

# **Structural and functional studies of eukaryotic translation initiation factors**

Dissertation

for the award of the degree

“Doctor of Philosophy”

Division of Mathematics and Natural Sciences  
of the Georg-August-Universität Göttingen

submitted by

**Yi Liu**

born in Sichuan, China

Göttingen, 2014

## **Thesis committee**

**Prof. Dr. Ralf Ficner (reviewer)**

Georg August University Göttingen  
Institute for Microbiology and Genetics  
Department of Molecular Structural Biology  
Göttingen, Germany

**Prof. Dr. Wolfgang Wintermeyer (reviewer)**

Max Planck Institute for Biophysical Chemistry  
Department of Physical Biochemistry  
Göttingen, Germany

**Prof. Dr. Holger Stark**

Max-Planck-Institut für Biophysikalische Chemie  
3D-Cryo Electron Microscopy Group  
Göttingen, Germany

**Prof. Dr. Reinhard Lührmann**

Max-Planck-Institut für Biophysikalische Chemie  
Department of Cellular Biochemistry  
Göttingen, Germany

**Prof. Dr. Jörg Stülke**

Georg August University Göttingen  
Institute for Microbiology and Genetics  
Department of General Microbiology  
Göttingen, Germany

**Prof. Dr. Kai Tittmann**

Georg August University Göttingen  
Albrecht von Haller Institute  
Department of Bioanalytics  
Göttingen, Germany

Date of oral examination: July 29<sup>th</sup>, 2014



I hereby declare that this thesis entitled "Structural and functional studies of eukaryotic translation initiation factors" has been written independently and with no other sources and aids than quoted. This thesis (wholly or in part) has not been submitted elsewhere for any academic award or qualification.

Yi Liu

June 2014, Göttingen

# Preface

Parts of this thesis were published in:

**Yi Liu**, Piotr Neumann, Bernhard Kuhle, Thomas Monecke, Stephanie Schell, Ashwin Chari, Ralf Ficner (2014) Translation Initiation Factor eIF3b Contains a Nine-Bladed  $\beta$ -Propeller and Interacts with the 40S Ribosomal Subunit. ***Structure***. 22, 1-8. DOI: [http:// dx.doi.org/10.1016/j.str.2014.03.010](http://dx.doi.org/10.1016/j.str.2014.03.010)

Parts of this thesis were presented in the following international conferences:

Yi Liu, Ralf Ficner. Expression, purification and crystallization of eukaryotic translation initiation factor 3 subunit j (eIF3j/Hcr1). EMBO Conference on Protein Synthesis and Translational Control. Heidelberg, Germany, September 2013

Yi Liu, Piotr Neumann, Bernhard Kuhle, Thomas Monecke, Ralf, Ficner. Eukaryotic translation initiation factor 3 subunit b (eIF3b) contains a novel nine-bladed WD40 domain. 22<sup>nd</sup> Annual Conference of the German Crytallographic Society. Berlin, Germany, March 2014

# Abstract

Protein synthesis can be divided into three distinct subprocesses, namely initiation, elongation and termination. During the initiation phase, the elongation competent ribosome, which harbors the initiator tRNA in its P-site base-paired with the start codon of the mRNA, is assembled. Translation initiation is the most regulated and rate-limiting step. In eukaryotes, at least 12 distinct translation initiation factors (eIFs) facilitate this critical process.

The largest among these factors is the multi-subunit complex eIF3 (6 subunits in yeast, 13 subunits (a-m) in mammals). eIF3 is involved in nearly all steps of the initiation process. However, the structure of eIF3 and the mechanism of its versatile functions are still largely elusive. In this work, the functional eIF3 core that is formed by 5 universally conserved subunits (a, b, c, g, and i) was reconstituted *in vitro*, and the interaction network was further explored. Within the 5-subunit eIF3 core, the b subunit (eIF3b) is shown to serve as the major scaffold. The crystal structure of the WD40 domain of *Chaetomium thermophilum* eIF3b was solved *de novo* by means of MAD. The structure revealed an unexpected nine-bladed WD40  $\beta$ -propeller fold. It possesses a central channel that exhibits a circular truncated cone shape rather than the common cylindrical form. Sequence analysis suggests that this nine-bladed propeller architecture is adopted by all eIF3b orthologs. Based on its unique shape, the WD40  $\beta$ -propeller of eIF3b was unambiguously fitted into the recently published cryo-EM map of the 43S preinitiation complex (PIC) (EMDB code: 5658). Due to this ribosomal location, underneath the shoulder of the 40S subunit, eIF3b is indicated to directly interact with the 40S ribosomal subunit. Consistently, stable interactions between eIF3b and the isolated ribosomal protein S9e, as well as the 40S subunit were observed using *in vitro* binding assays. Hence, these results strongly indicate a direct involvement of eIF3b in the assembly of the 43S PIC.

The initiator tRNA bound 43S PIC is recruited to the 5' untranslated region of the mRNA in a manner facilitated by eIF4F. eIF4F is a heterotrimeric complex,

consisting of the m<sup>7</sup>G cap-binding protein eIF4E, the ATP-dependent RNA helicase eIF4A and the multi-scaffold eIF4G. eIF4A melts secondary structures in the 5' UTR for recruitment of the 43S PIC and subsequent scanning. However, the helicase and ATPase activities of isolated eIF4A are very low. High activities occur only when eIF4B and eIF4G are present. In this thesis, a potential RNA-binding fragment of eIF4G is shown to be indispensable for the efficient stimulation. The RNA binding ability and ATPase activity of eIF4A was dramatically increased in the presence of an eIF4G variant that contains eIF4A- and RNA-binding regions. By contrast, an eIF4G truncation containing the eIF4A-binding domain only did not show any enzymatic enhancement. These findings suggest that eIF4G enhances the activities of eIF4A through simultaneous interactions with eIF4A and RNA, and providing the RNA substrate for eIF4A. This model was further supported by site-specific crosslinking experiments.

# Contents

<b>Preface .....</b>	<b>iv</b>
<b>Abstract.....</b>	<b>v</b>
<b>1 Introduction .....</b>	<b>1</b>
1.1 A overview of protein synthesis.....	1
1.2 Eukaryotic translation initiation and the role of initiation factors .....	2
1.3 Eukaryotic translation initiation factor 3.....	7
1.3.1 Composition and interaction network of eIF3 .....	7
1.3.2 The role of eIF3 in translation initiation .....	9
1.4 Eukaryotic translation initiation factor 4.....	12
1.4.1 eIF4E .....	12
1.4.2 eIF4A .....	13
1.4.3 eIF4B and eIF4H.....	17
1.4.4 eIF4G .....	18
1.5 Translational control via eIF4F .....	20
1.5.1 eIF4E regulation .....	20
1.5.2 eIF4A regulation .....	21
1.5.3 eIF4G regulation .....	22
1.6 Aims of this thesis.....	22
<b>2 Materials and methods.....</b>	<b>25</b>
2.1 Cloning.....	25
2.1.1 Materials.....	25
2.1.1.1 Chemicals.....	25
2.1.1.2 Buffers.....	25
2.1.1.3 Cell media .....	26
2.1.1.4 <i>E. coli</i> host strains .....	26
2.1.1.5 Enzymes and reagents .....	26
2.1.1.6 Genomic DNA .....	27
2.1.1.7 Instruments .....	28
2.1.1.8 Molecular biology kits .....	29
2.1.2 Methods .....	29
2.1.2.1 Determination of nucleic acid concentration .....	29
2.1.2.2 Polymerase chain reaction (PCR) .....	29

2.1.2.3 Agarose gel electrophoresis.....	30
2.1.2.4 Preparation of chemically competent <i>E. coli</i> cells .....	31
2.1.2.5 Molecular cloning .....	31
2.1.2.6 Site-directed mutagenesis.....	32
2.1.2.7 Transformation.....	32
2.1.2.8 DNA sequencing .....	33
2.1.3 Plasmids.....	34
2.2 Protein expression and purification .....	37
2.2.1 Materials.....	37
2.2.1.1 Instruments .....	37
2.2.1.2 Chromatography systems and columns .....	38
2.2.1.3 Buffers.....	38
2.2.1.4 Enzymes and reagents .....	39
2.2.1.5 Protein expression cells.....	39
2.2.1.6 Cell media for protein expression .....	40
2.2.2 Methods .....	41
2.2.2.1 Protein expression using bacterial cells .....	41
2.2.2.2 Protein expression using insect cell system.....	42
2.2.2.3 SDS-PAGE .....	44
2.2.2.4 Protein concentration determination .....	45
2.2.2.5 Western blot.....	45
2.2.2.6 Cell harvest and lysis .....	46
2.2.2.7 Protein purification .....	46
2.3 Protein characterization .....	48
2.3.1 Thermal shift assay.....	48
2.3.2 Multi-angle light scattering combined with size exclusion chromatography ....	48
2.3.3 Limited proteolysis.....	49
2.3.4 Electrophoretic mobility shift assay (EMSA) .....	49
2.3.5 GST pull-down.....	49
2.3.6 Cobalt pull-down .....	50
2.3.7 Analytical gel filtration analysis .....	50
2.3.8 Co-sedimentation .....	50
2.3.9 Crosslinking.....	50
2.3.10 Characterization of ATPase activity .....	51
2.3.11 Characterization of helicase activity.....	53
2.3.12 RNA sequences .....	54
2.4 Crystallization and data collection .....	54

2.4.1 Initial crystallization screens .....	54
2.4.2 Optimization of crystals .....	55
2.4.3 Data collection .....	55
2.5 Structure determination and refinement .....	56
<b>3 Results.....</b>	<b>57</b>
3.1 Eukaryotic translation initiation factor 3 (eIF3) .....	57
3.1.1 Preparation of eIF3 subunits.....	59
3.1.1.1 Purification of eIF3a, b, c, g and i .....	59
3.1.1.2 Purification of eIF3j .....	60
3.1.1.3 Formation of eIF3b-RRM/eIF3j complex.....	61
3.1.2 Interaction studies and complexes formation.....	62
3.1.2.1 Reconstitution of eIF3 functional core complex.....	62
3.1.2.2 Mapping the interacting regions in eIF3b-g-i subcomplex .....	63
3.1.3 Crystallization and structure determination of eIF3 subunits and subcomplexes .....	66
3.1.3.1 Crystals of eIF3j .....	66
3.1.3.2 Crystallization of eIF3b <sup>74-170</sup> /eIF3j subcomplex .....	67
3.1.3.3 Crystallization and structure determination of eIF3b .....	67
3.1.3.4 Crystallization of eIF3b-g-i complex .....	69
3.1.4 Structural analysis of eIF3b-WD40 .....	72
3.1.4.1 Overall structure .....	72
3.1.4.2 Sequence and structural conservation of the WD40 repeats .....	73
3.1.4.3 Conservation of eIF3b orthologs.....	74
3.1.5 Localization of eIF3b on the ribosome .....	77
3.1.6 Interaction of eIF3b with the ribosome.....	80
3.1.6.1 Interaction with the 40S ribosomal subunit.....	80
3.1.6.2 Interaction with the ribosomal protein S9e .....	81
3.1.6.3 Formation of rpS9e-eIF3b-g-i complex.....	81
3.2 Eukaryotic translation initiation factor 4 (eIF4) .....	83
3.2.1 Preparation and crystallization of the full-length eIF4F complex .....	83
3.2.1.1 Purification of eIF4A, eIF4B and eIF4E .....	83
3.2.1.2 Expression and purification of full-length eIF4G .....	84
3.2.1.3 Preparation of the full-length eIF4F complex .....	87
3.2.1.4 Crystallization of the eIF4F complex .....	88
3.2.2 Preparation of minimal functional eIF4F complex .....	89
3.2.2.1 Activities of eIF4A stimulated by eIF4G and eIF4B.....	90

3.2.2.2 RNA-binding ability of eIF4A and eIF4G variants .....	92
3.2.2.3 Interaction test between eIF4A and eIF4B.....	95
3.2.3 Monitoring the conformational change of eIF4A by crosslinking.....	97
3.2.3.1 Binding of eIF4G induces conformational changes in eIF4A.....	99
3.2.3.2 Conformational changes of eIF4A upon ADP binding.....	100
3.2.3.3 Conformational changes in eIF4G-RBR2 upon RNA binding.....	102
<b>4 Discussion .....</b>	<b>105</b>
4.1 Reconstitution of <i>C. thermophilum</i> eIF3 core complex.....	106
4.2 The nine-bladed $\beta$ -propeller fold of eIF3b-WD40 .....	107
4.3 Localization of eIF3 on the 40S ribosomal subunit.....	108
4.4 RNA substrate of eIF4A.....	110
4.5 Regulation of eIF4A activity by eIF4G and eIF4B.....	111
<b>References .....</b>	<b>115</b>
<b>Abbreviations .....</b>	<b>130</b>
<b>Acknowledgements .....</b>	<b>133</b>
<b>Curriculum vitae .....</b>	<b>135</b>



## 1

## Introduction

## 1.1 A overview of protein synthesis

Within living organisms, proteins play crucial roles in catalyzing metabolic reactions, transporting molecules, responding to environmental stimuli, and maintaining cellular structure *etc.* Therefore, protein synthesis is one of the most important cellular processes. Translation of messenger RNAs (mRNAs) into proteins consists of three subprocesses: initiation, elongation and termination. In eukaryotes, translation initiates with the assembly of the 43S preinitiation complex (PIC), which contains the small ribosomal subunit carrying the initiator transfer RNA (tRNA) in its P-site and several initiation factors (IFs). Subsequently, the 43S PIC binds to the 5' cap-proximal region of the mRNA and scans the whole 5' untranslated region (5' UTR) for the AUG start codon. After the start codon identification, the large ribosomal subunit joins and the translation elongation competent ribosome is formed. To efficient and accurate assemble this 80S ribosome with initiator tRNA complemented with the mRNA start codon in its P-site, up to a dozen of initiation factors are required. In the decoding stage of the elongation cycle, an aminoacyl-tRNA coupled with an elongation factor (i.e. eEF1A) binds to the mRNA codon in the A-site through its complementary anticodon. Then, the peptide chain on the P-site tRNA is transferred to the aminoacyl group in the A site tRNA. In the next steps, the ribosome moves towards the 3' end of the mRNA by one codon

with the help of eEF2, resulting in the translocation of the P-site deacylated tRNA and A-site peptidyl tRNA to the E-site and P-site, respectively. This elongation reaction cycle continues until the stop codon is reached. The termination codon is recognized by translation release factors (RFs). On binding to the corresponding stop codon, release factors induce the release of the newly synthesized polypeptide. Ribosome subunits as well as translation auxiliary protein factors are recycled for further rounds of translation.

## 1.2 Eukaryotic translation initiation and the role of initiation factors

Translation initiation in eukaryotes requires the assistance of at least 12 different accessory protein factors, the so-called eukaryotic initiation factors (eIFs), which are summarized in Table 1.1. As outlined in Figure 1.1, the eukaryotic translation initiation pathway via scanning mechanism can be divided into several steps.

**Formation of 43S preinitiation complex** Eukaryotic translation initiation commences with the binding of the initiator tRNA ( $\text{Met-tRNA}_i^{\text{Met}}$ ), in a complex with GTP-bound eIF2, to the P-site of the 40S ribosomal subunit. This step is promoted by eIF3 complex, eIF5, 1 and 1A, which further form the 43S preinitiation complex (43S PIC) with the eIF2·GTP·Met-tRNA<sub>i</sub><sup>Met</sup> ternary complex (TC) and the 40S subunit [1-3]. More and more evidence has emerged that eIF1, 3, 5 and the TC interact with each other prior to associating with the 40S subunit, forming a multi-factor complex (MFC) [4-6]. Interactions among the MFC are known to enhance the assembly and stability of the 43S PIC [4,6].

**Recruitment of 43S PIC to mRNA** The 43S PIC is recruited to the 5' end region of the mRNA, in which the heterotrimeric complex eIF4F plays important roles [7,8]. eIF4F consists of the mRNA cap recognition protein eIF4E, the DEAD-box RNA helicase eIF4A as well as the scaffolding protein eIF4G. eIF4A is known to unwind duplex structures in the 5' UTR of the mRNA to facilitate the recruitment of the 43S PIC and subsequent scanning. eIF4G contains binding sites

**Table 1.1:** Eukaryotic initiation factors

Name	Number & mass (kDa) of subunits	Functions
eIF1	1 (13)	Ensures to select the correct initiation codon; facilitates ribosomal scanning; prevents premature hydrolysis of eIF2-bound GTP and $P_i$ release; promotes the recruitment of eIF2·GTP·Met-tRNA <sup>Met</sup> ternary complex (TC) to the 40S subunit
eIF1A	1 (17)	Promotes binding of the TC to the 40S subunit; mediates ribosomal scanning and start codon recognition
eIF2	3 (36, 38 and 51)	Binds to GTP and the initiator tRNA (Met-tRNA <sup>Met</sup> ), stimulating the recruitment of Met-tRNA <sup>Met</sup> to the 40S subunit
eIF3	13 (800 total)	Interacts with the 40S subunit and initiation factors 1, 4 and 5; promotes the recruitment of the TC to the 40S subunit; involved in attachment of the 43S preinitiation complex (PIC) to mRNA and subsequent scanning; and participates in post-termination of translation and prevents ribosomal subunits re-joining
eIF4E	1 (24)	Binds to the m <sup>7</sup> G ‘cap’ structure of mRNA
eIF4A	1 (46)	DEAD-box RNA helicase and also possesses ATPase activity, unwinding the secondary structures in the 5′ UTR of the mRNA
eIF4G	1 (176)	Multi-scaffold protein, interacting with eIF4E, eIF4A, PABP, eIF3, MNK1 and RNA; also possesses the ability to enhance the activities of eIF4A
eIF4B	1 (69)	Enhances the activities of eIF4A
eIF4H	1 (27)	Stimulates the activities of eIF4A and is homologous to the N-terminal domain of eIF4B
eIF5	1 (49)	The GTPase-activating protein (GAP) of GTP-bound eIF2, inducing hydrolysis of eIF2-bound GTP
eIF5B	1 (139)	A ribosome-dependent GTPase that mediates the joining of the 60S subunit
eIF2B	5 (34, 39, 50, 60 and 80)	A guanine nucleotide exchange factor (GEF), promoting change of GDP to GTP on eIF2

The molecular mass, which gives a rough idea about the size of the proteins, was calculated from human initiation factors.

