Scholz of the EMBL crystallisation platform for her patience when I was concentrating protein in preparation for crystallisation.

I acknowledge all the core facilities at EMBL, including the Media Kitchen for all their help and support.

I thank all the Müller lab members, past and present, for all their help, support and humour over the past 4 years and for creating a great atmosphere in which to work.

I thank the members of my TAC: Carsten Sachse, Christian Häring and Bernd Bukau for all their advice and support during my PhD. I also thank the EMBL Graduate Office for all their assistance and support.

I thank Nicole, Ana, Dominic, Franka, Dan, Plamena, Paul, Lionel, Kasia, Anna, Erik, Rory, Esme, MACUMBA and the 'Predocs 2010' for their friendship over the past 4 years.

Finally, I thank my family and Sophie for all their support in getting me to this stage.

Summary

Transcription Factor IIIC (TFIIIC) is a six-subunit protein complex that recognises highly conserved promoter elements within genes transcribed by RNA polymerase III. These promoter elements, called 'A box' and 'B box', are recognised by two subcomplexes of TFIIIC, termed 'τA' and 'τB'. The τA subcomplex is formed from three subunits: τ131, τ95 and τ55. The τB subcomplex is also composed of three subunits: τ138, τ91 and τ60. The binding of TFIIIC to these promoter elements leads to the recruitment of another transcription factor, TFIIIB, which subsequently coordinates the recruitment of RNA polymerase III to the transcription start site.

In this thesis, I present my work on the characterisation of a novel interface between the τA and τB subcomplexes, which is located between the subunits τ131 and τ138. Chemical cross-linking coupled with mass spectrometry provides evidence for an important interaction between these two essential proteins, which I have mapped in detail using biochemical, biophysical and structural methods. The structure of a highly conserved tetra-trico peptide repeat (TPR) domain within τ131 is presented and is shown to bind with high affinity to a disordered region of τ138. In addition, I present the first high resolution structure of an extended winged helix domain of τ138 that I have solved by X-ray crystallography. Finally, I present work detailing efforts to produce the TFIIIC complex recombinantly in insect cells. The work presented here will further our understanding of the architecture of TFIIIC and provide insight into TFIIIC structure and function.

Zusammenfassung

Der Transkriptionsfaktor IIIC (TFIIIC) ist ein Proteinkomplex der aus sechs Untereinheiten aufgebaut ist und stark erhaltene Promotorelemente in von RNA Polymerase III transkribierten Genen erkennt. Diese als „A box" und „B box" bezeichneten Promotorelemente werden durch zwei TFIIIC - Subkomplexe erkannt, die „τA" und „τB" genannt werden. Der τA Subkomplex setzt sich aus den Untereinheiten τ131, τ95 und τ55 zusammen. Der τB Subkomplex besteht ebenfalls aus drei Untereinheiten, τ138, τ91 und τ60. Die Assoziation von TFIIIC an die Promotorelemente A Box und B Box durch die τA und τB Subkomplexe leitet die Rekrutierung eines weiteren Transkriptionsfaktors, TFIIIB, ein, welcher darauffolgend die RNA Polymerase III an den Transkriptionsstartpunkt rekrutiert.

In dieser Doktorarbeit stelle ich meine Arbeit über die Charakterisierung einer neuen Interaktionsstelle zwischen dem τA und τB Subkomplex vor, die zwischen den Untereinheiten τ131 und τ138 liegt. Chemische Quervernetzung gekoppelt mit Massenspektroskopie beweisen eine wichtige Interaktion zwischen diesen zwei essentiellen Proteinen welche ich detailliert mit biochemischen, biophysikalischen und strukturbiologischen Methoden charakterisiert habe. Die Struktur einer hochkonservierten „tetratricopeptid repeat" Domäne aus dem τ131 Protein, welche an eine unstrukturierte Region von τ138 bindet, wird in dieser Arbeit vorgestellt. Desweiteren präsentiere ich die erste hochauflösende Struktur einer „extended winged helix" Domäne des Proteins τ138 die ich mit Röntgenkristallographie bestimmt habe. Abschliessend stelle ich meine Bemühungen zur rekombinanten Produktion von TFIIIC in Insektenzellen dar. Die hier gezeigte Arbeit präsentiert neue Erkenntnisse über die Struktur und Funktion von TFIIIC und wird unser Verständnis über die Architektur dieses Transkriptionsfaktors erweitern.

Table of Contents

1. Introduction

Accessing the genetic information stored within a cell requires the coordinated action of many proteins, which often act together as parts of molecular machines [1]. The information is stored within the double-stranded and helical DNA molecule [2]. In eukaryotes, the DNA is associated with proteins called histones and is packaged within the nucleus in the form of chromatin [3]. For this thesis, I have focussed on a protein complex called transcription factor (TF) IIIC. This protein must coordinate with a molecular machine called RNA polymerase III in order to access specific genetic information stored within a eukaryotic cell.

1.1 Transcription initiation and RNA polymerases through evolution

Transcription, the process whereby DNA is 'read' and RNA is subsequently synthesised, relies on a highly conserved molecular machine called RNA polymerase. Bacteria and archaea utilise only one form of RNA polymerase to read their genomes, whereas eukaryotes have divided the task between several [4]. The process of transcription consists of three key stages in all three domains of life: initiation, elongation and termination. These stages require the concerted action of associated factors, working together to ensure the RNA polymerase can progress along the DNA template and synthesise an RNA molecule.

Initiation of transcription requires the formation of a pre-initiation complex (PIC) and the subsequent procession from a closed complex to an open complex [5, 6]. In brief, key promoter elements are recognised and bound by a transcription factor, before DNA is 'melted' and then stabilised in the active site of the RNA polymerase. A process known as abortive initiation then proceeds [7, 8], whereby short RNA transcripts are generated, before a sufficient length is reached and the PIC can escape the promoter and RNA synthesis can begin. The next stage in transcription, elongation, is not discussed here but has been well reviewed elsewhere [9, 10]. Transcription termination will be discussed in a later section.

There is a high degree of similarity between the catalytic properties of RNA polymerase in the three domains of life. This is also true for the mechanisms that lead to transcription initiation. However, differences exist between the architecture of the

RNA polymerase and the associated transcription factors, which will be discussed below.

1.1.1 Bacterial RNA polymerase

The elucidation of a high-resolution structure of the bacterial RNA polymerase gave the first insights into the conserved core of this enzyme family [11]. The 'crab-claw' shape of the polymerase along with the catalytic centre are preserved through evolution, with homologues of the 5 subunits of bacterial RNA polymerase (ββ'αασ) present in all three domains of life [12] (*Figure 1.1*).

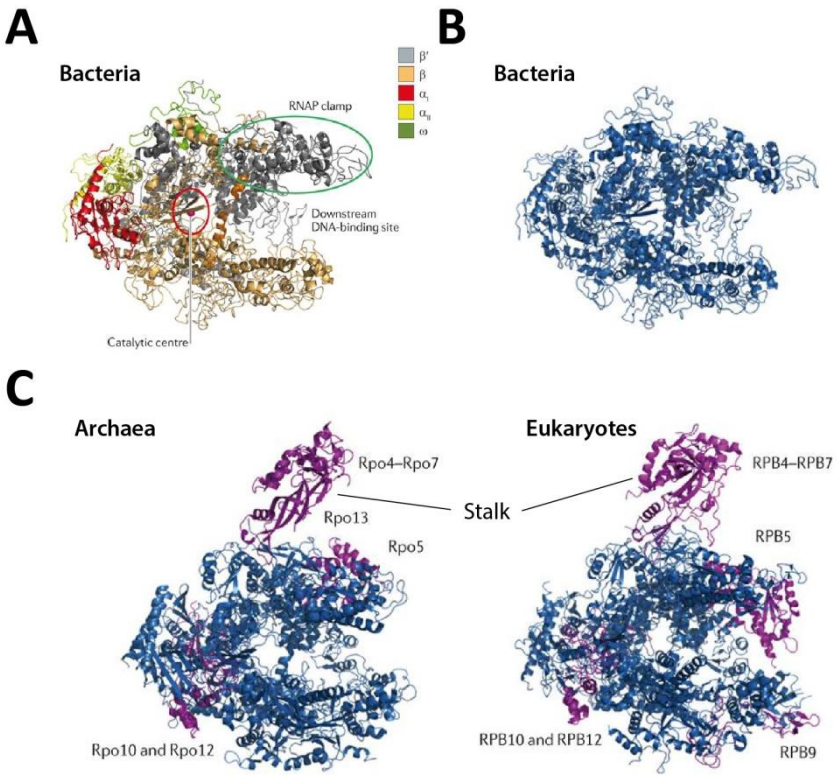

Figure 1.1 **The conserved core of RNA polymerase**. (A) The bacterial RNA polymerase is composed of 5 subunits. (B) The conserved core of bacterial RNA polymerase. (C) The conserved core (blue) is present in archaeal and eukaryotic RNA polymerases. Archaea and eukaryotes contain additional subunits, including a 'stalk' as indicated; see main text and *Table 1.1* for details. Figure modified from [12].

In order for bacterial RNA polymerase to begin transcribing DNA, it must first associate with a transcription factor known as a 'σ-factor' [13, 14]. These bacterial-specific proteins recognise promoter elements known as '-10' and '-35' [6] and enable transcription initiation in what is generally an energy-independent event, although

examples of σ-factors which require energy from ATP hydrolysis are known [15]. The key steps leading to promoter clearance by the RNA polymerase have been well reviewed elsewhere [6].

1.1.2 Archaeal RNA polymerase

As highlighted in *Figure 1.1*, archaeal and eukaryotic RNA polymerases contain additional subunits. A full list and comparison of these subunits and the associated transcription factors between bacteria, archaea and the eukaryote *Saccharomyces cerevisiae* (*S. cerevisiae*) is summarised in *Table 1.1*. These additional subunits that are present in archaeal and eukaryotic RNA polymerase are essential to the organism, but do not seem necessary for RNA synthesis *per se* [12]. The additional 'stalk' subunits, for example, are important for other functions such as binding to transcription factors [16, 17].

Archaeal RNA polymerase transcription can be viewed as a 'mosaic' of bacterial and eukaryotic features [18]. Whilst the structure of the archaeal RNA polymerase [19, 20] reveals a close similarity to eukaryotic polymerases with regards to additional subunits, many features of transcription initiation show more striking similarity to bacteria [6, 21]. The most obvious similarity however between archaea and bacteria is the possession of only one RNA polymerase.

1.1.3 Eukaryotic RNA polymerases

The identification of multiple eukaryotic polymerases was first established in work by Roeder and Chambon [22, 23]. It is now established that eukaryotes have at least three RNA polymerase enzymes, with plants possessing two additional, non-essential RNA polymerases [24, 25]. Each RNA polymerase recognises a unique set of promoters and is associated with a specific set of transcription factors (reviewed in [26-28]). The best characterised of the eukaryotic polymerases is RNA polymerase II, which is principally responsible for synthesising messenger RNA (mRNA). The crystal structure of *S. cerevisiae* RNA polymerase II was solved in 2000 [29]. Since then, several more structural snapshots of this polymerase with its associated transcription factors and at various stages of transcription have been solved [28, 30-33]. Recently, two groups have

studied the entire RNA polymerase II PIC using an integrated approach that combined crystal structures, cryo-EM reconstructions and cross-linking [34, 35].

Table 1.1 **The conserved RNA polymerase subunits and associated transcription factors**

Bacteria	Archaea	Eukaryotes		
RNA polymerase	*RNA polymerase*	*RNA polymerase II*	*RNA polymerase I*	*RNA polymerase III*
RNA polymerase core				
β subunit	Rpo1	RPB1	A190	C160
β subunit	Rpo2	RPB2	A135	C128
α subunit	Rpo3	RPB3	AC40	AC40
α subunit	Rpo11	RPB11	AC19	AC19
		RPB9	A12.2	C11
	Rpo5	RPB5	RPB5	RPB5
ω subunit	Rpo6	RPB6	RPB6	RPB6
	Rpo8	RPB8	RPB8	RPB8
	Rpo10	RPB10	RPB10	RPB10
	Rpo12	RPB12	RPB12	RPB12
RNA polymerase stalk				
	Rpo4	RPB4	A14	C17
	Rpo7	RPB7	A13	C25
General transcription factors and homologous subcomplexes				
		Tfg1 (TFIIFα)	A49 (N-terminus)	C37
		Tfg2 (TFIIFβ)	A34.5	C53
	TFEα	Tfa1 (TFIIEα)[1]		C82[1]
	TFEβ	Tfa2 (TFIIEβ)[2]	A49 (C-terminus)[2]	C34[2]
				C31
	TBP	TBP[3]	TBP[5]	TBP[6]
		TAFs[3,4]		
	TFB	TFIIB	Rrn7[5]	Brf1[6]

[1,2]Evolutionary relationship remains to be definitively established. [3]Part of TFIID (along with TAFs). [4]TBP-associated factors (part of TFIID). [5]TBP, Rrn7, Rrn6 and Rrn11 form the core factor. [6]TBP, Brf1 and Bdp1 form TFIIIB. Table and footnotes modified from [12, 36].

RNA polymerase I contains two additional subunits (A49/34.5) in comparison to RNA polymerase II and is responsible for the transcription of the ribosomal RNA precursor gene [37]. The presence of additional subunits in RNA polymerase I hinted at mechanistic differences with regards to the transcription mechanism. However, the solving of the structure of this subcomplex by Geiger and colleagues [38] and its location in the context of the RNA polymerase I enzyme (as recently solved by two groups [39, 40]) reveals a striking similarity to the general transcription factor TFIIF of

RNA polymerase II [41]. These results suggest that RNA polymerase I has incorporated this transcription factor as part of its core structure. In fact, further homology is observed between these additional subunits and the general transcription factor TFIIE, adding weight to the notion that RNA polymerase I and also RNA polymerase III have permanently recruited general transcription factors of RNA polymerase II into their core structure [4, 36].

The 17 subunit RNA polymerase III is the largest of the RNA polymerase enzymes and is responsible for transcribing genes encoding for small, untranslated RNAs such as transfer RNA (tRNA), 5S ribosomal RNA (5S rRNA) and U6 small nuclear RNA (U6 snRNA) (reviewed in [42, 43]). Although a crystal structure for RNA polymerase III remains currently unavailable, extensive protein-protein interaction networks have been established for the enzyme [44], as well as cross-linking studies [45] and a cryo-EM reconstruction at 9.9 Å resolution [46]. As with RNA polymerase I, the additional subunits of RNA polymerase III appear to show significant homology to the general transcription factors of RNA polymerase II. The subcomplex C37/53 [47, 48] appears to show homology to TFIIF (and A49/34.5) [36] and can be positioned into the cryo-EM density [46] in a similar location as for RNA polymerase II and I [39]. Furthermore, a trimeric subcomplex, C82/34/31 [49], shows extensive similarities to general transcription factor TFIIE of RNA polymerase II. The structure of the human (*Homo sapiens* or *H. sapiens*) C82 orthologue (hRPC62) contained a four-tandem extended winged helix domain which is structurally homologous to the largest TFIIE subunit, TFIIEα (and the archaeal homologue TFE) [50, 51]. The C34 subunit also appears to show homology to the TFIIEβ subunit of TFIIE and to the C-terminus of A49 from RNA polymerase I (*see Table 1.1*) [36].

In summary, eukaryotes have divided the task of transcription between at least three RNA polymerases, but appear to have retained some core homologies within each enzyme. These homologies extend not only to the catalytic core (which is also shared with bacteria and archaea), but also to the additional subunits that exist as either permanently or non-permanently recruited factors. This introduction will now focus on RNA polymerase III and its associated promoter elements and transcription factors.

1.2 The promoters of RNA polymerase III

With the aid of associated transcription factors RNA polymerase III (hereafter referred to as Pol III) recognises three types of promoters (reviewed in [42, 43]) (*Figure 1.2*).

1.2.1 Type 1 promoter

The sole example from this class is located in the gene for 5S rRNA, which was first characterised from *Xenopus laevis* [52, 53]. An unusual feature of this gene is the presence of intragenic promoter elements. The promoter for the 5S rRNA gene contains an A box, an intermediate element (IE) and a C box, which together form the internal control region (ICR) [54]. The ICR is recognised by the transcription factor TFIIIA. This binding then enables the recruitment of TFIIIC, TFIIIB and ultimately Pol III.

1.2.2 Type 2 promoter

The tRNA genes form the majority of this class, and were initially characterised around the same time as the type 1 promoter [55, 56]. As with the type 1 promoters, type 2 promoters also contain internal promoter elements. These elements, called A box and B box, are highly conserved [57] and correspond to the D and T loops of the subsequently transcribed tRNA molecule. It should be noted that, whilst the A box of type 2 promoters and type 1 promoters are similar and interchangeable [58], they in fact recruit different transcription factors: TFIIIA in the case of type 1 promoters and TFIIIC in the case of type 2 promoters. A more detailed discussion of the A and B boxes of tRNA genes will follow (*Section 1.3*).

1.2.3 Type 3 promoter

The type 3 promoters are different with respect to type 1 and type 2 promoters, in that their promoter elements are gene external. They were first identified in the U6 small nuclear RNA (snRNA) genes of mammals [59-61] and the 7SK gene of humans [62]. The well characterised human U6 gene contains three important elements, a distal sequence element (DSE), a proximal sequence element (PSE) and a TATA box. Finally, there are also hybrid forms of Pol III promoters which exist somewhere between type 2 and type 3 promoters. The *S. cerevisiae* U6 snRNA gene for example contains a gene-external TATA box but a gene internal A box (as well as a gene external

B box). The picture becomes even more complicated when studying *Schizosaccharomyces pombe* (*S. pombe*) promoters, which reveals that nearly all type 1 and type 2 promoters contain an additional TATA box which is necessary for transcription [63].

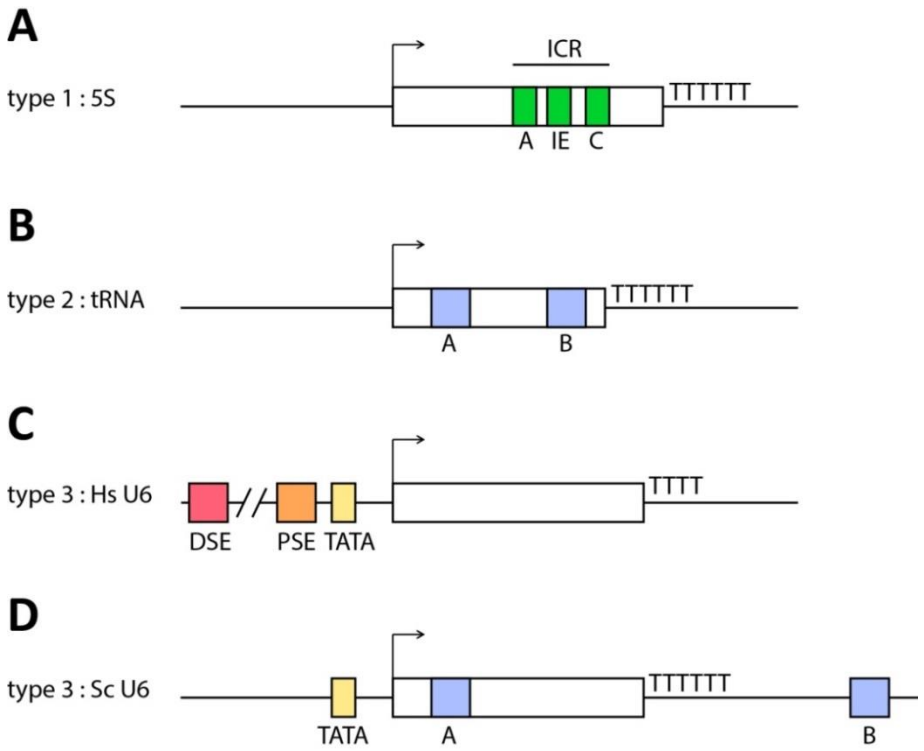

Figure 1.2 **RNA polymerase III promoters.** (A) Type 1 (5S rRNA) promoter. (B) Type 2 (tRNA) promoter. (C) *H. sapiens* type 3 (U6 snRNA) promoter. (D) *S. cerevisiae* type 3 (U6 snRNA) promoter. Details can be found in the main text. Arrows indicate transcription start site. Hs = *H. sapiens*; Sc = *S. cerevisiae*; ICR = internal control region; IE = intermediate element; A, B and C = respective boxes; DSE = distal sequence element; PSE = proximal sequence element; TATA = TATA box; TTTT(TT) = termination signal. Figure modified from [42].

In summary, three classes of promoter are transcribed by Pol III. Each promoter contains different elements which all ultimately lead to the same outcome: the recruitment of the transcription factor TFIIIB and subsequently Pol III. In type 1 and type 2 promoters, TFIIIB must be assembled though interactions with another transcription factor (TFIIIA in the case of type 1 and TFIIIC in the case of type 2). In type 3 promoters, TFIIIB can directly bind the TATA box (and make additional interactions with proteins that bind the PSE). These various scenarios have also been summarised elsewhere [42, 43].

1.3 TFIIIC mediated transcription of tRNA genes

The small genes that encode tRNAs require only two transcription factors in order to recruit Pol III. Initially recognised by chromatography of a HeLa cell extract [64], these two factors are called transcription factor IIIC (TFIIIC) and transcription factor IIIB (TFIIIB). This following section will summarise the currently available information regarding TFIIIC and its recognition of A box and B box containing promoters (type 2).

1.3.1 TFIIIC binds the A and B boxes

The binding of TFIIIC to tRNA genes is mediated through the highly conserved A box and B box promoter elements [65, 66]. Consensus sequences for these promoters have been defined [57] (*Figure 1.3*). Mutations of different key bases within the B box element have been described and have been shown to abolish TFIIIC binding [67]. The A box is located at +12 bases from the transcription start site, a position which appears highly conserved. The B box is separated from the A box by a variety of lengths, which is dependent on the presence of introns, or differing variable arm lengths within different tRNA genes [57]. TFIIIC can accommodate these variations and is assumed therefore to harbour a relatively high degree of flexibility [68, 69]. How this flexibility is achieved is currently poorly understood.

The binding of TFIIIC to the B box is a high-affinity interaction (pM-nM) that appears to be more critical for TFIIIC-tDNA binding than the A box interaction [70, 71]. The A box interaction is weaker (μM) and is inhibited by single-stranded DNA [72]. The mechanisms and identities of the crucial domains for TFIIIC-tDNA binding have been long sought after.

1.3.2 The molecular architecture of TFIIIC – τA and τB

In *S. cerevisiae*, TFIIIC consists of six subunits that form two individual subcomplexes called τA and τB, which both harbour DNA binding activity [65, 73] (and reviewed in [74]) (*Figure 1.3*). Proteolysis studies [75] and scanning transmission electron microscopy [68] first indicated the existence of these two subcomplexes, which resemble a 'dumb-bell' shaped molecule of two equal sized components that appear to be connected by a 'flexible linker'. Recent *in vivo* genome-wide footprinting analyses have corroborated that view, again suggesting that the 'two-lobed' τA-τB structure sits

over the entire tRNA gene [76]. The τA subcomplex has been shown to bind the A box, whilst the τB subcomplex binds the B box [65].

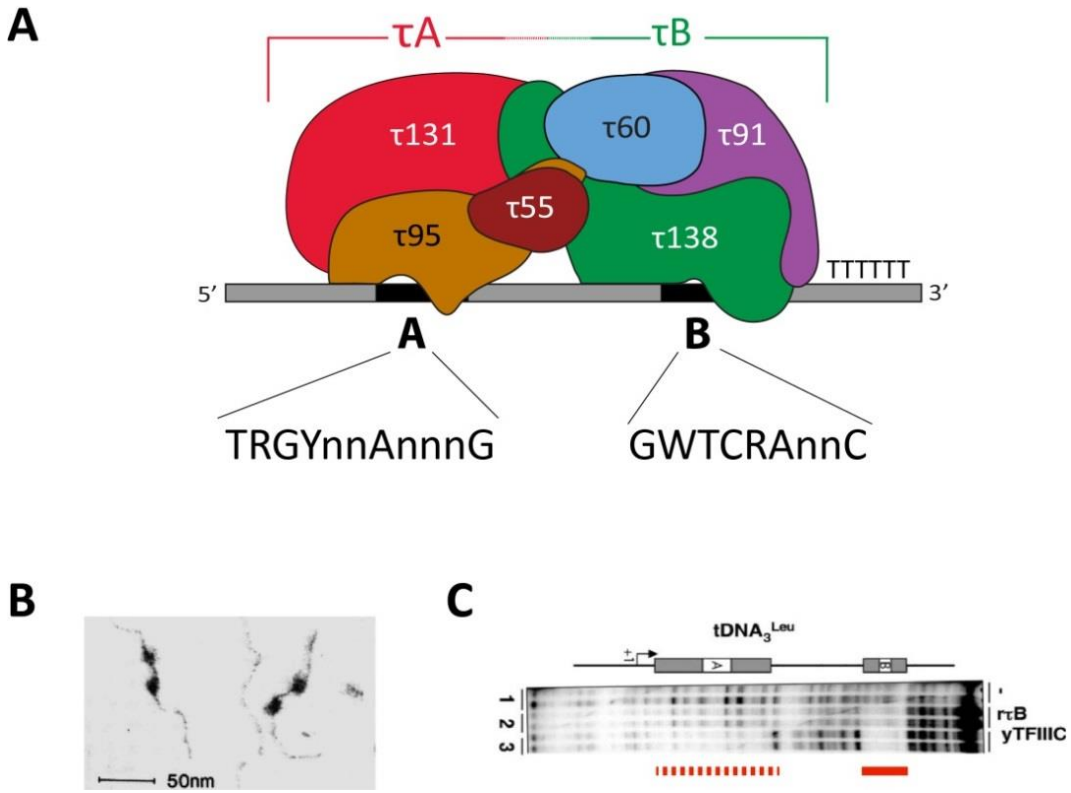

Figure 1.3 **The A and B box promoter elements are recognised by TFIIIC**. (A) Schematic diagram that shows the 6 subunits of TFIIIC (forming the τA and τB subcomplexes) binding to the A and B box promoters of a tRNA gene. The termination signal (TTTTTT) is also highlighted. The consensus sequences for these promoters are indicated as defined by [57]. R = purine; Y = pyrimidine; n = nucleotide; W = weak binding (A or T). (B) Scanning transmission electron microscopy image of TFIIIC bound to tDNA. Two densities are visible which correspond to the τA and TB subcomplexes. Figure modified from [68]. (C) Footprints of τB and TFIIIC on tDNA. The thick red line indicates the τB footprint, whilst the dashed red line indicates the more diffuse τA footprint. Figure modified from [65].

The τA subcomplex consists of the subunits τ131 (Tfc4), τ95 (Tfc1) and τ55 (Tfc7), whilst the τB subcomplex contains τ60 (Tfc8), τ91 (Tfc6) and τ138 (Tfc3) (*Figure 1.3*). All of these subunits are essential for cell viability [77-82]. Their names correspond to their approximate molecular weight as assessed by SDS-PAGE, but the names in parentheses are also found in the literature. A full list of the subunits with their orthologues in *S. pombe* and *H. sapiens* is displayed in *Table 1.2*. Early photo cross-linking studies have given indications of the positions of these subunits on the tRNA gene [83, 84].

Table 1.2 **TFIIIC subunits and orthologues**

S. cerevisiae	*S. pombe*	*H. sapiens*
τ131 (Tfc4)	Sfc4	TFIIIC102 (γ)
τ95 (Tfc1)	Sfc1	TFIIIC63 (ε)
τ55 (Tfc7)	Sfc7	TFIIIC35
τ60 (Tfc8)	Sfc8	TFIIIC90 (δ)
τ91 (Tfc6)	Sfc6	TFIIIC110 (β)
τ138 (Tfc3)	Sfc3	TFIIIC220 (α)

Modified from [42]. For *H. sapiens*, the names in parentheses correspond to pseudonyms in the literature.

1.3.3 τ131

The largest τA subunit, τ131, is predicted to contain many tetra-trico peptide repeats (TPRs) [77]. These repeats are protein-protein interaction motifs that are composed of 34 amino acids that form two antiparallel α-helices [85]. They are often found in large protein complexes acting as scaffolds for other subunits [86]. τ131 cross-links upstream of the 5' border of the tRNA gene and is crucial for the assembly of TFIIIB (reviewed in [87]). Within human TFIIIC, τ131 has been shown to interact with the τ95 subunit [88].

Unpublished work from Nicholas Taylor, a former PhD student in our lab, led to the structure determination of a fragment of τ131 containing several N-terminal TPRs. Details will be discussed in the *Results* section.

1.3.4 τ95 and τ55

The τ95 and τ55 subunits were shown to be located in the proximity of the A box and τ95 was subsequently postulated to be the genuine A box-binding subunit [83]. The *S. pombe* τ95 orthologue has subsequently been shown to bind single-stranded and double-stranded DNA, though with low affinity and no sequence specificity [89], thus consistent with previous findings of A box binding [72]. The crystal structure of the DNA-binding domain revealed a winged helix domain with a neighbouring 'winged helix interacting' domain [89]. τ95 may also indirectly affect B box binding, as a point mutant (E447K) was shown to disrupt B box binding [90]. The N-terminus of τ95 forms a hetero-dimer with the C-terminus of the τ55 subunit [89]. The structure revealed a

strikingly similar fold to the two subunits of the RNA polymerase II transcription factor, TFIIF [41, 89]. It is intriguing that Pol III would then possess another copy of this TFIIF-like fold in addition to C37/53. The τ95-τ55 subcomplex has also been proposed to exist independently from the TFIIIC complex [79].

In addition to interacting with τ95, τ55 also contains a non-essential and non-conserved histidine phosphatase domain in *S. cerevisiae* which may link the activity of TFIIIC to metabolic pathways [91].

1.3.5 τ60 and τ91

τB subunits, τ60 and τ91, both contain WD40 repeats that form canonical seven-bladed β-propellers, which pack perpendicularly to form a dimer [92]. τ60 also contains an α/β fold which has been shown to bind the TFIIIB subunit, TBP [80, 92]. τ60 is the only subunit of TFIIIC that could not be detected in photo cross-linking experiments on tRNA genes [84].

The τ91 subunit could be cross-linked to tRNA genes and is located at the 3' end of the gene near the termination sequence [83]. τ91 has been shown to bind single- and double-stranded DNA non-specifically through the disordered region at its N-terminus, which contains a so-called 'AT hook' [81, 92]. The τ60-τ91 scaffold is hypothesised to form a platform for the largest TFIIIC subunit, τ138, which is predicted to be the B-box binding subunit [92].

1.3.6 τ138

τ138 remains the least well understood and structurally characterised of the TFIIIC subunits. Cross-linking of τ138 near the B box [84] provided the first evidence that this subunit binds to this important promoter region. In addition, a temperature-sensitive point mutant in τ138 (G349E) was identified that reduces the binding of TFIIIC to tRNA genes [93]. This residue is completely conserved from yeast to human [94]. Furthermore, τ138 is predicted to contain several winged helix domains and high mobility group (HMG) boxes, which are both characterised as DNA binding domains in other proteins [95].

The precise mechanisms of how τ138 binds the B box are still unknown. τ91 may play an auxiliary role, as a point mutant (E447K) was shown to suppress the effects of the G349E point mutation [81].

1.3.7 Conservation of TFIIIC subunits

The evolutionary relationship between the six subunits of *S. cerevisiae* and *H. sapiens* TFIIIC is not immediately obvious in all cases. Human TFIIIC can be separated into two fractions by ion exchange chromatography [96, 97]. The first fraction is poorly characterised, whilst the second fraction, TFIIIC2, contains the orthologues of all six yeast TFIIIC subunits. Sequence conservation is highest with the two τA subunits τ131 and τ95. The remaining τA subunit, τ55, was the last human TFIIIC subunit to be characterised [98] and was revealed not to contain the histidine phosphatase domain described earlier.

Conservation of the τB subunits is much weaker and it is largely through the efforts of cloning and characterising *S. pombe* orthologues that a relationship between *S. cerevisiae* and *H. sapiens* τB subunits could be established [94, 99]. The human τ60 and τ91 orthologues, TFIIIC90 and TFIIIC110, have been shown to possess histone acetlytransferase activity [100, 101], implicating a role in chromatin modification. These adaptations, which are not apparent in yeast orthologues, may explain the significant divergence in sequence. The τ138 subunit shows the least conservation amongst the TFIIIC subunits. Proteolysis experiments revealed that the human TFIIIC subunit, TFIIIC220 (α), contains an 83 kDa DNA binding domain that recognises the B box [102] indicating that this is the τ138 orthologue in human, despite the sequences being poorly conserved. Later bioinformatics analysis concluded there is similarity within the N-terminal third of τ138 orthologues [103].

In summary, τ131 shows the highest overall conservation over the entirety of its sequence, possibly due to its essential role in TFIIIB recruitment (*Section 1.4.1*). τ138, in contrast, shows very weak conservation despite its hypothesised role in binding the critical B box promoter.

1.3.8 The identity of the flexible linker in TFIIIC

An important open question concerns which subunits are involved in forming the 'flexible linker', which allows TFIIIC to recognise differently spanned promoter regions using its flexibility between the presumably rigid τA and τB subcomplexes (*Section 1.3.2*). Both τ60 and τ95 have been postulated as potential candidates [80, 90]. τ60 has been suggested owing to its role in binding TBP and the fact that it does not cross-link to DNA. The human orthologue has also been shown to bind human orthologues of τ131 and τ95 [101]. τ95 is proposed as it has been shown to bind τ91 and τ138 by co-immunoprecipitation experiments [90]. Finally, whilst screening for suppressors of the temperature-sensitive G349E mutation of τ138, Rozenfeld and Thuriaux isolated several mutations within the TPR-containing region of τ131 [104]. This genetic interaction between τ131 and τ138 has not been directly demonstrated or investigated biochemically.

The precise identity and nature of the linker remains an open question and no direct interaction between τA and τB has been demonstrated yet using purified proteins to the authors' knowledge.

1.4 The recruitment of TFIIIB and Pol III to tRNA genes

With TFIIIC bound to the A box and B box promoters of the tRNA gene, TFIIIB can then be assembled. TFIIIB is composed of three proteins: Brf1, Bdp1 and TBP (reviewed in [42, 43]). TBP can direct the recruitment of TFIIIB to TATA-containing type 3 promoters. However, in the case of type 1 and type 2 promoters, the additional 'assembly' factors TFIIIA and TFIIIC are required, respectively. TBP is a protein shared by all three RNA polymerase systems [105], whilst Brf1 and Bdp1 are specialised for the Pol III transcription system and can be cross-linked to DNA [106]. Brf1 shows sequence and structural homology to the general transcription factor TFIIB, whilst Bdp1 shows no homology to any other known transcription factor.

1.4.1 TFIIIC recruits and assembles TFIIIB onto type 2 promoters

As previously described, τ60 and τ131 have been shown to interact with TFIIIB subunits. τ60 interacts with TBP and τ131 is thought to be the primary TFIIIC subunit responsible for the assembly of TFIIIB on type 2 promoters. The role of τ131 in TFIIIB

assembly has been extensively studied and a pathway for TFIIIB recruitment has been proposed (reviewed in [87]). Yeast two-hybrid analysis initially revealed that the TBP and Brf1 proteins interact and that Brf1 makes additional contacts with τ131 [107]. Subsequently, various biochemical and biophysical techniques have been used to map the binding of Brf1 and Bdp1 to a conserved N-terminal TPR array in τ131 [108-114]. Mutational analysis and biochemical mapping have led to the conclusion that Brf1 and Bdp1 compete for TPR binding sites within τ131 [109]. *In vivo*, it is thought that Brf1 must first be bound to TFIIIC (along with TBP) before Bdp1 can bind. The C-terminal half of Brf1 provides a binding platform for TBP and Bdp1 [115]. The binding of Bdp1 leads to a stable TFIIIB-TFIIIC-DNA complex [116] that can initiate transcription with Pol III. This step-by-step recruitment of TFIIIB components presumably requires flexibility within τ131 and/or TFIIIB [117].

1.4.2 Pol III recruitment and transcription initiation

Yeast two-hybrid studies, biochemical pull-downs and photo cross-linking have all contributed to elucidating the network of interactions that occurs between TFIIIC, TFIIIB and Pol III (*Figure 1.4*). With TFIIIB assembled on the type 2 promoter, the polymerase can be recruited and PIC formation can be completed. The C37/53 Pol III dimer appears to play a central role in connecting components of the PIC, with C37 shown to interact with Bdp1 and C53 interacting with τ131 [118]. Brf1 has been shown to interact with the Pol III subunit C17 [17] as well as with C34, C160 and C128 [115] via sites in its TFIIB-like N-terminus. τ131 has also been shown to interact with ABC10α [119]. In yeast, τ131 appears to be the only TFIIIC subunit which interacts with Pol III and thus may play a crucial role in its recruitment to type 2 promoters, again explaining its high conservation. It should be noted that in humans however, other TFIIIC subunits have been shown to interact with Pol III (reviewed in [42]).

A time-resolved 'movie' of how these interactions occur in the context of PIC assembly is still unresolved. However, although TFIIIB is still considered the initiation factor for 'proper' Pol III transcription [120], TFIIIC may still be important for directing Pol III accurately to transcription start sites through interactions with τ131.

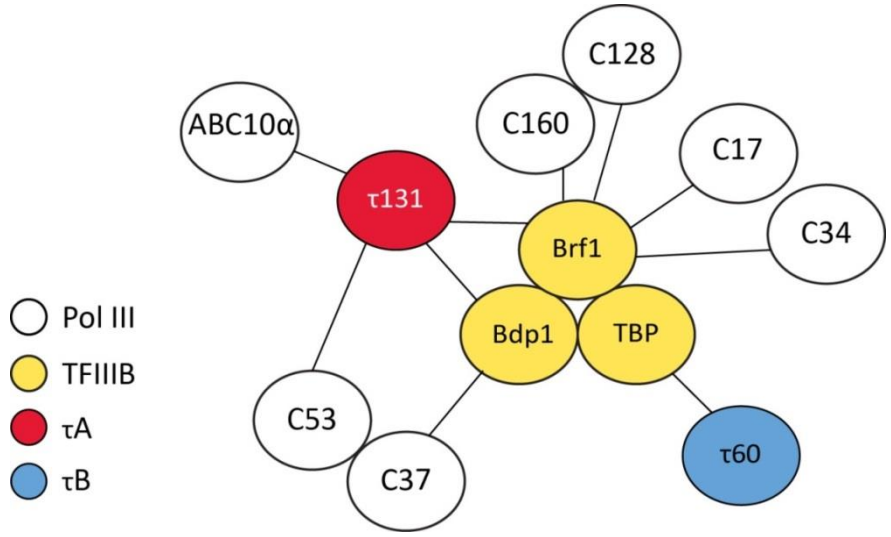

Figure 1.4 **Known interactions within the *S. cerevisiae* PIC**. Lines indicate interactions that have been determined by yeast two-hybrid, pull-downs or photo cross-linking experiments. See text for details and references.

1.4.3 Transcription termination and recycling

The mechanisms of transcription termination in Pol III are poorly understood, despite the apparent simplicity of the termination signal. The termination signal is a run of consecutive thymine residues in the non-template strand, the length of which is species dependent [121]. The process of termination has been described as consisting of two phases. Firstly, a pausing step upon the encountering of the poly-thymine stretch by Pol III and secondly a release step of the RNA [122]. A very recent study has also demonstrated the requirement for hairpin formation within the synthesised RNA to ensure termination [123]. Pol III subunits C53 and C37 have been implicated in the process of termination [48] as well as the TFIIIC subunit τ91, which cross-links to the termination signal [84].

One fascinating aspect of Pol III transcription is the predicted mechanism of facilitated recycling [124, 125]. This pathway allows the terminating Pol III to be efficiently re-recruited to the transcription start site and initiate a second round of transcription. The precise details of the pathway remain unclear [125] but several proteins (such as PC4, NF1 and DNA topoisomerase I) have been implicated in the stimulation of this process [126, 127]. Interestingly, the presence of TFIIIB alone appears sufficient to direct re-initiation of transcription, with TFIIIC only required for recycling at longer genes [128].

The ‘dispensability’ of TFIIIC for re-initiation and the paradox that TFIIIC sits over the entire gene that Pol III must transcribe (and so ultimately acts as a barrier to transcription), raises the question of the role of TFIIIC after PIC establishment. These topics, amongst others, are discussed in the next section.

1.5 TFIIIC within the genome

1.5.1 Genome occupancy of TFIIIC during transcription

It has been estimated that 70-85% of the nucleotides present in a cell are designated for the production of tRNAs, rRNAs, and mRNAs that encode ribosomal proteins [129]. As a consequence, these processes must be regulated to ensure the cell does not use up vast amounts of resources and energy unnecessarily. Interestingly, tRNA genes and 5S rRNA genes are co-localised in specific areas of the nucleus, suggesting a coordinated regulation [130].

When the cell faces a situation of favourable growth conditions (i.e. sufficient nutrients) transcription rates increase, when the situation is less favourable the transcription rate decreases. This scenario has been examined genome-wide with regards to the occupancy of TFIIIC [131]. In a study by Roberts and colleagues, it was observed that Pol III occupancy dramatically reduced on selected tRNA genes upon repression of nutrient availability. Strikingly, TFIIIC occupancy greatly increased, whereas TFIIIB occupancy remained constant. Reversion to conditions of plentiful nutrients increased Pol III occupancy and caused a decrease in the presence of bound TFIIIC at these genes. These observations link back to the questions raised in *Section 1.4.3*: what happens to TFIIIC during transcription elongation? It is assumed that TFIIIC must be released from the actively transcribed gene, in order for Pol III to gain access to the template strand. This idea is supported by another study which also showed far lower occupancy of TFIIIC at actively transcribing genes when compared to Pol III and TFIIIB [132].

The precise molecular mechanisms that are involved at actively transcribed Pol III loci throughout the genome remain elusive. Although evident that TFIIIC is essential for initially assembling TFIIIB and the formation of a PIC, the subsequent requirements for this factor in elongation, termination and re-initiation remain unresolved.

1.5.2 The non-transcriptional roles of TFIIIC

The non-transcriptional roles of TFIIIC, its association with chromatin and binding occupancy over the genome have recently gained particular attention and have been well summarised elsewhere [66, 133, 134]. This research was initiated by the identification of isolated B box elements in the genome of *S. cerevisiae* [135]. It was a surprise to observe that these canonical B box elements were associated with TFIIIC but not with TFIIIB or Pol III. These sites were later identified in *S. pombe* [136] and have since been termed 'extra TFIIIC' (ETC) sites. It was observed that one ETC site is situated in the promoter region of *TFC6*, a gene encoding the τ91 subunit of TFIIIC itself, and that TFIIIC binding leads to a decrease in transcription of this gene [137]. This hints at a cross-talk between the RNA polymerase II and Pol III transcription apparatuses, but many other non-transcriptional roles have been attributed to TFIIIC as well (*Figure 1.5*) [133].

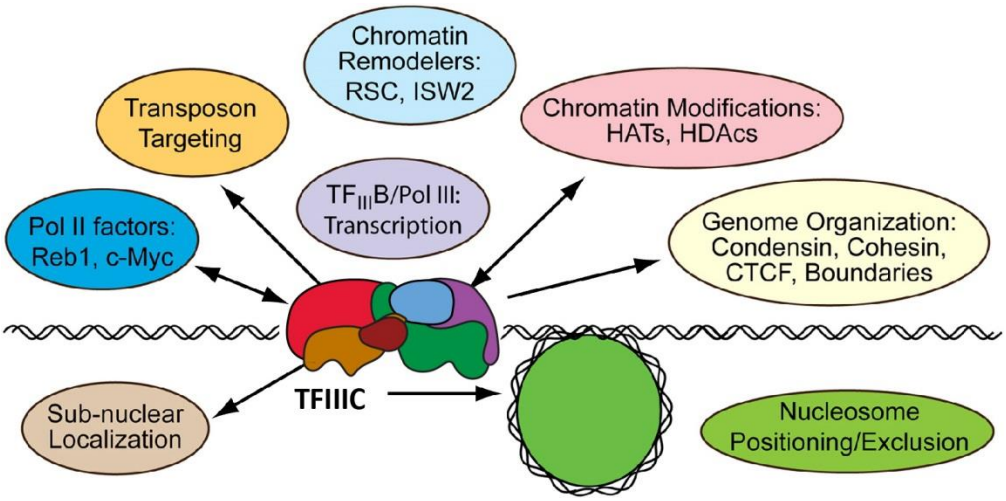

*Figure 1.*5 **TFIIIC as a platform for chromatin processes.** Figure modified from [133].

