Structural studies on subunits τ60/τ91 of yeast transcription factor IIIC (TFIIIC).

Anastasia Mylona

Université Grenoble I – Joseph Fourier

THÈSE

présentée par

Anastasia MYLONA

Pour obtenir le grade de

Docteur de l'Université Grenoble I – Joseph Fourier

Spécialité : Biologie

Structural studies on subunits τ60/τ91 of yeast transcription factor IIIC (TFIIIC)

Soutenue publiquement le 18 Juillet 2005

devant le jury composé de:

Dr. Christoph W. Müller	**Directeur de thèse**
Pr. Hans Geiselmann	**Président du Jury**
Dr. Jürg Müller	**Rapporteur**
Pr. Herman Van Tilbeurgh	**Rapporteur**
Dr. Andrea Dessen	**Examinatrice**

Thèse préparée au sein du Laboratoire Européen de Biologie Moléculaire

(European Molecular Biology Laboratory, Grenoble Outstation)

Στους γονείς μου

Acknowledgements

I am grateful to my supervisor Dr. Christoph Müller for his guidance throughout my thesis.

Many thanks to Dr. Carlos Fernández-Tornero for all the precious help and work that he did in the crystallographic part of my PhD thesis and for being a very good teacher.

I also thank Martine Moulin and Uli Steuerwald for their help and advice, Dr. Pierre Legrand for helping to solve the 'phase problem', Dr. Carlo Petosa for many suggestions and for proofreading the manuscript of my thesis, Dr. Patrice Morand for correcting all the french texts in the manuscript, and all the staff, scientific and non-scientific, at the EMBL for their help.

I would like to thank Dr. Joël Acker for all the help and advice in biochemistry, and Dr. Christian Marck for all the interesting discussions and Dr. André Sentenac for the fruitful collaboration we had.

I would also like to thank the members of my thesis advisory committee Dr. Jürg Müller, Dr. Winfried Weissenhorn and Prof. Hans Geiselmann for all the suggestions and discussions.

Σε όλους μου τους φίλους ένα ευχαριστώ γιατί είσαστε πάντα εδώ αυτά τα τέσσερα χρόνια αν και μακριά και γιατί χωρίς εσάς η ζωή δε θα ήταν τόσο ωραία!

Στους γονείς μου ένα μεγάλο ευχαριστώ γιατί είναι πάντα δίπλα μου αλλά και γιατί χωρίς αυτούς οι επιλογές του μυαλού και της ψυχής μου μπορεί να ήταν διαφορετικές.

Στο Νίκο για την αγάπη και την υποστήριξή του όλα αυτά τα χρόνια αλλά και γιατί μου έμαθε να είμαι αισιόδοξη και να ξεχωρίζω και να εκτιμώ τα σημαντικά από τα ασήμαντα πράγματα στη ζωή.

Résumé

TFIIIC est un facteur de transcription de classe III qui se lie spécifiquement et de façon stable sur les boîtes A et B des promoteurs des gènes de l'ARNt. TFIIIC est composé de 6 sous-unités: τ138, τ131, τ95, τ91, τ60 et τ55. Ce travail présente la structure du complexe entre τ60 et la partie C-terminale de τ91 (Δτ91) qui a été résolue à 3.2 Å. La structure comporte trois régions. a) Δτ91 qui est un β-propeller; b) une partie N-terminale de τ60 qui est aussi un β-propeller et qui se trouve entre Δτ91 et c) le domaine C-terminal de τ60, qui a un nouveau rempliement. L'interaction entre Δτ91 et τ60 est formée par les deux β-propellers et la partie C-terminale de τ60 est complètement indépendante. Cette nouvelle interaction β-propeller - β-propeller apparaît importante pour l'assemblage d'un complexe τB stable, capable de se fixer à l'ADN. Nos résultats donnent des informations sur le mécanisme d'assemblage du complexe τB et sa liaison à l'ADN.

Mots de clés :

- transcription
- ARN polymérase III
- TFIIIC
- structure de cristal
- β-propeller

Summary

TFIIIC is a DNA binding complex that serves as an assembly factor in class III genes. It is a multisubunit protein of about 600 kDa organized in two large domains, τA and τB, of similar size and mass. TFIIIC consists of six polypeptides τ138, τ131, τ95, τ91, τ60 and τ55. τ138, τ91 and τ60 make up the τB subcomplex and τ131, τ95 and τ55 belong to the τA subassembly. This work presents the X-ray crystal structure of yeast τ60 complexed with the C-terminal part of τ91 at 3.2 Å resolution. τ60 consists of a N-terminal β-propeller domain and a C-terminal domain that is an overall new fold and very likely is

the part that links τA with τB and contacts the TBP. Δτ91 is also a β-propeller. The interacting region between the two proteins involves residues in the two β-propellers, whereas the C-terminal domain of τ60 does not contribute at all to the interaction. This novel β-propeller – β-propeller interaction appears to be of great importance for the formation of a very stable τB complex able to bind with high affinity and stability to the B box. Our results provide a starting point for further structural and functional studies aimed at elucidating the mechanism of preinitiation complex formation in the RNA polymerase III transcription machinery.

Keywords:

- transcription
- RNA polymerase III
- TFIIIC
- crystal structure
- β-propeller

Contents

Part I

Introduction

Résumé

L'expression de l'information génétique est un processus complexe dont la première étape est la transcription de l'ADN en ARN, catalysée par une ARN polymérase. Cette première étape se décompose en trois phases :

- L'initiation consistant en l'assemblage de la machinerie transcriptionnelle sur un site spécifique du gène à transcrire.
- L'élongation correspondant à la synthèse de l'ARN.
- La terminaison permettant la libération du brin d'ARN et de l'ARN polymérase de la matrice d'ADN.

Chez les eucaryotes, on distingue trois formes d'ARN polymérases, définissant trois classes de gènes :

- L' ARN polymérase I transcrit les gènes de classe I, codant pour l'ARN ribosomique.
- L'ARN polymérase II transcrit les gènes de classe II, codant pour les ARN messagers ainsi que plusieurs petits ARN nucléolaires non traduits.
- L'ARN polymérase III transcrit les gènes de classe III, codant pour l'ARN ribosomal 5S et des petits ARN stables et non traduits tels que les ARN de transfert.

L'ARN polymérase III de la levure *S. cerevisiae* est un complexe protéique de haut poids moléculaire (~700 kDa) avec 17 sous-unités. Les gènes transcrits par l'ARN polymérase III codent pour une variété des petites molécules d'ARN stables (< 400 pb) et non traduits, comme les ARNt, l'ARNr 5S, 7SL, U6, et VA. Beaucoup de ces petits ARN ont une fonction essentielle dans le métabolisme cellulaire. Les promoteurs des gènes transcrits par l'ARN polymérase III ont la caractéristique de contenir fréquemment des séquences essentielles, localisés dans les régions transcrites des gènes. Ils ont pour la plupart une structure discontinue dénommée IRC (Région Interne de Contrôle). Elles sont constituées de blocs de séquences très conservés entre les différents genres et espèces.

Les facteurs généraux de transcription de classe III sont: TFIIIA, TFIIIC et TFIIIB. TFIIIB et TFIIIC sont nécessaires et suffisants pour transcrire la plupart des gènes de classe III in vitro. En revanche, la transcription du gène de l'ARNr 5S nécessite également TFIIIA. TFIIIB est le facteur d'initiation et ne peut se lier qu'à un complexe TFIIIC-ADN. TFIIIC, encore appelé tau (τ), est un complexe protéique de 600 kDa. Il se lie spécifiquement et de façon stable sur les boîtes A et B des promoteurs des gènes de l'ARNt. La boîte B est responsable de l'ancrage de TFIIIC sur le gène et active la transcription en augmentant la capacité de TFIIIC à fixer la boîte. De plus, l'interaction avec la boîte A est une étape importante pour la détermination du site d'initiation de l'ARN polymérase III. TFIIIC a 6 sous-unités: τ138, τ131, τ95, τ91, τ60, τ55.

Cette thèse présente la structure du complexe entre τ60 et la partie C-terminal de τ91. Nos résultats donnent des informations intéressantes sur le mécanisme de l'assemblage de τB complexe et sur sa liaison à l'ADN.

Eukaryotic transcription

The most important and critical step in the transcription of eukaryotic genes is the recognition of promoter elements by RNA polymerase. In eukaryotes, there are three RNA polymerases which are highly related: RNA polymerase I, II and III. Each of these RNA polymerases transcribes specific genes, and each depends on protein factors, known as transcription factors. The role of the transcription factors is to recognize the promoter sequences of the genes transcribed by the RNA polymerases.

RNA polymerase I transcribes the ribosomal RNA genes, which contain a single promoter structure. RNA polymerase II is responsible for the transcription of the mRNA genes as well as of some small nuclear RNA (snRNA) genes. The RNA polymerase II promoters consist of a core region, which is sufficient for directing transcription *in vitro*, and a regulatory region. The regulatory regions show a high diversity in structure, reflecting the large variety in mechanisms by which cellular proteins are synthesized and the need for complex regulation of these mechanisms. The genes transcribed by RNA polymerase III are generally shorter than 400 base pairs (bp) and encode RNA molecules involved in translation, splicing and tRNA processing, such as tRNA, 5S RNA, U6 spliceosomal RNA, 7SL, and RNase P RNA, as well as RNAs of unknown function (White, 1998; Roberts et al., 2003; Harismendy et al., 2003; Moqtaderi and Struhl, 2004)**.** They display a diversity of promoter organization with internal and/or external elements, but are not as diverse as the RNA polymerase II promoters. RNA polymerase III can be specifically recruited at the transcription start site via a cascade of protein-DNA and protein-protein interactions, which vary each time according to the promoter elements.

1.1. RNA polymerase III promoters

In general, the transcription of RNA polymerase III genes is under the control of intragenic DNA sequences (Ciliberto et al., 1983; Sharp et al., 1981). These internal control regions (ICR) are discontinuous structures, with essential "boxes" separated by nonessential nucleotides, and show a high sequence conservation (Pieler et al., 1985a & 1985b). Sequences located upstream of the transcriptional start site also affect the activities of many RNA polymerase III genes in numerous organisms (White, 1998). The main characteristics of these sequences are that they are not conserved between organisms or between RNA polymerase III genes of the same organism, and that the degree to which they determine activity of the genes varies a lot.

According to the existence of different promoter elements and the participation of different transcription factors, RNA polymerase III genes have been divided into three types (reviewed in Geiduschek and Kassavetis, 2001; Willis, 1993; Schramm and Hernandez, 2002) (Figure 1.1):

- Type 1 genes include the 5S RNA genes whose promoters are characterized by three intragenic sequence elements: a 5' A box, an intermediate element (IE) and a 3' C box, which are conserved in the 5S promoters of different species. These elements together span approximately 50 bp, begin at about position +45 and make up the ICR. In the *S. cerevisiae* 5S genes only the C box is required for transcription.
- Type 2 genes include tRNA genes, adenovirus virus-associated (VA) genes, Alu sequences and other short elements. The promoters of these genes are also internal and contain two highly conserved sequence elements, a proximal A box and a more distal B box, within the transcribed region. The conservation between various species probably is partly due to the fact that the two boxes encode the D and T-loops of tRNAs which are required for the function of tRNAs. The A box is located 10-20 bp from the transcription start site, and this reflects a role in start site selection. The distance between the A and B boxes varies. Distances between ~ 30-60 bp are optimal for transcription, although a distance of 365 bp can be tolerated. Another control element, the

polymerase III terminator, is located 20-25 bp downstream from the B box of tRNA genes.

- Type 3 genes include the vertebrate U6, 7SK, hY4, hY5 and H1 snRNA genes, which contain an upstream TATA element. The well-characterized promoter of the human U6 snRNA gene consists of an upstream TATA box, a proximal sequence element (PSE) and a distal sequence element (DSE). The PSE together with the TATA element are involved in the recruitment of the transcription factors (TFs) SNAPc/PTF and a TFIIIB-like activity (TFIIIB-a). The SNAPc complex is also used by RNA polymerase II as well as TFs for the transcription of U1 (and several other) snRNA genes, although the latter do not contain TATA elements. The transcriptional activator Oct-1 is recruited by the upstream DSE and functions in part by promoting binding of the SNAPc complex to the PSE. Interactions between TFIIIB, SNAPc and Oct-1 bound to their promoter elements lead to a stable initiation complex, which is in part mediated by a nucleosome that is located between the DSE and the PSE.

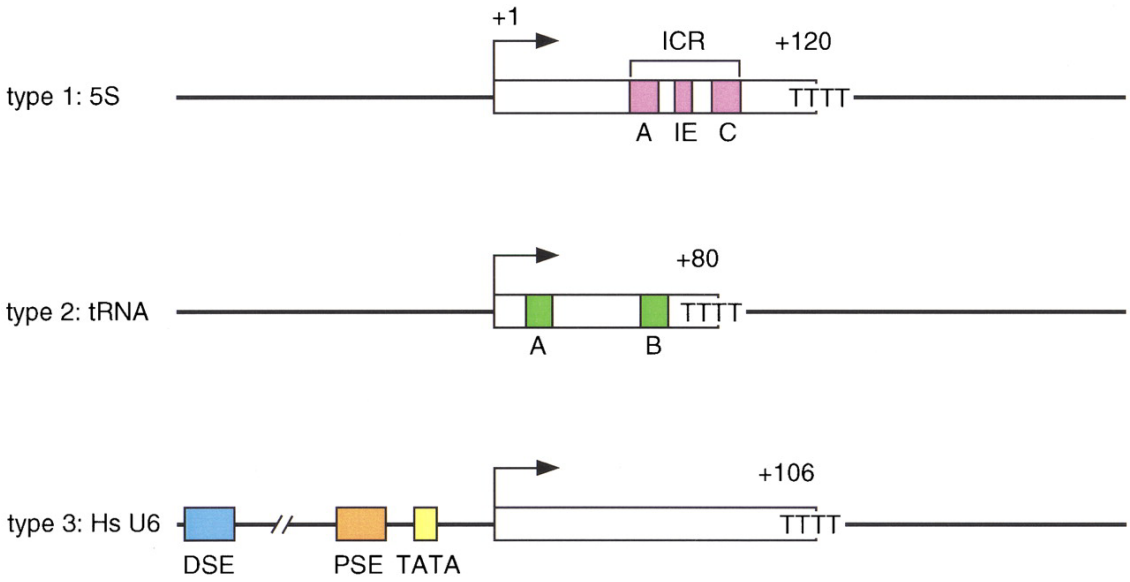


Figure 1.1 Different types of RNA polymerase III promoters. The type 1 promoter of the *Xenopus laevis* 5S RNA gene consists of an internal control region (ICR), which can be subdivided into an A box (+50 to +60), intermediate element (IE, +67 to +72), and C box (+80 to +90). The type 2 promoter of the *X. laevis* $tRNA^{Leu}$ gene consists of an A box (+8 to +19) and a B box (+52 to +62). The type 3 promoter of the *Homo sapiens* U6 snRNA gene consists of a distal sequence element (DSE, −215 to −240) that enhances transcription and a core promoter composed of a proximal sequence element (PSE, −65 to −48) and a TATA box (−32 to −25).

1.2. RNA polymerase III

RNA polymerase III is well defined in *S. cerevisiae* and consists of 17 subunits. All the corresponding genes except for C37 have shown to be essential for cell viability (reviewed in Chedin et al., 1998). Of the 17 subunits, five are unique to RNA polymerase III, and the rest either have paralogues or are identical to RNA polymerase II or I subunits (Table 1.1).

C160 and C128 form the major surface for DNA binding and the active site of the enzyme is located between these two subunits. Subunits AC40, ABC23 and AC19 also belong to the catalytic core. Four additional common polypeptides (ABC27, ABC14.5, ABC10α and ABC10β) are located in the periphery of the complex. The C82, C34, and C31 subunits dissociate from a yeast enzyme carrying a mutation within the zinc finger domain of the largest subunit, and each subunit interacts with the other two in a yeast two-hybrid assay. This suggests that these three subunits form a subcomplex which can be detached from the rest of the enzyme (Werner et al., 1992, 1993). The subunits in the subcomplex are required for initiation (Thuillier et al., 1995; Brun et al., 1997; Wang and Roeder, 1997). C17 and C25 are similar to the RNA polymerase II Rpb4 and Rpb7 subunits (Siaut et al., 2003). C11 is required for the RNA cleavage activity of RNA polymerase III (Chedin et al., 1998) and is thought to interact with subunits C37 and C53.

Table 1.1. Subunit composition of RNA polymerase III.

Module	RNAPIII	RNAPII	Sequence identity (%)
Core subunits	C160	Rpb1	34
	C128	Rpb2	39
	AC40	Rpb3	21
	ABC23	Rpb6	100
	AC19	Rpb11	18
Peripheral subunits	ABC27	Rpb5	100
	ABC14.5	Rpb8	100
	ABC10α	Rpb12	100
	ABC10β	Rpb10	100
Stalk subunits	C25	Rpb7	20
	C17	Rpb4	22
RNAPIII specific complex I	C53	-	-
	C37	-	-
	C11	Rpb9	26
RNAPIII specific complex II	C82	-	-
	C34	-	-
	C31	-	-

1.3. The general factors of transcription system III

General transcription factors specific for RNA polymerase III have been best studied in *S. cerevisiae* and human. Ten essential genes encode the subunits of the *S. cerevisiae* RNA polymerase III transcription factors: TFIIIA; TFIIIC, composed of six subunits; and TFIIIB, consisting of the TATA-binding protein (TBP), the TFIIB-related factor (Brf), and B''(B double prime, Bdp), which is the central RNA polymerase III recruiting factor. TFIIIC recruits TFIIIB onto the DNA upstream of the transcriptional start. In the case of the 5S rRNA genes TFIIIA binds first to DNA and forms the platform for binding TFIIIC to these genes (Table 1.2).

Table 1.2. RNA polymerase III transcription factors.

TFIIIC (6 subunits)	Assembly factor	A & B box (tRNA & 5S RNA genes)	assembly of TFIIIB upstream of the transcription start site
TFIIIA (1 subunit)	Assembly factor	C box (5S RNA genes)	assembly of TFIIIB upstream of the transcription start site
TFIIIB (3 subunits)	Initiation factor	Upstream (tRNA & 5S RNA genes)	responsible for RNA polymerase III recruitment

1.3.1. TFIIIA

In natural RNA polymerase III promoters, TFIIIB is recruited to the DNA mainly through protein-protein interactions with TFIIIC. The type 1 5S promoters and type 2 promoters both use TFIIIC, but on the 5S promoters TFIIIC is recruited through TFIIIA. TFIIIA belongs to the C_2H_2 zinc finger family of DNA-binding proteins (Miller et al., 1985) and contains nine C_2H_2 zinc fingers. In *S. cerevisiae*, the only essential role of TFIIIA is the transcription of the 5S RNA gene, because strains engineered to express 5S rRNA from a tRNA promoter and lacking TFIIIA are viable (Camier et al., 1995). *X. laevis* TFIIIA was the first eukaryotic transcription factor that was purified (Engelke et al., 1980) and the first for which the cDNA was isolated (Ginsberg et al., 1984). Little is known about how TFIIIA recruits TFIIIC to the

DNA. *X. laevis* TFIIIA contains a 14-amino acid domain located at the C-terminal end of the ninth zinc-finger that is not necessary for DNA binding but essential for transcription (Mao and Darby, 1993). In *S. cerevisiae* TFIIIA, a hydrophobic segment in an 84-amino-acid region located between zinc fingers 8 and 9 is also required for cell viability and transcription but not for DNA binding (Rowland and Segall, 1998). These protein domains may play a role in the recruitment of TFIIIC.

1.3.2. TFIIIC

In tRNA genes, the first step during preinitiation complex formation is the binding of TFIIIC to the internal A and B boxes, which is then followed by TFIIIB and RNA polymerase III recruitment, through interactions with TFIIIC. TFIIIB by itself does not bind to RNA polymerase III genes lacking a TATA element but, once it is assembled by TFIIIC, it interacts with DNA, recruits RNA polymerase III and is sufficient for directing accurate initiation by RNA polymerase III during multiple rounds of transcription (reviewed in Geiduschek and Kassavetis, 2001; Schramm and Hernandez, 2002). Hence, TFIIIC acts as an assembly factor (Figure 1.2). However, it has been shown that TFIIIC is a multifunctional protein, involved not only in promoter recognition and TFIIIB recruitment but also in the displacement of nucleosomes to relieve the repression of transcription by chromatin (Burnol et al., 1993). In addition, it is also involved in the termination and reinitiation process, and also very likely in RNA polymerase III recruitment (see 1.5 & 1.6).

Biochemical and genetic studies have led to an extensive characterization of TFIIIC in yeast. *S. cerevisiae* TFIIIC is a multisubunit protein of about 600 kDa organized in two large subassemblies, τA and τB, of similar mass (300 kDa) and size (10 nm diameter), as shown by scanning transmission electron microscopy (Schultz et al., 1989). τB binds with high affinity to the highly conserved B block of tRNA genes, whereas τA binds to the second well conserved A block. This binding is weaker and mostly B block dependent (Baker et al., 1986; Stillman et al., 1984). The link between the τA and τB subcomplexes shows remarkable flexibility, as it can adapt to different distances separating the A and B blocks in natural tRNA genes (Baker et al., 1987). The chromatographic separation of τA and τB has never been observed, but limited

proteolysis yields a τB subcomplex which is active for B block binding (Marzouki et al., 1986). Affinity-purified yeast TFIIIC consists of six polypeptides: τ138, τ131, τ95, τ91, τ60 and τ55 (Parsons and Weil, 1990; Gabrielsen et al., 1989; Bartholomew et al., 1990) (Table 1.3).

τ138, τ91 and τ60 belong to the τB subcomplex (Arrebola et al., 1998; Deprez et al., 1999). Based on protein-DNA cross-linking experiments, τ138 was mapped over the B block and τ91 at the most 3' location of TFIIIC-5S RNA gene complexes (Bartholomew et al., 1990; Braun et al., 1992). These two subunits appear to be predominant for the DNA binding activity (Gabrielsen et al., 1989; Arrebola et al., 1998). τ138 is predicted to have two HMG (High Mobility Group) motifs (Lefebvre et al., 1992) which are DNA binding motifs and it cooperates with τ91 in DNA binding, which is predicted to contain an AT-hook (HMG DNA binding motif) domain at the N-terminal region. A temperature sensitive mutation in τ138 (tfc3-G349E) decreases the affinity of TFIIIC for the B box (Lefebvre et al., 1994). A point mutation in τ91 (τ91-E330K) suppresses the thermosensitive phenotype of the tfc3-G349E mutant. This suppressor mutation rescued the DNA binding and transcription defects of mutant TFIIIC *in vitro* (Arrebola et al., 1998). τ60 is also required together with τ138 and τ91 for DNA binding (Acker J., personal communication), although it is the only TFIIIC polypeptide that could not be cross-linked to DNA. Thus, schematic models for TFIIIC-DNA complexes show τ60 as being unbound to DNA, connecting the τ95 and τ138 subunits and bridging the A and B block. A direct interaction between the TATA Binding Protein (TBP) and τ60 is suggested by coimmunoprecipitation experiments. τ60 very likely participates in TFIIIB assembly via this interaction (Deprez et al., 1999).

τ131, τ95 and τ55 form the τA subassembly (Marck et al., 1993; Rameau et al., 1994; Conesa et al., 1993; Manaud et al., 1998). The main role of τA is transcription activation and start site selection, since it directs the assembly of TFIIIIB (Baker et al., 1987; Joazeiro et al., 1996).

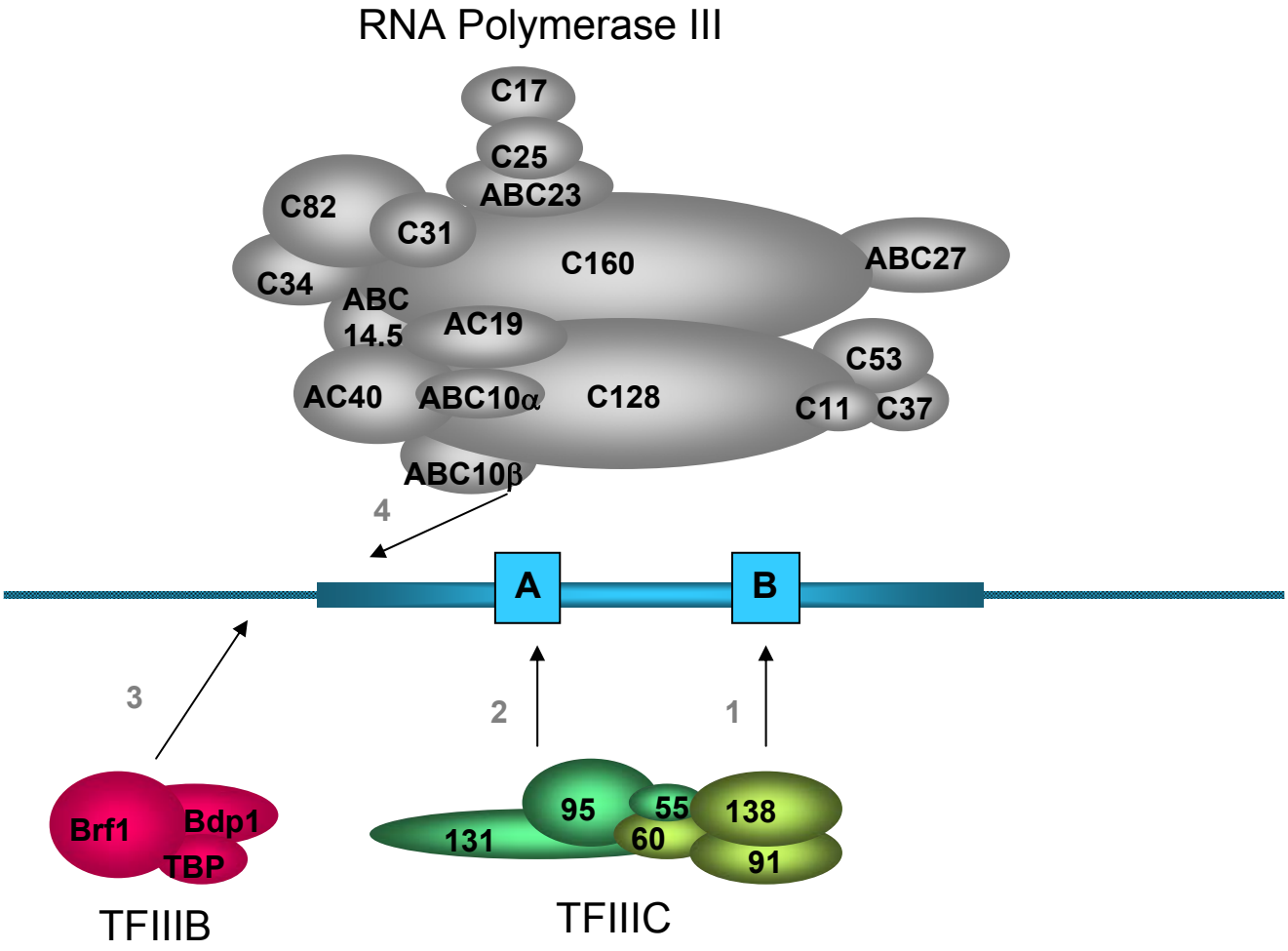


Figure 1.2 Preinitiation complex formation on a tRNA gene. TFIIIC binds to the internal promoters (A and B block) and then recruits TFIIIB upstream of the transcription start site, which finally (possibly together with TFIIIC) recruits RNA polymerase III at the transcription start site.

This is mediated mainly through its subunit τ131, which can move upstream in the TFIIIB binding region (Bartholomew et al., 1990), possibly due to the flexibility provided by its tetratricopeptide repeats (TPR). τ131 interacts directly with both TFIIIB subunits Brf1 and Bdp1 (Chaussivert et al., 1995; Willis, 2002). The assembly of TFIIIB is a dynamic process that involves significant conformational changes in τ131, the TFIIIB subunits as well as the DNA (see 1.3.4). τ95 was mapped over the A block and is thought to be responsible for A block recognition (Gabrielsen et al., 1989; Bartholomew et al., 1990; Hsieh et al., 1999b; Huang et al., 2000). Coimmunoprecipitation experiments suggest that τ95 bridges τA with τB via its interactions with τ138 and τ91 (Jourdain et al., 2003). Little is known about the smallest polypeptide of 55 kDa. τ55, together with τ95, was located by photo-cross-linking experiments in the vicinity of the A block of tRNA genes, on opposite sides of the DNA helix (Bartholomew et al., 1990), and interact with each other (Manaud et al., 1998).

Table 1.3. *Saccharomyces cerevisiae* TFIIIC components and orthologues in *Schizosaccharomyces pombe* and *Homo sapiens*.

TFIIIC homologs/orthologs[1]			**DNA-partner**	**Protein partners**	**Motifs present**
S. cerevisiae	*S. pombe*	*H. sapiens*			
τ95	Sfc1	TFIIIC63	A box	τ55 TFIIIC102, HsBrf1, HsTBP, and HsRPC62	
τ55	None	None	A box	τ95	
τ131	Sfc4	TFIIIC102	upstream of the start site	ScBrf1, ScBdp1, and ABC10α HsBrf1, HsTBP, TFIIIC63	TPR repeats
τ138	Sfc3	TFIIIC220	B box	τ91, τ60, τ95 TFIIIC110	HMG
τ91	Sfc6	TFIIIC110	B box	τ60, τ138, τ95 TFIIIC220	HMG, WD40 repeats
τ60	Sfc9	TFIIIC90		τ91, τ138, TBP TFIIIC220, 110, 63, HsBrf1, HsRPC62, and HsRPC39	

[1] *S. cerevisiae* proteins are indicated in blue, *S. pombe* in orange and *H. sapiens* in pink.

1.3.3. Human TFIIIC

Human TFIIIC was separated by chromatography into two fractions, TFIIIC1 and TFIIIC2 that were both required for transcription of the Ad2 VAI gene (Yoshinaga et al., 1987). However, recently it was reported that human Bdp is the major component of TFIIIC1 (Weser et al., 2004). Thus, human TFIIIC2 is the protein complex which corresponds to yeast TFIIIC. TFIIIC2 consists of five subunits named TFIIIC220, TFIIIC102, TFIIIC63, TFIIIC110, and TFIIIC90 (Table 1.3). The TFIIIC220 subunit is the homolog of τ138 (Matsutani, 2004), which recognizes the B box. TFIIIC220 does not bind DNA on its own, but it is present in TFIIIC subcomplexes that result from cleavage with proteinase C during poliovirus infection and that are still able to bind DNA (Clark et al., 1991). Some of these subcomplexes only consist of the N-terminal part (83 kDa) of TFIIIC220 together with the TFIIIC110 subunit or a fragment of it. This indicates that the DNA-binding domain is located at the N-terminal region of TFIIIC220 and that TFIIIC220 and TFIIIC110 are sufficient for DNA binding (Shen et al., 1996). The TFIIIC110 subunit corresponds to the yeast τ91 protein, although similarity between these two proteins is only evident when they are both compared with the *S. pombe* ortholog (Huang et al., 2000). TFIIIC110 has intrinsic HAT (Histone Acetyl-Transferase) activity, and acetylates free and nucleosomal histones H3 and H4 as well as nucleosomal histone H2B (Kundu et al., 1999). TFIIIC90 is thought to be the functional homolog of τ60 because it interacts with TFIIIC220, TFIIIC110, and TFIIIC63, as well as with the TFIIIB subunit Brf1 (Hsieh et al., 1999a). Thus, τ60/TFIIIC90 appears to bridge the TFIIIC τA and τB subcomplexes as well as TFIIIB. However, TFIIIC90 has the additional property of an intrinsic HAT activity for both free and nucleosomal histone H3, and preferentially acetylates histone H3 Lys 14 (Hsieh et al., 1999a). TFIIIC102 and TFIIIC63 are the τ131 and τ95 homologs respectively. TFIIIC102 interacts with TFIIIC63, and both of them interact with TBP and human Brf1 (Hsieh et al., 1999b).

1.3.4. TFIIIB

Three proteins form the budding yeast TFIIIB complex: TBP, Brf1 and Bdp1. Brf1 was named TFIIB related factor 1 due to its similarity with the RNA polymerase II transcription initiation factor TFIIB. TBP and Brf1 form a very stable complex

(known as B'), from which Bdp1 can be separated by chromatography (B'') (Kassavetis et al., 1991). This indicates that the Bdp1 subunit is less tightly incorporated into the TFIIIB complex. TFIIIB is a DNA binding protein and is considered to function as initiation factor because the TFIIIB-DNA complex by itself is able to efficiently recruit RNA polymerase III *in vitro* (Kassavetis et al., 1990). TFIIIB requires TFIIIC to bind class III genes lacking the TATA box, like most tDNAs, but it can be recruited by itself, *in vitro*, onto the TATA element of the SNR6 gene (Joazeiro et al., 1994; Margottin et al., 1991; Moenne et al., 1990), although TFIIIC enhances the transcription efficiency (Eschenlauer et al., 1993). *In vivo* the transcription of the SNR6 snRNA gene requires TFIIIC to direct TFIIIB binding (probably to help TFIIIB finding the correct orientation for binding to DNA) and probably to overcome the repression of chromatin (Burnol et al., 1993; Gerlach et al., 1995) or the competition of TFIIIB with other soluble DNA-binding proteins. Some tRNA genes that contain a TATA-like element can also be transcribed *in vitro* without TFIIIC (Dieci et al., 2000). The mechanism by which TFIIIC recruits TFIIIB onto class III genes, which do not contain a TATA element, involves different steps of protein-protein interactions and conformational changes. The TFIIIC-DNA complex is thought to interact first with Brf1 via τ131 (Chaussivert et al., 1995; Khoo et al, 1994). Incorporation of TBP in the complex, mediated by Brf1 (Kassavetis et al., 1992; Librizzi et al., 1996) and by τ60 (Deprez et al., 1999), stabilizes the complex via minor groove interactions similar to those seen in RNA polymerase II transcription complexes (Persinger et al., 1999). A masked DNA-binding domain at the C-terminal of Brf1 (covered by domain interactions within Brf1, but unmasked by conformational changes during transcription complex assembly) may contribute to the stabilization of the complex (Huet et al., 1997). Addition of Bdp1 further stabilizes the TFIIIC-TFIIIB-DNA complex (Grove et al., 1999) and leads to more drastic changes in the conformation and stability of the complex: (i) the TFIIIB-DNA complex becomes resistant to high salt or polyanion concentrations (Kassavetis et al., 1989) and (ii) the pattern of cross-linking of τ131, Brf1 and TBP changes (Kassavetis et al., 1992; Persinger et al., 1999; Kumar et al., 1997).