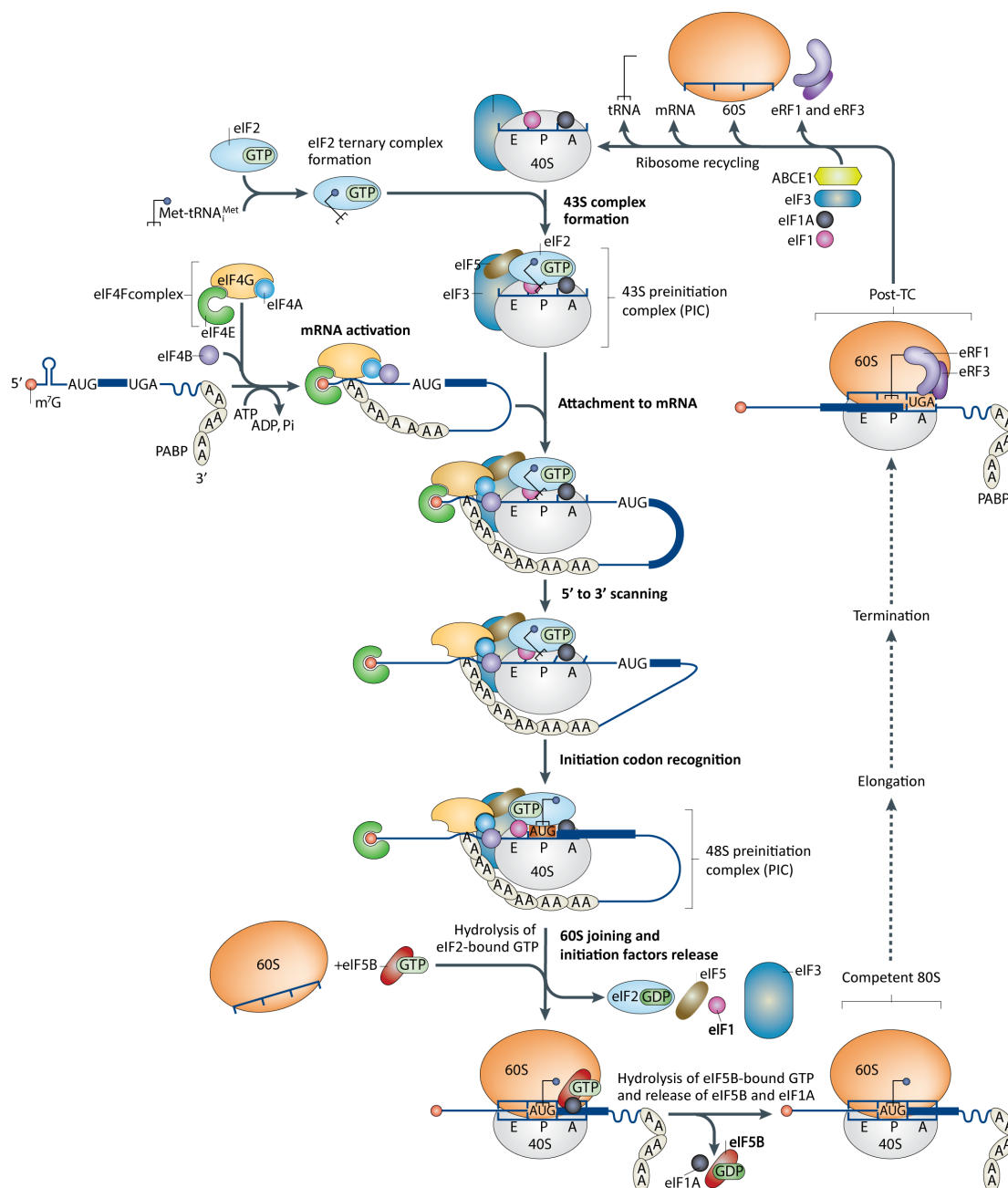
for eIF4E, eIF4A, poly(A)-binding protein (PABP), RNA, and in mammals for eIF3 as well as the eIF4E kinase MNK1 [9-13]. The binding sites for eIF4E, PABP and RNA in eIF4G enable the formation of the stable circular cap·eIF4E·eIF4G·PABP·

poly(A)-tail messenger ribonucleoprotein particle (mRNP) [8]. The eIF4G-eIF3 protein interaction (eIF4G-eIF5 in yeast) is thought to bridge the activated mRNP and the 43S PIC [12,14]. Notably, eIF3 complex and the 40S itself also possess RNA-binding activity, which could contribute to the attachment of the 43S PIC to the mRNA [15,16].

**Scanning of the mRNA 5' UTR** Once attached to the vicinity of the mRNA cap, the 43S PIC scans the downstream of the cap for the correct AUG start codon. The ribosomal movement along the mRNA requires unwinding of duplexes in the 5' UTR and a scanning-competent conformation of the 40S itself. This ribosomal conformation is induced by two small initiation factors, eIF1 and eIF1A [17,18]. eIF1 binds to the 40S subunit close to the P-site and strictly monitors the scanning for an AUG codon in an optimal sequence context (GCC(A/G)CCAAUGG) [19,20]. The universally conserved eIF1A is situated at the ribosomal A-site with its compact oligonucleotide binding domain (OB domain). This ribosomal situation of eIF1A prevents tRNAs from accessing the A-site during initiation phase. In addition, eIF1A bridges the head and the body of the 40S subunit through an  $\alpha$ -helix and the long unstructured N-terminal tail (NTT) [21,22]. Connection between the ribosomal head and body generates an active 40S subunit with closed mRNA channel around the A-site, which would hold the incoming mRNA and promote the scanning.

The single-stranded mRNA regions are generated by ATP-dependent RNA helicases and their auxiliary factors. eIF4A, eIF4G and eIF4B as well as ATP are needed even in the scanning of 5' UTRs that contains weak secondary structures [18,23]. In addition to eIF4A, other RNA helicases (*e.g.* human DHX29, DDX3, yeast Ded1) have been shown to be involved in scanning, especially in long and highly structured 5' UTRs [24-27].

**Start codon selection** eIF1 plays a crucial role in stringent recognition of the correct start codon that is usually the first AUG triplet in a suitable context (GCC(A/G)CCAAUGG) [28,29]. eIF1 is positioned on the top of the 18S rRNA helix 44 and would sterically clash with the anti-codon stem loop (ASL) in the P-site and with the D-loop of the initiator tRNA [20,30]. The special localization and high affinity to the TC bound 40S subunit (with a  $K_d \sim 1.4$  nM [17]) of eIF1 ensure the



**Figure 1.1:** Overview of eukaryotic translation initiation via scanning. The process of initiation begins with the assembly of the 43S preinitiation complex (PIC), which contains the 40S ribosomal subunit with initiator tRNA in its P-site and initiation factors 1, 1A, 2, 3 and 5. mRNA is activated by the binding of eIF4E to the m<sup>7</sup>G cap of the mRNA, accompanied with unwinding the duplexes in the 5' UTR by eIF4A, and interaction of PABP with the mRNA poly(A) tail. The ATP-dependent helicase activity of eIF4A is facilitated by eIF4B and eIF4G. The 43S PIC attaches to the 5' end of the activated mRNA and scans the mRNA 5' UTR of which the secondary structures are melted by eIF4A and other helicases. Once the AUG start codon base pairs with the anticodon of Met-tRNA<sub>i</sub><sup>Met</sup>, hydrolysis of eIF2-bound GTP triggers dissociation of initiation factors from the 40S subunit. Subsequently, eIF2-GDP is released from the initiator tRNA, and GTP bound eIF5B takes part in the joining of the 60S subunit, forming the elongation competent 80S ribosome. This figure is modified from Jackson *et al.*, 2010 [31].

discrimination against non-AUG triplets as well as AUG in non-optimal context, and trigger the dissociation of the ribosomal complex aberrantly assembled at such triplets. Establishment of the AUG codon-anticodon base pairing, which significantly increases the affinity of TC to the 40S subunit ( $K_d < 1$  nM), displaces eIF1 from near the P-site [32,33]. Subsequently, the 40S rotates its head domain and locks the initiator tRNA in a stable P-site state. The NTT and CTT of eIF1A both reach into the ribosomal P-site and would also participate in the start codon selection [34].

**Dissociation of initiation factors and joining of the 60S ribosomal subunit** The displacement of eIF1 (probably partially) upon AUG recognition triggers the irreversible eIF5-stimulated GTP hydrolysis by eIF2 in the TC [35]. GTP hydrolysis induces a conformational change of eIF2 and reduces its affinity for the initiator tRNA [36]. The Met-tRNA<sub>i</sub><sup>Met</sup> unbound eIF2-GDP eventually dissociates from the 40S subunit and is recharged with GTP for another cycle of initiation. Recharging eIF2 with GTP is mediated by its guanine nucleotide exchange factor (GEF) eIF2B [37]. Joining of the 60S subunit and complete dissociation of initiation factors bound to the 40S interface from the 40S subunit are mediated by the ribosome-dependent GTPase eIF5B [38,39]. eIF5B is situated in the intersubunit cleft and extensively contacts with both ribosomal subunits and the initiator tRNA, which enables it to promote the formation of the elongation-competent 80S ribosome [40,41]. The release of eIF5B is mediated through the GTP hydrolysis by its own. Those initiation factors that bind to the solvent exposed side of the 40S subunit (*e.g.* eIF4G and eIF3) could delay their dissociation from the 40S ribosomal subunit [31].

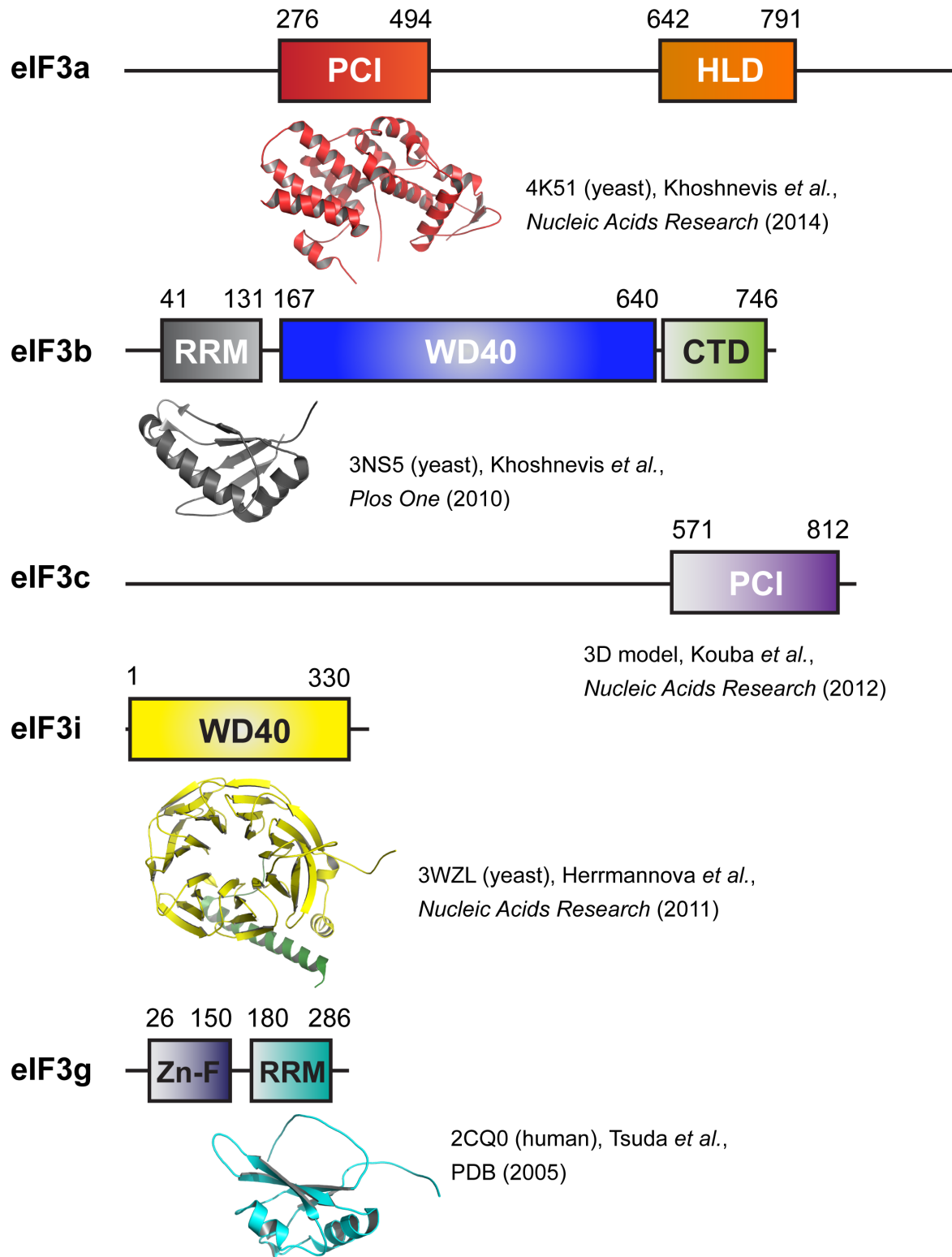
## 1.3 Eukaryotic translation initiation factor 3

### 1.3.1 Composition and interaction network of eIF3

Among the dozen of initiation factors, eIF3 is the largest multi-subunit complex. In mammals, it consists of 13 nonidentical subunits (designated eIF3a-m). However, not all but only five of eIF3 subunits (a/Tig32, b/Prt1, c/Nip1, g/Tif35 and i/Tif34) are conserved in all organisms and are crucial for translation *in vivo* [42-44]. These five subunits form a functional ‘core’ of eIF3 and thus represent the simplest model for structural and functional investigations. In this thesis, a main focus will be put on the eIF3 core, of which the subunits are outlined in more details in Figure 1.2.

Consistent with its complex composition, eIF3 generates an elaborate subunit-subunit interaction web. Within the core eIF3, the b subunit (eIF3b) serves as the major scaffold protein, providing interaction platforms for other subunits [4,45]. The ~90 kDa eIF3b contains an N-terminal structurally canonical RNA recognition motif (RRM), a middle WD40 domain, which has been predicted to fold into two tandem  $\beta$ -propellers, and a following C-terminal domain (CTD) [46,47]. eIF3b is able to associate with the CTD of eIF3a and the N-terminal domain (NTD) of eIF3j through its NTD, while eIF3b-CTD is required for the interaction of eIF3g and eIF3i subunits [4,45,47,48]. The pivotal role of eIF3b as a versatile interaction partner is further emphasized by studies in mammalian cells, which indicate a function in tumour growth, identifying eIF3b as a potential therapeutic drug target [49-51]. So far, the structural information about eIF3b is limited to the RRM domain (~100 amino acids) and a ~40 residue  $\alpha$ -helix of the CTD [47,48,52]. In contrast, neither functional nor structural information is available for the central WD40 repeats, which form ~65% of the protein and are highly conserved from yeast to humans.

eIF3a is the largest subunit of eIF3 and consists of ~1000-1400 amino acid residues from yeast to humans. Up to now, only a PCI domain of eIF3a is well defined [53]. PCI domains are known to play roles in the 26S proteasome lid, COP9 signalosome (CSN) and eukaryotic initiation factor 3 by serving as a principal



**Figure 1.2:** Overview of eIF3 functional core subunits, including structures of domains in cartoon representation. The green  $\alpha$ -helix bound to eIF3i (yellow) is from eIF3b-CTD (residues 700-738). The boundaries of the domains are indicated for *S. cerevisia* eIF3a and c, and for *C. thermophilum* eIF3b, i and g.



scaffold [54]. The PCI domain of eIF3a is located at the NTD and is proposed to interact with the ribosomal protein S0 (rpS0) located in the vicinity of the mRNA exit pore [1,55]. The structure of eIF3a PCI domain was recently solved by Khoshnevis *et al.* [53]. They also demonstrated that this domain mediates the binding of eIF3a to eIF3c and is indispensable (but not sufficient) for interacting with RNA. In addition, eIF3a contains a region that is related in sequence to eIF3j, called HCR1 (a yeast homolog of human eIF3j) like domain (HLD). The HLD domain of eIF3a is thought to associate with eIF3b [4]. The eIF3a-CTD was reported to interact with rpS2e, rpS3e, as well as helices 16-18 of the 18S rRNA, all constituents of the mRNA entry channel [1,15,30].

eIF3c is a ~100 kDa protein and contains a PCI domain at the extreme CTD, which was reported to directly interact with the seven-bladed WD40  $\beta$ -propeller ribosomal protein ASC1, a yeast homolog of human RACK1 [56]. The PCI domain of eIF3c is the first example that exhibits strong but unspecific RNA binding ability, apart from the interaction with protein [56]. The 3 dimensional (3D) structure of eIF3c's PCI domain was modeled *in silico* [56]. The NTD of eIF3c is thought to interact with eIF3a, eIF1 and eIF5 [57].

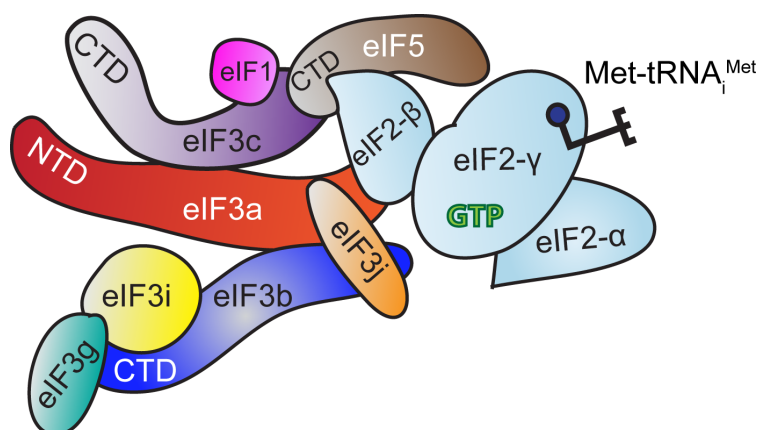
eIF3g and eIF3i are two small proteins with molecular mass of ~35 kDa each. eIF3g contains a predicted Zn-finger domain in its NTD and an RRM in its CTD. eIF3g-NTD has been reported to mediate the association with eIF3b and eIF3i, while the C-terminal RRM is necessary for interacting with 18S rRNA of 40S [4,44]. Additionally, eIF3g has been reported to bind to rpS3e and rpS20e [58]. The structure of eIF3g-RRM was determined by NMR and reveals a canonical  $\beta\alpha\beta\beta\alpha\beta$  fold (PDB code: 2CQ0). eIF3i adopts a seven-bladed WD40  $\beta$ -propeller architecture and interacts with eIF3b-CTD via the bottom side of the propeller [48]. eIF3i is also able to bind eIF3g, however, the binding site is not known yet [59].

### 1.3.2 The role of eIF3 in translation initiation

The structural complexity provides a prerequisite for crucial involvement of eIF3 in almost all translation initiation steps.

**Promoting the assembly of 43S PIC** eIF3 binds to eIF2 $\beta$ -NTD, and thus to eIF2-GTP-Met-tRNA<sub>i</sub><sup>Met</sup> ternary complex, through its subunit a (eIF3a) [42]. In addition, it interacts with eIF1 and eIF5-CTD via the c subunit (eIF3c) [57,60]. Yeast, plant and human eIF3, 1, 5 and the TC have been demonstrated in succession to form a multifactor complex (MFC) (free of the 40S ribosomal subunit), which can be isolated from whole cell extracts (Figure 1.3) [5,6,57]. The formation of MFC could facilitate the cooperative binding of these initiation factors to the 40S subunit, generating the 43S preinitiation complex (PIC). eIF3 plays an important role in the assembly of the 43S PIC. Besides interacting with eIF1, 5 and the TC, eIF3 makes extensive contacts with the small ribosomal subunit. The NTD of eIF3a binds to the ribosomal protein rpS0 and rpS1e, while eIF3a-CTD forms intermolecular bridges with rpS2e, rpS3e as well as the helices 16-18 of 18S RNA [1,15,55,61,62]. The c subunit also directly interacts with ribosome via the association with RACK1 [56]. The interaction between eIF3b and the 40S subunit is mediated by another eIF3 subunit, eIF3j [47]. The small subunits g and i also function in recruiting initiation factors to the ribosome. eIF3g is anchored in the 'beak' region of 40S, involving ribosomal proteins rpS3e and rpS20e [58]. A mutant with deletion of eIF3g and i showed lower 40S binding affinity. The ribosomal association of eIF3 has also been determined by cryo-electron microscopy (cryo-EM) reconstitutions [61,63]. Consistent with biochemical investigations, the cryo-EM structures of eIF3 bound to the 40S ribosomal subunit reveal eIF3 on the solvent-exposed side of the 40S.