The binding of TFIIIC to tRNA genes, through isolated or Pol III-associated B boxes, may act as a boundary to heterochromatin spreading and act to organise genome architecture (reviewed in [133, 138]). In addition, the association of TFIIIC with condensin, a protein crucial for the maintenance of chromosome architecture, has been implicated in these processes (reviewed in [139]). Work carried out in *S. cerevisiae* and *S. pombe* has demonstrated co-localisation of condensin with TFIIIC at both Pol III-transcribed B boxes and ETC B boxes [140, 141]. Furthermore, a mutation

in the τ138 subunit (the temperature-sensitive G349E introduced in *Section 1.3.6*) rescues a mutation in condensin that causes a decrease in association of tRNA genes with centromeres [141]. The interplay of condensin and TFIIIC may thus be important for maintaining accurate chromatin architecture in the nucleus of eukaryotic cells.

1.6 Scope of this thesis

I have outlined our current knowledge and understanding concerning the structure and function of the TFIIIC complex. High-resolution structural information of TFIIIC sub-domains has been obtained and there are accumulating insights into the positioning of individual subunits on DNA and the recruitment cascade of TFIIIB and Pol III.

There are however key questions that remain:

- **What is the identity of the linker between the τA and τB subcomplexes?**
- **Which domain or domains bind the B box of type 2 promoters?**
- **What is the overall molecular architecture of TFIIIC?**

During the course of my PhD I attempted to address these questions by using a variety of biochemical, biophysical and structural techniques. Here I present mass-spectrometry coupled cross-linking analyses of the entire TFIIIC complex and identify and characterise the first link between a τA and a τB subunit. I also present the first high resolution structure of a τ138 domain and investigate the identity of the B box binding domain. Finally, I also present my work on producing TFIIIC recombinantly in insect cells.

2. Results - Linking τA and τB

This chapter describes the biochemical and structural characterisation of a link between the two largest subunits of TFIIIC, namely τ131 in the τA subcomplex and τ138 in the τB subcomplex. The current working hypothesis in the field is based on a 'flexible linker' between the τA and τB subcomplexes, that involves one or more TFIIIC subunits.

In our lab, the crystal structure of τ131 (123-566), hereafter referred to as the 'TPR array' had been solved to 3.4 Å resolution by a former PhD student, Nicholas Taylor . Here, I present experiments which reveal that this domain is involved in a link with the τB subunit τ138. I will show the precise mapping of the interaction by GST pull-down assay, isothermal titration calorimetry (ITC), cross-linking and mutational analysis. Unexpectedly, the mapped interaction site within the TPR array of τ131 overlaps with previously characterised binding sites for TFIIIB.

2.1 TFIIIC cross-linking

As a first step in identifying TFIIIC subunits that may bridge the gap between the τA and τB subcomplexes, I carried out chemical cross-linking of endogenously purified TFIIIC (coupled with mass spectrometry) in collaboration with the Beck group at EMBL Heidelberg.

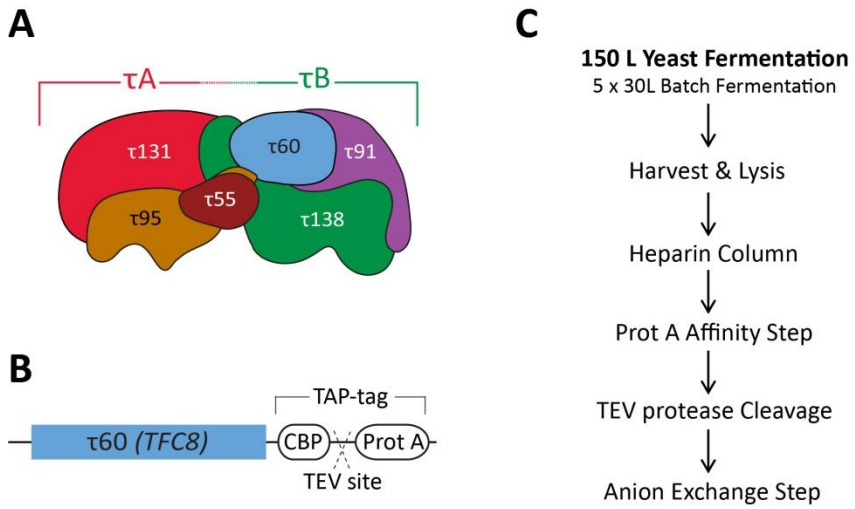

Figure 2.1 **TFIIIC can be purified from *S. cerevisiae***. (A) Schematic diagram of TFIIIC. (B) Schematic diagram of the TAP-tag at the C-terminus of the τ60 gene. CBP = calmodulin binding peptide; Prot A = protein A; TEV = Tobacco Etch Virus. (C) Main purification steps. Details can be found in the *Appendix*.

2.1.1 Purification of endogenous TFIIIC from *S. cerevisiae*

TFIIIC was purified from *S. cerevisiae* by means of a tandem affinity purification (TAP) tag at the C-terminus of τ60. The initially established protocol was modified and optimised by Nicholas Taylor and myself. *Figure 2.1* summarises the purification strategy, with further details available in the *Appendix*. In preparation for lysine-lysine cross-linking, the sample eluted from the final purification step is buffer exchanged into HEPES buffer (pH 7.5) using an analytical Superose 6 10/300 gel filtration column (GE Healthcare). This step was necessary to ensure that no Tris is present in the buffer, as the amine groups of Tris molecules can compete for cross-linking and thus quench the reaction. The analytical gel filtration step also serves as a quality control step for the integrity and purity of the preparation (*Figure 2.2*).

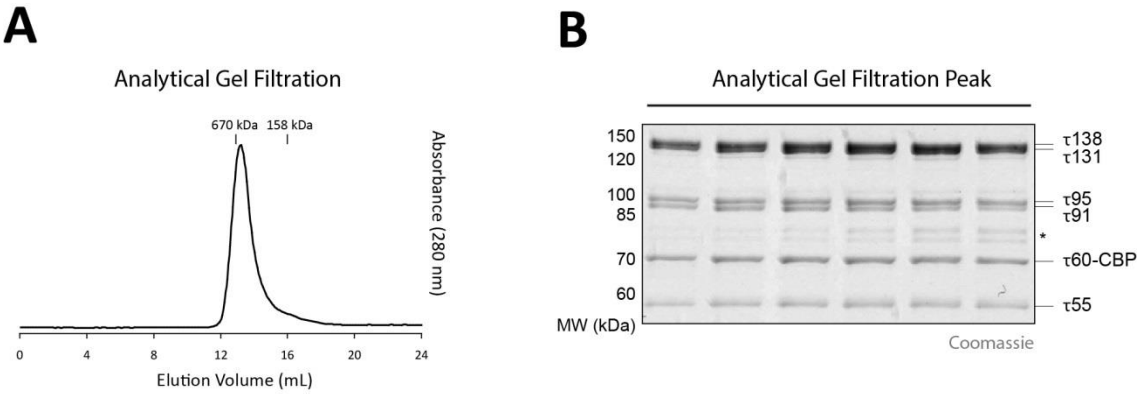

Figure 2.2 **Endogenous TFIIIC can be purified to homogeneity.** (A) The samples were injected onto a Superose 6 10/300 column (GE Healthcare) and elution position was analysed. Positions of known molecular weight standards are indicated at 670 kDa and 158 kDa. (B) 12% SDS-PAGE of gel filtration peak. * = τ91 degradation product, confirmed by mass spectrometry; CBP = calmodulin binding peptide; MW = molecular weight; kDa = kiloDalton.

2.1.2 Chemical cross-linking of purified TFIIIC

The principles behind chemical cross-linking coupled with mass spectrometry have been well reviewed [142]. In short, native protein complexes are incubated with a cross-linking reagent that can covalently link primary amines that are present on the surface of a protein complex. These primary amines, which occur in lysine side chains and the N-termini of proteins, will be cross-linked if they occur within the length of the cross-linking reagent. After digestion of the cross-linked protein complex by proteases, cross-linked peptides can be identified using mass spectrometry, thus providing restraints that can give information on the architecture of subunits within a protein

complex. In collaboration with the Beck group, we used state of the art liquid chromatography (LC)-coupled tandem mass spectrometry (MS/MS) with chemical cross-linking in order to obtain a map of distance restraints for TFIIIC that could indicate binding interfaces.

A detailed protocol of the cross-linking method can be found in *Materials and Methods*, *6.1.2*. Briefly, 30 µg (1 µg/µl) of purified TFIIIC was cross-linked by addition of an iso-stoichiometric mixture of H12/D12 isotope-coded, di-succinimidyl-suberate (DSS, Creative Molecules). Equal amounts of cross-linker were added ten times every 4 min to a final concentration of 2 mM. The reactions were then quenched and digested, before cross-linked peptides were enriched using size-exclusion chromatography [143]. The use of an isotope-labelled cross-linker aids mass spectrometry identification in the next step, as modified peptides occur as isotope pairs on the MS1 level. LC-MS/MS was employed for analysis of the cross-linked samples, with a final list of identified peptides filtered by the following criteria: false discovery rate (FDR) = 0.05, minimum delta score = 0.95, MS1 tolerance window of 4 to 7 ppm, linear discriminant (ld)-score > 25 (*Figure 2.3* and *Appendix*). An ld score of over 25 was chosen to reduce false-positives, and was based on analysis of test cases where obtained cross-links were mapped onto an already solved crystal structure (Jan Kosinski, personal communication).

2.1.3 Cross-linking reveals a link between subunits τ131 and τ138

We identified 121 unique cross-links (33 inter-subunit/88 intra-subunit) for TFIIIC that provide interesting new insights into various aspects of TFIIIC architecture. For this chapter, I will focus on the interface observed between τ131 and τ138. In this region a dense network of cross-links extend from the TPR array at the N-terminus of τ131 to a central region of τ138 (*Figure 2.3*). With the exception of τ95, this region represents the only link between a τA and a τB subunit identified by cross-linking. Importantly, the cross-links occurred within the TPR array of τ131, a domain of which we already had a structure, which provided a detailed framework in which to investigate the interaction in detail.

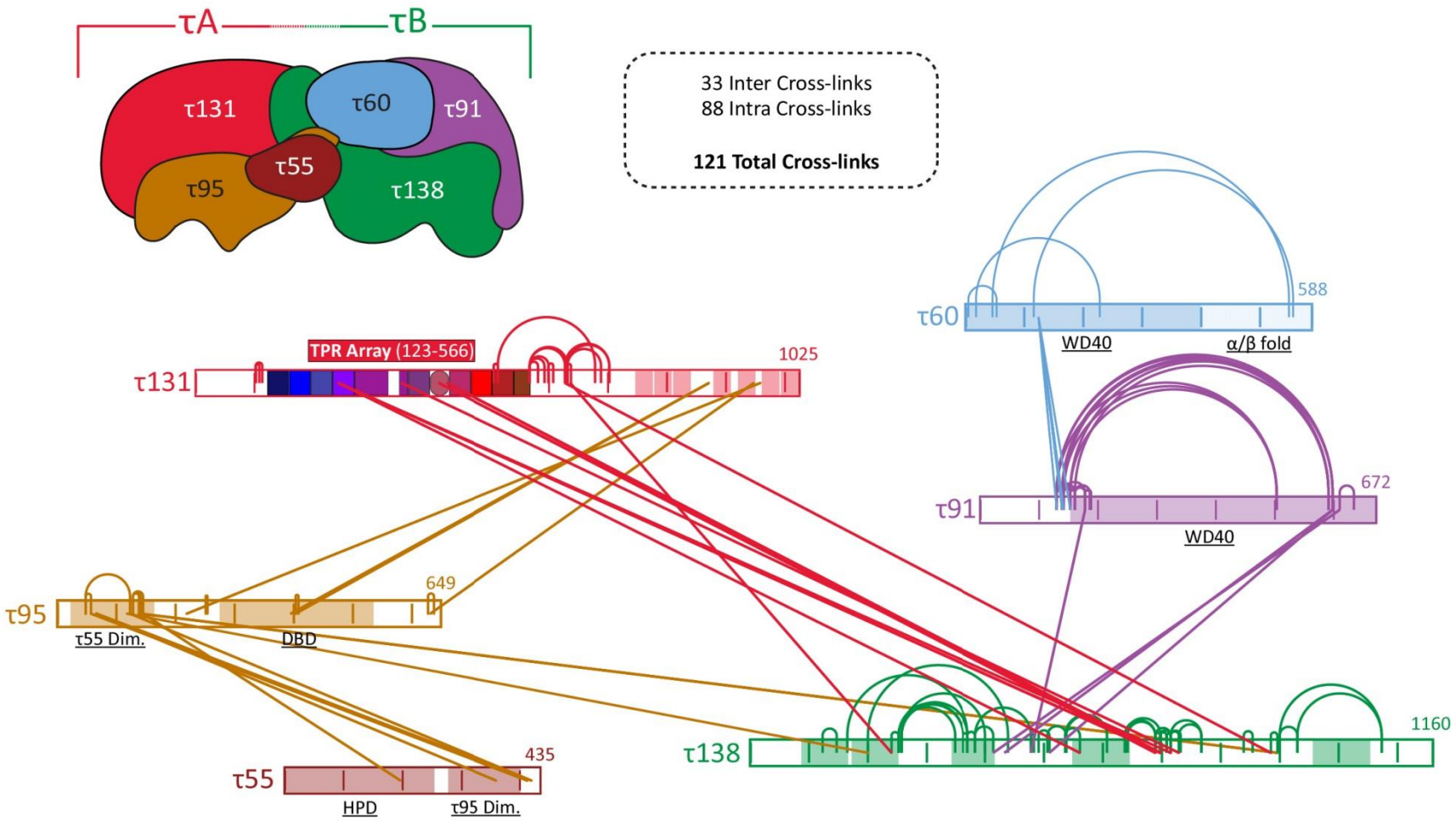

Figure 2.3 **Cross-linking map of TFIIIC .** Domains of which crystal structures are available are underlined. τ55 and τ95 Dim. = τ55 and τ95 dimerisation domains; DBD = DNA-binding domain; HPD = histidine phosphatase domain. The TPR array of τ131 is highlighted. Shaded green boxes on τ138 indicate predicted winged helix domains. Shaded red boxes on τ131 indicate predicted TPRs. Inter cross -links are depicted by lines connecting different subunits. Intra cross-links are depicted by arcs which connect residues within the same subunit.

2.2 Mapping the τ131 – τ138 interaction

The structure of the TPR array of τ131 is shown in *Figure 2.4*. The TPR array is highly conserved, as can be seen from the schematic structure predictions of τ131 orthologues in *S. pombe* and *H. sapiens*. τ131 is predicted to contain many TPR repeats, in addition to the ones within the TPR array at the N-terminus. A survey of different structure prediction servers suggests that up to 7 extra TPRs might exist within the C-terminal half of the protein. That the cross-links obtained occurred between the TPR array and a largely disordered region of τ138 is not so surprising, as TPR-containing proteins often stabilise extended, disordered regions within protein complexes [85].

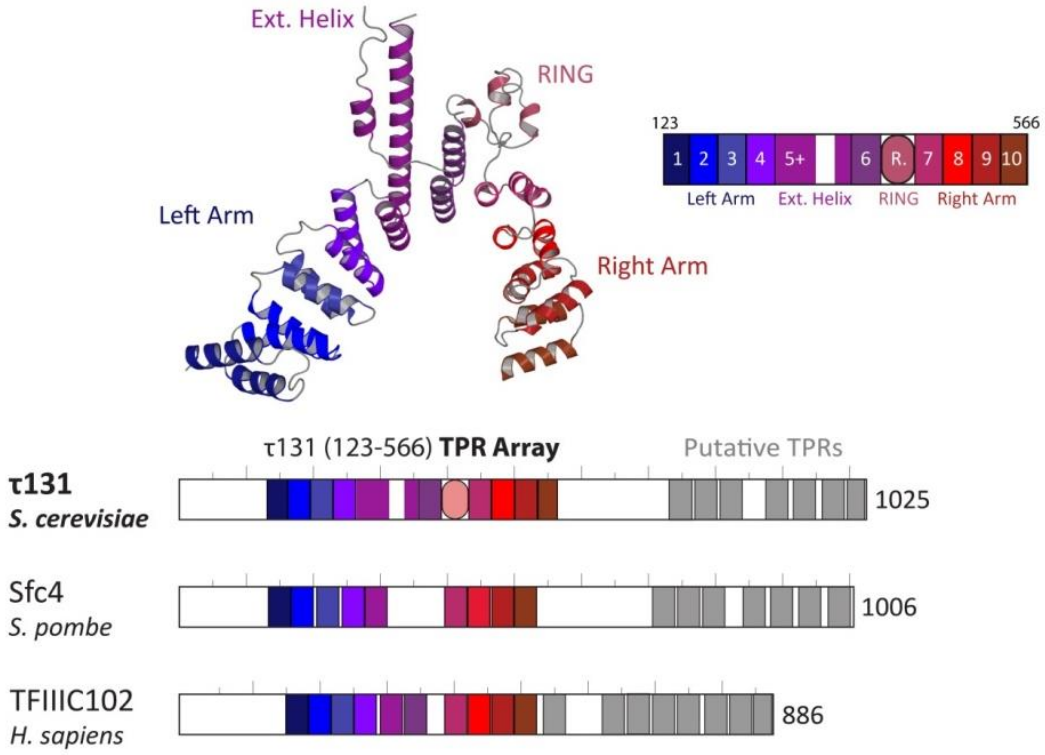

Figure 2.4 **Overview of the tetra-trico-peptide repeat (TPR) domains found in τ131 orthologues from *S. cerevisiae*, *S. pombe* and *H. sapiens***. The TPR array is shown in ribbon representation and is coloured according to the scheme displayed. Putative additional TPRs are coloured grey. Ext. = extended.

Figure 2.5 shows predicted structural motifs found within τ138 and orthologues in *S. pombe* and *H. sapiens*. Bioinformatic predictions revealed the presence of three winged helix domains at the N-terminus of the protein, one or two at the C-terminus and one central winged helix, immediately adjacent to the cross-links. Winged helix domains are usually implicated in binding DNA, but have also been characterised as

protein-protein interaction domains [95]. I therefore set out to identify the minimal region in τ138 required for interaction with τ131, focussing on constructs which included the predicted central winged helix.

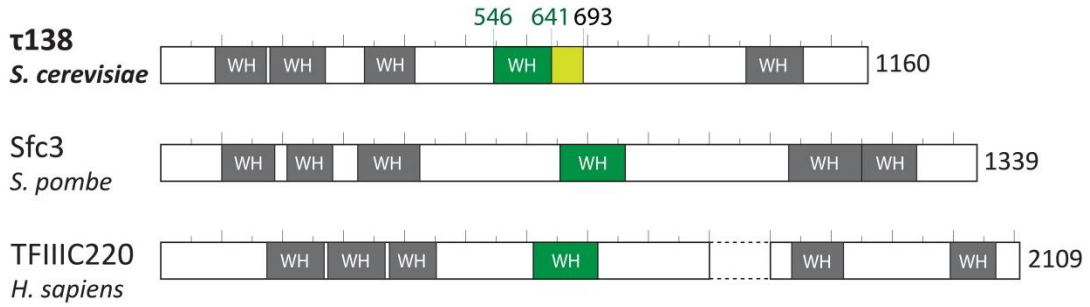

Figure 2.5 **Overview of the winged helix (WH) domains predicted to be found in τ138 orthologues from *S. cerevisiae*, *S. pombe* and *H. sapiens***. The central winged helix and the additional region investigated are highlighted in dark green and light green, respectively. Dashed lines in TFIIIC220 indicate a non-conserved region which is removed here for simplicity.

2.2.1 Biochemically mapping a critical τ131- τ138 interaction region

To map which part of the central τ138 region interacts with the TPR array of τ131, I performed a Glutathione S-Transferase (GST) pull-down assay. I purified the TPR array for these studies according to established protocols. Based on the obtained cross-links, I designed and cloned various τ138 fragments as GST-fusions. I used the pETM30 vector, which introduces an N-terminal, Tobacco Etch Virus (TEV) protease-cleavable, 6x-histidine-GST affinity tag. The GST-fusions were expressed and purified as detailed in *Materials and Methods, 6.1.4*.

The results of the GST pull-down assay are shown in *Figure 2.6*. Constructs that included the winged helix up to residue 693 and a construct that just represented the disordered region up to this residue (641-693) showed a stoichiometric interaction with the τ131 TPR array, but the winged helix (WH) alone did not pull-down the τ131 TPR array. Extending the winged helix to residue 681 did reveal an interaction with τ131, but it appeared sub-stoichiometric. A construct including only the disordered region up to this residue (641-681) also showed a very weak, sub-stoichiometric interaction. A negative control of GST alone was included to control for interaction of the τ131 TPR array with GST. The first lane in the 'Pull Down' gel shows that τ131 does not interact with GST.

The results from this assay suggested that the region 641-693 (hereafter referred to as τIR – τ131-Interacting Region) of τ138 is necessary and sufficient to form a stable, stoichiometric interaction with the TPR array of τ131.

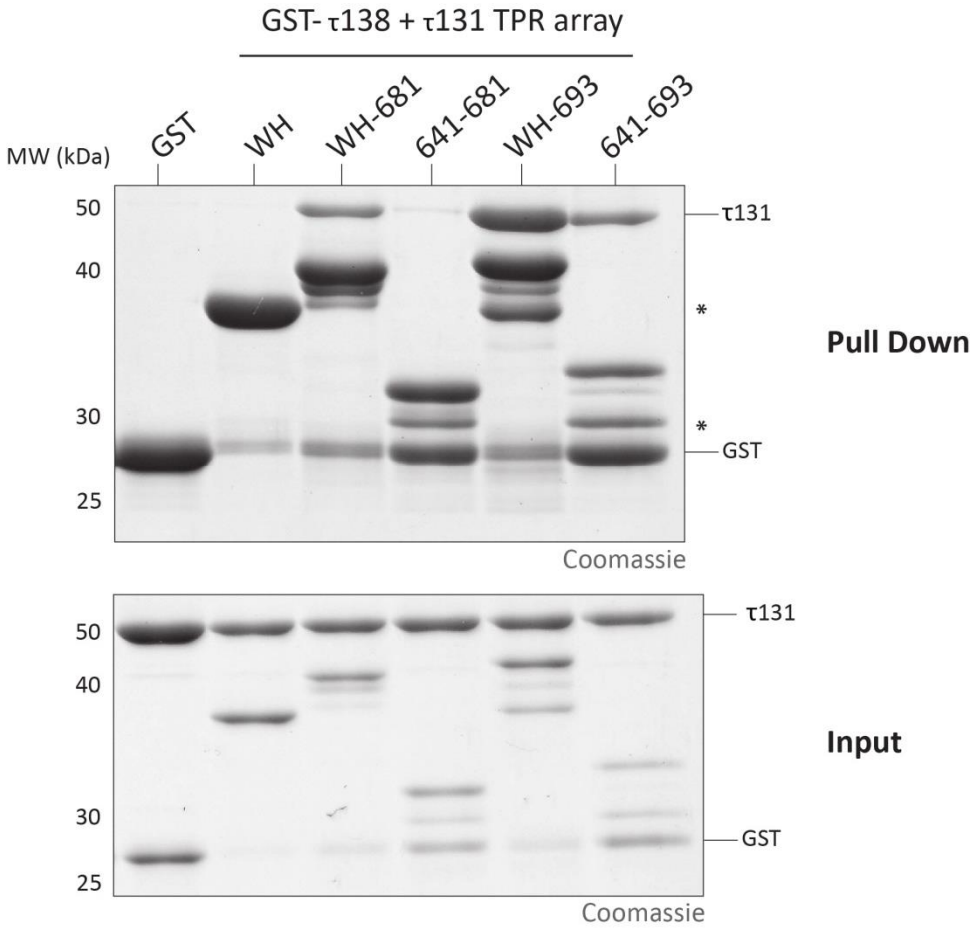

Figure 2.6 **GST pull-down assay confirms a τ131-τ138 interaction.** GST pull-down assay of recombinant GST-τ138 protein and GST with untagged τ131 TPR array. 12% SDS-PAGE of pull-down (top) and input (bottom) samples are shown. WH = winged helix (546-641). * = degradation products of τ138.

2.2.2 Biophysically mapping a critical τ131- τ138 interaction region

ITC was used to confirm the interaction site mapped with GST pull-down and to obtain precise binding affinities of the interacting fragments in the absence of affinity tags. ITC was carried out in collaboration with Vladimir Rybin, Biophysical Support Scientist at EMBL Heidelberg. I purified 'WH', 'WH-681' and 'WH-693' as before, but this time cleaving the GST tag with TEV protease and using gel filtration to purify pure, cleaved proteins (data not shown). As the constructs '641-681' and '641-693' were short and predicted to be disordered, I carried out a different purification protocol that is detailed in *Materials and Methods, 6.1.5*. In this protocol, the proteins were purified

by the 6x-histidine tag (rather than the GST tag), using nickel-NTA agarose resin (Qiagen). After TEV protease cleavage, the cleaved proteins were passed through a second nickel column that bound the TEV protease and un-cleaved protein, but allowed cleaved protein to pass in the flow-through (*Figure 2.7A*). I then used a MonoS cation exchange column for a final purification round (*Figure 2.7B*). The proteins were checked for aggregation on a Superdex Peptide 10/300 GL analytical gel filtration column (GE Healthcare) (data not shown). Despite being predicted as generally unstructured (*Figure 2.7C*), these proteins were not observed to form aggregates and so were suitable for ITC measurement.

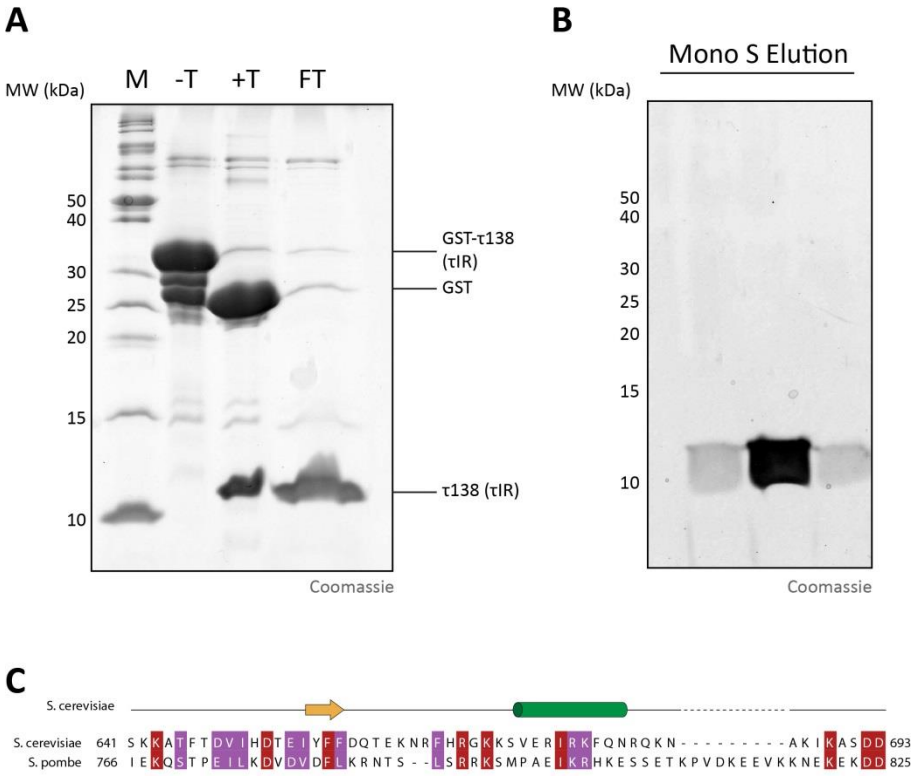

Figure 2.7 **The τ138 'τIR' (641-693): Purification and sequence conservation.** (A) 18% SDS-PAGE of the TEV protease cleavage of GST-τ138 τIR. M = marker; T = TEV protease; FT = flow-through. (B) Elution fractions from MonoS column. (C) Alignment of *S. cerevisiae* and *S. pombe* 'τIR' sequences. Identical residues are boxed in brick red. Highly conserved residues are boxed in purple. The secondary structure prediction is highlighted above the alignment. Cylinders denote α-helices; arrows denote β-strands.

The construct containing the winged helix with the τIR region binds to the τ131 TPR array with a K_d of 100 nM (*Figure 2.8A*). A similar value (80 nM) is obtained for the τIR alone (*Figure 2.8B*). A construct containing the winged helix extended to residue 681 still showed binding, but at a weaker affinity of 2.6 μM (*Figure 2.8C*). The construct

encompassing residues 641-681 showed no binding (data not shown). Finally, the winged helix alone showed no binding to the τ131 TPR array by ITC (*Figure 2.8D*).

In summary, the findings from ITC experiments are consistent with those from the GST pull-down assay: the τIR is necessary and sufficient for binding the τ131 TPR array with nanomolar affinity.

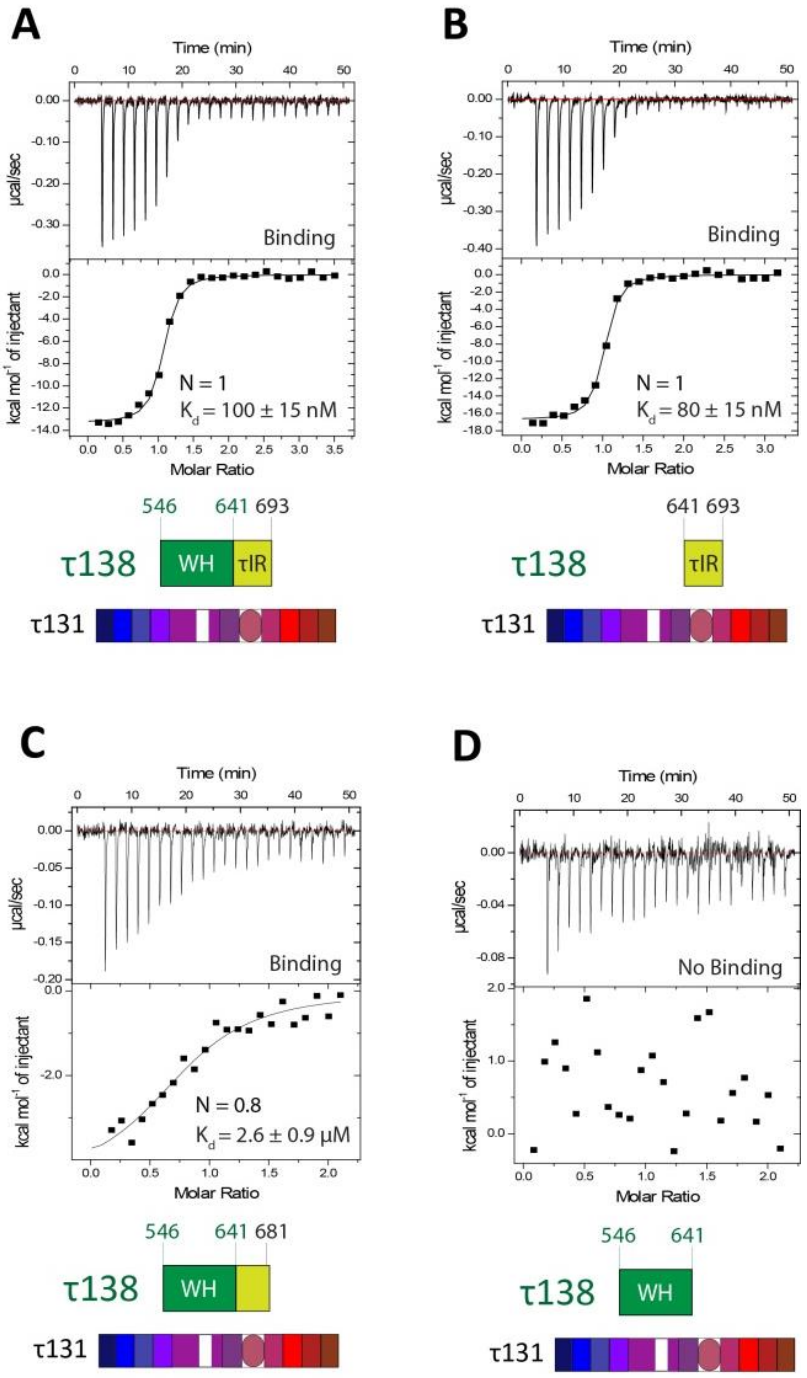

Figure 2.8 **Determining binding affinities for the τ131-τ138 interaction using ITC.** (A-D) See text for explanation. N = binding stoichiometry between τ131 and τ138; K_d = dissociation constant.

2.3 Crystallisation attempts of a τ131 TPR array - τ138 τIR complex

Having defined a minimal region of τ138 that interacted with τ131 and intrigued by how this disordered region is stabilised by the TPR array, I next set out to purify a complex of these two regions in order to crystallise the complex and solve its structure.

2.3.1 Purifying the τ131 - τ138 complex

The reader is referred to *Materials and Methods* for full details of cloning, expression and purification protocols. Vectors encoding the sequences for the τ131 TPR array and τ138 τIR were co-expressed in p*RARE *Escherichia coli* (*E. coli*) cells. The complex was purified by means of an N-terminal, TEV protease-cleavable, GST affinity tag present on the τ138 subunit. *Figure 2.9A* shows the elution peak of the τ131-τ138-GST complex. As can be easily appreciated, a large amount of material could be obtained (~200 mg from 12 L of *E. coli*). The complex was incubated with GST-tagged TEV protease to remove the GST tag and the flow-through was collected. *Figure 2.9B* shows the result of the TEV protease cleavage, indicating the successful collection of a τ131-τ138 cleaved complex in the flow-through.

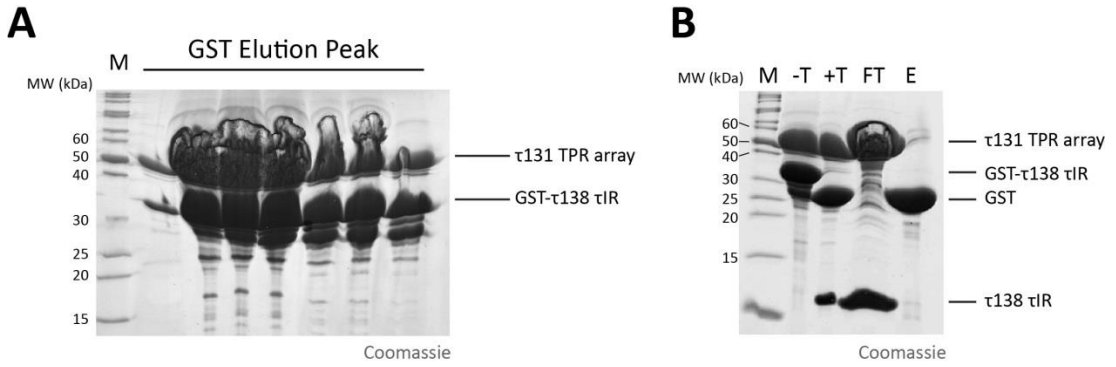

Figure 2.9 **GST-purification of the τ131 TPR array - τ138 τIR complex**. (A) 12% SDS-PAGE showing elution fractions from a GST-affinity column. (B) 18% SDS-PAGE showing TEV protease cleavage of the τ131-τ138 complex. M = marker; T = TEV protease; FT = flow-through; E = 2nd elution of free GST and GST-TEV protease

At this stage, it was apparent that this purification strategy resulted in a large yield of a highly pure sample. The complex was injected onto a Superdex 200 preparative gel filtration column (GE Healthcare) and eluted as a single species (*Figure 2.10A*). The elution volume corresponds to a higher molecular weight than a monomer (~57 kDa);

this is probably due to the extended shape of the TPR array and was also observed for the apo-τ131 TPR array. The complex was evaluated by SDS-PAGE (*Figure 2.10B*) and judged to be appropriate for crystallisation trials, as it was highly pure and showed a stoichiometric relationship between τ131 and τ138.

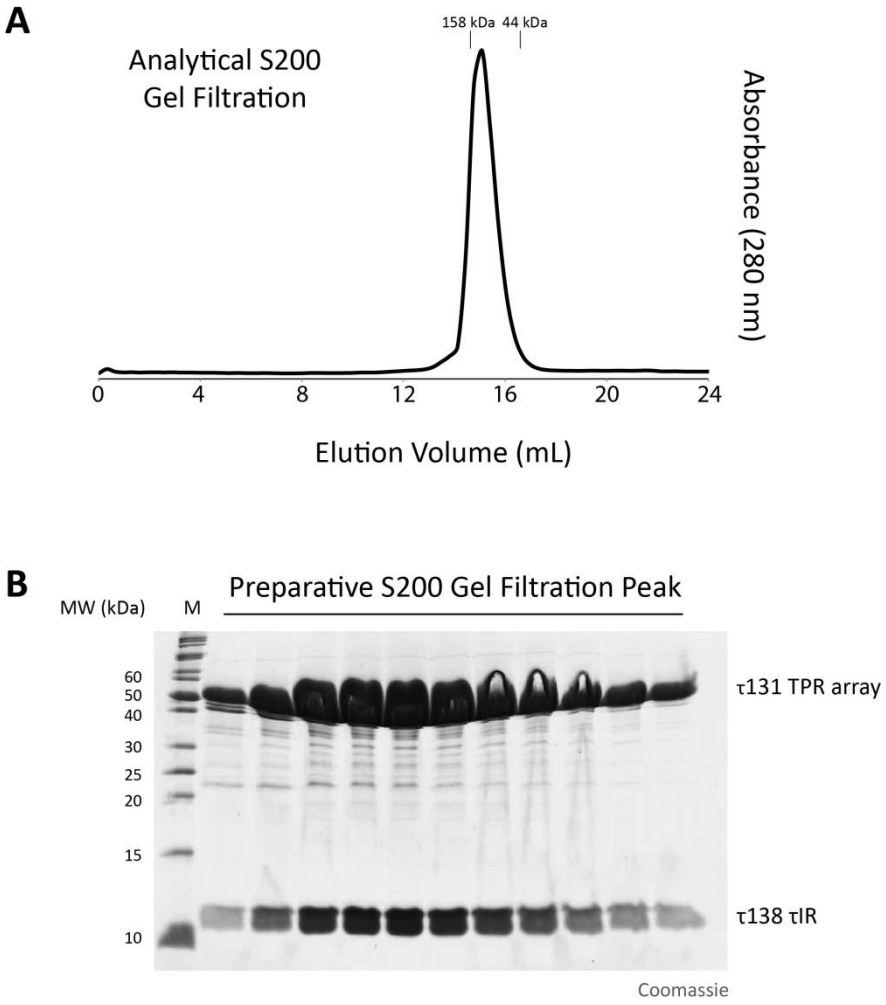

Figure 2.10 **The τ131 TPR array - τ138 τIR complex can be purified to homogeneity.** (A) Analytical gel filtration profile of the τ131-τ138 complex. The sample was injected onto a Superdex 200 10/300 GL column (GE Healthcare) and elution position analysed. Positions of known molecular weight standards are indicated at 158 kDa and 44 kDa. (B) 18% SDS-PAGE of a preparative S200 gel filtration peak. M = marker.

2.3.2 Crystallisation of the τ131 - τ138 complex

The complex was concentrated to 60 mg/mL for initial crystallisation trials. Crystals were grown at 20°C using the sitting drop method in an initial broad screening approach. Initial crystals were obtained in the JCSG screen (Molecular Dimensions) with a reservoir solution of 0.2 M $MgCl_2$, 0.1 M Tris pH 8.5 and 50% ethylene glycol as

the precipitant (*Figure 2.11A*). Manual refinement screens were set up around this condition with the best hits (i.e. single, well-defined crystals) occurring in 0.2 M $MgCl_2$, 0.1 M Tris pH 8.3 and 42.5% ethylene glycol. The high percentage of ethylene glycol in the condition meant that the crystals were already cryo-protected and therefore could be flash-frozen immediately in liquid nitrogen. One observation at this stage was that there was a large amount of quite thick skin or precipitate formed around some of the crystals although it was not clear whether this was protein or a component of the crystallisation cocktail. It was removed from the crystals for data collection (see *Discussion*).

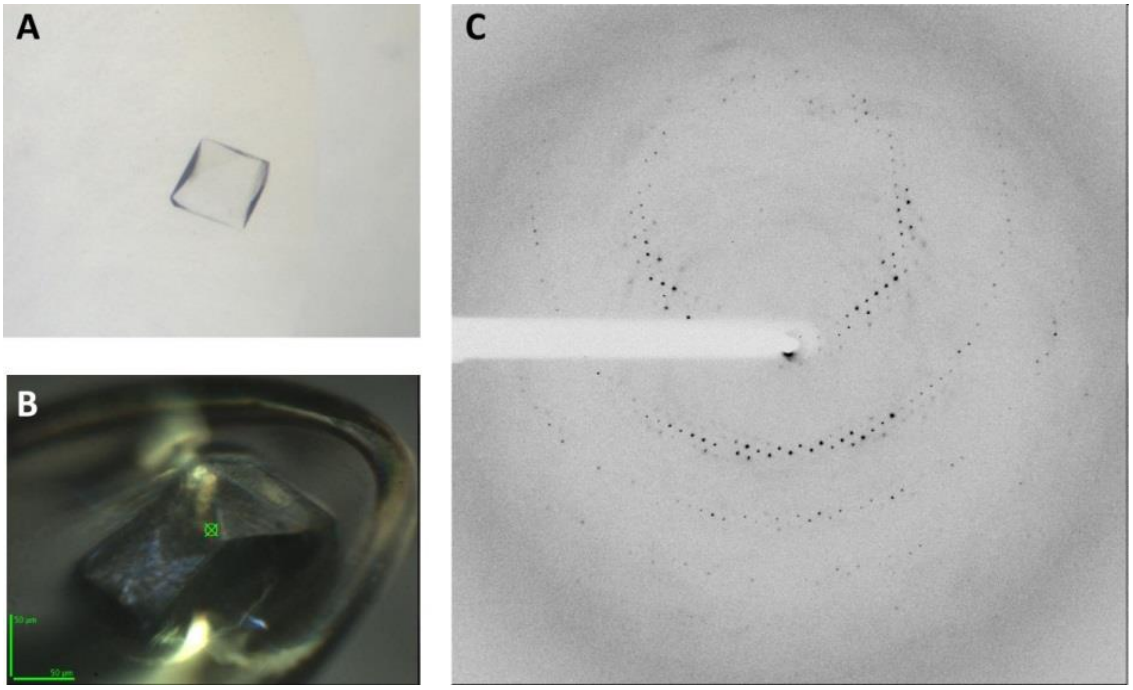

Figure 2.11 **Crystallisation and diffraction of the τ131-τ138 complex.** (A) Typical crystal obtained for the τ131-τ138 complex. (B) Collected crystal in a loop for data collection. (C) Example of a diffraction image taken for the crystal fished in (B).

2.3.3 X-ray diffraction from τ131 - τ138 complex crystals

Crystals were collected for native data collection, as it was hoped that a molecular replacement solution would be found using the apo-τ131 TPR array structure. *Figure 2.11B* and *Figure 2.11C* show a typical collected crystal and diffraction image. Initial processing using XDS [144] suggested that the space group for these crystals was $P6_2$ (see *Materials and Methods, 6.4* for further details on data processing). This was different to that of the apo-τ131 TPR array ($P4_3$). It was apparent from early on that

resolution would be a limiting factor for these crystals. Thus, in order to increase the signal to noise ratio for higher resolution reflections, a highly redundant data set (7200 images, 0.1° rotation) was collected. The crystals did not display signs of radiation damage up to this point. The details of the collected data are summarised in *Table 2.1*. The best dataset obtained extended to 3.3 Å resolution.