1.4. RNA polymerase III recruitment by TFIIIB

Eight RNA polymerase III subunits can be cross-linked to DNA in a transcription initiation complex (Bartholomew et al., 1993). Of these, C34 is required for transcription initiation (Werner et al., 1993; Wang and Roeder, 1997) and has been localized the furthest upstream from the transcription start site between positions –17 and +6, close to TFIIIB (Bartholomew et al., 1993). Yeast Brf1 interacts *in vivo* and *in vitro* with C34, while human Brf1 interacts with the human C34 homolog, RPC39, *in vitro* (Werner et al., 1993; Khoo et al., 1994; Wang and Roeder, 1997). Yeast Brf1 also interacts with C17 (Ferri et al., 2000). Human TBP interacts with RPC39 *in vitro* (Wang and Roeder, 1997). Interactions between Bdp1 and RNA polymerase III subunits have not been described (Figure 1.3). TBP and Brf1 alone are sufficient to recruit RNA polymerase III and direct multiple rounds of transcription, although the efficiency is only 5% to 10% of that observed with the full TFIIIB complex. Thus, Bdp1 plays an essential role in the opening of promoters (Kassavetis et al., 1998).

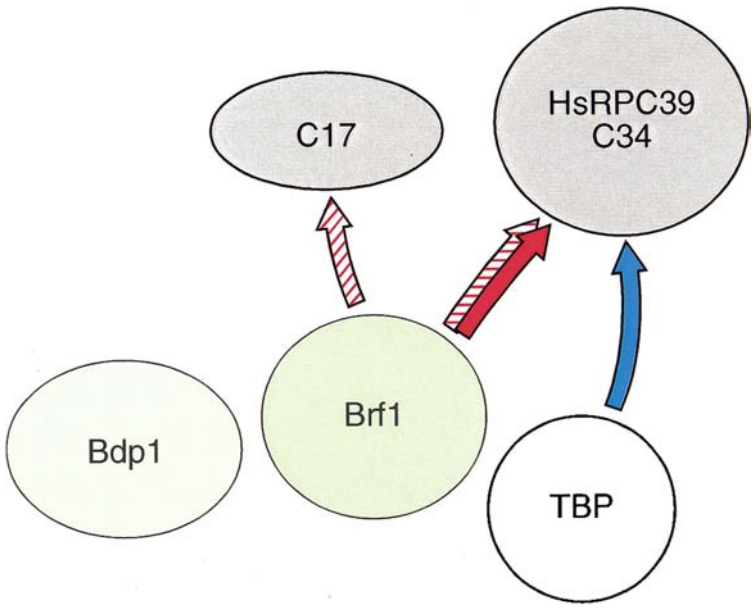


Figure 1.3 Protein-protein interactions between TFIIIB and RNA polymerase III subunits. The solid arrows represent contacts with human subunits, the stippled arrows show contacts with *Saccharomyces cerevisiae* subunits.

1.5. TFIIIC contacts RNA polymerase III

TFIIIC τ131 subunit has been shown to interact directly with RNA polymerase III subunits. τ131 interacts with the RNA polymerase III C53 subunit in a two-hybrid assay (Flores et al., 1999), and with the conserved C-terminal domain of the universal RNA polymerase III subunit ABC10α in a yeast two-hybrid assay as well as *in vitro*. This interaction is likely to be functionally significant because a thermosensitive mutation within the C-terminal domain of ABC10α that makes the interaction weaker can be suppressed by overexpression of a variant form of τ131 that makes the interaction stronger (Dumay et al., 1999). These interactions between TFIIIC and RNA polymerase III show that TFIIIC not only recruits TFIIIB, but may also participate in the recruitment of RNA polymerase III, and/or the cycle of elongation, termination and reinitiation.

1.6. Termination and recycling

RNA polymerase III is unique among the eukaryotic RNA polymerases in recognizing a simple run of T residues as a termination signal. *S. cerevisiae* RNA polymerase III requires a minimum of six or seven dT residues for efficient termination, while vertebrate RNA polymerase III can terminate very efficiently with four dT residues (Geiduschek and Tcchini-Valentini, 1988; Allison and Hall, 1985; Bogenhagen and Brown, 1981). Mutation or deletion of the termination signal yields incomplete transcripts. In addition, mutation of T residues in the VAI gene decreased the efficiency of single- and multiple- round transcription in a HeLa cell extract (Wang and Roeder, 1996; Wang et al., 2000). Similarly, in an *S. cerevisiae* system in which RNA polymerase III was limited, deletion of the terminator affected multiple-round transcription, although single-round transcription was not affected (Dieci and Sentenac, 1996). These results suggest that the T residues can contribute to the efficiency of initiation and reinitiation, and therefore that they may play a role in RNA polymerase recycling. However, the reinitiation process seems to be more complicated, since not only RNA polymerase III, but also TFIIIC, and TFIIIB have been shown to participate in this process. (Ferrari et al., 2004). In higher eukaryotes, a

number of factors seem to be involved in termination and RNA polymerase III recycling, including the La protein, NF1 polypeptides, DNA topoisomerase I, and PC4.

1.7. Scope of this thesis

A thorough understanding of the molecular details underlying preinitiation complex formation on class III genes has been hampered by a lack of structural information. In tRNA genes, which occupy more than 80% of the total number of class III genes, the most critical step during the formation of the preinitiation complex is binding of the τB subcomplex to the B block. This appears to be the most important step since it determines all the following steps until the recruitment of RNA polymerase III, and thus the binding to the B block has to be done with very high affinity and 'fidelity'. Therefore, the scope of this thesis was to shed some light into this mechanism from a structural point of view. The initial goal of this project was the structure determination of the τB complex bound to DNA. However, due to problems with the expression of the τ138 subunit, we decided to work only with the other two subunits, τ60 and τ91, which interestingly we found out that they form a very stable binary complex. In this thesis, the crystal structure at 3.2 Å resolution of the yeast full-length τ60 in complex with a C-terminal domain of τ91 (Δτ91) is presented. Our findings provide insights into the recruitment of the τB complex and its binding to the very well conserved B block.

Part II

Materials and Methods

Résumé

Cette partie est consacrée aux techniques et méthodes utilisées pour la résolution de la structure de la protéine τ60 complexée au domaine C-terminal de la protéine τ91 (Δτ91). Elle comprend l'expression des protéines recombinantes, les étapes de la purification et les méthodes de caractérisation biochimique (protéolyse limitée) ainsi que les principes de la cristallographie et les méthodes utilisées pour le traitement des données de la diffraction, la résolution du problème de phases, et le raffinement du model tridimensionnel.

2.1. Molecular Biology and Biochemical Techniques

2.1.1. Cloning

The cDNA of full-length τ91 was kindly provided by J. Acker. For subcloning τ91, oligos of 25-30 bp length were designed to contain NcoI-XhoI restriction sites. Oligos for PCR were ordered from MWG. For PCR the annealing temperature T_m was chosen to be approximately 5 ^oC below the theoretical T_m of the oligos and the elongation was performed at 72 ^oC. 30 cycles were carried out. The PCR products were purified with the PCR purification kit (Qiagen), digested with restriction enzymes and ligated to the target vector, previously digested with NcoI-XhoI. The ligation was carried out with the Rapid Ligation kit (La Roche) according to the manufacturer's instructions. The ligation product was directly transformed into chemically competent XL1-blue cells (Stratagene) according to the manufacturer's instructions and plated on LB-agarose plates supplemented with the appropriate antibiotics. Plasmids were purified from colonies by using the Wizard mini-prep kit (Promega) and analysed on agarose gels after double-digestions. The sequence of the Δτ91 (residues 159-672) clone, as well as the sequences of all the other τ91 clones, were confirmed by sequencing.

2.1.2. Protein Expression, Purification and Complex Assembly

Recombinant baculoviruses containing full-length τ60 and τ91 genes were kindly provided by J. Acker. The virus was used to infect High Five insect cells at a multiplicity of infection of 1 for τ60 and 2 for τ91. Infection was carried out in TC100 medium (Life Technologies) complemented with 10% fetal bovine serum (FBS). Insect cells were harvested after 3 days by centrifuging at 2800 g for 10 minutes.

The pellet was resuspended in lysis buffer containing 50 mM HEPES [pH 7.4], 30 mM NaCl for τ60 (and 150 mM NaCl for τ91), 5 mM β-mercaptoethanol, 20% glycerol and a Complete protease inhibitor cocktail (Roche Diagnostics), and cells were lysed by three steps of freeze-thawing and sonication at 4 ^{o}C (in the case of τ91, the protein was treated with DNase I before sonication). The cell debris was spun down by centrifugation at 39000 g for 40 min.

For the purification of τ60, the τ60 supernatant was loaded onto a Heparin Hi Trap column (Pharmacia) and eluted with a linear gradient 30-1000 mM NaCl in 50 mM HEPES [pH 7.4], 20% glycerol, 5 mM β-mercaptoethanol, 1 mM PMSF. τ60 was then dialyzed against buffer 1 (50 mM Tris [pH 7.4], 10% glycerol, 5 mM β-mercaptoethanol) and further purified on a MonoQ anion-exchange column (Pharmacia) using a 0-500 mM NaCl gradient in buffer 1.

The crude extract containing τ91 was loaded onto a Heparin Hi Trap column (Pharmacia) and eluted with a linear 200-800 mM NaCl gradient in 50 mM HEPES [pH 7.4], 5 mM β-mercaptoethanol, 1 mM PMSF. τ91 was then dialyzed against 50 mM Tris [pH 7.4], 50 mM NaCl, 5 mM β-mercaptoethanol and further purified on a MonoS cation-exchange column (Pharmacia) using a 200-800 mM NaCl gradient. Finally, both proteins were concentrated and additionally purified over a Superdex 200 HR10/30 gel filtration column (Pharmacia) in buffer 2 (50 mM Tris [pH 7.4], 150 mM NaCl, 5 mM β-mercaptoethanol, 1 mM PMSF).

A fragment encoding Δτ91 residues 159-672 with a C-terminal non-cleavable His tag was expressed from a pET28a vector (Novagen) in *E. coli* strain BL21(DE3)pLysS (Stratagene). This plasmid contains a kanamycin resistance gene for selection of transformed cells and a pBR322 replication origin. The target gene is placed under the control of the T7 RNA polymerase promoter. Bacteria were grown in LB medium at 37 ^{o}C to an optical density of 0.6 at 600 nm, cooled to 30 ^{o}C, induced with 0.3 mM isopropyl-thiogalactoside (IPTG), and grown for an additional 5 hr before harvesting. The bacteria were harvested by centrifugation at 4300 g for 15 minutes. The pellet was resuspended in 50 mM Tris [pH 7.4], 300 mM NaCl, 5 mM β-mercaptoethanol, 3 mM imidazole and a Complete EDTA-free protease inhibitor cocktail (Roche Diagnostics). Cells were lysed by sonication at 4 ^{o}C.

The cell debris was spun down by centrifugation at 48300 g for 30 minutes and the supernatant was incubated with Co-TALON affinity resin (Clontech) previously equilibrated with lysis buffer but without β-mercaptoethanol. Incubation was done at 4 ^{o}C for 1 h under gentle agitation. The resin was then transferred to a column and washed extensively with 50 mM Tris [pH 7.4], 300 mM NaCl, 1 mM β-mercaptoethanol, 8 mM Imidazole, 1 mM PMSF. The protein was eluted with 50 mM Tris [pH 7.4], 300 mM NaCl, 1 mM β-mercaptoethanol, 250 mM imidazole, 1 mM PMSF. Finally, the protein was concentrated and further purified on a Superdex 75 HR10/30 gel filtration column (Pharmacia) in buffer 2.

All the other N-terminal τ91 constructs (residues 168-672, 175-672, 181-672) were cloned in the same way as Δτ91 using the same restriction sites, but were inserted into the pETM11 vector (provided by the Protein Expression and Purification Unit in EMBL/Heidelberg), which contains an N-terminal TEV protease cleavable 6xHis-tag. These constructs were expressed and purified in the same way as Δτ91 except that after the Co-TALON affinity resin, they were dialyzed overnight against 50 mM Tris [pH 7.4], 300 mM NaCl, 1 mM DDT, 1 mM PMSF to remove imidazole. The next day they were cleaved with TEV protease (1:50 w/w) for 7 h at room temperature in the presence of 0.5 mM EDTA. After cleavage, they were loaded again onto the Co-resin, in order to remove the TEV protease and then further purified on a Superdex 75 gel filtration column in the same conditions as described above for τ60, τ91 and Δτ91. For the purification of digested Δτ91-159 and 168 constructs, after the Co-resin they were digested with 1:500 (w/w) elastase at room temperature for 4 h. The reaction was stopped with PMSF (10 mM final concentration) and the proteins were further purified on a Superdex 75 gel filtration column as already described.

The complex was formed at 4 ^{o}C by mixing pure τ60 with τ91 in a ratio 1:1 and was purified over a Superdex 200 HR10/30 column (Pharmacia) in 50 mM Tris [pH 7.4], 75 mM NaCl, 5 mM DTT, 1 mM PMSF. The final concentration was achieved by centrifugation using an amicon concentrator. The interaction between τ60 and τ91 constructs was tested as in the case of the full-length complex.

2.1.2.1. Production of seleno-methionine substituted protein

Seleno-methionine substituted Δτ91 was expressed in the same BL21(DE3)pLysS *E. coli* strain as used for the native protein. Cells were grown at 37 ^{o}C, in minimal medium M9 (supplemented with 20% glucose as unique carbon source) until an OD_{600} of 0.6. The following amino acids and seleno-methionine were added to the culture at the given final concentrations:

- Lys, Phe, Thr: 100 mg/ml
- Ile, Leu, Val: 50 mg/ml
- Se-Met : 60 mg/ml

with an additional growth of 15 minutes at 37 ^{o}C. Then the culture was cooled to 30 ^{o}C and induced with 0.3 mM IPTG and grown for an additional 8 hr. Purification of seleno-methionine labeled Δτ91 was performed as already described for the non-labeled protein, except that 10 mM β-mercaptoethanol was added to all purification buffers to prevent seleno-methionine oxidation.

2.1.3. Determination of protein concentration

Protein concentration was determined in all cases by photometric measurement at λ = 280 nm according to the Beer – Lambert Law:

$$A = \varepsilon I C$$

where A is the measured absorbance.
ε is the wavelength-dependent molar extinction coefficient with units of L mol^{-1} cm^{-1}.
I is the path length of the cuvette in which the sample is contained.
C is the concentration of the protein in solution, expressed in mol L^{-1}.

The molar extinction coefficient ε was calculated for each protein by program ProtParam (www.expasy.org). The ε for the complex was calculated by the equation:

$$\varepsilon = \varepsilon_1 * MW_1 + \varepsilon_2 * MW_2 / MW_1 + MW_2$$

2.1.4. Limited proteolysis

Digestion reactions were performed at room temperature with a protease:protein ratio of 1:500 (w/w). Reactions were started by the addition of freshly dissolved protease, and aliquots of the reaction mixture were withdrawn at regular intervals (5, 10, 20, 45 and 120 minutes). Reactions were stopped by the addition of PMSF to a final concentration of 10 mM and 4X sodium dodecyl sulfate (SDS) protein sample buffer. Samples were boiled for 4 min and analysed on a 10% SDS-polyacrylamide gel electrophoresis (PAGE) followed by coomassie blue staining.

2.1.5. Southern - Western blot

For the preparation of DNA probes, specific or non-specific DNA fragments were PCR amplified and purified on agarose gel. 100 ng of this purified fragment was phosphorylated with PNK T4 kinase (Biolabs) and ATP γ32P, during 1 h at 37 ^{o}C. The DNA was then purified with a PCR purification kit (Qiagen). To obtain a single strand probe, the probe was boiled for 5 minutes and then incubated on ice for 10 minutes, just before hybridization.

10-20 μg of protein sample was loaded on a 8% SDS PAGE. Protein samples were then transferred to a nitrocellulose membrane previously incubated in transfer buffer (25 mM Tris, 40 mM glycine, 10% [v/v] methanol). The membrane containing the protein samples was washed with 20 mM Hepes pH 7.5, 0.1 mM EDTA, 5 mM $MgCl_2$, 100 mM KCl for 30 minutes at 20 ^{o}C and then incubated in blocking buffer (2.5% (v/v) NP40, 1% (w/v) gelatin, 20 mM PBS 1X pH 7.2, 40 mM NaCl, 0.5 mM EDTA, 10% (v/v) glycerol) for 30 minutes at 20 ^{o}C. For the hybridization, the membrane was first washed with hybridization buffer (20 mM PBS 1X pH 7.2, 40 mM NaCl, 0.05% (w/v) gelatin) for 10 minutes at 4 ^{o}C. Then the membrane was incubated with 10 ml of hybridization buffer containing the probe (10^7 cpm) for 1 h at 4 ^{o}C. Finally, the membrane was washed three times for 5 minutes with hybridization buffer and autoradiographed.

2.2. Crystallogenesis and Crystallographic Methods

2.2.1. Crystallization

The most critical step in determining the three-dimensional structure of a macromolecule by X-ray crystallography is obtaining well-diffracting single crystals. Crystallization of macromolecules involves three main steps: nucleation, growth and cessation of growth. It is usually achieved by varying the physical parameters that affect the solubility of the macromolecule in order to achieve a supersaturation state, which will force the macromolecules into the solid state, i.e. the crystal (McPherson, 1992). There are a number of devices, procedures and methods to achieve supersaturation in a protein solution, generally by a slow increase in the concentration of some precipitant such as salt or PEG. Among them, the method of vapor diffusion is undoubtedly the most widely used approach at present. It is ideal for screening a lot of crystallization conditions, and it can be scaled up to obtain large crystals suitable for X-ray diffraction analysis. Furthermore, it is not expensive, it is convenient, and takes advantage of a number of different physical arrangements. The main principle of the method is that a small volume of protein solution is brought into a closed system with a large reservoir solution. A concentration difference between the reservoir and the sample causes vapor diffusion between the two solutions until the vapor pressure in the system is at equilibrium. Hence, the change of conditions in the protein solution can result in the precipitation of the protein. Under the right conditions this happens by the formation of crystals; in most cases, however, by amorphous aggregation. In practice, 0.1-10 μl of the purified protein is mixed with the reservoir solution, which contains a precipitant at a certain concentration. If the ratio is 1:1 both protein and precipitant concentration are halved upon mixing and will return to the initial concentration at equilibrium. However, due to the presence of the precipitant the solubility of the protein can now be lower so that it precipitates.

The two most common vapor diffusion techniques are the sitting drop and the hanging drop method. In the sitting drop method the protein-precipitant mixture is placed onto a small depression or bridge on top of the reservoir. In the hanging drop method, the protein sample is prepared on a cover slip that is turned upside-down before sealing the well (Figure 2.1).

Optimal conditions of crystal growth for biological macromolecules in general are very difficult to determine a priori, since the number of parameters affecting crystallization and the total number of possible solution conditions to be tested is very large (McPherson, 1992). Jancarik and Kim developed a set of 50 screening conditions for initial experiments in protein crystallization, chosen from known or published crystallization conditions of various proteins (Jancarik and Kim, 1991). This and other commercially available crystallization screens (Hampton Research) were automatically dispersed with the Cartesian robot (available at the EMBL/Grenoble) and were used in the first trials to grow the crystals described in this thesis. In all cases, the parameters that were tested were: nature and concentration of the precipitant, protein concentration, composition of protein and reservoir solutions (pH, ionic strength etc.), temperature.

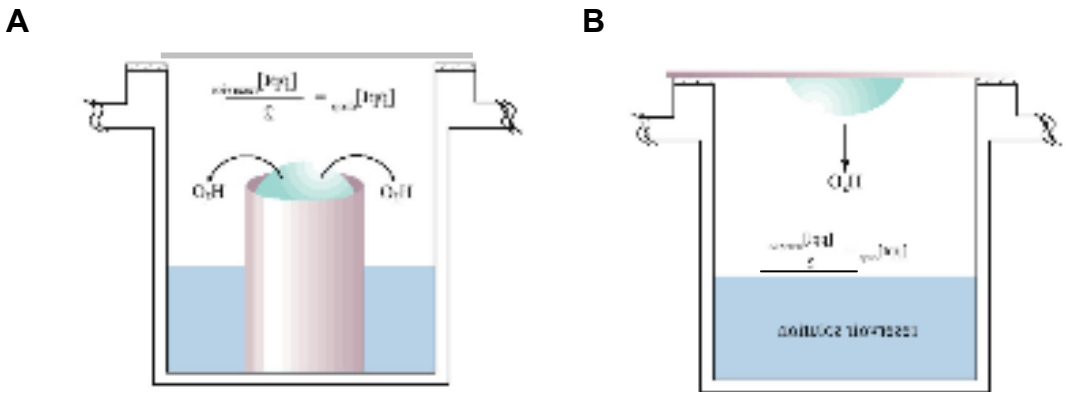


Figure 2.1 A) Sitting drop method, B) Hanging drop method

Crystallization trials were carried out as hanging drops with complex concentrations between 17-20 mg/ml. The mixing ratio was usually 1:1 with 2 μl total drop volume and 0.5 ml reservoir solution. For the full-length complex all the available screens from Hampton research were tested. For the truncated complex, the initial screening was performed by the Cartesian robot (which in our case prepared drops with 0.4 μl total volume) and the conditions that gave the best crystals were repeated manually in 2 μl total drop volume. Conditions where crystals were appearing were further refined by varying the precipitant concentration, protein concentration and pH. The effect of temperature (20 and 4 ^{o}C) was also tested.

Once initial crystallization conditions had been determined, additive screens (Hampton Research) were used according to the manufacturer's instructions (the additive was mixed with the precipitant in a ratio 1:5 and this mixture was subsequently mixed with the protein solution in a 1:1 ratio) with a reservoir solution of 0.8 M ammonium sulfate, 0.1 M Hepes pH 7.0.

For micro-seeding the protein complex (10-20 mg/ml) in 50 mM Tris pH 7.4, 75 mM NaCl, 5 mM DTT was prepared as a hanging drop mixed with 0.8 M ammonium sulfate, 0.1 M Hepes pH 7.0. After the drops were equilibrated, each drop was streak-seeded with a cat whisker. A crystal from a previous preparation was touched with the whisker, the whisker was washed in reservoir solution and immersed into the fresh drop. For micro-seeding of drops containing the τ60/Δτ91 complex, where subunit Δτ91 was SeMet labeled, native crystals were used as seeds.

For co-crystallization with heavy atom derivatives, each heavy atom compound was dissolved in harvesting buffer (0.9 M ammonium sulfate, 0.1 M Hepes pH 7.0) to a final concentration of 1 and 10 mM for all derivatives, apart from the Hg derivatives, where the final concentrations were 0.5 and 5 mM. Protein solution (18 mg/ml) was mixed with each heavy atom solution in a ratio 1:1, to a total 2 μl drop volume. The heavy atom compounds that were tested are listed in Table 2.1.

Table 2.1. Heavy atom derivatives used for co-crystallization with the τ60/Δτ91 complex.

Name	Molecular Formula	Concentration (mM)
Potassium tetrachloroplatinate	K_2PtCl_4	1, 10
Potassium tetracyanoplatinate	$K_2Pt(CN)_4$	1, 10
cis-Dichlorodiamine platinum	Cis-$Pt(NH_3)_2Cl_2$	1, 10
Di-μ-iodobis(ethylenediamine)-di-platinum nitrate	$C_4H_{16}I_2N_6O_6Pt_2$	1, 10
Potassium dicyanoaurate	$KAu(CN)_2$	1, 10
Potassium tetrachloroaurate	$KAuCl_4$	1, 10
Mercury chloride	$HgCl_2$	0.5, 5
Sodium 4-(Chloromercuri)benzensulfonate	$C_6H_4ClHgO_3SNa$	0.5, 5
Phenyl mercury acetate	$CH_3CO_2HgC_6H_5$	0.5, 5
Mercury acetate	$Hg(CH_3COO)_2$	0.5, 5
Ethyl mercurithiosalicylate (Thimerosal)	$C_9H_9HgNaO_2S$	0.5, 5
Potassium hexachloroiridate	K_2IrCl_6	1, 10
Silver nitrate	$AgNO_3$	1, 10
Gadoteridol	$C_{17}H_{29}GdN_4O_7$	1, 10

2.2.2. Introduction to X-ray crystallography

X-ray crystallography is one of the techniques that allow to obtain structural information on biological macromolecules at the atomic or near-atomic level. The most important limitation of the technique is the requirement for diffracting crystals.

A crystal is an ordered three-dimensional array of atoms or molecules. The simplest portion in the crystal which is repeated and shows its fully symmetry is defined as the unit cell. The location of an atom in the unit cell (real space) is usually given by a set of three-dimensional Cartesian coordinates, x, y, and z. One of the vertices (a lattice point or any other convenient point) is used as the origin of the unit cell's coordinate system, and is assigned the coordinates x=0, y=0, z=0 (0,0,0).

According to Bragg's definition of diffraction, during X-ray diffraction of a crystal the diffracted waves from the crystal are treated as reflections from an imaginary set of planes through the crystal. The angle of reflection must be in accordance with the following equation:

$$2d \sin\theta = n\lambda$$

where d is the distance between the planes, θ the diffraction angle, λ the wavelength, and n an integer. Each reflection can be assigned three coordinates or indices (hkl), called Miller indices. The diffraction pattern produced by a crystal constitutes the reciprocal space and its dimensions are the inverse of the real space dimensions. The central reflection in the diffraction pattern is taken as the origin and assigned the coordinates h=0, k=0, l=0, (0,0,0).

The X-ray scattering of an atom is proportional to the electron density of each atom. Every atom in the unit cell contributes to every reflection in the diffraction pattern. The structure factor F_{hkl} for a reflection (hkl) is the sum of the scattering from all the atoms in the unit cell and is a vector:

$$F_{hkl} = \Sigma_{j=1}^{atoms} f_j \exp [2\pi * i(hx_j + ky_j + lz_j)]$$

The f_j is the scattering factor of atom j and depends on the kind of atom and the diffraction angle of the corresponding reflection (hkl). The exponent is complex with x, y, z the fractional coordinates of each atom in the summation, and (hkl) the Miller indices of the corresponding reflection. This complex exponential function is periodic, and with the above parameters it is limited between –1 and 1 for its real part, and –i and i for the imaginary part. In such cases of periodic functions a Fourier transformation can be applied. The Fourier transformation with a set of structure factors as its Fourier coefficient is the electron density:

$$\rho_{xyz} = (1/V) \Sigma_{hkl} |F_{hkl}| \exp [-2\pi * i(hx_j + ky_j + lz_j - \alpha_{hkl})]$$

where V is the volume of the unit cell.

The minus sign preceding the exponent means that the reciprocal space (h,k,l) is transformed into real space (the electron density at point x,y,z). If the structure factors (their amplitudes $|F_{hkl}|$ and their phases α_{hkl}) are known, the actual real structure (the density of the electrons in real space) can be calculated. However, during a diffraction experiment, only the intensities and the positions of reflections are measured. From the position of these reflections, Miller indices (hkl) can be determined and the appropriate intensity assigned. This intensity is proportional to the square of the structure factor amplitude, $|F_{hkl}|^2$. Thus, from the diffraction data we have no information about the phase α_{hkl} of each structure factor. This loss of phase information during data collection is known as the "Phase Problem", which can be solved experimentally by a number of methods, such as multiple isomorphous replacement (MIR), single anomalous dispersion (SAD), multiwavelength anomalous dispersion (MAD), or molecular replacement. Some of these methods will be discussed below (see 2.2.4).

2.2.3. Data collection and processing

2.2.3.1 Cryogenic protection of crystals

Diffraction data collection at cryogenic temperatures from flashed-cooled protein crystals has become routine in macromolecular crystallography. Reduced radiation damage of the crystals, the requirement of a smaller number of crystals to collect complete data sets and solve the crystal structure, the use of intense beamlines at synchrotron radiation facilities for collecting data on weakly diffracting crystals, the ability to transport crystals at the temperature of liquid nitrogen, the possibility to mount and store crystals for extended periods, and the sometimes higher diffraction resolution limit have made flash-cooling of crystals an extremely important technique in macromolecular crystallography (Garman and Schneider, 1997).

For flash-cooling protein crystals prior to data collection, a crystal is mounted on a small loop of non-diffracting material, which is then plunged into a cryogen (such as liquid nitrogen or propane) or directly mounted in a stream of cryogenic gas (Teng, 1990). During data collection crystals are maintained at approximately 100 K in the stream of a cryogenic gas. When a crystal is cooled to cryogenic temperatures the formation of ice must be avoided, because it could lead to the disruption of the crystal lattice, due to the increase of the specific volume during transition from water to ice. In order to avoid ice formation the addition of cryoprotectants to the solution is required (Rubinson et al., 2000).

2.2.3.2. Data collection

Two main types of X-ray sources exist in protein crystallography: X-ray tubes (i.e. rotating anode) and synchrotrons. The main principle of X-ray tubes is that heat applied to a metal cathode produces free electrons. The X-rays are produced when the rapidly moving electrons are suddenly stopped as they strike the metal target of the tube anode. A rotating anode produces more powerful X-rays than fixed tubes. In this case the target is a rapidly rotating metal disk, which prevents overheating of the anode material.

Synchrotrons are enormous rings, where electrons circulate at velocities near the speed of light and are maintained in circular motion by powerful magnets. Because the electrons are forced on their circular orbit, photons are emitted in tangential direction. All the energy is concentrated in a narrow beam yielding X-rays of very high brilliance. Thus, the most important advantage of synchrotron radiation for X-ray diffraction is its high brilliance. This is very advantageous when collecting data on weakly diffracting crystals. Furthermore, in synchrotrons X-rays are produced over a certain range of wavelengths and the optimal wavelength can be selected with the monochromator. This property of synchrotron radiation is for example employed for MAD experiments.

During data collection the intensities of all reflections within a given resolution range are measured. For measuring the reflection intensities in our experiments we used the ADSC Q4 CCD detector. These detectors contain a thin screen of phosphor which emits photons when hit by X-rays. The photons are collected through an optical taper and detected on the CCD array. Phosphoimage detectors are very sensitive to X-rays and are characterized by fast readout times. The 4-module Quantum 4 detector is the latest and most advanced of the commercial CCD detectors from ADSC.

In all our experiments, data were collected using the oscillation technique, where the diffraction image is taken while the crystal is rotated by a rotation angle around an axis (the oscillation axis) that is perpendicular to the X-ray beam. For a complete data set collection, rotation images are recorded until the crystal has been rotated through enough reciprocal space to collect all unique data. The choice of the rotation angle per image depends on the unit cell size (the larger the unit cell the smaller the rotation angle), spot size or mosaic spread, and resolution (the higher the resolution the smaller the rotation angle).

2.2.3.3. Data processing

Data processing involves the following steps:

1. Indexing an image, which means assigning correct Miller indices to each reflection of the image. The positions of the theoretical diffraction spots can be estimated if the accurate cell dimensions, crystal orientation, space group, wavelength and detector distance are known.
2. Refinement of the crystal and detector parameters using least-squares minimization. The differences between the experimental reflection centers are minimized with respect to the theoretical positions.
3. Study of the crystal symmetry and determination of the space group. The crystal symmetry can be deduced from the experimental diffraction pattern and the systematic absence of certain specific reflections.
4. Integration of the diffraction spots and evaluation of each reflection.
5. Crystal parameters post-refinement using all acquired data.
6. Calculation of the relative scale factors among different images in the data set.
7. Statistical analysis of the collected reflection intensities and errors.

The CCP4 suite (CCP4, 1994) is a collection of disparate programs covering most of the computation required for macromolecular crystallography. MOSFLM, SCALA, and TRUNCATE are included in this suite of programs. MOSFLM is a program used to process diffraction images developed by A. G. W. Leslie (Leslie, 1991). This software provides all the tools necessary to carry out the integration of each reflection observed in the experimental X-ray diffraction image. It generates the reflection list, reads the image, integrates the spots and writes the intensities and standard deviations into the mtz file. SCALA (Evans, 1993) is a program that calculates the scale factors among the collected images, adds the partially recorded reflections, rejects the wrong measurements and averages them. In addition, this program performs a statistical analysis of the scaled data. To determine the space group, statistic parameters of symmetry correlation are essential.

The R_{sym} value is an estimation of the degree of accordance among all the reflections related by symmetry operations:

$$R_{sym} = \Sigma_{hkl} \Sigma_j(| I_j - \langle I \rangle |) / \Sigma_{hkl} \Sigma_j I_j$$

where <I> is the average intensity of equivalent reflections and I_j is a general reflection contributing to the average intensity <I>.

TRUNCATE (French and Wilson, 1978) transforms the reflection intensities to structure factor amplitudes and determines the overall temperature factor of the data set from the fall-off in intensity as a function of the resolution.

2.2.4. Methods for solving the phase problem

As already discussed there are several methods to solve "the phase problem" in macromolecular crystallography. Only multiple isomorphous replacement (MIR), single and multiwavelength anomalous dispersion (SAD/MAD) will be discussed here.

1. MIR

This method involves collection of data from crystals of the protein alone, and crystals soaked in various heavy atom compounds. To form a useful derivative, heavy atoms have to bind specifically and with sufficient occupancy to the protein without introducing other changes in the crystal. In this case their locations can be identified, and the phase problem can be solved from the differences in the structure factors between the protein and its heavy-atom derivatives. These differences are referred to as isomorphous differences. Thus, this method requires production of heavy atom derivatives isomorphous to the native crystals.