**Functioning in mRNA attachment** Mammalian eIF3 facilitates the recruitment of mRNA to the 40S subunit *in vitro* [42]. A mutation in yeast eIF3b as well as mutants with deletion of eIF3a and eIF3b impair mRNA binding to the 40S subunit *in vivo* [64,65]. Direct interaction of mammalian eIF3 with eIF4G, a scaffold subunit of the mRNA activation complex eIF4F, would enable eIF3 to function as a 'bridge' between the 40S subunit and the eIF4F-mRNA complex [14,66]. However, this eIF3-eIF4G interaction is not universally conserved. For example, yeast eIF4G lacks the eIF3-binding domain and thus is unable to interact with eIF3, which indicates that eIF3 can facilitate mRNA binding to the 40S independently of eIF4G.



**Figure 1.3:** Schematic representation of the multifactor complex (MFC). Depicted interactions are based on biochemical investigations.

Indeed, a yeast mutant lacking eIF4G led to accumulation, rather than depletion, of mRNAs in the native 48S PICs [65]. eIF4B might also serve as an adaptor between mRNA and eIF3, since mammalian and yeast eIF4B have been demonstrated to interact with eIF3a and eIF3g, respectively [67,68]. In addition, recent evidences showed that eIF3 could directly bind to mRNAs and promote their attachment to and/or stabilize their binding on the 40S subunit [42,53,56].

**Involvement in scanning and AUG start codon recognition.** eIF1 is the gatekeeper of the start codon identification, while eIF5 triggers the GTP hydrolysis by TC upon the start codon recognition. eIF3 is able to cooperatively interact with eIF1 and 5 through its c subunit's NTD [69]. Mutations in yeast eIF3c-NTD stimulate translation initiation at UUG codons *in vivo* [69]. The major scaffold subunit of eIF3, eIF3b, is believed to be involved in this process as well. Mutation within eIF3b-RRM increases leaky scanning, which indicates a role of eIF3b in AUG initiation selection. Strong leaky scanning is also observed if the NTD of eIF3j is deleted [42]. All these findings strongly suggest that the multi-subunit eIF3 remains bound to the ribosome after mRNA recruitment and participates in scanning and start codon recognition.

In addition to playing role in translation initiation, eIF3 is also involved in post-termination of proteins synthesis, promoting dissociation and preventing re-joining of ribosomal subunits. For comprehensive review see [42].

## 1.4 Eukaryotic translation initiation factor 4

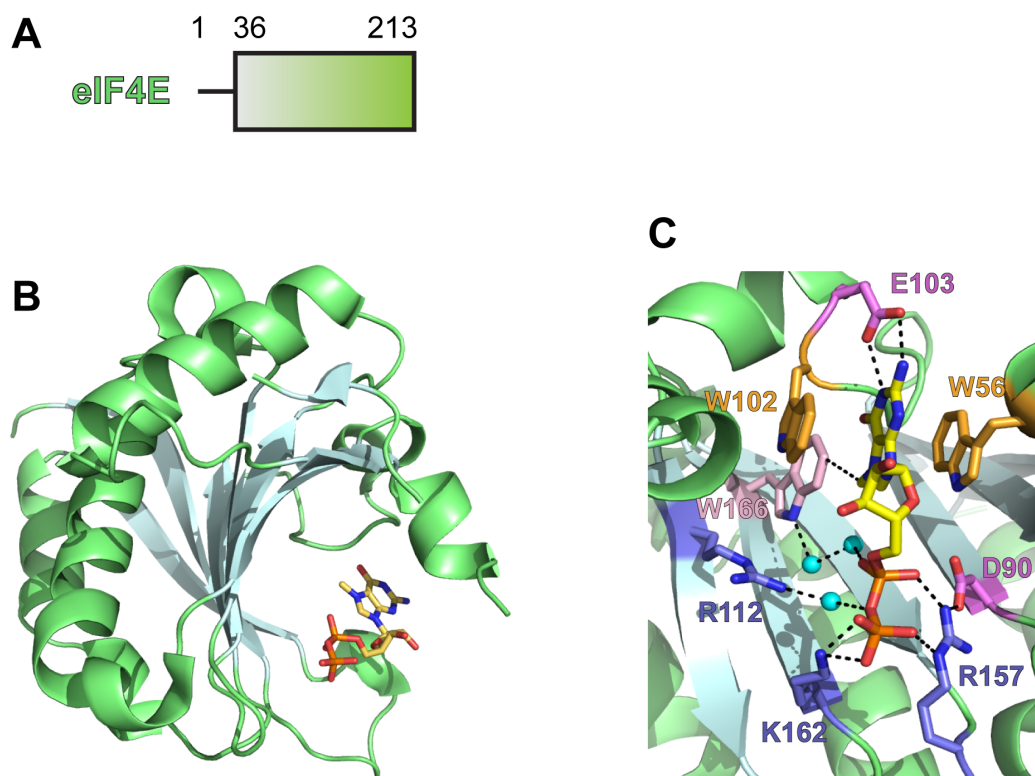
The multi-subunit eukaryotic translation initiation factor 4 cooperatively functions in the activation of the mRNA to be translated and the subsequent scanning. It includes the 7-methyl guanosine triphosphate ( $m^7GTP$ ) cap recognition protein eIF4E, the ATP-dependent RNA helicase eIF4A, eIF4A auxiliary proteins eIF4B and eIF4H (in mammalian cells), as well as the large scaffold protein eIF4G. eIF4E, eIF4G and eIF4A form the heterotrimeric complex eIF4F.

### 1.4.1 eIF4E

eIF4E is a ~24 kDa protein and was firstly identified in 1978 by its ability to specifically cross-link to the 5'-terminal cap ( $m^7GTP$ ) in the mRNAs [70]. It binds the cap so strongly that it can be purified to apparent homogeneity by one affinity chromatography on an  $m^7GDP$ -coupled matrix [71]. eIF4E recognizes the cap structure of the mRNA and thus is crucial for cap-dependent translation. Depletion of eIF4E significantly reduces translation of capped mRNA and this reduction can be restored by addition of purified eIF4E [72]. eIF4E is universally conserved (with at least 30% sequence identity) and mammalian eIF4E can rescue the yeast lethality caused by mutation of yeast eIF4E gene, *CDC33* [73].

**eIF4E Structure** The three-dimensional structure of eIF4E has been extensively investigated by X-ray crystallography and nuclear magnetic resonance (NMR). So far, up to 30 eIF4E structures from different organisms either in *apo*,  $m^7GDP$ - or  $m^7GTP$ -bound states or in complex with inhibitors or eIF4E-binding proteins are present in PDB database. These proteins adopt the same overall baseball glove-like architecture, consisting of one single  $\alpha/\beta$  domain (Figure 1.4) [74,75]. The eight antiparallel  $\beta$ -strands form the 'palm' side of the structure, while the  $\alpha$ -helices situate as the back convex surface. eIF4E orthologs across yeast to human utilize almost identical modes of cap binding. The  $m^7GTP$  cap or the cap analog  $m^7GDP$  is captured on the concave palm surface of eIF4E by  $\pi$ - $\pi$  stacking and cation- $\pi$  interactions, hydrogen bonds, salt bridges as well as a van der Waals

contact (Figure 1.4). All of the residues that are involved in the cap-binding are universally well conserved.



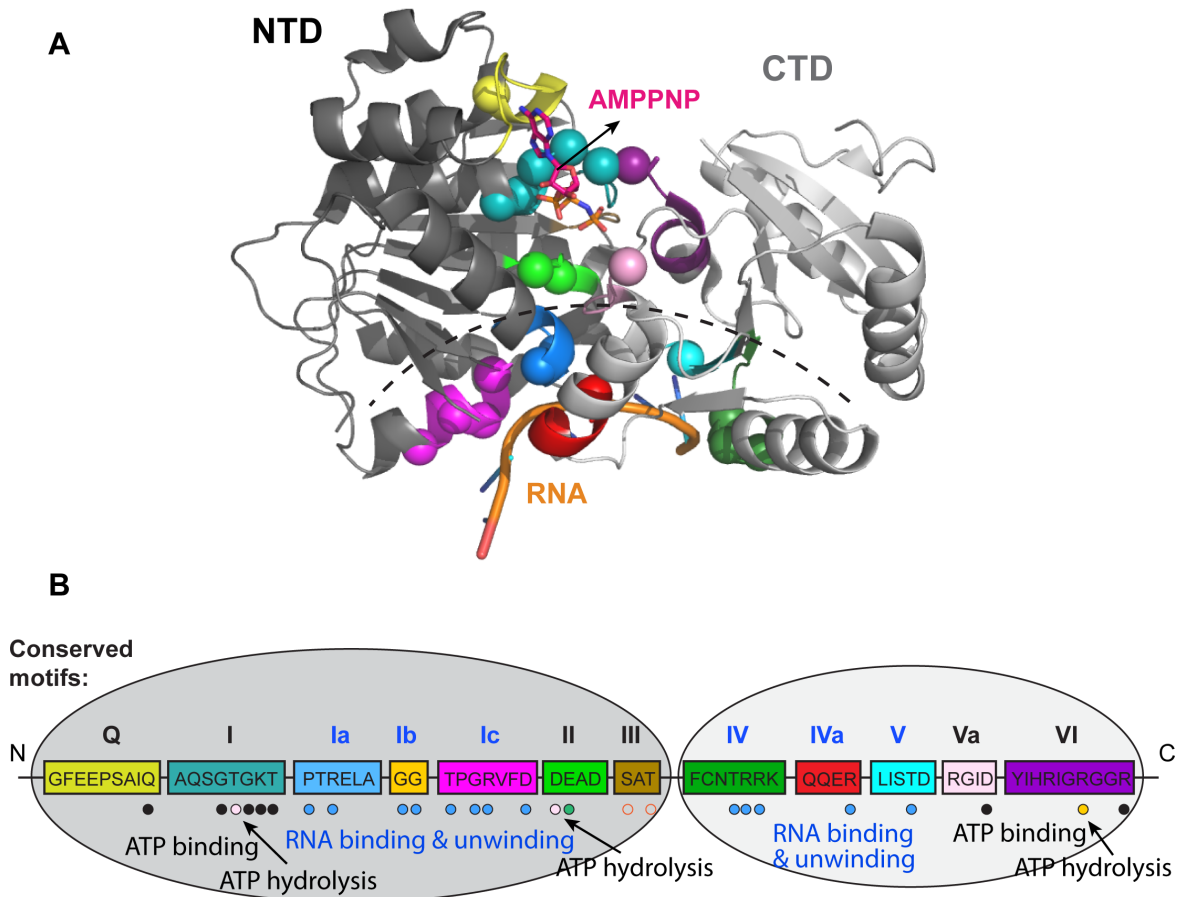
**Figure 1.4:** Structure of eIF4E (green and light cyan) bound to the cap analog m<sup>7</sup>GDP (yellow and orange) (PDB code: 1EJ1). (A) Schematic domain organization of eIF4E. (B) Overall structure. The cap is held in the concave side. (C) Magnified view of the cap-binding area. The amino acid side chains involved in the interaction are highlighted and shown in sticks.

### 1.4.2 eIF4A

The initiation factor eIF4A (~45 kDa) is thought to be the ‘godfather’ of the DEAD-box family of ATP-dependent RNA helicases [76,77]. This family is named after its conserved Asp-Glu-Ala-Asp (DEAD) sequence motif in which the first Asp coordinates a magnesium ion and the Glu functions in activation of the catalytic water molecule to hydrolyze the  $\gamma$  phosphate of the ATP [78]. In addition to exhibiting the RNA unwinding activity, DEAD-box helicases also possess RNA-dependent ATPase activity [79]. The ATP and RNA binding, ATP hydrolysis

as well as RNA unwinding motifs are highly conserved in all helicases from this family. These conserved motifs are arranged into two recA-like domains, which, in many other DEAD-box proteins, are flanked by N- and/or C-terminal domains that would facilitate substrate binding [80-82]. eIF4A contains only the two core recA-like domains, lacking any accessory domain. However, other initiation factors that are able to enhance the activities of eIF4A (e.g. eIF4G and eIF4B) may play some of the same roles.

The highly conserved functional motifs in DEAD-box containing helicases are outlined in Figure 1.5. Their relative sequential positions (shown for *Saccharomyces*

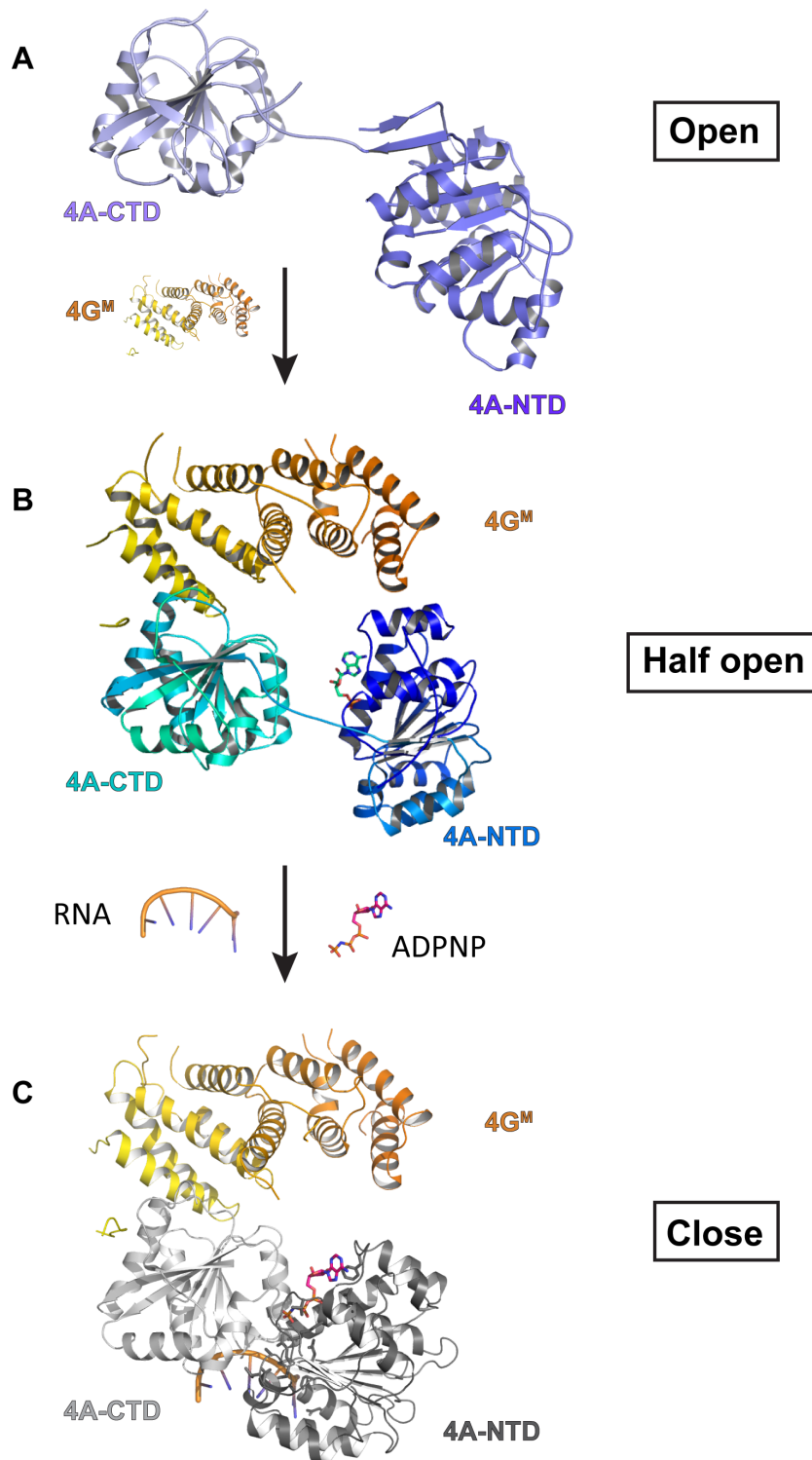


**Figure 1.5:** Conserved functional DEAD-box helicase motifs in *sceIF4A*. (A) The 3D structure of eIF4A bound RNA and AMPPNP is modeled based on the crystal structure of eIF4A III (2HYI). (B) Schematic functional motif organization of eIF4A. Residues directly involved in ATP and RNA binding, ATP hydrolysis as well as RNA unwinding are depicted as spheres in (A) and small circles in (B), respectively. Small circles in (B): blue: RNA binding and unwinding; black: ATP binding; pink: coordinating a magnesium ion; green: activating the catalytic water; orange: stabilizing the transition state

*cerevisiae* eIF4A) and localizations on a 3D structural model of *sce*IF4A are shown in the bottom and top panel, respectively. The 3D model is made based on the crystal structure of the related protein eIF4A III (PDB code: 2HYI), which plays role in RNA processing, but not in translation initiation [83,84]. The bound RNA (orange) and ATP analog AMPPNP (hotpink) in this model are taken directly from eIF4A III structure. The N-terminal domain of eIF4A (also other DEAD-box proteins) harbors the eponymous DEAD segment and most of motifs involved in the ATP binding and hydrolysis, including Q, I, II, and III. eIF4A-NTD also contains RNA binding and unwinding motifs Ia, Ib and Ic. The smaller CTD contains motifs IV, IVa and V, which are implicated in RNA binding or in helicase activity, and motifs VI responsible for binding ATP [80]. Although these conserved functioning fragments are distributed throughout the whole eIF4A sequence, they are all localized at the interface between the two RecA-like domains as revealed by the modeled structure.

The activities of DEAD-box helicases are based on conformational changes. In the absence of ATP and RNA, the two recA-like domains are far apart, forming an ‘open’ conformation. By contrast, when ATP and RNA are present, the two core domains move together, assembling the ‘closed’ conformation, to cooperatively hydrolyze ATP and to unwind the double-stranded RNA. So far, there is no crystal structure of eIF4A in its closed form. But the structures of yeast eIF4A in the *apo* state (Figure 1.6 A) and in complex with a HEAT repeat domain of eIF4G (eIF4G<sup>572-853</sup>) (Figure 1.6 B) have been crystallographically determined [85,86]. Compared to the known structures of other DEAD-box proteins in their closed conformations with bound RNA and nucleotide (e.g. human eIF4A III and *drosophila* VASA), the structure of yeast eIF4A/eIF4G<sup>572-853</sup> shows a ‘half open’ conformation of eIF4A, in which the functional motifs are pre-organized for activation.

eIF4A itself shows very low ATPase and helicase activities. However, the ATP hydrolysis and RNA unwinding efficiency are significantly stimulated by eIF4B, eIF4H and eIF4G [79,87,88]. It has been proposed that these auxiliary proteins enhance the activities of eIF4A by promoting a closed and more active state of the helicase (Figure 1.6 C). However, the underlying mechanism is still elusive.

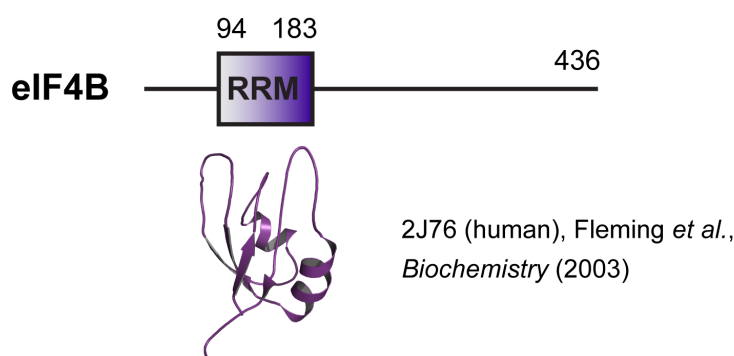


**Figure 1.6:** Conformational changes of eIF4A. Crystal structures of yeast eIF4A in its open (1FUU) (A) and half open (2VSO) (B) conformations. A modeled structure of the closed state of *S. cerevisiae* eIF4A (based on human eIF4A III (2HYI)) is provided (C). During modeling, eIF4A-CTD, which is the primary eIF4G-interacting domain, is fixed, while the NTD of eIF4A is moved and superimposed onto eIF4A III-NTD.