Table 2.1 **τ131 (123-566) Data collection and refinement statistics**

	τ131 (123-566) Native	**τ131 (123-566) SeMet**
Data collection[1]		
Beamline	P14 (PETRAIII)	ID23-1 (ESRF)
Space group	$P6_2$	$P6_2$
Cell dimensions *a*, *b*, *c* (Å)	116.08, 116.08, 95.75	121.71, 121.71, 97.78
Wavelength (Å)	0.97626	0.97924
Resolution (Å)	50-3.29 (3.37-3.29)	50-4.50 (4.61-4.50)
CC 50	1.0 (0.56)	1.0 (0.47)
R_{merge} (%)	6.7 (164.0)	9.6 (183.3)
I/σI	37.62 (2.13)	16.65 (1.62)
Completeness (%)	99.7 (97.0)	99.1 (89.9)
Redundancy	38.1 (23.6)	20.3 (16.6)
Sites	-	8 SeMet
Refinement		
Resolution (Å)	44.5 – 3.29	
No. reflections	11247	
R_{work}/ R_{free} (%)	29.6/31.7	
No. atoms		
Protein	3491	
B-factors (Å^2)		
Protein	138.36	
R.m.s deviations		
Bond lengths (Å)	0.003	
Bond angles (°)	0.715	

[1] Values in parentheses correspond to the highest-resolution shell

2.3.4 Solving the structure of the τ131 - τ138 complex

The structure was solved by molecular replacement using Phaser [145] but only after fairly extensive modification of the τ131-apo search model. In particular, the extreme N-terminal and C-terminal TPRs and the extended helix (*Figure 2.4*) had to be truncated, indicating variability in these regions of the protein within these crystals. After iterative rounds of model building and refinement using Coot [146] and the

Phenix package [147] the model was refined to a R_{work}/ R_{free} of 29.6%/31.7% with good stereochemistry (*Table 2.1*). However, no additional density which would be indicative of the τ138 protein was visible, and the reasons for why this might be will be discussed later (*Section 2.6.1*). A simulated-annealing composite omit map of the TPR array from the 3.3 Å dataset is shown in *Figure 2.12A*. In general, the map is of good quality. In less flexible regions, in the helices of the TPRs, side-chains could be placed with confidence (*Figure 2.12B*). The situation for more flexible regions such as at the N- and C-termini and the loops between some of the TPR helices was quite different. Here, owing to high temperature factors (*Figure 2.14B*) the electron density was harder to interpret, with many side-chains showing no clear density.

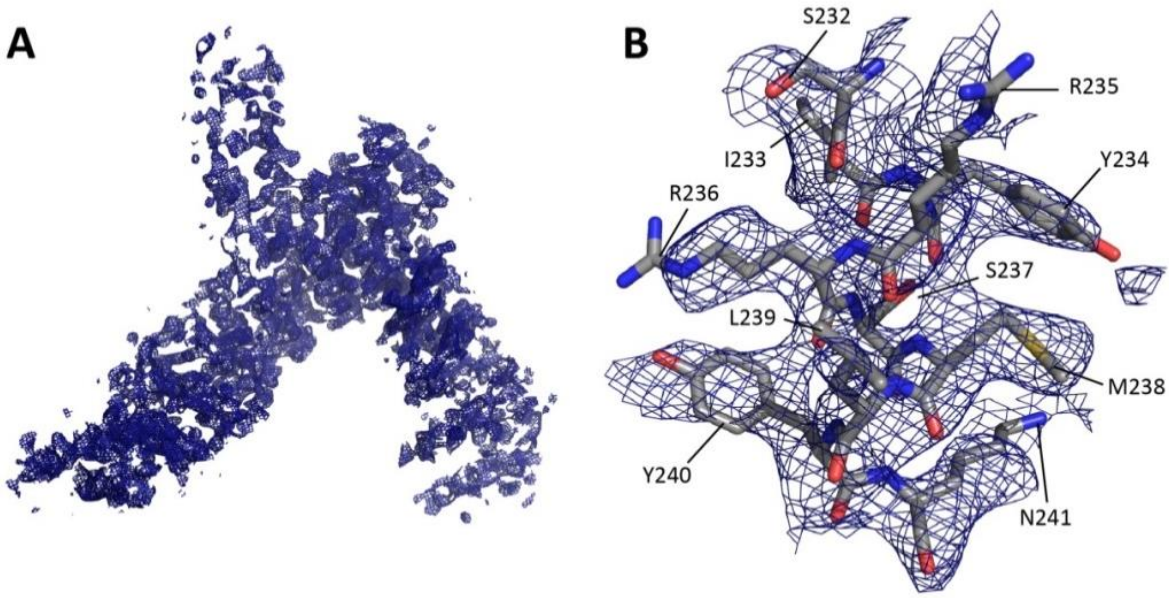

Figure 2.12 **Electron density of the 3.3 Å dataset of the TPR array.** (A) Simulated-annealing composite omit map of the TPR array form the 3.3 Å native dataset contoured at 1σ. (B) Close-up view of a simulated-annealing omit map (blue) of the 'A' helix of TPR 4 (grey) with side-chains labelled (oxygen atoms, red; nitrogen atoms, blue; sulphur atoms, yellow). Map contoured at 1σ.

2.3.5 Selenomethionine incorporation into τ131 - τ138 complex crystals

As an additional validation that the τ131 structure was correct, crystals of a selenomethionine (SeMet) derivative of the τ131-τ138 complex were grown. The complex was expressed and purified as for the native complex, albeit with modifications at the expression stage to ensure that methionine residues were substituted for SeMets (see *Materials and Methods, 6.1.9,* for details). Mass

spectrometry was used to confirm the incorporation of SeMet into the protein. The complex was concentrated to 30 mg/mL for crystallisation trials. Crystals were grown at 20°C by the hanging drop method using refinement screens already established for the native complex. Crystals grew in 0.2 M $MgCl_2$, 0.1 M Tris pH 8.2-8.5 and 42.5% - 45% Ethylene Glycol. Only a few crystals were obtained in comparison with the native complex and they were significantly smaller. The morphology was the same.

A dataset of 3600 images at 0.2° rotation was collected and the statistics of the collected data are displayed in *Table 2.1*. The overall resolution extended to 4.5 Å but more importantly anomalous signal from the selenium atoms was recorded to ~5.8 Å. The structure was solved by single-wavelength anomalous diffraction (SAD) combined with density modification using the program autoSHARP [148], with the program able to locate the positions of 8 SeMet sites. This is consistent with 7 methionine residues and 1 additional cloning artefact at the N-terminus. Upon inspection of the electron density, 6 of the 7 selenium peaks overlapped with built methionine residues (see the legend of *Figure 2.13*). In conclusion, a SeMet dataset confirmed the validity of our molecular replacement solution and indicated that the model had been built correctly. However, again there was no electron density for the τ138 τIR.

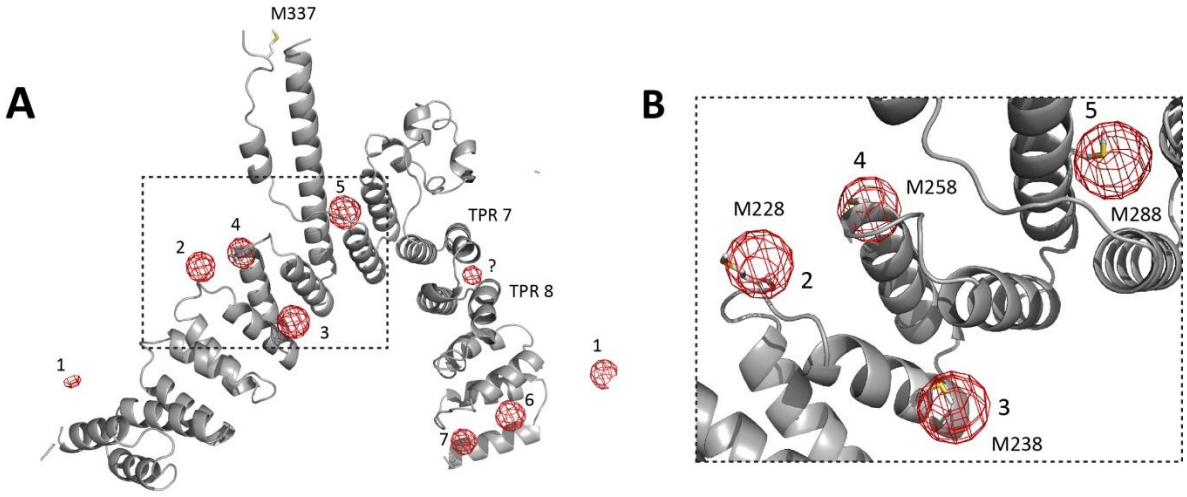

Figure 2.13 **Selenium peaks indicate the positions of methionine residues in the T131 TPR array.** (A) Anomalous difference Fourier map (red) showing selenium positions contoured at 5*σ*. Peak 1 is most likely the N-terminal methionine residue from cloning as it does not overlap with a built methionine residue. A peak between TPR 7 and TPR 8 '?' occurs in a region of no methionine residues and could possibly come from a flexible N-terminal methionine in a symmetry mate (the size of the peak is much smaller in comparison to the others). No selenium peak was observed at M337, probably owing to the high degree of flexibility in this region. (B) Close-up of boxed region from A. Methionine residue are labelled as sticks and numbered accordingly.

2.3.6 Comparing the $P4_3$ and $P6_2$ τ131 TPR array structures.

The $P6_2$ structure of the τ131 TPR array is highly similar to that of the $P4_3$ structure solved originally (RMSD Cα-380 = 1.83 Å) (*Figure 2.14A*). Additional residues are visible at the N and C termini of the $P6_2$ structure and the TPRs at the N and C termini are shifted when compared to the $P4_3$ structure. These regions are also where the most flexibility occurs within the structure (as indicated by the temperature factors displayed in *Figure 2.14B*). In general, the structures are in good agreement, thereby showing that even in two different crystallographic environments and lattices the TPR repeats are arranged in an almost identical shape.

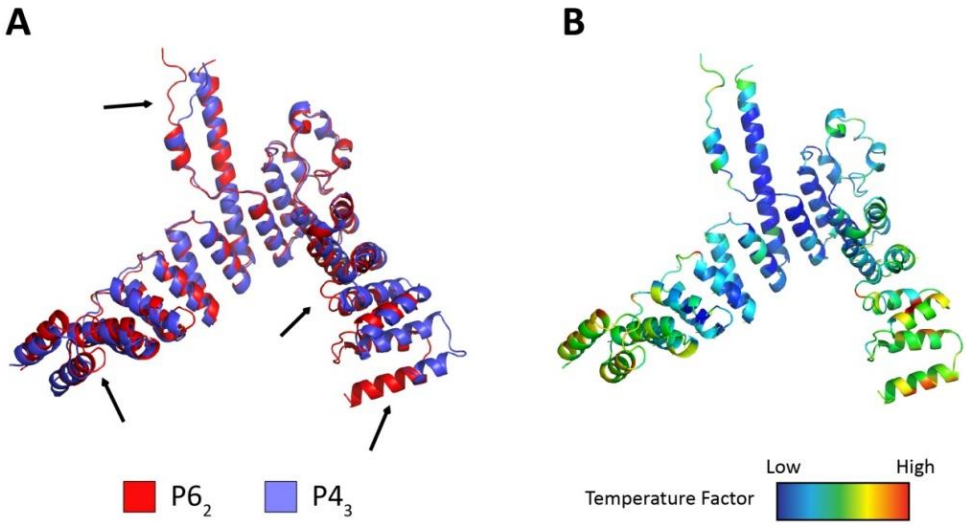

Figure 2.14 **Flexibility within the TPR array.** (A) Overlay of the $P6_2$ TPR array (red) and the $P4_3$ structure solved previously (lilac). Arrows indicate regions of greatest difference. (B) Temperature factors of the $P6_2$ structure.

2.4 Identifying the τ138 - τ131 interaction site

As the efforts to obtain the structure of the τ131-τ138 complex were not successful, I decided to use existing structural information on the τ131 TPR array and analyse the surface conservation and validate known mutational data from the literature biochemically. It was hoped that this would provide clues as to which residues or regions may be involved in an interaction with τ138.

2.4.1 Conservation and mutational analysis of the τ131 TPR array

Surface representations of the τ131 TPR array are displayed in *Figure 2.15*. Conserved regions occur on both arms of the array and in the groove between the two arms. The

extended helix and RING are poorly conserved (*Figure 2.16* and also *Figure 2.4*). The TPR array contains a mixture of acidic and basic patches, some of which overlap with regions of conserved residues. The τIR region of τ138 is largely basic (*Figure 2.7*) suggesting that acidic residues of the TPR array could form part of the interaction site.

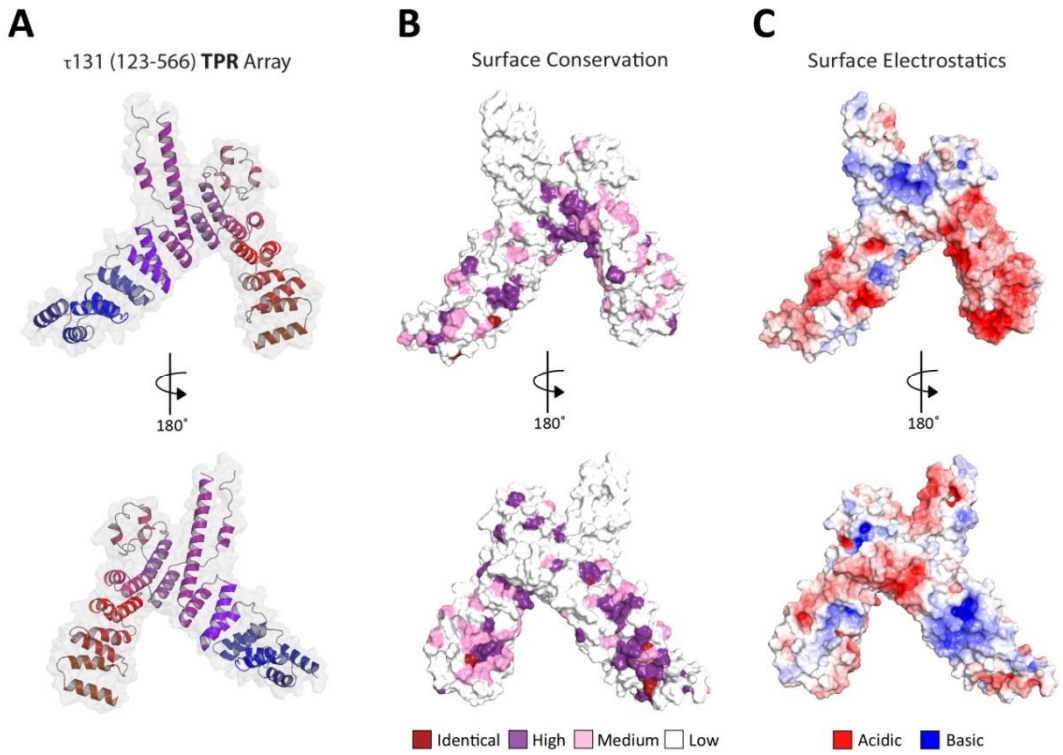

Figure 2.15 **TPR array surface conservation and electrostatics.** (A) View reproduced from *Figure 2.4*. (B) Surface conservation representation. Conserved residues are coloured as shown in the legend. (C) Surface electrostatics coloured as shown in the legend.

Mutation of residues involved in protein-protein interactions may disrupt binding. Therefore, I mapped previously characterised mutations of τ131 onto the TPR array structure (*Figure 2.17*). The mutations that increase Pol III transcription cluster mostly on TPR 2, those that decrease Pol III transcription spread over TPRs 6-9 and those that rescue a τ138 temperature-sensitive mutation, G349E (see *Introduction*), map to TPRs 7-8. Of particular interest were residues which, when mutated, rescued the τ138 temperature-sensitive mutation, as this could imply a physical contact between τ131 and τ138 at this region. Two of these residues – D468 and E498 – are both acidic and solvent-exposed, and so may make contact with the basic region of τ138 τIR. These residues occur on TPR 8 of the τ131 TPR array and so I decided to focus on this region in my next experiments.

Figure 2.16 **Sequence and structural alignment of the TPR array of τ131.** Residues are coloured according to the legend in *Figure 2.15*. A structural alignment using predictions of the *S. pombe* and *H. sapiens* sequences are displayed above the sequences. Other aligned species featured are: *Candida albicans, Gallus gallus and Arabidopsis thaliana*. Coloured asterisks (*) indicate highlighted mutants displayed in *Figure 2.17*.

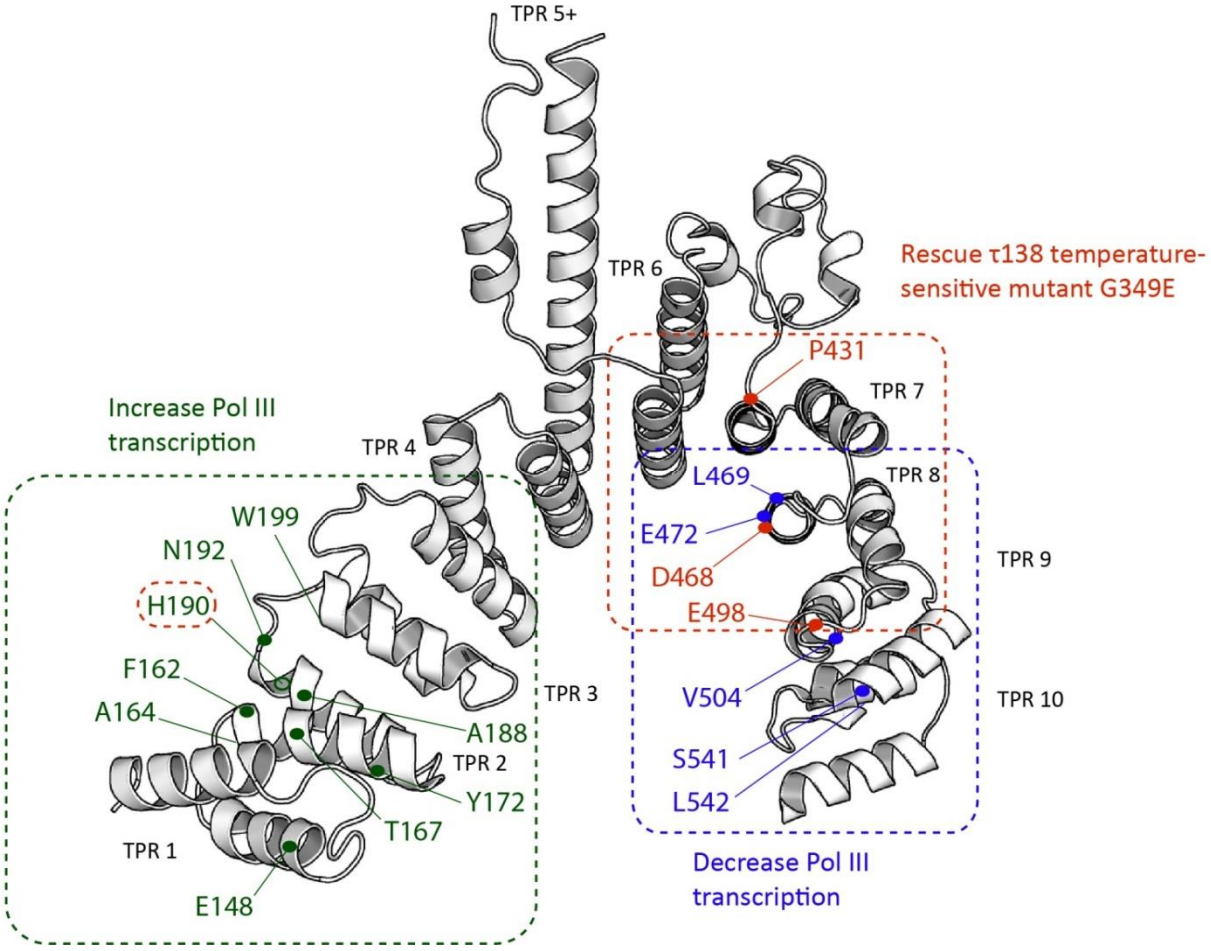

Figure 2.17 **Mapping known mutants onto the τ131 TPR array.** Boxed in green are residues associated with an increase in Pol III transcription. Boxed in red are residues associated with rescuing a temperature-sensitive τ138 mutant. Boxed in blue are residues associated with a decrease in Pol III transcription. Note that residue H190 was also shown to rescue the τ138 temperature-sensitive mutation.

2.4.2 Selecting and purifying τ131 TPR 8 point mutants

To further map the interaction site for the τ138 τIR, I generated five point mutants in TPR 8 of τ131 (D468K, E498K, L468K, E472K, and E497K). These residues were selected based on the analysis just described and are highlighted in *Figure 2.18*. The residues were all mutated to lysine to create a charge-swap in most cases. Mutagenesis cloning was carried out as detailed in *Materials and Methods, 6.1.10*. The untagged proteins were purified by the same protocol as used for the wildtype τ131 (123-566).

Analytical gel filtration was used as a quality control to ensure that the proteins remained folded correctly and were not aggregated (*Figure 2.19*).

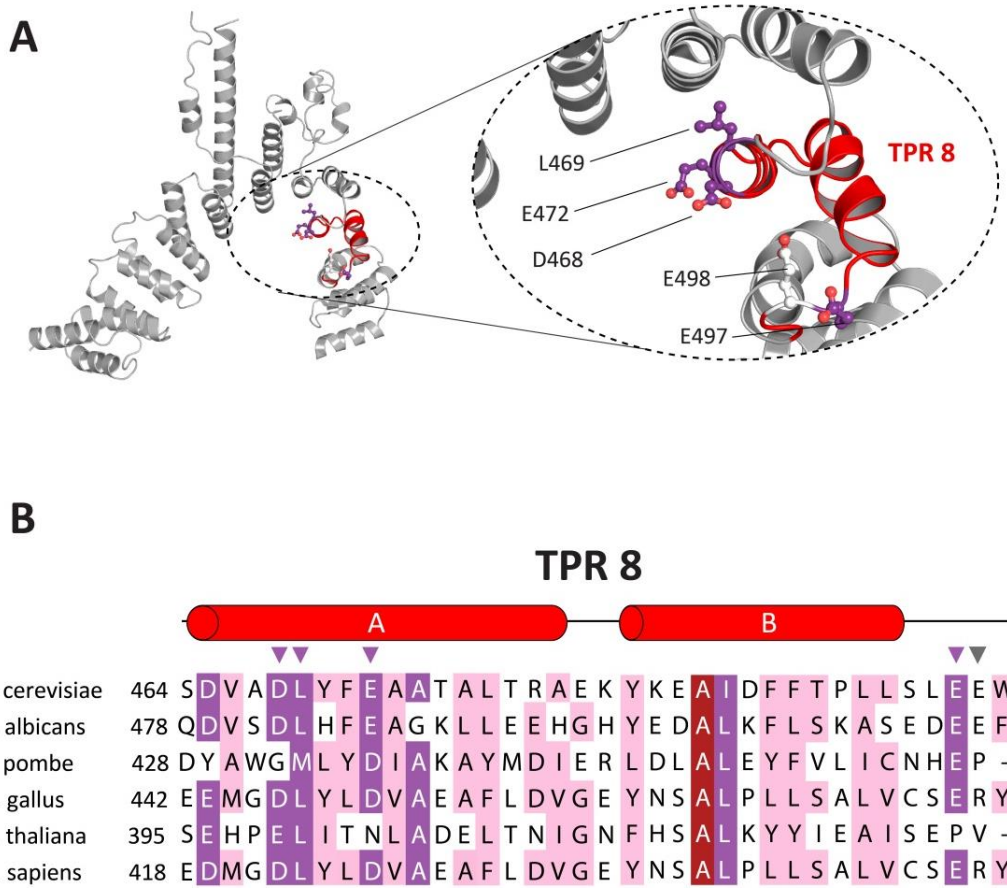

Figure 2.18 **Conservation of residues in TPR 8 of the τ131 TPR array.** (A) Close-up view of residues selected for mutation, located in TPR 8 (red) of the TPR array (grey). Residues D468, L469, E472, E497 and E498 are displayed as sticks and coloured according to conservation. These residues are highlighted on the sequence alignment in B by purple (D468, L469, E472, E497) or dark grey (E498) arrowheads. (B) Sequence alignment of TPR 8. Colour code for conservation is the same as in *Figure 2.15*.

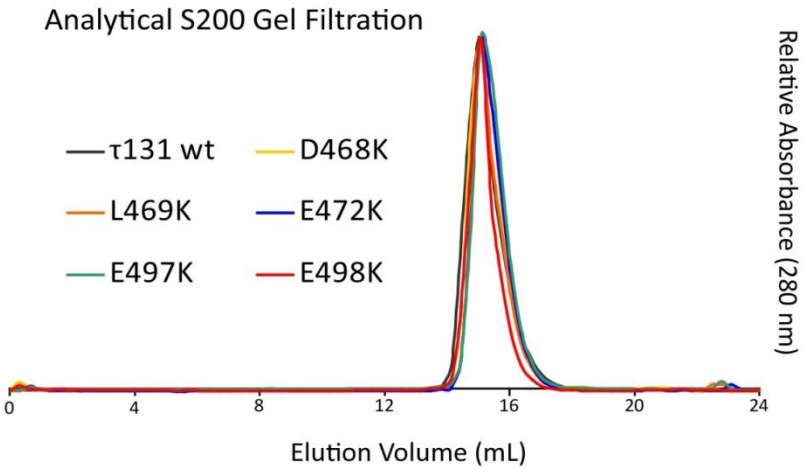

Figure 2.19 **Analytical gel filtration profiles of the TPR array of τ131 and various mutants.** The samples were injected onto a Superdex 200 10/300 GL column (GE Healthcare) and elution position was analysed. wt = wildtype.

2.4.3 Testing τ131 mutants by GST pull-down

The same assay that I had established for initially mapping the τ131-τ138 interaction was used to test the τ131 mutants for interaction with τ138. The GST-tagged protein used in the assay was τ138-τIR. *Figure 2.20* shows the results of the assay.

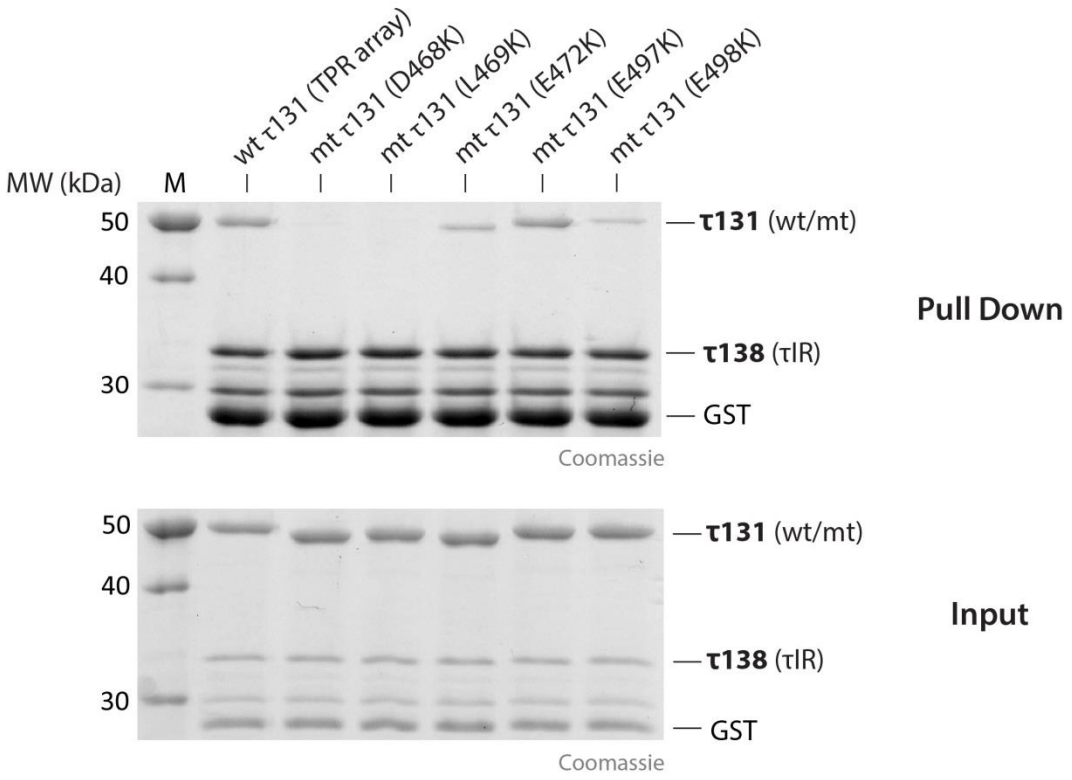

Figure 2.20 **Residues D468 and L469 are important for τ138 interaction.** GST pull-down assay. 12% SDS-PAGE of pull-down (top) and input (bottom) samples are shown. M = marker; wt = wildtype; mt = mutant.

The first lane is the positive control, showing the previously obtained result that τ138 τIR pulls down the τ131 TPR array. Mutants D468K and L469K are not pulled down by τ138-τIR. Mutants E472K and E498K are pulled down, but the intensity of the τ131 band is weaker than for the wildtype. The E497K mutant appears to be pulled down as efficiently as the wildtype. It can be concluded from this assay that the D468 and L469 residues are very important for the τ138 interaction, as the binding is abolished when these two residues are mutated.

2.4.4 Testing τ131 mutants by ITC

I again used ITC in order to quantitatively analyse the interaction between the τ131 mutants and τ138. I tested both the τ138 τIR and the τ138 winged helix + τIR (546-693)

by ITC for binding to the τ131 mutants. The results were again consistent with the results I observed for the GST pull-down. The ITC results are shown in *Figure 2.21* and a summary of binding affinities is displayed in *Table 2.2.*

Table 2.2 **Summary of Kd values from ITC experiments between τ131 TPR mutants and τ138**

τ131 TPR array	τ138 (τIR)	τ138 (WH+τIR)
wildtype	80 nM	100 nM
D468K	N.B.	N.B.
L469K	N.B.	N.B.
E472K	833 nM	415 nM
E497K	89 nM	99 nM
E498K	833 nM	971 nM

N.B. = No binding

Again, the D468K and L469K mutations lead to a loss of binding with τ131. Mutations E472K and E498K cause an approximate 10-fold reduction in binding, whilst mutant E497K shows almost identical binding affinity as the wildtype. In summary, the findings from ITC experiments fully confirm the observations from the GST pull-down assay. The results suggest that the τ138 τIR binds, at least in part, to residues in TPR 8 of the τ131 TPR array.

2.5 Investigating τ131-Bdp1 interaction.

Unexpectedly, the binding site for τ138 overlapped with a reported binding site for Bdp1, an essential subunit of TFIIIB. Bdp1 is hypothesised to form contacts with both arms of the τ131 TPR array, competing with another TFIIIB subunit, Brf1 [109]. The mutation D468N causes a reduced two-hybrid response to Bdp1 [104] whilst L469K decreases the interaction with Bdp1 *in vitro* and *in vivo* [109]. I decided to test by GST pull-down assay whether the TPR array could interact with Bdp1 full-length protein, and whether this interaction could be disrupted by the τ131 mutants I had already characterised for the τ138 interaction.

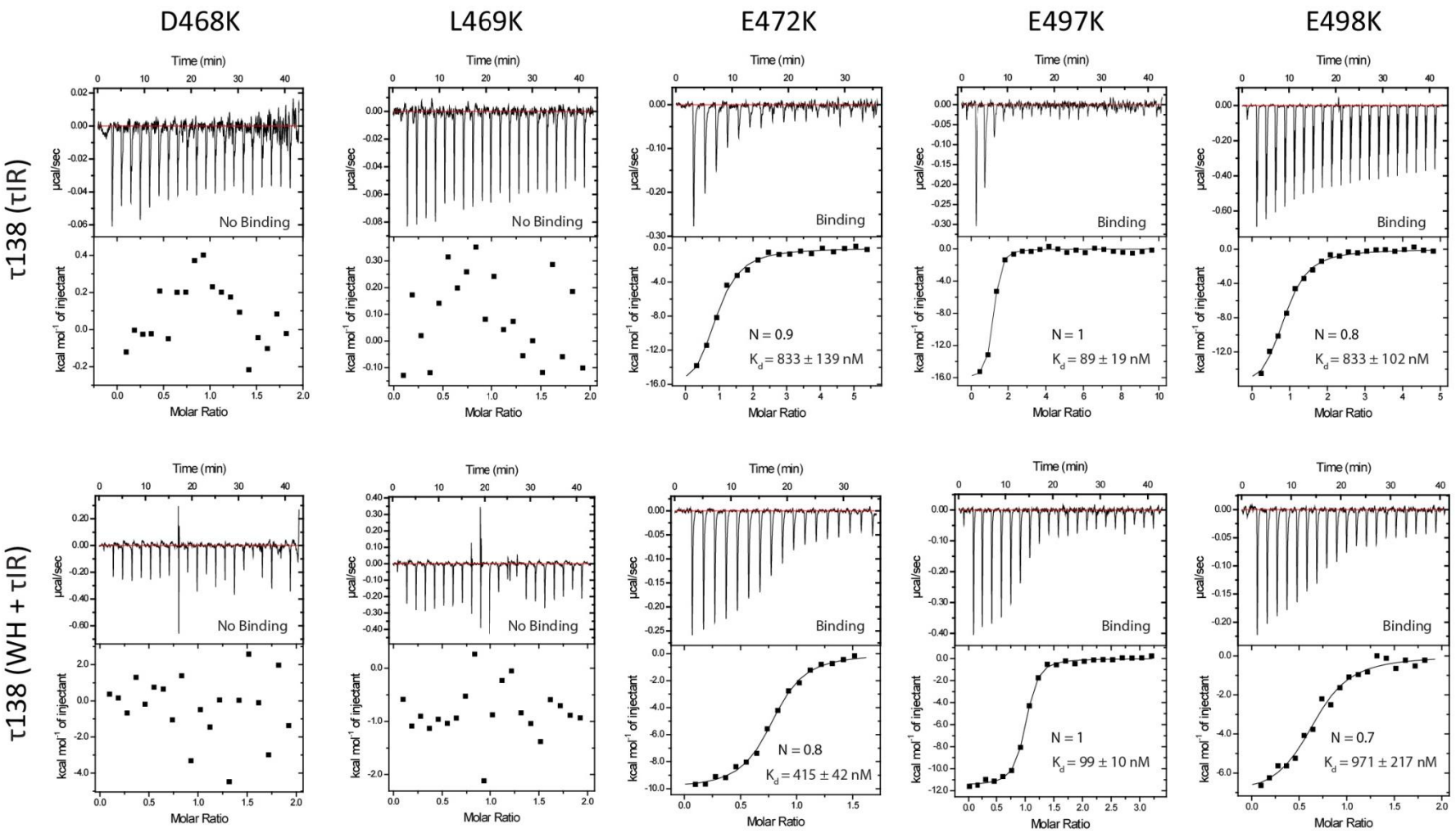

Figure 2.21 **ITC confirms that residues D468 and L469 are important for τ138 interaction**. ITC data that is summarised in the table above. N = binding stoichiometry between τ131 and τ138; K_d = dissociation constant.

2.5.1 Testing τ131-Bdp1 interaction by GST pull-down

To test for τ131-Bdp1 interaction, I used full-length Bdp1 protein, kindly provided by Niklas Hoffmann, a PhD student in our lab. I first tested the interaction between the τ131 TPR array and Bdp1 by GST pull-down.

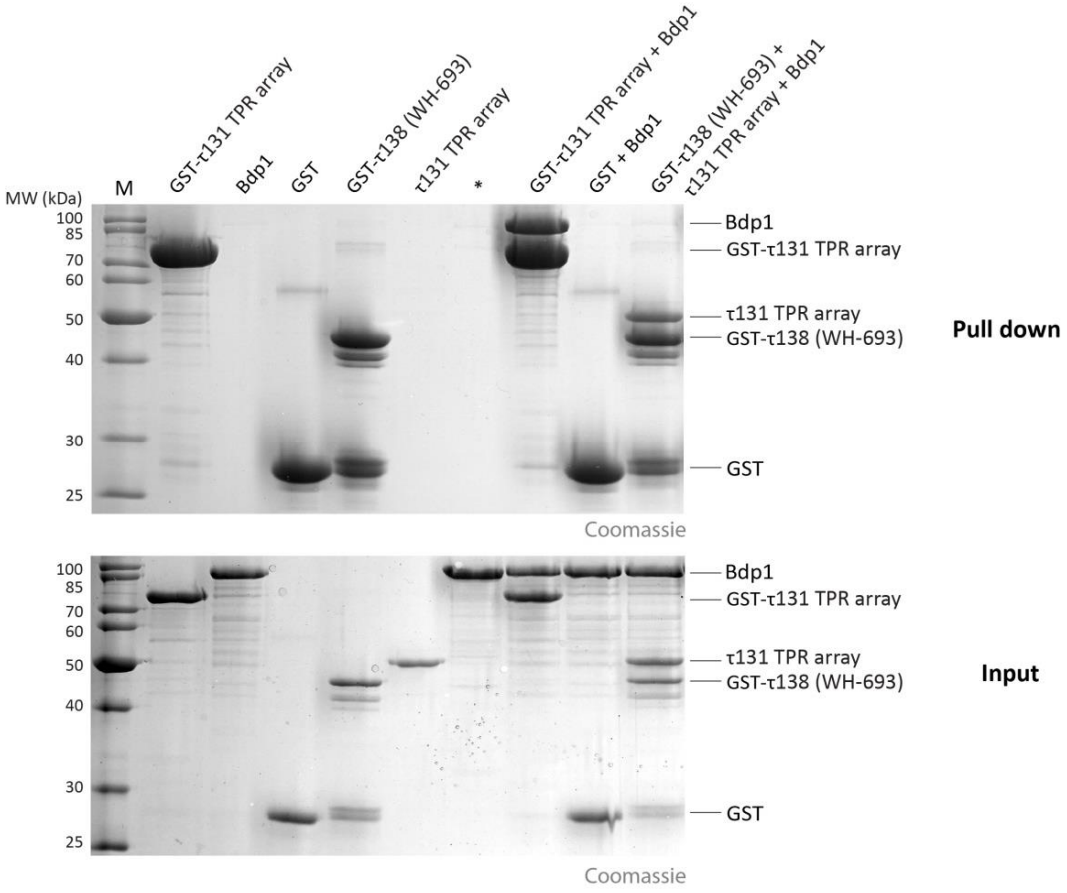

Figure 2.22 **τ131 binds full-length Bdp1 but not in the presence of τ138 (WH-693).** GST pull-down assay to investigate τ131-Bdp1 interaction. 12% SDS-PAGE pull down and input gels. * = sample investigated for different experiment; M = marker.

The τ131 TPR array, fused to GST, pulled down full-length Bdp1 (*Figure 2.22*). This result is expected based on previous literature but is the first time, to the authors' knowledge, that this has been demonstrated using recombinant, purified proteins. Bdp1 alone did not bind to the GST-affinity resin, or to GST acting as negative controls.

I then checked if the τ138 τIR could out-compete the Bdp1 for binding to the TPR array. To this end, I incubated GST-τ138 (WH-693), τ131 (123-566) and full-length Bdp1 together with the GST-affinity resin and performed the pull-down experiment as before. *Figure 2.22* shows that the τ138 pulls down τ131 as expected but with no pull-

down of Bdp1. This suggests that the τ138 τIR can indeed out-compete Bdp1 for binding to the τ131 TPR array.

2.5.2 Overlapping binding sites of τ131, Bdp1 and τ138

I next tested whether the residues D468 and L469 were significant for the interaction between τ131 and Bdp1, as I had shown for the interaction between τ131 and τ138. To this end, I repeated the GST pull-down assay but this time tested the GST-tagged Bdp1 for binding with τ131 D468K and L469K. The results are shown in *Figure 2.23*.

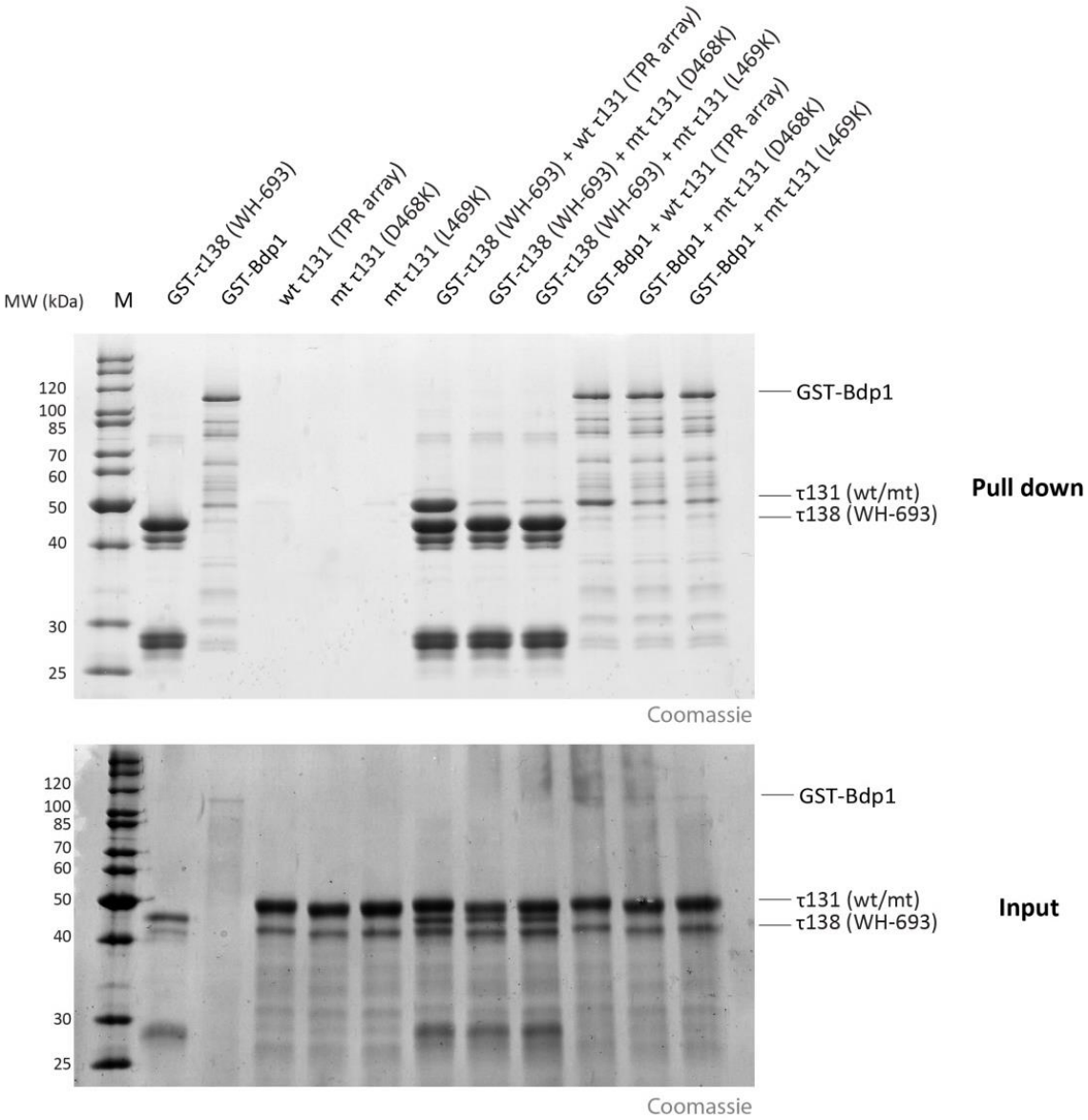

Figure 2.23 **The τ131-Bdp1 interaction is also be affected by mutants D468K and L469K.** GST pull-down assay to investigate τ131 point mutants-Bdp1 interaction. 12% SDS-PAGE pull down and input gels. τ138 (WH-693) proteins are included for comparison. M = marker.

One problem with this experiment is the stability of the GST-Bdp1 construct. A degradation band which is visible at ~50 kDa overlaps with the position of the τ131 band, thus interfering with the conclusion from this experiment. However, it would appear that the two point mutants (D468K and L469K) reduce the interaction between Bdp1 and τ131, owing to the fainter bands present at the τ131 position for the two mutants. This mirrors the situation observed for τ138, suggesting overlapping binding sites on the τ131 TPR array for τ138 and Bdp1.

2.6 Discussion

This chapter has detailed my findings that the TPR 8 of τ131 specifically interacts with a 50 amino acid region of τ138, called 'τIR'. My findings identified the first direct interaction site between τA and τB. Interestingly, this site overlaps with τ131's binding site for the TFIIIB subunit Bdp1.