The exact positions of the heavy atoms can be determined by difference Patterson or difference Fourier techniques. The Patterson function is a Fourier series described as:

$$P(uvw) = (1/V) \Sigma_{hkl} |F(h,k,l)|^2 \exp [-2\pi i (hu + kv + lw)]$$

where u, v, w are the Patterson fraction coordinates. Peaks in the Patterson function correspond to interatomic vectors. Peaks corresponding to vectors between heavy atoms can be calculated by a difference Patterson function using the difference between the structure factor amplitude F_{ph} and the native F_p: $(|F_{ph}| - |F_p|)^2$. Once the positions of the heavy atoms are known the structure factor of the heavy atom F_h can be calculated and subsequently the phases for the native structure factors can be determined.

2. SAD and MAD

These methods take advantage of the heavy atom's capacity to absorb X-rays in function of the energy of the incident X-ray beam. The strongest changes are observed near the characteristic X-ray absorption edges of the heavy atoms. The scattering factor of an atom j can be described as:

$$f_{anom\ j} = f_j + Df_j' + iDf_j''$$

The values for Df_j' and Df_j'' vary with wavelength and type of atom.

Friedel's law states that Friedel pair reflections have equal amplitudes and opposite phases. If there is an anomalous scatterer in the crystal then Friedel's law does no longer hold, and the reflections hkl and –h-k-l are not equal in intensity. This inequality of symmetry-related reflections is caused by the anomalous scattering (or anomalous dispersion). The structure factors of Friedel's pairs will be different in amplitude and phase and by measuring accurately the differences in their structure factor amplitudes we can determine the phase of the native structure factor.

In this approach during a MAD or SAD experiment, the wavelength is varied around the absorption edge of the heavy atom in the crystal. At wavelengths near the absorption edge, there is significant variation in the real and imaginary components f' and f'' of the anomalous scattering of the heavy atoms.

SAD is the simplest case of anomalous dispersion because it only requires a single dataset. The wavelength for the data collection is usually chosen to correspond to the maximum of the absorption edge of the heavy atom present in the crystal which maximizes the anomalous scattering differences f'. However, a SAD experiment does not allow to unambiguously determine the correct phases but instead results in a bimodal phase distribution. Other methods like solvent flattening or NCS averaging are required to resolve the phase ambiguity. In a MAD experiment several data sets are collected around the absorption edge of the heavy atom to maximize differences in f' and f''. In a SAD or MAD experiment all data are collected from the same crystal and thus systematic errors can be reduced (very accurate estimates of the intensity differences are made) and non-isomorphism is not a problem like in the case of MIR.

SHARP (de la Fortelle and Bricogne, 1997) is a computer program that can be used for phasing in macromolecular crystallography. It operates on reduced, merged and scaled data from SIR(AS), MIR(AS) and MAD experiments, refines the heavy-atom model, helps detect minor or disordered sites using likelihood-based residual maps, and calculates phase probability distributions for all reflections in the dataset.

3. Non-crystallographic symmetry (NCS) averaging and Solvent flattening

NCS averaging is one of the most powerful density modification techniques. The presence of multiple copies of the same molecule in the asymmetric unit can be very advantageous. The redundancy present can be used to yield additional phase information. In order to proceed with NCS averaging, firstly we assume that the multiple identical molecules are all in the same conformation and thus should have identical electron density. The individual copies in the asymmetric unit are related by non-crystallographic symmetry operations such as rigid body rotations and translations. It is then assumed that any observed deviations from equivalent densities at corresponding points within all copies must be due to errors. These errors can be removed by introducing the same density value at all related points within the electron density map, and inverting the modified map (averaged over the non-crystallographic symmetry elements) to obtain new structure factors. The new phase information can be combined with prior phase information.

The general procedure for NCS averaging contains the following steps:

1. Determination of the non-crystallographic symmetry.
2. Determination of the volume(s) ("envelopes") within the asymmetric unit in which NC symmetry operations can be applied.
3. Refinement of parameters describing the NC symmetry operators.
4. Obtaining the average electron density at equivalent points within the appropriate envelopes.
5. Replacement of the density at each copy with the corresponding average density.
6. Inversion of the NC symmetry averaged electron density map to obtain new structure factors.
7. Combination of the new structure factor information with prior information.
8. Iteration of steps 4-7.

Solvent flattening, which also is a density modification technique, is usually performed at the same time as NCS averaging. The assumption that is made here is that the solvent in protein crystals is disordered and therefore has a uniform density value.

Any phase information contributes to, and therefore helps to determine the electron density at every point in the unit cell. Likewise, the density at every point contributes to, and helps to determine the phase of every structure factor. If there are errors in the phases of the structure factors the features of the resulting map are unrealistic or even physically impossible. Improvement of this situation can be achieved by eliminating such features in the map or by replacing them with more realistic ones. Fourier inversion of the modified electron density map might then provide better phase estimates for all reflections. These new phase estimates could then be combined with the observed amplitudes and used to create an improved map. By iterating this procedure the electron density map can be significantly improved.

The program DM from the CCP4 suite (CCP4, 1994) combines NCS averaging and solvent flattening and the two procedures run in parallel. In our case the two techniques were performed continuously using program DM.

2.2.5. Building, refinement and validation of the model

2.2.5.1. Model building

Programs O (Jones and Kjeldgaard, 1991) and TURBO (Roussel and Cambilleau, 1989) can be used for building a protein model. They are both graphic programs which allow to visualize the model in stereo. Furthermore, they read and display electron density maps and allow coloring molecules according to different criteria, superimpose different proteins, evaluate the geometry of the molecules, mutate a protein or chemically modify it, and evaluate the resulting conformational changes. The model movements in both programs can be controlled from a dial-box.

2.2.5.2. Refinement

The initial molecular model is manually built into the best possible electron density map. The initial model is normally associated with errors. Consequently, once the first model is obtained, it has to be refined and completed. The refinement minimizes the differences between the experimental structure factors and the ones calculated from the coordinates and temperature factors in the model. The usual indicator used to monitor the quality of the model during the refinement is the crystallographic R factor:

$$R_{cryst} = \Sigma_{hkl} \, || F_{obs} | - | F_{calc} || \, / \, \Sigma_{hkl} \, | F_{obs} |$$

Where each $| F_{obs} |$ is derived from a measured reflection intensity and each $| F_{calc} |$ is the amplitude of the corresponding structure factor calculated from the current model. In cases where the parameters are overfitted to the observed data, the R factor can overestimate the quality of the model. This risk is increased by the fact that during the building of the model, the phases computed from the model are often used to calculate the electron density map for the building cycle. As a result the map is biased by the model used in the refinement. To overcome this bias, the set of reflections are separated into two subsets usually called the working set and the free set, which comprise 5-10% of the total number of reflections.

The R factor calculated using the working set is called R_{work} and the R factor calculated using the free set R_{free}. Only reflections from the working set are used in the refinement. Thus, R_{free} is not model biased and decreases only if the model improves.

2.2.5.3 Electron density maps

The original experimental electron density map obtained from heavy-atom derivatives followed by solvent flattening and sometimes NCS-averaging form the starting point for building the model. The program CNS allows to combine the initial experimental phases with phases calculated from the refinement models. This feature is particular useful at the start of the refinement when the model is still incomplete or contains many errors. Towards the end of the refinement the calculated phases generally are much more accurate than the initial experimental phases and it is no longer useful to combine experimental and calculated phases.

During the refinement of the τ60/Δτ91 complex, we mainly used 2Fo-Fc and Fo-Fc electron density maps (together with the combined density modification maps). Difference Fourier (Fo-Fc) maps are particular useful to detect errors in the model. Positive peaks indicate regions where the model is incomplete and where atoms need to be added, whereas negative densities indicate regions where there is too much scattering material and where atoms need to be removed. Fo-Fc maps emphasize errors in the current model but lack the familiar appearance found in a 2Fo-Fc map, which more resembles a molecular surface. In areas where the maps continue to be ambiguous, it is helpful to examine the original experimental electron density maps.

2.2.5.4. Validation

Once the final molecular model is obtained, it is necessary to analyze its geometry because it can still contain certain errors due to misinterpretations of the electron density maps, particularly in those regions where the electron density is weak. PROCHECK (Laskowski et al., 1993) is a program for validation of protein models, which we used for the validation of the τ60/Δτ91 model. PROCHECK allows a detailed analysis of the stereochemical parameters of a protein structure, by comparing our final model with other structures refined at the same resolution. The most important analyses are the Ramachandran plot (Ramachandran and Sassiekharan, 1968), the χ1-χ2 plots of side chain conformations and the plot of distorted bonds. In a polypeptide the main chain N-C_α and C_α-C bonds are relatively free to rotate. These rotations are represented by the torsion angles phi and psi, respectively. The Ramachandran plot shows the phi-psi torsion angles for all residues of a protein.

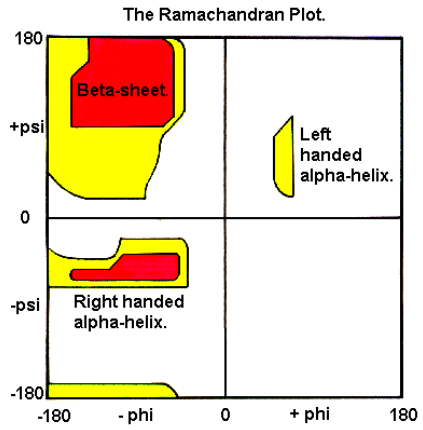


Figure 2.2 The Ramachandran plot

G. N. Ramachandran used computer models of small polypeptides to systematically vary phi and psi with the objective of finding stable conformations. For each conformation, the structure was examined for close contacts between atoms. Atoms were treated as hard spheres with dimensions corresponding to their van der Waals radii. Therefore, phi and psi angles which cause spheres to collide correspond to sterically disallowed conformations of the polypeptide backbone. In the diagram above (Figure 2.2) the white areas correspond to conformations where atoms in the

polypeptide come closer than the sum of their van der Waals radi. These regions are sterically disallowed for all amino acids except glycine which is unique in that it lacks a side chain. The red regions correspond to conformations where there are no steric clashes, i.e. these are the allowed regions namely the alpha-helical and beta-sheet conformations. The yellow areas show the allowed regions if slightly shorter van der Waals radi are used in the calculation, i.e. the atoms are allowed to come a little closer together. This brings out an additional region which corresponds to the left-handed alpha-helix.

L-amino acids cannot form extended regions of left-handed helix but occassionally individual residues adopt this conformation. These residues are usually glycine but can also be asparagine or aspartate where the side chain forms a hydrogen bond with the main chain and therefore stabilizes this otherwise unfavourable conformation. The 3(10) helix occurs close to the upper right of the alpha-helical region and is on the edge of allowed region indicating lower stability.

Disallowed regions generally involve steric hindrance between the side chain C_β methylene group and main chain atoms. Glycine has no side chain and therefore can adopt phi and psi angles in all four quadrants of the Ramachandran plot. Hence, it frequently occurs in turn regions of proteins where any other residue would be sterically hindered.

Part III

Results

Résumé

Le projet a débuté par la caractérisation biochimique des deux protéines τ60 et τ91, indépendamment, et du complexe qu'elles forment. Nous avons observé que isolément et dans le complexe, la partie N-terminal de τ91 était flexible et pas importante pour l'interaction avec τ60. Nous avons alors cloné un fragment de τ91, sans cette partie N-terminal (Δτ91) et nous avons cristallisé le complexe entre τ60 et Δτ91. Nous avons obtenu des cristaux en présence de 0.8 M ammonium sulfate et de 0.1 M Hepes pH 7.0 avec de la sulfobetaine. Les cristaux diffractaient jusqu'à une résolution de 2.8 Å. Le complexe a été co-cristallisé avec $KAu(CN)_2$ et la structure du complexe a été résolue par diffusion anomale (SAD) et par moyennation due aux symétries non cristallographiques et lissage de solvent. La structure a été affinée à 3.2 Å de résolution. La structure est formée par trois régions: Δτ91 qui est un β-propeller; une partie N-terminal de τ60 qui est aussi un β-propeller et qui se trouve entre Δτ91 et le domaine C-terminal de τ60, qui est un repliement nouveau. L'interaction entre Δτ91 et τ60 est formée par les deux β-propellers et la partie C-terminal de τ60 est complètement indépendante. Les deux propellers se ressemblent beaucoup mais Δτ91 est caractérisé par une extension N-terminal qui traverse la surface du propeller d'une coté. L'interaction entre les deux protéines est electrostatique. La partie C-terminal de τ60 est formée par des hélices alpha et par des feuillets beta qui ensemble participent à un core hydrophobique. Ce domaine n'interagit avec aucun propeller. Elle est liée à la partie N-terminal de τ60 par un linker.

3.1 Expression, proteolytic analysis, reconstitution and crystallization of the τ60/τ91 complex

3.1.1. Expression and purification of τ60 and τ91 proteins

TFIIIC subunits τ60 and τ91 were expressed in High Five insect cells and purified independently (see Materials and Methods). Typically, from 400 ml cell culture (which was the minimum culture volume used for each protein during one complex preparation) the final yield after 3 conventional chromatographic steps was 5-6 mg pure protein for τ91 and 3-4 mg for τ60. Both proteins elute as a monomer on a Superdex 200 gel filtration column. Attempts to crystallize the two proteins separately were unsuccessful. For each protein all available Hampton screens were tested. Only in the case of τ60 very small crystals were obtained, but unfortunately these showed no diffraction.

3.1.2. τ60 and τ91 form a stable binary complex

Although τ91, τ60 and τ138 are known to form the τB subcomplex (see Introduction), no stable binary complexes of these proteins were known to exist. Since it was more interesting to try to crystallize the complex of τ60 and τ91 rather than the individual components, we tested the interaction of the two pure proteins by mixing them, concentrating them and injecting them on a Superdex 200 column. The two proteins co-elute as a single peak on gel filtration, indicating the formation of a stable complex (Figure 3.1). According to the elution volume of the τ60/τ91 peak, the complex forms a heterodimer. The τ60/τ91 complex can also be reconstituted by co-expressing (co-infection of the insect cells with the two recombinant viruses) the two proteins in insect cells and co-purifying them, but in this case the purity of the complex was not sufficient for crystallization.

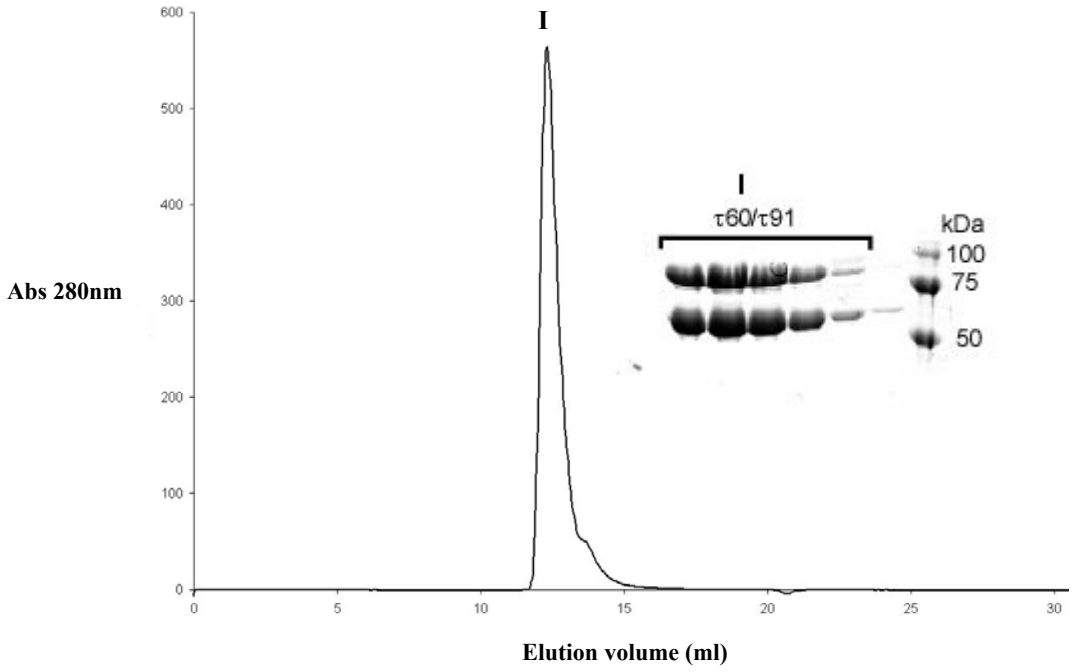


Figure 3.1 Subunit τ60 forms a stable complex with τ91: size exclusion chromatography profile of the τ60/τ91 complex. Peak fractions were analysed by SDS-PAGE.

3.1.3. Crystallization of the τ60/τ91 full-length complex

For crystallization trials the τ60/τ91 complex was concentrated to 16 mg/ml and all available screens from Hampton Research were tested. 1 µl of the complex was mixed with 1 µl of the reservoir solution in all cases. The τ60/τ91 complex crystallized in the presence of 0.8 M ammonium sulfate, 0.1 M Tris pH 8.0. Crystals grew as plates with maximal dimensions of 50µm x 50µm x 20µm, after 5 days at 20 ^{o}C (Figure 3.2). At the microfocus beamline ID13 of the European Synchrotron Radiation Facility (Grenoble, France) these crystals maximally diffracted at 4 Å resolution. The space group was assigned as $P2_1$ with unit cell parameters of a = 61.8 Å, b = 126.9 Å, c = 217.3 Å, β = 96.1º. However, crystals usually grew as stacked clusters and were not reproducible.

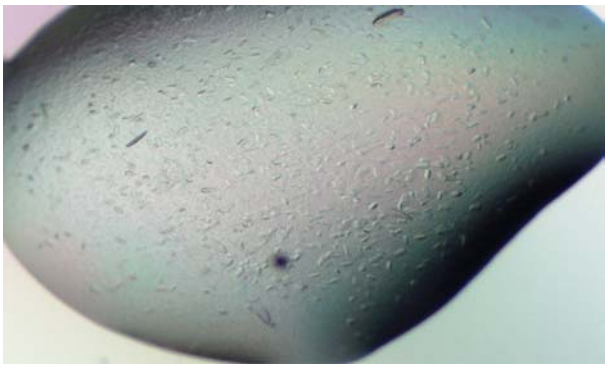


Figure 3.2 Microcrystals of the full-length τ60/τ91 complex.

3.1.4. Instability of the τ60/τ91 complex at 20 °C

The irreproducibility of the τ60/τ91 crystals could be due to the presence of flexible regions in the complex which inhibit crystallization. For this reason the possibility of crystallizing a more stable subcomplex was explored. τ91 is known to contain WD40 repeats in its C-terminal region (Huang et al., 2000) but due to the variation in sequence of the repeats, it is very difficult to predict the number of repeats and to detect their precise limits within the sequence (Andrade et al., 2001). No structural domains at all could be predicted for τ60 using different domain databases. Due to the lack of structural information, the sensitivity of the protein complex to proteolysis was explored, in order to define a minimal binary complex. The pure τ60/τ91 complex (in 50 mM Tris pH 7.4, 75 mM NaCl, 5 mM DTT) was incubated at 20 °C for 5 days (which corresponds to the time required for first crystals to appear) and subsequently analyzed by SDS PAGE. τ60 was resistant to proteolysis but τ91 seemed rather sensitive and was degraded to a fragment which migrated with an apparent molecular weight of 60 kDa (Figure 3.3).

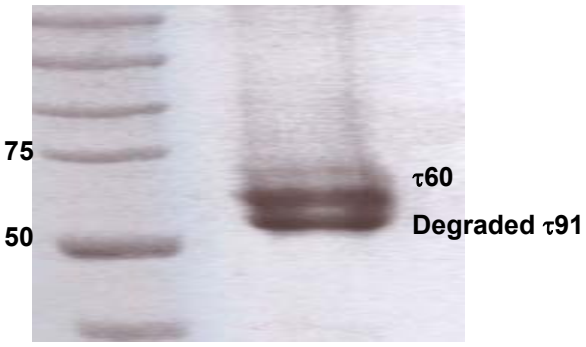


Figure 3.3 τ91 is degraded at 20 °C to a 60 kDa fragment whereas τ60 appears to be stable

The same proteolytic fragment could be obtained by subjecting τ91 to limited proteolysis using various proteases at 20 °C. Trypsin, chymotrypsin, elastase and subtilisin were tested in a 1:500 (w/w) protease:protein ratio at room temperature. Trypsin cleaved most specifically and reproducibly, yielding a proteolytic fragment of approximately 60 kDa that was stable even after 2 hours of digestion (Figure 3.4A). N-terminal sequencing (performed at the Institut de Biologie Structurale (IBS) /Grenoble) and mass spectrometry (performed at EMBL Heidelberg) showed that the cleavage occurred between residues 158 and 159, leading to the N-terminally truncated fragment Δτ91 (residues 159-672, MW=57789 Da). We also observed several fragments with molecular masses smaller than 20 kDa, which probably resulted from the digestion of the N-terminal moiety of τ91 (Figure 3.4A). However, the intensities of these bands were weak and no attempts were made to further characterize these fragments.

Proteolysis of τ60 using trypsin resulted in two fragments of approximately 40 kDa (τ60-I) and 28 kDa (τ60-II) molecular mass (Figure 3.4B), which together approximately correspond to the total molecular weight of τ60. N-terminal sequencing revealed that the larger fragment starts at amino acid residue 2, whereas the smaller fragment starts at residue 360. Tryptic digestion of the τ60/τ91 complex under the same conditions also resulted in the C-terminal τ91 fragment and the two τ60 fragments (Figure 3.4C).

Using gel filtration, τ91 and the N-terminally truncated fragment Δτ91 elute at volumes corresponding to their expected molecular weights of ca. 80kDa and 55kDa, respectively. Interestingly, proteolytically cleaved τ60 eluted by gel filtration at the same volume as the intact protein (Figure 3.5). This suggests that the protease cleaves a flexible loop on the protein surface without further modifying the structure of τ60. Thus, τ91 and τ60 show distinct sensitivities towards proteolysis: τ91 has a well structured C-terminal moiety resistant to proteolysis, and an apparent flexible N-terminal moiety, which is rapidly degraded during digestion. In contrast, τ60 is a rather well structured protein, without any N- or C-terminal extensions susceptible to proteolysis (Figure 3.6). The complex of the two proteins is sensitive to proteolysis, but it seems that the most flexible part of the complex is the N-terminal part of τ91, since τ60 is stable to degradation at room temperature, even after 5 days.

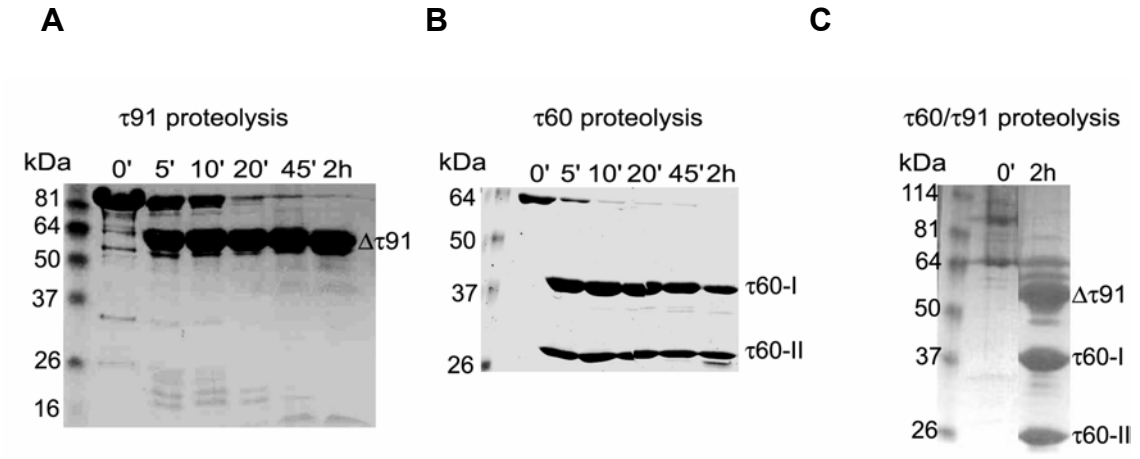


Figure 3.4 Limited proteolysis of τ91, τ60 and the τ60/τ91 complex. Analysis of recombinant (A) τ91, (B) τ60 and (C) τ91/τ60 complex fragments following digestion with trypsin. Digestion times are given above each lane. Proteolysis defines a stable C-terminal domain of τ91 and two distinct domains, τ60-I and τ60-II, for τ60. The same proteolytic fragments are obtained by limited proteolysis of the complex.

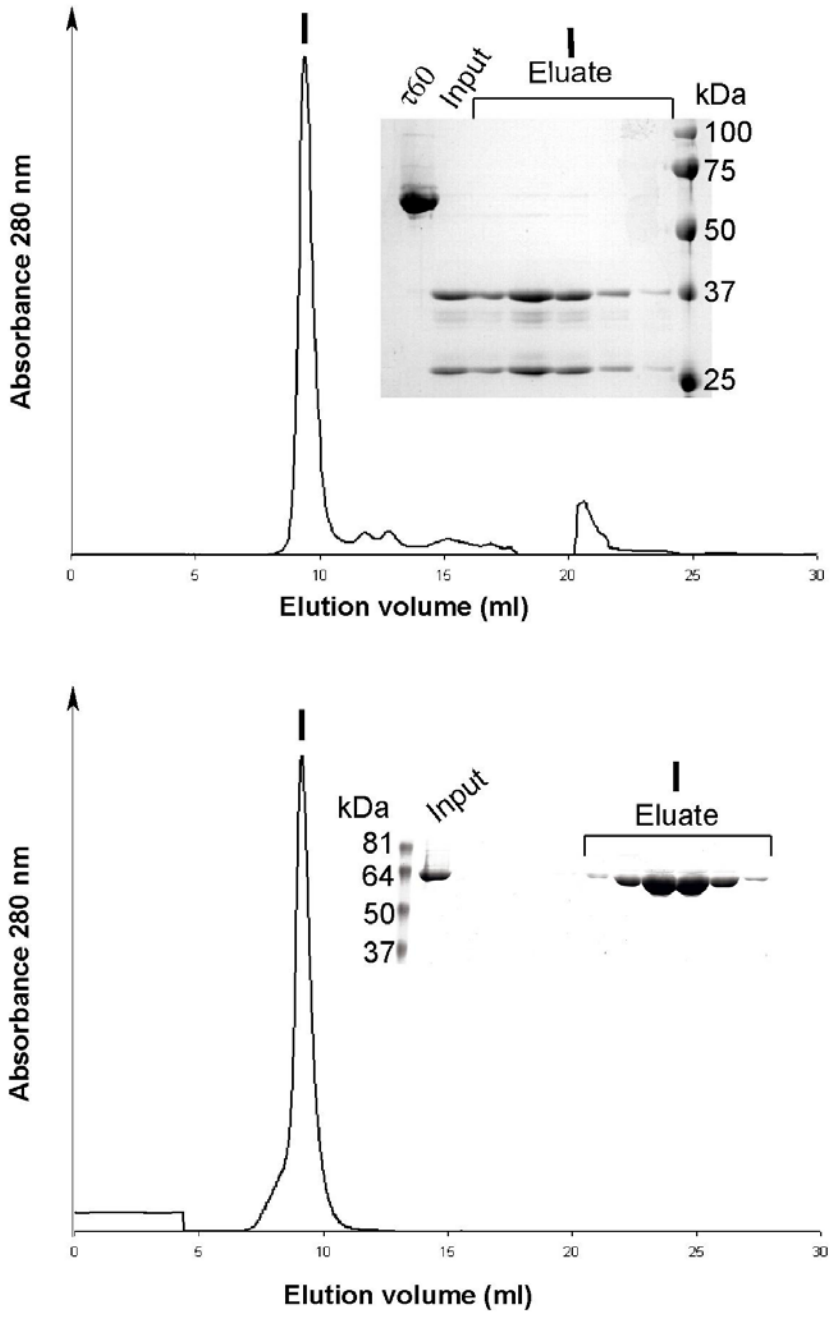


Figure 3.5 Fragments τ60-I and τ60-II co-elute on a Superdex 75 HR10/30 gel filtration column at the same elution volume as full-length τ60, suggesting that the two fragments remain associated after cleavage.

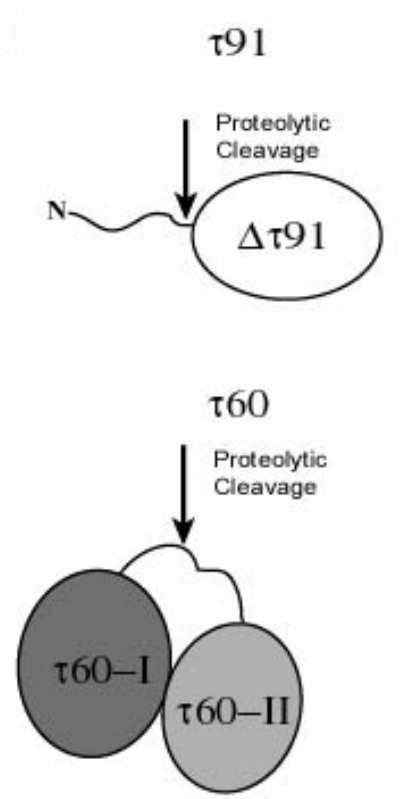


Figure 3.6 Schematic representation of the different domain structures of τ91 and τ60 revealed by limited proteolysis.

3.1.5. Expression and purification of the protease resistant C-terminal τ91 fragment (Δτ91) followed by τ60/Δτ91 complex formation

The results from limited proteolysis were used to subclone the C-terminal fragment of full-length τ91. A gene coding for Δτ91 was constructed by deleting the first 158 amino acids of full-length τ91. The gene was fused to the T7 promoter sequence of the pET28a plasmid, and Δτ91 was expressed in and purified from *E. coli*. The final yield for pure Δτ91 was ~ 2 mg/L culture. For the reconstitution of the τ60/Δτ91 complex the same approach was used as for the full-length complex. After purifying each protein independently (see Materials and Methods) the two proteins were mixed and purified further on a Superdex 200 gel filtration column. This yielded two clear

peaks in the elution profile, the first one corresponding to the truncated complex and the second one corresponding to the excess of τ60 (Figure 3.7). This result indicated that removal of the first 158 residues of τ91 does not change the interaction with τ60. It is very possible that all the interactions between τ91 and τ60 come only from the C-terminal part of τ91 and that the first 158 residues do not contribute to this interaction. If this is the case, it would explain why this part remains sensitive to proteolysis even when complexed with τ60, implying that binding of τ60 does not cause steric hindrance that would protect the N-terminal fragment of τ91 from cleavage.

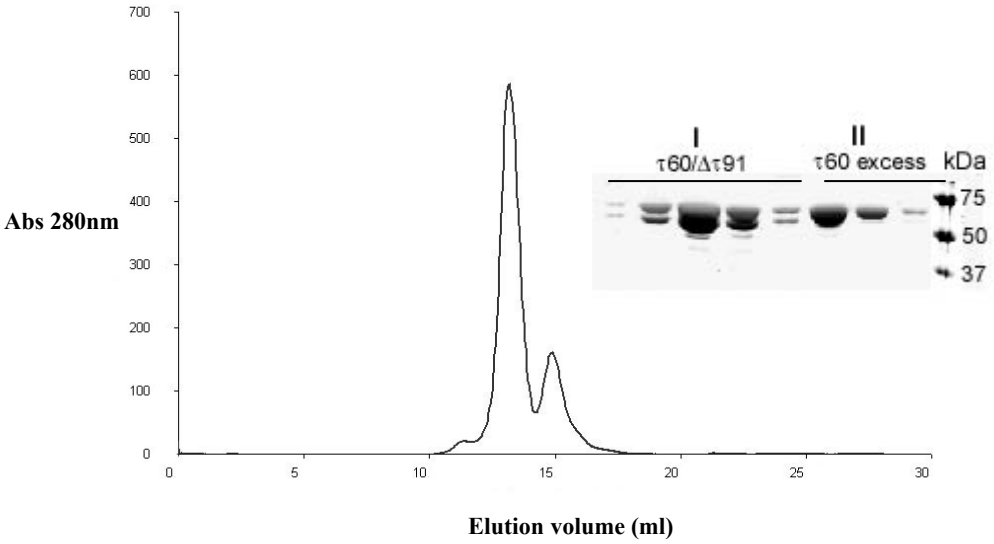


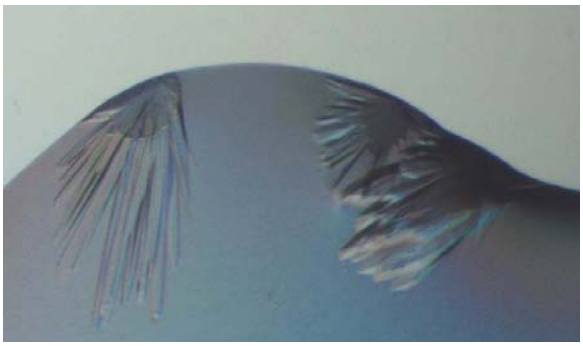
Figure 3.7 Size exclusion chromatography showing the formation of the τ60/Δτ91 complex. Peak fractions were analysed by SDS-PAGE.