### 1.4.3 eIF4B and eIF4H

eIF4B (~50 kDa) plays an auxiliary, but not essential role in translation initiation because the 48S preinitiation complex (48S PIC) can be assembled in its absence and a yeast strain with deletion of eIF4B gene, *TIF3*, is viable [89,90]. Experiments have shown that eIF4B enhances the duplex unwinding activity and ATPase activity of eIF4A [87,88]. However, the underlying mechanism of the enhancement is unclear. The sequences of eIF4B orthologs differ a lot, with identities of 20-30% across species. Mammalian eIF4B forms a homodimer via the middle seven repeats that are rich in aspartic acid, arginine, tyrosine and glycine rich (termed DRYG domain), which also mediate the interaction of eIF4B with eIF3 [91,92]. The DRYG domain is missing in yeast eIF4B. In addition, eIF4B possesses two RNA-binding domains. The canonical RNA recognition motif (RRM) localized in the N-terminus is thought to bind to the 18S rRNA, whereas the C-terminal basic region rich in arginine and serine residues strongly interacts with any RNAs in a sequence-independent manner [67,93,94]. Up to now, a solution structure of the RRM domain, which employs a typical  $\beta\alpha\beta\beta\alpha\beta$  fold, is the only known structural information of eIF4B (PDB code: 2J76) [95].



**Figure 1.7:** Solution structure of human eIF4B-RRM domain, displaying a canonical  $\beta\alpha\beta\beta\alpha\beta$  folding.

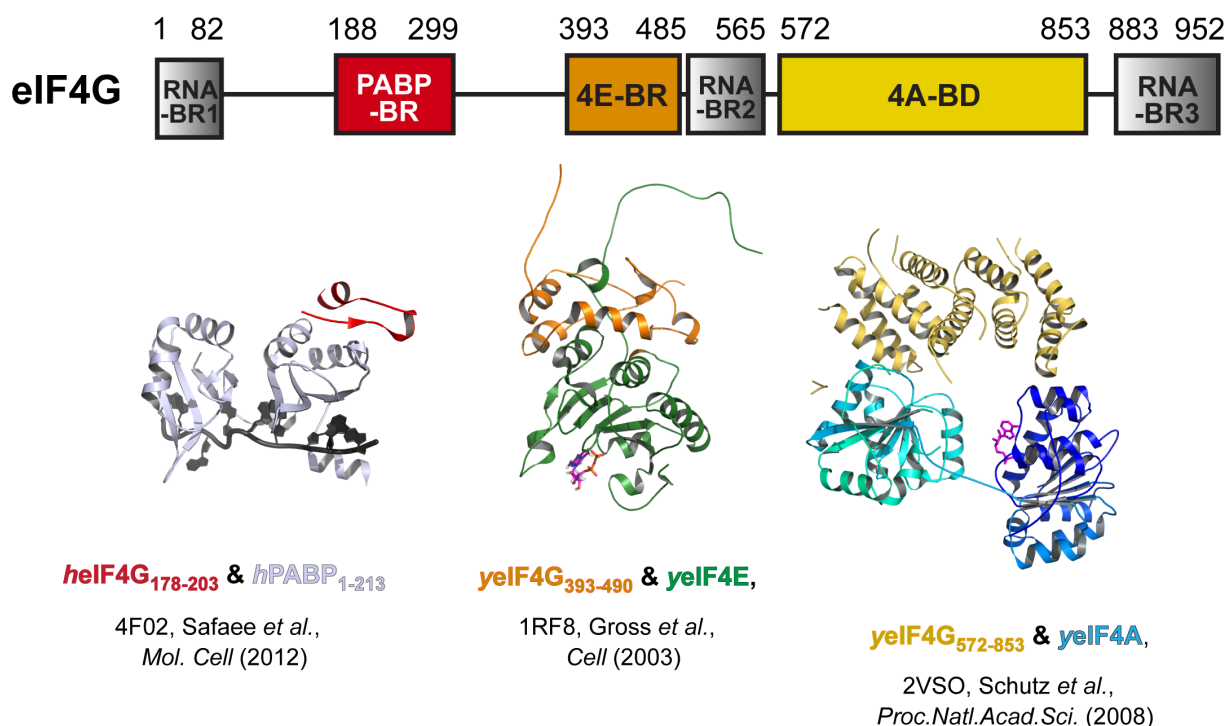
eIF4H (~25 kDa) was discovered based on its stimulatory ability to an *in vitro* translation [96]. This protein exhibits a sequence homology to the N-terminal region of eIF4B that contains the first RRM domain but not the DRYG repeats nor the second RNA-binding basic domain. eIF4H enhance activities of eIF4A but cannot completely substitute for eIF4B [88,96]. In yeast cells, an eIF4H homolog is missing.

#### 1.4.4 eIF4G

The translation initiation factor 4G (eIF4G) is a multifactor binding protein at the center of eIF4F complex and is vital in both cap-dependent and IRES-driven translation [9]. This protein has been conceptually divided into three parts based on the products of picornaviral protease cleavage [9]: The N-terminal third is the most variable and flexible part of eIF4G, containing a binding site for the poly(A)-binding protein (PABP), as well as one for the cap-binding protein eIF4E [9,75,97,98]. The middle third of eIF4G associates with the RNA helicase eIF4A and in mammals with eIF3, which directly binds to the ribosome [11,12,14,86]. The C-terminal third, which is missing in yeast eIF4G, contains the second eIF4A binding site and a binding site for the eIF4E kinase MNK1 [11,13,99]. The first two thirds of eIF4G have been identified as crucial for cap-dependent translation, whereas the last third is not required for translation and only plays some modulatory roles [99,100]. In addition to binding sites for proteins, eIF4G also possesses a conserved RNA-binding region firstly noted in yeast eIF4G [101]. These properties of eIF4G are the basis for its two main functions: mRNA circularization and ribosome recruitment.

**PABP-binding site** The interaction between eIF4G and PABP has been determined to be necessary for the stimulation of mRNA translation [102]. eIF4G binds the RRM2 of PABP, which contains four RNA recognition motifs (RRM1-4) and a C-terminal MLLE domain, through a ~25-residue fragment in its N-terminus (residues 178-203 in human eIF4G) as shown by both NMR and crystal structure. The binding of PABP induces a conformational change of eIF4G from an unstructured form to a folded state (Figure 1.8) [98].

**eIF4E-binding site** eIF4E-binding site on eIF4G is the first best characterized polypeptide-binding site. eIF4G utilizes the conserved  $Y-X-X-X-X-L-\Phi$  motif (where  $X$  is variable and  $\Phi$  represents hydrophobic) to recognize eIF4E [103]. This segment is also present in a family of translational repressors, the eIF4E-binding proteins (4E-BPs), which compete with eIF4G to bind eIF4E (see Section 1.5). NMR and crystal structures show that the eIF4G fragment interacts with the dorsal convex surface of eIF4E, behind the cap-binding groove [75,103].



**Figure 1.8:** Overall information of eIF4G. (A) Schematic domain organization of *S. cerevisiae* eIF4G. The boundaries of the domains are indicated. RNA-BR: RNA binding region; PABP-BR: PABP binding region; 4E-BR: eIF4A binding region; 4A-BD: eIF4A binding domain. (B) The known domain structures of eIF4G in complex with PABP (pale-blue, left), eIF4E (green, middle) and eIF4A (cyan and blue, right), respectively. *h*: human; *y*: yeast.

**eIF4A-binding site** The eIF4A-binding region of eIF4G forms a HEAT repeat domain, with helically stacked  $\alpha$ -helical hairpins (Figure 1.8) [86]. There is only one 4A-binding site in yeast eIF4G, but two in mammalian. Binding of eIF4A to eIF4G is not as strong as that of eIF4E to eIF4G (a  $K_d$  of  $\sim 10$  nM) [104]. eIF4A can be readily dissociated from eIF4G on ion exchange columns and natively purified eIF4F complexes usually lack eIF4A [105-107]. The structure of yeast eIF4A/eIF4G-M complex (where eIF4G only contains the 4A-binding domain) has also been solved [86]. eIF4G associates with both recA-like domains of eIF4A, but primarily with the C-terminal recA-like domain. This interaction triggers eIF4A to adopt a half open conformation that is thought to be correlated with the stimulation of eIF4A activities [108].

**RNA-binding site** Yeast eIF4G contains three regions that are rich in basic amino acids. These three basic clusters are located in the N-, middle and C-terminal domain of eIF4G, respectively, and are able to bind RNA [109,110]. Deletion of each one individually still allows for tight binding of eIF4G to RNA. It has been shown that the three RNA binding fragments of eIF4G cooperatively stimulate RNA unwinding by eIF4A in the 5' to 3' direction [110]. Like its yeast counterpart, human eIF4G also contains RNA-binding fragments. However, it does not show a preference to 5'-overhangs, indicating that this mechanism is not universal.

## 1.5 Translational control via eIF4F

Translational control in eukaryotic cells is critical for regulating gene expression. Compared to regulation at the point of transcription, translational control of existing mRNAs allows for more rapid response to environmental changes. Controlling translational rates at initiation phase rather than elongation or termination makes the regulation particularly effective, thus initiation is the most regulated and the rate-limiting stage of protein synthesis. Two general types of regulation occur at translational initiation phase: by mechanisms which influence initiation factors (eIFs) or the mRNA itself either through specific RNA-binding proteins or via microRNAs. The best-understood instances in the eIFs-dependent regulation are control of the availability of translational active eIF2 and eIF4F by phosphorylation. eIF3 and eIF4B are also phosphorylated under certain conditions, which may alter their activities [31]. However, not much is known about this regulation. In this section the translational regulation via eIF4F (a main focus of this thesis) will be described in detail.

### 1.5.1 eIF4E regulation

eIF4E can be phosphorylated *in vivo* [7,111]. The phosphorylation of eIF4E is catalyzed by mitogen-activated protein kinase (MAPK)-interacting serine/threonine kinase 1 (Mnk1), which directly binds to eIF4G in mammalian cells [112]. It has

been reported that phosphorylated eIF4E exhibits up to fourfold stronger affinity for the cap analogs m<sup>7</sup>GTP and m<sup>7</sup>GpppG, as well as for capped mRNA than its nonphosphorylated counterpart [113]. eIF4E phosphorylation has been shown to promote translational upregulation of several proteins that are implicated in tumorigenesis [7].

In addition, eIF4E activity is also regulated through its association with a family of repressor proteins, eIF4E-binding proteins (4E-Bps). 4E-Bps (~10 to 12 kDa) inhibit cap-dependent (but not cap-independent) translation, both in cell-free translation assays and when overexpressed in cells [114]. Binding of 4E-BPs to eIF4E does not affect eIF4E affinity for the mRNA cap. Instead, 4E-BPs prevents the interaction of eIF4E with eIF4G, thus suppressing the formation of the mRNA activation complex eIF4F. Both eIF4G and 4E-BPs contains the  $Y-X-X-X-X-L-\Phi$  motif (where  $X$  is any amino acid and  $\Phi$  is a hydrophobic residue) that is the binding site for eIF4E, and compete for interaction with eIF4E. 4E-BPs and eIF4G do bind the same region on eIF4E as revealed by the crystal structures of eIF4E in complex with a fragment of 4E-BPs or eIF4G [103].

Binding of 4E-BPs to eIF4E is modulated by 4E-BPs phosphorylation. Superphosphorylated 4E-BPs, which is induced by extracellular stimuli (including hormones, growth factors, cytokines, mitogens, *etc.*), exhibit lower affinity to eIF4E. In contrast, under stress conditions (*e.g.* starvation or infection with poliovirus), the phosphorylation of eIF4E is decreased, resulting in efficient binding to eIF4E and thus reduction of protein expression.

### 1.5.2 eIF4A regulation

eIF4A is also an important target for regulation. The tumor suppressor protein, programmed cell death 4 (PDCD4), has been demonstrated to inhibit translation by interacting with eIF4A [115]. The structure of eIF4A-PDCD4 complex has been determined, revealing that one PDCD4 molecule binds two eIF4A molecules, traps eIF4A in an inactive conformation and prevents the interaction of eIF4A with eIF4G to restrain the formation of the enzymatic active eIF4F complex [116,117]. In

addition, several studies have reported that two natural products, hippuristanol and pateamine A, inhibit translation initiation through directly modulating eIF4A activity [118-120]. Moreover, the non-coding RNA BC1 specifically binds to eIF4A and promotes the RNA-dependent ATPase activity of eIF4A but block its RNA separation activity, and thus inhibits translation initiation in neurons [121].

### **1.5.3 eIF4G regulation**

eIF4G is phosphorylated in response to various stimuli [122-125]. Increased phosphorylation of eIF4G is correlated with an increase in the formation of eIF4F complex, indicating that eIF4G phosphorylation regulates the level of eIF4F [126]. The eIF4F concentration is also impaired by an eIF4G structural homolog, death-associated protein 5 (DAP5, also called p97). DAP5 shares ~30 identity with C-terminal two thirds of eIF4G but lacks the N-terminal eIF4E and PABP binding sites. It competitively interacts with eIF4A (and eIF3) and promotes the IRES-driven translation, but prevents the cap-dependent translation [127-129]. Similar inhibition of cap-dependent translation occurs under certain stressful conditions, for example, viral infection or apoptosis. eIF4G is proteolytically cleaved into two or more fragments, resulting in an eIF4F-complex lacking eIF4E (and thus the cap-recognition component) [7,130].

## **1.6 Aims of this thesis**

Despite the extensive and crucial involvement of the multi-subunit eIF3 complex in translation initiation (43S PIC assembly, mRNA recruitment, scanning and start codon selection), its overall structure is limited to the low-resolution cryo-EM reconstructions [61,63]. In addition, only several structures of domains of eIF3 subunits (the RRM domain of eIF3b and eIF3g, the PCI domain of eIF3a and eIF3k) and of the seven-bladed  $\beta$ -propeller protein eIF3i have been determined at atomic level either by X-ray crystallography or NMR [47,48,52,53,131]. However, in order to understand the mechanism by which eIF3 works in translation initiation more

high-resolution structures of eIF3 complex, subcomplexes or domains are needed. Hence, a major goal of this thesis is to establish strategies for reconstitution of eIF3 subcomplexes *in vitro*, as well as obtain crystals of these complexes and individual subunits. Proteins from different organisms are chosen as the targets of interest. Promising crystallizable subcomplexes of eIF3 and subunit variants are systematically designed and screened.

Although structural information about eIF4E- and eIF4A-binding domain of eIF4G in complex with eIF4E and eIF4A, respectively, are available, the molecular organization in eIF4F subunits is still elusive. Therefore, one aim of this thesis is to crystallize full-length eIF4F complex, and further figure out why eIF4E, eIF4G, eIF4A and eIF4B work cooperatively on mRNA activation and the reason for the regulation of their activity. As identification of crystallization condition would fail, other structural probing methods (for example, cross-linking) should be performed.





## 2

## Materials and methods

## 2.1 Cloning

### 2.1.1 Materials

#### 2.1.1.1 Chemicals

All standard chemicals, which are not listed following, were purchased from AppliChen, Bio-Rad, Biotech, Fluka, Merck, Roth or Sigma-Aldrich. The purity grade *pro analysis* was chosen.

#### 2.1.1.2 Buffers

TAE (1 ×) buffer was used for making the agarose gels and as the running buffer of agarose gel electrophoresis.

**Table 2.1:** Standard buffers for cloning

Buffer	Composition
TAE (50 ×)	2 M Tris acetate, pH 8.0, 50 mM EDTA
DNA loading buffer (6 ×)	0.25% Bromophenole blue, 0.25% xylene cyanol FF, 30% Glycerol

### 2.1.1.3 Cell media

LB,  $2 \times$  YT and SOC media were used for growing cells intended for cloning. To make agar plates, 1.5% (w/v) (final concentration) agar was supplemented prior to autoclaving.

**Table 2.2:** Media for bacterial cells

Medium	Composition
LB	1% (w/v) Tryptone, 0.5% (w/v) Yeast extract, 0.5% (w/v) NaCl
$2 \times$ YT	1.6% (w/v) Tryptone, 1% (w/v) Yeast extract, 1% (w/v) NaCl
SOC medium	2% (w/v) Tryptone, 0.55% (w/v) yeast extract, 10 mM NaCl, 10 mM KCl, 10 mM MgCl <sub>2</sub> , 10 mM MgSO <sub>4</sub> , 20 mM Glucose

### 2.1.1.4 *E. coli* host strains

*Escherichia coli* (*E. coli*) strain XL1-Blue was used for routine cloning, while DH10Bac<sup>TM</sup> was utilized for generation of the recombinant bacmid that contains the gene of interest (i.e. full-length eIF4G, see 2.2.2.2).

**Table 2.3:** Cells used for cloning

Cells	Key features	Reference
XL1-Blue	Endonuclease and recombination deficient; an excellent host strain for routine cloning	Agilent
DH10Bac <sup>TM</sup>	Contains baculovirus shuttle vector (bacmid) and a transposition help plasmid; allows generation of a recombinant bacmid	Invitrogen

### 2.1.1.5 Enzymes and reagents

Appropriate buffers provided by the suppliers were used for the enzymes. All antibiotics were prepared in a 1000  $\times$  stock solutions. Antibiotics were purchased from Roth.

**Table 2.4:** Enzymes and reagents used for cloning

Enzyme and standard protein	Supplier
dNTPs (10mM each)	Fermentas
<i>Pfu</i> DNA polymerase	Fermentas
<i>Phusion</i> DNA polymerase	Finnzymes
<i>Taq</i> DNA polymerase	Promega
GelRed	VWR
Gene Ruler™ 1kb DNA ladder	Fermentas
<i>Bam</i> HI	Fermentas
CIAP	NEB
<i>Hind</i> III	Fermentas
<i>Nde</i> I	Fermentas
<i>Not</i> I	Fermentas
<i>Xho</i> II	Fermentas
T4 DNA ligase	Fermentas

**Table 2.5:** Antibiotics

Antibiotics	Final concentration used
Ampicillin	100 µg/ml
Chloramphenicol	34 µg/ml
Gentamicin	7 µg/ml
Kanamycin	50 µg/ml
Tetracycline	15 µg/ml

### 2.1.1.6 Genomic DNA

*Saccharomyces cerevisiae* S288c genomic DNA was a collection of our group (Department for Molecular Structure Biology, MSB). The organism *Chaetomium*

*thermophilum* DSM 1495 was a kind gift of Prof. Dr. Ed Hurt (Heidelberg University). The genomic DNA of *Chaetomium thermophilum* was extracted by Bernhard Kuhle.

### 2.1.1.7 Instruments

Machines used in the procedure of cloning are listed in Table 2.7.