2.6.1 The TPR array of τ131 links τA and τB

TFIIIC is described in the literature as a 'dumb-bell' shaped molecule formed of two DNA-binding subcomplexes called τA and τB. These two subcomplexes are proposed to be connected by a flexible linker. A low resolution view of this linker was obtained by early scanning electron microscopy images and supported by proteolysis studies [68, 75]. To identify the subunits involved in this linker, I first used chemical cross-linking studies coupled with mass spectrometry. These results provided a clear indication that τ131 and τ138 were in close proximity to each other. This was consistent with a previous study that identified point mutations within the τ131 TPR array that suppress the temperature-sensitive τ138 mutation, G349E.

I have shown that the highly conserved TPR array within τ131 binds and stabilises a central region of τ138 I call 'τIR'. It is unclear how τIR interacts at the molecular level with the TPR array. It may form an extended interface as with other TPR-'peptide' structures [149], possibly wrapping around both the convex and concave surface of the TPR array to exploit the larger surface area. In the absence of a crystal structure of the complex between the TPR array and the τ138 τIR, this remains an open question. Despite extensive efforts, I was not able to locate additional density for the

τ138 τIR within the 'complex' crystals. A crystal of the complex was run on SDS-PAGE to ensure that the τ138 τIR was present in the crystals (*Figure 2.24*).

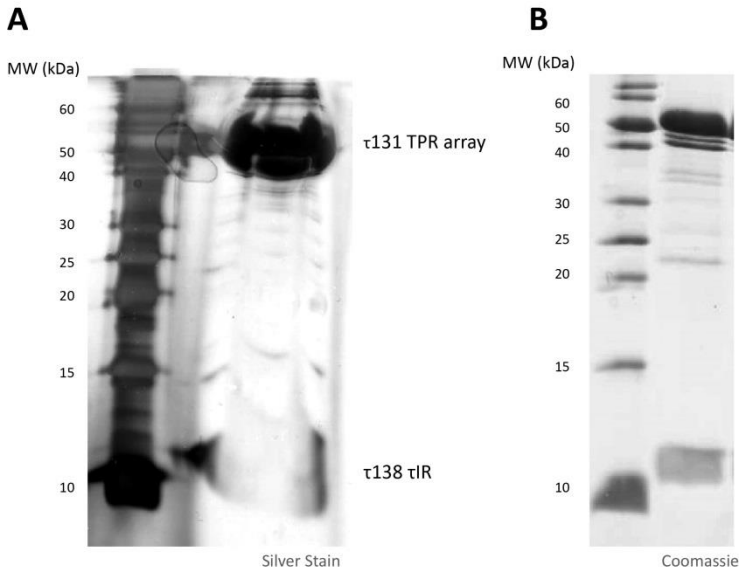

Figure 2.24 **Confirmation that the τ138 τIR is present in the complex crystals**. (A) A complex crystal was dissolved in 5x SDS buffer and ran on an 18% SDS-PAGE. A smeared band at ~10kDa indicates the presence of the τ138 τIR. (B) A lane from *Figure 2.10B* showing the purified complex as a comparison.

With the gel indicating the presence of the τ138 τIR, what explanations are there for the lack of electron density in the crystal lattice? One explanation could be that the observed precipitate surrounding the crystals is in fact the denatured τ138 τIR and that the protein degrades over time, although this is not indicated by the gel in *Figure 2.24*. Another explanation could be that within the crystal lattice there is not a sufficient number of τ131 molecules bound to the τ138 τIR, and that the two components should instead be purified separately and co-crystallised with a molar excess of the τ138 τIR to increase the likelihood of observing electron density for this part of the complex.

Despite the lack of structural information for this interaction, I showed that the τ138 τIR binds to TPR 8 of τ131 and I have identified key residues involved in this interaction. As the τIR comprises 50 amino acids and as it appears that the entire length is required for high affinity-affinity binding, it would seem reasonable that the τIR also binds at other locations within the TPR array. It was surprising to observe a complete loss of binding upon mutation of either residue D468 or L469 to lysine. These highly conserved residues occur on TPR 8 (*Figure 2.18*): D468 points into the inner groove, whilst L469 makes hydrophobic contacts with the A helix of TPR 7. Both

residues are located close to the 'hinge' between the two arms of the TPR array and so it might be possible that mutation of these residues leads to a change in the positioning of the two arms relative to each other. This could explain why they are important for the interaction with τ138 and also Bdp1 (see below), although direct interaction through these residues is still a possibility.

A more extensive mutational analysis could be carried out, firstly *in vitro*, in order to assess where else the τ138 τIR may bind on the TPR array. However, this task would be made far easier by a structure-guided mutagenesis approach. As the τ131-τ138 complex does not appear to be amenable to studying by X-ray crystallography in *S. cerevisiae*, studying the interaction in other species may represent a possible future direction.

2.6.2 Conservation of the link between τA and τB

As highlighted in the *Introduction* there are a number of apparent differences between TFIIIC in *S. cerevisiae* and the TFIIIC of other eukaryotes, such as humans. The divergence of subunits between the two is most pronounced within the subunits of τB. Whilst τA subunits show high sequence and structural conservation the τB subunits are less conserved, particularly τ138. *Figure 2.5* shows that the τ138 orthologue in humans (TFIIIC220) is in fact double the size and poorly conserved on an overall sequence level. There is however some structural conservation, particularly at the N-terminus, which appears to contain three winged helix domains.

The TPR array of τ131 is highly conserved, with the overall architecture of the TPR array apparently required from yeast to human (see *Figure 2.4*, *Figure 2.15* and *Figure 2.16*). On the other hand, the τIR of τ138 is not well conserved. There is some sequence conservation, when comparing *S. cerevisiae* with *S. pombe* (*Figure 2.7C*), but the τIR aligns very poorly with sequences from other eukaryotes, such as human. What is clear is that the region after this winged helix is relatively disordered in all species, however the size and sequence of this region varies.

As suggested in *Section 2.6.1*, one could attempt to produce equivalent constructs in other species such as *S. pombe* or *Chaetomium thermophilum* and test if a complex is still formed. This would probably require new construct screening projects

for homologues from higher eukaryotes, as the τIR region of τ138 is not very well conserved and so predicting and delineating boundaries would prove difficult.

2.6.3 Do τ131 and τ138 form the only link between τA and τB?

Although I have shown that τ131 and τ138 form a link between τA and τB, other TFIIIC subunits may be involved as well. Previous work has cited τ60 and τ95 as making links between τA and τB [90, 101]. However, no cross-links were observed between τ60 and τA subunits in the experiments I presented in this chapter. In contrast, there were cross-links between τA component τ95 and τB component τ138, which supports a role for the involvement of this subunit in linking τA and τB. A previous study has shown binding of τ95 to τ138 by co-immunoprecipitation, and described a point mutant within τ95 (E447K) that reduces the affinity of TFIIIC towards tDNA [90]. The τ95 subunit may bind to τ138 directly, or be brought into close proximity to τ138 by the interface with the τ131 TPR array, as τ95 has been shown to bind this TPR array in humans [88] and in *S. pombe* in unpublished data from our lab.

Although no other cross-links were observed between τA and τB subunits the participation of other subunits, or indeed the already characterised subunits at additional sites, cannot be ruled out at this point. An absence of cross-linking may be due to low accessibility of the cross-linker to a tightly buried interface (such as the WD40s of τ60-τ91, where no cross-links are observed) or the absence of required lysine residues in the respective interface, and so there may be other regions within TFIIIC which do indeed interact but are undetectable by this method.

2.6.4 The overlapping binding sites of τ138 and Bdp1 on τ131

The TPR array of τ131 and its crucial role in the assembly of TFIIIB has been the subject of extensive studies. It was intriguing that the residues that I identified as being crucial for the τ131-τ138 interaction had already been implicated in Brf1 and Bdp1 binding [104, 109], suggesting that the binding sites for these subunits overlap on the TPR array. I have presented preliminary pull-down analysis comparing the effect of τ131-Bdp1 interaction with τ131-τ138 interaction. It is clear that the TPR array provides a platform over which Bdp1 binds stoichiometrically, which is consistent with previous studies. In addition, the pull-down suggests that Bdp1 binding is affected by the

mutations D468K and L469K *in vitro* and *in vivo*. For this last experiment it is difficult to make a firm conclusion however, as the position of the τ131 band overlaps with a degradation band of Bdp1. Bdp1 is predicted to be a largely disordered protein and producing recombinant full-length protein has proved difficult. Nevertheless, these initial findings are consistent with the literature and are supported by the experiment shown in *Figure 2.22*. Here, incubation of Bdp1 and τ131 with GST-τ138 led to only the pull-down of τ131 suggesting that Bdp1 is outcompeted for binding when τ138 is present. Further work will be required in order to investigate precisely the regulatory mechanisms involved.

In summary, I have demonstrated that τA subunit τ131 uses its TPR array to link τA with τB, binding with high affinity to a disordered region C-terminal to a conserved winged helix domain in τ138. This is the first structurally and biochemically mapped interaction site between a τA and a τB component of TFIIIC, which in addition overlaps with a TFIIIB binding site.

3. Results - Structural and Functional Studies on τ138

From the project described in the last chapter, a region of τ138 was identified that formed a high-affinity interaction with τ131. N-terminal to this region was a domain that was predicted to be a winged helix (546-641). I set out to crystallise this fragment and learn more about the role of τ138 within TFIIIC.

3.1 Solving the structure of τ138 (546-641)

3.1.1 Purification of τ138 (546-641)

The reader is referred to *Materials and Methods* for full details of cloning, expression and purification protocols. A pETM30 vector encoding the sequence for τ138 (546-641) was expressed in BL21 (DE3) Gold *E. coli* cells. The protein was purified by means of an N-terminal, TEV protease-cleavable 6x-histidine tag adjacent to a GST-tag. After elution from nickel resin, the 6x-histidine-GST tag was cleaved by TEV protease and the cleaved protein was collected. The protein was injected onto a S75 Superdex preparative gel filtration column (GE Healthcare) and eluted at an expected position for a monomer and as a single species (*Figure 3.1A*). Analytical ultracentrifugation (AUC) was carried out in collaboration with Vladimir Rybin at EMBL Heidelberg, again confirming a single species at the expected molecular weight of ~11 kDa (*Figure 3.1B*). The protein was evaluated by SDS-PAGE and judged to be appropriate for crystallisation trials (*Figure 3.1A*).

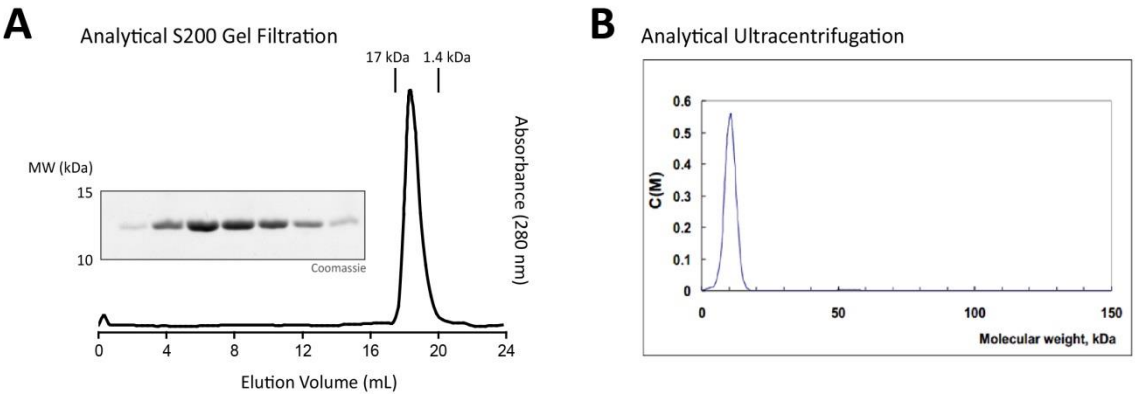

Figure 3.1 **Biophysical characterisation of τ138 (546-641).** (A) Analytical gel filtration profile of τ138 (546-641). The sample was injected onto a Superdex 200 10/300 GL column (GE Healthcare) and elution position analysed. Positions of known molecular weight standards are indicated at 17 kDa and 1.4 kDa. Inset, 15% SDS-PAGE showing elution peak. (B) Analytical ultracentrifugation profile of τ138 (546-641).

3.1.2 Crystallisation of τ138 (546-641)

The protein was concentrated to 30 mg/mL and used for initial crystallisation trials. Crystals were grown at 20°C using the sitting drop method in an initial broad screening approach. Initial crystals were obtained in the JCSG+ screen (Molecular Dimensions) with a reservoir solution of 1 M sodium citrate pH 6.5 and 0.1 M sodium cacodylate (*Figure 3.2A*). Manual refinement screens were set up around this condition with the best hits occurring in 1.15 M sodium citrate pH 6.2 and 0.1 M sodium cacodylate (*Figure 3.2B*). Crystals were cryo-protected by soaking in mother liquor supplemented with 15% glycerol for a few seconds before being flash-frozen in liquid nitrogen.

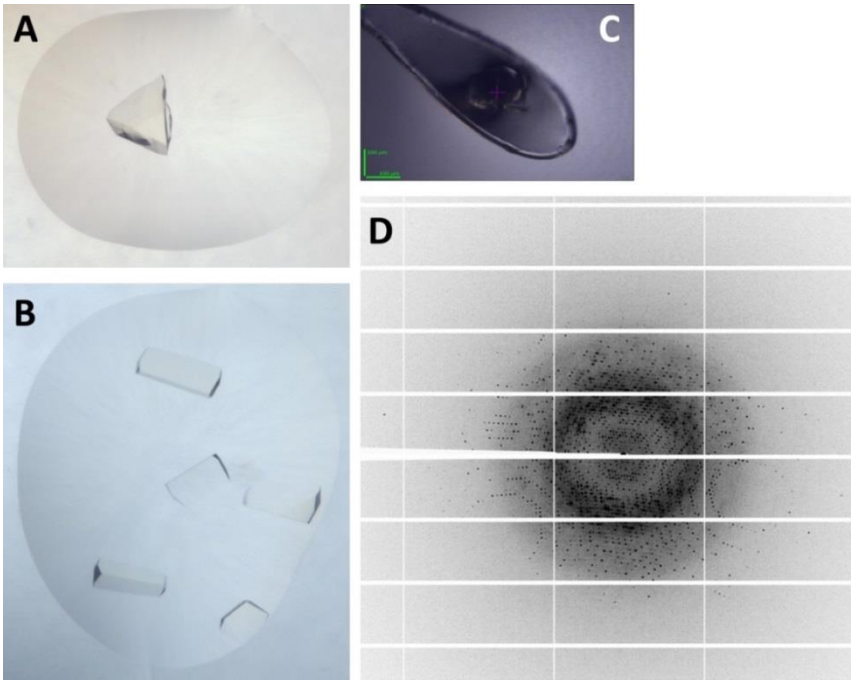

Figure 3.2 **Crystallisation and diffraction of τ138 (546-641).** (A) Initial crystal obtained for τ138 (546-641). (B) Refined crystals that were used for data collection. (C) Collected crystal in a loop for data collection. (D) Example of a diffraction image taken for the crystal collected in C.

3.1.3 Solving the structure of τ138 (546-641) by sulphur-SAD

The crystals diffracted to ~1.9 Å on the rotating anode X-ray home source at the EMBL Heidelberg. As initial molecular replacement solutions were not successful, an uncommon approach to structure solving was tried, namely sulphur-SAD. This technique for solving structures was first demonstrated in 1981 [150]. Sulphur-SAD relies on the very small but useable anomalous signal from sulphur atoms that are present in methionine or cysteine residues of proteins. The anomalous signal arises

from the excitation of specific electrons of the sulphur atom, which occurs at a so-called 'absorption edge' (different elements possess different edges). This edge, where a sharp increase in absorption of X-rays by the electrons of a specific element occurs, is accessed at specific X-ray wavelengths. The wavelength of X-rays produced by a copper anode (present in our home source) is relatively close to the sulphur edge. The signal produced is very small, especially when compared to the anomalous signal from selenium that is recorded in selenomethionine substituted protein crystals (*Figure 3.3*).

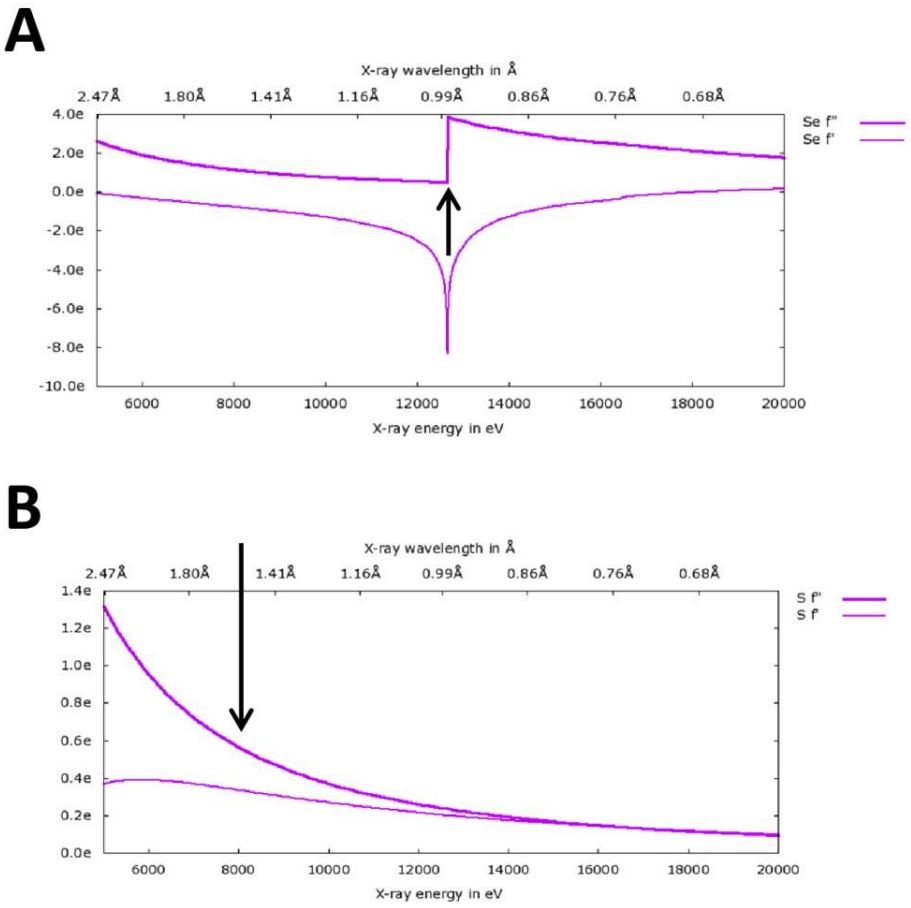

Figure 3.3 **The anomalous signal from the sulphur-edge is very small when compared to the selenium-edge**. (A) Edge plot for selenium. The so-called absorption edge can be easily appreciated, indicated by the arrow. (B) Edge plot for sulphur. The arrow indicates the wavelength at which experiments were performed. F'' (bold magenta line) represents anomalous differences, f' (light magenta line) represents dispersive differences. Both differences contribute to the overall scattering by the atom. e = electrons. Edge plots were generated using the webserver: http://skuld.bmsc.washington.edu/cgi-bin/edgeplots.

A highly redundant data collection (540 images, 1° rotation) was carried out – essential for accurately recording the small anomalous signal – and data was processed using XDS [144] (see *Materials and Methods, 6.4* for further details on data processing).

Table 3.1 **τ138 (546-641) Data collection, phasing and refinement statistics**

	τ138 (546-641) Sulphur-SAD	**τ 138 (546-641) Native**
Data collection[1]		
X-ray source	Rotating anode (EMBL)	ID23-1 (ESRF)
Space group	R32	R32
Cell dimensions *a*, *b*, *c* (Å)	129.05, 129.05, 68.11	129.09, 129.09, 68.04
Wavelength (Å)	1.54180	0.97630
Resolution (Å)	50-1.86 (1.91-1.86)	50-1.40 (1.44-1.40)
CC 50	1.0 (0.92)	0.99 (0.65)
R_{merge} (%)	5.9 (63.7)	5.1 (137.9)
I/σI	40.31 (4.57)	21.31 (1.73)
Completeness (%)	99.9 (99.6)	99.9 (99.3)
Redundancy	15.6 (13.6)	10.1 (9.4)
Sites	6 S	-
Refinement		
Resolution (Å)		43.2 – 1.40
No. reflections		42552
R_{work}/ R_{free} (%)		16.3/19.4
No. atoms		
Protein		1545
Water		246
Ligand		12
B-factors (Å^2)		
Protein		27.3
Water		37.1
Ligand		57.8
R.m.s deviations		
Bond lengths (Å)		0.017
Bond angles (°)		1.656

[1] Values in parentheses correspond to the highest-resolution shell

The statistics of the integrated reflection data showed that the data was highly precise (as judged by low R_{merge} values) and that the overall resolution extended to 1.9 Å carrying reliable anomalous signal up to 3.4 Å. The structure was solved by SAD combined with density modification using the program autoSHARP [148], with the program able to locate the positions of 6 sulphur atoms, indicating 2 molecules in the asymmetric unit (*Figure 3.4B*). The two molecules of the asymmetric unit are related by a non-crystallographic dyad.

An initial build using AutoBuild [151] in the Phenix suite [147] was followed by iterative rounds of real-space model building in Coot [146] and reciprocal-space refinement in Phenix. Water molecules could be modelled owing to the high-resolution density. The resolution was later extended to 1.4 Å using a native dataset collected at beamline ID23 at the ESRF (*Figure 3.2D*). The quality of this high-resolution electron

density is highlighted in *Figure 3.4C*. The final model could be refined to R_{work}/R_{free} values of 16.3%/19.4% with good stereochemistry (*Table 3.1*).

3.1.4 τ138 (546-641) is an extended winged helix domain

The crystal structure revealed that this domain is in fact an 'extended' winged helix (eWH). The canonical winged helix fold (α1, β1, α2, α3, β2, β3) is observed but an additional C-terminal alpha-helix (α4) is present (*Figure 3.4A*). The name 'extended' winged helix was used originally by Meinhart et al. [51]. In that example, the canonical winged helix of TFIIE was extended both at the N-terminus and the C-terminus.

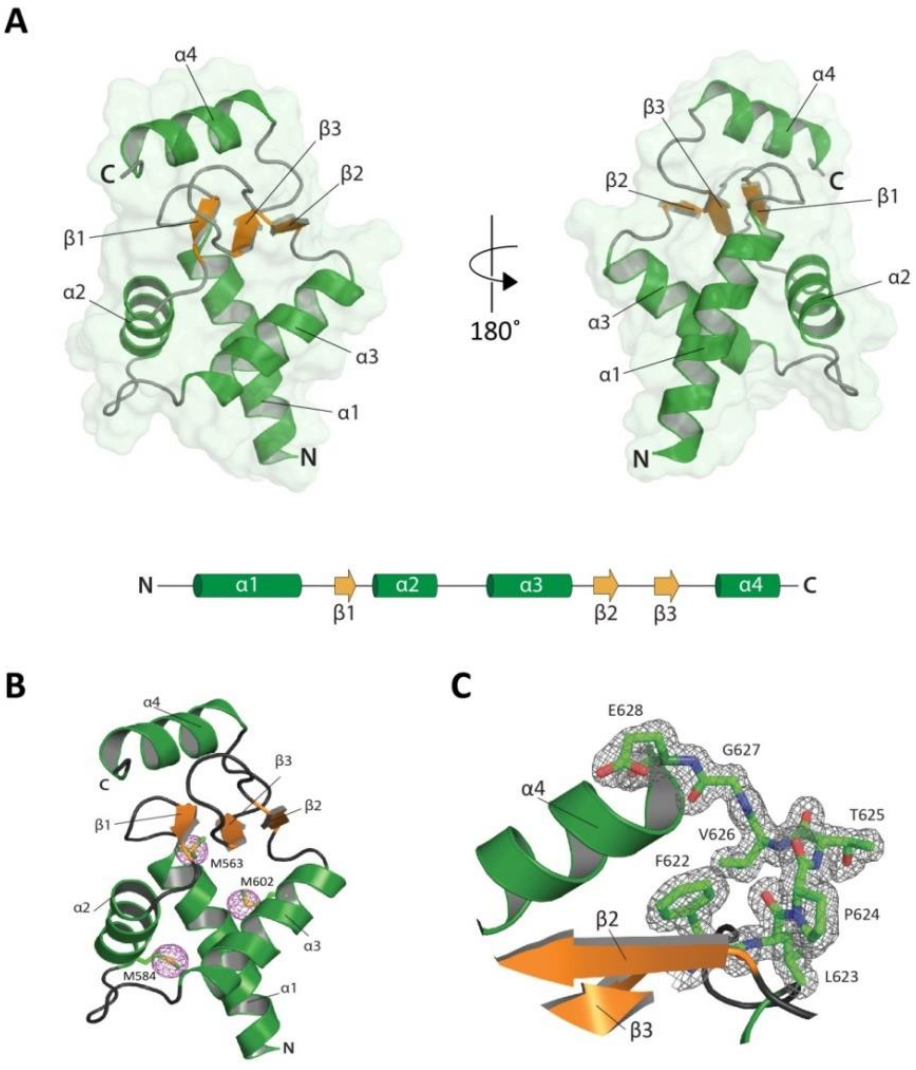

Figure 3.4 **Structure of the central eWH domain (546-641) of τ138.** (A) Two views of the eWH domain are displayed in ribbon representation with a transparent surface representation superimposed. A schematic of the arrangement of alpha-helices and beta strands is displayed underneath. Cylinders depict α-helices, arrows depict β-strands. (B) The eWH domain with sulphur electron density peaks highlighted in magenta at the three methionine positions. (C) A close-up of the loop between β3 and α4 highlights the excellent electron density. 2mFo-DFc map contoured at 1σ.

3.2 Functional analysis of the central eWH domain of τ138

A DALI sever search [152] revealed many winged helix domains, but none that superimposed well in terms of the extended α4 helix. I thus investigated the structure in more depth to provide clues of a potential function.

3.2.1 Surface conservation and electrostatics of the eWH domain

The structure is relatively well conserved at a sequence level from yeast to human (*Figure 3.5*). Moreover, the eWH fold is almost completely preserved when comparing the structure prediction of *S. cerevisiae, S. pombe and H. sapiens* (*Figure 3.5*).

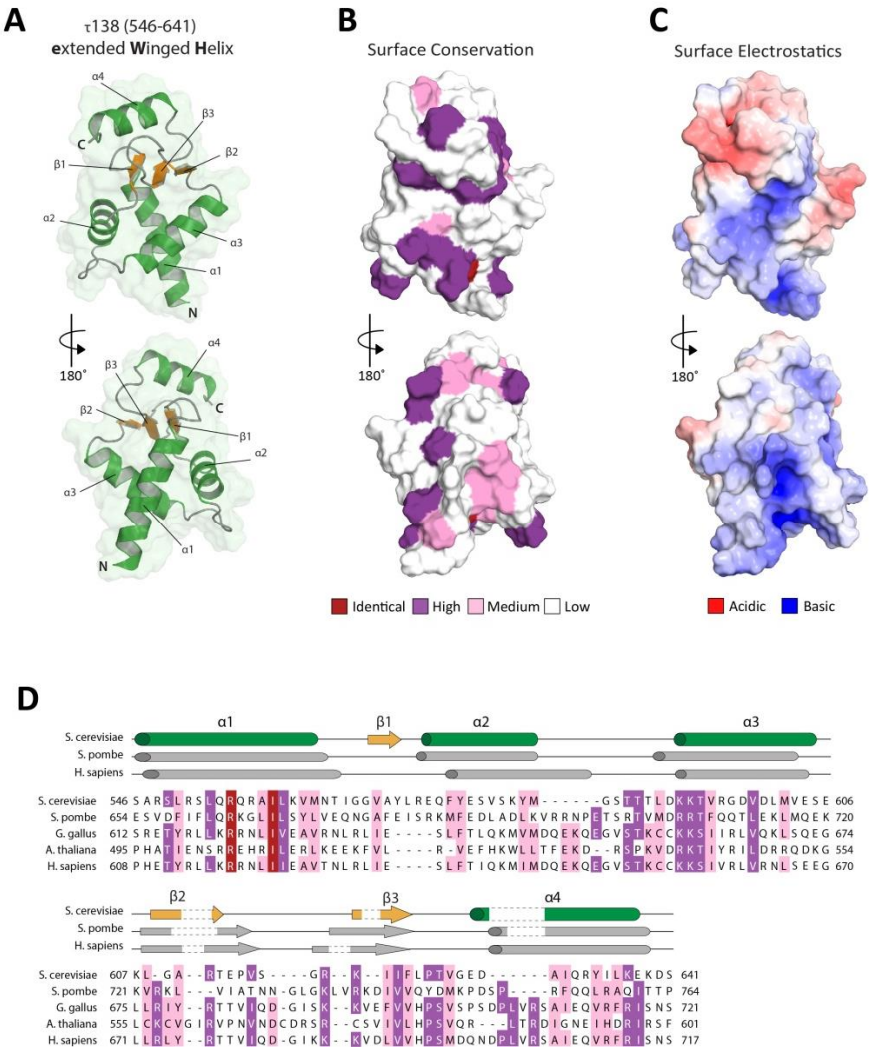

Figure 3.5 **eWH domain surface conservation and electrostatics.** (A) View reproduced from *Figure 3.4*. (B) Surface conservation representation. Conserved residues are coloured as shown in the legend. (C) Surface electrostatics coloured as shown in legend. (D) Sequence alignment of the eWH domains of five selected species. Residues are coloured according to the legend in B. Structural alignment using predictions of the *S. pombe* and *H. sapiens* sequences are displayed above the sequences. Cylinders depict α-helices, arrows depict β-strands.

In terms of surface conservation, there are patches of conserved residues on both faces of the eWH domain (*Figure 3.5B*). Of particular interest are the conserved basic residues at the start of α3 (the so called 'recognition-helix') and the 'wing' between β1 and β2. These sites are often associated with binding nucleic acids. The eWH domain also contains a large acidic patch at the top of the molecule (*Figure 3.5C*), in part from contribution of the additional helix α4. However, this appears to be non-conserved.

3.2.2 Testing for DNA-binding of the eWH domain

With several basic patches present on the surface of the eWH domain I next decided to test the eWH domain for nucleic acid binding using electrophoretic mobility shift assay (EMSA). I used a radioactively labelled (^{32}P) oligonucleotide that was 66 bases in length and represented the sequence of a tDNA encoding tRNA^{Glu}, see *Materials and Methods*, *6.2.4* for details. This oligonucleotide had been shown to bind holo-TFIIIC with moderately high affinity and thus appeared suitable for testing fragments of τ138 (*Figure 3.12*). The eWH domain was tested for binding to this single-stranded (ss) and double-stranded (ds) oligonucleotide at varying concentrations (*Figure 3.6*).

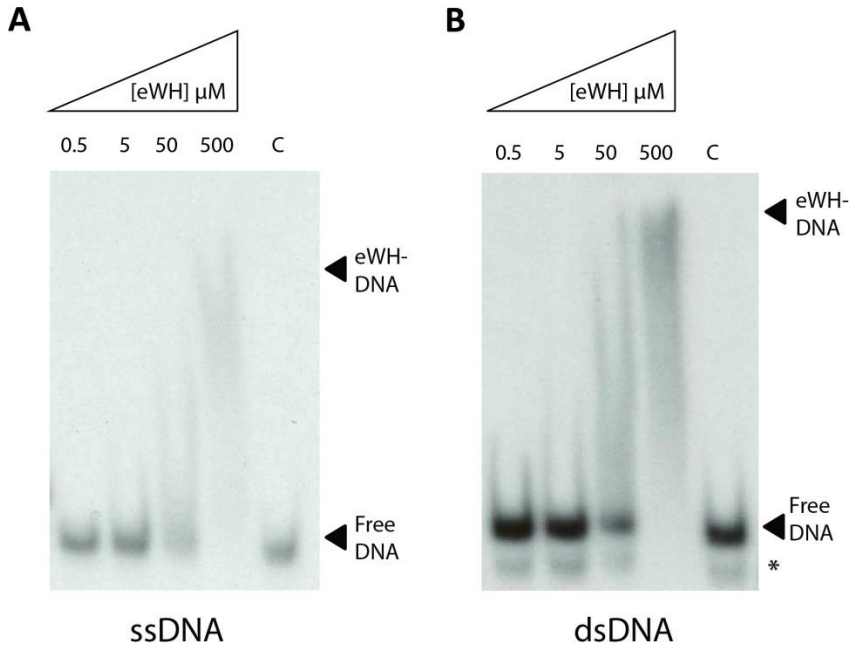

Figure 3.6 **The eWH domain does not bind specifically to single-stranded (ss) or double-stranded (ds) tDNA^{Glu}.** (A) EMSA experiment of eWH domain binding to a ssDNA 66bp tDNA^{Glu} oligonucleotide. (B) EMSA experiment of eWH domain binding to a dsDNA 66bp tDNA^{Glu} oligonucleotide. Top arrows indicate the shifted eWH-DNA complex and lower arrows indicate free DNA. * = excess of single-stranded DNA after annealing. C = control experiment (no protein added).

Binding to both ssDNA and dsDNA was not observed until a protein concentration of 500 μM was reached, indicating non-specific binding of the eWH domain to DNA. I therefore concluded that the central eWH domain of τ138 does not specifically bind to tDNA.

In summary, I solved the structure of a conserved central extended winged helix domain of τ138. In the former chapter I described that this domain was not involved in the τ131 TPR array interaction. Here, I have shown that this domain is also not involved in binding DNA. Curious as to which part of τ138 is responsible for B box binding, I set out to generate further constructs of τ138 for structural and functional studies.

3.3 Construct mapping of τ138

Full-length τ138 from *S. cerevisiae* could be expressed in *E. coli* and insect cells but was mostly insoluble and rapidly degraded after lysis. Consequently, I turned to designing smaller fragments of τ138 for structural and functional studies in *E. coli* with the aim of identifying the B box binding region.

3.3.1 Secondary structure prediction and alignment of τ138

By comparing various structure prediction servers (e.g. Phyre2 [153] , HHPred [154] and GeneSilico [155]) and alignment of orthologues found in other species, a consensus for secondary structure elements for τ138 was reached. *Figure 3.7* highlights predicted secondary structure elements for *S. cerevisiae* τ138. As highlighted in the previous chapter, τ138 is predicted to contain many winged helix domains, an arrangement which appears to be conserved from yeast to human, particularly for the three N-terminal winged helix domains and the central eWH domain (*Figure 2.5*). HMG box domains are also predicted at the extreme N and C-terminal ends of the protein, whilst the region after the disordered τIR (641-693) is largely helical. HMG-box domains, as with winged helix domains, are implicated in binding and bending DNA, either sequence specifically or non-specifically [95, 156]. Both domains have also been shown to mediate protein-protein interactions.

I designed constructs to 'scan' structured regions of τ138, attempting larger constructs in the first instance before focussing on smaller regions that would encompass isolated or tandem domains. Particular attention was paid to the

boundaries of these predicted elements when designing constructs. I ensured that constructs had small hydrophilic residues, usually serine residues, at the termini whenever possible to minimise risk of aggregation.

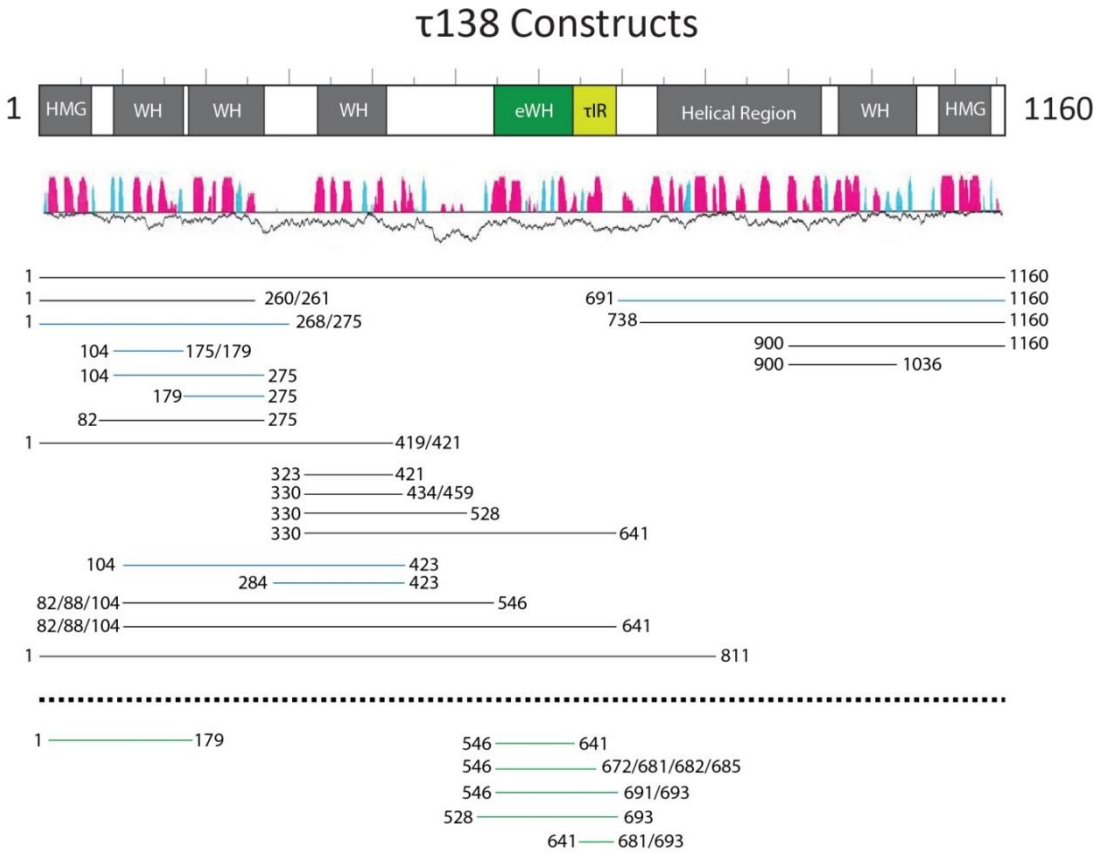

Figure 3.7 **Predicted secondary structure elements of τ138**. The upper bar represents the τ138 sequence, with predicted secondary structure elements highlighted. HMG = high mobility group; WH = winged helix; τIR = τ131-Interacting Region. The central eWH domain and τIR are highlighted in dark green and light green respectively. Underneath the top bar is a representation of predicted α-helices (pink) and β-strands (blue). Predicted disorder is represented underneath this by a jagged line, the lower line the higher the predicted confidence of disorder. Constructs coloured black were expressed but were not soluble. Constructs in blue were expressed and soluble but only with the GST tag. Constructs below the dashed line and coloured green were soluble without the GST tag and were stable on gel filtration.

3.3.2 Workflow for τ138 construct screening

I employed an iterative workflow for this project (*Figure 3.8*). Cloning was carried out using the pETM30 system, which introduces an N-terminal, TEV protease-cleavable 6x-histidine-GST tag. Expression and solubility tests were carried out in parallel, on a small-scale, with promising constructs (i.e. soluble and well-expressed) then taken to large scale for purification and analysis (*Materials and Methods*, *6.2*). It should be noted that the biggest challenge for this project was protein solubility, with most

proteins either ending up in the pellet after cell lysis or becoming insoluble after the GST tag was cleaved off. In total I tested 41 constructs, of which 11 were soluble without the GST tag and were stable on gel filtration. This approach focused on a semi-high-throughput construct pipeline and the majority of the iterations of my workflow were centred on the N-terminus of τ138. This was because the N-terminal half of TFIIIC appeared most conserved when comparing yeast TFIIIC to human TFIIIC (*Figure 2.5*) [103] and so seemed a more suitable candidate for B box binding.

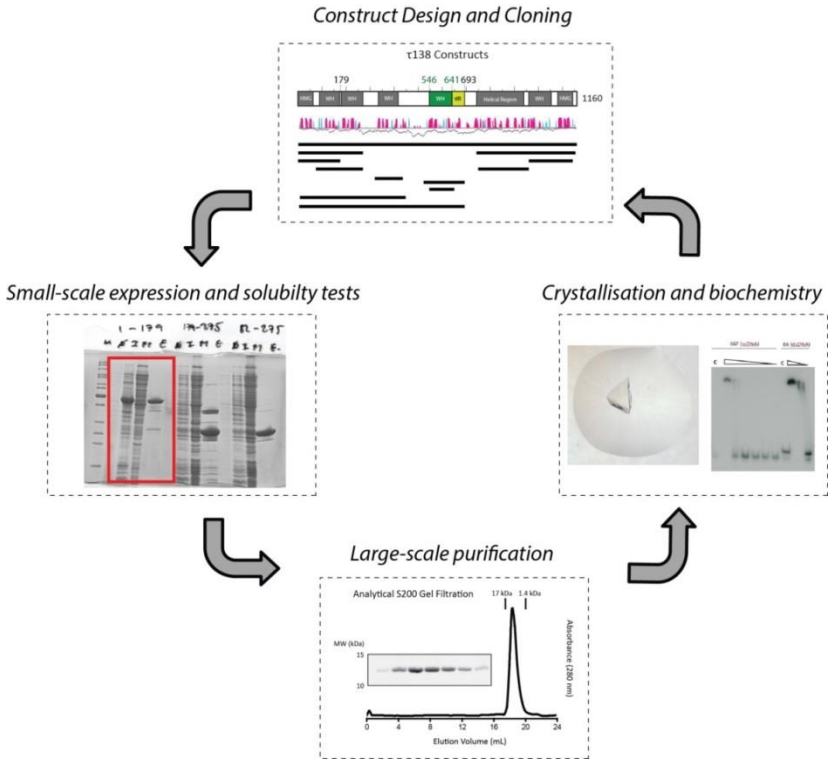

Figure 3.8 **Workflow for screening τ138 constructs.**

3.4 Characterising an N-terminal τ138 fragment (1-179)

After several iterative rounds, an N-terminal construct (1-179) was identified as expressed and soluble by my workflow. The construct is predicted to contain a HMG-box domain and a winged helix domain (*Figure 3.7*).

3.4.1 Purification of τ138 (1-179)

Cloning, expression and purification of this protein was carried out using the same protocol as for the eWH domain (*Section 3.1.1*). After elution from nickel resin, the 6x-histidine-GST tag was cleaved by TEV protease and the cleaved protein was collected

(*Figure 3.9A*). The protein was then injected onto a Superdex 75 preparative gel filtration column (GE Healthcare) and eluted at an expected position for a monomer and as a single species (*Figure 3.9B*). The protein was evaluated by SDS-PAGE and judged to be appropriate for crystallisation trials.

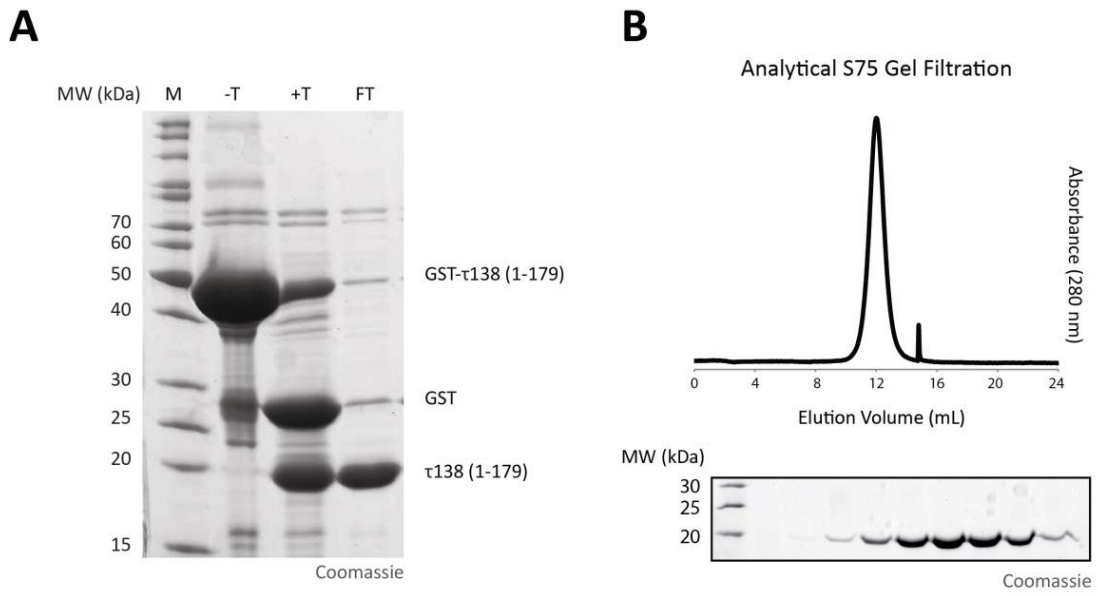

Figure 3.9 **τ138 (1-179) can be purified to homogeneity**. (A) 12% SDS-PAGE showing TEV protease cleavage of τ138 (1-179). T = TEV protease; FT = flow-through. M = marker. (B) Analytical gel filtration of τ138 (1-179). The sample was injected onto a Superdex 75 10/300 GL column (GE Healthcare) and elution position analysed. N.B. the second narrow peak was as a consequence of an overpressure on the system. 15% SDS-PAGE of elution peak is shown underneath the gel filtration profile.