3.1.6. Crystallization of the τ60/Δτ91 subcomplex

For crystallization trials the τ60/Δτ91 complex was concentrated to 18 mg/ml. In general, a concentration above 15 mg/ml was essential for obtaining crystals. Initial screening was performed using the Cartesian robot (in this case the protein complex was mixed with the precipitant in a 1:1 ratio with 0.4 µl total drop volume) at 20 and at 4 ^{o}C. Conditions that gave crystals were then repeated and refined manually, where the complex was mixed with the precipitant in a ratio 1:1 using 2 µl total drop volume. Crystals of the truncated complex appeared with many different precipitants: ammonium sulfate, phosphate, sodium formate, PEG 400 (15-20%), PEG 8000 (0.5-

5%), KNaTartrate, and isopropanol. However, crystals were either 2-dimensional plates (pH 7.0-7.5) or needle-like (pH 8.0). The best-diffracting plates grew out of 0.8 M ammonium sulfate, 0.1 M Hepes pH 7.0 at 20 ^{o}C after 5 days (Figure 3.8A), diffracting at 3.2 Å when tested on the ESRF beamlines (Figure 3.8B). The space group was assigned as $P2_1$ with unit cell parameters a = 60 Å, b = 126.2 Å, c = 210 Å and β = 94.3^{o} using program MOSFLM (Leslie, 1991). Unfortunately, the τ60/Δτ91 crystals were very sensitive to radiation damage and we were not able to collect a complete data set.

A



B

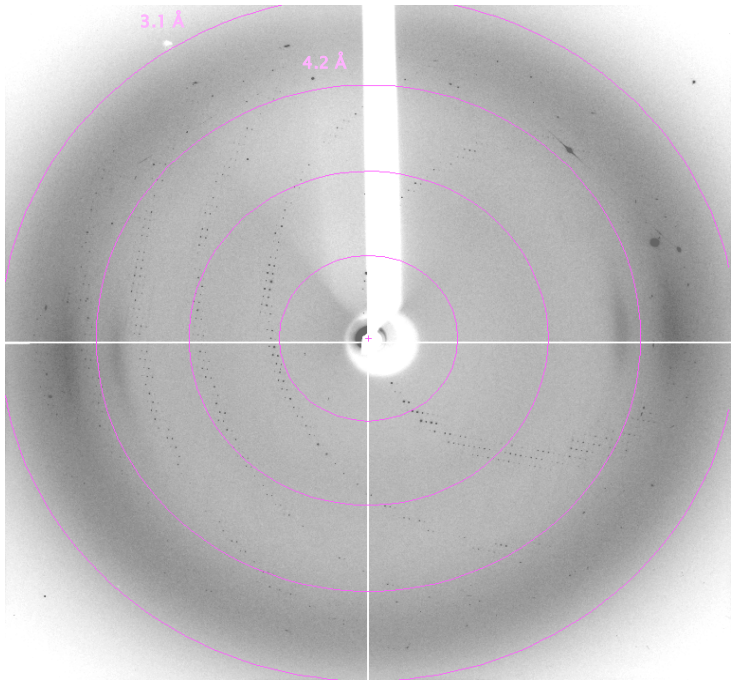
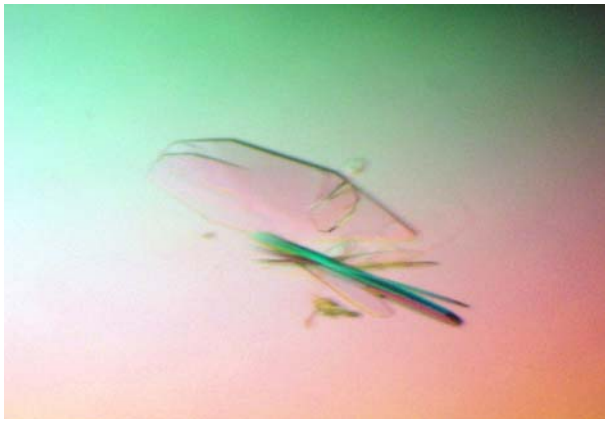


Figure 3.8 A) Crystals and B) Diffraction pattern of the τ60/Δτ91 complex.

In order to obtain better crystals with a third dimension and resistant to radiation damage, we varied the protein and precipitant concentrations, the pH, and the temperature, and we also tried seeding, but there was no improvement in the crystals. In a next step, we tried additive screens I, II and III (Hampton research) using as precipitant 0.8 M ammonium sulfate and 0.1 M Hepes pH 7.0. The non-detergent sulfobetaine-195 additive yielded plates that were much bigger and thicker. These crystals reached a maximum size of 700µm x 400µm x 50µm after 10 days (Figure 3.9A). Crystals diffracted to 2.8 Å in the a* direction and to about 3.2 Å in the b* and c* directions (Figure 3.9B). The space group of these crystals is $P2_1$ with a = 61.4 Å, b = 125.8 Å, c = 210.5 Å, β = 94.5º. The difference in unit cell parameters between these crystals and those grown without sulfobetaine is that in the latter case the a-axis is shorter by 1.4 Å, or 2.3%. Interestingly, crystals of the τ60/τ91 and the τ60/Δτ91 complex also show very similar cell parameters. The main difference is observed along the c-axis, which in the τ60/Δτ91 complex crystals is shorter by 7 Å, or 3%. This indicates that the packing is similar in both crystal forms and that the N-terminal domain of τ91 does not mediate important crystal contacts. From the crystals grown with sulfobetaine we collected a full data set at 3.2 Å resolution that was used to solve the τ60/Δτ91 structure (see 3.2).

A



B

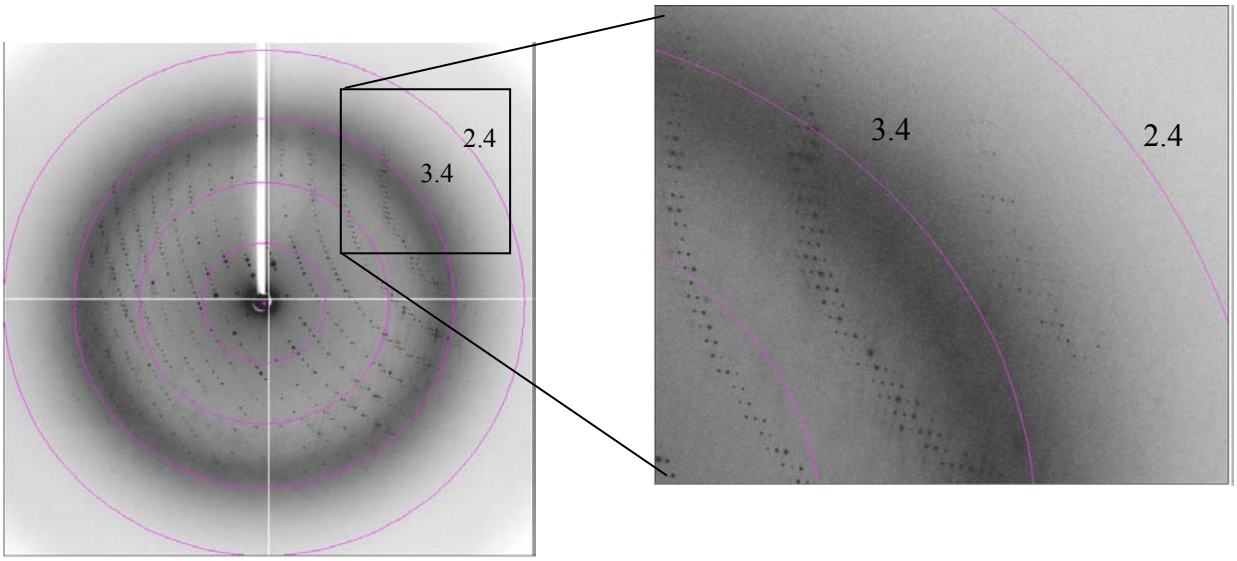


Figure 3.9 (A) Crystals of the τ60/Δτ91 complex grew as two-dimensional plates of 0.7 mm x 0.4 mm x 0.05 mm. (B) 0.75^{o} oscillation image of these crystals collected at beamline ID29 of the ESRF synchrotron at 100 K. The insert shows the presence of reflections extending beyond 3.0 Å resolution.

3.1.6.1. Attempts to improve the resolution of the τ60/Δτ91 crystals

The quality and size of crystals was better for certain protein preparations, in which I could observe by SDS-PAGE a band of ~52 kDa that co-eluted with the binary complex as well as with the excess of Δτ91 during the final gel filtration (Figure 3.10).

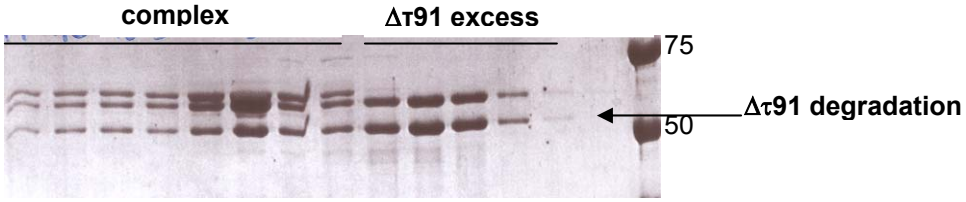


Figure 3.10 A 52 kDa Δτ91 degradation fragment co-eluting with the τ60/Δτ91 complex and with the excess of Δτ91 on a Superdex 200 gel filtration column.

N-terminal sequencing and mass-spectrometry analysis revealed that this protein fragment was due to an N-terminal truncation of Δτ91. This indicated that very likely there was still a flexible part at the N-terminal region of Δτ91, and that by further shortening the N-terminus of the protein, the complex could pack better inside the crystals. Based on the N-terminal sequencing results, three more constructs of τ91 were prepared (168-672, 175-672**,** and 181-672 respectively). All constructs had a cleavable His-tag at the N-terminus and were produced in *E.coli.* Whereas the 181-672 construct did not express at all, the other two expressed well and were soluble. The construct starting at position 168 was better expressed than the one starting at position 175, and was chosen for reconstitution of the complex. Complex preparations with and without the His-tag were purified and screened again for crystallization using the Hampton screens. The crystals that were obtained were either plates with the same space group and unit cell dimensions as the original crystals or were voluminous crystals with a very large unit cell (c ~ 700 Å!), which diffracted only weakly (Figure 3.11A).

As a next step, limited proteolysis experiments were performed on Δτ91 (the original construct starting at position 159) to define an even more stable fragment for crystallization. Different proteases were tested and of these elastase gave the clearest and most specific result. N-terminal sequencing and mass spectrometry showed that elastase led to a C-terminal truncation of Δτ91, again yielding a molecular weight of 52 kDa. Based on these results, the protein complex was prepared using full-length τ60 and elastase-treated Δτ91 and drops were set up. Bigger plates were obtained but again the resolution, space group and unit cell dimensions were the same as with the original τ60/Δτ91 complex. Δτ91-168 was also subjected to elastase treatment, in the hope that the combination of an N- and C-terminal truncation would yield a better crystal form. By using this strategy, nice 3-dimensional crystals were obtained (Figure 3.11B) but unfortunately they belonged to the same crystal form with very large cell parameters (described above). In all cases crystals were grown at 0.8 M ammonium sulfate and 0.1 M Hepes pH 7.0-8.0.

These experiments were taking place while we had already started building the model using native data at 3.2 Å resolution. During model building, we did not observe any disordered regions in τ60, which indicated that there was no need to truncate this protein. This was an additional reason for which we were concentrating our efforts on the optimisation of the Δτ91 protein.

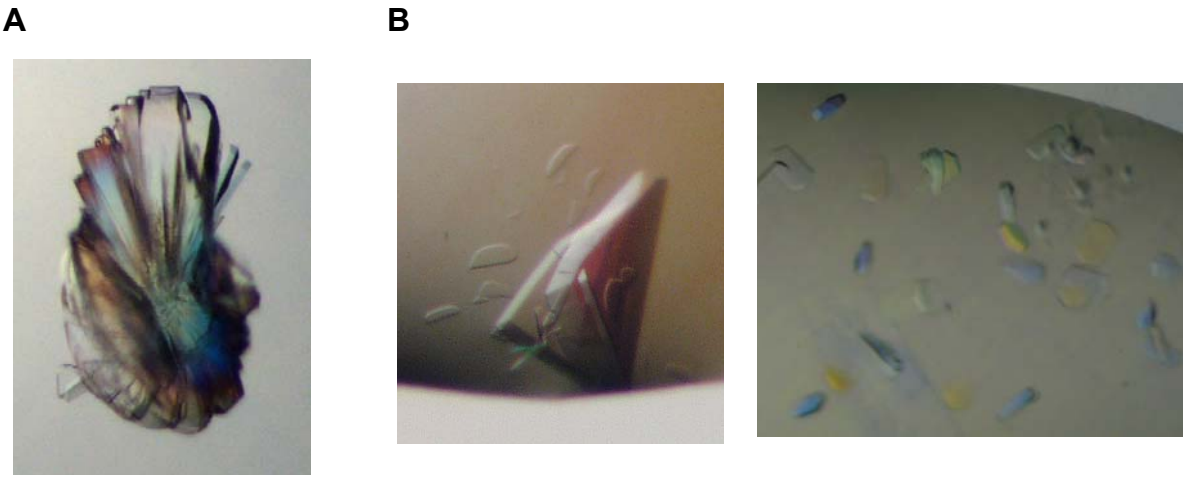


Figure 3.11 A) Crystals of the τ60/Δτ91-168 and B) τ60/Δτ91-168 elastase digested complexes.

3.1.6.2. Co-crystallization with heavy atom derivatives

Attempts to soak the τ60/Δτ91 crystals in different heavy atom solutions were unsuccessful because in all cases there was a dramatic loss in resolution. Moreover, the use of τ60/Δτ91-selenomethionine (SeMet) labeled complex (only Δτ91 was labeled with SeMet - see Materials and Methods) never yielded crystals, even after extensive screening (many different conditions were tested in terms of protein/precipitant concentration and pH). Seeding of such drops with native crystals in general was not successful. Only in one drop a small crystal grew after 10 days under the same conditions as the native crystals (Figure 3.12A). We collected a data set at 4.5 Å resolution from this crystal. Although this data set was not used for phasing, it was helpful during the model building to locate the methionine residues in Δτ91 (see below). As a next step, we co-crystallized the complex using different heavy atom derivatives (see Materials and Methods). Crystals were obtained in the presence of the following heavy atoms: 1 mM $AgNO_3$, 1 & 10 mM $K_2Pt(CN)_4$, 10 mM Gadoteridol, 10 mM K_2OsCl_6, 10 mM $KAu(CN)_2$ (Figure 3.12B,C,D,E). All these crystals were tested at ESRF beamlines. Crystals co-crystallized with $AgNO_3$ and K_2OsCl_6 showed poor diffraction to about 8 Å resolution. Crystals grown in the presence of Gadoteridol showed diffraction to 3.5 Å resolution, but due to radiation damage we only collected a data set at 4.5 Å resolution. In addition, we also collected a data set at 5.5 Å resolution from crystals grown in the presence of $K_2Pt(CN)_4$. Both the crystals grown in the presence of Gadoteridol and of $K_2Pt(CN)_4$ finally did not contain the heavy atom (see below). The best diffracting crystals which also contained the heavy-atom bound to the protein were those grown in the presence of 10 mM $KAu(CN)_2$ (Figure 3.12E). From these crystals we collected a data set at 3.5 Å resolution, from which initial phases were obtained.

A

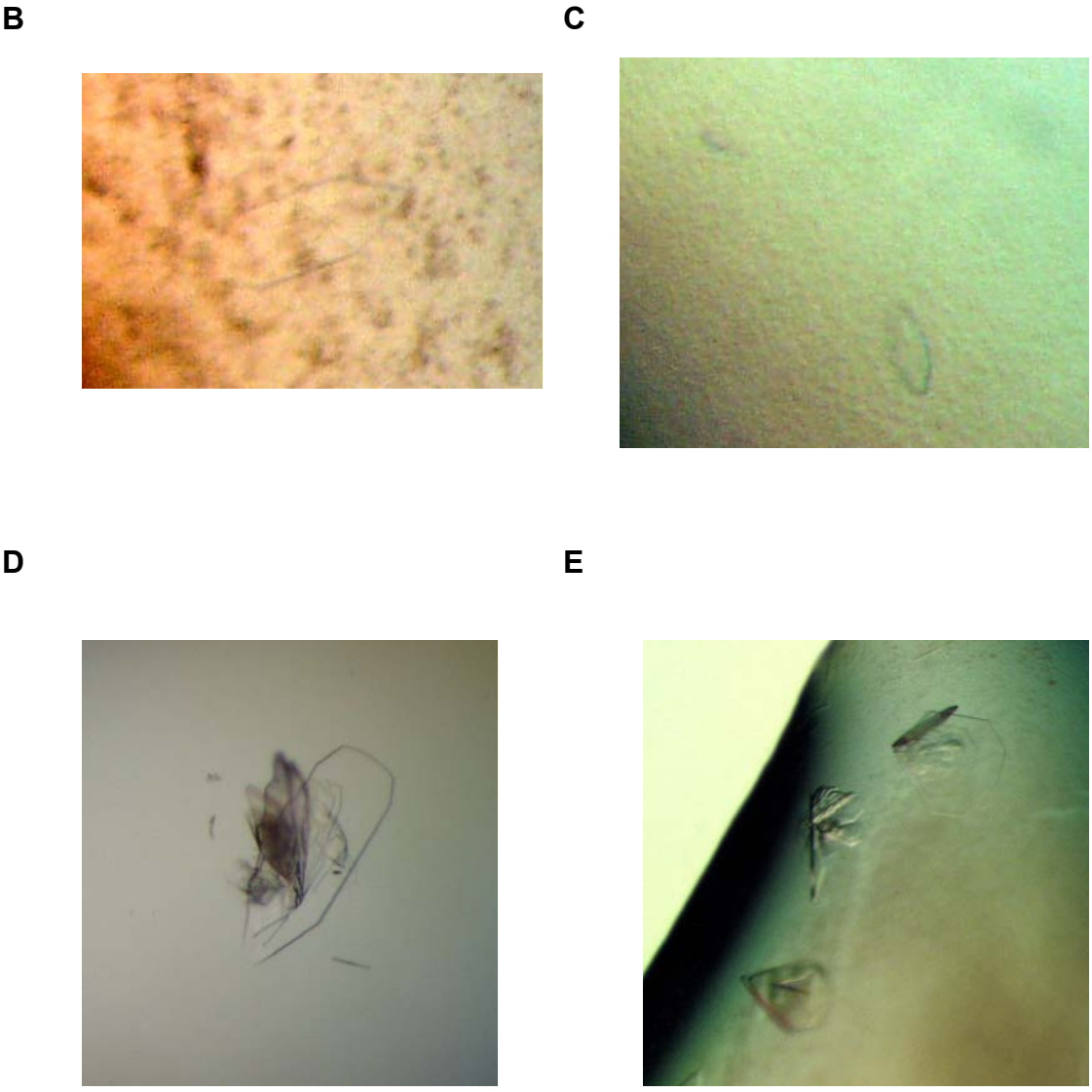


Figure 3.12 A) τ60/Δτ91-SeMet labeled Crystals. Crystals of the τ60/Δτ91 complex co-crystallized with B) 1 mM $AgNO_3$, C) 10 mM $K_2Pt(CN)_4$, D) 10 mM Gadoteridol and E) 10 mM $KAu(CN)_2$.

3.2. Data collection and structure determination

3.2.1. Cryo-crystallography

All data used for the X-ray structure determination of the τ60/Δτ91 complex were collected from crystals which were flash-frozen in a stream of nitrogen gas at 100 K using 30% glycerol in 0.8 M ammonium sulfate, 0.1 M Hepes pH 7.0. Crystals that grew in the presence of heavy atom derivatives that can covalently bind to the protein were first back-soaked in harvesting buffer (0.9 M ammonium sulfate, 0.1 M Hepes pH 7.0) and then flash-cooled in a stream of nitrogen gas at 100 K using 30% glycerol in harvesting buffer.

3.2.2. Data collection and processing

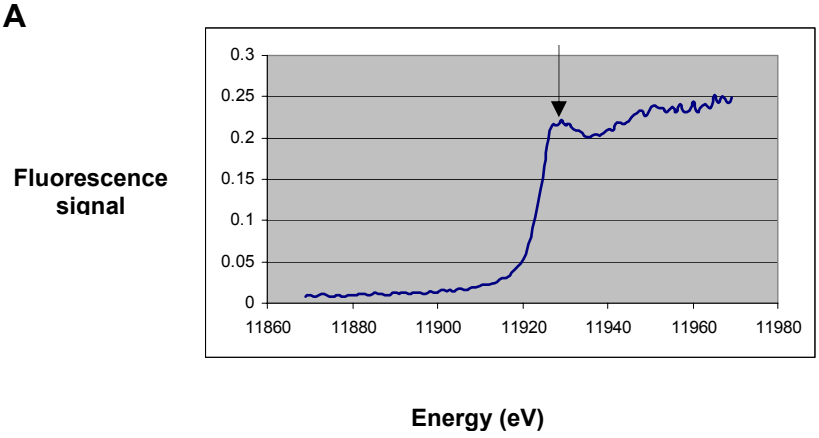
All data were collected at the ESRF beamline ID-29 at 100 K. The parameters for each data collection are summarized in Table 3.1. The ESRF software to control data collection is ProDC.

Table 3.1. Parameters used for data collection.

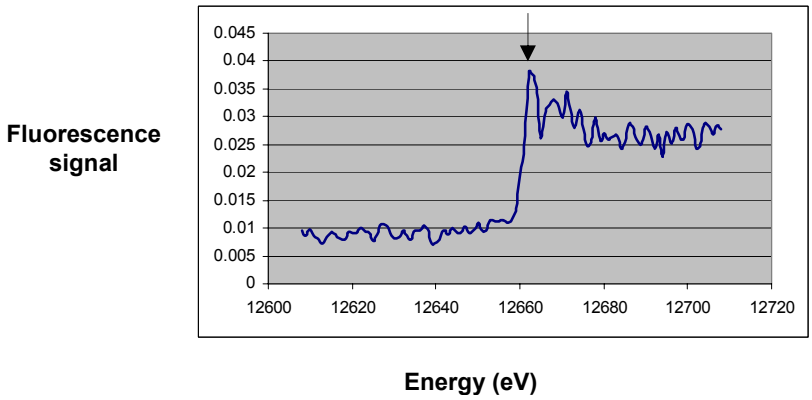
Data set	Native	Au	Semet	Gd	Pt
Beam line	ID-29	ID-29	ID-29	ID-29	ID-29
Detector	ADSC Q4	ADSC Q4	ADSC Q4	ADSC Q4	ADSC Q4
Distance (mm)	250	280	300	180	340
λ (Å)	0.9793	1.0394	0.9791	1.7118	0.9330
$\Delta\Phi_{image}$ (^{o})	0.75	1.0	1.0	1.0	0.75
$\Delta\Phi_{total}$	250	240	178	180	202
Exposure time (seconds)	2	2	5	5	5
Attenuators	-	256 (Al, 0.5 mm)	-	-	-

The best native diffraction data set was collected from a crystal grown in the presence of sulfobetaine, as already mentioned, at ESRF beamline ID29 (λ=0.9793) on an ADSC Quantum 4 CCD detector. Although these native crystals showed diffraction up to 2.8 Å resolution, we finally managed to collect a complete data set only at 3.2 Å resolution due to radiation damage. Radiation damage was in general one of the biggest problems we had to deal due to the weak diffraction of the crystals and because of the low symmetry of the crystals which required collecting between 120°-180° in order to obtain a complete data set. Therefore, the crystals had to be exposed for a long time in the X-ray beam and the resolution was usually decreasing a lot after having collected ~100°. Only in the case of the Au data collection, the crystal was resistant to radiation damage probably due to the use of attenuators which allowed us to expose for longer time without damaging the crystal. However, also in this case the radiation damage was evident after we had collected the first data at the absorption edge. This obliged us to perform a SAD and not a MAD experiment since we were not able to collect data also at the inflection and the remote point of the Au absorption edge. For all other derivatives we also could only collect a single data set at the absorption edge for the same reason. The wavelength for data collection from crystals grown in the presence of heavy atoms was chosen to be at the absorption edge of each heavy atom in order to maximize the anomalous scattering differences. The exact wavelength for each heavy atom data collection was determined based on the fluorescence scan recordered before data collection. In particular, in the case of Au prior to data collection a fluorescence scan at the absorption edge (L3) of Au was recorded, indicating the presence of the heavy atom (Figure 3.13A) and a SAD data set of the Au derivative was collected at a wavelength corresponding to the peak of the experimental Au L3-absorption edge (λ=1.0394). For the Se, again prior to the data collection a fluorescence scan at the absorption edge (K-edge) of the Se was recorded, indicating the presence of the Se (Figure 3.13B) and again a one-wavelength data set of the SeMet labelled crystal was collected at a wavelength corresponding to the peak of the experimental selenium K-absorption edge (λ=0.9791).

A



B



Heavy atom	Absorption edge	Energy (eV)	Wavelength (Å)
Au	L3	11918	1.0403
Se	K	12658	0.9795

Figure 3.13 Fluorescence scans of a A) Gold and B) SeMet labelled crystal. The fluorescence scans were taken at the absorption edge of each heavy atom (see table for theoretical values), prior to data collection, in order to confirm the presence of gold (A) and selenium (B) atoms. Arrows indicate the energies where data were recordered.

All data sets were processed with MOSFLM (Leslie, 1999) and SCALA (CCP4, 1994) (Table 3.2). After scaling the observed intensities were transformed in structure factor amplitudes using program TRUNCATE. The space group was determined as monoclinic $P2_1$ based on the statistics during data reduction and the systematic absences along the 2_1 axis. The τ60/Δτ91 complex crystallizes with 2 complexes per asymmetric unit and a solvent content of 55%.

Table 3.2. Crystallographic Statistics

$P2_1$ $a = 61.4$, $b = 25.8$, $c = 210.5$, $\beta = 94.5^o$
2 molecules/asymmetric unit

Data Collection and Phasing	Native	$KAu(CN)_2$	Semet	Gd	Pt
Wavelength	0.9793	1.0394	0.9791	1.7118	0.9330
Resolution (Å)*	35.0-3.2 (3.4-3.2)	35.0-3.5 (3.7-3.5)	40.0-4.5 (4.9-4.5)	40.0-4.5 (4.74-4.5)	30.0-5.5 (5.8-5.5)
ESRF Beamline	ID-29	ID-29	ID-29	ID-29	ID-29
Multiplicity*	3.2 (3.0)	4.2 (3.7)	3.0 (2.9)	3.5 (3.4)	3.0 (3.1)
Completeness (%)*	98.8 (97.6)	98.7 (96.6)	89.6 (80.9)	99.8 (99.7)	97 (97.4)
I/σ(I)*	5.7 (1.6)	6.2 (2.4)	4.5 (1.9)	4.1 (2.4)	3.4 (2.9)
R_{sym} (%)*	10.2 (45.5)	9.5 (29.3)	13.2 (32.7)	13.4 (23.3)	12 (24)
No of sites		12	6		
Phasing power		0.739			
R_{cullis}		0.915			
FOM (Before/After DM with AVE)		0.273/0.922			

*Values in parentheses correspond to the highest resolution shell.

3.2.3. Structure determination

For all the derivative crystals the anomalous signal (defined as the anomalous signal/noise ratio based on variances of F+ and F-) was calculated using program XPREP. The anomalous signal was calculated at resolution ranges ∞ to 8 and 6 to 8 Å. At these resolutions any value for anomalous signal below 1.5 is considered as noise. As shown in Table 3.3, the anomalous signal was high only in the case of Au indicating the presence of the heavy atom in the crystals, whereas in all the other cases the anomalous signal was very low. The isomorphous differences varied from 15-35%, even in the cases where the heavy atoms could not be detected in the crystals. Isomorphous and anomalous Patterson differences were calculated for all derivatives and the difference Patterson maps were analysed manually and automatically.

Table 3.3. Anomalous signals and isomorphous differences for crystals grown in the presence of different heavy atoms as calculated by XPREP.

	Native	**Au**	**Se**	**Gd**	**Pt**
Resolution (Å)	3.2	3.5	4.5	4.5	5.5
Unit cell(Å, °)	61.42, 125.80, 210.47, 94.49	61.11, 125.50, 212.73, 94.27	61.71, 126.00, 212.34, 95.34	61.10, 125.05, 210.80, 94.69	61.69, 126.25, 214.09, 95.88
Change in cell parameters (%)		0.5	0.6	0.4	1
Anomalous Signal[1]	∞ -8.0 / 8.0-6.0 1.09 1.17	∞ -8.0 / 8.0-6.0 1.74 1.60	∞-8.0 / 8.0- 6.0 1.27 1.20	∞-8.0 / 8.0-6.0 1.24 1.22	∞-8.0 / 8.0-6.0 1.30 1.11
Isomorphous Difference[2] (%)		35.2	20.1	14.8	32.1

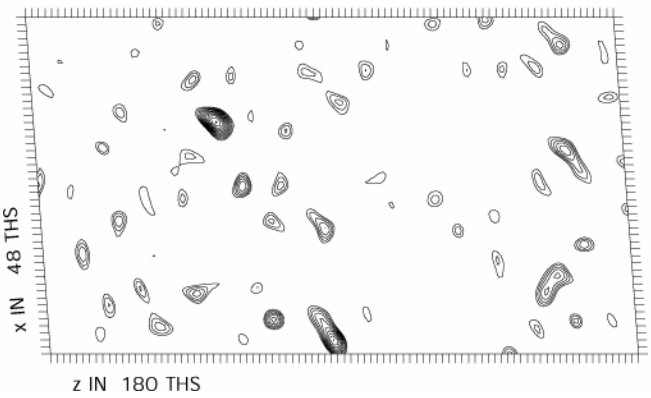
[1]defined as the anomalous signal/noise ratio based on variances of F+ and F-

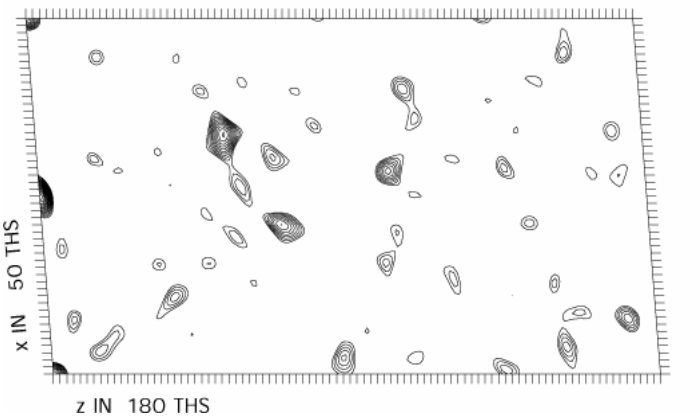
[2]defined as $R = \Sigma_{hkl} |F_{ph}(hkl) - F_p(hkl)| / \Sigma_{hklp}(hkl)$

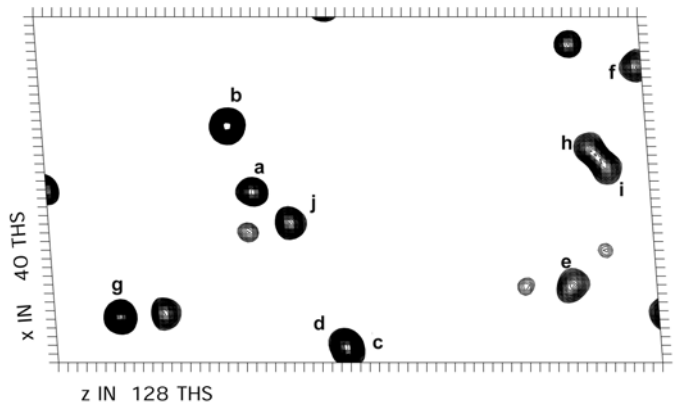
The absence of clear peaks in the case of Gd, Pt and Se indicated that either these heavy atoms did not bind at defined positions in the crystals or that the data quality was not sufficient to locate their positions. Only in the case of Au some clear peaks were observed. However, Mass-spectrometry performed on the SeMet substituted Δτ91 protein revealed that all six methionines present in the protein were substituted by Se. This indicated that very likely the seleniums were actually present in the crystals and indeed we managed later to locate the Se by cross-difference Fourier analysis (see below).

The analysis of all the above data sets revealed that the Au crystals, which diffracted best among all derivative crystals, very likely contained the heavy atom. To search for the Au atoms, the scaled Au data were used as an input in SHELXD (Schneider and Sheldrick, 2002) (at 5 Å resolution) and 10 gold sites were located in the asymmetric unit (Figure 3.14B). In addition, we performed the same site search with SHELXD but with the use of the isomorphous difference between the native and the Au data sets (in this case the input for SHELXD were the Au data merged with the native data with program CAD (CCP4, 1994) and scaled together with SCALEIT (CCP4, 1994)) (Figure 3.14B). The presence of the same 10 gold sites in both cases confirmed that the Au atoms occupied defined positions. To confirm the results of SHELXD we also calculated an anomalous Patterson map using the anomalous signal of the Au derivative and a Patterson map using the isomorphous difference between the native and the Au data sets. The theoretical Patterson map of these 10 sites was compared to the experimental ones (Figure 3.14A).