**Table 2.6:** Machines used in the procedure of cloning

Machine	Supplier
Adjustable pipettes	Eppendorf
Agarose gel electrophoresis chamber	Bio-Rad
BioPhotometer	Eppendorf
Balance	Sartorius
GelDoc EQ System	Bio-Rad
HLC block thermostat BT 1302	Techne
Incubator Mytron	Schütt
Innova 4230 incubator	New Brunswick
Magnetic stirrer IKAMAG REO	IKA
MiniSpin plus	Eppendorf
PCR mastercycler gradient	Eppendorf
PCR-Whatman Biometra T personal	Biometra
pH-Meter Beckman	Beckman Coulter
Pipette Controller Accu-Jet <i>pro</i>	Brand
Table centrifuge 5415R	Eppendorf
Thermomixer comfort	Eppendorf
Unitron incubator shaker	Infors

### 2.1.1.8 Molecular biology kits

Kits were used following the user manuals provided by the suppliers.

**Table 2.7:** Commercial molecular biology kits

Kit	Supplier
peqGOLD Gel Extraction Kit	PEQLAB
peqGOLD MicroSpin Cycle-Pure Kit	PEQLAB
peqGOLD Plasmid Miniprep Kit	PEQLAB
QIAquick Gel Extraction Kit	Qiagen
QIAquick PCR Purification Kit	Qiagen
QIAprep Spin Miniprep Kit	Qiagen
In-fusion® HD Cloning Kit	Clontech
BigDye® Terminator v1.1 Cycle Sequencing Kit	Life technologies

## 2.1.2 Methods

### 2.1.2.1 Determination of nucleic acid concentration

The concentration of nucleic acids in solution was determined by measuring the absorption at the wavelength of 260 nm, using a spectrophotometer. The concentration of DNA or RNA was then calculated as follows [132]:

- dsDNA Concentration ( $\mu\text{g/ml}$ ) =  $(\text{OD}_{260}) \times (\text{dilution factor}) \times 50$
- ssDNA Concentration ( $\mu\text{g/ml}$ ) =  $(\text{OD}_{260}) \times (\text{dilution factor}) \times 33$
- ssRNA Concentration ( $\mu\text{g/ml}$ ) =  $(\text{OD}_{260}) \times (\text{dilution factor}) \times 40$

### 2.1.2.2 Polymerase chain reaction (PCR)

Genes or gene fragments were amplified by applying the polymerase chain reactions

(PCR) [133]. The compositions of a typical PCR and the procedure of PCR cycles are given below.

**Table 2.8:** Recipe of a standard PCR

Item	Volume ( $\mu$ l)
5 $\times$ (or 10 $\times$ ) buffer	10 (or 5)
dNTPs (10 mM each)	1
Forward Primer (10 pmol)	2.5
Reverse Primer (10 pmol)	2.5
Template (100 ng/ $\mu$ l)	0.5
Polymerase (10 U/ $\mu$ l)	1
H <sub>2</sub> O	32.5 (or 37.5)
Total volume	50

**Table 2.9:** Procedure of PCR cycles

Cycles	Temperature ( $^{\circ}$ C)	Time
1 $\times$	95 (initial denaturation)	3 min
20-30 $\times$	95 (denaturation)	30 sec
	55-61 (annealing)	30 sec
	72 (elongation)	30 sec - 6 min
1 $\times$	72 (final elongation)	10 min

PCR products were purified either using a PCR purification kit (if the samples were pure enough), or a gel extraction kit. In the later case, an agarose gel electrophoresis (see 2.1.2.3) was first performed, and the bands containing the DNA of interest were excised from the gel.

### 2.1.2.3 Agarose gel electrophoresis

To separate DNA fragments from PCR or a restriction digest reaction, agarose gel electrophoresis was used for both analytical visualization and purification. Gels were prepared with 1  $\times$  TAE buffer, 0.5  $\times$  GelRed and agarose in different

concentrations, ranging from 0.8% to 2% according to the size of the nucleic acid fragments. Gels were run at a constant current of 400 mA in  $1 \times$  TAE buffer. DNA bands were visualized by using GelDoc EQ System at UV 254 nm.

In necessary cases, DNA bands of interest were excised from the gels with a sterile blade. DNA fragments were extracted from the gel by using gel extraction kit following the user manuals.

#### 2.1.2.4 Preparation of chemically competent *E. coli* cells

Chemically competent *E. coli* cells were generated using the  $\text{CaCl}_2$  method [134]. 2.5 ml of cells cultured overnight were inoculated into 500 ml  $2 \times$  YT medium supplemented with appropriate antibiotics. Cells were grown at 37 °C with shaking at 220 rpm until an  $\text{OD}_{600}$  of 0.5-0.6 was reached and then incubated on ice for 30 min. Cells were collected in sterile bottles by centrifugation ( $3000 \times g$ , 10 min, 4 °C). The cell pellet was resuspended in 125 ml pre-cooled 0.1 M  $\text{MgCl}_2$ , placed on ice for 20 min and harvested by centrifugation ( $2500 \times g$ , 12 min, 4 °C). Cells were additionally washed with 250 ml 0.1 M  $\text{CaCl}_2$  (20 min incubation on ice and then centrifugation) and finally resuspended in 5 ml pre-cooled solution containing 0.1 M  $\text{CaCl}_2$  and 15 % glycerol. The competent cells were aliquoted in cold Eppendorf tubes, flash-frozen in liquid nitrogen and stored at -80 °C.

#### 2.1.2.5 Molecular cloning

**Digestion-ligation based cloning** Target DNA fragments produced by PCR and selected vectors were digested with identical restriction enzymes, resulting in compatible ends in vectors and DNA fragments. Buffers, temperatures as well as the amount of vector, DNA and enzymes for the digestion reactions were chosen according to the instruction of suppliers. In the case of vector digestion, Calf intestinal alkaline phosphatase (CIAP, 0.01 unit per pmol DNA) was added 30 min before stopping the reaction. The digested products were purified using either PCR purification kits or gel extraction kits.

Digested plasmid DNA (usually 100 ng) and insert DNA (3-5 fold molar excess) were ligated by T4 DNA ligase (500 units). Typically, ligation reactions were incubated at room temperature for 1-1.5 hours. The ligation reaction mixture was transformed into XL1-Blue *E. coli* cells.

**In-fusion cloning** In-fusion HD Cloning kit (Clontech) allows fast, directional cloning of several DNA fragments into one vector at the same time at any location. Therefore, it has great advantage when the genomic DNA, of which the complete genes are usually separated by several introns, is used as the PCR template. In this study, this cloning method was extensively used (see Table 2.12). Primers with 15 nucleotides overlap to the adjacent DNA fragment were designed according to the detailed description of the In-fusion user manual. 100 ng plasmid DNA and 3-fold molar excess of insert DNA fragments were mixed and incubated at 50 °C for 15 min. Reaction mixture was then transformed into *E. coli* XL1-Blue cells.

#### 2.1.2.6 Site-directed mutagenesis

Desired point mutations were introduced according to the instruction of QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies). Different from normal PCR (see 2.1.2.2), 1.5 µl DMSO was additionally supplied. The resulting clones were verified by DNA sequencing (see 2.1.2.8).

#### 2.1.2.7 Transformation

Approximately 100 ng plasmid DNA, or reaction mixtures from ligation or In-fusion were individually added into ~100 µl just thawed chemically competent cells. After 30 min incubation on ice, cells were heat-shocked at 42 °C for 90 sec and placed on ice for 3 min. Then 1 ml 2 × YT medium (without antibiotics) was supplied and the cell suspension was incubated at 37 °C with shaking for 45 min to recovery. Finally, the cell suspension was distributed onto a 2 × YT agar plate that contains the appropriate antibiotic(s). Transformation of pFastBac HT B into DH10Bac™ cells is described in Section 2.2.2.4.



### 2.1.2.8 DNA sequencing

The sequencing of DNA was performed to assure the correct sequences of the desirable inserts. Compositions of the sequencing PCR and procedure of the PCR cycles are given in Table 2.10 and Table 2.11, respectively.

**Table 2.10:** Recipe of sequencing PCR. The sequencing primers are used according to the requirement of the vectors.

Item	Volume ( $\mu$ l)
Plasmid	1
Sequencing primer (10 pmol)	1
BigDye Terminator v1.1	1.5
Sequencing buffer	1.5
H <sub>2</sub> O	5
Total volume	10

**Table 2.11:** Procedure of sequencing PCR cycles

Cycle	Temperature ( $^{\circ}$ C)	Time
1 $\times$	95 (initial denaturation)	2 min
25 $\times$	95 (denaturation)	20 sec
	55 (annealing)	20 sec
	60 (elongation)	4 min

DNA was precipitated by addition of 1  $\mu$ l 3 M NaOAc, 1  $\mu$ l 125 mM EDTA and 50  $\mu$ l ethanol. The mixture was incubated for 5 min and then centrifuged at 20,000  $\times$  g for 20 min. The DNA pellet was washed with 70  $\mu$ l of 70 % (v/v) ethanol and dried at room temperature. The pellet was finally resuspended in formamide and sent to Department of Developmental Biochemistry, University of Göttingen, for sequencing.

### 2.1.3 Plasmids

Genes were individually cloned into vector pGEX6P-1 (GE Healthcare), modified pET15b (MSB), modified pET22b (MSB) and modified pETDuet (MSB) for expression in *E. coli* cell. Yeast eIF4G was also cloned into pFastBac HT B (Invitrogen) for expression in insect cells. DNA sequencing verified that the sequences of inserts are correct. In order to amplify plasmid DNAs, the plasmids were respectively transformed into XL1-Blue *E. coli* cells and cultured at 37 °C overnight. Plasmids isolation was carried out using Miniprep kits. Table 2.12 lists all plasmids generated in this study.

**Table 2.12:** Expression vectors designed and generated in this study. *ct*: *Chaetomium thermophilum*; *sc*: *Saccharomyces cerevisiae*; DL: digestion-ligation based cloning; IF: In-fusion cloning; PP: PreScission protease; TEV: Tobacco Etch Virus protease; Thrb: Thrombin protease.

Plasmid	Tag(s)	Resistance	Cloning procedure	Protease
pGEX6P-1_ <i>ct</i> rpS9e FL (1-190)	N-GST	Ampicillin	IF	PP
pGEX6P-1_ <i>ct</i> rpS9e (1-183)	N-GST	Ampicillin	IF	PP
pGEX6P-1_ <i>ct</i> eIF3a FL (1-1086)	N-GST	Ampicillin	IF	PP
pGEX6P-1_ <i>ct</i> eIF3a (1-906)	N-GST	Ampicillin	IF	PP
pGEX6P-1_ <i>ct</i> eIF3b FL (1-746)	N-GST	Ampicillin	IF	PP
pGEX6P-1_ <i>ct</i> eIF3b (24-135)	N-GST	Ampicillin	IF	PP
mpETDuet_ <i>ct</i> eIF3b (41-746)	N-His <sub>6</sub>	Ampicillin	DL (BamHI, Hind3)	TEV
mpETDuet_ <i>ct</i> eIF3b (167-746)	N-His <sub>6</sub>	Ampicillin	DL (BamHI, Hind3)	TEV
pGEX6P-1_ <i>ct</i> eIF3b (167-625)	N-GST	Ampicillin	IF	PP
pGEX6P-1_ <i>ct</i> eIF3b (167-640)	N-GST	Ampicillin	IF	PP
pGEX6P-1_ <i>ct</i> eIF3b (167-670)	N-GST	Ampicillin	IF	PP
pGEX6P-1_ <i>ct</i> eIF3b (167-704)	N-GST	Ampicillin	IF	PP
pGEX6P-1_ <i>ct</i> eIF3b (167-720)	N-GST	Ampicillin	IF	PP
pGEX6P-1_ <i>ct</i> eIF3g FL (1-289)	N-GST	Ampicillin	IF	PP

Plasmid	Tag(s)	Resistance	Cloning procedure	Protease
pGEX6P-1 <sub>ct</sub> eIF3g (26-286)	N-GST	Ampicillin	IF	PP
pGEX6P-1 <sub>ct</sub> eIF3g (26-286) ( $\Delta$ 151-179)	N-GST	Ampicillin	IF	PP
pGEX6P-1 <sub>ct</sub> eIF3g (26-150)	N-GST	Ampicillin	IF	PP
pGEX6P-1 <sub>ct</sub> eIF3g (180-286)	N-GST	Ampicillin	IF	PP
pGEX6P-1 <sub>ct</sub> eIF3i (1-330)	N-GST	Ampicillin	IF	PP
pGEX6P-1 <sub>ct</sub> eIF3j FL (1-279)	N-GST	Ampicillin	IF	PP
pGEX6P-1 <sub>ct</sub> eIF3j (53-238)	N-GST	Ampicillin	IF	PP
pGEX6P-1 <sub>sc</sub> eIF3j FL (1-265)	N-GST	Ampicillin	DL ( <i>Bam</i> HI, <i>Xho</i> I)	PP
pGEX6P-1 <sub>sc</sub> eIF3j (20-218)	N-GST	Ampicillin	DL ( <i>Bam</i> HI, <i>Xho</i> I)	PP
pGEX6P-1 <sub>sc</sub> eIF3j (56-218)	N-GST	Ampicillin	DL ( <i>Bam</i> HI, <i>Xho</i> I)	PP
mpET15b <sub>sc</sub> eIF4A FL (1-395)	N-His <sub>6</sub>	Ampicillin	DL ( <i>Nde</i> I, <i>Xho</i> I)	TEV
pGEX6P-1 <sub>sc</sub> eIF4A (14-395)	N-GST	Ampicillin	IF	PP
pGEX6P-1 <sub>sc</sub> eIF4A (14-395) 23K→Bpa	N-GST	Ampicillin	QuikChange	PP
pGEX6P-1 <sub>sc</sub> eIF4A (14-395) 25D→Bpa	N-GST	Ampicillin	QuikChange	PP
pGEX6P-1 <sub>sc</sub> eIF4A (14-395) 43E→Bpa	N-GST	Ampicillin	QuikChange	PP
pGEX6P-1 <sub>sc</sub> eIF4A (14-395) 67Q→Bpa	N-GST	Ampicillin	QuikChange	PP
pGEX6P-1 <sub>sc</sub> eIF4A (14-395) 125I→Bpa	N-GST	Ampicillin	QuikChange	PP
pGEX6P-1 <sub>sc</sub> eIF4A (14-395) 205P→Bpa	N-GST	Ampicillin	QuikChange	PP
pGEX6P-1 <sub>sc</sub> eIF4A (14-395) 284K→Bpa	N-GST	Ampicillin	QuikChange	PP
pGEX6P-1 <sub>sc</sub> eIF4A (14-395) 296Q→Bpa	N-GST	Ampicillin	QuikChange	PP
pGEX6P-1 <sub>sc</sub> eIF4A (14-395) 297Q→Bpa	N-GST	Ampicillin	QuikChange	PP
pGEX6P-1 <sub>sc</sub> eIF4A (14-395) 322R→Bpa	N-GST	Ampicillin	QuikChange	PP
pGEX6P-1 <sub>sc</sub> eIF4A (14-395) 340N→Bpa	N-GST	Ampicillin	QuikChange	PP
pGEX6P-1 <sub>sc</sub> eIF4A (14-395) 342E→Bpa	N-GST	Ampicillin	QuikChange	PP
pGEX6P-1 <sub>sc</sub> eIF4B FL(1-436)	N-GST	Ampicillin	DL ( <i>Bam</i> HI, <i>Not</i> I)	PP
mpET15b <sub>sc</sub> eIF4E FL(1-213)	N-His <sub>6</sub>	Ampicillin	DL ( <i>Nde</i> I, <i>Xho</i> I)	TEV
pGEX6P-1 <sub>sc</sub> eIF4E FL(1-213)	N-GST	Ampicillin	DL ( <i>Nde</i> I, <i>Xho</i> I)	PP

## Materials and methods

Plasmid	Tag(s)	Resistance	Cloning procedure	Protease
pGEX6P-1_sc eIF4E (28-213)	N-GST	Ampicillin	DL ( <i>Bam</i> HI, <i>Xho</i> I)	PP
pGEX6P-1_sc eIF4E (36-213)	N-GST	Ampicillin	DL ( <i>Bam</i> HI, <i>Xho</i> I)	PP
pGEX6P-1_sc eIF4G FL (1-952)	N-GST	Ampicillin	DL ( <i>Bam</i> HI, <i>Xho</i> I)	PP
mpET22b_sc eIF4G FL (1-952)	N-Strep, C-His <sub>6</sub>	Ampicillin	DL ( <i>Bam</i> HI, <i>Xho</i> I)	Thrb(N)
pFastBac HTB_sc eIF4G FL	N-His <sub>6</sub>	Ampicillin Gentamycin	DL ( <i>Bam</i> HI, <i>Xho</i> I)	TEV
pGEX6P-1_sc eIF4G (1-617)	N-GST	Ampicillin	DL ( <i>Bam</i> HI, <i>Xho</i> I)	PP
pGEX6P-1_sc eIF4G (410-853)	N-GST	Ampicillin	DL ( <i>Bam</i> HI, <i>Xho</i> I)	PP
pGEX6P-1_sc eIF4G (410-883)	N-GST	Ampicillin	DL ( <i>Bam</i> HI, <i>Xho</i> I)	PP
pGEX6P-1_sc eIF4G (410-952)	N-GST	Ampicillin	DL ( <i>Bam</i> HI, <i>Xho</i> I)	PP
pGEX6P-1_sc eIF4G (468-617)	N-GST	Ampicillin	DL ( <i>Bam</i> HI, <i>Xho</i> I)	PP
pGEX6P-1_sc eIF4G (468-853)	N-GST	Ampicillin	DL ( <i>Bam</i> HI, <i>Xho</i> I)	PP
pGEX6P-1_sc eIF4G (468-883)	N-GST	Ampicillin	DL ( <i>Bam</i> HI, <i>Xho</i> I)	PP
pGEX6P-1_sc eIF4G (468-952)	N-GST	Ampicillin	DL ( <i>Bam</i> HI, <i>Xho</i> I)	PP
mpET Duet_sc eIF4G (542-883)	N-His <sub>6</sub>	Ampicillin	DL ( <i>Bam</i> HI, <i>Hind</i> III)	TEV
mpET Duet_sc eIF4G (542-952)	N-His <sub>6</sub>	Ampicillin	DL ( <i>Bam</i> HI, <i>Hind</i> III)	TEV
mpET Duet_sc eIF4G (572-853)	N-His <sub>6</sub>	Ampicillin	DL ( <i>Bam</i> HI, <i>Hind</i> III)	TEV
mpET Duet_sc eIF4G (572-952)	N-His <sub>6</sub>	Ampicillin	DL ( <i>Bam</i> HI, <i>Hind</i> III)	TEV

## 2.2 Protein expression and purification

Materials and methods that are provided in the Section 2.1 will not be redundantly listed here and later.

### 2.2.1 Materials

#### 2.2.1.1 Instruments

**Table 2.13:** Machines used in protein expression and purification

Machine	Supplier
Centrifuge Allegra 21R	Beckman Coulter
Centrifuge Avanti J-20 XPI	Beckman Coulter
Centrifuge Avanti J-30 I	Beckman Coulter
Rotor JA-20	Beckman Coulter
Rotor JA-30.5	Beckman Coulter
Rotor JA-8.1000	Beckman Coulter
Rotor S4180	Beckman Coulter
Axiovert 25 inverse microscope	Carl Zeiss
Branson sonifier 250	Fisher Scientific
Micro fluidizer M110S	Microfluidics
Platform shaker Promax 1020	Heidolph
Prettl Telstar Typ Bio II A	TelStar
SDS-PAGE system Hoefer miniVE	Amersham Pharma
Table mixer RM5	Schütt
Unitron incubator shakers	Infors
Ultra Centrifugal Mill ZM	Retsch
UV lamp (365 nm)	Vilber Lourmat

### 2.2.1.2 Chromatography systems and columns

Äkta Prime and Äkta Purifier (GE Healthcare) were used for purifying proteins. Samples were loaded on the columns by using loops with volume of 0.5 ml, 1 ml, 5 ml, 10 ml, 50 ml and 150 ml, respectively. Columns used in this study are listed below (Table 2.14).