3.4.2 Crystallisation of τ138 (1-179)

The protein was concentrated to 30 mg/mL for initial crystallisation trials. Crystals were grown at 20°C and at 4°C using the sitting drop method in an initial broad screening approach. Initial crystals were obtained in the Wizard I and II screen (EMBL Heidelberg Crystallisation Platform) with a reservoir solution of 0.1 M sodium acetate pH 4.5 with 0.8 M sodium phosphate and 1.2 M di-potassium phosphate at 20°C. Refinement screens were set up around this condition, both at 20°C and at 4°C. *Figure 3.10* provides typical examples of crystals that were tested for this protein.

Reproducibility of these crystals was very challenging. Using flash-frozen and then thawed protein resulted in reduced crystal quality, thus requiring a fresh preparation before each crystallization trial. Furthermore, several concentrations of glycerol and other compounds such as PEG400 had to be tested for cryo-protection as

the crystals appeared to suffer from cracking upon freezing. Crystals grown at 4°C appeared more robust and larger, but provided un-interpretable diffraction patterns. Crystals grown at 20°C grew in clusters that proved difficult to collect, but did occasionally produce clear diffraction patterns. Resolution was limited to 4-6 Å however.

For now, the structure of this protein remains unsolved. The best dataset currently extends to 3.8 Å but no molecular replacement solution could be found. Furthermore, no selenomethionine incorporated protein crystals were successfully grown up to now.

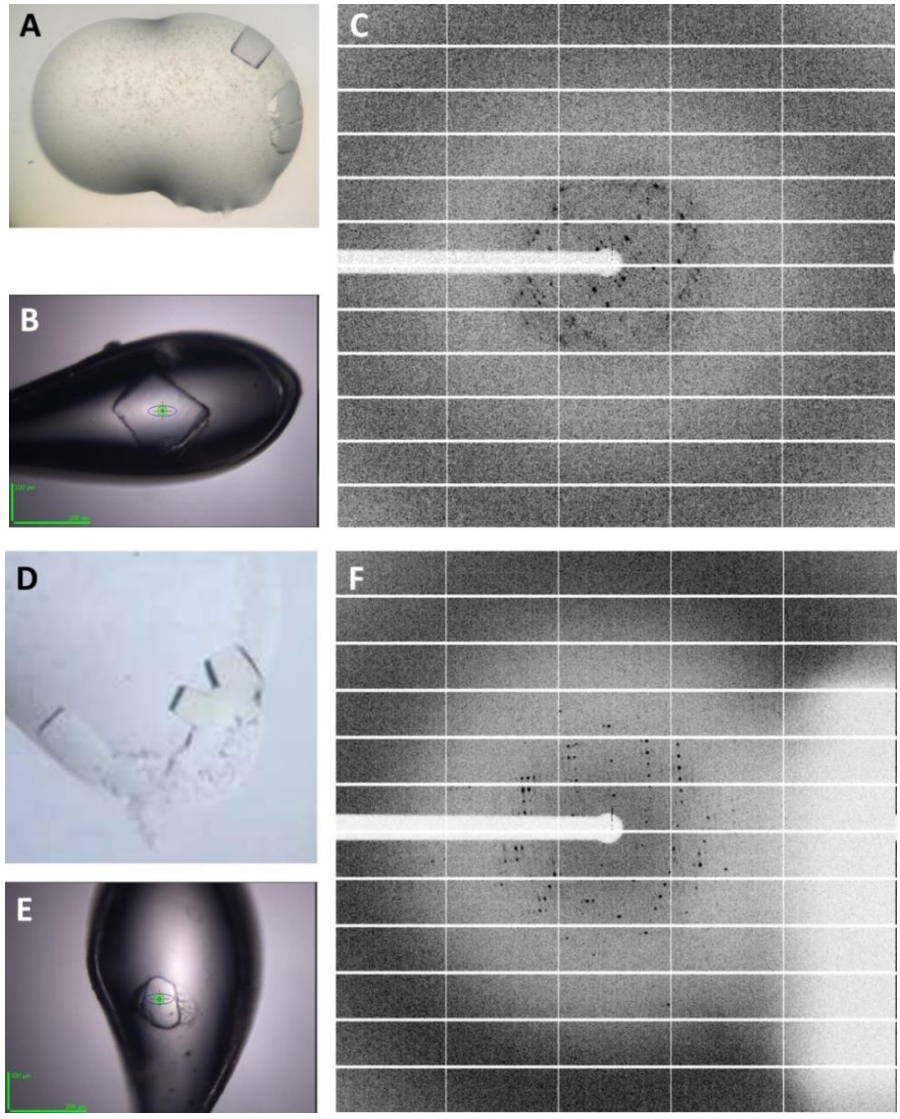

Figure 3.10 **Crystallisation and diffraction of τ138 (1-179).** (A) Crystal grown at 4 °C. (B) Collected crystal in a loop for data collection. (C) Example of a diffraction image taken for the crystal collected in B. (D) Crystal grown at 20 °C. (E) Collected crystal in a loop for data collection. (F) Example of a diffraction image taken for the crystal collected in E.

3.4.3 Testing for DNA-binding of τ138 (1-179)

With no structure available, but with the construct predicted to contain two potential DNA-binding domains, I carried out an EMSA experiment as I had done for the eWH domain (*Section 3.2.2*). I used the same 66 base oligonucleotide as previously (see *Materials and Methods, 6.2.4*). The protein was tested for binding to ssDNA and dsDNA oligonucleotides at varying concentrations. The results of the single-stranded experiment are displayed in *Figure 3.11*. Binding to ssDNA was not observed until a protein concentration of 500 μM was reached, indicating non-specific binding of the eWH domain to DNA. The same was observed for dsDNA (data not shown). I therefore concluded that τ138 (1-179) does not bind specifically to tDNA.

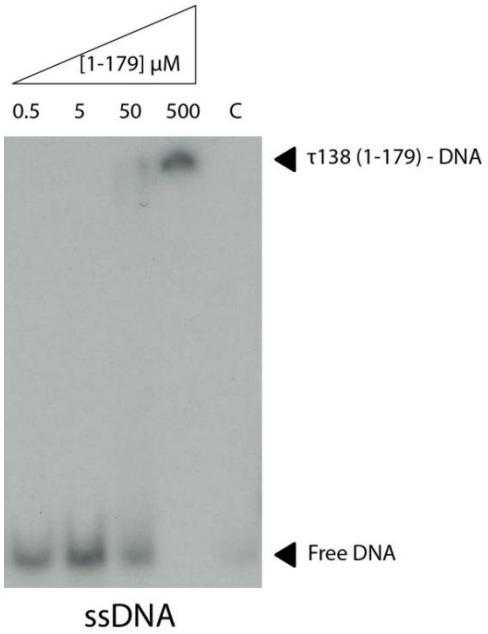

Figure 3.11 **τ138 (1-179) does not specifically bind single-stranded (ss) tDNA^{Glu}.** EMSA experiment of τ138 (1-179) binding to a ssDNA 66 base tDNA^{Glu} oligonucleotide. Top arrow indicates the shifted τ138 (1-179)-DNA complex and the lower arrow indicates free DNA. C = control experiment (no protein added).

3.5 Cross-linking TFIIIC when bound to tDNA

Producing soluble fragments of τ138 to test for DNA-binding, proved more challenging than expected and the two fragments that could be tested (546-641 and 1-179) did not bind tDNA. I reasoned that using the cross-linking coupled with mass spectrometry approach that was employed previously could provide useful information instead. The

rationale was that it might be possible to identify regions within TFIIIC that are altered upon DNA-binding, by comparing the cross-links obtained with unbound TFIIIC.

3.5.1 Binding TFIIIC to tDNA

The 66 base $tDNA^{Glu}$ oligonucleotide used in the previous EMSAs had been selected based on the fact that TFIIIC binds to it with moderately high affinity. *Figure 3.12* shows a clear band-shift at 500 nM $tDNA^{Glu}$ indicating the specific binding of TFIIIC to the previously used tDNA oligonucleotide.

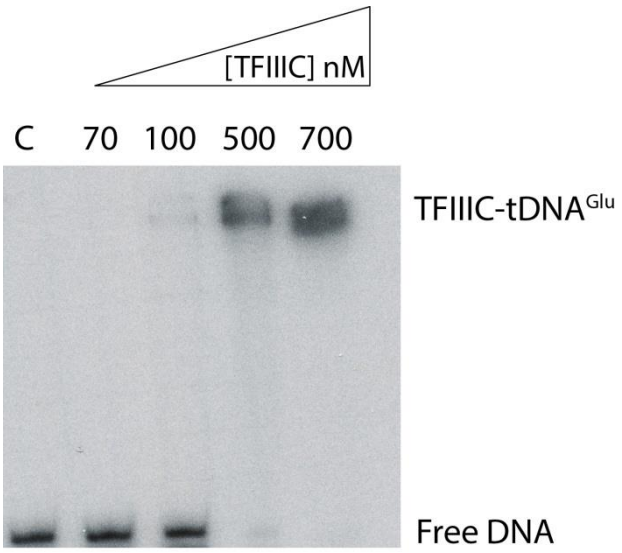

Figure 3.12 **TFIIIC binds the 66bp double-stranded $tDNA^{Glu}$ with high affinity.** EMSA experiment of TFIIIC binding to a dsDNA 66bp $tDNA^{Glu}$ oligonucleotide. Top arrow indicates the shifted TFIIIC-DNA complex and the lower arrow indicates free DNA. C = control experiment (no protein added).

3.5.2 Comparing TFIIIC cross-links when bound to tDNA

TFIIIC was incubated with the oligonucleotide as described in *Materials and Methods*, *6.2.5*. Chemical cross-linking coupled with mass spectrometry was performed as described before (see also *Materials and Methods, 6.1.2*).

Table 3.2 lists the obtained inter-subunit cross-links from both DNA-bound and DNA-unbound TFIIIC cross-linking conditions. Of the 41 inter-subunit cross-links obtained over both conditions, 19 are detected in both conditions, 14 are unique to the DNA-unbound condition and 8 are unique to the DNA-bound condition. *Table 3.3* lists the obtained intra-subunit cross-links from both DNA-bound and DNA-unbound TFIIIC cross-linking conditions. Of the 107 intra-subunit cross-links obtained over both

conditions, 45 are detected in both conditions, 44 are unique to the DNA unbound condition and 18 are unique to the DNA bound condition. *Figure 3.13* displays cross-linking maps of TFIIIC unbound and bound to DNA.

Table 3.2 **– Inter-subunit cross-links of TFIIIC**

		Both conditions		- DNA only		+ DNA only	
Prot. 1	***Prot. 2***	***Lys. 1***	***Lys. 2***	***Lys. 1***	***Lys. 2***	***Lys. 1***	***Lys. 2***
τ60	τ91	124	130	124	153		
		124	139				
		124	142				
τ60	τ138					124	506
τ91	τ138			178	478		
				598	433		
				598	509		
				610	414		
τ131	τ138	242	697	242	561	113	619
		242	698	412	689	242	642
		242	715	426	689	426	708
		349	729	627	885	426	715
		412	697	636	240		
		426	697				
		426	698				
τ131	τ55					113	211
τ95	τ138	138	896	118	200		
		141	896				
τ95	τ131	219	958	403	871	403	412
		395	871	631	947		
τ95	τ55	57	412	66	412	57	359
		66	359				
		66	419				
		123	419				
		145	196				

Prot. = Protein / Lys. = Lysine

The link between τ131 and τ138 that was discussed in the previous chapter is maintained in both DNA-bound and DNA-unbound conditions. However, there are some differences observed between the cross-links obtained in other regions of these two subunits. There are two lysine residues outside the TPR array, K636 and K627, which cross-link to τ138 at the N and C termini respectively in the DNA-unbound condition that are not present when DNA is bound. These lysine residues are situated in a predicted disordered region of τ131 that contains many intra cross-links in the DNA-unbound condition. Over half of these intra cross-links are not present in the

DNA-bound condition, suggesting that local conformational changes take place within this region of τ131 upon binding to DNA.

Table 3.3 – **Intra-subunit cross-links of TFIIIC**

		Both conditions		- DNA only		+ DNA only	
Prot. 1	***Prot. 2***	***Lys. 1***	***Lys. 2***	***Lys. 1***	***Lys. 2***	***Lys. 1***	***Lys. 2***
τ138	τ138	240	252	123	147	123	170
		348	252	153	139	348	247
		362	252	370	165	400	362
		362	257	383	478	401	362
		370	257	400	257	506	478
		370	362	414	383	506	619
		370	383	439	200	619	592
		478	483	491	509	619	637
		607	538	506	509	619	715
		689	637	506	538	708	729
		689	715	536	561	900	890
		697	715	536	607		
		698	689	689	642		
		698	715	698	639		
		708	697	732	715		
		708	715	767	715		
		729	715	839	854		
		767	729	900	1073		
		896	885	930	900		
				930	1073		
τ91	τ91	139	130	142	591	178	130
		142	130	142	598	178	139
		142	161	153	130	503	139
		142	178	153	161	598	134
		142	503	153	503		
		161	139	153	598		
		161	178	161	130		
		161	186	188	178		
		178	186	591	130		
		188	178	598	130		
		503	130	634	610		
		591	139				
		591	161				
τ60	τ60	46	549	18	228	160	116
		53	5	116	556		
τ131	τ131	102	113	102	108		
		513	506	513	678		
		592	567	567	631		
		637	631	581	567		
		688	631	599	567		
		702	627	599	607		
				627	567		
				702	631		
τ95	τ95	48	123	48	57	243	251
		128	138	251	254	266	251
		128	145	408	403		
		128	141				
		637	627				

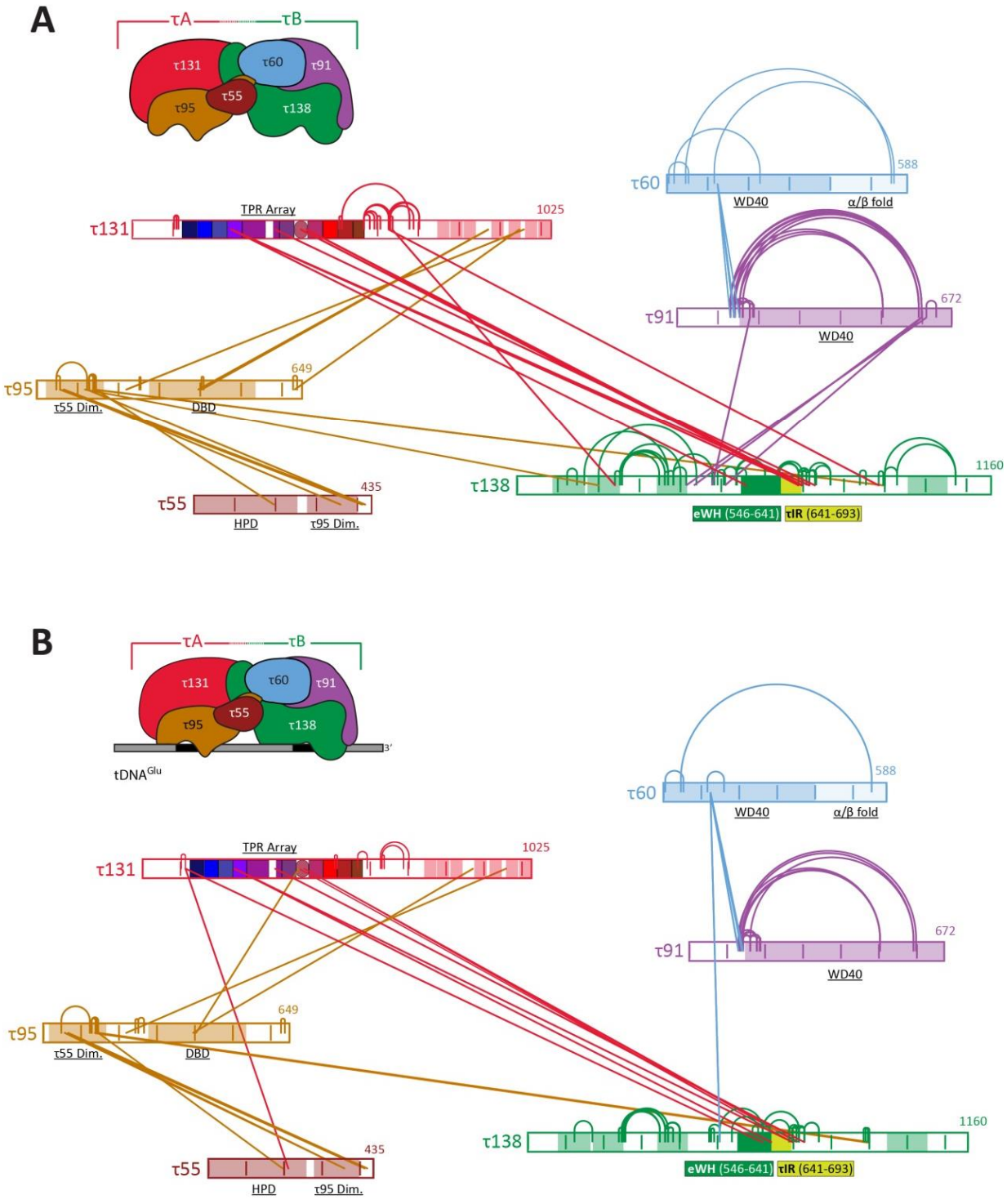

Figure 3.13 **Cross-linking map of TFIIIC unbound (A) and bound to tDNA^{Glu} (B).** TFIIIC subunits are represented as bars. Protein length is labelled at the C-terminus. Domains of which crystal structures are available are underlined. τ55 and τ95 Dim. = τ55 and τ95 dimerisation domains; DBD = DNA-binding domain; HPD = histidine phosphatase domain. The eWH domain and τIR are highlighted. Shaded green boxes on τ138 indicate predicted winged helix domains. Shaded red boxes on τ131 indicate predicted TPRs. Inter cross-links are depicted by lines connecting different subunits. Intra cross-links are depicted by arcs which connect residues within the same subunit.

47% of all intra cross-links (50) were detected for the τ138 subunit of TFIIIC. It is striking to observe that a large number of these intra cross-links occur between the predicted winged helix domains within the N-terminus of τ138, suggesting that the individual domains are in close spatial proximity to one another, perhaps even stabilising each other in the context of the entire protein. 40% of the τ138 intra cross-links are lost when TFIIIC is bound to DNA, which could support the notion that these predicted winged helix domains bind DNA, potentially using lysine residues that may then not be able to react with the cross-linker.

Both the τ60-τ91 and τ95-τ55 dimers do not appear to undergo drastic conformational changes with respect to the subunits within each dimer, as the inter and intra cross-links are generally present in both DNA-bound and DNA-unbound conditions (τ55 contains no intra cross-links in fact). I validated the intra cross-links of τ60 as they all occurred between lysine residues that are present in the previously solved crystal structure (*Figure 3.14*). The distances between these residues were all under 30 Å, thus satisfying the maximum distance permitted by the length of the cross-linker. The τ95-τ55 dimerisation domain cross-links could not be validated as they generally occurred outside the structured interface or between residues that are not modelled in the structure (i.e. in disordered loops).

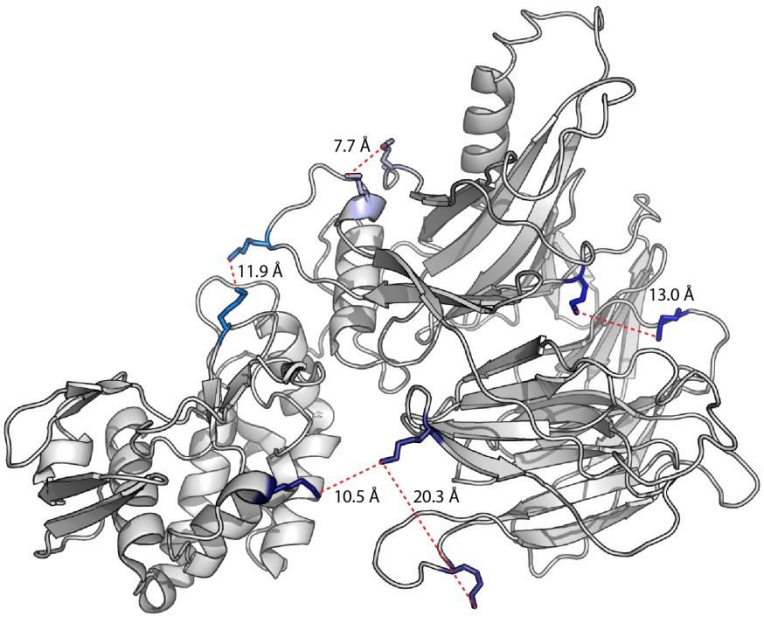

Figure 3.14 **Validating the intra cross-links of τ60**. Distances between cross-linked lysine pairs were measured and are displayed as dashed red lines.

The most striking difference between the DNA-bound and DNA-unbound cross-linking conditions is a region of cross-links between τ91 and τ138. 4 cross-links are observed between the WD40 domain of τ91 and a disordered region of τ138 (414-509). None of these cross-links are detected when TFIIIC is bound DNA. This is highly intriguing given the predicted roles of τ138 and τ91 in contributing to B-box binding and will be discussed further below.

3.6 Discussion

This chapter has presented the first high resolution structure for the τ138 subunit of TFIIIC, an extended winged helix domain, which is conserved from yeast to human. In addition, initial crystals were obtained for the N-terminal fragment (1-179) of τ138. Cross-linking experiments have provided initial insights into a region of τ138 that may be responsible for B box binding and will be discussed below.

3.6.1 What binds the B box of tRNA genes?

The chemical cross-linking coupled to mass spectrometry results provided an intriguing finding which is consistent with previous data. Cross-links were observed between subunits τ91 and τ138 of TFIIIC that were subsequently not detected when TFIIIC was bound to tDNA (*Figure 3.13*). Several factors imply that this is not simply an experimental difference between the two conditions. For example, the ld scores of the cross-links when detected in the DNA-unbound state are above 25, strongly suggesting they are not false-positives. Furthermore, it is stochastically unlikely that all cross-links between τ91 and τ138 were simply not detected in the DNA-bound condition. What could explain this disappearance of cross-links? One explanation is that the binding of DNA in this region leads to steric hindrance of the cross-linker for solvent-exposed lysine residues in this region. Another explanation is that a significant rearrangement of domains occurs at this region upon DNA binding, which is further supported by the change in intra cross-links within τ138. Both explanations thus suggest that this region of τ138 is involved in binding to DNA. The region (414-509) where the cross-links are located is largely disordered and does not contain any known DNA binding domains, and so it is highly likely that DNA binding is mediated by one of the winged helix domains adjacent to it. As an aside, this situation is similar as for the τ131-τ138 cross-

links previously discussed, as the links occur generally adjacent to the actual interacting region, probably because the actual interaction region is buried and so not accessible to cross-linking (also observed with τ60-τ91, see *Section 2.6.3*). As I have shown that the eWH domain does not bind DNA, the most likely candidate for DNA binding is thus the third predicted winged helix domain (*Figure 3.15*).

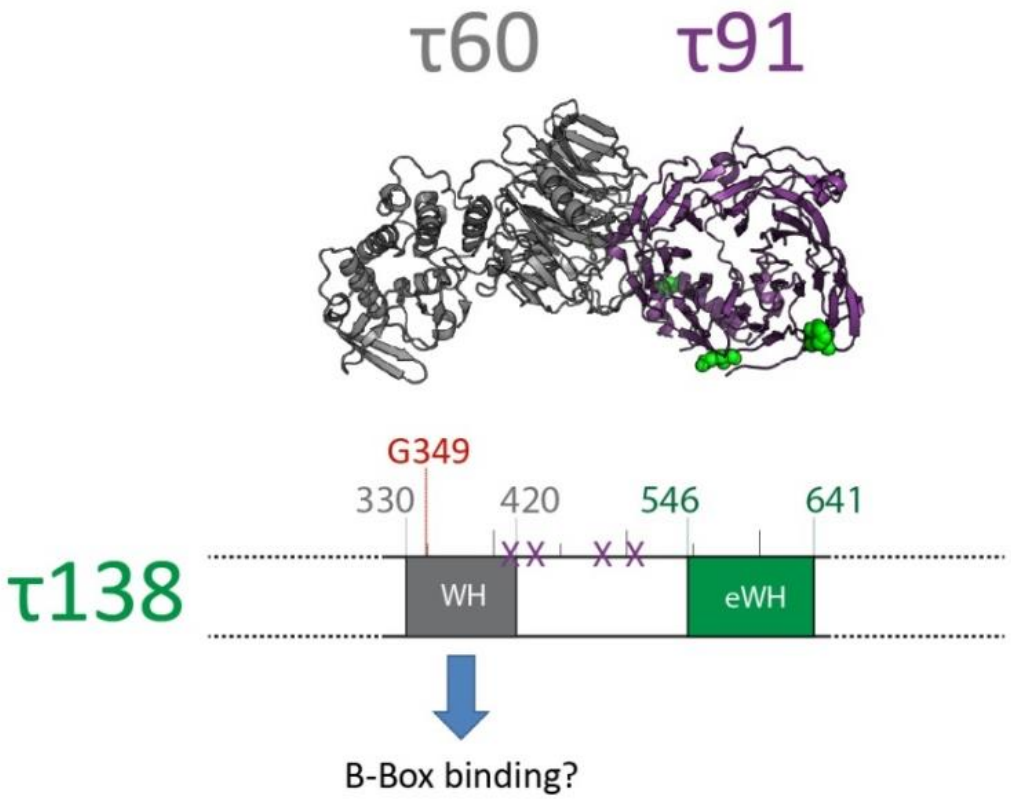

Figure 3.15 **A potential role for the third predicted winged helix domain in B-box binding.** Lysine residues involved in τ91-τ138 crosslinking (without tDNA) are represented as green spheres on the τ91 WD40 (purple) and as purple X's on the τ138 schematic. The predicted position of the third winged helix is indicated between residues 330 and 420, and is consistent with secondary structure predictions. Residue G349 is highlighted (see main text for details).

This is interesting for two reasons: 1) A temperature-sensitive mutation, G349E (see *Introduction*) occurs within this predicted winged helix domain and has been shown to reduce TFIIIC-tDNA binding [93]; 2) In a bioinformatics study, which also looked for B box binding conservation in bacteria and archaea, this area has been identified as a highly conserved region within the relatively poorly conserved τ138 [157]. I have screened for constructs around this predicted winged helix but have not obtained soluble protein so far. I have also tried expressing constructs in combination with τ91,

based on the cross-links and consistent with a previous study (see also *Introduction, 1.3.6*). Identifying the minimal B-box binding region could allow for analysis by X-ray crystallography, perhaps by co-crystallisation of the minimal domain with the B-box sequence. Alternatively, it could be possible to covalently cross-link TFIIIC to the tDNA by UV, in order to identify the region of TFIIIC that remains bound (after extensive washing and digestion) by mass spectrometry as has been reported for RNA-binding proteins [158]. Further work will be required in order to definitively answer the question that leads this section.

3.6.2 What is the role of the other domains within τ138?

The eWH domain and 1-179 fragments did not bind DNA, as shown by EMSA. However, as these are isolated domains, one cannot conclude that they do not bind DNA *in vivo*. Perhaps the eWH domain and/or 1-179 bind DNA in combination with each other, or other winged helix domains present in the N terminus. They may also mediate protein-protein interactions. A structure of a tandem array of winged helix domains, interestingly extended winged helix domains, has been reported. Lefèvre et al. determined the crystal structure of hRPC62 (the human homologue of the yeast Pol III-specific subunit C82) which contains four eWH domains arranged around a coiled-coil [50]. The eWH domains were shown to bind ssDNA, dsDNA and bind other subunits of Pol III. The authors propose that hRPC62 acts as an 'adaptor platform' within the context of the Pol III PIC. Such a situation could occur for τ138, particularly as many intra cross-links occur between the N-terminal predicted winged helix domains. An intriguing hypothesis is that the τ138 winged helix domains could stabilise the transcription bubble transiently before elongation proceeds, with individual winged helix domains contributing low affinity interactions with protein and DNA partners. It may also be possible that the winged helix domains surrounding the possible B box binding domain help to stabilise B box recognition.

Finally, what is the role of the predicted HMG boxes at the N and C termini? A very recent bioinformatics study and previous work has revealed that these two domains are very well conserved [103, 157]. This ~75 amino acid domain of three α-helices often binds kinked, or non B-type DNA [156]. It is possible that these HMG boxes unwind the promoter DNA, perhaps combining with the winged helix domain for

specificity. A structure of the 1-179 fragment would be of some benefit here as it contains a predicted HMG box and a winged helix domain. Crystals have been obtained but the structure remains unsolved (*Section 3.4.2*). Further screening of crystallisation conditions will be required to produce crystals that produce an interpretable diffraction at sufficiently high resolution.

4. Results - Producing Recombinant TFIIIC

The work in this chapter describes efforts to produce recombinant TFIIIC complex for structural and functional studies. The work presented uses the MultiBac method [159] in order to produce the complex in insect cells (*Spodoptera frugiperda)*. I present purifications of recombinant τA subcomplex and recombinant TFIIIC holo-complex.

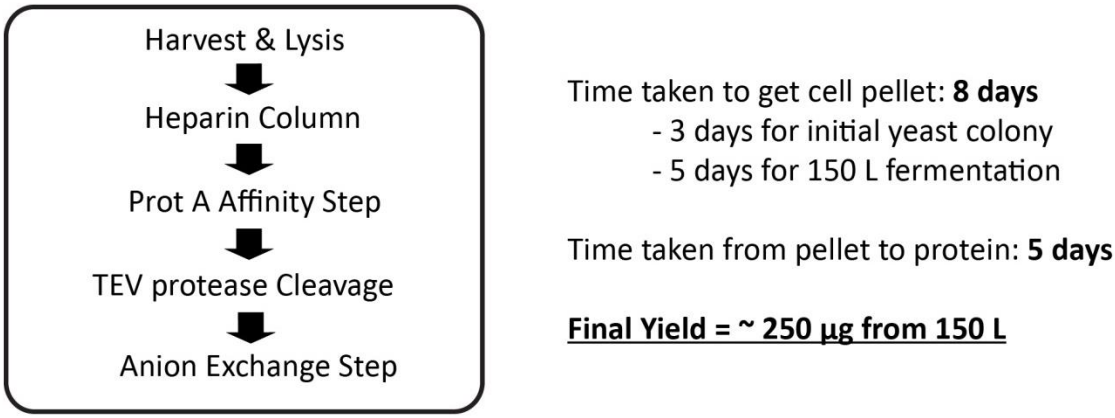

Figure 4.1 **The time required for endogenous TFIIIC purification.** Boxed text is taken from *Figure 2.1* and indicates the major steps taken during endogenous TFIIIC purification.

Our previous work on the full TFIIIC complex has required the purification of this protein endogenously from yeast using a TAP-tag strategy. Due to the low abundance of TFIIIC in yeast cells, as confirmed by proteomic studies [160, 161], the yield is very low (~250 µg of TFIIIC per 150 L yeast culture). As a consequence, the production of recombinant TFIIIC would be a significant step in increasing our workflow and allowing us to address the types of questions that rely on large amounts of purified protein.

4.1 MultiBac and polyproteins

In collaboration with the Berger group at EMBL Grenoble, I utilised the MultiBac system in efforts to produce the τA and τB subcomplexes and ultimately TFIIIC recombinantly in insect cells. The protocols included in this chapter are based on the Handbook 'Eukaryotic Expression Facility at EMBL Grenoble' from Frederic Garzoni and Imre Berger, and on published protocols [159, 162] (see also *Materials and Methods*, *6.3*). The MultiBac system utilises fluorescent proteins to analyse protein expression

and virus performance in order to improve reproducibility and provide a faster means of assessing successful protein expression. A yellow fluorescent protein (YFP) is encoded in the DNA (Bacmid) of the generated virus as a marker for virus performance. Another fluorescent protein (e.g. cyan fluorescent protein, CFP) is also present at the C-terminus of the open reading frame to assess protein expression levels. When both fluorescence levels reach a plateau (i.e. maximal expression level of protein has been reached) the cells can be harvested and the purification proceeds (*Figure 4.2*). For τA, a novel specific polyprotein strategy was used. Specifically, the genes for each subunit are contained within the same open reading frame, and expression is under the control of one promoter. This strategy allows stoichiometric expression of each subunit from the complex. The sequences for each subunit are separated by TEV protease-cleavable sites, with the TEV protease-coding gene also expressed within the same open reading frame (*Figure 4.2*).

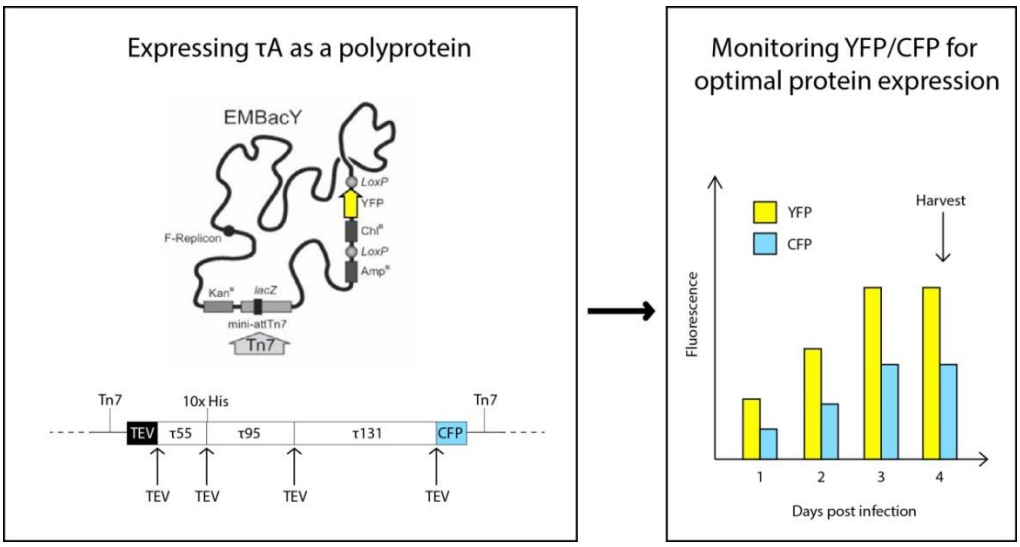

Figure 4.2 **Using MultiBac for recombinant τA production.** *Left,* the EMBacY Bacmid DNA with the YFP protein and the Tn7 site for integration of the polyprotein construct highlighted. Figure modified from [159]. The τA polyprotein is indicated underneath the Bacmid with the encoded proteins, including TEV protease and CFP. *Right*, monitoring the YFP and CFP values over the course of an infection allows optimal harvesting time for cells.

4.2 Recombinant production of τA

Cloning of constructs was carried out by Christoph Bieniossek, in the Berger group at EMBL Grenoble. The τA polyprotein showed most promise in initial expression trials

and so I focused on this subcomplex first. Viruses were generated as described in *Materials and Methods, 6.3.1*.

4.2.1 Expression and purification of τA

The reader is referred to *Materials and Methods* for full details of expression and purification protocols. Typically, 6 L of Sf21 insect cells were infected with V_1 virus at a dilution of 1 in 10000. When YFP/CFP levels had plateaued the cells were harvested for purification. The τA subcomplex was isolated by means of a 10x-histidine tag on the τ95 subunit. Soluble lysate was circulated for several hours on a TALON nickel affinity column (Clontech) in order to achieve maximum yield yet still with high purity. The eluted complex can be seen in *Figure 4.3*.

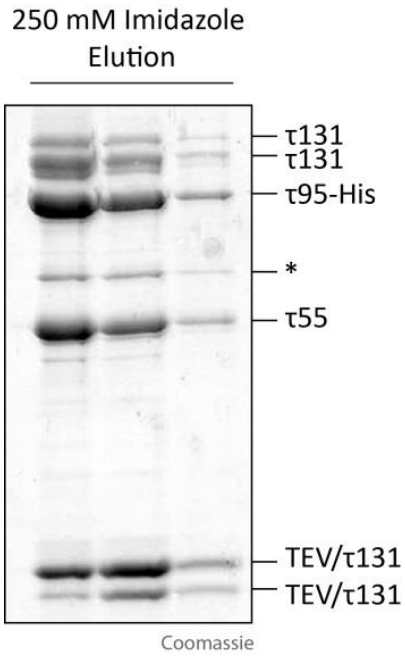

Figure 4.3 **Pull-down of τA subcomplex from insect cells.** 12% SDS-PAGE of histidine-eluted τA subcomplex. Note the two bands present for τ131 and the double band of TEV protease. * = Sf21 endogenous protein (determined by mass spectrometry).

All bands were analysed and identified by the Mass Spectrometry Core Facility at EMBL Heidelberg. It can be observed that the τ131 subunit degrades readily into a lower molecular weight species (~10-20 kDa smaller). The τ95 and τ55 subunits appear not to be affected by degradation. A double band, which was identified as TEV protease, is also visible. The TEV protease is histidine-tagged and so would be pulled down in addition to the complex.

Subsequently, the complex was injected onto a Superose 6 gel filtration column (GE Healthcare) in order to test for aggregation. As can be seen in *Figure 4.4A*, the complex elutes a single peak, indicating a mono-disperse sample. The elution position corresponds approximately to the correct molecular weight of ~250 kDa.

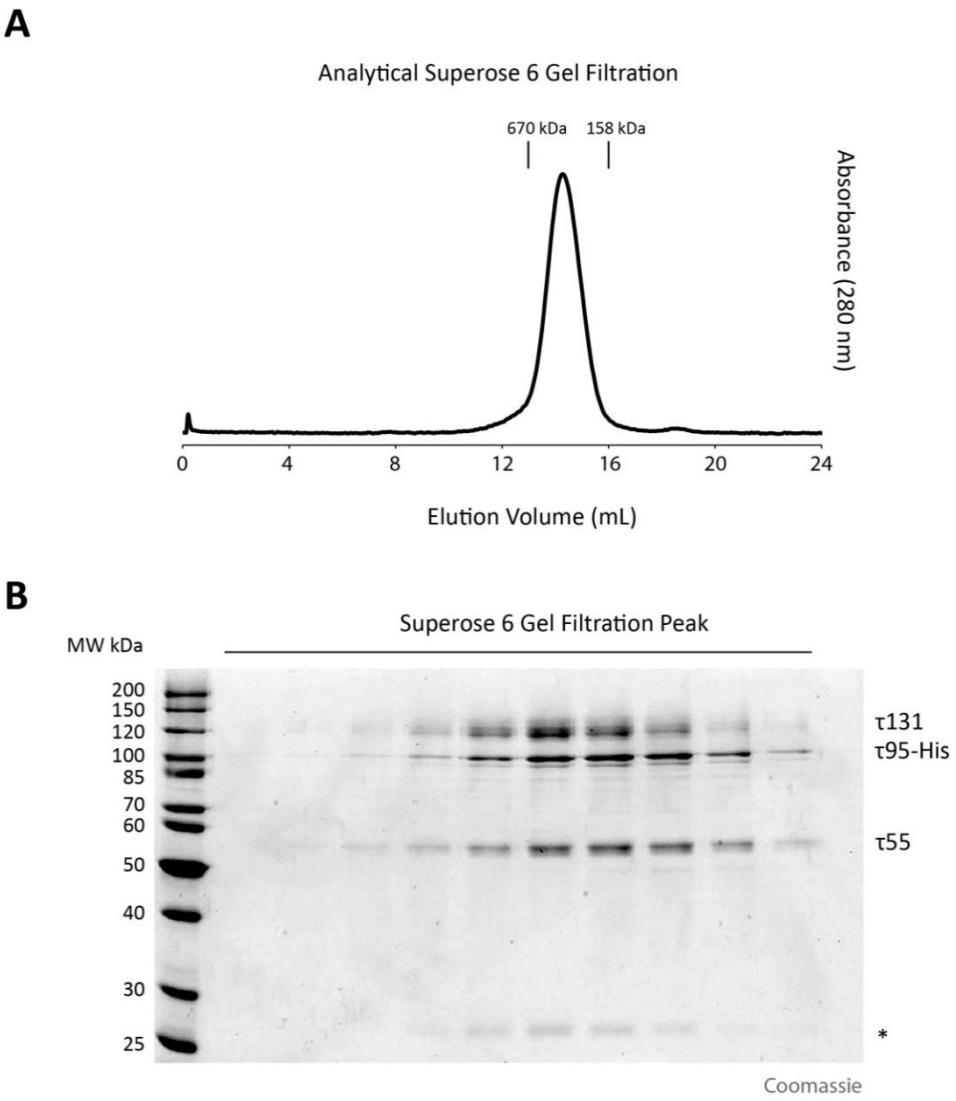

Figure 4.4 **Analytical Gel Filtration of τA**. (A) The sample was injected onto a Superose 6 10/300 GL column (GE Healthcare) and elution position analysed. Positions of known molecular weight standards are indicated at 670 kDa and 158 kDa. (B) 12% SDS-PAGE showing elution peak. * = unidentified band.

Figure 4.4B shows the elution peak fractions run on SDS-PAGE. The degradation of τ131 is still evident, yet the bands appear to migrate together as part of the complex, rather than eluting at separate positions distinct from the rest of τA. It could be that the cleaved TPRs of τ131 remain attached within the τA subcomplex, thus causing their co-migration. A band at approximately 25 kDa is indicated on *Figure 4.4B*. Although

mass spectrometry could not confirm the identity of this species, it is probably a degradation band of τ131, owing to its co-migration behaviour with the τA subcomplex.

In summary, a quite substantial yield of τA can be purified from insect cells (~2-5 mg from 6 L), albeit with degradation of the τ131 subunit.

4.2.2 Functional characterisation of τA

I first tested whether the τA subcomplex could bind DNA by EMSA (*Figure 4.5* and *Materials and Methods*, *6.3.4*). The subcomplex could bind to a 256 base-pair tDNA oligonucleotide at a concentration of 2-4 µM.

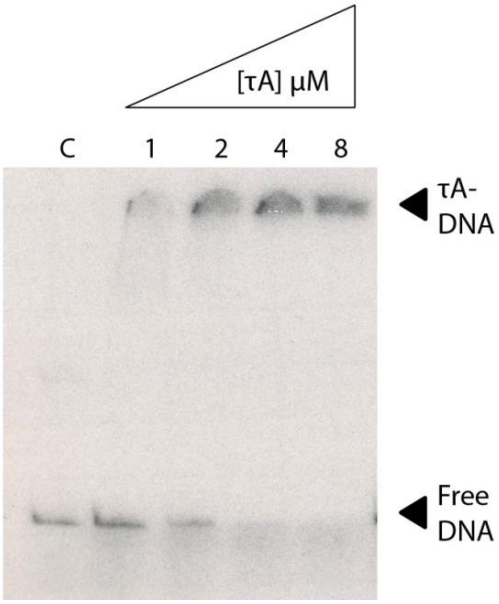

Figure 4.5 **EMSA experiment of τA binding to a double-stranded DNA 256bp tDNA^{Glu} oligonucleotide.** Top arrow indicates the shifted τA-DNA complex and the lower arrow indicates free DNA. C = control experiment (no protein added).