A


x IN 48 THS
z IN 180 THS


x IN 50 THS
z IN 180 THS


b
a
j
f
h
i
e
g
d
c
x IN 40 THS
z IN 128 THS

B

SHELX – anomalous

```
------------------------------------------------------------------------------
R = 0.348, Min.fun. = 0.530, <cos> = 0.417, Ra = 0.364
Try    67, CC All/Weak 30.89 / 18.46, best 31.24 / 19.35, best PATFOM 19.34
PATFOM  19.21
------------------------------------------------------------------------------

 a*  HETATM    1  S   HAT     1      13.760  18.515  18.064 1.000 20.00
 b   HETATM    2  S   HAT     2     -10.681  -9.952  16.240 0.805 20.00
 c   HETATM    3  S   HAT     3      -6.674 -25.488  80.342 0.789 20.00
 d   HETATM    4  S   HAT     4     -38.726 -23.109 131.548 0.767 20.00
 e   HETATM    5  S   HAT     5     -49.562 -27.926 166.319 0.703 20.00
 f   HETATM    6  S   HAT     6      -8.323  -1.596  52.330 0.598 20.00
 g   HETATM    7  S   HAT     7     -34.931 -41.320 115.845 0.530 20.00
 h   HETATM    8  S   HAT     8     -30.669 -25.593 164.065 0.507 20.00
 i   HETATM    9  S   HAT     9     -16.910 -17.929  49.235 0.472 20.00
 j   HETATM   10  S   HAT    10      11.530 -40.752  84.571 0.436 20.00
```

SHELX - isomorphous

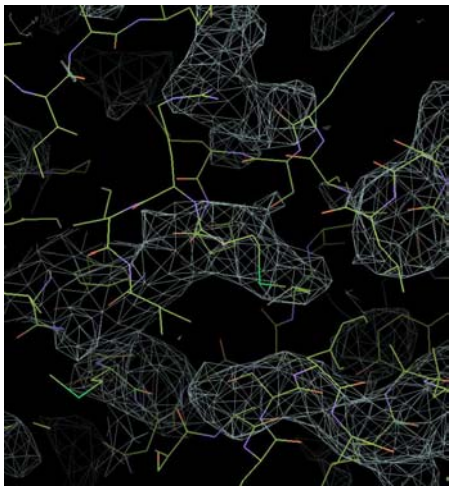
```
------------------------------------------------------------------------------
R = 0.368, Min.fun. = 0.449, <cos> = 0.444, Ra = 0.369
Try    65, CC All/Weak 30.51 / 21.46, best 31.36 / 22.42, best PATFOM 15.63
PATFOM  15.72
------------------------------------------------------------------------------

 a*  HETATM    1  S   HAT     1      13.530  61.742  18.092 1.000 20.00
 b   HETATM    2  S   HAT     2     -10.767  33.265  16.153 0.875 20.00
 c   HETATM    3  S   HAT     3      -7.476  17.930  79.802 0.815 20.00
 d   HETATM    4  S   HAT     4     -39.202  19.907 129.958 0.771 20.00
 e   HETATM    5  S   HAT     5     -50.138  15.422 164.671 0.684 20.00
 f   HETATM    6  S   HAT     6      -8.421  41.410  51.627 0.596 20.00
 g   HETATM    7  S   HAT     7     -35.363   1.807 114.571 0.576 20.00
 j   HETATM    8  S   HAT     8      11.723   2.553  83.802 0.517 20.00
 h   HETATM    9  S   HAT     9     -31.659  17.729 162.582 0.494 20.00
 i   HETATM   10  S   HAT    10     -17.573  23.733  47.837 0.410 20.00
```

*Letters a-j indicate identical sites in both cases. The arbitrary difference in the b dimension between the two choices is 43 in all cases.

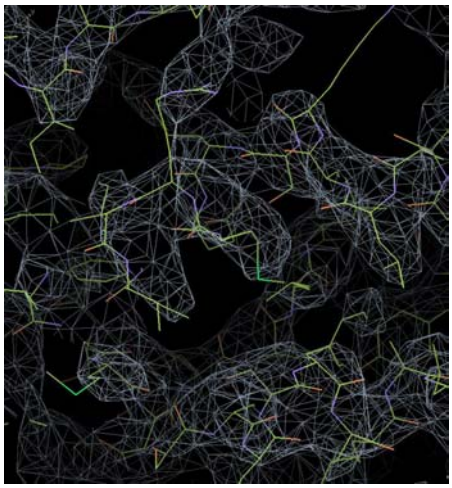
Figure 3.14 Patterson maps and positions of the gold sites predicted by SHELXD. A) Experimental anomalous Patterson map (v=1/2) using the Au data set at 5 Å (Top). Experimental Patterson map (v=1/2) calculated using the isomorphous difference between the native and the Au crystal at 5 Å (Middle). Predicted difference Patterson map (Bottom) B) The 10 Au sites predicted by SHELXD, with anomalous differences (Top). The 10 Au sites predicted by SHELX, with isomorphous difference (Bottom). Identical sites are labelled a-j.

Two more Au sites were subsequently located by performing difference Fourier synthesis. Initial phases were calculated at 3.5 Å resolution using the scaled Au data set and the 12 gold sites as input for program SHARP (de la Fortelle and Bricogne, 1997) with standard parameters for SAD phasing (AUTOSHARP). As expected the figure of merit obtained after this step was quite low (0.273) (Table 3.2) and the calculated electron density map was looking very noisy without any secondary structure features (Figure 3.15A). From Matthews coefficient calculations (Matthews, 1968) we were expecting that most likely the crystals would contain two or three complexes in the asymmetric unit with corresponding V_M values of 3.2 or 2.2 $Å^3$/Da and a solvent content of 62 or 42%, respectively. We, therefore, looked whether we could detect a non-crystallographic symmetry between the 12 gold sites that would allow us to perform NCS averaging in order to obtain better phases. By visual inspection of the 12 Au sites we could detect two clusters of heavy atoms, which were related by a dyad (Figure 3.16). As a next step, we used program Lsqkab (CCP4, 1994) and superimposed the two clusters of Au sites. The NCS matrix obtained with Lsqkab was introduced to program DM (CCP4, 1994), which performed NCS averaging and solvent flattening at 3.5 Å. After this step the figure of merit considerably improved (0.922) (Table 3.2) yielding an interpretable electron density map with recognisable secondary structure features (Figure 3.15B). Finally, phases were transferred to the nearly isomorphous native data set and extended to 3.2 Å resolution, followed by 2-fold NCS averaging.

A



B



C

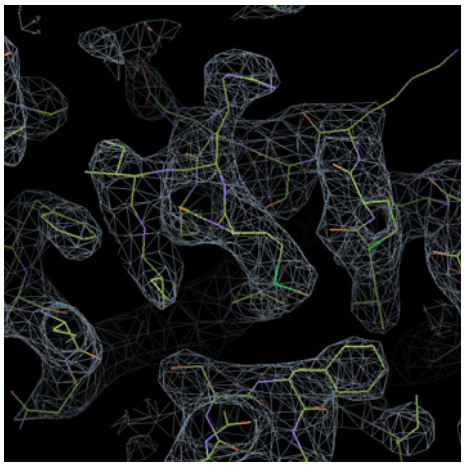


Figure 3.15 A) Electron density map after SAD phasing. The absence of secondary structure elements and connectivity is evident. B) Electron density map after NCS averaging and solvent flattening. This map reveals clear secondary structure elements. C) The final refined 2Fo-Fc map. In Figures A and B the final model does not fit very well into the electron density maps due to the differences in the cell parameters between the native and the derivative crystals.

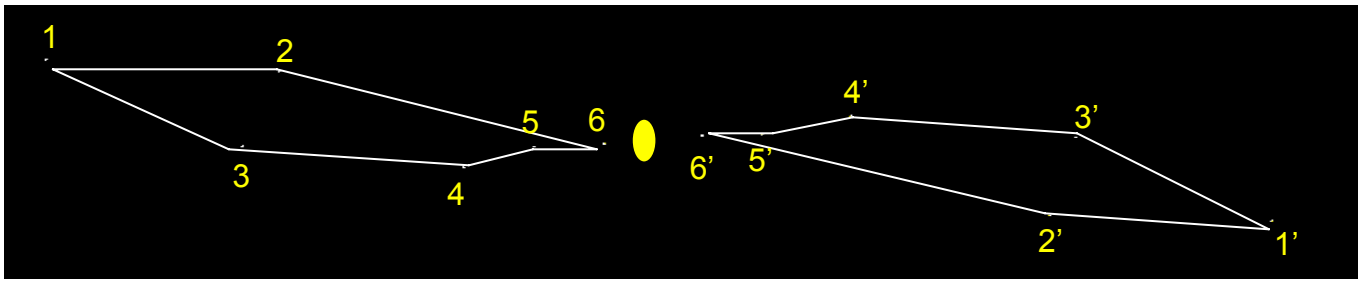


Figure 3.16 Au sites (1-6) and (1'-6') are related by a non-crystallographic dyad.

3.2.4. Model building and refinement

The initial electron density map of the τ60/Δτ91 complex was not easy to interpret due to the low resolution. First, we started building manually a polyalanine model in the parts where the electron density was clearer using program O (Jones and Kjeldgaard, 1991). The structure of the 7-bladed β-propeller of Tup1 (pdb code: 1erj) protein and the secondary structure prediction (calculated consensus secondary structure prediction from http://pbil.univ-lyon1.fr) for each protein, were used as a guide for tracing the polyalanine model. As already mentioned in section 3.1.4. WD40 repeats were predicted for τ91, whereas for τ60 none of the domain databases could detect any structural motifs. Therefore, when we obtained the first electron density map we did not know that τ60 also folds into a β-propeller. In the electron density map we could clearly see two similar densities, where we could fit the Tup1 structure. This made us wonder whether the crystals contained a dimer of Δτ91 or whether both proteins had similar fold. Once we started building a polyalanine model into the two densities corresponding to the propellers, we realized that one density, which later turned out to correspond to the τ60 propeller, was showing a higher structural similarity to Tup1. Furthermore, this density was linked to a smaller domain, which in the final model corresponds to the C-terminal domain of τ60. Therefore, we concluded that both proteins were present in the crystals and that each of them had a β-propeller fold, which was confirmed once we could assign the sequence.

For τ60 it was easier to build the polyalanine model and it was built first, which is probably due to the fact that τ60 shows a higher structural similarity with Tup1 and that its C-terminal part was mostly helical, and thus easier to be traced at lower resolution. The model for Δτ91 was very difficult to build, especially because half of the molecule was structurally unrelated with Tup1 and also because the outer part of the propeller was showing poor density probably due to the few crystal contacts observed in this region.

In general, because of the low resolution, the high content of β-sheets, and the size of the complex the sequence assignment was very difficult. Therefore, we performed cross-difference Fourier synthesis for the location of the Se atoms. From the six Semethionine residues present in Δτ91 we located only 4 residues because two were in disordered regions. We also performed cross-difference Fourier synthesis for the Gd and the Pt derivatives, just in case Pt and Gd atoms were bound to the protein despite the low anomalous signal. However, no cross-difference Fourier anomalous peaks were observed in both cases. Secondary structure prediction combined with Se cross-difference anomalous Fourier peaks and the Au sites (located close to Cys) provided useful markers for the sequence assignment (Figure 3.17). Side chains were added using program TURBO (Roussel and Cambilleau, 1989).

Iterative cycles of phase combination, model building, and refinement led to a complete model. Refinement was carried out with program CNS (Brunger at al., 1998). During the refinement, 10% of the reflections were omitted from the refinement and used to calculate the free R factor. The initial model had an R value of 51 and a free R value of 51.6. As general strategy we built a polyalanine model in the parts of the map where the density was clear enough, which was then refined by minimization. Phases from the partial model were combined with the experimental phases followed by NCS averaging and solvent flattening, but also used to calculate 2Fo-Fc and Fo-Fc maps. During building we used all these maps as well as the original experimental map in order to prevent model bias. As shown in Figure 3.18, the R factor decreased almost by 3% (and the free R factor by 2%) when the first side chains were added to the model (30 alanine residues were substituted by side chains in τ60). Until the 6^{th} cycle of refinement we were building at new parts of the protein

backbone and were mutating alanines wherever it was possible, until all the alanines were mutated to side chains. The calculation of a sharpen for the B factors (B factors were sharpened to 20) density modification map using program Rstats (CCP4, 1994) helped a lot to the sequence assignment. After the 6th refinement cycle all the alanines were mutated to side chains and from this step on we were performing minimization and individual B-factor refinement, which further improved the R and free R values. In the next cycle we no longer used the experimental phases as restrains which led to a further improvement of the R and free R factors and in the following cycle we also dropped the NCS. This also yielded a slightly better R and free R values, since the two molecules are not exactly the same (for example there is a slight movement of the C-terminal domain of τ60 at ~3^{o}). In the final four refinement steps we were performing corrections in the model (mostly in the stereochemistry), which yield an even lower free R.

Validation using program PROCHECK was performed during model building and refinement to identify errors in the model.

Currently, the structure has been refined at 3.2 Å resolution to a crystallographic R value of 24.2% and a free R value of 30.4% with reasonable stereochemistry (Table 3.4 – Figure 3.15C). The model consists of two τ60 chains (A and C), each in complex with a Δτ91 chain (B and D, respectively) (Figure 3.19B). All residues are well ordered except Δτ91 residues 320-337, 591-596 and 625-637.

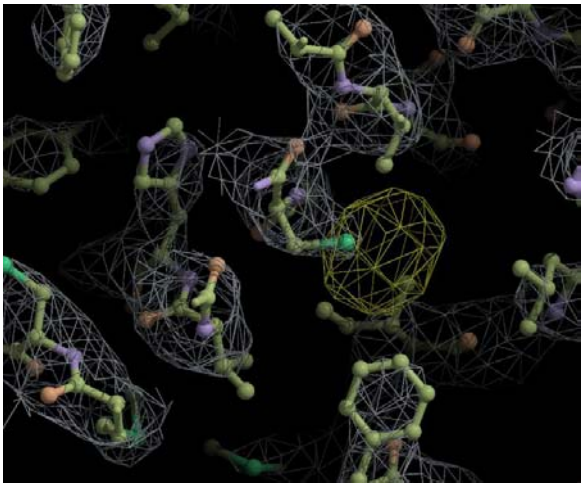
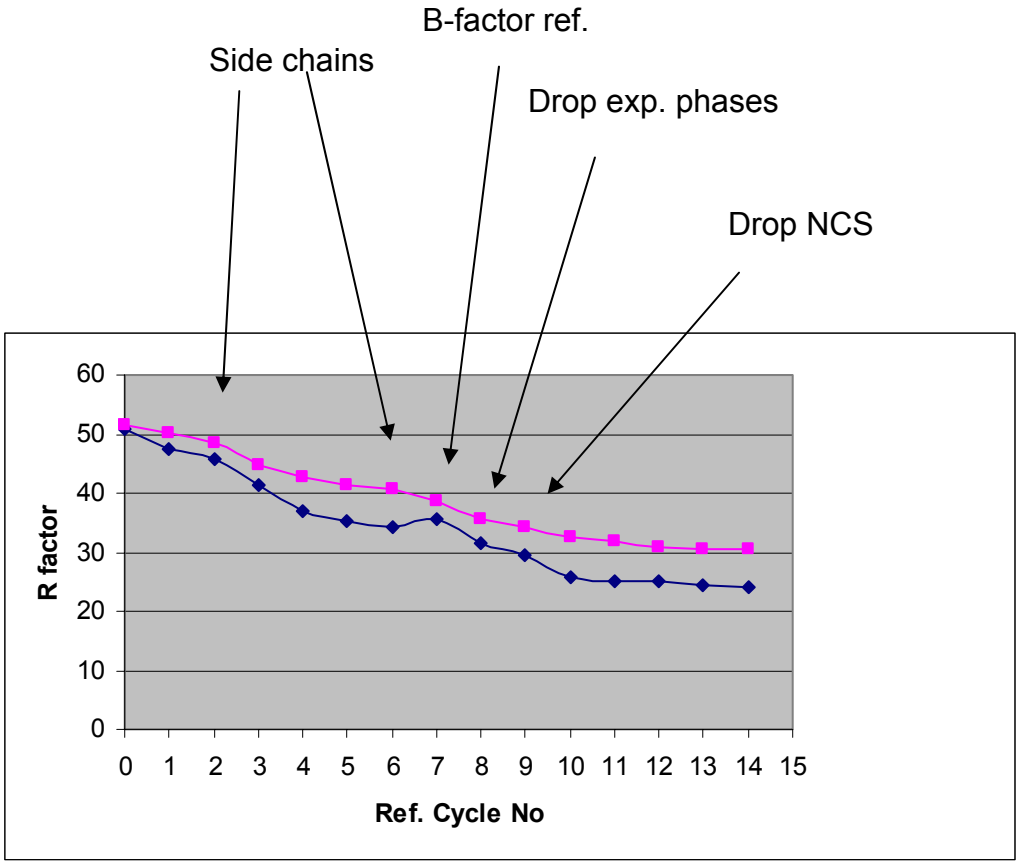


Figure 3.17 Au sites (yellow density) located close to cysteines were used as markers for sequence assignment.



Ref. Cycle No	No. of atoms
0	3129
1	3129
2	7372
3	9538
4	10718
5	12822
6	13784
7	13474
8	14998
9	15986
10	16698
11	16894
12	16942
13	16943
14	16943

Figure 3.18 Statistic analysis of the R and free R values during the different refinement steps. The R value is represented in blue and the free R value in pink. The table shows the number of atoms present in the model at each refinement cycle.

Table 3.4. Structure Refinement	
Resolution (Å)	500 - 3.2
Overall reflections (working and free set)	52 096
Number of protein atoms	16 894
Number of water molecules	5
R_{cryst} (%)	24.2
R_{free} (%)	30.4
R_{msd} bonds (Å)	0.009
R_{msd} angles (degrees)	1.700
Average B factor (Å^2)	62.1
Residues in Ramachandran plot (%)	
Most favored regions	64.6
Allowed	27.1
Generous	5.9
Disallowed	2.4

3.2.5. Crystal packing

The τ60/Δτ91 complex crystallizes with two molecules per asymmetric unit. Figure 3.19A shows the crystal packing. Along the a direction there are quite a lot of crystal contacts whereas along the b and c directions we can observe important solvent channels and only few crystal contacts. This is consistent with the anisotropy observed in the diffraction pattern, where the best diffraction is observed along the a* direction (see 3.1.6).

Although τ60/Δτ91 crystallizes with two molecules in the asymmetric unit (Figure 3.19B), there is no evidence that the contacts observed between NCS-related molecules reflect the formation of a biologically relevant dimer of the complex. In solution, both full-length τ60/τ91 as well as the truncated τ60/Δτ91 complex are heterodimeric, as determined by gel filtration.

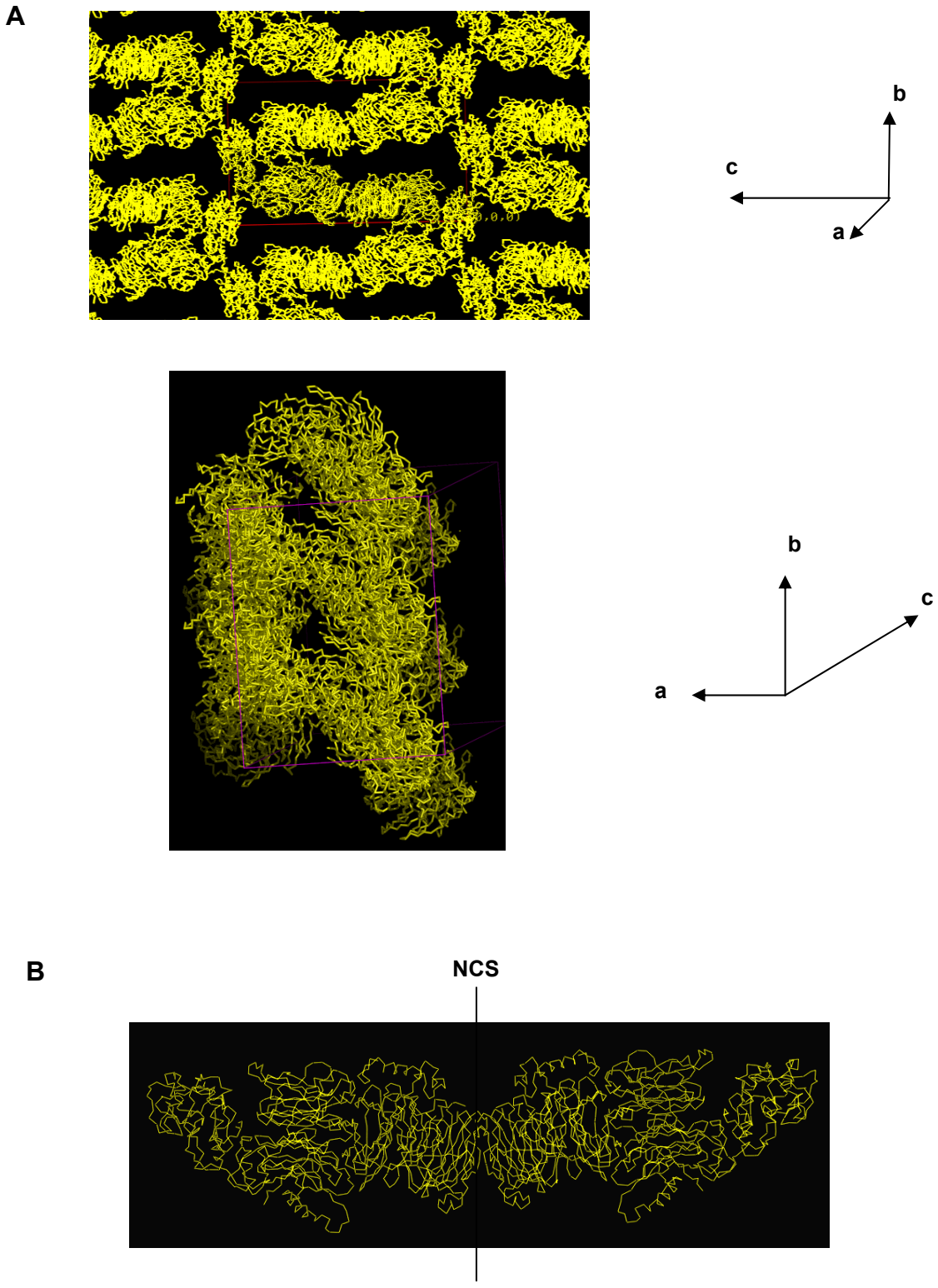


Figure 3.19 A) Crystal packing. The majority of the crystal contacts is along the a direction, whereas along the b and c directions there are few crystal contacts. B) The τ60/Δτ91 complex crystallizes with two molecules per asymmetric unit.

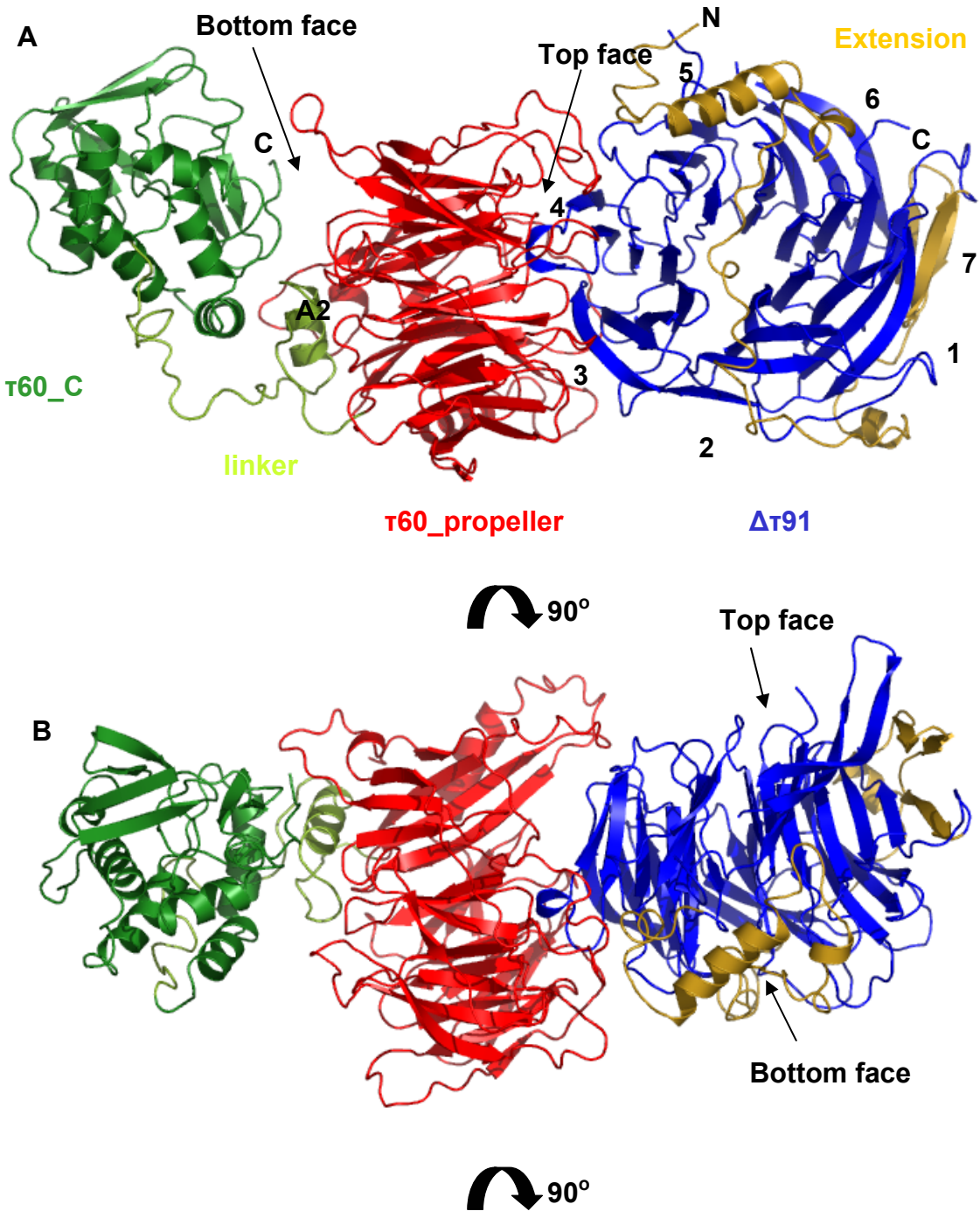
3.3. Structure of the τ60/Δτ91 complex

3.3.1. Overall Structure of the τ60/Δτ91 complex

The τ60/Δτ91 complex contains one molecule of τ60 (residues 2-588) and the C-terminal part of τ91 (residues 165-672). The structure has an elongated shape and consists of three clearly defined regions: a) Δτ91, which forms a seven-bladed β-propeller (residues 269-672) preceded by a 104 residue N-terminal moiety (residues 165-268) that packs against one face of the propeller; b) the N-terminal domain of τ60 (residues 2-372), which also forms a seven-bladed β-propeller, interacts with Δτ91 and forms the bridge between Δτ91 and the τ60 C-terminal domain; and c) the C-terminal domain of τ60 (residues 420-588) which adopts a novel fold formed by helices and β-sheets, and is located distal to Δτ91 (Figure 3.20). The two domains of τ60 are connected through a linker (residues 373-419), which contains a helix (helix A2) (residues 383-395), parallel to each domain. The two domains are independent from each other, and it is the linker, which is sandwiched between the two domains, that establishes most of the interactions with each of them. The τ60 C-terminal domain does also not interact with Δτ91. The interaction between τ60 and Δτ91 is mediated exclusively by the two β-propellers and is a 'face to edge' interaction. Two long loops of τ60 embrace the Δτ91 propeller, which is in a perpendicular orientation relative to that of the τ60 propeller (Figure 3.20A,B,C).

The long N-terminal extension of Δτ91 crosses along the bottom (see 3.3.2 for definition) surface of the Δτ91 propeller, leaving the central channel of the propeller closed from this side. N-terminal extensions have been observed in other β-propellers (Sprargue et al., 2000; Weirich et al., 2004), but none has previously been described which crosses one side of the propeller, spanning the complete diameter of the domain. This extension is not an independent domain but rather a subdomain, which interacts intimately with the propeller. It starts at the N-terminus with a helix (residues 178-193), which packs orthogonally against blades 5 and 6, leads to an additional helix (residues 240-246), at the opposite edge of the domain**;** and finally adds additional strands to blades 6 and 7 of the propeller. It interacts with the propeller

mainly though hydrophobic contacts, which are quite conserved among different species, as well as through some polar interactions. Hence, it seems rather unlikely that this moiety changes conformation or becomes displaced upon the putative binding of a partner to this side of the propeller (Figure 3.20A). In addition, this extension appears important for Δτ91 protein expression and solubility, since an N-terminal deletion which removes part of the first helix severely reduces the level of soluble expression (see 3.1.6.1).



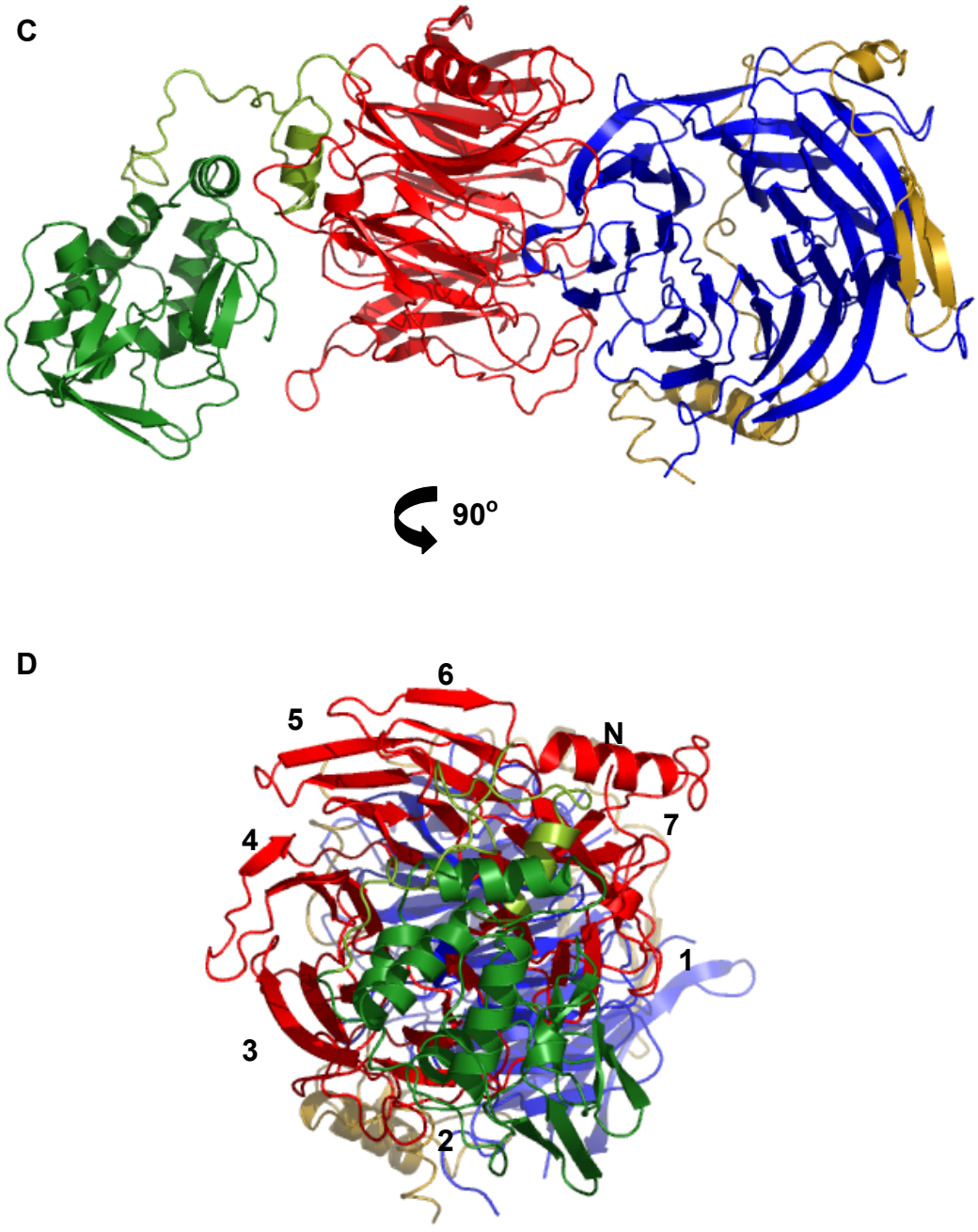
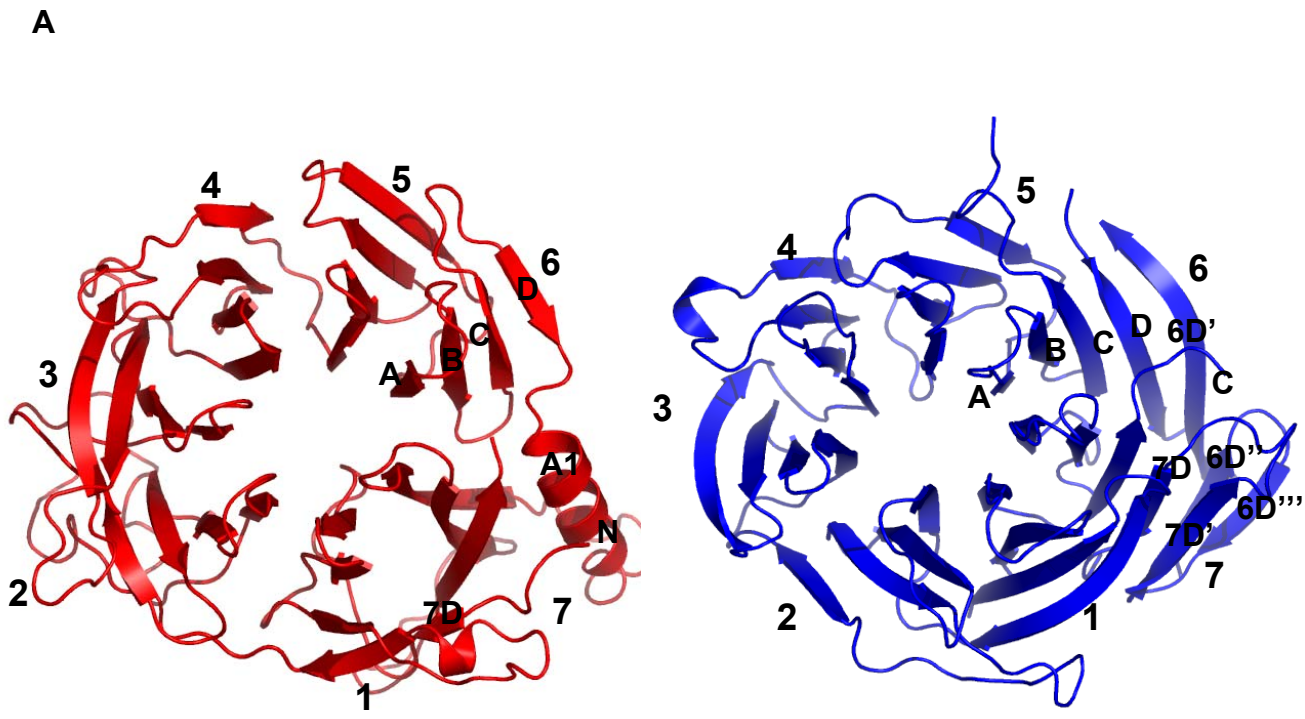


Figure 3.20 Overall structure of the τ60/Δτ91 complex. (A, B, C and D) Different views of the binary τ60/Δτ91 complex. In Figure A Δτ91 blades and in Figure D τ60 blades are labeled from 1 to 7, respectively.