**Table 2.14:** Chromatographic resins and columns

Name	Supplier
GSH Sepharose/GSTrap 5ml	GE Healthcare
Ni-NTA Sepharose/HisTrap 5ml	GE Healthcare
HiPrep Desalting 26/10	GE Healthcare
Source 30Q	GE Healthcare
SP Sepharose	GE Healthcare
StrepTactin HP Sepharose 5ml	GE Healthcare
Superdex 75 (26/60)	GE Healthcare
Superdex 200 (26/60)	GE Healthcare
Superdex 200 (16/60)	GE Healthcare
Superdex 200 (10/300)	GE Healthcare

### 2.2.1.3 Buffers

**Table 2.15:** Standard buffer routinely used in the work regarding proteins

Buffer	Composition
SDS loading buffer (2 ×)	100 mM Tris pH 6.8, 4% (w/v) SDS, 4% (v/v) β-ME, 20% (v/v) Glycerol, 0.02% (w/v) Bromophenol blue
SDS loading buffer (5 ×)	250 mM Tris pH 6.8, 10% (w/v) SDS, 10% (v/v) β-ME, 50% (v/v) Glycerol, 0.05% (w/v) Bromophenol blue
SDS running buffer (10 ×)	250 mM Tris pH 6.8, 2 mM Glycine, 1% (w/v) SDS
Coomassie staining solution	0.025% (w/v) Coomassie (R250); 0.025% (w/v) Coomassie (G250); 30% (v/v) isopropanol; 7.5% (v/v) acetic acid

Buffer	Composition
Destaining solution	40% (v/v) ethanol, 20% (v/v) acetic acid
Blot transfer buffer (1 ×)	25 mM Tris, 192 mM Glycin, 20 % (v/v) methanol
TBST	150 mM NaCl, 20 mM Tris/HCl, pH 8.0, 0.2 % Tween 20
Blot Buffer A	100 mM Tris/HCl, pH 9.5, 100 mM NaCl, 5 mM MgCl <sub>2</sub>

#### 2.2.1.4 Enzymes and reagents

The stock concentration of DNAase and ALP (a mixture of Aprotinin, Leupeptin and Pepstatin) is 2 mg/ml and 100 mg/ml, respectively.

**Table 2.16:** Enzymes and special reagents for protein expression and purification

Name	Supplier
β-Mercaptoethanol (β-ME)	Roth
Bradford	Roth
DNAase	Appligen
Dithiothreitol (DTT)	Appligen
IPTG <sup>a</sup>	Roth
PMSF <sup>b</sup>	Appligen
Proteinase inhibitor mixture ALP	MSB
Proteinase inhibitor tablets (Complete)	Roche
PreScission protease (PP)	MSB
TEV protease	MSB

<sup>a</sup> Isopropyl β-D-1-thiogalactopyranoside; <sup>b</sup> PMSF: Phenylmethylsulfonyl fluoride.

#### 2.2.1.5 Protein expression cells

*Escherichia coli* (*E. coli*) strains BL21 (DE3) and Rosetta 2 (DE3) were used to express proteins. B834 (DE3) was used to produce selenomethionine derivative

*cteIF3b*. Sf9 insect cells were used to propagate recombinant baculoviral stocks and preliminary expression experiments. High Five™ cells were used for high-level expression of His tagged full-length yeast eIF4G protein.

**Table 2.17:** Cells used for expressing recombinant proteins

Cells	Key features	Reference
BL21 (DE3)	T7 RNA polymerase controlled by Lac regulatory construct; IPTG inducible	Novagen
Rosetta 2 (DE3)	Enhance the expression of eukaryotic proteins; pRARE2 (Cam); IPTG inducible	Novagen
B834 (DE3)	Methionine auxotroph; seleno-Met labeling; IPTG inducible	Novagen
Sf9	Isolate and propagate recombinant baculoviral stocks; produce recombinant protein	Invitrogen
High Five™	High-level expression of recombinant protein using Baculovirus Expression Vector system (BEVS)	Invitrogen

### 2.2.1.6 Cell media for protein expression

Insect cell media SF900 III SFM and Express Five® SFM were bought from Invitrogen. LB and 2 × YT are provided in 2.1.1.3

**Table 2.18:** Media for bacterial cells used for protein expression

Medium	Composition
ZY	10 g N-Z-amine AS, 5 g yeast extract, 950 ml of deionized water
ZYM-5052	950 ml ZY media supplemented with 2 ml MgSO <sub>4</sub> (1 M), 200µl trace elements (1000× solution), 20 ml of 50× 5052, 20 ml of 50× M
50 × 5052	25% glycerol, 2.5% glucose, 10% α-lactose
50 × M	1.25 M Na <sub>2</sub> HPO <sub>4</sub> , 1.25 M KH <sub>2</sub> PO <sub>4</sub> , 2.5 M NH <sub>4</sub> Cl, 0.25 M Na <sub>2</sub> SO <sub>4</sub>
1000 × Trace elements	50 mM FeCl <sub>3</sub> , 20 mM CaCl <sub>2</sub> , 10 mM MnCl <sub>2</sub> , 10 mM ZnSO <sub>4</sub> , 2 mM CoCl <sub>2</sub> , 2 mM CuCl <sub>2</sub> , 2 mM NiCl <sub>2</sub> , 2 mM Na <sub>2</sub> MoO <sub>4</sub> , 2 mM Na <sub>2</sub> SeO <sub>3</sub> , 2 mM H <sub>3</sub> BO <sub>3</sub> in ~60 mM HCl
M9 minimal medium	1L medium contains: 20 ml of 50× M, 20% glucose, 1 mM MgSO <sub>4</sub> , 0.3 mM CaCl <sub>2</sub> , 1 µg biotin, 1 µg thiamin, 1 ml trace elements (1000×)



## 2.2.2 Methods

### 2.2.2.1 Protein expression using bacterial cells

**IPTG induction** Recombinant proteins were individually expressed in *E. coli* strain BL21 (DE3) (all proteins, except eIF4B and eIF4G) or Rosetta 2 (DE3) (all proteins) grown in 2 × YT medium supplemented with appropriate antibiotic(s). Expression of target proteins was induced by addition of 0.25 mM isopropyl β-D-thiogalactopyranoside (IPTG) when an OD<sub>600</sub> of ~1.0 was reached. After induction, cells were cultured at 16 °C for 4-6 hours (for eIF4A and rpS9e) or overnight. The speed of the shaking was set up to 220 rpm.

**Auto-induction** In order to obtain the full-length yeast eIF4G protein in a reasonable amount, auto-induction was also tried [135]. Cells were grown in ZYM-5052 auto-inducing complex medium containing 100 mg/L ampicillin and 34 mg/L chloramphenicol at 16 °C for two days.

#### **Expression of unnatural amino acid Bpa labeled proteins**

Plasmids that harbor *sce*IF4A mutants were individually transformed into a special BL21 (DE3) *E. coli* strain (kind gift of Heinz Neumann, Göttingen), which hold a pSup-Bpa plasmid (resistant to chloramphenicol). pSup-Bpa encodes a suppressor tRNA and an engineered aminoacyl-tRNA synthetase that charges the suppressor tRNA with the unnatural amino acid Bpa (*p*-benzoyl-L-phenylalanine). The suppressor tRNA is designed to base pair with the amber stop codon (UAG) that would be genetically mutated from a residue of interest. Therefore, the unnatural amino acid Bpa is able to be site-specifically inserted into a target protein [136]. Cells were grown in LB medium with addition of 50 mg/L ampicillin and 17 mg/L chloramphenicol, as well as 1 mM Bpa (kind gift of Heinz Neumann, Göttingen) at 37 °C. When an OD<sub>600</sub> of ~0.6 was reached, cells were transferred to 20 °C. In order to induce the expression of target protein, 0.5 mM IPTG (final concentration) was added at an OD<sub>600</sub> of ~1.0. Cells were further incubated at 20 °C for ~5 hours.

**Expression of SeMet substituted protein** The selenomethionine (SeMet) derivative *cte*IF3b was overexpressed in *E. coli* strain B834 (DE3), which is methionine auxotrophic. The cells were initially grown in M9 minimal medium with

the supply of 50 mg/L methionine. When an OD<sub>600</sub> of ~0.6 was reached, this methionine-containing medium was then removed by centrifugation (2,500 × g, 15 min, 4°C) and the cells were resuspended in new M9 medium. After the depletion of the residual methionine, 50 mg/L selenomethionine was supplied. Target protein expression was induced by addition of 0.25 mM IPTG. Cells were incubated with shaking at 220 rpm at 16 °C overnight.

#### 2.2.2.2 Protein expression using insect cell system

Expression of a gene of interest using the Bac-to-Bac Baculovirus Expression System (Invitrogen) is far more complicated than expression in *E. coli* cells. It can be divided into following steps: (i) clone gene of interest into a chosen pFastBac<sup>TM</sup> donor plasmid; (ii) transformation of the pFastBac<sup>TM</sup> recombinant into DH10Bac<sup>TM</sup> *E. coli* host cells for generation of a recombinant bacmid; (iii) transfection of the isolated recombinant bacmid DNA into a insect cell line to propagate a recombinant baculovirus; (iv) Amplification and titration of the recombinant baculoviral stock, and infection of insect cells to produce the target protein.

**Cultivation of insect cells** All operation concerning insect cells was carried out under sterile conditions, using a laminar flow hood. Sf9 cells were cultured in SF900 III SFM medium, while High Five<sup>TM</sup> cells were grown in Express Five<sup>®</sup> SFM medium supplemented with 1× L-glutamine. Both cells were incubated at 27 °C in the dark. The suspension cultures were maintained in an appropriate flask, of which the capacity is usually four times bigger than the volume of the cells. The cells are shaking at 100 rpm and subcultured every day to a density of  $0.5 \times 10^6$  cells/ml. Adherent monolayer cultures were cultured in plates with 70% to 90% occupancy. Cell viability was assessed every day under an inverse microscope.

**Preparation of recombinant bacmid** The gene of yeast eIF4G was cloned into pFastBac<sup>TM</sup> HT B vector and transformed into DH10Bac<sup>TM</sup> *E. coli* cells. Following heat-shocking, cells were incubated at 37 °C for 3h in SOC medium for recovery and plated on LB-agar media that contain 100 µg/ml ampicillin, 50 µg/ml kanamycin, 10 µg/ml tetracycline, 7 µg/ml gentamicin, as well as 100 µg/ml Bluo-Gal and 1 mM IPTG. Cells were cultured for 2 days. White positive colonies

were picked and restreaked on a new LB-agar plate supplemented with all the antibiotics mentioned above, IPTG as well as Bluo-Gal. The blue-white screening was performed twice, and subsequently single white colonies were inoculated into 5 ml LB medium for amplifying the recombinant bacmid. Recombinant bacmid DNAs verified by PCR were isolated using a miniprep kit, omitting the column purification step. Instead of column purification, the large bacmid DNAs were precipitated by addition of 0.6 volume of isopropanol and pelleted by centrifugation ( $15,000 \times g$  for 30 min). Pellets were washed with 50  $\mu$ l of 70% ethanol.

**Transfection of Sf9 cells** The dried recombinant bacmid pellets were dissolved in 20  $\mu$ l autoclaved ddH<sub>2</sub>O and mixed with 200  $\mu$ l of SF900 III medium. The transfection reagent, FuGENE 6, was also diluted in SF900 III medium (10  $\mu$ l FuGENE 6 in 100  $\mu$ l medium) and added into the bacmid containing tube. The mixture was incubated for 45 min and equally distributed into two wells of a 6-well plate (each 165  $\mu$ l). Each well contains 3 ml Sf9 cells at a density of  $0.3 \times 10^6$  cells/ml. The plates were incubated at 27 °C protected against light for 60 hours, and then the supernatant containing the V<sub>0</sub> virus was collected.

**Virus amplification** The V<sub>0</sub> virus was used to infect a suspension culture of Sf9 cells (3 ml virus to 25 ml cells). Infected Sf9 cells were counted and diluted to  $0.6 \times 10^6$  cells/ml every 24 h until cell proliferation was stopped. After proliferation arrest,  $10^6$  infected cells were sampled every 12 h for preliminary protein expression test. When the density of the cells reached  $0.2\text{-}0.3 \times 10^6$  cells/ml, the V<sub>1</sub> virus present in the supernatant was harvested by centrifugation ( $2,000 \times \text{rpm}$ , 20 min) and stored at 4 °C. Protein expression was analyzed by SDS-PAG. V<sub>2</sub>, V<sub>3</sub> or V<sub>4</sub> virus was produced by infecting Sf9 cells with previous generation of virus. Normally, the ratio of virus to cells was 1 to 1000 during the amplification of virus.

**Protein expression in insect cell** To determine the appropriate concentration of virus to be used for large scale protein expression and the time scale of infection, 4 flasks that each contains 25 ml High Five™ cells at density of  $0.6\text{-}0.7 \times 10^6$  cells/ml were infected with varying volumes of virus (1:100, 1:200, 1:500, 1:1000). After 24 h infection, cells were checked, counted and sampled every 4 h. The amount of virus and the expression period that yielded high protein

production were chosen for large-scale expression.

To express the target protein (i.e. eIF4G) in large scale, each 500 ml suspension High Five™ cells cultured in a 2 L shaker flask at  $0.6\text{-}0.7 \times 10^6$  cells/ml were infected with 1 to 2 ml virus. The cells were harvest after 36 h infection, normally with cell viability of 90%.

### 2.2.2.3 SDS-PAGE

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyze proteins. Depending on the size of the proteins to be separated, 8-12.5% (v/v) acrylamide containing gel were prepared. The recipe for the gel solutions is shown in Table 2.19.

**Table 2.19:** Composition of separating gel (40 ml for 4 gels)

Item	8 %	10 %	12.5 %	15 %
1.5 M Tris Base/HCl 8.8	10 ml	10 ml	10 ml	10 ml
H <sub>2</sub> O	18.6 ml	15.9 ml	12.6 ml	9.3 ml
Rotiphorese Gel 30 (37.5:1)	10.7 ml	13.3 ml	16.7 ml	20 ml
10 % SDS	400 µl	400 µl	400 µl	400 µl
10 % APS	300 µl	300 µl	300 µl	250 µl
TEMED	40 µl	40 µl	40 µl	30 µl

**Table 2.20:** Composition of stacking gel (6%, 20 ml for 4 gels)

Item	6 %
0.625 M Tris Base/HCl 6.8	4 ml
H <sub>2</sub> O	11.6 ml
Rotiphorese Gel 30 (37.5:1)	4 ml
10 % SDS	200 µl
10 % APS	200 µl
TEMED	10 µl

Samples were mixed with laemmli loading buffer and heated for 2 min at 95 °C prior to loading onto the gels. Gels were run vertically in SDS-PAGE running buffer for ~1 h at a constant current of 40 mA. To visualize protein bands, SDS-PAGE gels were first stained with staining buffer and subsequently destained. The recipes of the buffers are provided in 2.2.1.3. Gels were scanned and documented.

#### **2.2.2.4 Protein concentration determination**

Protein concentrations were determined by performing the Bradford assay [137]. This assay is based on the absorbance shift of the dye coomassie Brilliant Blue G-250 (from 465 nm to 595 nm) if it binds to protein. 20 µl of protein solution was mixed with 980 µl of Bradford reagent (1:5 diluted with H<sub>2</sub>O). Before measuring the OD<sub>595</sub> of the sample using a previously calibrated photometer, the mixture was incubated at room temperature for 1 min. The concentration of the protein in mg/ml ranging from 0.1 to 0.9 was directly read from the monitor of the photometer.

#### **2.2.2.5 Western blot**

Proteins were separated by SDS-PAGE and wet-blotted onto a nitrocellulose membrane at 90 V for 2 h at 4 °C, using a blotting chamber (BioRad). The gel, membrane and filter papers were pre-wetted in cold transfer buffer prior to transfer. The materials were stacked in the following order:

- Case (black side) (-)
- Sponge
- Whatman paper
- Gel
- Membrane
- Whatman paper
- Sponge
- Case (white side) (+)

Air bubbles trapped between the layers of the assembly were carefully removed. In order to decrease the non-specific binding of antibodies, the membrane was first blocked in blocking buffer for 2 h with shaking. After three 10-minute washes, the

membrane was incubated with the primary antibody (mouse anti-His) 1:1000 diluted in blotting buffer overnight, 4 °C or 2-4 h at room temperature. The excess of the first antibody was washed away with TBST (three 10-minute washing). The membrane was then incubated with 1:5000 diluted alkaline phosphatase (AP) conjugated secondary antibody (goat anti-mouse AP) for 1 h at room temperature. After three 5-minute washes with buffer A, proteins were detected by directly dying. The membrane was incubated with 66 µl NTB stock, 33 µl BCIP stock and 10 ml buffer A for 2-5 min and washed with H<sub>2</sub>O.

#### 2.2.2.6 Cell harvest and lysis

Cells were harvested from expression cultures by centrifugation (4,800 × g, 20 min, 4 °C). The cell pellets were washed with 50 mM HEPES/NaOH, pH 7.5, 200 mM NaCl and then pelleted again for either immediate lysis or storage at -20 °C. The lysis buffer usually contained 20 mM HEPES/NaOH, pH 6.8-8.0, 500 mM NaCl, 2.5 mM β-mercaptoethanol (β-ME) and 5-10% glycerol. *E. coli* cells were ruptured using a Microfluidizer in which the cell suspension was pushed through a chamber at a high pressure of 0.55 MPa for five times. By contrast, insect cells were gently disrupted using an ultra centrifugal mill (Retsch ZM 200). The debris was separated via centrifugation (30,000 × g, 35 min, 4 °C). The supernatant containing the protein of interest was then loaded onto a column for purification.

#### 2.2.2.7 Protein purification

For GST tagged proteins the supernatant after cell rupture and centrifugation was loaded onto a GSTrap column (GE Healthcare) equilibrated in lysis buffer. Once loaded in the column, samples were extensively washed with lysis buffer supplemented with 1 M LiCl (especially for these RNA-binding proteins, *e.g.* eIF4A, eIF4B, eIF4G). Then the protein-bound column was equilibrated in a low salt buffer (25 mM HEPES/NaOH, pH 7.5, 100 (200 for eIF3g, eIF4G<sup>468-853</sup> and eIF4G<sup>486-853</sup>) mM NaCl, 5% glycerol and 5 mM β-ME). The GST fusion proteins were eluted with the low salt buffer supplemented with 30 mM reduced glutathione. The GST-tag

was cleaved with PreScission protease overnight at 4 °C, with a 1:100 mass ratio of protease to fusion protein.

For the His<sub>6</sub>-tagged proteins the supernatant was applied to an HisTrap column (GE Healthcare). After loading in the column, samples were extensively washed with lysis buffer containing additional 1 M LiCl. Then the protein-bound column was equilibrated in a low salt buffer (25 mM HEPES/NaOH, pH 7.5, 100 (200 for eIF3a) mM NaCl, 5% glycerol and 5 mM  $\beta$ -ME). The His<sub>6</sub> fusion proteins were eluted applying a linear gradient to the 400 mM imidazole-containing low salt buffer containing 400 mM imidazole. The His<sub>6</sub>-tag was cleaved with tobacco etch virus (TEV) protease overnight at 4 °C, with a 1:100 mass ratio of protease to fusion protein.