Next, I investigated the architecture of the τA subcomplex. Chemical cross-linking coupled with mass spectrometry was performed in collaboration with the Beck group at EMBL Heidelberg. The procedure has been discussed before and further details can be found in *Materials and Methods, 6.1.2*. There are similarities and differences when comparing the cross-links obtained for TFIIIC with the cross-links within isolated τA (*Figure 4.6*).

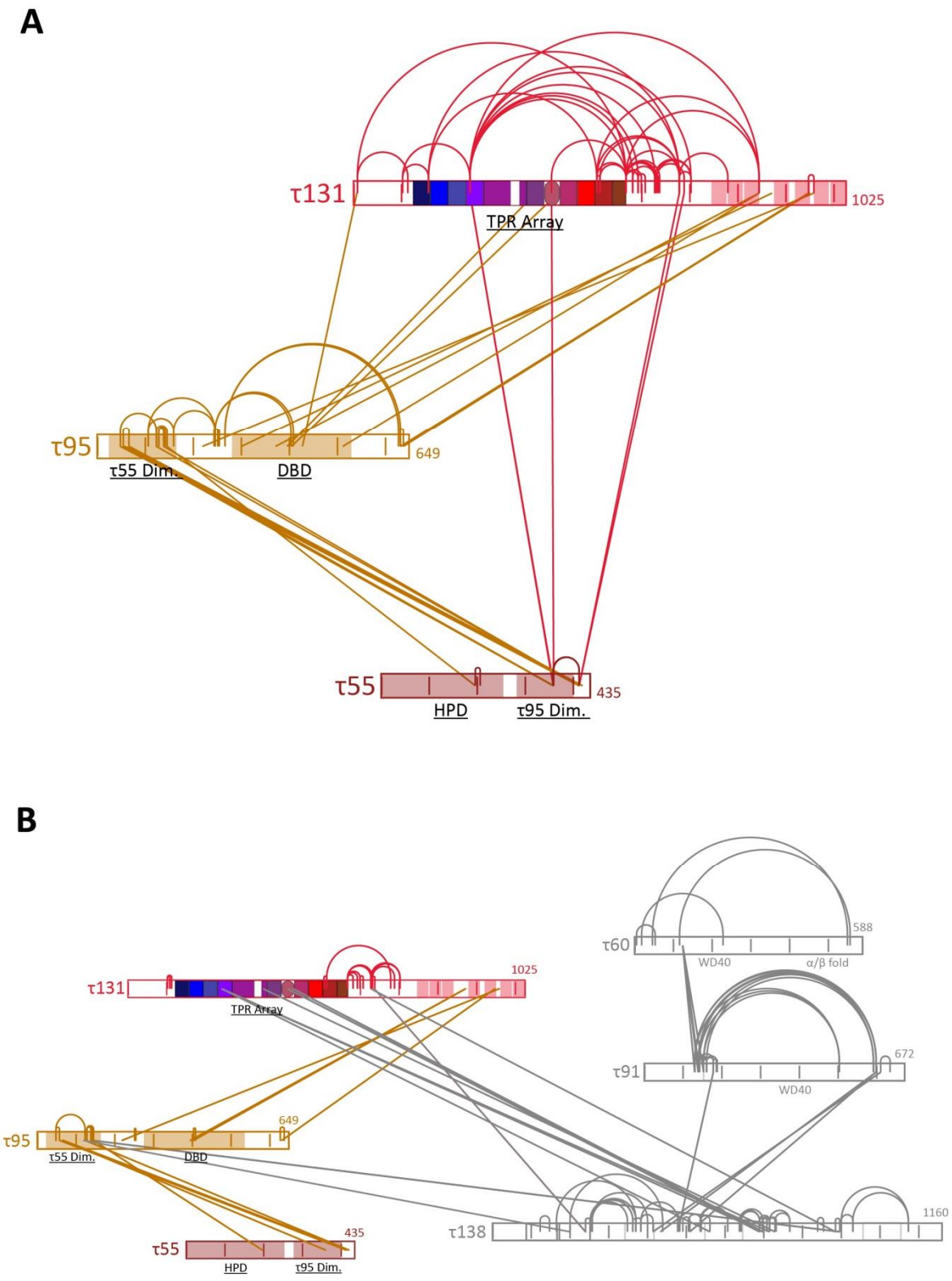

Figure 4.6 **Comparing chemical cross-linking of isolated τA and as part of TFIIIC**. (A) Chemical cross-linking map of isolated τA. Subunits are represented as bars. Protein length is labelled at the C-terminus. Domains of which crystal structures are available are highlighted. τ55 and τ95 Dim. = τ55 and τ95 dimerisation domains; DBD = DNA-binding domain; HPD = histidine phosphatase domain. Shaded red boxes on τ131 indicate predicted TPRs. Inter cross-links are depicted by lines connecting different subunits. Intra cross-links are depicted by arcs which connect residues within the same subunit. (B) Comparison with the τA cross-links obtained from holo-TFIIIC. Figure modified from *Figure 3.12*.

In detail, cross-links are obtained between the τ95 and τ55 dimerisation domains for both TFIIIC and τA. Furthermore, there are cross-links obtained between τ95 and the predicted TPRs of τ131 in the C-terminus in both cases. In τA alone, two cross-links appear between τ95 and the TPR array of τ131 (which would be predicted based on pull-down experiments of human homologues of τ95 and τ131 [88]) but these are not observed with cross-linking of TFIIIC. Furthermore, there are cross-links between τ55 and τ131 in the isolated τA subcomplex that are not observed in TFIIIC. Most strikingly, there are many more intra cross-links obtained within τ95 and τ131 for τA alone (see *Section 4.4.1*).

In summary, EMSA and cross-linking suggest that the recombinant τA is able to bind to DNA and possesses a similar architecture as in the context of TFIIIC. However, owing to the degradation of τ131 and the poor appearance of the subcomplex on negative stain electron microscopy grids (data not shown), I opted not to pursue structural characterisation of τA, but to focus on purification of the full TFIIIC complex.

4.3 Recombinant production of TFIIIC

A co-infection strategy was attempted for recombinant production of TFIIIC, using τA virus described in the previous section and a τB virus. The fluorescence strategy could still be employed, as the τB polyprotein is fused to mCherry at the C-terminus, thus expression of both subcomplexes can be monitored. In initial tests, Sf21 insect cells were infected with V_1 virus of both τA and τB at a dilution of 1 in 5000. When YFP/CFP/mCherry levels had plateaued the cells were harvested for purification, see *Materials and Methods, 6.3.3*. I had observed previously when attempting to purify the τB subcomplex alone, that expression was significantly lower than for τA (data not shown). Therefore, I opted to purify the complex by means of the CBP tag on the τ91 subunit of the τB subcomplex, in order to enhance the recovery of τB subunits. The soluble lysate was incubated with CBP resin for 2 hours at 4°C. *Figure 4.7A* shows the eluted complex.

All six subunits of the TFIIIC complex were successfully isolated by this purification. The τ60, τ91 and τ138 subunits of τB are over-represented owing to the presence of the CBP tag on τ91. The complex is highly pure after this one-step

purification and was immediately injected onto a Superose 6 gel filtration column (GE Healthcare) to assess for mono-dispersity and aggregation (*Figure 4.7B*).

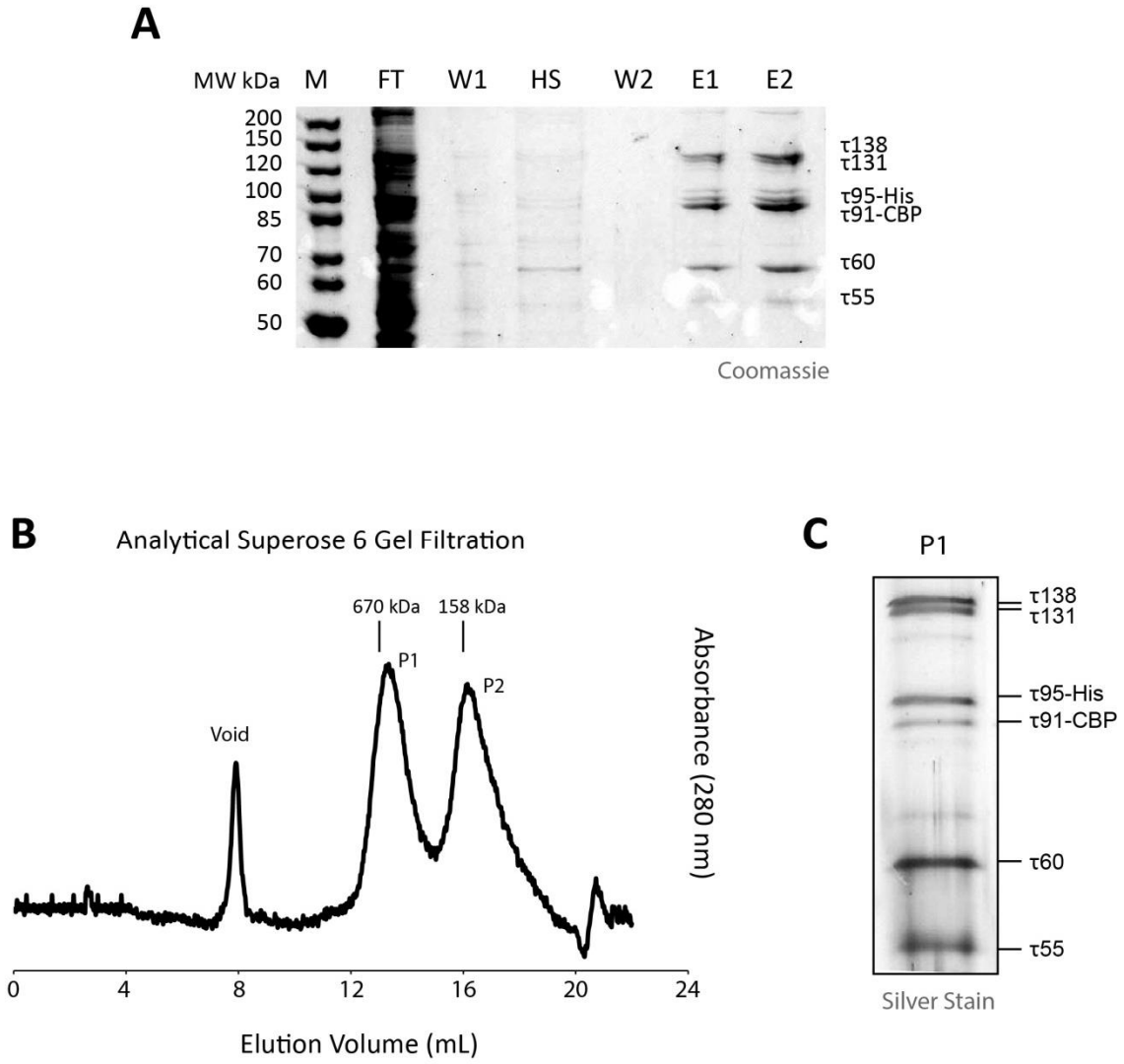

Figure 4.7 **Recombinant TFIIIC can be purified from insect cells**. (A) 12% SDS-PAGE of CBP-eluted TFIIIC subcomplex. M = marker; FT = flow-through; W1 = wash 1; HS = high salt wash; W2 = wash; E1 = elution 1; E2 = elution 2. See *Materials and Methods (6.3.3)* for details. (B) Analytical gel filtration of TFIIIC. The sample was injected onto a Superose 6 10/300 GL column (GE Healthcare) and elution position analysed. Positions of known molecular weight standards are indicated at 670 kDa and 158 kDa. The positions of the void peak and peak 1 (P1) and peak 2 (P2) are indicated (see main text for details). (C) A silver-stained SDS-PAGE of a P1 fraction.

There were three peaks observed during the gel filtration run. The first peak corresponded to the void volume and, although not displayed on the chromatogram in *Figure 4.7B*, showed a high 260 nm absorbance reading, indicating that this species was most likely DNA contamination, as no protein was visible by SDS-PAGE (data not shown). The first peak, P1, corresponded to the TFIIIC complex and eluted at the

expected molecular weight of ~520 kDa. Analysis by SDS-PAGE revealed the presence of all 6 subunits (*Figure 4.7C*). The second peak, P2, corresponded to the τ60-τ91 dimer and probably reflects an excess of these subunits based on the CBP tag location on τ91 (data not shown). In summary, the purification protocol can produce recombinant, mono-disperse TFIIIC.

4.4 Discussion

This chapter has detailed efforts to produce TFIIIC recombinantly using the MultiBac expression system in insect cells. A bottleneck for structural and functional studies on TFIIIC has been the very low yield of TAP-tagged protein from yeast. Early results look promising but will require further work in order to produce the desired yield. The work has also provided insight into the τA and τB subcomplexes and highlighted once again the need for a τ131-τ138 link which was characterized in the earlier chapters.

4.4.1 τ131 and τ138 degrade in the isolated τA and τB subcomplexes

The two DNA-binding subcomplexes of TFIIIC have been defined as τA and τB, with τA consisting of the subunits τ131, τ95 and τ55, and τB consisting of the subunits τ138, τ91 and τ60. This definition was used when designing our MultiBac TFIIIC constructs. As a short note, a full TFIIIC polyprotein was also synthesised but never showed detectable expression. The purification of the τA subcomplex that I have presented in this chapter shows that the trimeric τA complex can be expressed, pulled down and stays stably associated on gel filtration. However, the τ131 degrades rapidly and is clearly unstable in the absence of a τB subunit or subunits. This stabilisation is most likely mediated through τ138, as I have shown before that these two subunits directly interact. The larger number of intra cross-links observed within τ131 in the isolated τA subcomplex may be a consequence of this degradation, but is most likely due to the increased flexibility of τ131 in the absence of τB subunits or because these residues would normally interact directly with τB subunits in the context of TFIIIC. Unfortunately, I could not purify sufficient quantity of the τB subcomplex alone for similar cross-linking studies on the isolated τB subcomplex. It was observed that the largest subunit τ138 readily degraded, indicating that the τ60-τ91 scaffold is not sufficient for τ138 stabilisation. Again, based on my former findings, the τ131 subunit

of τA would probably be required for stabilisation of this subunit. In the context of recombinant TFIIIC, no degradation is observed for the τ131 and τ138 subunits, indicating that the combination of τA and τB is required for stabilisation of the τ131 and τ138 subunits.

4.4.2 Optimisation of recombinant TFIIIC purification

As shown in *Figure 4.7*, a recombinant TFIIIC complex can be produced in insect cells by using the protocols I have established. This procedure significantly cuts the time, effort and cost associated with purifying endogenous TFIIIC from yeast (*Figure 4.8*).

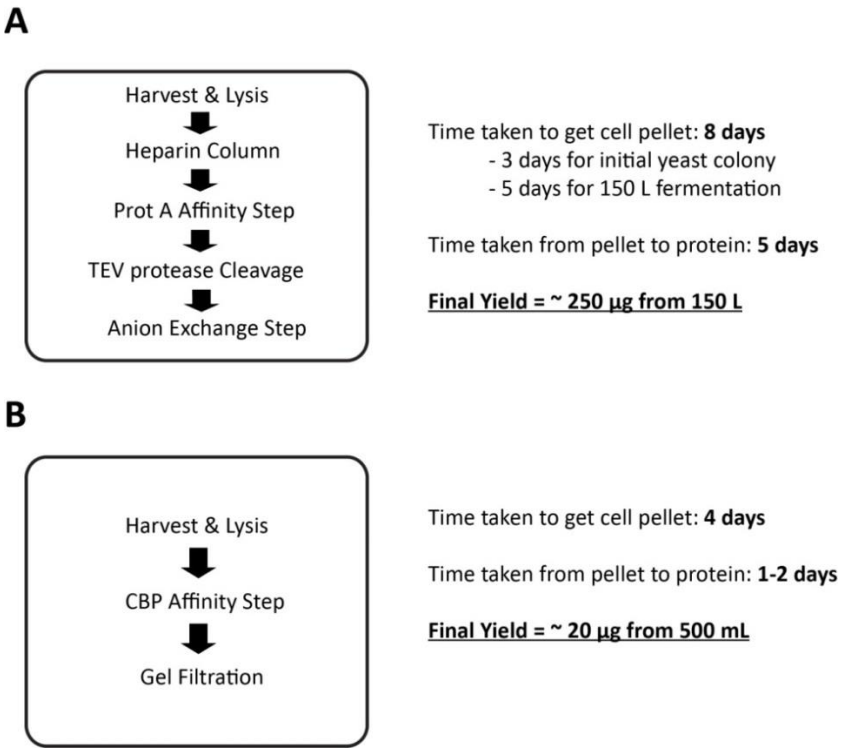

Figure 4.8 **Comparing endogenous and recombinant purification of TFIIIC.** (A) Overview of the endogenous TFIIIC purification as shown in *Figure 4.1*. (B) Overview of the recombinant TFIIIC purification.

In addition, it will allow the specific truncation or deletion of subdomains and introduction of point mutants that would lead to lethality in yeast cells as TFIIIC is an essential macromolecular complex. However, it is also apparent at this stage that producing *large* amounts of TFIIIC for structural and functional studies is still tricky. There is a significant difference in expression levels when comparing the τA and τB subcomplexes. It is not clear why the τB construct expresses so poorly but may be due

to the differences in vector backbone (pKL as opposed to pPBac for τA). One could try different vectors for τB expression and/or change the affinity tag on τB to improve the yield, although clearly the CBP tag allows the purification of a highly pure sample. Tagging another subunit may also be appropriate. τ60 for instance, at the C-terminus, has been shown to work in our TAP-tag purifications of TFIIIC. The results presented here are from relatively small-scale experiments (500 mL of insect cell culture) and so can be scaled up to produce larger quantities of material, but clearly resolving the issue with expression of τB is of paramount importance.

5. Outlook and Final Discussion

In this chapter, I will draw some overall conclusions based on the results and discussions presented in the previous chapters. I will also provide some future directions and perspectives for this exciting research field.

5.1 Linking the τA and τB subcomplexes of TFIIIC

The major finding of this thesis is the identification of a high-affinity interaction between the largest τA subunit, τ131, and the largest τB subunit, τ138. This is a novel finding which has been confirmed biochemically, biophysically using isolated proteins, and in the context of the entire TFIIIC complex. These two subunits could indeed form the flexible linker between the τA and τB subcomplexes, that has remained a molecular mystery for the last decades. Complementing *in vivo* mutational experiments in yeast are ongoing, in order to assess whether this link is the major connection between the τA and τB subcomplexes.

Another question regards the definition of the τA and τB subcomplexes and whether or not they actually exist *in vivo*. The concept of τA and τB stemmed from early electron microscopy studies and proteolysis experiments, and was further defined by biochemical and genetic experiments. However, for yeast TFIIIC it has been shown that the complex cannot be chromatographically separated into the two subcomplexes without proteolysis, suggesting that the two subcomplexes are extensively stabilised by each other. It is also clear from my *in vitro* purification that the τ131 subunit is not stable in the context of the subcomplex τA alone. The picture becomes more complicated with additional evidence that suggests the τ95-τ55 subcomplex exists independently from holo-TFIIIC [79]. Further structural work will be required to understand exactly which subunits or parts of subunits constitute the τA and τB subcomplexes and whether this definition will remain relevant.

5.2 τ131 - a hub for PIC formation

The TPR array of τ131 not only links τA and τB via τ138, it also provides a platform for the binding of τ95 [88], TFIIIB [109] and Pol III [119], as discussed in the *Introduction*. I

confirmed that Bdp1 binds the TPR array by GST pull-down experiments and that the site overlaps with that of τ138 (*Figure 2.22* and *Figure 2.23*). It is apparent that both arms of the TPR array are required for the binding of the Brf1 and Bdp1 subunits of TFIIIB [87], and that some degree of conformational flexibility must exist. The flexibility of τ131 has indeed been demonstrated, as photo cross-linking studies revealed that τ131 adopted different positions on the tRNA gene upon binding of TFIIIB [163]. The temperature factors displayed in *Figure 2.14B* also indicate flexibility at the N and C termini of the structure. Furthermore, the point mutants I have identified that reduce τ138 and Bdp1 binding are situated between the two arms and could suggest that changes to the architecture of the TPR array are important for ligand binding. The TPR array of τ131 may thus act as a flexible scaffold at which various subunits involved in Pol III transcription could assemble. *Figure 5.1* displays the sequential events that could occur at the τ131 TPR array during pre-initiation complex (PIC) formation and is based on the model proposed in [87] and on the findings I have presented in this thesis.

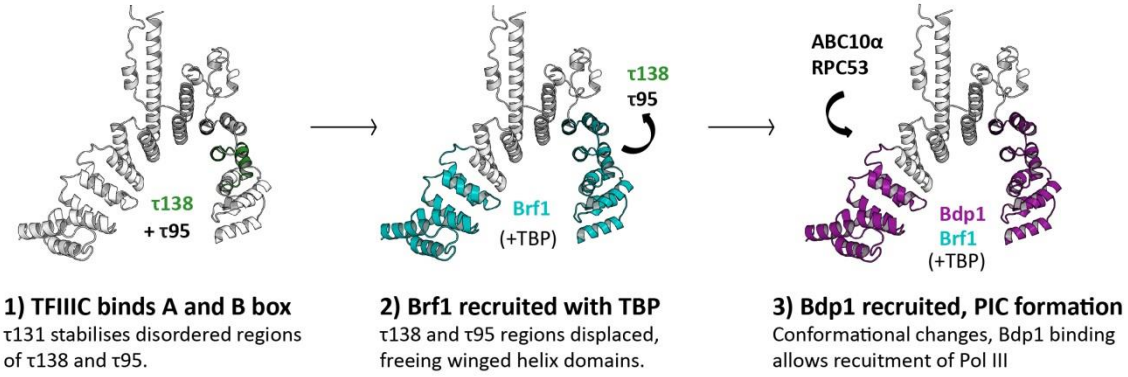

Figure 5.1 **The TPR array as a hub for PIC formation.** The TPR array is coloured according to binding regions of the highlighted subunit.

The TPRs would, in a first step, stabilise the disordered regions of τ95 and τ138, perhaps supporting A box and B box binding. With TFIIIC bound at the promoter elements, TFIIIB can then be assembled. Brf1 and TBP would be recruited first, consistent with the literature, with Brf1 binding at various points over both of the arms. τ138 and τ95 would be displaced, allowing interaction with other components of the PIC. In a final step, Bdp1 is recruited and stabilises the TFIIIB-TFIIIC-DNA complex

[43], allowing recruitment of Pol III, with subunits ABC10α and RPC53 contacting the TPR array. Access to these overlapping sites would require conformational flexibility and time-resolved binding of the various subunits.

In order to confirm the proposed series of events, an integrated structural biology approach will be required. This will require the combination of high resolution structures of various subcomplexes of the PIC together with cryo-electron microscopy reconstructions of larger assemblies of TFIIIC, TFIIIB and Pol III. This could be coupled with photo chemical cross-linking studies that have already been reported [118] and additional lysine-lysine cross-linking in order to arrive at a model for PIC formation.

5.3 Towards a quasi-atomic model of TFIIIC

An ongoing research aim is the elucidation of the molecular architecture of TFIIIC. The work in this thesis will strongly support this aim. Firstly, the high resolution structure of an extended winged helix domain of τ138 that I have presented could be used as a model for the remaining winged helix domains within this subunit (*Figure 5.2*). Secondly, the characterisation of an interface between the τ138 and τ131 subunits will provide additional restraints with which to model the interface between the subcomplexes τA and τB. Finally, the cross-links obtained for TFIIIC will also act as distance restraints for modelling the entire complex. In order to arrive at a consistent model for TFIIIC architecture, an integrated structural biology approach will again be needed. Ideally, a cryo electron microscopy reconstruction of the entire complex would be obtained, within which one can confidently position the high resolution structures (guided by cross-linking and biochemical data). One could use all these restraints to model the complex and then validate the resulting modelled interfaces *in vitro* or *in vivo* by mutations.

One of the big challenges for arriving at a TFIIIC model is the availability of the TFIIIC protein itself for structural and functional studies. As highlighted in this thesis, the purification of endogenous TFIIIC results in a very low yield of protein, after an expensive and time-consuming procedure. Producing TFIIIC recombinantly would be of huge benefit for the project and I have presented work that, with some optimisation still required, should provide a stable protocol for producing sufficient amounts of this protein complex in insect cells for future studies.

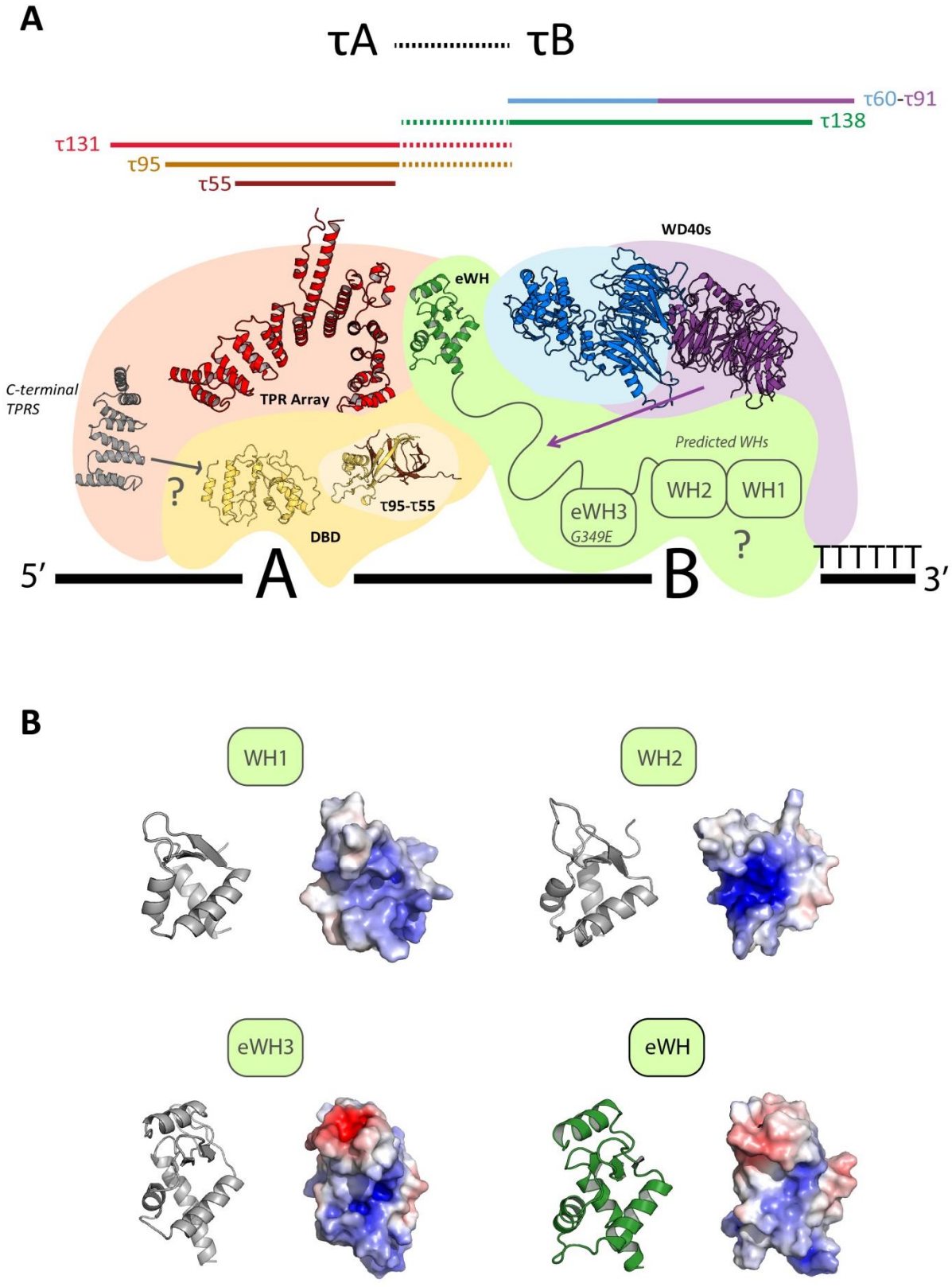

Figure 5.2 **A model for TFIIIC.** (A) Current crystal structures of TFIIIC and their approximate positions within TFIIIC (note the τ55 HPD structure is omitted for clarity). Predicted C-terminal TPRs of τ131 and predicted winged helix domains of τ138 are included. The G349E mutant is highlighted in eWH3. Arrows represent cross-links that have been obtained and could indicate interactions. (B) Models of N-terminal winged helix domains of τ138 with surface electrostatic representations. The eWH domain is included for comparison. The models were generated using MODexplorer [164, 165].

This thesis has presented new information about the link between the τA and τB subcomplexes of TFIIIC. In addition, cross-linking studies together with crystallography and biochemistry have provided novel insights into the architecture of this intriguing protein complex (*Figure 5.2*). The τA and τB subcomplexes are linked through interactions between τ131 and τ138. Cross-linking also indicates a role for τ95. Questions remain over how the remaining TPRs of τ131 function within TFIIIC: cross-linking indicates they bind τ95. Questions also remain over the interaction of τ91 with the winged helix domains of τ138 in the context of B box recognition. Models of the remaining N-terminal winged helix domains of τ138 have been generated using MODexplorer [164, 165]. The third predicted winged helix is likely to also be an extended winged helix based on the structure predictions and so was modelled on the eWH I have solved. Each of the modelled winged helix domains contains a large basic patch which could bind nucleic acids – although I have demonstrated that WH1 (as part of the 1-179 construct) and the eWH do not bind tDNA. A combinatorial binding of several of the winged helix domains may thus be required for B box binding as has been discussed earlier. Future work will need to bring the remaining unsolved parts of this protein together in order to fully understand its architecture and function in type 2 promoter recognition.

6. Materials and Methods

6.1 Results - linking τA and τB

6.1.1 Yeast TFIIIC fermentations

Fermentations for TAP-tagged TFIIIC from *S. cerevisiae* were performed by Emmanuel Poilpre (EMBL protein expression and purification core facility) using a 30 litre fermenter. I also performed two 150 litre fermentations. The strain used – YPL007C – contains a TAP-tag on the τ60 *(TFC8)* subunit and was provided by the Anne-Claude Gavin group at EMBL Heidelberg.

6.1.2 Chemical cross-linking coupled with mass spectrometry

The following protocol was kindly provided by Alexander von Appen, a PhD student in the Martin Beck Group at EMBL Heidelberg.

30 µg (1 µg/µl) of purified TFIIIC complex was cross-linked by addition of an iso-stoichiometric mixture of H12/D12 isotope-coded, di-succinimidyl-suberate (DSS, Creative Molecules). Equal amounts of cross-linker were added ten times every 4 min to a final concentration of 2 mM. The crosslinking reactions were allowed to proceed for 40 min at 37°C and quenched by addition of ammonium bicarbonate to a final concentration of 50 mM for 10 min at 37°C. Cross-linked proteins were denatured using urea and Rapigest (Waters) at a final concentration of 4 M and 0.05% (w/v), respectively. Samples were reduced using 10 mM DTT (30 min at 37°C) and cysteines were carbamidomethylated with 15 mM iodoacetamide (30 min in the dark). Protein digestion was performed first using 1:100 (w/w) LysC (Wako Chemicals GmbH, Neuss, Germany) for 3.5 hours at 37°C then finalised with 1:50 (w/w) trypsin (Promega GmbH, Mannheim, Germany) overnight at 37°C, after the urea concentration was diluted to 1.5 M. Samples were then acidified with 10% (v/v) TFA and desalted using MicroSpin columns (Harvard Apparatus) following the manufacturer's instructions. Cross-linked peptides were enriched using size exclusion chromatography (SEC) as described by [143]. In brief, desalted peptides were reconstituted with SEC buffer (30% (v/v) ACN in 0.1% (v/v) TFA) and fractionated using a Superdex Peptide PC 3.2/30 column (GE) on a Ettan LC system (GE Healthcare) at a flow rate of 50 µl/min. Fractions eluting between 1 and 1.5 ml were evaporated to dryness and reconstituted in 50 µl 5% (v/v) ACN in 0.1% (v/v) FA. Between 2 and 10% of the amount contained in the collected SEC fractions were analysed by liquid chromatography (LC)-coupled tandem mass spectrometry (MS/MS) using a nanoAcquity UPLC system (Waters) connected online to LTQ-Orbitrap Velos Pro instrument (Thermo). Peptides

were separated on a BEH300 C18 (75 mm x 250 mm, 1.7 mm) nanoAcquity UPLC column (Waters) using a stepwise 60 min or 120 min gradient between 3 and 85% (v/v) ACN in 0.1% (v/v) FA. Data acquisition was performed using a TOP-20 strategy where survey-MS scans (m/z range 375-1,600) were acquired in the *Orbitrap* (R = 30,000) and up to 20 of the most abundant ions per full scan were fragmented by collision-induced dissociation (normalised collision energy = 40, activation Q = 0.250) and analysed in the LTQ. In order to focus the acquisition on larger cross-linked peptides, charge states 1, 2 and unknown were rejected. Dynamic exclusion was enabled with repeat count = 1, exclusion duration = 60 s, list size = 500 and mass window ± 15 ppm. Ion target values were 1,000,000 (or 500 ms maximum fill time) for full scans and 10,000 (or 50 ms maximum fill time) for MS/MS scans. All the samples were analysed in technical duplicates. To assign the fragment ion spectra, raw files were converted to centroid mzXML using the Mass Matrix file converter tool and searched using xQuest [166] against a FASTA database containing sequences of the cross-linked proteins. Posterior probabilities were calculated using xProphet [166] and results were filtered accordingly: FDR = 0.05, min delta score = 0.95, MS1 tolerance window of 4 to 7 ppm, ld-score > 25.

6.1.3 Cloning of τ138 constructs

Codon optimised constructs of *S. cerevisiae* τ138 were cloned into pETM30, which introduces an N-terminal, TEV protease-cleavable, 6x-histidine GST affinity tag as described below.

(i) Primer Design

Primers were designed so as to generate a polymerase chain reaction (PCR) fragment that included the desired coding sequence, preceded by an NcoI restriction enzyme site at the 5' end and an Acc65I restriction site at the 3' end. Overhang sequences at the 5' and 3' ends were included to ensure an efficient restriction enzyme cleavage in a later step. Melting temperatures (Tm) for both primers were kept as close as possible and were designed to be between 60 and 65°C.

(ii) PCR and PCR purification

PCR reactions were set up as outlined below:

Reagent	Volume (Final Concentration)
H_2O	37.9 µL
5x HF buffer	10 µL
10 mM dNTP mix	1 µL (200 µM of each dNTP)
Template[1]	0.1 µL (1-10 ng)
Fw Primer	0.25 µL (0.5 µM)
Rv Primer	0.25 µL (0.5 µM)
Phusion[2] Polymerase	0.5 µL

[1] τB MultiBac Vector [2] Phusion High Fidelity DNA Polymerase (Finnzymes)

The reaction was mixed thoroughly and then placed into a PCR machine (Eppendorf Mastercycler Gradient). The reaction was carried out as follows (and as recommended by the manufacturer of Phusion Polymerase). Note that the table below outlines a protocol that was generally applicable for my cloning, occasionally changes to annealing temperature or extension times were required.

Step	Temperature	Time	Cycles
Initial Denaturation	98°C	30 s	1
Denaturation	98°C	10 s	35
Annealing	Primer Tm -5°C	30 s	
Extension	72°C	30 s per kilobase	
Final Extension	72°C	5 min	1

5 µL of the PCR product was analysed on a 1% agarose gel to assess purity. If pure, the remaining PCR product was purified using the Qiaquick PCR Purification Kit (Qiagen). In the case of contaminant bands, the remaining PCR product was ran on a 1% agarose gel and purified using the Qiaquick Gel Extraction Kit (Qiagen). In both scenarios the final PCR product was eluted in 30 µL H_2O with a final concentration of usually 100 ng/µL.

(iii) Restriction Digest

Restriction double digests were set up as follows, with buffer requirements determined by the use of the NEB double digest online tool (www.neb.com).

Reagent	Volume
H_2O	12.5 µL
PCR Product (100ng/ µL)	30 µL
NEB 3 reaction buffer	5 µL
BSA (100x)	0.5 µL
NcoI[1]	1 µL
Acc65I[1]	1 µL

[1]New England Biolabs

The reaction was mixed thoroughly and restriction digests were incubated at 37°C overnight. The reactions were then analysed by running them on a 1% agarose gel. A single band corresponding to the digested product was cut from the gel and the DNA 'insert' was extracted using the Qiaquick Gel Extraction kit (Qiagen). 5 µg of pETM30 vector was digested in parallel by the same methods. Both vector and insert were eluted in 30 µL H_2O for ligation. The vector was dephosphorylated (to reduce the chances of re-ligation without insert) using Antarctic Phosphatase (New England Biolabs) for 1 hour at 37°C before inactivation at 65°C for 15 mins.

(iv) Ligation, Transformation and Positive Clone Selection.

Ligation reactions were set up as follows. A Molar ratio of 1:3 (vector:insert) was desired. As vector stock concentration was constant at 100 ng/µL and insert concentrations were in the

range of 20-40 ng/µL a ratio of 1 µL vector and 3 µL insert was usually sufficient for a successful ligation reaction, based on the base pair sizes of the pETM30 vector and my inserts. DNA concentrations were always checked and adjusted however if this was not the case.

Reagent	Volume (µL)
H_2O	4
pETM30 Vector (100 ng/µL)	1
DNA Insert (20-40 ng/µL)	3
T4 Ligase buffer (10x)	1
T4 DNA Ligase[1]	1

[1]New England Biolabs

A control ligation was also carried out, whereby no insert was included. Ligation reactions were mixed thoroughly and incubated overnight at room temperature. 2 µL of ligation and control reaction were then transformed into 50 µL XL1 Blue electrical competent cells (Stratagene) before incubation for 1 hour at 37°C in 950 µL SOC media. 200 µL of cells were then plated out on to LB-Agar plates containing the antibiotic kanamycin (50 µg/mL) of which cells carrying the ligated pETM30 vector would have resistance to. Plates were incubated overnight at 37°C. A successful ligation can initially be assessed by the presence of colonies on the ligation plate but none, or very few on the control plate. Colonies were picked and inoculated in 2 mL of LB medium containing kanamycin and incubated overnight at 37°C. DNA was purified using the Qiaprep Spin miniprep kit (Qiagen) and a small aliquot tested by restriction digest and agarose gel analysis for the presence of vector and insert. Positive clones were then sent for sequencing (GATC Biotech) to confirm the fidelity of the insert.

6.1.4 Expression and purification of GST-tagged τ138 constructs

Vectors were transformed into BL21 (DE3) Gold *E. coli* cells (Invitrogen) and plated out on LB-Agar plates containing kanamycin (50 µg/mL), and were then incubated overnight at 37°C. 50 mL pre-cultures, supplemented with kanamycin, were grown in TB media till an OD_{600} of 0.8. 12 L of TB media supplemented with kanamycin were then inoculated with the pre-culture and grown at 37°C until an OD_{600} of 0.8. Cells were induced overnight at 18°C with 0.5 mM IPTG. Lysis and purification was the same as for the τ131-τ138 complex (*Section 6.1.8*), with the exception that the GST tags were not cleaved and so after elution the proteins were immediately injected onto a Superdex S200 prep grade gel filtration column (GE Healthcare). The eluted proteins were analysed by SDS-PAGE and then concentrated.

6.1.5 Expression and purification of τ138 '641-681' and '641-693'

The constructs were codon-optimised and were expressed in BL21 (DE3) Gold *E. coli* cells (Invitrogen) using the same strategy as described in *Section 6.1.8*. The cells were harvested by centrifugation and resuspended in lysis buffer (5% glycerol, 50 mM Tris pH 7.5 (at 4°C), 500 mM NaCl, 20 mM imidazole, 4 mM $MgCl_2$, Complete EDTA-free protease inhibitor cocktail (1 tablet/100 mL – Roche), DNase1 (Roche), 2 mM β-mercaptoethanol). After lysis and centrifugation (*Section 6.1.8*) the soluble fraction was incubated with 1 mL nickel-NTA agarose resin (Qiagen) for 1 hour at 4°C. The resin was packed into a 10 mL disposable column (ThermoScientific) and the flow-through collected. The resin was washed with 20 column volumes of binding buffer (50 mM Tris pH 7.5 (at 4°C), 300 mM NaCl, 20 mM imidazole, 2 mM β-mercaptoethanol), 20 column volumes of high salt buffer (50 mM Tris pH 7.5 (at 4°C), 1 M NaCl, 20 mM imidazole, 2 mM β-mercaptoethanol) and then 20 column volumes of binding buffer again. The protein was eluted with 5 column volumes of elution buffer (50 mM Tris pH 7.5 (at 4°C), 300 mM NaCl, 250 mM imidazole, 2 mM β-mercaptoethanol). The eluate was analysed by SDS-PAGE and then dialysed with TEV protease (EMBL protein expression and purification core facility) overnight, in a ratio of 1 µg TEV protease to 100 µg eluate, in 2 L of binding buffer. The resin was stored in elution buffer overnight. The eluate was then re-incubated with the nickel resin that had been cleaned with H_2O and re-equilibrated in binding buffer and incubated at 4°C for 1 hour. The resin was re-applied to the column and the flow-through containing the cleaved protein was collected. The flow-through was then applied to a 5 mL HiTrap SP HP column (GE Healthcare) that had been pre-equilibrated in buffer 'A' (50 mM Tris pH 7.5 (at 4°C), 50 mM NaCl, 2 mM DTT). The column was washed with 20 column volumes of buffer 'A', before a gradient of 40 column volumes into buffer 'B' was applied (50 mM Tris pH 7.5 (at 4°C), 1 M NaCl, 2 mM DTT). The proteins eluted as a single peak and were analysed by SDS-PAGE before being concentrated. The proteins were then dialysed overnight into ITC buffer (20 mM Tris pH 7.5 (at 4°C), 150 mM NaCl, 1 mM β-mercaptoethanol).

6.1.6 GST pull-down experiments

15 µg of GST-tagged τ138 protein and 900 nM τ131 (123-566) protein were incubated with 25 µL Glutathione Sepharose 4B beads in pull-down buffer (50 mM Tris pH 7.5 (at 4°C), 150 mM NaCl, 1 mM DTT and 0.1 % Tween20) at 4°C for 4 hours. The beads were then washed three times with pull-down buffer, before being heated to 100°C in Laemmli sample buffer for 5 min. Samples were analysed by SDS-PAGE.

6.1.7 Isothermal Titration Calorimetry (ITC) experiments

ITC was performed using a MicroCal ITC200 System (GE Healthcare). For experiments mapping the τ131-τ138 interaction, all samples were dialysed into ITC buffer (20 mM Tris pH 7.5 (at 4°C), 150 mM NaCl and 2 mM β-mercaptoethanol). The protein concentration in the cell was 15 μM and the protein concentration in the syringe was 150 μM. Experiments were performed at room temperature.