3.3.2. Comparison between the two β-propellers

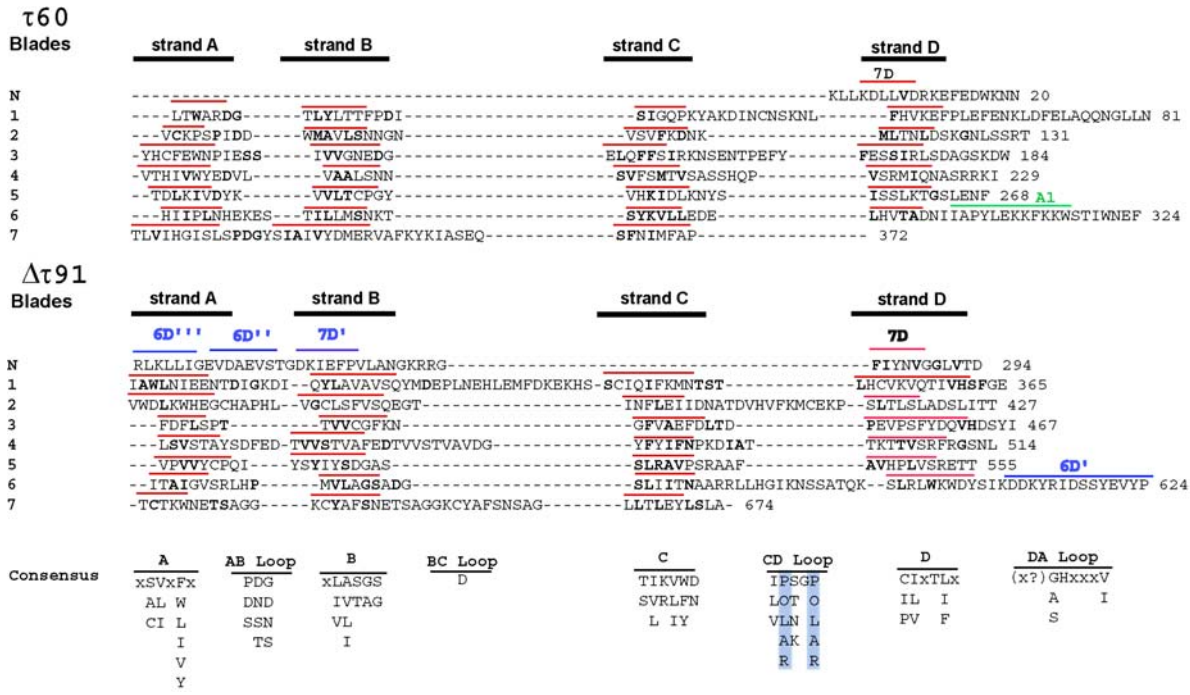
The N-terminal domain of τ60 (residues 2-373) and Δτ91 both form a WD40 seven-bladed β-propeller with dimensions of 44 Å x 51 Å x 46 Å and 45 Å x 50 Å x 40 Å, respectively (Figure 3.21A). Each blade for both proteins consists of four antiparallel β strands and is arranged about a central axis. We have labelled these strands A-D following standard convention (Weirich et al., 2004). The 'top' surface of the propeller is defined as the face that contains the DA loops connecting consecutive blades whereas the 'bottom' face is the opposite surface (Figure 3.20A,B). The fourth β strand of blade 7, in both cases, is contributed by the N-terminus of the protein, forming a "velcro closure" that is typical for β-propellers. Neither Δτ91 nor τ60 display high internal sequence conservation among the individual blades, but they contain amino acids commonly found at certain positions in WD40 repeat sequences, as defined by Smith et al. (Smith et al., 1999) (Figure 3.21B). For τ60 and Δτ91 the individual blades superimpose well on each other (Figure 3.21C). A database search of DALI (Holm and Sander, 1995) showed that both proteins are structurally similar to other β-propellers. Most specifically, Δτ91 shows the highest structural similarity with the actin interacting protein 1 (aip1), which has 14 WD40 repeats forming two 7-bladed β-propellers, with a DALI Z score of 25.7, and τ60 with the actin-like protein 3 (arp3) which is also a WD40 seven-bladed β-propeller, with a Z score of 22.5. Superposition of Δτ91 and τ60 propellers shows that they are also structurally very similar. The two proteins superimpose with an r.s.m.d. of 3.1 Å over 194 residues. However, Δτ91 has a more 'squeezed' shape, whereas τ60 is more round, and this is due to blade 6 of Δτ91 which shows the lowest structural similarity with its corresponding blade from τ60 (Figure 3.21D). Another important difference between the two propellers is the additional strands in blades 6 and 7 of Δτ91. Blade 6 is extended by two antiparallel beta strands (strand 6D'' and 6D''') (residues 253-258 and 261-266), which come from the N-terminal extension. In addition, this extension forms one short strand that extends blade 7 (strand 7D'). C-terminal blades extended by strands coming from N-terminal extensions have been seen before (Sprargue et al., 2000; Weirich et al., 2004), but not from internal insertions. Such a case is observed in blade 6 of Δτ91, where an extra strand (strand 6D') is formed by residues 611-624.

It is likely that this extra strand is responsible for the fact that blades 6 from the two proteins show the lowest structural similarity. In τ60 this additional strand is absent but instead there is a helix (helix A1) (residues 306-318), in the DA loop connecting blades 6 with 7. τ60 in general has more extended insertions than Δτ91, and especially the loops connecting strands C with D and the DA loops, which embrace Δτ91 from both the top and the bottom side (Figure 3.20A,B). These insertions also appear to be well-structured oppositely to Δτ91, where the largest loops are very flexible and therefore disordered in the crystal structure.



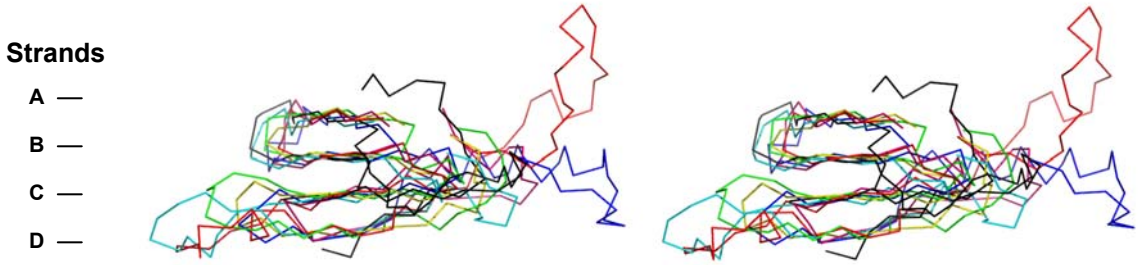
A) τ60 (red) and Δτ91 (blue) WD40 seven-bladed β-propellers. In both cases the blades are labelled 1-7 and the strands A-D. Additional secondary elements are also labelled in each case.

B



B) Structure-based sequence alignment of the WD40 repeats in τ60 and Δτ91. Amino acids commonly found at each position of the WD40 repeat sequence, as defined by Smith et al. 1999 are shown underneath. Conserved residues as defined by the 'consensus' are indicated in bold. Strands are indicated in red lines apart from the extra strands that are depicted in blue. Due to space considerations residues 82-89 and 325-329 from τ60 and 625-640 from Δτ91 are not listed.

C

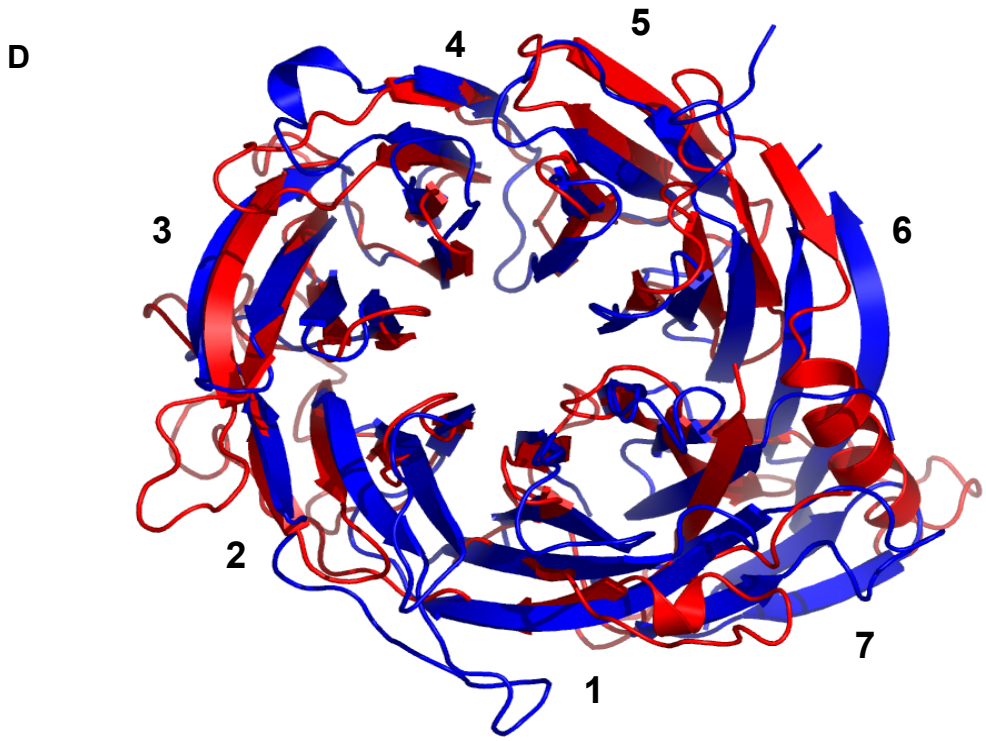


τ60

Δτ91

Blade 1 Blade 2 Blade 3 Blade 4 Blade 5 Blade 6 Blade 7

C) Superposition of the individual blades of each propeller (stereo Cα trace).



D) Superposition of τ60 and Δτ91 β-propellers. The main difference between the two propellers is in the loops, whereas the strands of the blades superimpose well, apart from blades 6.

Figure 3.21 The τ60 and Δτ91 WD40 β-propellers.

3.3.3. The C-terminal domain of τ60: a new fold

The C-terminal domain of τ60 (residues 420-588) is a globular region of dimensions 44 Å x 37 Å x 27 Å that shows no interaction with neither of the two β-propellers (Figure 3.20). It appears to be an overall new fold since no significant similarity between this domain and any known structure was detected when we queried the Protein Data Bank using programs DALI (Holm and Sander, 1995) and Combinatorial Extension (CE) (Shindyalov and Bourne, 1998). It can be subdivided into two parts: a five-helix bundle part and a second, mostly β-sheet region. Helix α1 runs antiparallel to helix α3. Helix α4 packs next to helix α3 and is oriented at 45^{o} to the other two helices (α1 and α3), and therefore establishes very few interactions only with helix α3. Helix α2 is located at the bottom left of the domain and it establishes interactions mostly with the linker connecting the two domains of τ60 and very few interactions with helix α3 and with the very C-terminal end of the domain. Helix α5 is formed by the very C-terminal part of the primary sequence and packs with one side against helix α1 and with another side against helices α3 and α4. It interacts with all three of these helices, and 'closes' the overall structure of the τ60 C-terminal domain (Figure 13.22A). Helix α5 residues F566, L570 and residues F574, V579 and Y580 close to this helix, F423 and L427 of helix α1 and I451, F454 and L458 of helix α3 are part of a hydrophobic core, which tightly connects the N- and C-terminal parts of the domain (Figure 3.22B). The second part of the C-terminal domain contains three pairs of antiparallel β strands (β1 and β2, β3 and β4, β5 and β6). Strands β1 and β2 are oriented horizontally to the rest of the C-terminal domain, forming a platform-like configuration on the top of the domain. Strand β1 interacts with helix α3 and strand β5 interacts with the very C-terminal extension, so that the two parts are rather associated into a single overall domain.

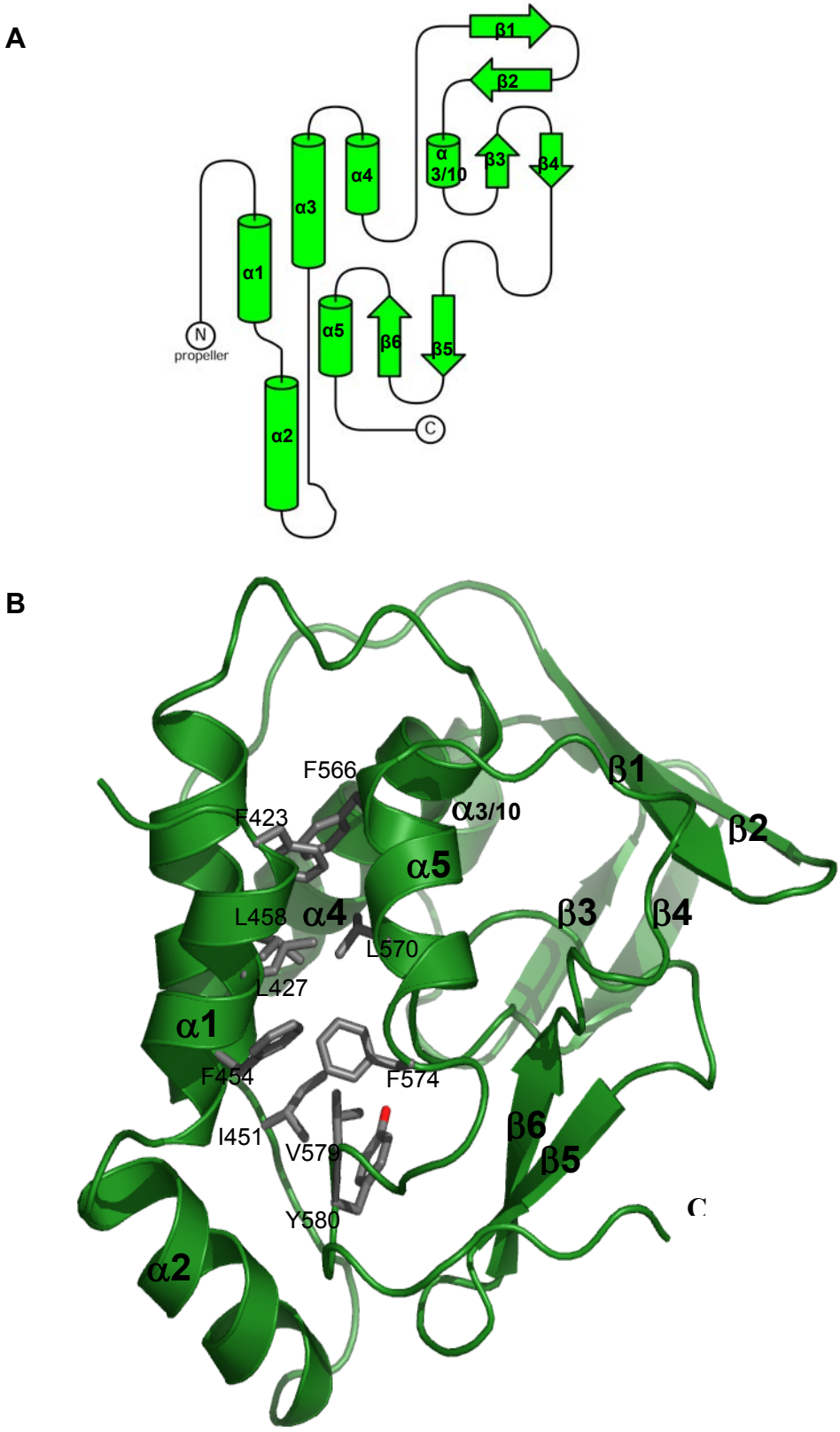


Figure 3.22 The τ60 C-terminal domain. (A) Topology diagram. Helices (green cylinders) are labelled α1-α5 and strands (green arrows) are labelled β1-β6 (for secondary structure sequence assignment see also Figure 4.1 in Discussion and Perspectives). (B) Cartoon diagram showing the side chains of helices α1, α3 and α5 that partake in the hydrophobic core.

3.3.4. τ60/Δτ91 interaction: A β-propeller – β-propeller interaction

The top face of the propeller and in particular the region around the main channel of τ60 is the anchoring surface for the CD loop in blade 4 from Δτ91, which contains a 3/10 helix. τ60 embraces Δτ91 via two long loops – one connecting blades 1 and 2 and the other connecting strands 7B and 7C, both of which are well-structured. This interface involves the side of the Δτ91 propeller covering blades 3-5. The total interaction surface buried upon complex formation is 3640 $Å^2$ and is reinforced mainly by hydrogen bonds, with only two pairs of interacting hydrophobic residues between the two molecules (Table 3.5) (Figure 3.23A). The main interaction interface involves mostly residues located in the loops of the blades. In particular, residues in the DA loops following blades 4 and 5 from τ60 establish interactions with residues in strand D of blade 3 and in the CD loop of blade 4. In addition, there are also interactions observed between residues from the DA loop following blade 1 of τ60 and residues from strand D and the loops connecting strands of blade 4 as well as from the CD loop of blade 5 of Δτ91. Moreover, more interactions are observed between residues in the BC and DA loops of blade 7 of τ60 and residues in strands D of blades 3 and 4 of Δτ91. According to the calculated electrostatic potential of Δτ91, the area covered by blades 3 and 4 of Δτ91 is acidic and the area around blade 5 is basic. The corresponding interacting interface of τ60 is basic in the region covered by blades 4 and 5 and acidic in the region of blade 1 (Figure 3.23B). So electrostatic complementary appears to be important in the recognition and docking of one protein to the other.

Table 3.5. Interactions between τ60 and Δτ91

τ60		Δτ91	
Blades	**Hydrogen bonds**		**Blades**
1 (DA loop)	A74 (O)	R541 (NH2)	5 (CD loop)
1 (DA loop)	Q76 (O)	D479 (N)	4 (AB loop)
1 (DA loop)	S82 (OG)	A543 (N)	5 (CD loop)
1 (DA loop)	R89 (NH1)	K503 (O)	4 (strand D)
1 (DA loop)	R89 (NH2)	K503 (O)	4 (strand D)
1 (DA loop)	R89 (NH2)	T501 (O)	4 (CD loop)
1 (DA loop)	R89 (NE)	T501 (O)	4 (CD loop)
3 (DA loop)	K182 (O)	K497 (NZ)	4 (CD loop)
4 (DA loop)	R226 (NH1)	V455 (O)	3 (strand D)
4 (DA loop)	K228 (NZ)	K497 (O)	4 (CD loop)
4 (DA loop)	R226 (O)	S457 (O)	3 (strand D)
4 (DA loop)	K228 (N)	S457 (O)	3 (strand D)
5 (DA loop)	N267 (ND2)	D460 (OD1)	3 (strand D)
7 (BC loop)	Y358 (OH)	F509 (O)	4 (strand D)
7 (BC loop)	Y358 (OH)	R510 (O)	4 (DA loop)
7 (DA loop)	E15 (O)	K503 (NZ)	4 (strand D)
7 (DA loop)	W17 (NE1)	Q461 (OE1)	3 (strand D)
Hydrophobic interactions			
5 (DA loop)	L265	L420	3 (strand A)
5 (DA loop)	L265	P456	3 (strand D)

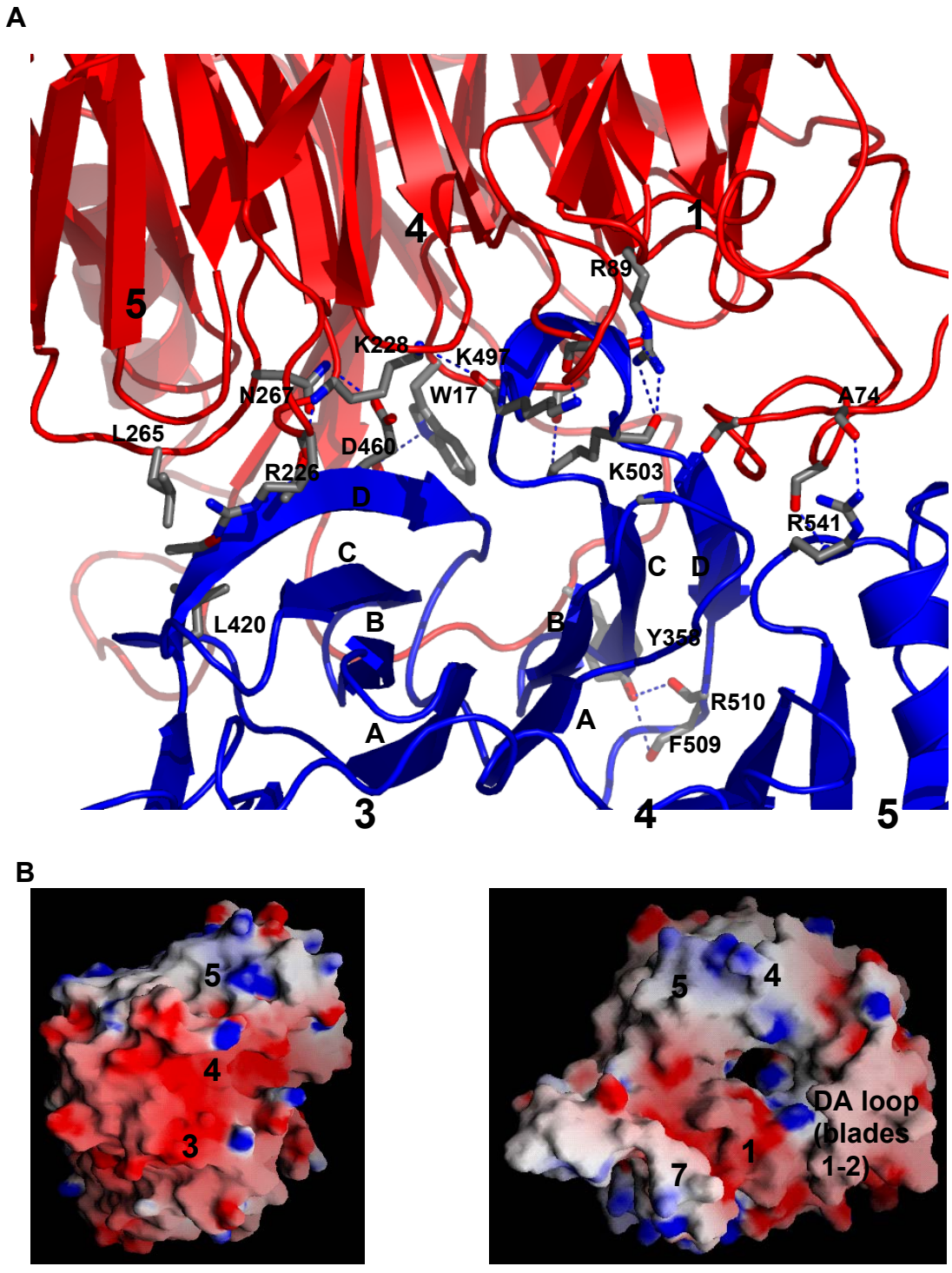


Figure 3.23 (A) Cartoon diagram of the interactions between τ60 and Δτ91. Hydrogen bonds are indicated in dashed lines. The blades of each protein involved in the interaction are indicated for both proteins and the strands for Δτ91. (B) Electrostatic representation of the interface between Δτ91 (left) and τ60 (right). Blue and red coloring represent the positive and negative electrostatic potential. The blades of each protein involved in the interaction are indicated for both proteins. The figures were generated with GRASP (Nicholls et al., 1991).

3.3.5. DNA binding activity of τ91, τ60 and Δτ91

τ138 is considered to be the DNA binding subunit containing two HMG (High Mobility Group) domains (Lefebvre et al., 1992). τ91 is thought to collaborate with τ138 in DNA binding (Arrebola et al., 1998). The N-terminal part of τ91 contains an AT-hook (DNA-binding) motif conserved in *S. pombe* and human (Huang et al., 2000). τ91 also contains a cluster of 13 cysteine and histidine residues (residues 328-411) which was suggested to bind zinc ions and to form a DNA-binding domain (Arrebola et al., 1998). Although residues 320-337 are disordered, nothing in the structure indicates that τ91 contains a Zn-binding motif. The reconstitution of a functional τB subcomplex able to contact DNA requires the presence of all three subunits, τ138, τ91, and τ60. Reconstitution of a partial τB subcomplex containing only two subunits leads to the loss of DNA binding affinity (Acker J., personal communication).

In order to better understand the role and the importance of the β-propeller – β-propeller interaction in the formation of a functional τB complex able to bind DNA, we tested whether full-length τ91 and τ60 as well as truncated Δτ91 bind DNA by southern-western blot experiments. We assayed for specific and non-specific DNA binding using probes derived from yeast tRNAGlu3 and *E. coli* LacZ genes respectively, in both their double- and single-stranded forms. τ60 showed no DNA binding affinity and this is in accordance with cross-linking experiments, where τ60 is the only TFIIIC subunit that cannot be cross-linked to DNA (Deprez et al., 1999). τ91 showed non-specific DNA-binding affinity for both single and double stranded DNA. In contrast, Δτ91 showed no DNA binding affinity (Figure 3.24). These results indicate that it is the N-terminal extension of τ91 which non-specifically contacts DNA presumably through the AT-hook domain.

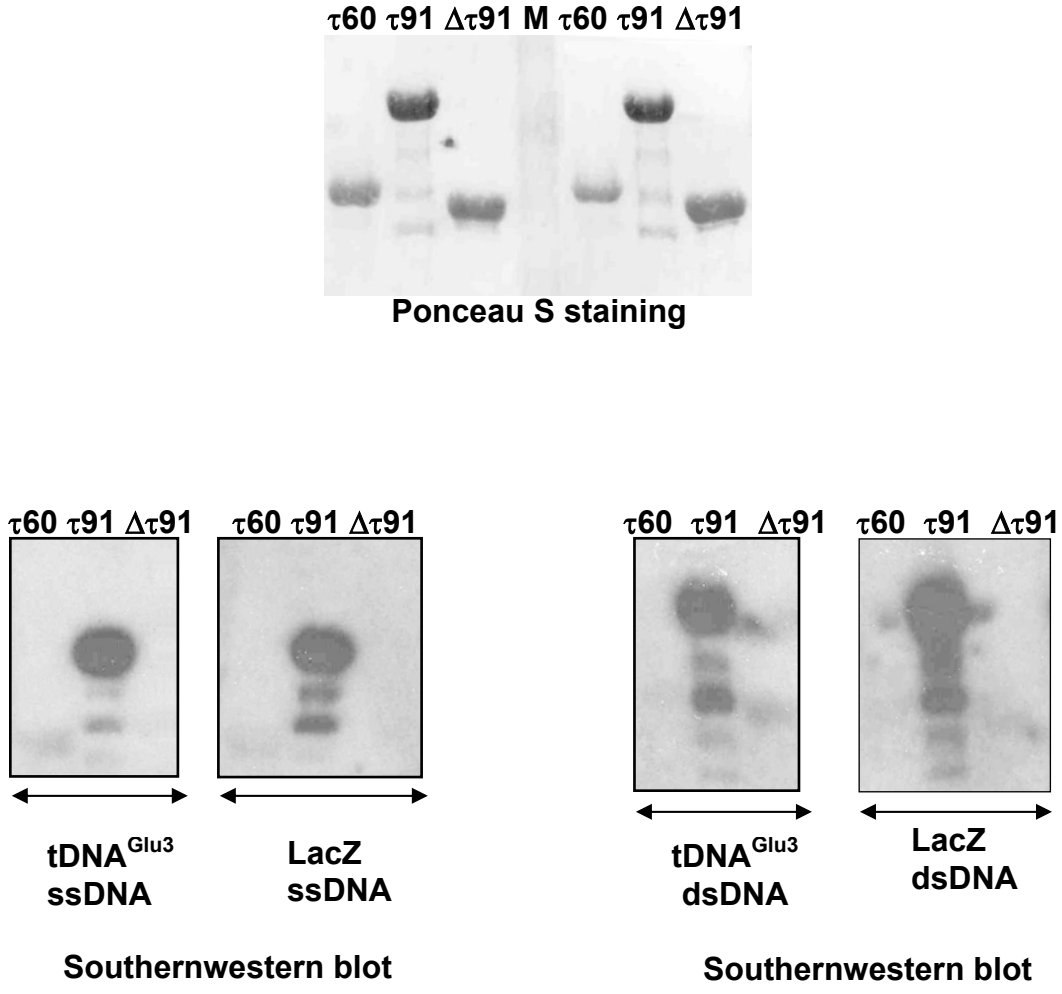


Figure 3.24 The Nterminal part of τ91 is important for its DNA binding activity. Purified recombinant τ91, Δτ91 and τ60,were analyzed on a 8% SDS PAGE, transferred onto nitrocellulose and then incubated with a specific (tRNAGlu3 gene) or non specific (LacZ fragment) single-stranded (ss) DNA and double-stranded (ds) DNA probe as indicated. Bound DNA was revealed by autoradiography.

Part IV
Discussion and Perspectives

Résumé

Dans ce chapitre l'importance de la structure du τ60/Δτ91 complexe est discutée. Les β-propeller – β-propeller interactions entre les deux protéines apparaissent très importantes pour l'assemblage d'un complexe τB très stable pour fixer à l'ADN et pour aider à démasquer les régions des protéines responsables de la liaison à l'ADN. Il n'est pas sur que la partie C-terminal de τ60 participe à l'assemblage de τB. A cause de son orientation vers le point d'initiation, c'est peut-être la partie de τ60 qui interagit avec la protéine TBP. En conclusion, notre structure permet de mieux connaître le mécanisme de la formation du complexe de pré-initiation d'ARN polymérase III. Cependant, d'autres études fonctionnelles et structurales sont encore nécessaires pour comprendre définitivement ce mécanisme.

4.1. The overall architecture of the complex is conserved among different organisms

Both yeast τ60 and τ91 show minimal sequence similarity to their human homologs. However, several lines of evidence indicate that also in other species the C-terminal part of τ91 and the N-terminal part of τ60 form β-propellers. Firstly, it has been reported that *S. cerevisiae* τ91 and its *S. pombe* Sfc6p and human TFIIIC110 homologs contain WD40 repeats in their C-terminal domains (Huang et al., 2000; Sinn et al., 1995). Essentially all proteins containing WD40 repeats adopt a β-propeller fold (reviewed in Smith et al., 1999). Secondly, an alignment of τ91 with homologous sequences from other fungi and human with MultiAlin (Corpet, 1988) revealed that the homology is very low between the N-terminal part of all proteins, whereas the C-terminal part of τ91 that corresponds to the propeller aligns well with the C-terminal part of the homologous proteins. Moreover, 80% of invariant or highly conserved residues are predicted to be within a β strand or short turn, and only 19% are predicted to be within loops. In addition, the majority of sequence deletions or additions fall into predicted loop structures (Figure 4.1A).

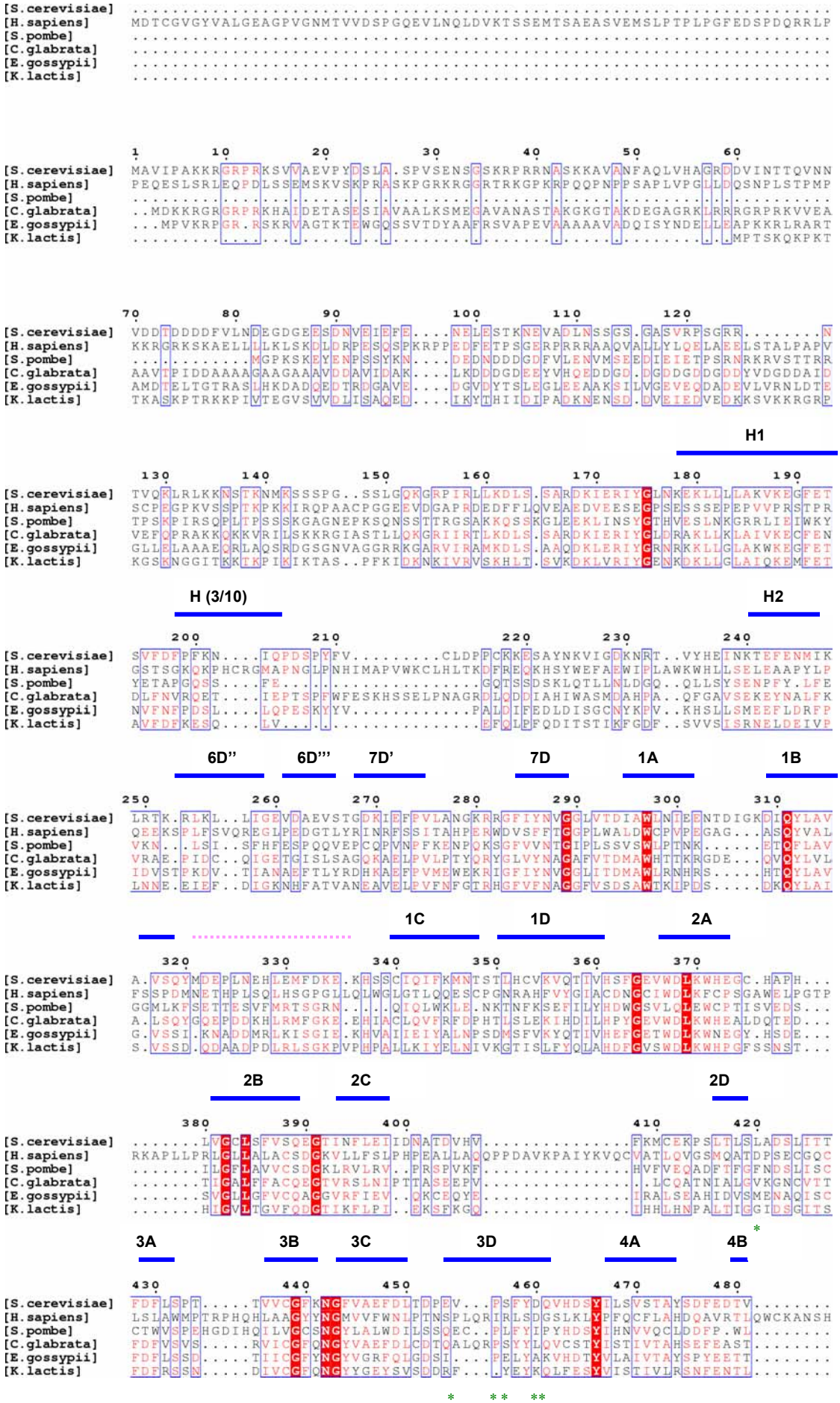
For τ60, the homology with humans is even weaker and this could be in part due to the fact that the sequences of higher organisms are much longer. This could mean that τ60 homologs in higher organisms are more complicated proteins with additional domains and functions. For example, human TFIIIC90 has the added property of an intrinsic HAT (Histone Acetyl-Transferase) activity (Hsieh et al., 1999a), which is absent from yeast τ60. In the case of τ91, not until the *S. pombe* homolog Sfc6p was cloned did the sequence similarity between yeast τ91 and human TFIIIC110 become apparent (Huang et al., 2000). Likewise, in the case of τ138, identification of the *A. thaliana* homolog revealed the homology between yeast and vertebrate τ138 subunits (Matsutani, 2004). Thus, it is possible that in the case of τ60 a 'sequence-linker' from an organism is missing in order to firmly establish the homology between yeast τ60 and its vertebrate counterparts.