To further purify full-length eIF3b, eIF3g, eIF4A, eIF4B, eIF4G, as well as eIF4G<sup>468-853</sup> and eIF4G<sup>486-853</sup>, respectively, the protein was applied onto an ion exchange column (GE healthcare) (Source 30Q for eIF3b, eIF4A and eIF4B, SP Sepharose for eIF3g, full-length eIF4G, eIF4G<sup>468-853</sup> and eIF4G<sup>486-853</sup>). Prior to sample loading, the ion exchange column was equilibrated in buffer A (25 mM HEPES/NaOH, pH 7.5, 100 mM NaCl (200 mM NaCl for eIF3g, eIF4G<sup>468-853</sup> and eIF4G<sup>486-853</sup>), 5% glycerol and 5 mM  $\beta$ -ME). The target protein was eluted with a linear gradient of buffer B (25 mM HEPES/NaOH, pH 7.5, 1 M NaCl, 5% glycerol and 5 mM  $\beta$ -ME).

To further purify eIF3a the protein was loaded onto a HiTrap Heparin HP column (GE healthcare) that was equilibrated in buffer A (20 mM HEPES/NaOH, pH 7.3, 200 mM NaCl, 5% glycerol and 5 mM  $\beta$ -ME). eIF3a was eluted with a linear gradient of buffer B (20 mM HEPES/NaOH, pH 7.5, 1 M NaCl, 10% glycerol and 5 mM  $\beta$ -ME).

Finally, a Superdex 75 (26/60) or a Superdex 200 (26/60) gel filtration column (GE Healthcare) was used to purify all the proteins, including individual full-length subunits, truncations, as well as subcomplexes. The gel filtration buffer contained 10 mM HEPES/NaOH, pH 7.5, 100-150 mM NaCl, 2% glycerol and 2 mM DTT. Fractions that contained the target proteins were pooled, concentrated, flash-frozen in liquid nitrogen and stored at -80 °C.

## **2.3 Protein characterization**

### **2.3.1 Thermal shift assay**

Thermal shift assay, also called differential scanning fluorimetry (DSF), was used to choose conditions in which the target protein is stable over long periods of time. ~5  $\mu$ M purified protein supplemented with 1  $\times$  SYPRO orange fluorescent dye was pipetted into a 96-well plate (20  $\mu$ l in each well). The temperature controlled by a thermocycler was increased from 25  $^{\circ}$ C to 95  $^{\circ}$ C, and the fluorescence emission was read in steps of 1  $^{\circ}$ C/min with 30 sec hold between reads. The fluorescence intensity was plotted as a function of temperature. Different buffers (varying the chemical composition and pH) and salt concentration were tested for their ability to stabilize the target protein. The curve obtained directly from each condition of the experiment was normalized and adjusted against the background signal of the fluorophore in the buffer. The melting temperatures ( $T_m$ ) reflected by the inflection points of the curves were compared, and the reasonable condition with the highest  $T_m$  was chosen.

### **2.3.2 Multi-angle light scattering combined with size exclusion chromatography**

Multi-angle light scattering (MALS) combined with size exclusion chromatography was applied to determine the absolute molar mass of the protein of interest. ~50  $\mu$ g protein sample in a volume of 400  $\mu$ l was injected onto an analytical Superdex 200 (10/300) column (GE healthcare), which was equilibrated with 25 mM HEPES/NaOH, pH 7.5, 100 mM NaCl. The elution from the column was monitored by a multi-angle light scattering detector, Wyatt miniDAWN TREOS. The data were analyzed using supplied ASTRA6 software.



### 2.3.3 Limited proteolysis

The flexible regions of the target proteins (*e.g.* exposed loops, disordered termini, or dynamic domain linkers) usually impair the formation of crystals. Removing such flexible parts from a protein using limited proteolysis results in more compact and conformationally homogeneous molecules or compact single domains. Hence, the crystallization becomes possible and easier, as well as the crystals diffract better.

Different proteases (*e.g.* chymotrypsin, trypsin, thermolysin, Proteinase K, gluC) were tested for each protein or complex of interest. Enzyme/substrate ratio (1:500, 1:1000, 1:2500, 1:5000), temperature (4 °C and 20°C) and time of digestion (5 min, 20 min, 1h, 2h, 4h, 6h, 1d, 2d, 4d) were varied to find the best-digested product. ~7 µg protein was used for each reaction in gel filtration buffer. The reactions were stopped by addition of SDS-PAGE loading buffer. The digested samples were evaluated by SDS-PAGE.

### 2.3.4 Electrophoretic mobility shift assay (EMSA)

In each reaction, 1 µg RNA oligonucleotides were mixed with 5-fold molar excess of protein(s). Before loading onto a 2 % agarose gel, the mixture was incubated at 4 °C for 1 h. Agarose gel electrophoresis was performed as described in 2.1.2.3.

### 2.3.5 GST pull-down

GST pull-down assays were performed as described in Liu Y et al., 2014 [138]. 50 µg GST fusion protein was mixed with 2-fold molar excess of non-tagged protein in a buffer containing 25 mM HEPES/NaOH, pH 7.5, 100 mM NaCl, 5% glycerol and 2.5 mM β-ME and incubated with 100 µl glutathione beads for 30 min. After washing four times with 1 mL buffer, bound protein was eluted with the same buffer supplemented with 30 mM reduced glutathione.

### 2.3.6 Cobalt pull-down

50 µg of His-tagged protein was mixed with 2-fold molar excess of non-tagged protein(s) in a buffer that contained 25 mM HEPES/NaOH, pH 7.5, 100 mM NaCl, 5% glycerol and 2.5 mM β-ME. The mixture was incubated with 100 µl of cobalt charged TALON (Clontech) beads for 30 min. After washing with 1 mL buffer for four times, bound protein was eluted with the same buffer supplemented with 500 mM imidazole.

### 2.3.7 Analytical gel filtration analysis

An analytical Superdex 200 (10/300) column (GE healthcare) was used for *in vitro* protein interaction test. In each case, ~50 µg protein in a volume of 400 µl was injected on the column equilibrated to the buffer that contained 25 mM HEPES/NaOH, pH 7.5, 100-150 mM NaCl, 5% glycerol and 2.5 mM β-ME.

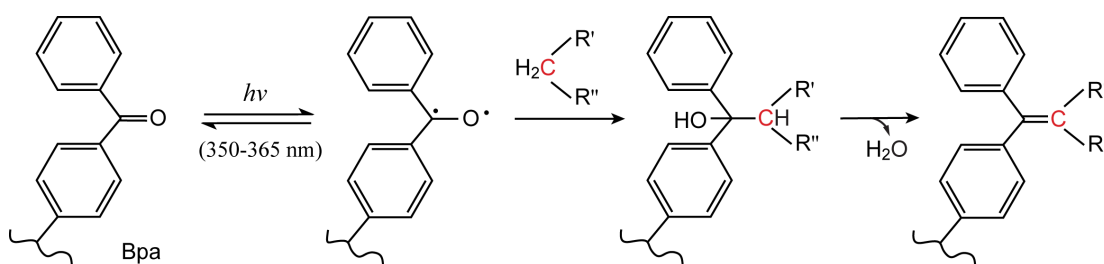
### 2.3.8 Co-sedimentation

Interaction between 40S ribosomal subunit and full-length eIF3b or its truncations was tested by sedimentation experiment (in collaboration with S. Schell and A. Chari). Purified 40S was incubated with 4, 8, 12, 16, 20, or 5, 15-fold molar excess of eIF3b proteins for 30 minutes at room temperature. Mixtures were then loaded on a 7.5% sucrose cushion and centrifuged through a 30% sucrose cushion with a 100,000 rpm speed for 2 hr at 4 °C using a MLA-130 rotor (Beckman). Supernatants were carefully removed, and the pellets were resuspended. Both sucrose cushions and resuspension buffer contained 50 mM KCl, 50 mM BisTris, pH 6.8, 10 mM MgCl<sub>2</sub>, and 5 mM DTT.

### 2.3.9 Crosslinking

Unnatural amino acid *p*-benzoyl-L-phenylalanine (Bpa) was site-specifically incorporated into *sceIF4A* (see Section 2.2.2.1). Upon UV irradiation, Bpa forms a

diradical and reacts with nearby C-H bonds (in 3 Å radius) [139] (Figure 2.1). Crosslinking was performed in a self-made 36-well metal plate on ice in buffer containing 20 mM HEPES/NaOH, pH 7.5, 100 mM NaCl, 2 mM MgCl<sub>2</sub> and 1 mM DTT. The crosslinking reactions were induced by high intensity long-wave (365 nm) ultraviolet light (Vilber Lourmat) for 5 min and 30 min. ~3 µg of protein was used per reaction. The samples were separated by SDS-PAGE.



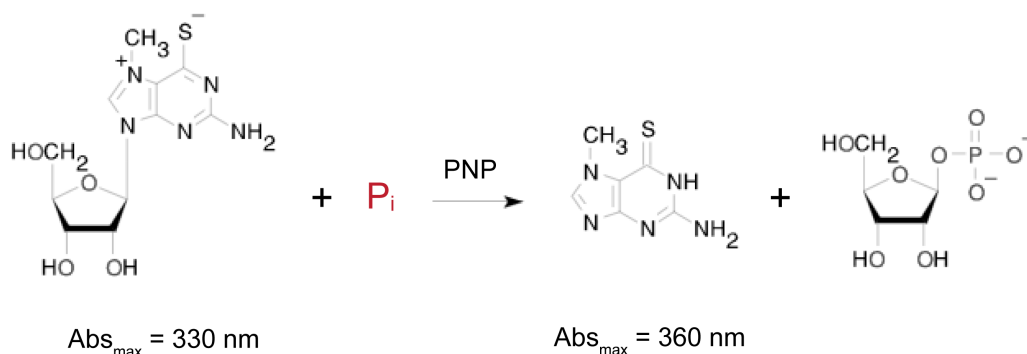
**Figure 2.1:** Structure of Bpa and the mechanism of the UV0induced crosslinking. Upon UV exposure, Bpa forms a diradical and further inserted into a C-H bond within 3 Å radius.

### 2.3.10 Characterization of ATPase activity

The ATPase activity of eIF4A was determined using the *EnzCheck*<sup>®</sup> Phosphate Assay kit (Invitrogen). When inorganic phosphate (P<sub>i</sub>) is present in solution, 2-amino-6-mercapto-7-methypurine riboside (MESG) is converted to ribose 1-phosphate and 2-amino-6-mercapto-7-methypurine, which is catalyzed by purine nucleoside phosphorylase (PNP) (Figure 2.2). The accompanying change in absorption from 330 nm (substrate) to 360 nm (product) enables to quantify the free phosphate ion consumed in the reaction. Thus the increase of absorption at 360 nm is directly proportional to the rate of dephosphorylation reaction that stimulated by an ATPase [140].

1 µM eIF4A was mixed with 1 µl PNP, 20 µl MESG as well as 5 µM RNA into a buffer that contains 20 mM HEPES/NaOH, pH 7.0, 100 mM NaCl, 2 mM MgCl<sub>2</sub>. The total volume of the reaction was 100 µl. Mixtures were incubated at 20 °C for 10 min and measurements were started by addition of ATP to a final concentration ranging from 0 to 1 mM. The absorbance at 360 nm was recorded every two seconds

for 10 min, using an Ultrospec 2100 pro UV/visible spectrophotometer (GE Healthcare).



**Figure 2.2:** Mechanism of the EnzCheck® Phosphate Assay kit. Purine nucleoside phosphorylase (PNP) catalyzes 2-amino-6-mercapto-7-methylthio-7H-purin-6-yl-ribose (MESG) to 2-amino-6-mercapto-7-methylthio-7H-purine and ribose-1-phosphate. The accompanying absorption maximum change, from 330 nm to 360 nm, allows the quantitation of inorganic phosphate released from an ATP hydrolysis reaction.

The initial velocity ( $V_0$ ) of a reaction was determined by measuring the slope of the absorption at very beginning time point. In order to reach the velocity maximum ( $V_{\text{max}}$ ), the concentration of the substrate (ATP) was titrated and each concentration was measured independently for three times. Negative controls without ATP or without protein were performed to assure that the signal origin was due to the ATPase activity of the protein. The amount of consumed and thus released inorganic phosphate was calculated according to the standard absorption curve of phosphate. The Michaelis-Menten equation was employed to determine the kinetic parameters, i.e. the Michaelis constant  $K_M$  and the maximal velocity  $V_{\text{max}}$ . This calculation was carried out using SigmaPlot (Systat Software Inc.). The catalytic constant ( $k_{\text{cat}}$ ) equals  $V_{\text{max}}/\text{enzyme concentration}$ .

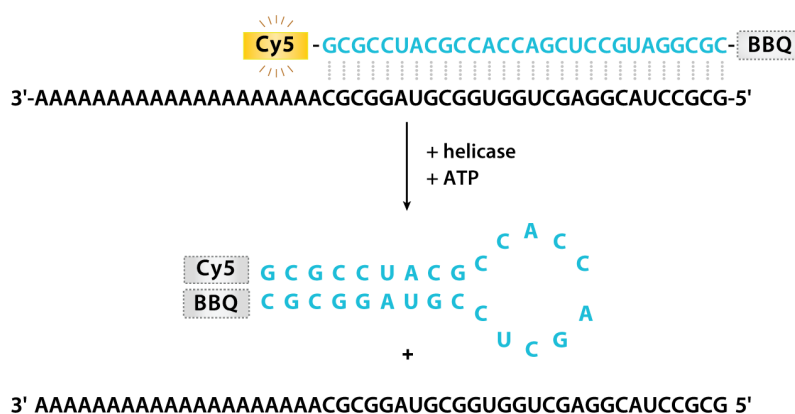
The Michaelis-Menten equation:

$$V_0 = \frac{V_{\text{max}} [S]}{K_m + [S]}$$

[S]: concentration of the substrate

### 2.3.11 Characterization of helicase activity

The helicase activity of eIF4A was primarily tested based on a principle published by Belon and Frick [141]. Two RNA strands (from Henning Christian, our group), one of which was labeled with a fluorescent dye Cyanine 5 (Cy5) at the 5'-end and a Black Berry Quencher 650 (BBQ) at the 3'-end, were annealed together. Upon excitation at 643 nm, Cy5 emits light at 667 nm. If a strand separation of the annealed RNA occurs, the labeled RNA oligonucleotide would form a stable hairpin structure in which the fluorophore Cy5 becomes very close to the quencher BBQ, thus the fluorescent signal is reduced. Hence, the rate of the signal decrease is directly proportional to the rate of strand separation stimulated by the RNA helicase. The sequence of the two RNA oligonucleotides and the mechanism of this reaction are provided in Figure 2.3.



**Figure 2.3:** Mechanism of the helicase assay. The dsRNA substrate consists of a non-labeled strand (black) and a Cy5- and BBQ-labeled strand (cyan). The strand separation causes the labeled strand to form a stable hairpin structure in which the dye Cy5 gets close to the quencher BBQ and thus quenches the signal.

0.1  $\mu$ M eIF4A was mixed with 0.5  $\mu$ M RNA into buffer containing 20 mM HEPES/NaOH, pH 7.0, 100 mM NaCl, 2 mM  $MgCl_2$ , with total volume of 150  $\mu$ l. Mixture was incubated at 20  $^{\circ}$ C in the dark for 15 min and measurement was started by addition of ATP to a final concentration of 1 mM. The excitation wavelength was set to 643 nm and the fluorescence change at 667 nm was recorded every two seconds for 10 min.

## 2.3.12 RNA sequences

**Table 2. 21:** Sequences of RNAs used in this thesis

Name	Sequence	Company
poly(A) <sub>5</sub>	pAAAAA	IBA
poly(A) <sub>10</sub>	pAAAAAAAAAA	IBA
poly(A) <sub>15</sub>	pAAAAAAAAAAAAAA	IBA
poly(A) <sub>20</sub>	pAAAAAAAAAAAAAAAAAAAA	IBA
ssRNA <sub>15161</sub>	pUGAGGUAGUAGGUUGUAUAGUU	Roche
ssRNA <sub>15162</sub>	pUUGUCUCUGGUCCUUACUU	Roche
ssRNA <sub>a-eGFP</sub>	pCAGGUAGUGGUUGUCGGGCAG	Roche
slRNA <sub>AHC1</sub>	p(Cy5)-GCGCCUACGCCACCAGCUCCGUAGGCGC-(BBQ)	IBA
slRNA <sub>AHC2</sub>	pGCGCCUACGGAGCUGGUGGCGUAGGCGCAAAAAAAAAA AAAAAAAAAAAAA	IBA
dsRNA	pCAGGUAGUGGUUGUCGGGCAG GAGUCCAUCACCAACAGCCCGp	Roche

## 2.4 Crystallization and data collection

### 2.4.1 Initial crystallization screens

Initial screening of crystallization conditions was performed using sitting drop vapor diffusion technique. 0.25 µl of proteins were pipetted into 96-3-well MRC plates (Jena Bioscience) and mixed with either 0.25 µl (1:1) or 0.125 µl (2:1) of the reservoir solution using a Phoenix RE pipetting robot (Art Robbins Instruments). The volume of the reservoir is 35 µl. In total, 11 commercially available screens (listed in Table 2.22) at 4 °C and 20 °C have been tested for each protein or complex.

**Table 2.22:** Initial crystallization screens. Each screen contains 96 different conditions

Name	Supplier
Ammonium sulfate Suite	Qiagen
JBScreen Classic (1, 2, 4, 5)	Jena BioScience
JBScreen Classic (6, 7, 8, 10)	Jena BioScience
JBScreen Nuc-Pro HTS (1, 2, 4, 5)	Jena BioScience
JCSG <sup>+</sup> Suite	Newman <i>et al.</i> (2005) [142]
MIDAS	Molecular Dimensions
MIDAS diluted (0.5×)	Molecular Dimensions
Morpheus	Molecular Dimensions
Natrix HT	Hampton Research
PGA	Molecular Dimensions
Proplex	Radaev <i>et al.</i> (2006) [143]

## 2.4.2 Optimization of crystals

Initial hits were usually optimized manually. Commercial “Additive Screen” (Hampton Research), “Silver Bullets Screen” (Hampton Research) as well as “pH buffer screen” (Emerald Bio) were routinely tested to improve crystallization conditions. The concentration of the precipitants and the their types were also varied.

## 2.4.3 Data collection

Crystals were fished with appropriate nylon loops. Prior to being exposed to X-ray, they were transferred into the cryo-protectant buffer (typically crystallization condition supplemented with 20-30 % (v/v) of glycerol or ethylene glycol) for 5 seconds to 1 min and then flash cooled into liquid nitrogen. Diffraction of crystals was tested at a laboratory X-ray source (X-ray diffractometer MicroMax<sup>TM</sup>-007 HF,

Rigaku), and those giving highest resolution were selected for data collection in synchrotron. The exposure time for testing experiments was usually 15 min and with 1° rotation. Diffraction datasets were collected at 100 K at beamline P13, PETRA III, DESY (Hamburg, Germany) and at beamline 14.1, BESSY II (Berlin, Germany, (Mueller *et al.* 2012)).