6.1.8 Co-expression and purification of τ131 (123-566) - τ138 (641-693)

The τ131 (123-566) construct was sub-cloned into the untagged vector pCDF. The vectors were co-transformed into p*RARE *E. coli* cells (Invitrogen) and plated out on LB-Agar plates containing kanamycin (50 μg/mL) (pETM30), streptomycin (50 μg/mL) (pCDF) and chloramphenicol (33 μg/mL) (p*RARE), and incubated overnight at 37°C. 50 mL pre-cultures, supplemented with antibiotics, were grown in TB media till an OD_{600} of 0.8. 12 L of TB media supplemented with antibiotics were then inoculated with the pre-culture and grown at 37°C until an OD_{600} of 0.8. Cells were induced overnight at 18°C with 0.5 mM IPTG. Cells were harvested the next day by centrifugation and resuspended in the appropriate volume of lysis buffer (5% glycerol, 50 mM Tris pH 7.5 (at 4°C), 150 mM NaCl, 2 mM $MgCl_2$, Complete EDTA-free protease inhibitor cocktail, DNase1 (Roche), 2 mM DTT) for the weight of the pellet. All following steps were carried out at 4°C. The cells were lysed by two rounds of homogenisation (Avestin Emulsiflex-C3). The lysate was cleared by centrifugation at 20,000 rpm for 1 hour in a JA 25.50 Rotor (Beckmann). The soluble fraction was then incubated overnight on an Aekta PRIME using a GST preparative FF 16/60 column (GE Healthcare) which had been pre-equilibrated in lysis buffer, at a flow rate of 0.3 mL/min. The next day, the column was equilibrated with binding buffer (50 mM Tris pH 7.5 (at 4°C), 150 mM NaCl, 2 mM DTT). When the UV absorbance had decreased to baseline, the column was washed with 10 column volumes of high salt buffer (50 mM Tris pH 7.5 (at 4°C), 750 mM NaCl, 2 mM DTT). The column was then equilibrated with binding buffer until the conductivity had reached baseline. Elution was carried out at 1 mL/min using elution buffer (50 mM Tris pH 7.5 (at 4°C), 150 mM NaCl, 10 mM reduced glutathione, 2 mM DTT). 3 mL Fractions were collected and fractions which corresponded to a peak in absorbance were analysed by SDS-PAGE and pooled accordingly. The pooled eluate was dialysed together with GST-tagged TEV protease (EMBL protein expression and purification core facility) overnight, in a ratio of 1 μg TEV protease to 100 μg eluate, in 2 L of binding buffer. The column was washed with H_2O, 6 M guanidine hydrochloride and then H_2O again. The column was then equilibrated with binding buffer in

preparation for the second GST-affinity step. After overnight TEV protease cleavage, the eluate was circulated over the column at 0.3 mL/min for several hours. The flow-through was collected (cleaved protein) and concentrated to 8 mL for gel filtration. Gel filtration was carried out using a 320 mL Superdex 200 preparative grade column (GE Healthcare) that had been pre-equilibrated in gel filtration buffer (20 mM Tris pH 7.5 (at 4°C), 150 mM NaCl, 2 mM DTT). The protein was injected using a 10 mL loop. 2mL Fractions were collected and fractions which corresponded to a peak in absorbance were analysed by SDS-PAGE.

6.1.9 Selenomethionine-incorporated τ131 (123-566) - τ138 (641-693) expression

Expression was carried out in BL21 (DE3) Gold (Invitrogen) by feedback inhibition of methionine synthesis. This protocol has been established previously in our lab. A 5 mL pre-culture was grown in LB medium till an OD_{600} of 0.6. The culture was harvested and the pellet was resuspended and washed twice with M9 medium (1x M9 salts[1], 2 mM $MgSO_4$, 20% glucose, 1 µg/mL thiamine, 1x trace elements[2], 0.1 mM $CaCl_2$). 50 mL M9 pre-culture (per litre of cells) was grown overnight. Pre-culture was added to 1 L of M9 medium and cells were grown till an OD_{600} of 0.6. Several amino acids were then added to ensure feedback inhibition of methionine synthesis. 100 mg/L of lysine, phenylalanine, threonine and 50 mg/L of isoleucine, leucine and valine were added. Finally, 50 mg/L selenium-methionine was added. The cells were incubated for 20 minutes, before being induced overnight at 18°C with 0.5 mM IPTG. Purification of selenomethionine-incorporated τ131 (123-566) - τ138 (641-693) complex was as with the native protein (*Section 6.1.8*). A lower overall yield was obtained in comparison to the native complex but sufficient quantities for crystallisation were obtained.

[1]M9 salts (10x) – Na_2HPO_4 (70 g), KH_2PO_4 (30 g), NaCl (5 g), NH_4Cl (5 g).

[2]Trace elements – EDTA (5 g), $FeCl_3$ (0.83 g), $ZnCl_2$ (0.084 g), $CuCl_2x6H_2O$ (0.013 g), $CoCl_2x6H_2O$ (0.010 g), boric acid (0.010 g), $MnCl_2x6H_2O$ (1.6 mg).

6.1.10 Cloning and purification of τ131 point mutants

Cloning of the τ131 point mutants was carried out using the QuikChange Lightning Kit (Agilent Technologies). The PCR protocol was carried out following the manufacturer's guidelines. Primers introducing the specific point mutations were designed using an online tool (https://www.genomics.agilent.com/primerDesignProgram.jsp) from Agilent Technologies. After PCR, 1 µL of DpnI was added to each reaction and incubated for 15 min at 37°C in order to digest parental template. 10 µL of the DpnI-treated mixture was then transformed by heat-shock at 42°C into chemically-competent XL10-gold *E. coli* cells (Agilent Technologies). Cells

were then plated out onto LB-Agar plates containing kanamycin (50 µg/mL). Colonies were picked and inoculated in 2 mL of LB medium containing kanamycin and incubated overnight at 37°C. DNA was purified using the Qiaprep Spin miniprep kit (Qiagen). Clones were then sent for sequencing (GATC Biotech) to confirm the presence of the mutation.

6.2 Results – structural and functional studies on τ138

6.2.1 Small-scale expression and purification of τ138 constructs

Owing to the large number of constructs generated for this project, I tried wherever possible to increase the throughput of expression and solubility testing. In each 'round' of trials I took 6 constructs and purified them. I expressed the constructs using 1 L of BL21 (DE3) Gold cells (Invitrogen) in TB. Cells were grown at 37°C to an OD_{600} of 0.8 and expression was induced overnight using 0.5 mM IPTG at 18°C. Cells were harvested by centrifugation and re-suspended in lysis buffer (5% glycerol, 50 mM Tris pH 7.5 (at 4°C), 500 mM NaCl, 2 mM $MgCl_2$, Complete EDTA-free protease inhibitor cocktail (Roche), DNaseI (Roche), 2 mM DTT). Cells were lysed and then centrifuged (*Section 6.1.8*). The soluble fraction was incubated with Glutathione Sepharose 4B resin (GE Healthcare) overnight at 4°C. The next day, the samples were applied to a disposable column (ThermoScientific) and the flow-through was collected. The resin was washed with 20 column volumes of binding buffer (50 mM Tris pH 7.5 (at 4°C), 150 mM NaCl, 2 mM DTT) before elution was carried out with 5 column volumes of elution buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM reduced glutathione, 2 mM DTT). Samples were analysed by SDS-PAGE.

6.2.2 Expression and purification of τ138 (546-641)

The codon-optimised sequence encoding for τ138 (546-641) was cloned into pETM30 (*Section 6.1.3*) and transformed into BL21 (DE3) Gold *E. coli* cells (Invitrogen). Cells were plated out on LB-Agar plates containing kanamycin (50 µg/mL) and incubated at 37°C overnight. A 50 mL pre-culture, supplemented with kanamycin was grown in TB till an OD_{600} of 0.8. 12L of TB supplemented with kanamycin was then inoculated with the pre-culture and grown at 37°C until an OD_{600} of 0.8. Cells were induced overnight at 18°C with 0.5 mM IPTG. The cells were harvested and resuspended in lysis buffer (5% glycerol, 50 mM Tris pH 7.5 (at 4°C), 500 mM NaCl, 20 mM imidazole, 4 mM $MgCl_2$, Complete EDTA-free protease inhibitor cocktail (Roche), DNaseI (Roche), 2 mM β-mercaptoethanol). After lysis and centrifugation (*Section 6.1.8*) the soluble fraction was incubated with 2 mL nickel-NTA agarose resin (Qiagen) for 1 hour at 4°C. The resin was packed into a 10 mL disposable column (ThermoScientific) and the flow-through

collected. The resin was washed with 20 column volumes of binding buffer (50 mM Tris pH 7.5 (at 4°C), 300 mM NaCl, 20 mM imidazole, 2 mM β-mercaptoethanol), 20 column volumes of high salt buffer (50 mM Tris pH 7.5 (at 4°C), 1 M NaCl, 20 mM imidazole, 2 mM β-mercaptoethanol) and then binding buffer again. The protein was eluted with 5 column volumes of elution buffer (50 mM Tris pH 7.5 (at 4°C), 300 mM NaCl, 250 mM imidazole, 2 mM β-mercaptoethanol). The eluate was analysed by SDS-PAGE and then dialysed with TEV protease (EMBL protein expression and purification core facility) overnight, in a ratio of 1 µg TEV protease to 100 µg eluate, in 2 L of binding buffer. The resin was stored in elution buffer overnight. The eluate was then re-incubated with the nickel resin, which had been cleaned with H_2O and re-equilibrated in binding buffer, and incubated at 4°C for 1 hour. The resin was re-applied to the column and the flow-through containing the cleaved protein was collected. The flow-through was concentrated to 4 mL for gel filtration. The protein was injected using a 5 mL loop onto a Superdex 75 preparative column (GE Healthcare) that had been pre-equilibrated in gel filtration buffer (20 mM Tris pH 7.5 (at 4°C), 150 mM NaCl, 2 mM DTT). 2 mL Fractions were collected and fractions which corresponded to a peak in absorbance were analysed by SDS-PAGE.

6.2.3 Expression and purification of τ138 (1-179)

Methods and buffers were exactly the same as for τ138 (546-641). See *Section 6.2.2* for details.

6.2.4 EMSA of τ138 (546-641) and τ138 (1-179)

(i) Labelling of oligonucleotides

In a first step, the HPLC-purified, 66-base single-stranded (ss) $tDNA^{Glu}$ oligonucleotides (forward and reverse - 5' - CGA TAT AGT GTA ACG GCT ATC ACA TCA CGC TTT CAC CGT GGA GAC CGG GGT TCG ACT CCC CGT ATC - 3') were radioactively labelled. A 10 µL reaction was set up which included 3 µg of the oligonucleotide, 10 U/ µL of T4 PNK (New England Biolabs) and 3000 Ci/mmol of $ATP\gamma P_{32}$. The reaction was incubated at 37°C for 1 hour. Labelling was stopped by incubating the reaction at 95°C for 3 min. A 15% denaturing (8M Urea) SDS-PAGE was pre-run for 1 hour at 25 W before samples were loaded and run for approximately 1 hour at 25 W. The gel was placed onto plastic film and auto-radiographed to identify labelled oligonucleotides. The bands were removed, placed into a 1.5 mL Eppendorf tube and eluted with 300 µL elution buffer (0.6 M $AcNH_4$, 1 mM EDTA, 0.1% SDS) overnight. The eluted DNA was then precipitated by adding 3 volumes of 100% ethanol and 1 µL of glycogen (20 mg/mL)

at -20°C for 30 min. Samples were then centrifuged at 4°C for 20 min, before being resuspended in 20 µL H_2O.

(ii) EMSA

1 µL of labelled 66-base $tDNA^{Glu}$ (forward) in one reaction and 1 µL + 1 µL of labelled 66-base $tDNA^{Glu}$ (forward & reverse) in a second reaction were incubated at 95°C for 2 min, thus allowing annealing of forward and reverse to double-stranded (ds) DNA. 12 µL of EMSA buffer (20 mM Tris pH 7.5 (at 4°C), 150 mM NaCl, 1 mM $MgCl_2$, 2 mM DTT) was added to cool the reactions to room temperature. 1 µL of ssDNA or dsDNA was added to a tube containing τ138 (546-641) or τ138 (1-179) at a specific concentration (e.g. 500 µM). The tubes were incubated for 30 min at room temperature. 2 µL of native loading dye was added to each sample and the samples were loaded onto a native 4.5% acrylamide gel (4.5% acrylamide, 10% sucrose, 20 mM Tris HCl pH 8.0) which had been pre-run for 2 hours at 100 V. The gel was run at 120 V for 2 hours in a Tris-glycine running buffer. After migration, the gel was dried and auto-radiographed with X-ray film (Biomax, MR-film, Kodak).

6.2.5 Cross-linking TFIIIC bound and unbound to tDNA

Cross-linking reactions were performed as in *Section 6.1.2*. For TFIIIC bound to tDNA, the TFIIIC was first incubated with the 66 base-pair $tDNA^{Glu}$ oligonucleotide that had been shown to bind TFIIIC by EMSA. 30 µg (1 µg/µl or 2 µM) of TFIIIC was incubated with an excess (5 µM) of double-stranded $tDNA^{Glu}$.

6.3 Results – producing recombinant TFIIIC

The protocols included here are based on the Handbook 'Eukaryotic Expression Facility at EMBL Grenoble' from Frederic Garzoni and Imre Berger, and on published protocols [159, 162].

6.3.1 Virus generation for Sf21 insect cell expression

(i) Bacmid Generation

2 µL of plasmid (pPBac τA or pKL τB) was transformed into electrically competent EMBacYFP cells. 900 µL of SOC media was added and the cells were incubated overnight at 37°C. *This longer incubation before plating ensures higher success for the integration of the plasmid into the Tn7 sites in the Bacmid backbone.* Cells were then plated out onto LB-Agar plates containing kanamycin (50 µg/mL), tetracycline (12 µg/mL), gentamycin (10 µg/mL), ampicillin (100 µg/mL), IPTG (1 mM) and BluOGal (100 µg/mL) in serial dilutions. Plates were incubated at 37°C for 36-48 hours so that there was sufficient time to make a clear distinction between

blue and white colonies. *White colonies have successfully integrated the plasmid into the Bacmid DNA backbone*. 4 white colonies and 1 blue colony were re-streaked onto plates and incubated overnight at 37°C as a control. 4 white colonies were then picked and inoculated in 2 mL of LB medium containing antibiotics and incubated overnight at 37°C. The Bacmid DNA was isolated using the Qiaprep Spin miniprep kit (Qiagen) and stored at -20°C until required.

(ii) V_0 and V_1 Virus Generation

All steps are carried out in a cell-culture hood at 27°C.

One well of a six-well plate was seeded with 3 mL of Sf21 cells at a density of 0.5 x 10^6 and allowed to settle for at least 30 min. Two Eppendorf tubes were prepared: one containing 10 µL of Insectogene (Biontex) combined with 90 µL of Sf900 III SFM media (Invitrogen) and the second tube containing 2.5 µg Bacmid DNA in a total volume of 100 µL media. The two tubes were combined and left for 30 min to form virus particles. The supernatant from each well was removed, leaving a layer of Sf21 cells. 800 µL of media was added to each Bacmid DNA-Insectogene tube and then mixed, before being pipetted onto the cells. For a control well, 2 mL of media was added, in order to test for contamination from the cell stock. After 4 hours, the supernatant was removed and 2 mL of fresh media was added. V_0 virus was harvested after 60 hours (i.e. the supernatant was removed from the well and spun down to remove any remaining cell debris). The V_0 virus was then added to 25 mL of Sf21 cells at a density of 0.5 x 10^6. Cells were checked every 24 hours for cell proliferation arrest – i.e. when cells stop doubling. 72 hours after proliferation arrest (or when the YFP levels plateaued) cells were harvested by centrifugation at 800*g* for 10 min. V_1 virus supernatant was removed and stored at 4°C.

6.3.2 Expression and purification of τA

6 L of Sf21 cells were infected with V_1 virus at a dilution of 1 in 10000. When YFP/CFP levels had plateaued the cells were harvested at 800*g* for 20 min. Cell pellets were re-suspended in lysis buffer (25 mM Tris pH 7.5 (at 4°C), 150 mM NaCl, 10 mM imidazole, 1 mM $MgCl_2$, Complete EDTA-free protease inhibitor cocktail (Roche), DNaseI (Roche), 1 mM β-mercaptoethanol) (100 mL per 2L cell pellet). For lysis, cells were subjected to two rounds of flash-freezing in liquid nitrogen, followed by thawing. The lysate was then centrifuged at 20,000 rpm for 90 mins. The soluble fraction was circulated for at least 4 hours (or overnight) at 4°C on a TALON nickel affinity column (Clontech) which had been pre-equilibrated in lysis buffer. The column was washed with 20 column volumes of high salt buffer (25 mM Tris pH 7.5 (at 4°C), 1 M NaCl, 10 mM imidazole, Complete EDTA-free protease inhibitor cocktail (Roche),

DNaseI (Roche), 1 mM β-mercaptoethanol), before being re-equilibrated in binding buffer (25 mM Tris pH 7.5 (at 4°C), 150mM NaCl, 10 mM imidazole, Complete EDTA-free protease inhibitor cocktail (Roche), DNaseI (Roche), 1 mM β-mercaptoethanol). A step elution was performed using elution buffer (25 mM Tris pH 7.5 (at 4°C), 150 mM NaCl, 250 mM imidazole, 1 mM $MgCl_2$, Complete EDTA-free protease inhibitor cocktail (Roche), DNaseI (Roche), 1 mM β-mercaptoethanol). 3 mL fractions were collected and fractions which corresponded to a peak in absorbance were analysed by SDS-PAGE and pooled accordingly. The eluate was concentrated to 4mL for gel filtration. Gel filtration was carried out using a 120 mL Superose 6 column (GE Healthcare) that had been pre-equilibrated in gel filtration buffer (20 mM Tris pH 7.5 (at 4°C), 150 mM NaCl, 2 mM DTT). The protein was injected using a 5 mL loop. 3 mL fractions were collected and fractions which corresponded to a peak in absorbance were analysed by SDS-PAGE.

6.3.3 Co-expression and purification of τA + τB (TFIIIC)

V_1 Virus of both pPBac τA and pKL τB were used to infect 500 mL of Sf21 cells at a dilution of 1 in 5000. When YFP/CFP/mCherry levels had plateaued the cells were harvested at 800*g* for 20 min. Cell pellet was re-suspended in 70 mL of lysis buffer buffer (10% glycerol, 50 mM Tris pH 7.5 (at 4°C), 200 mM NaCl, 2 mM $CaCl_2$, Complete EDTA-free protease inhibitor cocktail (Roche), DNaseI (Roche), 2 mM β-mercaptoethanol). For lysis, cells were subjected to sonication (40% power, 90s – 10s on, 15 s off). The lysate was then centrifuged at 20,000 rpm for 90 mins. The soluble fraction was incubated with 100 µL CBP resin for 2 hours at 4°C. The resin was then applied to a 2 mL disposable column (ThermoScientific) and washed sequentially with 30 column volumes of binding buffer (W1) (50 mM Tris pH 7.5 (at 4°C), 200 mM NaCl, 0.1 mM $CaCl_2$, 2 mM β-mercaptoethanol), high salt buffer (50 mM Tris pH 7.5 (at 4°C), 1 M NaCl, 0.1 mM $CaCl_2$, 2 mM β-mercaptoethanol) and then back to binding buffer (W2). See also *Figure 4.7*. The protein was eluted from the beads in 400 µL elution steps using elution buffer 1 (50 mM Tris pH 7.5 (at 4°C), 200 mM NaCl, 2 mM EGTA, 2 mM β-mercaptoethanol) and then elution buffer 2 (50 mM Tris pH 7.5 (at 4°C), 200 mM NaCl, 5 mM EGTA, 2 mM β-mercaptoethanol). The eluate was concentrated to 200 µL for analytical gel filtration. Gel filtration was carried out using a 24 mL Superose 6 10/300 column (GE Healthcare) that had been pre-equilibrated in gel filtration buffer (20 mM Tris pH 7.5 (at 4°C), 150 mM NaCl, 2 mM DTT). The protein was injected using a 500 µL loop. 500 µL fractions were collected and fractions which corresponded to a peak in absorbance were analysed by SDS-PAGE.

6.3.4 EMSA of τA

EMSA experiments were performed as in *Section 6.2.4*, but instead using a 256 bp $tDNA^{Glu}$ which contained the same 66 base sequence as used in *Section 6.2.4* but was flanked by additional sequence.

6.4 X-ray crystallography

As many of the experiments described in this thesis are centred on X-ray crystallography experiments, this section will briefly describe some relevant theoretical and practical background to this technique.

In order to obtain atomic structures, one first requires protein crystals. This step is regarded as the major bottleneck in X-ray crystallography. If crystals are obtained, then they must produce interpretable diffraction patterns upon exposure to X-rays. These diffraction patterns contain recorded intensities in the forms of 'spots' or 'reflections' that represent the electron density in reciprocal space. However, in order to reconstruct electron density in real space, one still requires phase information (the position of electrons within the crystal), which is lost in the diffraction experiment. This 'phase problem' can be solved by molecular replacement or experimental phasing. Some aspects of the X-ray crystallography pipeline that are relevant for this thesis are discussed below.

6.4.1 Preparing protein for crystallisation

To begin crystallisation trials a protein purity of > 90% (as assessed by SDS-PAGE) is desired. Secondly, proteins should elute from a gel filtration column as a single, mono-disperse, peak. A highly pure, homogenous protein solution is desired for crystallisation, as this will increase the probability of protein molecules forming an ordered lattice. It is not possible to determine the optimal protein concentration for crystallisation *a priori*. When setting screens, I trialled two concentrations: the highest achievable, and then half of this. Although there are no fixed rules for which concentration to reach, it is usually optimal to test as high concentration as is reachable in order to increase the probability of the protein solution reaching super-saturation and for crystals to form.

6.4.2 Setting up crystallisation experiments

A number of sparse-matrix screenings (such as JCSG+, Wizard) were used during initial crystallisation trials. These 96-well screens contain a wide-range of crystallisation solutions that can be tested on a nanolitre scale using an automated pipetting robot. For each screen,

every crystallisation 'drop' required 100 nL of protein solution, which was mixed with 100 nL of crystallisation solution. Screens were usually stored at 20°C but trials at 4°C were also attempted in more challenging cases. If crystals were obtained (usually after 2-3 days in the experiments presented in this thesis) then a refinement screen was set up around the successful condition. For refinement, a 96-well plate was again used, with the initial condition screened by changing two variables. After an assessment of these results, manual screens were established. Here, 24-well screens were set up by hand, with 1 µL of protein solution mixed with 1 µL of the crystallisation solution. The setup used in all manual crystallisation trials within this thesis was the hanging-drop method.

6.4.3 Preparing crystals for data collection

To minimise radiation damage brought about by exposure to X-rays (especially at synchrotron sources), data collection was carried out at liquid nitrogen temperature. This meant that crystals had to be flash-frozen in liquid nitrogen after they were collected (or 'fished'). Crystals therefore had to be 'cryo-protected', to ensure that water in the protein crystals did not form ice upon freezing, which can significantly affect the quality of the subsequent diffraction. Data collection of the highest resolution datasets obtained in this thesis were carried out at synchrotron X-ray sources, using the rotation method. Crystals were rotated in 0.1° steps whilst exposed to X-rays with resulting diffraction patterns recorded.

6.4.4 Data processing pipeline

The major steps involved in processing data and solving a structure in this thesis are summarised in *Figure 6.1*. All data processing of diffraction images in this thesis was carried out using the program XDS [144]. The program begins by indexing the reflections in the *XDS* step, integrating the recorded intensities and indicating to the user the probable space group which best describes the recorded positions of the reflections. This step was repeated by modifying the input file with the calculated unit cell parameters and including the space group details. The parameters are refined and the cycle is repeated until convergence. The data was scaled in *XSCALE* and statistics describing the quality of the data are reported. In this thesis, particular attention was paid to the values of R_{merge}, which indicates the precision of data and whether the correct space group has been assigned. Resolution limits were determined at this step, with attention paid to the values of I/σI and $CC_{1/2}$ [167] in the highest resolution shell. Ultimately, the correct cut-off can only be truly be determined based on the final values of R_{work}/R_{free} at the end of refinement, and many alternative cut-offs may need to be tested. In this thesis, the I/σI was cut at ~2 and $CC_{1/2}$ at ~0.6 in the highest resolution shell. After scaling,

XDSCONV was used to convert the reflections to a necessary format for later processing and to generate a set of test reflections that were used to calculate R_{free} values during refinement. Finally, the file name was changed to .mtz for downstream phasing steps.

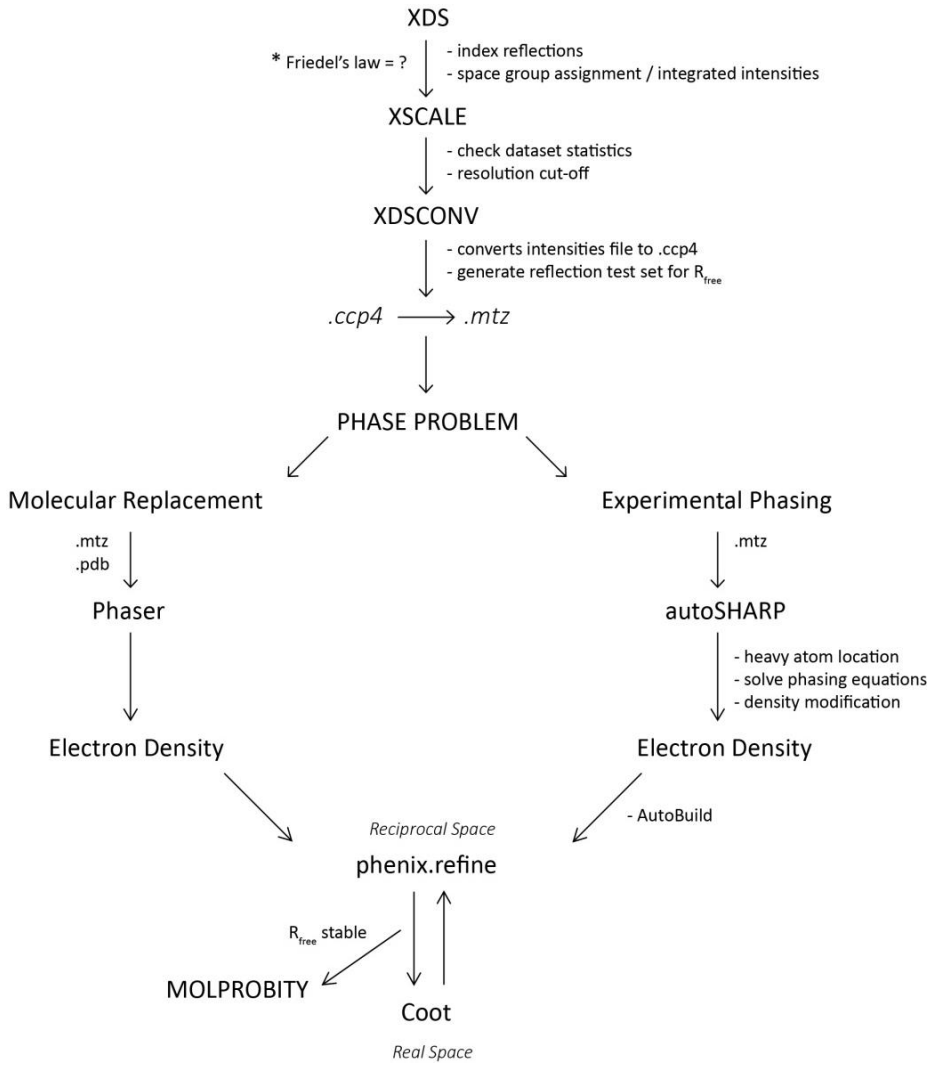

Figure 6.1 **Data processing pipeline.** The key steps in processing the diffraction data and phasing performed in this thesis. See main text for details of the various steps. *At the XDS step, Friedel's law must be set to either 'TRUE' or 'FALSE' in the input file. FALSE was selected if anomalous differences had been measured.

With intensities determined, phases were then required in order to complete the Fourier equations and calculate the electron density. Details of the phasing techniques used are described in the relevant results sections.

6.4.5 Model building, refinement and validation

In this thesis, model building was carried out using the program Coot [146] and refinement was carried out using Phenix [147]. For molecular replacement, a model is already present with

which to modify based on the interpreted electron density difference map. In the case of experimental phasing, a model must be built *de novo*. For the sulphur-SAD experiment performed in this thesis, the AutoBuild [151] feature in Phenix was implemented in order to provide a starting model that could be further modified (aided by the high resolution of the final dataset). During refinement, several parameters or restraints were included in order to ensure that the model was not over-fitted to the data and to reduce bias. These parameters were selected based on the quality of the data. For example, for low-resolution data it was not always possible to interpret side-chain density. As a consequence, refining B-factors (for example) for each individual atom was not suitable, whereas with high resolution data this strategy was appropriate. Selecting other strategies and parameters is highly dependent on the quality of data and was varied depending on the stage of the refinement.

In order to assess the progress of refinement, the R_{work} and R_{free} values were checked after each refinement cycle. R_{work} is a measure of the fit between the data and the model and should decrease after each refinement cycle. However, this value will keep decreasing if the model is over-fitted and must be cross-validated with the R_{free} value. The R_{free} value is a more accurate measure of whether the model is actually improving and if the model agrees with the original data, this value is generated by cross-validating with a test set of reflections that were set aside by *XDSCONV* earlier in the pipeline. In this thesis, once the R_{free} value had stabilised, the refinement was abandoned and stereochemistry and geometry was validated using MolProbity [168].

7. APPENDIX

7.1 TAP-tag purification of TFIIIC

The protocol for purifying TAP-tagged TFIIIC has been outlined in the Doctoral thesis of Nicholas Taylor and was based on the purification of RNA polymerase I [39]. A detailed protocol is provided here and is summarised in *Figure 2.1*.

Day 1: Preparations

- See Buffer tables 1 & 2
- Prepare all buffers – but do not add inhibitors/reducing agents until Day 2
- Check conductivity of all buffers
- Prepare 0.1 M PMSF (870 mg in 50 mL ethanol)
- Place bead beater (+ glass beads) in the cold room, along with blade homogeniser, tubes for TEV protease reaction, 250 mL measuring cylinder, 4 x 1 L beakers, 2 L bottle
- Place frozen cell paste in the cold room in a bucket covered in foil

Preparing the Heparin Column
Pump wash
Manual run at 0.3 MPa pressure limit with water
Run Stored Method, System No. 20: Tubes A and 8 go into 5 L water. Tube B into 'Hep Clean' Buffer. Run method at least twice.

Day 2: Cell lysis and applying lysate to heparin column

- Keep track of wet weights/dry weights
- Remember to add last minute additives to buffers (e.g. protease inhibitors, reducing agents) before using them
- Wash cells with 'Cell wash' buffer (for 100*g* add 100 mL) and homogenise
- Centrifuge for 7000 rpm, 10 min. This removes wet mass. Measure the dry pellet and re-suspend in an equal volume of 'Lysis' buffer (prepare more lysis buffer than required as more is needed for the re-lysis)
- Homogenise with blade homogenizer
- Fill ¾ of the lysis chamber with glass beads (250 ml), fill chamber to the top with resuspended cells (~200 ml) and add 100 mg Pefablock
- Assemble the bead beater and cool the chamber with ice / 'Cooling' buffer
- Beat 9 times for 45 s with 90 s pauses in-between
- Filter beads through metallic mesh (pore size 0.3 mm). Put filtered lysate into JA-14 tubes, recover beads and place back into lysis chamber
- Centrifuge with JA-14, for 1 hour at 14,000 rpm
- Recover and filter the supernatant – begin applying to heparin column
- Re-suspend un-lysed cells and repeat lysis as before
- Load lysate from second lysis on equilibrated heparin column (see below)
- Wash column overnight with 1 L 'Hep Bind' at 2-4 ml/min until UV is almost zero

Equilibration of heparin column
Wash AKTA pumps. A1, A8 & B tubes with buffers 'Hep A', ' Hep bind' & 'Hep B' respectively
Manual run with 100% B, 50mL, then 0% B 50mL – ensure you are on WASTE
Return to LOAD - equilibrate with 1 L Hep 'Bind' (A8 tube), until conductivity is flat

Day 3: Heparin elution and IgG binding steps

- Equilibrate column in 'Hep A' Buffer
- Elute with gradient 0-100% 'Hep B' for 500 ml at 2-4 ml/min, collecting 1 fraction of ~900 ml. The dead volume is about 200 mL

Bead preparation for IgG binding
Wash IgG beads (5 mL from 10 mL slurry) with H_2O to remove ethanol (when washing always use 1 column volume)
Then wash in the following order: 0.5 M Tris (pH 7.4), Acetic acid/NaOH (pH 3.4), 0.5M Tris (pH 7.4), Acetic acid/NaOH (pH 3.4), 0.5 M Tris (pH 7.4)
To make Acetic Acid/NaOH pH 3.4 (500 mL): fill up with H_2O to 479 mL Acetic acid: 14.3 mL NaOH: 6.25 mL

- Heparin eluate must be diluted with 'Hep A' buffer to reduce salt concentration for bead binding. Add a volume equivalent to the length of the peak from when B is at 100% to end of elution peak (about 500 mL)
- Add IgG resin to the heparin elution & incubate overnight at 4 °C with mild shaking

Day 4: TEV protease cleavage

- Pass incubated eluate through a yellow column, collecting the flow-through
- Wash with 200 ml 'IgG bind' buffer, re-suspend beads with 20 ml 'IgG bind' buffer and transfer to 2 clean tubes.
- Add 1 mL of TEV protease (EMBL protein expression and purification facility) to each tube.
- Top-up tubes with 'IgG bind' buffer (each tube holds approx. 15 mL)
- Incubate overnight on rotating wheel
- Clean MonoQ column in preparation for Day 5: 20 column volumes H_2O, 0.5 M Tris, water, 0.5 M NaOH, H_2O

Day 5: IgG 'elution' / MonoQ / concentration

- Add tube contents to yellow column, and collect FT!
- Pass 4x 'IgG elution' buffer in 5 mL steps through beads
- Add 'IgG elution' buffer to a final volume of 150 mL
- MonoQ preparation (use 10/100 column – 8 mL): wash column with H_2O, 'MonoQ A', 'MonoQ B', 'MonoQ A'
- Inject IgG eluate with the 50 mL superloop – collect fractions immediately
- Elute with a 350 mL gradient against 'MonoQ B'
- Collect 1 mL fractions
- TFIIIC elutes at ~ 18% B, ~ 36 mS/cm

Concentrate TFIIIC on a 30 kDa cutoff spinning concentrator (5mL)
Analyse fractions by 10-12% SDS-PAGE

Buffer Table 1 : Lysis & Heparin

Stocks	**Cell wash** (2 L)	**Lysis** (2 L)	**Cooling** (4 L)	**Hep clean** (5 L)	**Hep bind** (2.5 L)	**Hep A** (2 L)	**Hep B** (2 L)
1 M Tris pH 8	100 ml	500 ml	-	-	125 ml	100 ml	100 ml
glycerol	-	800 ml	-	-	1100 ml	-	-
Ammonium sulphate	52.86 g	52.86 g	-	-	66.08 g	52.86 g	264.28 g
0.5 M EDTA	-	4 ml	-	-	5 ml	2 ml	2 ml
5M NaCl	-	-	1200 g	585 g	-	-	-
β-EtSH	-	96µl / 100 mL	-	-	0.7 ml	150 µl	150 µl
'Complete' Tablets*	-	1 / 25 ml	-	-	-	-	-
Pefablock	-	1 mg/ml	-	-			
0.1 M PMSF	-	-	-	-	10 ml	10 ml	10 ml
DTT	-	-	-	-	-	-	-
Conductivity (ms/cm)	-	*12.23*	-	-	*8.28*	*38.2*	*138.4*

* (- EDTA) (Roche) Pre-dissolve in Tris

Ingredients in shaded boxes should be added on the day that the buffer is used

Buffer Table 2 : IgG and MonoQ

Stocks	**IgG bind** (2 L)	**IgG elution** (2 L)	**MonoQ A** (4 L)	**MonoQ B** (5 L)
1 M Tris pH 8	25 ml	25 ml	40 ml	40 ml
glycerol	100 ml	-	-	-
Ammonium sulphate	6.61 g	2.112g	-	-
0.5 M EDTA	500 µl	500 µl	1 ml	1 ml
5M NaCl	-	-	20 ml	200 ml
β-EtSH	75 µl	75 µl	-	-
'Complete' Tablets*	-	-	-	-
Pefablock	-	-	-	-
0.1 M PMSF	0.1mg/ml	0.1mg/ml	0.1mg/ml	0.1mg/ml
DTT	-	-	1.54g	1.54g
Conductivity (ms/cm)	*12.76*	*10.7*	*14.16*	*90.4*

* (- EDTA) (Roche) Pre-dissolve in Tris

Ingredients in shaded boxes should be added on the day that the buffer is used

7.2 Tables of cross-linked peptides

Appendix Table B1 – TFIIIC inter cross-links without tDNA

Peptide*	Prot. 1	Prot. 2	Lys. 1	Lys. 2	ld-score**
MLTNLDS**K**GNLSSR-NM**K**SSSPGSSLGQK	τ60	τ91	124	142	46.39
ISAST**K**DLSK-DSI**K**SLK	τ138	τ95	896	138	43.28
ISAST**K**DLSK-SL**K**DSNK	τ138	τ95	896	141	40.11
KSTSVNVSDGK-**K**TGQLAR	τ138	τ131	698	242	39.36
MLTNLDS**K**GNLSSR-NST**K**NMK	τ60	τ91	124	139	39.35
KTAGLLSPTEENETTNAGQSK-V**K**EAFK	τ55	τ95	359	66	38.54
ASDDAIS**K**K-**K**TGQLAR	τ138	τ131	697	242	38.51
DHTGDKEEVQSE**K**IYR-V**K**EAFK	τ55	τ95	419	66	38.33
NSSATQ**K**SLR-GFDELG**K**SR	τ91	τ138	598	433	37.53
F**K**NSTFDSLLAAEK-I**K**ASDDAISK	τ131	τ138	412	689	36.91
MLTNLDS**K**GNLSSR-NTVQ**K**LR	τ60	τ91	124	130	35.78
E**K**SYNIPIDIR-I**K**ASDDAISK	τ131	τ138	426	689	35.28
DHTGD**K**EEVQSEK-AI**K**MCGGIEK	τ55	τ95	412	57	34.50
DHTGD**K**EEVQSEK-V**K**EAFK	τ55	τ95	412	66	30.13
KVLVVSPK-YEKM**K**K	τ138	τ131	240	636	28.79
KLLSSPIVK-EVYNPN**K**K	τ95	τ131	395	871	28.67
E**K**SYNIPIDIR-**K**STSVNVSDGK	τ131	τ138	426	698	28.39
F**K**NSTFDSLLAAEK-ASDDAIS**K**K	τ131	τ138	412	697	28.37
SSSPGSSLGQ**K**GRPIR-MLTNLDS**K**GNLSSR	τ91	τ60	153	124	28.36
E**K**SYNIPIDIR-ASDDAIS**K**K	τ131	τ138	426	697	28.33
DHTGDKEEVQSE**K**IYR-VTMP**K**GTLSK	τ55	τ95	419	123	27.86
DESVIL**K**VTMPK-S**K**NGIR	τ95	τ138	118	200	27.62
LDP**K**TAETMK-YH**K**IRK	τ95	τ131	631	947	27.25
LFTAQNFQ**K**LTNAEDEISVPK-WDYSI**K**DDKYR	τ138	τ91	414	610	27.09
RI**K**LEQHVSTAQEPK-NSSATQ**K**SLR	τ138	τ91	509	598	26.67
KTGQLAR-**K**VSAGR	τ131	τ138	242	715	26.17
TTVVVENT**K**EDK-KIDA**K**YK	τ138	τ131	729	349	26.15
MSLNENGD**K**IR-DSN**K**LR	τ55	τ95	196	145	26.08
AIL**K**VMNTIGGVAYLR-**K**TGQLAR	τ138	τ131	561	242	25.72
IYGLN**K**EK-IPPNS**K**K	τ91	τ138	178	478	25.69
LLSSPIV**K**K-EVYNPN**K**K	τ95	τ131	403	871	25.54
SLYTDLE**K**QEADYNLGR-SV**K**IPSTDFQLPPPPK	τ131	τ95	958	219	25.40
ITA**K**VVDK-**K**TYTRK	τ131	τ138	627	885	25.24

Appendix Table B2 – TFIIIC intra cross-links without tDNA

Peptide*	Prot. 1	Prot. 2	Lys. 1	Lys. 2	ld-score**
ITG**K**EFQR-NPAI**K**IR	τ138	τ138	362	252	45.76
KVLVVSPK-NPAI**K**IR	τ138	τ138	240	252	41.79
LL**K**DLSSAR-NST**K**NMK	τ91	τ91	161	139	41.38
IYGLN**K**EK-LLLLA**K**VK	τ91	τ91	178	186	40.74
LLHGI**K**NSSATQK-LL**K**DLSSAR	τ91	τ91	591	161	40.34
KSTSVNVSDGK-I**K**ASDDAISK	τ138	τ138	698	689	39.79

ITA**K**VVDKYEK-**K**HQVDETLHR	τ131	τ131	627	567	39.12
SGL**K**GISTMDVVNR-NPAI**K**IR	τ138	τ138	348	252	39.12
I**K**ASDDAISK-**K**VSAGR	τ138	τ138	689	715	39.10
NM**K**SSSPGSSLGQK-LL**K**DLSSAR	τ91	τ91	142	161	38.13
AFT**K**SSEYYLESVDK-ITG**K**EFQR	τ138	τ138	370	362	38.05
GDVDLMVESE**K**LGAR-G**K**VVNFGGFSAR-	τ138	τ138	607	538	37.99
LL**K**DLSSAR-LLLLA**K**VK	τ91	τ91	161	186	37.61
ITG**K**EFQR-CV**K**YVK	τ138	τ138	362	257	37.52
A**K**IDAHGINITCTK-WDYSI**K**DDKYR	τ91	τ91	634	610	37.26
NEIEAL**K**NVGNESIDNVIMDMAK-VPLG**K**PFSR	τ138	τ138	930	1073	37.18
NM**K**SSSPGSSLGQK-LLHGI**K**NSSATQK-	τ91	τ91	142	591	36.96
DINCNS**K**NLFHVK-LL**K**DLLVDRK	τ60	τ60	53	5	36.70
TTVVVENT**K**EDK-**K**VSAGR	τ138	τ138	729	715	36.63
ISAST**K**DLSK-**K**TYTRK	τ138	τ138	896	885	36.61
LL**K**DLSSAR-IYGLN**K**EK	τ91	τ91	161	178	36.47
DLS**K**SQSDDYIR-VPLG**K**PFSR	τ138	τ138	900	1073	36.04
TAETM**K**SELK-IA**K**LDPK	τ95	τ95	637	627	35.94
GTLS**K**NNNSVK-DSI**K**SLK	τ95	τ95	128	138	35.70
SAEDSPSSNGGTVV**K**GK-GDVDLMVESE**K**LGAR	τ138	τ138	536	607	35.69
IPSLELPLNVST**K**HSSIQK-VTMP**K**GTLSK	τ95	τ95	48	123	35.51
LL**K**DLSSAR-NTVQ**K**LR	τ91	τ91	161	130	35.27
SSNDTSDISSKPLLEDS**K**FR-**K**HQVDETLHR	τ131	τ131	599	567	34.86
NSSATQ**K**SLR-NTVQ**K**LR	τ91	τ91	598	130	34.18
KSTSVNVSDG**K**IK-**K**VSAGR	τ138	τ138	708	715	33.99
CY**K**EIESYETAK-**K**FVGILR	τ131	τ131	513	678	33.83
CY**K**EIESYETAK-TTDVF**K**PLAR	τ131	τ131	513	506	33.47
IYDFEG**K**KK-CV**K**YVK	τ138	τ138	400	257	33.40
IT**K**LFPNNSLDNLK-TTVVVENT**K**EDK	τ138	τ138	767	729	33.32
NST**K**NMK-NTVQ**K**LR	τ91	τ91	139	130	33.11
STSVNVSDG**K**IK-ASDDAIS**K**K	τ138	τ138	708	697	32.98
I**K**ASDDAISKK-EKDS**K**K	τ138	τ138	689	642	32.85
SSNDTSDISS**K**PLLEDSKFR-**K**HQVDETLHR	τ131	τ131	592	567	32.74
NM**K**SSSPGSSLGQK-DIATT**K**TTVSR	τ91	τ91	142	503	32.68
SSSPGSSLGQ**K**GRPIR-NST**K**NMK	τ138	τ138	153	139	32.57
SSSPGSSLGQ**K**GRPIR-NSSATQ**K**SLR	τ91	τ91	153	598	32.49
KHQVDETLHR-VVD**K**YEK	τ131	τ131	567	631	32.39
KSTSVNVSDGK-**K**VSAGR	τ138	τ138	698	715	32.31
KFNTELDFQIER-VVD**K**YEK	τ131	τ131	688	631	32.27
SSSPGSSLGQ**K**GRPIR-NTVQ**K**LR	τ91	τ91	153	130	32.26
SSNDTSDISSKPLLEDS**K**FR-**K**RTPYDAER	τ131	τ131	599	607	32.09
NM**K**SSSPGSSLGQK-RNTVQ**K**LR	τ91	τ91	142	130	32.04
V**K**EGFETSVFDFPFK-IYGLN**K**EK	τ91	τ91	188	178	32.03
YA**K**DINCNSK-ICPVS**K**QR	τ60	τ60	46	549	32.01
EASNF**K**VK-G**K**SYGR	τ131	τ131	102	113	31.85
TDL**K**TLNEDNFVALNNTVR-RS**K**NGIR	τ138	τ138	439	200	31.73
EFEDW**K**NNLTWAR-**K**ITDLK	τ60	τ60	18	228	31.67
AFT**K**SSEYYLESVDK-CV**K**YVK	τ138	τ138	370	257	31.40
SSSPGSSLGQ**K**GRPIR-DIATT**K**TTVSR	τ91	τ91	153	503	31.33
RLLHGI**K**NSSATQK-NST**K**NMK	τ91	τ91	591	139	31.16
RLLHGI**K**NSSATQK-NTVQ**K**LR	τ91	τ91	591	130	30.94
I**K**ASDDAISKK-YIL**K**EK	τ138	τ138	689	637	30.94