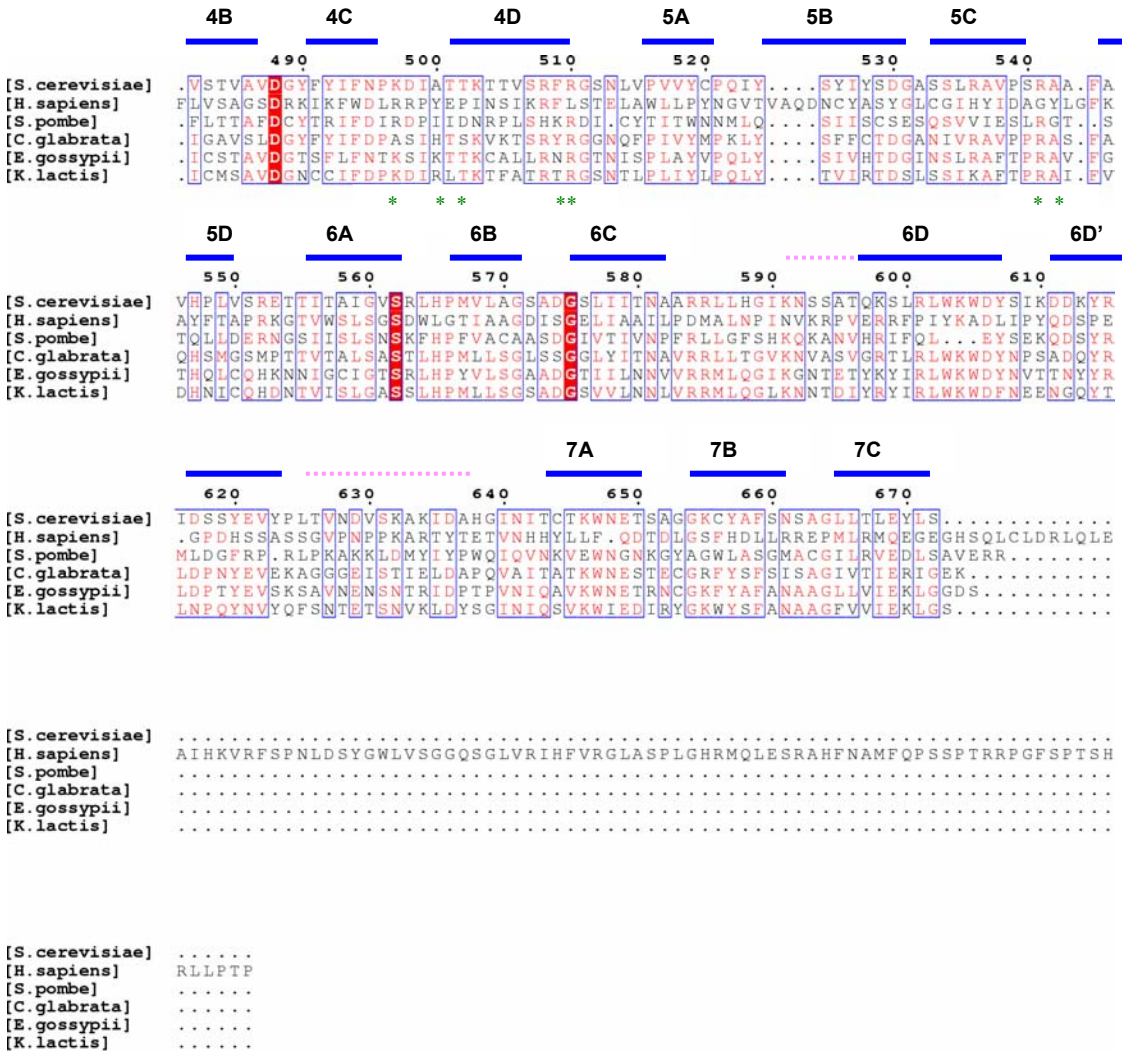
However, human TFIIIC90 is thought to be the functional homolog of yeast τ60, since both appear to bridge the τA and τB subcomplexes as well as to be involved in TFIIIB recruitment (Hsieh et al., 1999a; Deprez et al., 1999). Alignment of τ60 with homologous sequences from other organisms with MultiAlin (Corpet, 1988) shows that the C-terminal part is more conserved (Figure 4.1B). This conservation could imply a similar fold and/or function for this protein among different fungi and higher organisms. The N-terminal part that corresponds to the β-propeller shows low sequence similarity with higher eukaryotes (Figure 4.1B). However, this does not exclude that in humans this domain also adopts a β-propeller: Proteins with dissimilar sequences and with weak sequence repetition among the different β sheets can form a β-propeller (reviewed in Jawad et al. 2002; Andrade et al., 2001). Furthermore, when we submitted the N-terminal part of the sequence of the human homolog of τ60 to the SMART web server (Schultz et al., 1998), WD40 repeats are predicted. By submitting the same sequence to the 3D-pssm server for protein fold recognition (http://www.sbg.bio.ic.ac.uk/~3dpssm/), this part was predicted with a possibility of 80% to fold into a β-propeller.

The residues involved in the interactions between τ60 and Δτ91 are not very highly conserved among different fungi or higher organisms (see appendix & Figure 4.1 A,B). The most conserved residues are L265 and Y358 from τ60 and P456 and R510 from Δτ91. Interestingly, many of the τ60/Δτ91 contacts are formed through backbone contacts which are less affected by changes in the primary sequence. However the overall interacting regions are well conserved (Figure 4.1A,B). This could be again due to the fact that, as already mentioned above, proteins with very dissimilar sequences can fold into a β-propeller. This means that while the overall architecture and interacting regions of the homologous proteins can be the same or very similar, the residues involved in the interaction might not be very conserved. Moreover, the majority of the interacting residues are located in the loops and these are the parts that show the least sequence homology (Figure 4.1A,B).

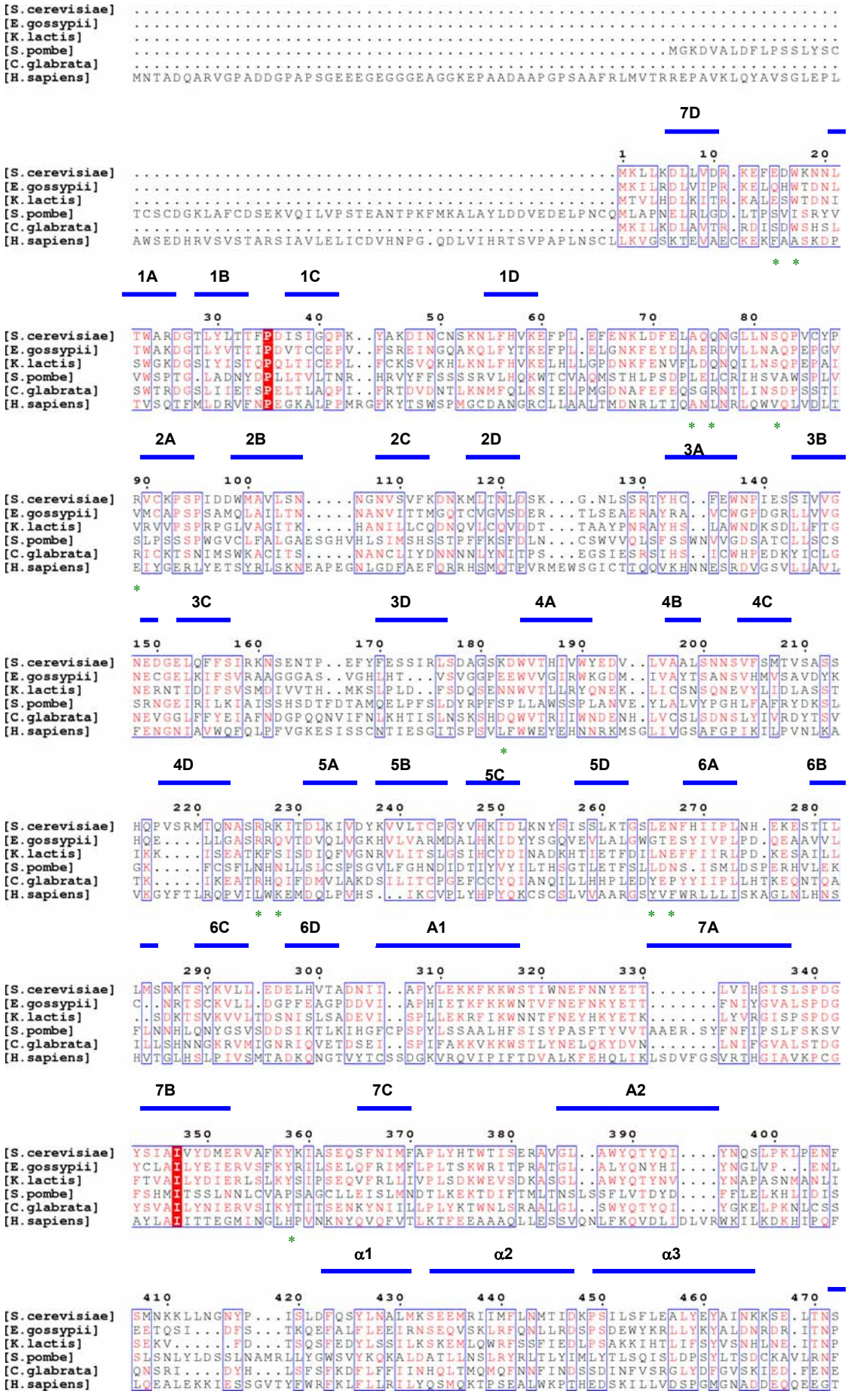
In summary, it is very likely that the N-terminal part of τ60 homologs and the C-terminal part of τ91 homologs form a β-propeller structure and that the β-propeller – β-propeller interaction is common in all organisms, something that would imply a common mechanism in τB assembly and in DNA binding for all organisms.

A: τ91 sequence alignment




4B
4C
4D
5A
5B
5C
5D
6A
6B
6C
6D
6D'
7A
7B
7C
[S.cerevisiae]
[H.sapiens]
[S.pombe]
[C.glabrata]
[E.gossypii]
[K.lactis]

B: τ60 sequence alignment



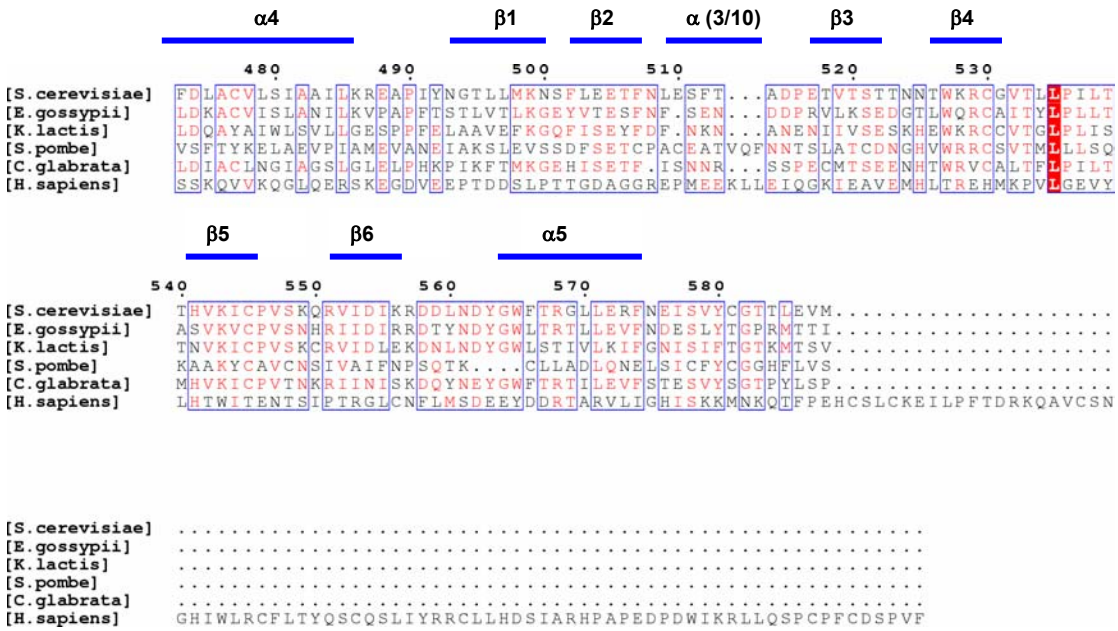


Figure 4.1 Sequence alignment of (A) τ91 and (B) τ60. Highly similar residues (with score >70%) as defined by Risler et al., 1988 are depicted in red. Residues that are 100% conserved are in red boxes. The secondary structure of both τ91 and τ60 is in blue and was assigned after visual inspection of the structure. Green stars indicate residues involved in the interaction with τ60 (A) and with τ91 (B). Disordered regions are indicated in dashed pink lines. The figures were generated with MultiAlin (Corpet, 1988) and ESPript (Gouet et al., 1999).

4.2. Importance of the β propeller – β propeller interaction

As already mentioned, in order to reconstitute a τB complex able to bind DNA, all three proteins τ138, τ91, and τ60, are required. In addition, only τ138 and τ91 display DNA binding activity, and in the case of τ91 the binding is non-specific. All these results lead to the conclusion that very likely there is not one single protein that binds specifically with high affinity to DNA, but DNA binding is rather a cooperative effect involving τ138 and τ91 directly contacting the DNA but also τ60 as part of the complex. Alternatively, the DNA binding could also be explained by an allosteric mechanism, where the domains of τ138 and τ91 displaying non-specific DNA binding activity are unmasked upon recruitment of the τB complex. Thus, in both mechanisms the interaction between τ60 and τ91 appears to play a crucial role. In addition, this novel β-propeller – β-propeller interaction could be essential for providing an

increased interacting region for a big protein like τ138 and for reassuring the recruitment of a very stable τB complex able to bind with very high affinity to the B box, which is the primary and maybe the most critical step in the preinitiation complex formation.

As for the C-terminal domain of τ60; it is not clear whether it participates in the τB-DNA recruitment. The fact that it does not interact with either of the two propellers and its upstream positioning make it very possible that it plays no role in this process. Moreover, as already mentioned, based on protein-DNA cross-linking experiments, τ91 has been mapped at the most 3' location of TFIIIC-5S RNA gene complexes (Braun et al., 1992) and τ60 appears to link τA and τB subcomplexes and to participate in TFIIIB assembly via its interaction with TBP (Deprez et al., 1999). Thus, the C-terminal domain of τ60 appears to be oriented upstream, whereas the two β-propellers are positioned downstream towards the 3' location of tRNA genes. Therefore, it could be simply the domain that links τA and τB subcomplexes and establishes the interactions with TBP. The linker that connects the τ60 β-propeller with the C-terminal domain of τ60 interacts with the propeller mostly through polar interactions, whereas with the C-terminal domain mostly through hydrophobic interactions (Figure 4.2). In particular, linker residues 400-410 show very few interactions with the two domains. Hypothetically, the linker might bend and the C-terminal domain could move slightly upon incorporation of TBP into the TFIIIC-TFIIIB-DNA complex, allowing the C-terminal domain to extend further upstream (Figure 4.3). This flexibility of the linker could also be a way of adapting to the different distances separating the A and B blocks in natural tRNA genes. However, further functional studies will be required in order to shed more light into the function of this domain.

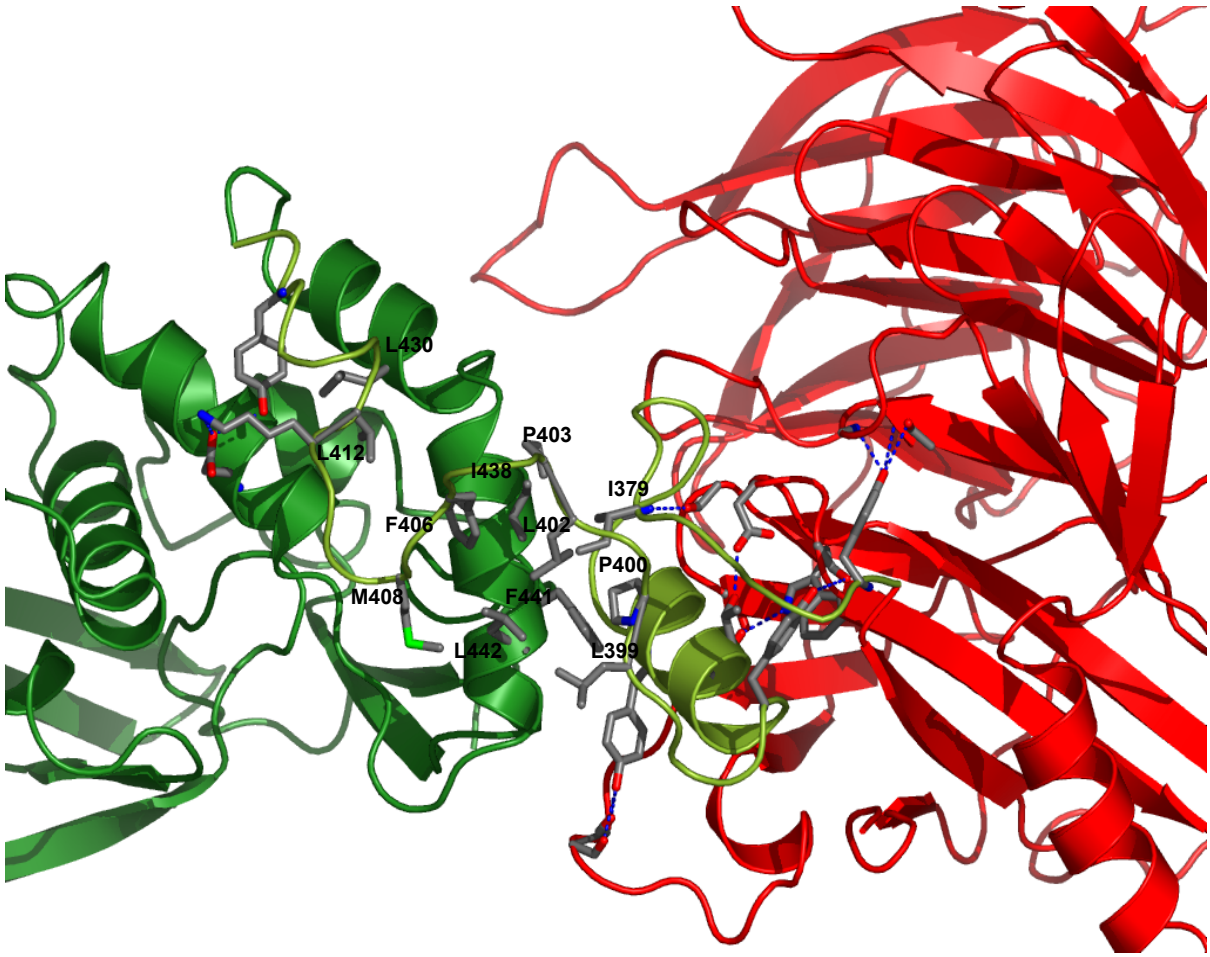


Figure 4.2 Cartoon diagram of the interactions between the linker and the two domains of τ60. Hydrogen bonds are indicated in dashed lines. Only hydrophobic residues are labelled.

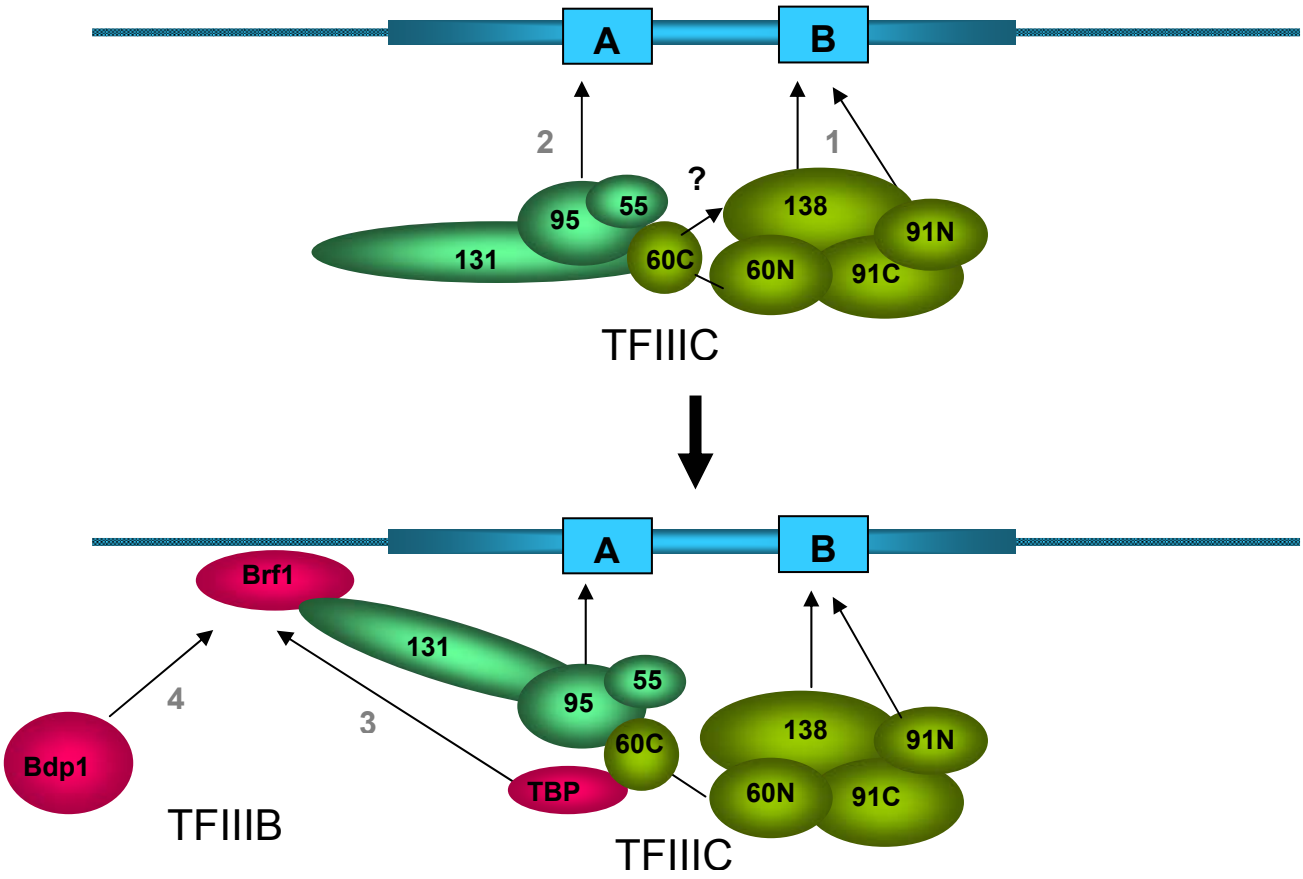


Figure 4.3 Schematic representation of a proposed model for the possible roles of τ91 and the two domains of τ60 during the preinitiation complex formation in class III genes.

4.3. Possible interactions of the τ60/Δτ91 complex with other TFIIIC subunits

In order to identify the possible surface of the τ60/Δτ91 complex involved in the interaction with τ138, a surface representation showing residue conservation of the τ60/Δτ91 complex was generated (Figure 4.4A). Interestingly, we observed two highly conserved regions located on the side of the complex where the channel of Δτ91 is open, whereas the opposite side of the complex displays no conserved regions (Figure 4.4B). The conserved regions might mediate interactions with τ138. A temperature sensitive mutation in the τ138 subunit (tfc3-G349E) decreases the affinity of TFIIIC for the B box element (Lefebvre et al., 1994). This mutation is located in one of the two small conserved HMG domains of τ138, and the wild-type Gly residue is actually invariant from yeast to human (reviewed in Huang et al., 2001). A point mutation in τ91 (E330K) was found to suppress the thermosensitive phenotype of tfc3-G349E mutant affected in τ138. The suppressor mutation alleviated the DNA binding and transcription defects of mutant TFIIIC *in vitro* (Arrebola et al., 1998). Although residues 320-337 of τ91 form a disordered loop in the crystal structure, the base of the loop localizes to the conserved region (Figure 4.4A). Therefore, this region is very likely to be the interacting surface of the τ60/τ91 complex with τ138, and possibly the channel of Δτ91 plays a role in this interaction. This region is highly acidic, with a small basic region on the top of the complex surface, implying a possible electrostatic interaction between the τ60/τ91 complex and τ138 (Figure 4.4C). It should also be mentioned that τ95 is another possible candidate that might bind to these regions via τ91, since it has been previously reported that there is a direct interaction between τ91 and τ95 (Jourdain et al., 2003).

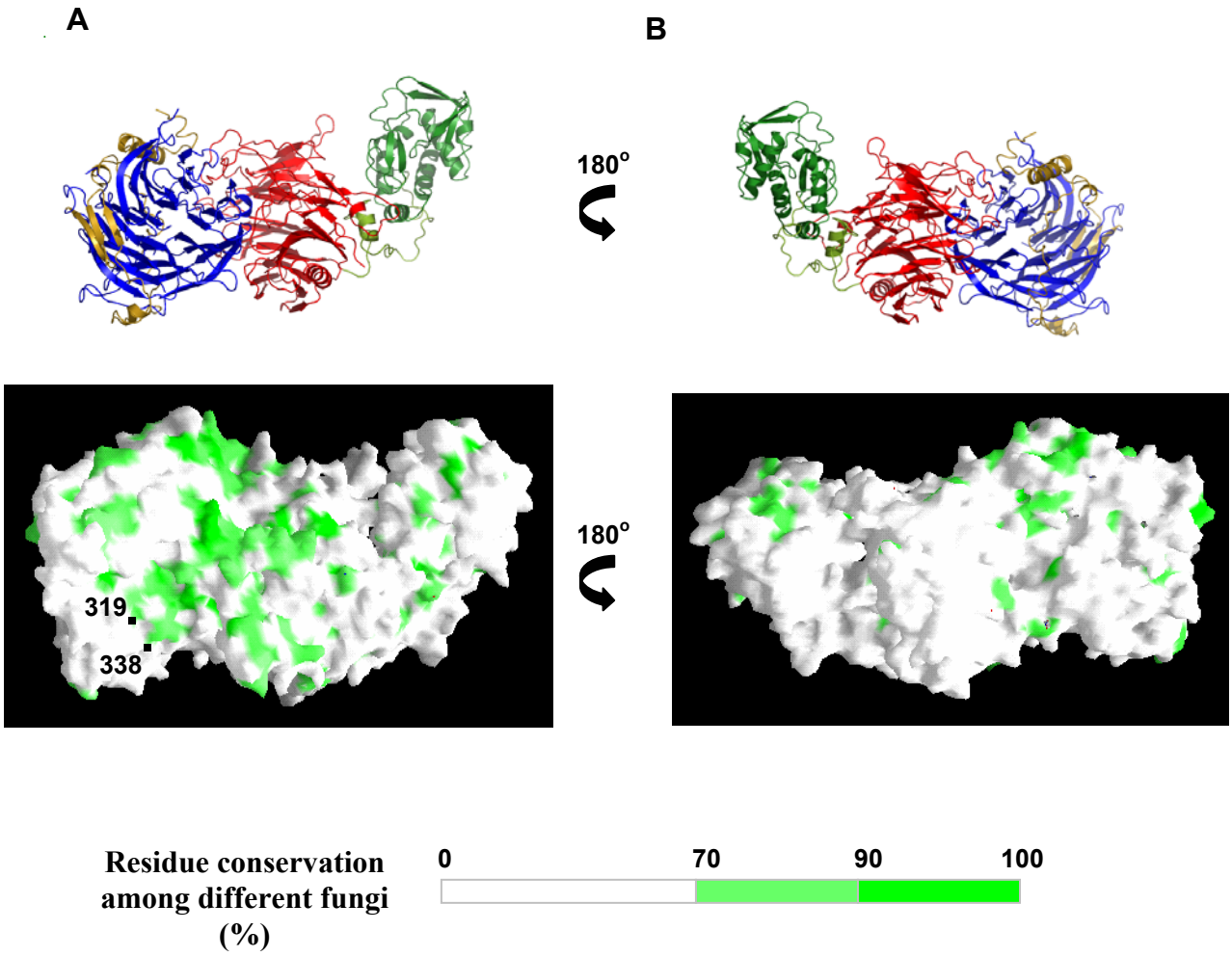


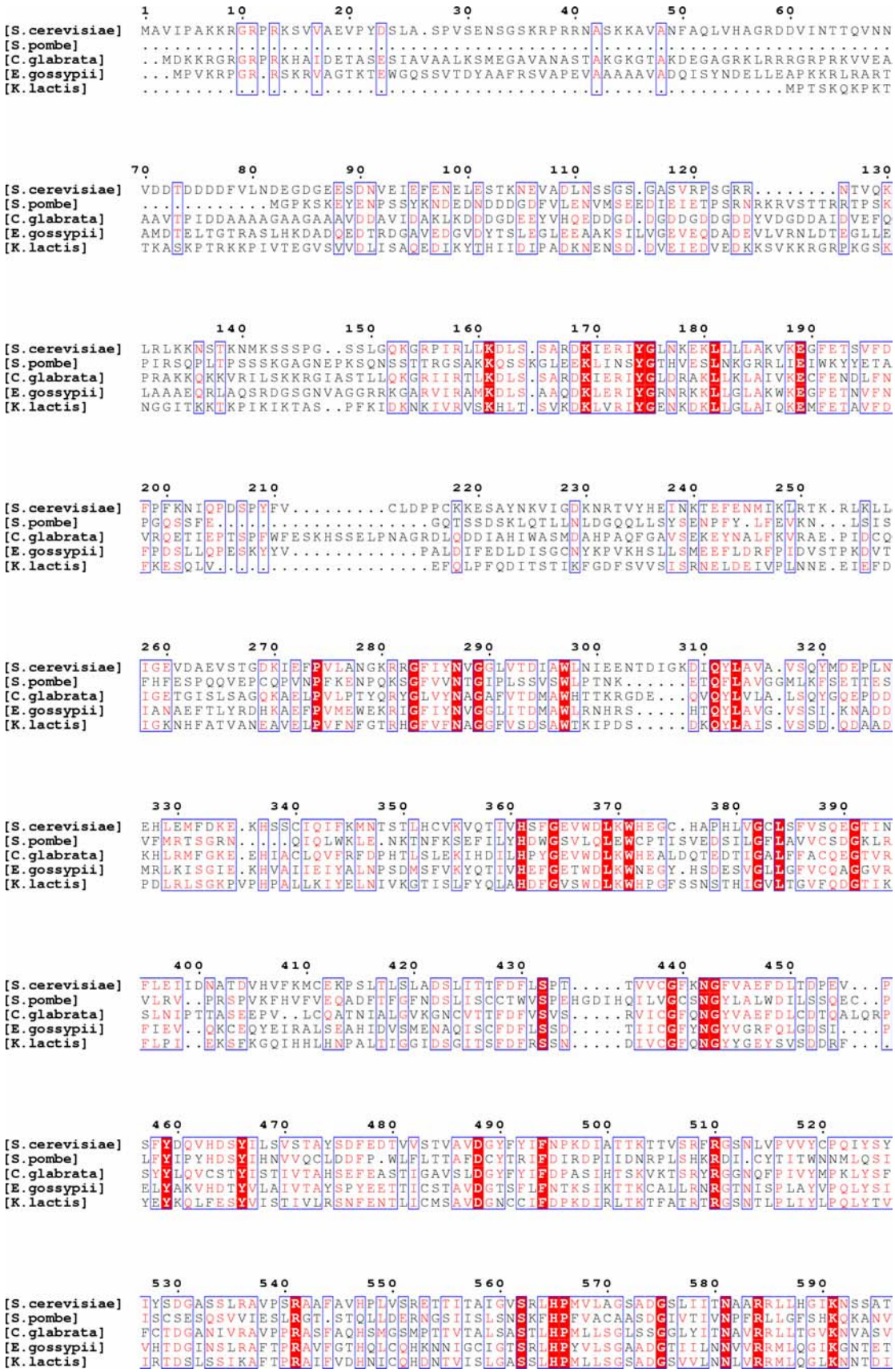
Figure 4.4 (A and B) Surface representation of residue conservation of the complex surface. Residue conservation above 70% is represented from light to darker green according to the scale. The representation was based on sequence alignments among different fungi (see appendix). The complex in Figure B is rotated 180° in respect with Figure A. Cartoon representations of the τ60/Δτ91 complex are shown to orient the reader. (C) Electrostatic representation of the complex surface. Blue and red coloring represent the positive and negative electrostatic potential. The orientation of the complex is the same as in Figure A. The figures were generated with GRASP (Nicholls et al., 1991).