LS**K**LAEGDSVFEGPLMEER-ITA**K**VVDKYEK	τ131	τ131	702	627	30.80
NSTNASVAGNISNP**K**R-I**K**LEQHVSTAQEPK	τ138	τ138	506	509	30.78
IPSLELPLNVST**K**HSSIQK-AI**K**MCGGIEK	τ95	τ95	48	57	30.59
RQV**K**NSTNASVAGNISNPK-RI**K**LEQHVSTAQEPK	τ138	τ138	491	509	30.26
ISNE**K**SSNDTSDISSKPLLEDSK-**K**HQVDETLHR	τ131	τ131	581	567	30.18
NSTNASVAGNISNP**K**R-G**K**VVNFGGFSAR	τ138	τ138	506	538	29.71
AFT**K**SSEYYLESVDK-GHVV**K**QLK	τ138	τ138	370	165	29.51
DN**K**MLTNLDSK-VIDI**K**R	τ60	τ60	116	556	29.31
IPPNS**K**K-TPN**K**NK	τ138	τ138	478	483	29.28
TTVVVENTKED**K**TVYHAGTK-**K**VSAGR	τ138	τ138	732	715	29.28
EI**K**RPLFLYK-**K**FTLVR	τ138	τ138	839	854	29.26
SAEDSPSSNGGTVV**K**GK-AIL**K**VMNTIGGVAYLR	τ138	τ138	536	561	29.22
DIATT**K**TTVSR-NTVQ**K**LR	τ91	τ91	503	130	29.12
AFT**K**SSEYYLESVDK-Q**K**ENTGGYR	τ138	τ138	370	383	28.95
LS**K**LAEGDSVFEGPLMEER-VVD**K**YEK	τ131	τ131	702	631	28.90
IT**K**LFPNNSLDNLK-**K**VSAGR	τ138	τ138	767	715	28.82
LFTAQNFQ**K**LTNAEDEISVPK-Q**K**ENTGGYR	τ138	τ138	414	383	28.40
NM**K**SSSPGSSLGQK-NSSATQ**K**SLR	τ91	τ91	142	598	27.99
KFELNSGLNEAK-VVD**K**YEK	τ131	τ131	637	631	27.81
NEIEAL**K**NVGNESIDNVIMDMAK-DLS**K**SQSDDYIR	τ138	τ138	930	900	27.70
ASDDAIS**K**K-**K**VSAGR	τ138	τ138	697	715	27.63
SSSPGSSLGQ**K**GRPIR-LL**K**DLSSAR	τ91	τ91	153	161	27.34
SGE**K**GINTMDLAQVTGQDPR-**K**INHLLTSSQLIYK	τ138	τ138	123	147	27.31
NVP**K**PPPLVFESDTPGGIDSR-KLLSSPIV**K**K	τ95	τ95	408	403	27.29
ANPFA**K**KK-**K**NGVTEVK	τ95	τ95	251	254	27.08
EASNF**K**VK-**K**NDKGK	τ131	τ131	102	108	26.84
GTLS**K**NNNSVK-SL**K**DSNK	τ95	τ95	128	141	26.59
KSTSVNVSDGK-E**K**DSKK	τ138	τ138	698	639	26.34
GTLS**K**NNNSVK-DSN**K**LR	τ95	τ95	128	145	26.15
NM**K**SSSPGSSLGQK-IYGLN**K**EK	τ91	τ91	142	178	26.09
Q**K**ENTGGYR-IPPNS**K**K	τ138	τ138	383	478	25.69

Appendix Table B3 – TFIIIC inter cross-links with tDNA

Peptide*	Prot. 1	Prot. 2	Lys. 1	Lys. 2	ld-score**
MLTNLDS**K**GNLSSR-NTVQ**K**LR	τ60	τ91	124	130	42.62
MLTNLDS**K**GNLSSR-NM**K**SSSPGSSLGQK	τ60	τ91	124	142	40.62
ISAST**K**DLSK-DSI**K**SLK	τ138	τ95	896	138	39.75
MLTNLDS**K**GNLSSR-NST**K**NMK	τ60	τ91	124	139	39.49
NSTNASVAGNISNP**K**R-MLTNLDS**K**GNLSSR	τ138	τ60	506	124	38.69
KIIFLPTVGEDAIQR-G**K**SYGR	τ138	τ131	619	113	36.55
KTAGLLSPTEENETTNAGQSK-V**K**EAFK	τ55	τ95	359	66	35.82
SGSCSLDKYEIL**K**K-G**K**SYGR	τ55	τ131	211	113	35.04
E**K**SYNIPIDIR-**K**STSVNVSDGK	τ131	τ138	426	698	34.80
KLLSSPIVKK-EVYNPN**K**K	τ95	τ131	395	871	34.77
KISAST**K**DLSK-SL**K**DSNK	τ138	τ95	896	141	33.85
E**K**SYNIPIDIR-ASDDAIS**K**K	τ131	τ138	426	697	33.80
F**K**NSTFDSLLAAEK-ASDDAIS**K**K	τ131	τ138	412	697	33.77
DHTGDKEEVQSE**K**IYR-V**K**EAFK	τ55	τ95	419	66	32.98
DHTGD**K**EEVQSEK-AI**K**MCGGIEK	τ55	τ95	412	57	31.98

KSTSVNVSDGK-**K**TGQLAR	τ138	τ131	698	242	30.83
KTAGLLSPTEENETTNAGQSK-AI**K**MCGGIEK-	τ55	τ95	359	57	30.05
MSLNENGD**K**IR-DSN**K**LR	τ55	τ95	196	145	29.03
E**K**SYNIPIDIR-**K**VSAGR	τ131	τ138	426	715	28.93
SLYTDLE**K**QEADYNLGR-SV**K**IPSTDFQLPPPPK	τ131	τ95	958	219	28.70
ASDDAIS**K**K-**K**TGQLAR	τ138	τ131	697	242	27.44
TTVVVENT**K**EDK-KIDA**K**YK	τ138	τ131	729	349	27.41
F**K**NSTFDSLLAAEK-LLSSPIV**K**K	τ131	τ95	412	403	27.12
STSVNVSDG**K**IK-E**K**SYNIPIDIR	τ138	τ131	708	426	26.64
KTGQLAR-EKDS**K**K	τ131	τ138	242	642	25.82
KTGQLAR-**K**VSAGR	τ131	τ138	242	715	25.04

Appendix Table B4 – TFIIIC intra cross-links with tDNA

Peptide*	Prot. 1	Prot. 2	Lys. 1	Lys. 2	ld-score**
NM**K**SSSPGSSLGQK-IYGLN**K**EK	τ91	τ91	142	178	44.25
NM**K**SSSPGSSLGQK-LL**K**DLSSAR	τ91	τ91	142	161	43.22
GTLS**K**NNNSVKDSIK-DSN**K**LR	τ95	τ95	128	145	41.40
KNSENTPEFYFESSIR-DN**K**MLTNLDSK	τ60	τ60	160	116	41.27
KVLVVSPK-NPAI**K**IR	τ138	τ138	240	252	41.23
SGL**K**GISTMDVVNR-NPAI**K**IR	τ138	τ138	348	252	40.52
GDVDLMVESE**K**LGAR-G**K**VVNFGGFSAR	τ138	τ138	607	538	40.42
LLHGI**K**NSSATQK-LL**K**DLSSAR	τ91	τ91	591	161	39.97
LL**K**DLSSAR-IYGLN**K**EK	τ91	τ91	161	178	39.91
ITG**K**EFQR-CV**K**YVK	τ138	τ138	362	257	39.80
NM**K**SSSPGSSLGQK-DIATT**K**TTVSR	τ91	τ91	142	503	39.66
GTLS**K**NNNSVK-DSI**K**SLK	τ95	τ95	128	138	39.23
ISAST**K**DLSK-**K**TYTRK	τ138	τ138	896	885	38.70
KSTSVNVSDGK-I**K**ASDDAISK	τ138	τ138	698	689	38.35
ASDDAIS**K**K-**K**VSAGR	τ138	τ138	697	715	38.33
DIATT**K**TTVSR-NTVQ**K**LR	τ91	τ91	503	130	38.15
IYGLN**K**EK-LLLLA**K**VK	τ91	τ91	178	186	38.13
GTLS**K**NNNSVK-SL**K**DSNK	τ95	τ95	128	141	38.11
NST**K**NMK-NTVQ**K**LR	τ91	τ91	139	130	38.10
NM**K**SSSPGSSLGQK-NTVQ**K**LR	τ91	τ91	142	130	38.01
DINCNS**K**NLFHVK-LL**K**DLLVDRK	τ60	τ60	53	5	37.14
TAETM**K**SELK-IA**K**LDPK	τ95	τ95	637	627	36.38
NSTNASVAGNISNP**K**R-**K**IIFLPTVGEDAIQR	τ138	τ138	506	619	35.91
I**K**ASDDAISK-**K**VSAGR	τ138	τ138	689	715	35.84
LLHGI**K**NSSATQK-NST**K**NMK	τ91	τ91	591	139	35.43
V**K**EGFETSVFDFPFK-IYGLN**K**EK	τ91	τ91	188	178	35.37
KFELNSGLNEAK-VVD**K**YEK	τ131	τ131	637	631	35.16
KIIFLPTVGEDAIQR-YIL**K**EK	τ138	τ138	619	637	35.10
AFT**K**SSEYYLESVDK-CV**K**YVK	τ138	τ138	370	257	34.92
KFNTELDFQIER-VVD**K**YEK	τ131	τ131	688	631	34.86
TTVVVENT**K**EDK-**K**VSAGR	τ138	τ138	729	715	34.78
DLS**K**SQSDDYIR-**K**ISASTK	τ138	τ138	900	890	34.68
SGL**K**GISTMDVVNR-VLVVSP**K**NPAIK	τ138	τ138	348	247	34.65
YA**K**DINCNSK-ICPVS**K**QR	τ60	τ60	46	549	34.63
KIIFLPTVGEDAIQR-**K**VSAGR	τ138	τ138	619	715	34.61

CY**K**EIESYETAK-TTDVF**K**PLAR	τ131	τ131	513	506	34.46
LL**K**DLSSAR-NST**K**NMK	τ91	τ91	161	139	34.18
STSVNVSDG**K**IK-ASDDAIS**K**K	τ138	τ138	708	697	34.14
KSTSVNVSDGK-**K**VSAGR	τ138	τ138	698	715	33.99
AFT**K**SSEYYLESVDK-ITG**K**EFQR	τ138	τ138	370	362	33.94
IYGLN**K**EK-NST**K**NMK	τ91	τ91	178	139	33.84
IT**K**LFPNNSLDNLK-TTVVVENT**K**EDK	τ138	τ138	767	729	33.81
SSNDTSDISS**K**PLLEDSK-**K**HQVDETLHR	τ131	τ131	592	567	33.40
I**K**ASDDAISK-YIL**K**EK	τ138	τ138	689	637	33.27
IPSLELPLNVST**K**HSSIQK-VTMP**K**GTLSK	τ95	τ95	48	123	33.08
STSVNVSDG**K**IK-**K**VSAGR	τ138	τ138	708	715	32.98
DIATT**K**TTVSR-NST**K**NMK	τ91	τ91	503	139	32.74
ITG**K**EFQR-NPAI**K**IR	τ138	τ138	362	252	32.66
LSMVGFPLLY**K**YK-ANPFA**K**K	τ95	τ95	243	251	32.50
LS**K**LAEGDSVFEGPLMEER-ITA**K**VVDKYEK	τ131	τ131	702	627	32.38
KIIFLPTVGEDAIQR-YMGSTTTLD**K**K	τ138	τ138	619	592	32.27
LL**K**DLSSAR-LLLLA**K**VK	τ91	τ91	161	186	32.21
QV**K**NSTNASVAGNISNPK-TPN**K**NK	τ138	τ138	491	483	32.15
SGE**K**GINTMDLAQVTGQDPR-QLKL**K**K	τ138	τ138	123	170	31.74
IYDFEGK**K**K-ITG**K**EFQR	τ138	τ138	401	362	31.60
STSVNVSDG**K**IK-TTVVVENT**K**EDK	τ138	τ138	708	729	31.51
IYDFEG**K**KK-ITG**K**EFQR	τ138	τ138	400	362	30.59
AFT**K**SSEYYLESVDK-Q**K**ENTGGYR	τ138	τ138	370	383	30.51
NSTNASVAGNISNP**K**R-IPPNS**K**K	τ138	τ138	506	478	30.43
NSSATQ**K**SLR-L**K**KNSTK	τ91	τ91	598	134	30.22
IYGLN**K**EK-NTVQ**K**LR	τ91	τ91	178	130	29.81
IPPNS**K**K-TPN**K**NK	τ138	τ138	478	483	29.65
EASNF**K**VK-G**K**SYGR	τ131	τ131	102	113	27.62
GTYI**K**NYQLFVHDLSDK-ANPFA**K**K	τ95	τ95	266	251	27.42

* Cross-linked lysines are underlined and in bold

** All values are listed in descending order of Id Score, only cross-links with an ld score >25 were selected for analysis

Prot. = Protein / Lys. = Lysine

8. References

1. Alberts, B., *The cell as a collection of protein machines: preparing the next generation of molecular biologists.* Cell, 1998. **92**(3): p. 291-4.
2. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid.* Nature, 1953. **171**(4356): p. 737-8.
3. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 A resolution.* Nature, 1997. **389**(6648): p. 251-60.
4. Carter, R. and G. Drouin, *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.* Mol Biol Evol, 2010. **27**(5): p. 1035-43.
5. Roeder, R.G., *The role of general initiation factors in transcription by RNA polymerase II.* Trends Biochem Sci, 1996. **21**(9): p. 327-35.
6. Murakami, K.S. and S.A. Darst, *Bacterial RNA polymerases: the wholo story.* Curr Opin Struct Biol, 2003. **13**(1): p. 31-9.
7. Kapanidis, A.N., et al., *Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism.* Science, 2006. **314**(5802): p. 1144-7.
8. Goldman, S.R., R.H. Ebright, and B.E. Nickels, *Direct detection of abortive RNA transcripts in vivo.* Science, 2009. **324**(5929): p. 927-8.
9. Kwak, H. and J.T. Lis, *Control of transcriptional elongation.* Annu Rev Genet, 2013. **47**: p. 483-508.
10. Vassylyev, D.G., *Elongation by RNA polymerase: a race through roadblocks.* Curr Opin Struct Biol, 2009. **19**(6): p. 691-700.
11. Zhang, G., et al., *Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution.* Cell, 1999. **98**(6): p. 811-24.
12. Werner, F. and D. Grohmann, *Evolution of multisubunit RNA polymerases in the three domains of life.* Nat Rev Microbiol, 2011. **9**(2): p. 85-98.
13. Campbell, E.A., L.F. Westblade, and S.A. Darst, *Regulation of bacterial RNA polymerase sigma factor activity: a structural perspective.* Curr Opin Microbiol, 2008. **11**(2): p. 121-7.
14. Feklistov, A. and S.A. Darst, *Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit.* Cell, 2011. **147**(6): p. 1257-69.
15. Wigneshweraraj, S., et al., *Modus operandi of the bacterial RNA polymerase containing the sigma54 promoter-specificity factor.* Mol Microbiol, 2008. **68**(3): p. 538-46.
16. Werner, F. and R.O. Weinzierl, *Direct modulation of RNA polymerase core functions by basal transcription factors.* Mol Cell Biol, 2005. **25**(18): p. 8344-55.
17. Ferri, M.L., et al., *A novel subunit of yeast RNA polymerase III interacts with the TFIIB-related domain of TFIIIB70.* Mol Cell Biol, 2000. **20**(2): p. 488-95.
18. Hirata, A. and K.S. Murakami, *Archaeal RNA polymerase.* Curr Opin Struct Biol, 2009. **19**(6): p. 724-31.
19. Hirata, A., B.J. Klein, and K.S. Murakami, *The X-ray crystal structure of RNA polymerase from Archaea.* Nature, 2008. **451**(7180): p. 851-4.

20. Korkhin, Y., et al., *Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure.* PLoS Biol, 2009. **7**(5): p. e1000102.
21. Hausner, W. and M. Thomm, *Events during initiation of archaeal transcription: open complex formation and DNA-protein interactions.* J Bacteriol, 2001. **183**(10): p. 3025-31.
22. Roeder, R.G. and W.J. Rutter, *Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms.* Nature, 1969. **224**(5216): p. 234-7.
23. Kedinger, C., et al., *Alpha-amanitin: a specific inhibitor of one of two DNA-pendent RNA polymerase activities from calf thymus.* Biochem Biophys Res Commun, 1970. **38**(1): p. 165-71.
24. Matzke, M., et al., *RNA-mediated chromatin-based silencing in plants.* Curr Opin Cell Biol, 2009. **21**(3): p. 367-76.
25. Pikaard, C.S., et al., *Roles of RNA polymerase IV in gene silencing.* Trends Plant Sci, 2008. **13**(7): p. 390-7.
26. Orphanides, G., T. Lagrange, and D. Reinberg, *The general transcription factors of RNA polymerase II.* Genes Dev, 1996. **10**(21): p. 2657-83.
27. Kornberg, R.D., *The molecular basis of eukaryotic transcription.* Proc Natl Acad Sci U S A, 2007. **104**(32): p. 12955-61.
28. Cramer, P., et al., *Structure of eukaryotic RNA polymerases.* Annu Rev Biophys, 2008. **37**: p. 337-52.
29. Cramer, P., et al., *Architecture of RNA polymerase II and implications for the transcription mechanism.* Science, 2000. **288**(5466): p. 640-9.
30. Kostrewa, D., et al., *RNA polymerase II-TFIIB structure and mechanism of transcription initiation.* Nature, 2009. **462**(7271): p. 323-30.
31. Liu, X., et al., *Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism.* Science, 2010. **327**(5962): p. 206-9.
32. Gnatt, A.L., et al., *Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution.* Science, 2001. **292**(5523): p. 1876-82.
33. Kettenberger, H., K.J. Armache, and P. Cramer, *Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS.* Mol Cell, 2004. **16**(6): p. 955-65.
34. He, Y., et al., *Structural visualization of key steps in human transcription initiation.* Nature, 2013. **495**(7442): p. 481-6.
35. Murakami, K., et al., *Architecture of an RNA polymerase II transcription pre-initiation complex.* Science, 2013. **342**(6159): p. 1238724.
36. Vannini, A. and P. Cramer, *Conservation between the RNA polymerase I, II, and III transcription initiation machineries.* Mol Cell, 2012. **45**(4): p. 439-46.
37. Kuhn, C.D., et al., *Functional architecture of RNA polymerase I.* Cell, 2007. **131**(7): p. 1260-72.
38. Geiger, S.R., et al., *RNA polymerase I contains a TFIIF-related DNA-binding subcomplex.* Mol Cell, 2010. **39**(4): p. 583-94.
39. Fernandez-Tornero, C., et al., *Crystal structure of the 14-subunit RNA polymerase I.* Nature, 2013. **502**(7473): p. 644-9.
40. Engel, C., et al., *RNA polymerase I structure and transcription regulation.* Nature, 2013. **502**(7473): p. 650-5.
41. Gaiser, F., S. Tan, and T.J. Richmond, *Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 A resolution.* J Mol Biol, 2000. **302**(5): p. 1119-27.

42. Schramm, L. and N. Hernandez, *Recruitment of RNA polymerase III to its target promoters.* Genes Dev, 2002. **16**(20): p. 2593-620.
43. Geiduschek, E.P. and G.A. Kassavetis, *The RNA polymerase III transcription apparatus.* J Mol Biol, 2001. **310**(1): p. 1-26.
44. Flores, A., et al., *A protein-protein interaction map of yeast RNA polymerase III.* Proc Natl Acad Sci U S A, 1999. **96**(14): p. 7815-20.
45. Wu, C.C., et al., *RNA polymerase III subunit architecture and implications for open promoter complex formation.* Proc Natl Acad Sci U S A, 2012. **109**(47): p. 19232-7.
46. Fernandez-Tornero, C., et al., *Conformational flexibility of RNA polymerase III during transcriptional elongation.* EMBO J, 2010. **29**(22): p. 3762-72.
47. Kassavetis, G.A., P. Prakash, and E. Shim, *The C53/C37 subcomplex of RNA polymerase III lies near the active site and participates in promoter opening.* J Biol Chem, 2010. **285**(4): p. 2695-706.
48. Landrieux, E., et al., *A subcomplex of RNA polymerase III subunits involved in transcription termination and reinitiation.* EMBO J, 2006. **25**(1): p. 118-28.
49. Wang, Z. and R.G. Roeder, *Three human RNA polymerase III-specific subunits form a subcomplex with a selective function in specific transcription initiation.* Genes Dev, 1997. **11**(10): p. 1315-26.
50. Lefevre, S., et al., *Structure-function analysis of hRPC62 provides insights into RNA polymerase III transcription initiation.* Nat Struct Mol Biol, 2011. **18**(3): p. 352-8.
51. Meinhart, A., J. Blobel, and P. Cramer, *An extended winged helix domain in general transcription factor E/IIE alpha.* J Biol Chem, 2003. **278**(48): p. 48267-74.
52. Sakonju, S., D.F. Bogenhagen, and D.D. Brown, *A control region in the center of the 5S RNA gene directs specific initiation of transcription: I. The 5' border of the region.* Cell, 1980. **19**(1): p. 13-25.
53. Bogenhagen, D.F., S. Sakonju, and D.D. Brown, *A control region in the center of the 5S RNA gene directs specific initiation of transcription: II. The 3' border of the region.* Cell, 1980. **19**(1): p. 27-35.
54. Bogenhagen, D.F., *The intragenic control region of the Xenopus 5 S RNA gene contains two factor A binding domains that must be aligned properly for efficient transcription initiation.* J Biol Chem, 1985. **260**(10): p. 6466-71.
55. Galli, G., H. Hofstetter, and M.L. Birnstiel, *Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements.* Nature, 1981. **294**(5842): p. 626-31.
56. Sharp, S., et al., *Internal control regions for transcription of eukaryotic tRNA genes.* Proc Natl Acad Sci U S A, 1981. **78**(11): p. 6657-61.
57. Marck, C., et al., *The RNA polymerase III-dependent family of genes in hemiascomycetes: comparative RNomics, decoding strategies, transcription and evolutionary implications.* Nucleic Acids Res, 2006. **34**(6): p. 1816-35.
58. Ciliberto, G., et al., *Common and interchangeable elements in the promoters of genes transcribed by RNA polymerase iii.* Cell, 1983. **32**(3): p. 725-33.
59. Das, G., et al., *Upstream regulatory elements are necessary and sufficient for transcription of a U6 RNA gene by RNA polymerase III.* EMBO J, 1988. **7**(2): p. 503-12.

60. Krol, A., et al., *Xenopus tropicalis U6 snRNA genes transcribed by Pol III contain the upstream promoter elements used by Pol II dependent U snRNA genes.* Nucleic Acids Res, 1987. **15**(6): p. 2463-78.
61. Kunkel, G.R. and T. Pederson, *Upstream elements required for efficient transcription of a human U6 RNA gene resemble those of U1 and U2 genes even though a different polymerase is used.* Genes Dev, 1988. **2**(2): p. 196-204.
62. Murphy, S., M. Tripodi, and M. Melli, *A sequence upstream from the coding region is required for the transcription of the 7SK RNA genes.* Nucleic Acids Res, 1986. **14**(23): p. 9243-60.
63. Hamada, M., et al., *Widespread use of TATA elements in the core promoters for RNA polymerases III, II, and I in fission yeast.* Mol Cell Biol, 2001. **21**(20): p. 6870-81.
64. Segall, J., T. Matsui, and R.G. Roeder, *Multiple factors are required for the accurate transcription of purified genes by RNA polymerase III.* J Biol Chem, 1980. **255**(24): p. 11986-91.
65. Ducrot, C., et al., *Reconstitution of the yeast RNA polymerase III transcription system with all recombinant factors.* J Biol Chem, 2006. **281**(17): p. 11685-92.
66. Orioli, A., et al., *RNA polymerase III transcription control elements: themes and variations.* Gene, 2012. **493**(2): p. 185-94.
67. Kaiser, M.W. and D.A. Brow, *Lethal mutations in a yeast U6 RNA gene B block promoter element identify essential contacts with transcription factor-IIIC.* J Biol Chem, 1995. **270**(19): p. 11398-405.
68. Schultz, P., et al., *The two DNA-binding domains of yeast transcription factor tau as observed by scanning transmission electron microscopy.* EMBO J, 1989. **8**(12): p. 3815-24.
69. Baker, R.E., et al., *Gene size differentially affects the binding of yeast transcription factor tau to two intragenic regions.* Proc Natl Acad Sci U S A, 1987. **84**(24): p. 8768-72.
70. Geiduschek, E.P. and G.P. Tocchini-Valentini, *Transcription by RNA polymerase III.* Annu Rev Biochem, 1988. **57**: p. 873-914.
71. Baker, R.E., O. Gabrielsen, and B.D. Hall, *Effects of tRNATyr point mutations on the binding of yeast RNA polymerase III transcription factor C.* J Biol Chem, 1986. **261**(12): p. 5275-82.
72. Stillman, D.J., P. Caspers, and E.P. Geiduschek, *Effects of temperature and single-stranded DNA on the interaction of an RNA polymerase III transcription factor with a tRNA gene.* Cell, 1985. **40**(2): p. 311-7.
73. Conesa, C., et al., *On the subunit composition, stoichiometry, and phosphorylation of the yeast transcription factor TFIIIC/tau.* J Biol Chem, 1993. **268**(24): p. 18047-52.
74. Acker, J., C. Conesa, and O. Lefebvre, *Yeast RNA polymerase III transcription factors and effectors.* Biochim Biophys Acta, 2013. **1829**(3-4): p. 283-95.
75. Marzouki, N., et al., *Selective proteolysis defines two DNA binding domains in yeast transcription factor tau.* Nature, 1986. **323**(6084): p. 176-8.
76. Nagarajavel, V., et al., *Global 'bootprinting' reveals the elastic architecture of the yeast TFIIIB-TFIIIC transcription complex in vivo.* Nucleic Acids Res, 2013. **41**(17): p. 8135-43.

77. Marck, C., et al., *The TFIIIB-assembling subunit of yeast transcription factor TFIIIC has both tetratricopeptide repeats and basic helix-loop-helix motifs.* Proc Natl Acad Sci U S A, 1993. **90**(9): p. 4027-31.
78. Swanson, R.N., et al., *Isolation of TFC1, a gene encoding one of two DNA-binding subunits of yeast transcription factor tau (TFIIIC).* Proc Natl Acad Sci U S A, 1991. **88**(11): p. 4887-91.
79. Manaud, N., et al., *A chimeric subunit of yeast transcription factor IIIC forms a subcomplex with tau95.* Mol Cell Biol, 1998. **18**(6): p. 3191-200.
80. Deprez, E., et al., *A subunit of yeast TFIIIC participates in the recruitment of TATA-binding protein.* Mol Cell Biol, 1999. **19**(12): p. 8042-51.
81. Arrebola, R., et al., *Tau91, an essential subunit of yeast transcription factor IIIC, cooperates with tau138 in DNA binding.* Mol Cell Biol, 1998. **18**(1): p. 1-9.
82. Lefebvre, O., et al., *TFC3: gene encoding the B-block binding subunit of the yeast transcription factor IIIC.* Proc Natl Acad Sci U S A, 1992. **89**(21): p. 10512-6.
83. Bartholomew, B., et al., *The subunit structure of Saccharomyces cerevisiae transcription factor IIIC probed with a novel photocrosslinking reagent.* EMBO J, 1990. **9**(7): p. 2197-205.
84. Braun, B.R., et al., *Topography of transcription factor complexes on the Saccharomyces cerevisiae 5 S RNA gene.* J Mol Biol, 1992. **228**(4): p. 1063-77.
85. Zeytuni, N. and R. Zarivach, *Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module.* Structure, 2012. **20**(3): p. 397-405.
86. Halbach, F., et al., *The yeast ski complex: crystal structure and RNA channeling to the exosome complex.* Cell, 2013. **154**(4): p. 814-26.
87. Moir, R.D. and I.M. Willis, *Tetratricopeptide repeats of Tfc4 and a limiting step in the assembly of the initiation factor TFIIIB.* Adv Protein Chem, 2004. **67**: p. 93-121.
88. Hsieh, Y.J., et al., *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.* Mol Cell Biol, 1999. **19**(7): p. 4944-52.
89. Taylor, N.M., et al., *RNA polymerase III-specific general transcription factor IIIC contains a heterodimer resembling TFIIF Rap30/Rap74.* Nucleic Acids Res, 2013. **41**(19): p. 9183-96.
90. Jourdain, S., et al., *The tau95 subunit of yeast TFIIIC influences upstream and downstream functions of TFIIIC.DNA complexes.* J Biol Chem, 2003. **278**(12): p. 10450-7.
91. Taylor, N.M., et al., *Structural and functional characterization of a phosphatase domain within yeast general transcription factor IIIC.* J Biol Chem, 2013. **288**(21): p. 15110-20.
92. Mylona, A., et al., *Structure of the tau60/Delta tau91 subcomplex of yeast transcription factor IIIC: insights into preinitiation complex assembly.* Mol Cell, 2006. **24**(2): p. 221-32.
93. Lefebvre, O., J. Ruth, and A. Sentenac, *A mutation in the largest subunit of yeast TFIIIC affects tRNA and 5 S RNA synthesis. Identification of two classes of suppressors.* J Biol Chem, 1994. **269**(37): p. 23374-81.

94. Huang, Y. and R.J. Maraia, *Comparison of the RNA polymerase III transcription machinery in Schizosaccharomyces pombe, Saccharomyces cerevisiae and human.* Nucleic Acids Res, 2001. **29**(13): p. 2675-90.
95. Harami, G.M., M. Gyimesi, and M. Kovacs, *From keys to bulldozers: expanding roles for winged helix domains in nucleic-acid-binding proteins.* Trends Biochem Sci, 2013. **38**(7): p. 364-71.
96. Yoshinaga, S.K., P.A. Boulanger, and A.J. Berk, *Resolution of human transcription factor TFIIIC into two functional components.* Proc Natl Acad Sci U S A, 1987. **84**(11): p. 3585-9.
97. Dean, N. and A.J. Berk, *Separation of TFIIIC into two functional components by sequence specific DNA affinity chromatography.* Nucleic Acids Res, 1987. **15**(23): p. 9895-907.
98. Dumay-Odelot, H., et al., *Identification, molecular cloning, and characterization of the sixth subunit of human transcription factor TFIIIC.* J Biol Chem, 2007. **282**(23): p. 17179-89.
99. Huang, Y., M. Hamada, and R.J. Maraia, *Isolation and cloning of four subunits of a fission yeast TFIIIC complex that includes an ortholog of the human regulatory protein TFIIICbeta.* J Biol Chem, 2000. **275**(40): p. 31480-7.
100. Kundu, T.K., Z. Wang, and R.G. Roeder, *Human TFIIIC relieves chromatin-mediated repression of RNA polymerase III transcription and contains an intrinsic histone acetyltransferase activity.* Mol Cell Biol, 1999. **19**(2): p. 1605-15.
101. Hsieh, Y.J., et al., *The TFIIIC90 subunit of TFIIIC interacts with multiple components of the RNA polymerase III machinery and contains a histone-specific acetyltransferase activity.* Mol Cell Biol, 1999. **19**(11): p. 7697-704.
102. Shen, Y., et al., *DNA binding domain and subunit interactions of transcription factor IIIC revealed by dissection with poliovirus 3C protease.* Mol Cell Biol, 1996. **16**(8): p. 4163-71.
103. Matsutani, S., *Similarities in transcription factor IIIC subunits that bind to the posterior regions of internal promoters for RNA polymerase III.* BMC Evol Biol, 2004. **4**: p. 26.
104. Rozenfeld, S. and P. Thuriaux, *Genetic interactions within TFIIIC, the promoter-binding factor of yeast RNA polymerase III.* Mol Genet Genomics, 2001. **265**(4): p. 705-10.
105. Cormack, B.P. and K. Struhl, *The TATA-binding protein is required for transcription by all three nuclear RNA polymerases in yeast cells.* Cell, 1992. **69**(4): p. 685-96.
106. Bartholomew, B., G.A. Kassavetis, and E.P. Geiduschek, *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.* Mol Cell Biol, 1991. **11**(10): p. 5181-9.
107. Chaussivert, N., et al., *Complex interactions between yeast TFIIIB and TFIIIC.* J Biol Chem, 1995. **270**(25): p. 15353-8.
108. Dumay-Odelot, H., et al., *Multiple roles of the tau131 subunit of yeast transcription factor IIIC (TFIIIC) in TFIIIB assembly.* Mol Cell Biol, 2002. **22**(1): p. 298-308.

109. Liao, Y., I.M. Willis, and R.D. Moir, *The Brf1 and Bdp1 subunits of transcription factor TFIIIB bind to overlapping sites in the tetratricopeptide repeats of Tfc4.* J Biol Chem, 2003. **278**(45): p. 44467-74.
110. Liao, Y., R.D. Moir, and I.M. Willis, *Interactions of Brf1 peptides with the tetratricopeptide repeat-containing subunit of TFIIIC inhibit and promote preinitiation complex assembly.* Mol Cell Biol, 2006. **26**(16): p. 5946-56.
111. Moir, R.D., K.V. Puglia, and I.M. Willis, *Autoinhibition of TFIIIB70 binding by the tetratricopeptide repeat-containing subunit of TFIIIC.* J Biol Chem, 2002. **277**(1): p. 694-701.
112. Rameau, G., et al., *A mutation in the second largest subunit of TFIIIC increases a rate-limiting step in transcription by RNA polymerase III.* Mol Cell Biol, 1994. **14**(1): p. 822-30.
113. Moir, R.D., et al., *A tetratricopeptide repeat mutation in yeast transcription factor IIIC131 (TFIIIC131) facilitates recruitment of TFIIB-related factor TFIIIB70.* Mol Cell Biol, 1997. **17**(12): p. 7119-25.
114. Moir, R.D., K.V. Puglia, and I.M. Willis, *A gain-of-function mutation in the second tetratricopeptide repeat of TFIIIC131 relieves autoinhibition of Brf1 binding.* Mol Cell Biol, 2002. **22**(17): p. 6131-41.
115. Khoo, S.K., et al., *Mapping the protein interaction network for TFIIB-related factor Brf1 in the RNA polymerase III preinitiation complex.* Mol Cell Biol, 2014. **34**(3): p. 551-9.
116. Cloutier, T.E., et al., *Kinetic trapping of DNA by transcription factor IIIB.* Proc Natl Acad Sci U S A, 2001. **98**(17): p. 9581-6.
117. Moir, R.D., K.V. Puglia, and I.M. Willis, *Interactions between the tetratricopeptide repeat-containing transcription factor TFIIIC131 and its ligand, TFIIIB70. Evidence for a conformational change in the complex.* J Biol Chem, 2000. **275**(34): p. 26591-8.
118. Wu, C.C., Y.C. Lin, and H.T. Chen, *The TFIIF-like Rpc37/53 dimer lies at the center of a protein network to connect TFIIIC, Bdp1, and the RNA polymerase III active center.* Mol Cell Biol, 2011. **31**(13): p. 2715-28.
119. Dumay, H., et al., *Interaction between yeast RNA polymerase III and transcription factor TFIIIC via ABC10alpha and tau131 subunits.* J Biol Chem, 1999. **274**(47): p. 33462-8.
120. Kassavetis, G.A., et al., *S. cerevisiae TFIIIB is the transcription initiation factor proper of RNA polymerase III, while TFIIIA and TFIIIC are assembly factors.* Cell, 1990. **60**(2): p. 235-45.
121. Hamada, M., et al., *Transcription termination by RNA polymerase III in fission yeast - A genetic and biochemically tractable model system.* Journal of Biological Chemistry, 2000. **275**(37): p. 29076-29081.
122. Campbell, F.E. and D.R. Setzer, *Transcription Termination by Rna Polymerase-Iii - Uncoupling of Polymerase Release from Termination Signal Recognition.* Molecular and Cellular Biology, 1992. **12**(5): p. 2260-2272.
123. Nielsen, S., Y. Yuzenkova, and N. Zenkin, *Mechanism of eukaryotic RNA polymerase III transcription termination.* Science, 2013. **340**(6140): p. 1577-80.
124. Dieci, G. and A. Sentenac, *Facilitated recycling pathway for RNA polymerase III.* Cell, 1996. **84**(2): p. 245-52.

125. Arimbasseri, A.G., K. Rijal, and R.J. Maraia, *Comparative overview of RNA polymerase II and III transcription cycles, with focus on RNA polymerase III termination and reinitiation.* Transcription, 2013. **4**(6): p. e27639.
126. Wang, Z., et al., *Nuclear factor 1 (NF1) affects accurate termination and multiple-round transcription by human RNA polymerase III.* EMBO J, 2000. **19**(24): p. 6823-32.
127. Wang, Z. and R.G. Roeder, *DNA topoisomerase I and PC4 can interact with human TFIIIC to promote both accurate termination and transcription reinitiation by RNA polymerase III.* Mol Cell, 1998. **1**(5): p. 749-57.
128. Ferrari, R., et al., *Distinct roles of transcription factors TFIIIB and TFIIIC in RNA polymerase III transcription reinitiation.* Proc Natl Acad Sci U S A, 2004. **101**(37): p. 13442-7.
129. Paule, M.R. and R.J. White, *Survey and summary: transcription by RNA polymerases I and III.* Nucleic Acids Res, 2000. **28**(6): p. 1283-98.
130. Thompson, M., et al., *Nucleolar clustering of dispersed tRNA genes.* Science, 2003. **302**(5649): p. 1399-401.
131. Roberts, D.N., et al., *The RNA polymerase III transcriptome revealed by genome-wide localization and activity-occupancy relationships.* Proc Natl Acad Sci U S A, 2003. **100**(25): p. 14695-700.
132. Soragni, E. and G.A. Kassavetis, *Absolute gene occupancies by RNA polymerase III, TFIIIB, and TFIIIC in Saccharomyces cerevisiae.* J Biol Chem, 2008. **283**(39): p. 26568-76.
133. Donze, D., *Extra-transcriptional functions of RNA Polymerase III complexes: TFIIIC as a potential global chromatin bookmark.* Gene, 2012. **493**(2): p. 169-75.
134. Dumay-Odelot, H., et al., *Cell growth- and differentiation-dependent regulation of RNA polymerase III transcription.* Cell Cycle, 2010. **9**(18): p. 3687-99.
135. Moqtaderi, Z. and K. Struhl, *Genome-wide occupancy profile of the RNA polymerase III machinery in Saccharomyces cerevisiae reveals loci with incomplete transcription complexes.* Mol Cell Biol, 2004. **24**(10): p. 4118-27.
136. Noma, K., et al., *A role for TFIIIC transcription factor complex in genome organization.* Cell, 2006. **125**(5): p. 859-72.
137. Kleinschmidt, R.A., K.E. LeBlanc, and D. Donze, *Autoregulation of an RNA polymerase II promoter by the RNA polymerase III transcription factor III C (TF(III)C) complex.* Proc Natl Acad Sci U S A, 2011. **108**(20): p. 8385-9.
138. Iwasaki, O. and K. Noma, *Global genome organization mediated by RNA polymerase III-transcribed genes in fission yeast.* Gene, 2012. **493**(2): p. 195-200.
139. Piazza, I., C.H. Haering, and A. Rutkowska, *Condensin: crafting the chromosome landscape.* Chromosoma, 2013. **122**(3): p. 175-90.
140. D'Ambrosio, C., et al., *Identification of cis-acting sites for condensin loading onto budding yeast chromosomes.* Genes Dev, 2008. **22**(16): p. 2215-27.
141. Iwasaki, O., et al., *Centromeric localization of dispersed Pol III genes in fission yeast.* Mol Biol Cell, 2010. **21**(2): p. 254-65.
142. Leitner, A., et al., *Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics.* Mol Cell Proteomics, 2010. **9**(8): p. 1634-49.

143. Leitner, A., et al., *Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography.* Mol Cell Proteomics, 2012. **11**(3): p. M111 014126.
144. Kabsch, W., *Xds.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 125-32.
145. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
146. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.
147. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
148. Vonrhein, C., et al., *Automated structure solution with autoSHARP.* Methods Mol Biol, 2007. **364**: p. 215-30.
149. Zhang, Y. and D.C. Chan, *Structural basis for recruitment of mitochondrial fission complexes by Fis1.* Proc Natl Acad Sci U S A, 2007. **104**(47): p. 18526-30.
150. Hendrickson, W.A. and M.M. Teeter, *Structure of the Hydrophobic Protein Crambin Determined Directly from the Anomalous Scattering of Sulfur.* Nature, 1981. **290**(5802): p. 107-113.
151. Terwilliger, T.C., et al., *Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard.* Acta Crystallogr D Biol Crystallogr, 2008. **64**(Pt 1): p. 61-9.
152. Holm, L. and P. Rosenstrom, *Dali server: conservation mapping in 3D.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W545-9.
153. Kelley, L.A. and M.J. Sternberg, *Protein structure prediction on the Web: a case study using the Phyre server.* Nat Protoc, 2009. **4**(3): p. 363-71.
154. Hildebrand, A., et al., *Fast and accurate automatic structure prediction with HHpred.* Proteins, 2009. **77 Suppl 9**: p. 128-32.
155. Kurowski, M.A. and J.M. Bujnicki, *GeneSilico protein structure prediction meta-server.* Nucleic Acids Res, 2003. **31**(13): p. 3305-7.
156. Stros, M., D. Launholt, and K.D. Grasser, *The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins.* Cell Mol Life Sci, 2007. **64**(19-20): p. 2590-606.
157. Matsutani, S., *Evolution of the B-Block Binding Subunit of TFIIIC That Binds to the Internal Promoter for RNA Polymerase III.* Int J Evol Biol, 2014. **2014**: p. 609865.
158. Castello, A., et al., *Insights into RNA biology from an atlas of mammalian mRNA-binding proteins.* Cell, 2012. **149**(6): p. 1393-406.
159. Bieniossek, C., et al., *MultiBac: expanding the research toolbox for multiprotein complexes.* Trends Biochem Sci, 2012. **37**(2): p. 49-57.
160. Huh, W.K., et al., *Global analysis of protein localization in budding yeast.* Nature, 2003. **425**(6959): p. 686-91.
161. Marguerat, S., et al., *Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells.* Cell, 2012. **151**(3): p. 671-83.
162. Berger, I., et al., *The multiBac protein complex production platform at the EMBL.* J Vis Exp, 2013(77): p. e50159.

163. Joazeiro, C.A., G.A. Kassavetis, and E.P. Geiduschek, *Alternative outcomes in assembly of promoter complexes: the roles of TBP and a flexible linker in placing TFIIIB on tRNA genes.* Genes Dev, 1996. **10**(6): p. 725-39.
164. Kosinski, J., A. Barbato, and A. Tramontano, *MODexplorer: an integrated tool for exploring protein sequence, structure and function relationships.* Bioinformatics, 2013. **29**(7): p. 953-4.
165. Soding, J., *Protein homology detection by HMM-HMM comparison.* Bioinformatics, 2005. **21**(7): p. 951-60.
166. Walzthoeni, T., et al., *False discovery rate estimation for cross-linked peptides identified by mass spectrometry.* Nat Methods, 2012. **9**(9): p. 901-3.
167. Karplus, P.A. and K. Diederichs, *Linking crystallographic model and data quality.* Science, 2012. **336**(6084): p. 1030-3.
168. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.