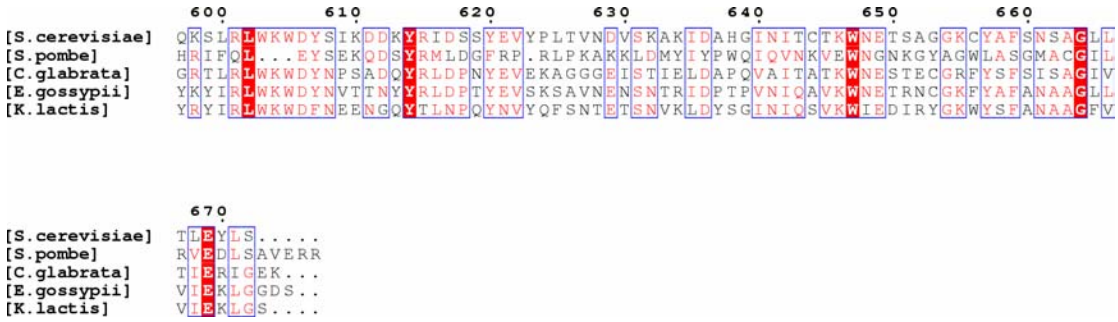
4.4. Concluding remarks and perspectives

We have determined the X-ray crystal structure of yeast τ60 complexed with the C-terminal part of τ91. τ60 consists of an N-terminal β-propeller domain and a C-terminal domain which has a novel fold and is very likely the part that links τA with τB and contacts TBP. Δτ91 also shows a β-propeller fold. Heterodimer formation involves residues in the two β-propellers, whereas the C-terminal domain of τ60 does not contribute at all to the interaction. This novel β-propeller – β-propeller interaction appears to be of great importance for the formation of a very stable τB complex able to bind with high affinity and stability to the B box. Our results provide a starting point for further structural and functional studies aimed at elucidating the mechanism of preinitiation complex formation in the RNA polymerase III transcription machinery. Further insights into the molecular basis of this process as far as it concerns TFIIIC, can be achieved by testing which of the domains of τ60 and Δτ91 are indispensable for binding to the B box both *in vivo* and *in vitro*. In addition, the role of the C-terminal domain of τ60 can be clarified by checking interactions with possible protein-partners oriented upstream, like the TBP and τ95. However, the most challenging and interesting approach is the reconstitution of the entire TFIIIC complex and determination of its structure by electron microscopy or X-ray crystallography. However, it should be mentioned that this goal could be very ambitious due to the fact that the linker connecting τA and τB subcomplexes shows remarkable flexibility. This flexibility could be a major problem for further structural studies. Therefore, the reconstitution of the τA and τB subcomplexes separately could be an alternative approach yielding an equivalent scientific result.

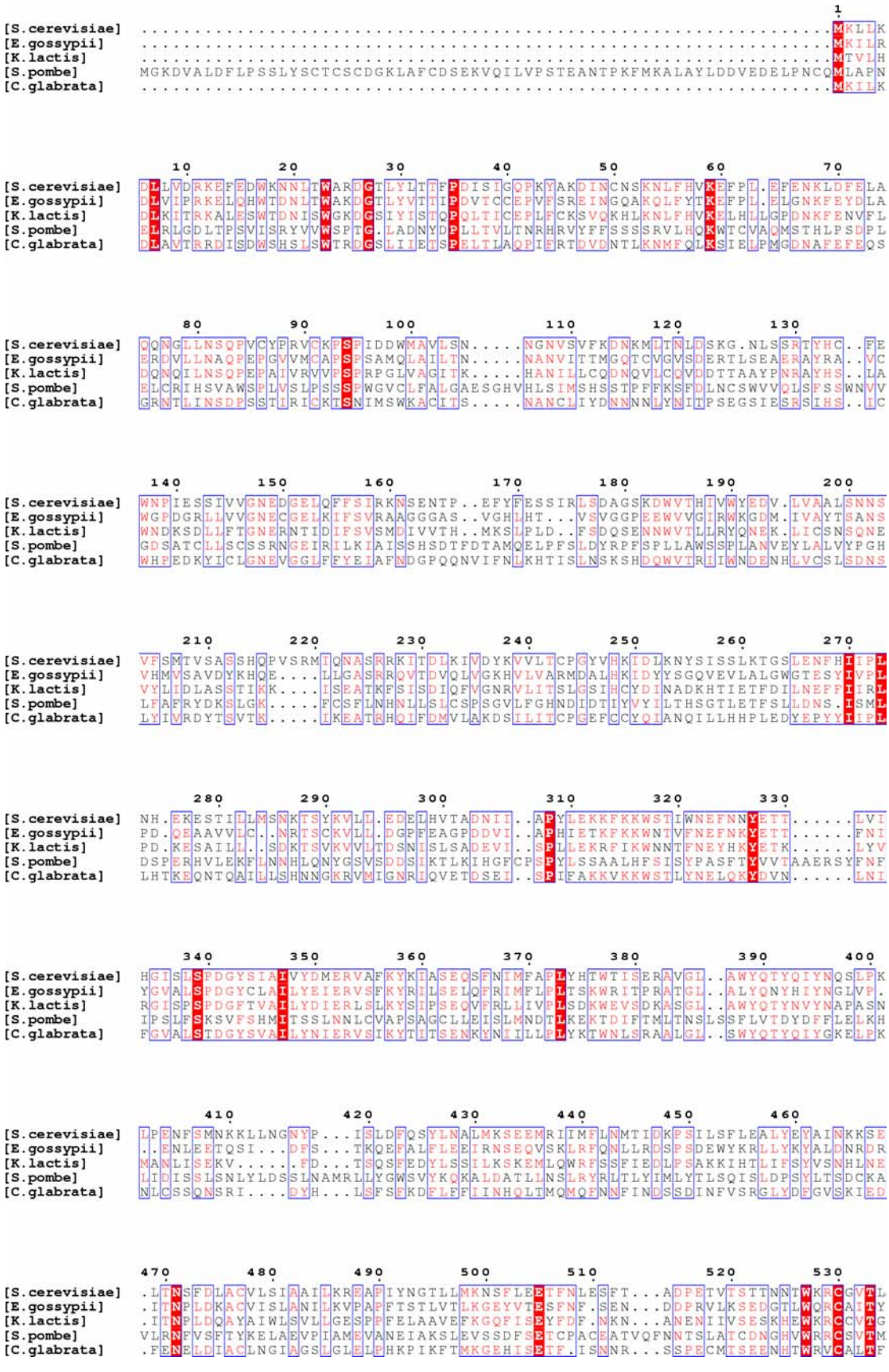
Appendix

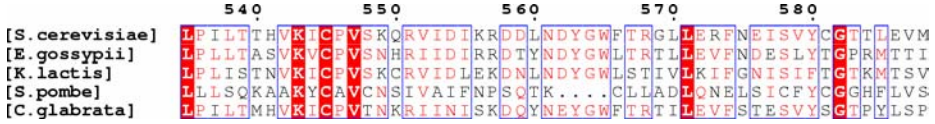
Sequence alignment of τ91 among different fungi




600
610
620
630
640
650
660
[S.cerevisiae] QKSLRLWKWDYSIKDDKYRIDSSYEVYPLTVNDVSKAKIDAHGINITCTKWNETSAGGKCYAFSNSAGLL
[S.pombe] HRIFQL...EYSEKQDSYRMLDGFRP.RLPKAKKLDMYIYPWQIQVNKVEWNGNKGYAGWLASGMACGIL
[C.glabrata] GRTLRLWKWDYNPSADQYRLDPNYEVEKAGGGEISTIELDAPQVAITATKWNESTECGRFYSFSISAGIV
[E.gossypii] YKYIRLWKWDYNVTTNYYRLDPTYEVSKSAVNENSNTRIDPTPVNIQAVKWNETRNCGKFYAFANAAGLL
[K.lactis] YRYIRLWKWDFNEENGQYTLNPQYNVYQFSNTETSNVKLDYSGINIQSVKWIEDIRYGKWYSFANAAGFV
670
[S.cerevisiae] TLEYLS.....
[S.pombe] RVEDLSAVERR
[C.glabrata] TIERIGEK...
[E.gossypii] VIEKLGGDS..
[K.lactis] VIEKLGS....

Sequence alignment of τ60 among different fungi




540
550
560
570
580
[S.cerevisiae] LPILTTHVKICPVSKQRVIDIKRDDLNDYGWFTRGLLERFNEISVYCGTTLEVM
[E.gossypii] LPLLTASVKVCPVSNHRIIDIRRDTYNDYGWLTRTLLEVFNDESLYTGPRMTTI
[K.lactis] LPLISTNVKICPVSKCRVIDLEKDNLNDYGWLSTIVLKIFGNISIFTGTKMTSV
[S.pombe] LLLSQKAAKYCAVCNSIVAIFNPSQTK....CLLADLQNELSICFYCGGHFLVS
[C.glabrata] LPILTMHVKICPVTNKRIINISKDQYNEYGWFTRTILEVFSTESVYSGTPYLSP

References

Allison, D. S. and Hall, B. D. (1985) Effects of alterations in the 3' flanking sequences on *in vivo* and *in vitro* expression of the yeast SUP4-o tRNATyr gene. EMBO J. **4**, 2657-2664.

Andrade, M. A., Perez-Iratxeta, C. and Ponting, C. P. (2001) Protein Repeats : Structures, Functions, and Evolution. J. Struct. Biol. **134**, 117-131.

Arrebola, R., Manaud, N., Rozenfeld, S., Marsolier, M. C., Lefebvre, O., Carles, C., Thuriaux, P., Conesa, C., and Sentenac, A. (1998) τ91, an essential subunit of yeast transcription factor IIIC, cooperates with τ138 in DNA binding. Mol. Cell. Biol. **18**, 1-9.

Bartholomew, B., Kassavetis, G. A., Braun, R., and Geiduschek, E. P. (1990) The subunit structure of *Saccharomyces cerevisiae* transcription factor IIIC probed with a novel photocrosslinking reagent. EMBO J. **9**, 2197-2205.

Bartholomew, B., Durkovich, D., Kassavetis, G. A., and Geiduschek, E. P. (1993) Orientation and topography of RNA polymerase III in transcription complexes Mol. Cell. Biol. **13**, 942-952.

Baker, R. E., Gabrielsen, O. S., and Hall, B. D. (1986) Effects of tRNA Tyr point mutations on the binding of yeast RNA polymerase III transcription factor C. J. Biol. Chem. **261**, 5275-5282.

Baker, R. E., Camier, S., Sentenac, A., Hall, B. D. (1987) Gene size differentially affects the binding of yeast transcription factor τ to two intragenic regions. Proc. Natl. Acad. Sci. USA **84**, 8768-8772.

Bogenhagen, D. F. and Brown, D. D. (1981) Nucleotide sequences in Xenopus 5S DNA required for transcription termination. Cell **24**, 261-270.

Braun, B. R., Bartholomew, B., Kassavetis, G. A., and Geiduschek, E. P. (1992) Topography of transcription factor complexes on the *Saccharomyces cerevisiae* 5S RNA gene. J. Mol. Biol. **228**, 1063-1077.

Brun, I., Sentenac, A., and Werner, M. (1997) Dual role of the C34 subunit of RNA polymerase III in transcription initiation. EMBO J. **16**, 5730-5741.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. and Warren G. L. (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Cryst. Sect. D **54**, 905-921.

Burnol, A.-F., Margottin, F., Huet, J., Almouzni, G., Prioleau, M.-N., Mechali, M., and Sentenac, A. (1993) TFIIIC relieves repression of U6 snRNA transcription by chromatin. Nature **362**, 475-477.

Camier, S., Dechampesme, A. M., and Sentenac, A. (1995) The only essential function of TFIIIA in yeast is the transcription of 5S rRNA genes. Proc. Natl. Acad. Sci. **92**, 9338-9342.

CCP4 (Collaborative Computational Project) Number 4, (1994) The CCP4 suite: programs for protein crystallography. Acta Cryst. Sect. D **50**, 760-763.

Chaussivert, N., Conesa, C., Shaaban, S., and Sentenac, A. (1995) Complex interactions between yeast TFIIIB and TFIIIC. J. Biol. Chem. **270**, 15353-15358.

Chedin, S., Riva, M., Schultz, P., Sentenac, A., and Carles, C. (1998) The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIS-like subunit and is important for transcription termination. Genes Dev. **12**, 3857-3871.

Ciliberto, G., Raugei, G., Costanzo, F., Dente, L., and Cortese, R. (1983) Common and interchangeable elements in the promoters of genes transcribed by RNA polymrase III. Cell **32**, 725-733.

Clark, M. E., Hammerle, T., Wimmer, E., and Dasgupta, A. (1991) Poliovirus proteinase 3C converts an active form of transcription factor IIIC to an inactive form: A mechanism for inhibition of host cell polymerase III transcrption by poliovirus. EMBO J. **10**, 2941-2947.

Conesa, C., Swanson, R. N., Schultz, P., Oudet, P., and Sentenac, A. (1993) On the subunit composition, stoichiometry and phosphorylation of the yeast transciption factor TFIIIC/tau. J. Biol. Chem. **268**, 18047-18052.

Corpet, F. (1988) Multiple sequence alignment with hierarchical clustering. Nucl. Acids Res. **16**(22), 10881-10890.

De La Fortelle, E., and Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement for the multiple isomorphous replacement and multiwavelength anomalous diffraction methods. Methods Enzymol. **276**, 472-494.

Deprez, E., Arrebola, R., Conesa, C., and Sentenac, A. (1999) A subunit of yeast TFIIIC participates in the recruitment of TATA-binding protein. Mol. Cell. Biol. **19**, 8042-8051.

Dieci, G. and Sentenac, A. (1996) Facilitated recycling pathway for RNA polymerase III. Cell **84**, 245-252.

Dieci, G., Percudani, R., Giuliodori, S., Bottarelli, L., and Ottonello, S. (2000) TFIIIC-independent *in vitro* transcription of yeast tRNA genes. J. Mol. Biol. **299**, 601-613.

Dumay, H., Rubbi, L., Sentenac, A., and Marck, C. (1999) Interaction between yeast RNA polymerase III and transcription factor TFIIIC via ABC10α and τ131 subunits. J. Biol. Chem. **274**, 33462-33468.

Engelke, D. R., Ng, S. Y., Shastry, B. S., and Roeder, R. G. (1980) Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. Cell **19**, 717-728.

Eschenlauer, J. B., Kaiser, M. W., Gerlach, V. L., and Brow, D. A. (1993) Architecture of a yeast U6 RNA gene promoter. Mol. Cell. Biol. **13**, 3015-3026.

Evans, P. R. (1993) Data reduction, Proceedings of CCP4 study weekend, on Data Collection and Processing, pp. 114-122, York, UK.

Ferrari, R., Rivetti, C., Acker, J., and Dieci, G. (2004) Distinct roles of transcription factors TFIIIB and TFIIIC in RNA polynerase III transcription reinitiation. Proc. Natl. Acad. Sci. USA **101**, 13442-13447.

Ferri, M. L., Peyroche, G., Siaut, M., Lefebvre, O., Carles, C., Conesa, C., and Sentenac, A., (2000) A novel subunit of yeast RNA polymerase III interacts with the TFIIIB-related domain of TFIIIB70. Mol. Cell. Biol. **20**, 488-495.

Flores, A., Briand, J. F., Gadal, O., Andrau, J. C., Rubbi, L., Van Mullem, V., Boschiero, C., Goussot, M., Marck, C., and Carles, C. (1999) A protein-protein interaction map of yeast RNA polymerase III. Proc. Natl. Acad. Sci. **96**, 7815-7820.

French, S. and Wilson, K. (1978) On the treatment of negative intensity observations. Acta Cryst. Sect. A **34**, 517-525.

Gabrielsen, O. S., Marzouki, N., Ruet, A., Sentenac, A., and Fromageot, P., (1989) Two polypeptide chains in yeast transcription factor tau interact with DNA. J. Biol. Chem. **264**, 7505-7511.

Garman, E. F. and Schneider, T. R. (1997) Macromolecular Cryocrystallography. J. Appl. Cryst. **30**, 211-237.

Geiduschek, E. P. and Tcchini-Valentini, G. P. (1988) Transcription by RNA polymerase III. Annu. Rev. Biochem. **57**, 873-914.

Geiduschek, E. P. and Kassavetis, G. A. (2001) The RNA polymerase III transcription apparatus. J. Mol. Biol. **310**, 1-26.

Gerlach, V. L., Whitehall, S. K., Geiduschek, E. P., and Brow, D. A. (1995) TFIIIB placement on a yeast U6 RNA gene *in vivo* is directed primarily by TFIIIC rather than by sequence-specific DNA contacts. Mol. Cell. Biol. **15**, 1455-1466.

Ginsberg, A. M., King, B. O., and Roeder, R. G. (1984) *Xenopus* 5S gene transcription factor, TFIIIA: Characterization of a cDNA clone and measurement of RNA levels throughout development. Cell **39**, 479-489.

Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) ESPript: multiple sequence alignments in PostScript. Bioinformatics **15**, 305-308.

Grove, A., Kassavetis, G. A., Johnson, T. E., and Geiduschek, E. P. (1999) The RNA polymerase III-recruiting factor TFIIIB induces a DNA bend between the TATA box and the transcriptional start site. J. Mol. Biol. **285**, 1429-1440.

Harismendy, O., Gendrel, C.-G., Soularue, P., Gidrol, X., Sentenac, A., Werner, M., and Lefebvre, O. (2003) Genome-wide location of yeast RNA polymerase III transcription machinery. EMBO J. **22**, 4738-4747.

Holm, L. and Sander, S. (1995) Dali: a network tool for protein structure comparison. Trends Biochem. Sci. **20**, 478-480.

Hsieh, Y. J., Kundu, T. K., Wang, Z., Kovelman, R., and Roeder, R. G., (1999a) 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. **19**, 7697-7704.

Hsieh, Y. J., Wang, Z., Kovelman, R., and Roeder, R. G., (1999b) Cloning and characterization of two evolutionary conserved subunits (TFIIIC102 and TFIIIC63) of human TFIIIC and their involvement in functional interactions with TFIIIB and RNA polymerase III. Mol. Cell. Biol. **19**, 4944-4952.

Huet, J., Conesa, C., Carles, C., and Sentenac, A. (1997) A cryptic DNA binding domain at the COOH terminus of TFIIIB70 affects formation, stability and function of preinitiation complexes. J. Biol. Chem. **272**, 18341-18349.

Huang, Y., Hamada, M., and Maraia, R. J. (2000) Isolation and cloning of four subunits of a fission yeast TFIIIC complex that includes an ortholog of the human regulatory protein TFIIICβ. J. Biol. Chem. **275**, 31480-31487.

Huang, Y. and Maraia, R. J. (2001) Comparison of the RNA polymerase III transcription machinery in *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae* and human. Nucleic Acids Res. **29**, 2675-2690.

Jancarik, J. and Kim, S. (1991) Sparse matrix sampling: a screening method for crystallization of proteins. J. Appl. Cryst. **24**, 409-411.

Jawad, Z. and Paoli, M. (2002) Novel sequences propel familiar folds. Structure **10**, 447-454.

Joazeiro, C. A., Kassavetis, G. A., and Geiduschek, E. P. (1996) Alternative outcomes in assembly of promoter complexes: the roles of TBP and a flexible linker in placing TFIIIB on tRNA genes. Genes Dev. **10**, 725-739.

Joazeiro, C. A., Kassavetis, G. A., and Geiduschek, E. P. (1994) Identical components of yeast transcription factor IIIB are required and sufficient for transcription of TATA box-containing and TATA-less genes. Mol. Cell. Biol. **14**, 2798-2808.

Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A **47**, 110-119.

Jourdain, S., Acker, J., Ducrot, C., Sentenac, A., and Lefebvre, O. (2003) The tau95 subunit of yeast TFIIIC influences upstream and downstream functions of TFIIIC-DNA complexes. J. Biol. Chem. **278**, 10450-10457.

Kassavetis, G. A., Riggs, D. L., Negri, R., Nguyen, L. H., and Geiduschek, E. P. (1989) Transcription factor IIIB generates extended DNA interactions in RNA polymerase III transcription complexes on tRNA genes. Mol. Cell. Biol. **9**, 2551-2566.

Kassavetis, G. A., Braun, B. R., Nguyen, L., H., and Geiduschek, E. P. (1990) *S. cerevisiae* TFIIIB is the transcription initiation factor proper of RNA polynerase III, while, TFIIIA and TFIIIC are assembly factors. Cell **60**, 235-245.

Kassavetis, G. A., Bartholomew, B., Blanco, J. A., Johnson, T. E. and Geiduschek, E. P. (1991) Two essential components of the *Saccharomyces cerevisiae* transcripion factor TFIIIB: transcription and DNA-binding properties. Proc. Natl. Acad. Sci. USA **88**, 7308-7312.

Kassavetis, G. A, Joazeiro, C. A. P., Pisano, M., Geiduschek, E. P., Colbert, T., Hahn, S., and Blanco, J. A. (1992) The role of the TATA-binding protein in the assembly and function of the multisubunit yeast RNA polymerase III transcription factor, TFIIIB. Cell **71**, 1055-1064.

Kassavetis, G. A., Kumar, A., Letts, G., and Geiduschek, E. P. (1998) A post-recruitment function for the RNA polymerase III transcription-initiation factor IIIB. Proc. Natl. Acad. Sci. **95**, 9196-9201.

Khoo, B., Brophy, B., and Jackson, S. P. (1994) Conserved functional domains of the RNA polymerase III general transcription factor BRF. Genes Dev. **8**, 2879-2890.

Kumar, A., Kassavetis, G. A., Geiduschek, E. P., Hambalko, M., and Brent, C. J. (1997) Functional dissection of B'' component of RNA polymerase III transcription factor TFIIIB: a scaffolding protein with multiple roles in assembly and initiation of transcription. Mol. Cell. Biol. **17**, 1868-1880.

Kundu, T. K., Wang, Z., and Roeder, R. G. (1999) Human TFIIIC relieves chromatin mediated repression of RNA polymerase III transcription and contains an intrinsic histone acetyl-transferase activity. Mol. Cell. Biol. **19**, 1605-1615.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. **26**, 283-291.

Lefebvre, O., Carles, C., Conesa, C., Swanson, R. N., Bouet, F., Riva, M., and Sentenac, A. (1992) TFC3: Gene Encoding the B-Block Binding Subunit of the Yeast Transcription Factor IIIC. Proc. Natl. Acad. Sci. USA **89**, 10512-10516.

Lefebvre, O., Ruth, J., and Sentenac, A. (1994) A mutation in the largest subunit of yeast TFIIIC affects tRNA and 5S RNA synthesis. Identification of two classes of suppressors. J. Biol. Chem. **269**, 23374-23381.

Leslie, A. (1991) Macromolecular data processing. In Crystal Computing, (Moras, V. D., Podjarny, A. D. and Thierry, J. C., eds.), pp. 27-38, Oxford University Press, Oxford, UK.

Librizzi, M. D., Moir, R. D., Brenowitz, M., and Willis, I. M. (1996) Expression and purification of the RNA polymerase III transcription specificity factor IIIB70 from *Saccharomyces cerevisiae* and its cooperative binding with TATA-binding protein. J. Biol. Chem. **271**, 32695-32701.

Mao, X., and Darby, M. K. (1993) A position-dependent transcription-activating domain in TFIIIA. Mol. Cell. Biol. **13**, 7496-7506.

Manaud, N., Arrebola, R., Buffin-Meyer, B., Lefebvre, O., Voss, H., Riva, M., Conesa, C., and Sentenac, A. (1998) A chimeric subunit of yeast transcription factor IIIC foms a subcomplex with tau95. Mol. Cell. Biol. **18**, 3191-3200.

Marck, C., Lefebvre, O., Carles, C., Riva, M., Chaussivert, N., Ruet, A., and Sentenac, A. (1993) The TFIIIB-assembling subunit of yeast transcription factor TFIIIC has both tetratricopeptide repeats and basic helix-loop-helix motifs. Proc. Natl. Acad. Sci. USA **90**, 4027-4031.

Margottin, F., Dujardin, G., Gerard, M., Egly, J. M., Huet, J. and Sentenac, A. (1991) Participation of the TATA factor in transcription of the yeast U6 gene by RNA polymerase C. Science **251**, 424-426.

Marzouki, N., Camier, S., Ruet, A., Moenne, A., and Sentenac, A. (1986) Selective proteolysis defines two DNA bining domains in yeast transcription factor tau. Nature **323**, 176-178.

Matsutani, S. (2004) Similarities in transcription factor IIIC subunits that bind to the posterior regions of internal promoters for RNA polymerase III. BMCB Evol. Biol. **4**, 1-12.

Matthews, B. W. (1968) The solvent content of protein crystals. J. Mol. Biol. **33**, 491-497.

McPherson, A. (1992) Preparation and analysis of protein crystals. Robert E. Krieger publishing Co. Inc, Malabar, USA.

Miller, J., McLachlan, A. D., and Klug, A. (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. EMBO J. **4**, 1609-1614.

Moenne, A., Camier, S., Anderson, G., Margottin, F., Beggs, J., and Sentenac, A. (1990) The U6 gene of *Saccharomyces cerevisiae* is transcribed by RNA polymerase C (III) *in vivo* and *in vitro*. EMBO J. **9**, 271-277.

Moqtaderi, Z., and Struhl, K. (2004) Genome-wide occupancy profile of the RNA polymerase III machinery in *Saccharomyces cerevisiae* reveals loci with incomplete transcription complexes. Mol. Cell. Biol. **24**, 4118-4127.

Nicholls, A., Sharp, A. K., and Honing, B. (1991) Protein folding and association: insight from the interfacial and thermodynamic properties of hydrocarbon. Proteins **11**, 281-296.

Parsons, M. C. and Weil, P. A. (1990) Purification and characterization of *Saccharomyces cerevisiae* transcription factor TFIIIC. Polypeptide composition defined with polyclonal antibodies. J. Biol. Chem. **265**, 5095-5103.

Persinger, J., Sengupta, S. M., and Bartholomew, B. (1999) Spatial organization of the core region of yeast TFIIIB-DNA complexes. Mol. Cell. Biol. **19**, 5218-5234.

Pieler, T., Appel, B., Oei, S. L., Mentzel, H., and Erdmann, V. A. (1985a) Point mutational analysis of *Xenopus laevis* 5S gene promoter. EMBO J. **4**, 1847-1853.

Pieler, T., Oei, S. L., Hamm, J., Engelke, U., and Erdmann, V. A. (1985b) Functional domains of the *Xenopus laevis* 5S gene promoter. EMBO J. **4**, 3751-3756.

Ramachandran, G. N. and Sassiekharan, V. (1968) Conformation of polypeptides and proteins. Adv. Prot. Chem. **23**, 283-438.

Rameau, R., Puglia, K., Crowe, A., Sethy, I., and Willis, I. M. (1994) A mutation in the second largest subunit of TFIIIC increases a rate-limiting step in transcription by RNA polymerase III. Mol. Cell. Biol. **14**, 822-830.

Risler, J. L., Delorme, M. O., Delacroix, H., and Henaut, A. (1988) Amino acid substitutions in structurally related proteins. A pattern recognition approach. Determination of a new and efficient scoring matrix. J. Mol. Biol. **204**, 1019-1029.

Roberts, D. N., Stewart, A. J., Huff, J. T., and Bradley, R. C. (2003) The RNA polymerase III transcriptome revealed by genome-wide localization and activity-occupancy relationships. Proc. Natl. Acad. Sci. USA **100**, 14695-14700.

Roussel, A. and Cambilleau, C. (1989) Turbo-Frodo. In silicon graphics geometry partners directory, pp. 77-79, Silicon Graphics, Mountain View, USA.

Rowland, O. and Segall, J. (1998) A hydrophobic segment within the 81-amino-acid domain of TFIIIA from *S. cerevisiae* is essential for its transcription factor activity. Mol. Cell. Biol. **18**, 420-432.

Rubinson, K. A., Ladner, J. E., Tordova, M. and Gilliland, G. M. (2000) Cryosalts: suppression of ice formation in macromolecular crystallography. Acta Crystallogr. Sect. D **56**, 996-1001.

Schneider, T. R. and Sheldrick, G. M. (2002) Substructure solution with SHELXD. Acta Crystallogr. Sect. D **58**, 1772-1779.

Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. (1998) SMART, a simple modular architecture research tool: Identification of signaling domains. Proc. Natl. Acad. Sci. USA **95**, 5857-5864.

Schultz, P., Marzouki, N., Marck, C, Ruet, A., Oudet, P., and Sentenac, A. (1989) The two DNA-binding domains of yeast transcription factor τ as observed by scanning transmission electron microscopy. EMBO J. **8**, 3815-3824.

Schramm, L. and Hernandez, N. (2002) Recruitment of RNA polymerase III to its target promoters. Genes Dev. **16**, 2593-2620.

Sharp, S., DeFranco, D., Dingermann, T., Farrel, P., and Soll, D. (1981) Internal control regions for transcription of eukaryotic tRNA genes. Proc. Natl. Acad. Sci. USA **78**, 6657-6661.

Shen, Y., Igo, M., Yalamanchili, P., Berk, A. J., and Dasgupta, A. (1996) DNA binding domain and subunit interactions of transcription factor IIIC revealed by dissection with poliovirus 3C protease. Mol. Cell. Biol. **16**, 4163-4171.

Shindyalov, I. N. and Bourne, P. E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Engineering **11**, 739-747.

Siaut, M., Zaros, C., Levivier, E., Ferri, M.-L., Court, M., Werner, M., Callebaut, I., Thuriaux, P., Sentenac, A., and Conesa, C. (2003) An Rpb4/Rpb7-like complex in yeast RNA polymerase III contains the orthologue of mammalian CGRP-RCP. Mol. Cell. Biol. **23**, 195-205.

Sinn, E., Wang, Z., Kovelman, R., and Roeder, R. G. (1995) Cloning and characterization of a TFIIIC2 subunit (TFIIICb) whose presence correlates with activation of RNA polymerase III-mediated transcription by adenovirus E1A expression and serum factors. Genes Dev. **9**, 675-685.

Smith, T. F., Gaitatzes, C., Saxena, K. and Neer, E. J. (1999) The WD repeat: a common architecture for diverse functions. TIBS **24**, 181-185.

Sprague, E. R., Redd, M. J., Johnson, A. D., and Wolberger, C. (2000) Structure of the C-terminal domain of Tup1, a corepressor of transcription in yeast. EMBO J. **19**, 3016-3027.

Stillman, D. J. and Geiduschek, E. P. (1984) Differential binding of a *S. cerevisiae* RNA polymerase III transcription factor to two promoter segments of a tRNA gene. EMBO J. 3, 847-853.

Teng, T. Y. (1990) Mounting of crystals for macromolecular crystallography in a free-standing thin film. J. Appl. Cryst. **23**, 387-391.

Thuillier, V., Stettler, S., Sentenac, A., Thuriaux, P., and Werner M. (1995) A mutation in the C31 subunit of *Saccharomyces cerevisiae* RNA polymerase III affects transcription initiation. EMBO J. **14**, 351-359.

Wang, Z. and Roeder, R G. (1996) TFIIIC1 acts through a downstream region to stabilize TFIIIC2 binding to RNA polymerase III promoters. Mol. Cell. Biol. **16**, 6841-6850.

Wang, Z. and Roeder, R. G. (1997) Three human RNA polymerase III-specific subunits form a subcomplex with a selective function in specific transcription initiation. Genes Dev. **11**, 1315-1326.

Wang, Z., Bai, L., Hsieh, Y. J., and Roeder, R. G. (2000) Nuclear factor 1 (NF1) affects accurate termination and multiple-round transcription by human RNA polymerase III. EMBO J. **19**, 6823-6832.

Weirich, C. S., Erzberger, J. P., Berger, J. M., and Weis, K. (2004) The N-terminal domain of Nup159 forms a β-propeller that functions in mRNA export by tethering the helicase Dbp5 to the nuclear pore. Molecular Cell **16**, 749-760.

Werner, M., Hermann-Le Denmat, S., Treich, I., Sentenac, A., and Thuriax, P. (1992) Effect of mutations in a zinc-binding domain of yeast RNA polymerase C (III) on enzyme function and subunit association. Mol. Cell. Biol. **12**, 1087-1095.

Werner, M., Chaussivert, N., Willis, I. M., and Sentenac, A. (1993) Interaction between a complex of RNA polymerase III subunits and the 70-kDa component of transcription factor IIIB. J. Biol. Chem. **268**, 20721-20724.

Weser, S., Gruber, C., Hafner, H M., Teichmann, M., Roeder, R. G., Seifart, K. H., and Meissner, W. (2004) Transcription factor (TF)-like nuclear regulator, the 250-kDa form of *Homo sapiens* TFIIIB'', is an essential component of human TFIIIC1 activity. J. Biol. Chem. **279**, 27022-27029.

White, R. J. (1998) RNA polymerase III transcription. Springer-Verlag, Berlin, Germany.

Willis, I. M. (1993) RNA polymerase III: Genes, factors and transcriptional specificity. Eur. J. Biochem. **212**, 1-11.

Willis, I. M. (2002) A universal nomenclature for subunits of the RNA polymerase III transcription initiation factor TFIIIB. Genes Dev. **16**, 1337-1338.

Yoshinaga, S. K., Boulanger, P. A., and Berk, A. J. (1987) Resolution of human transcription factor TFIIIC into two functional components. Proc. Natl. Acad. Sci. **84**, 3585-3589.