## 2.5 Structure determination and refinement

Diffraction data of *C. thermophilum* eIF3b were processed and scaled using the programs XDS and XSCALE [144]. Phase calculations and determination were carried out by using the SHELX suite [145]. Se positions were further refined using SHARP [146] followed by density modification by Solomon [147]. An initial poly-alanine model was built by ARP/wARP [148] and refined against Se-Met data with Refmac5 [149,150], which was used for manual rebuilding and sequence assignment in Coot [151]. Due to the lack of isomorphism between Se-Met and native crystals (the c axis differs in length by 14.7 Å, see Table 3.1), this initial model of the *cteIF3b* WD40 domain was positioned by molecular replacement into the unit cell of the native crystal. Further model optimization and completion has been performed against the native data. Prior to structural refinement, randomly selected 5% test set of the reflections were set aside for the calculation of  $R_{free}$  as a quality monitor [152,153]. Refinement was performed with the PHENIX package [154]. The electrostatic surface potential was calculated with PDB2PQR [155] and displayed using the APBS plugin [156] of the PyMol software.

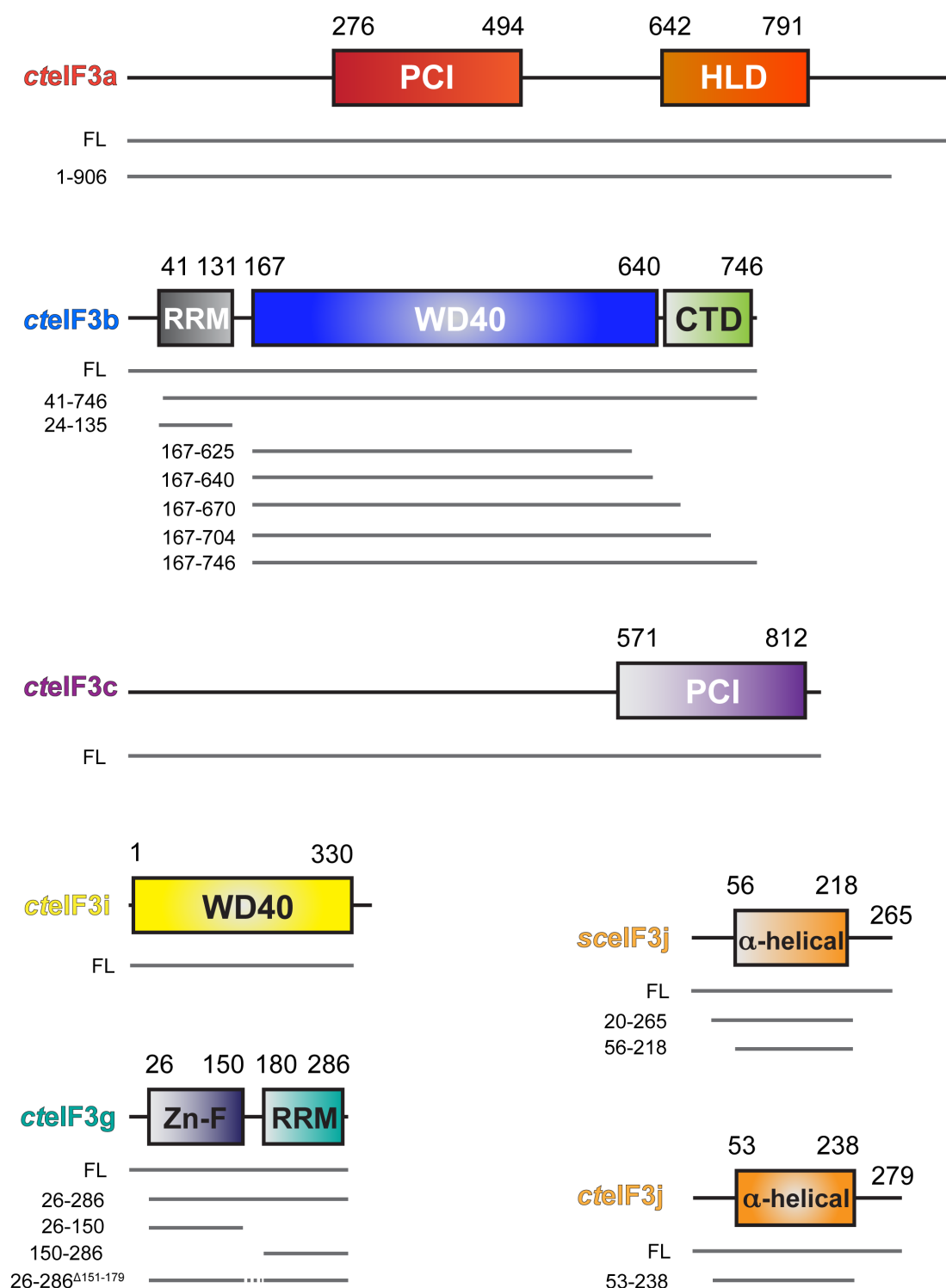


## 3

## Results

### 3.1 Eukaryotic translation initiation factor 3 (eIF3)

Depending on the organism eukaryotic translation initiation factor 3 (eIF3) contains in up to 13 non-identical subunits (eIF3a-m). In this thesis, only the simpler five-subunit functional core complex, consisting of eIF3a, b, c, g and i, has been mainly investigated. All the five subunits are universally conserved and essential for translation *in vivo* [42]. In addition to the functional core, the subunit j (eIF3j), which is involved in both translation initiation and post-termination, has also been extensively studied. All constructs (including full-length and truncations) of eIF3 subunits designed and studied in this thesis are summarized in Figure 3.1. Notably, besides eIF3 subunits from *Saccharomyces cerevisiae*, *Chaetomium thermophilum* proteins were chosen as the objects of study. The reason for this is described in detail in Section 4.1. Generally, it is observed that thermophilic proteins are more stable and easier to crystallize compared to their mesophilic orthologs.

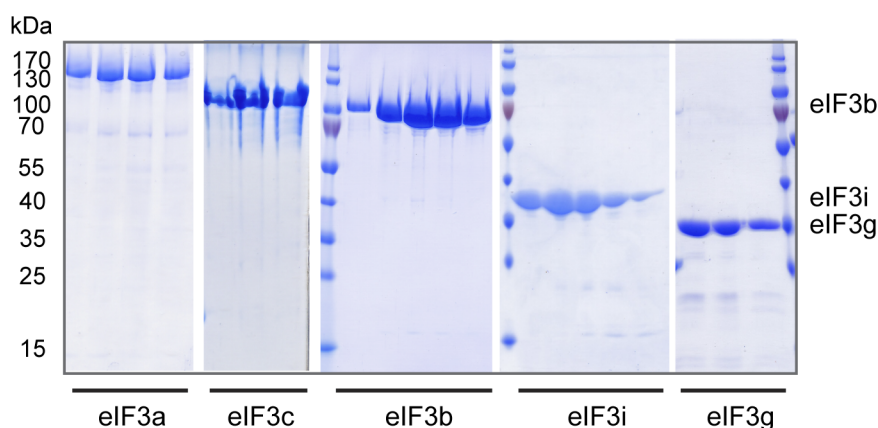


**Figure 3.1:** Diagram illustrating the designed constructs of eIF3 subunits. The boundaries of the constructs are indicated. Constructs design is based on secondary structure prediction, sequence conservation and functional domains.

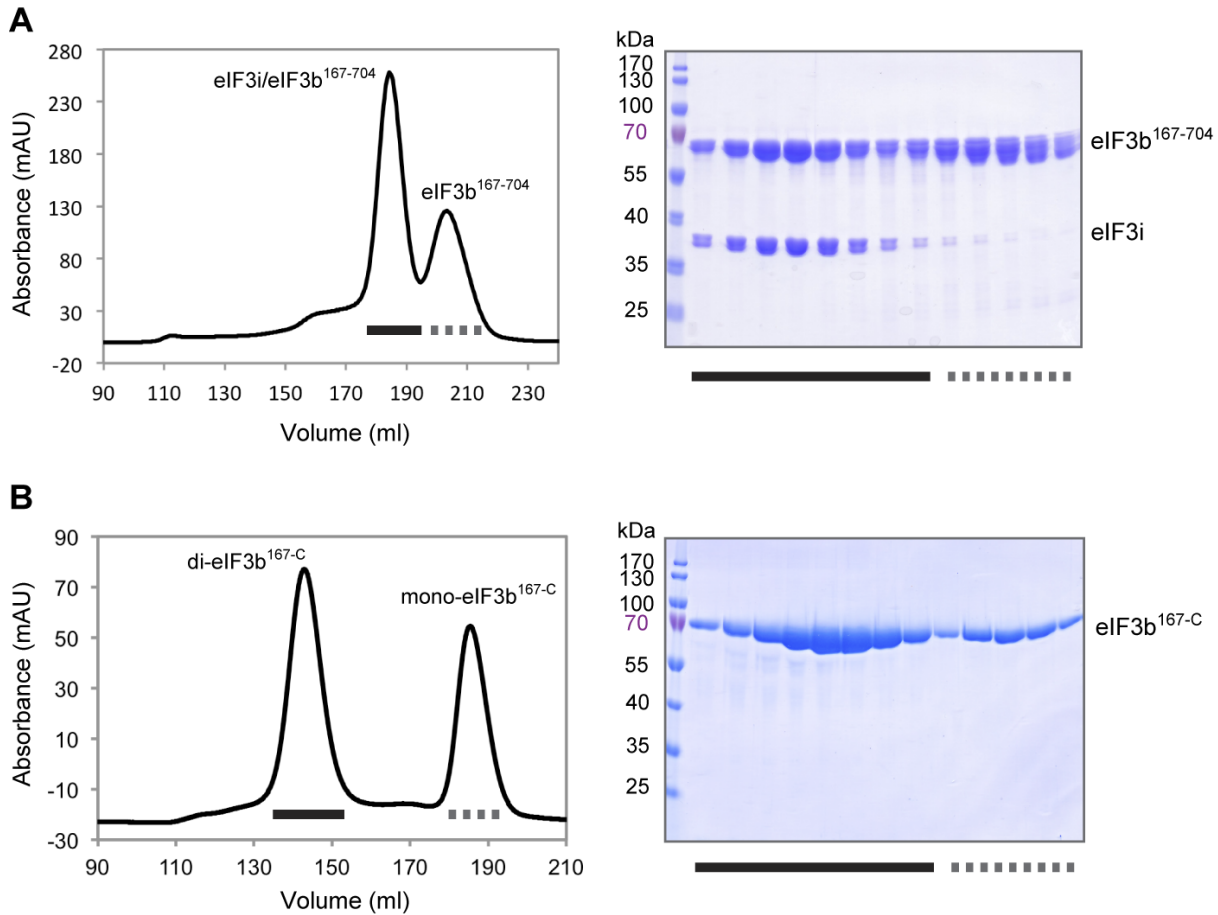
### 3.1.1 Preparation of eIF3 subunits

#### 3.1.1.1 Purification of eIF3a, b, c, g and i

All subunits of *C. thermophilum* eIF3 (*cteIF3*) core and their variants (Figure 3.1) (except eIF3g<sup>26-150</sup> and eIF3g<sup>26-286, Δ151-179</sup>), of which the affinity tag(s) are listed in Table 2.12, could be overexpressed individually in *E. coli* BL21 (DE3) and Rosetta 2 (DE3) cells. They could also be purified independently to high purity with the exception of eIF3a and eIF3b<sup>167-704</sup> (Figure 3.2, 3.3) (for detailed protein purification strategies for each protein refer to 2.2.2.7). eIF3a requires eIF3c or/and eIF3b to pass through the gel filtration column, while eIF3b<sup>167-704</sup> needs eIF3i to get rid of the degraded products (Figure 3.3 A). Interestingly, a truncated eIF3b (eIF3b<sup>167-746</sup>) forms a dimer (Figure 3.3 B), whereas full-length eIF3b and its other truncations behave as a monomer as revealed by the size exclusion chromatography (SEC). All other subunits exist as monomers in solution. The yield for eIF3a and eIF3c was ~2-3 mg protein per liter of cell culture, whereas ~10, ~20, ~50 mg protein per liter of culture could be obtained in the case of eIF3g, eIF3b and eIF3i, respectively.



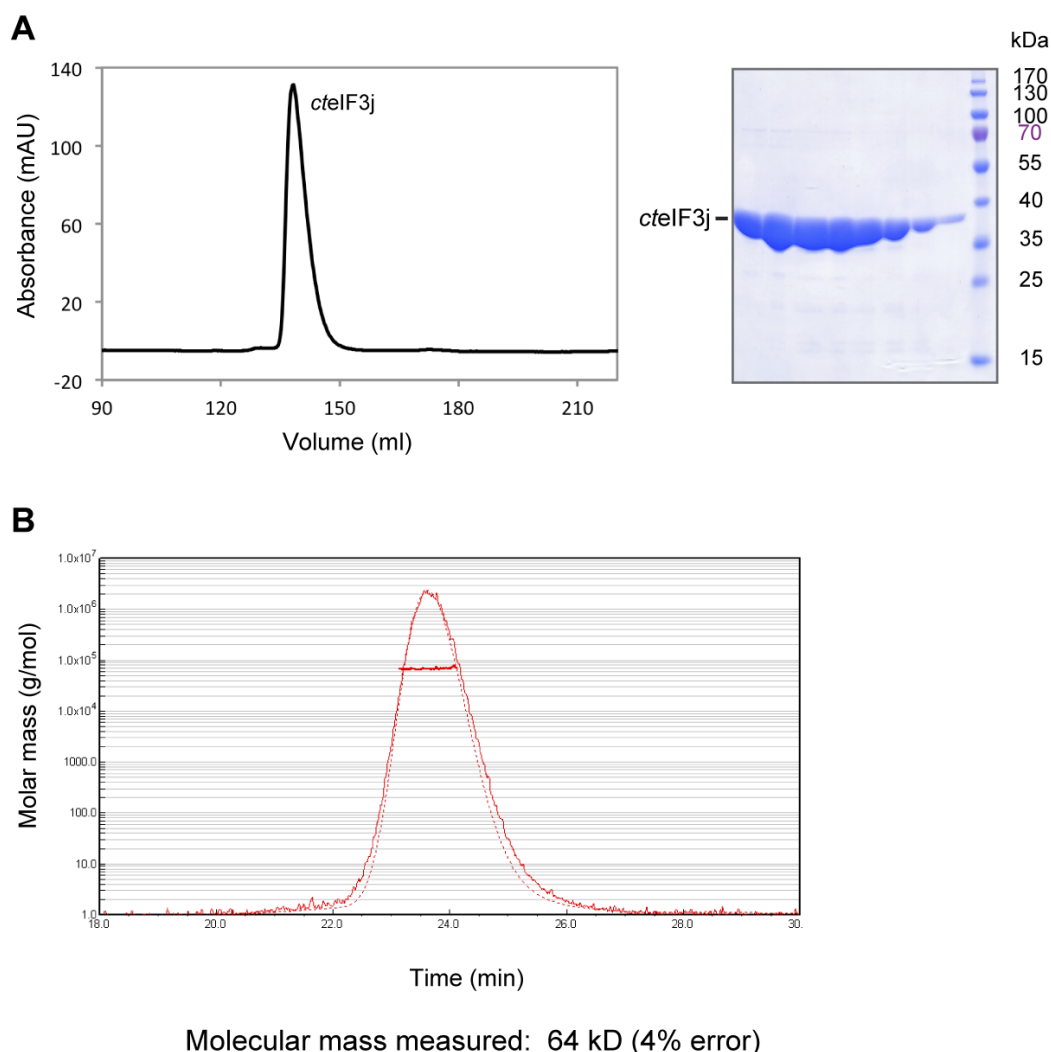
**Figure 3.2:** Coomassie-stained SDS gels show full-length eIF3b, c, g and i subunits after the gel filtration (as the last purification step) and full-length eIF3a from Heparin column. Before the gel filtration, subunits b, c, g and i were independently purified using a GSTrap affinity column followed by a Source 30Q for eIF3b and SP Sepharose for eIF3g. Prior to Heparin, eIF3a was passed through a HisTrap column. eIF3a and eIF3c were purified by Bernhard Kuhle.



**Figure 3.3:** Purification of eIF3b<sup>167-704</sup>/eIF3i complex (**A**) and eIF3b<sup>167-C</sup> (**B**). The left panels show the profile of the Superdex 200, while the right panels show the corresponding Coomassie-stained SDS gels. eIF3i only binds to the stable and complete eIF3b<sup>167-704</sup> but not the degraded eIF3b<sup>167-704</sup>. eIF3b<sup>167-C</sup> behaves as a dimer. However, the dimerization did not occur in the case of full-length eIF3b and all other truncations.

### 3.1.1.2 Purification of eIF3j

Both *S. cerevisiae* eIF3j (*sceIF3j*) and *C. thermophilum* eIF3j (*cteIF3j*) were overexpressed in *E. coli* cells. After the GStrap affinity column, a Source 30Q ion exchange chromatography was necessary to remove impurities. *sceIF3j* (29.6 kDa) and *cteIF3j* (30.8 kDa) eluted at the volume of ~140 ml on the Superdex 200 (26/60) column, corresponding to a molecular weight of ~240 kDa (Figure 3.4 A). This indicates that eIF3j orthologs behave as octamers in solution. However, the multi-angle light scattering (MALS), which is more precise than SEC, demonstrated that eIF3j orthologs form dimers in solution (Figure 3.4 B).

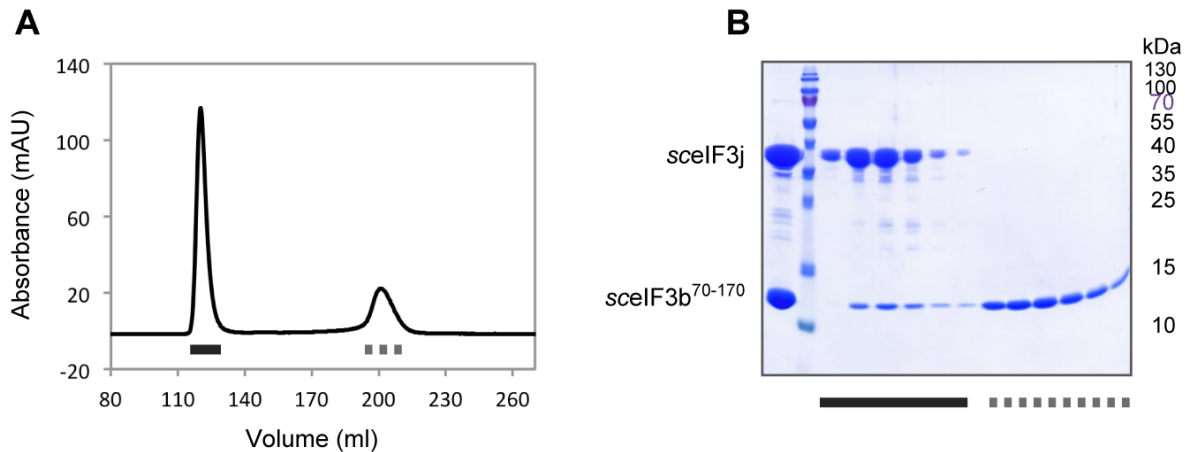


**Figure 3.4:** Purification of eIF3j. (A) Size exclusion chromatography (SEC) of *ctelF3j*. eIF3j orthologs behave as octamers on the Superdex column. The corresponding Coomassie-stained SDS gel is shown on the right panel. (B) Multi-angle light scattering (MALS) of *ctelF3j*, identifying 64 kDa (dimer size) molecular mass.

### 3.1.1.3 Formation of eIF3b-RRM/eIF3j complex

The interaction between eIF3b and eIF3j has been determined by biochemical probes and structural studies [47,157]. The complex structure of human eIF3b-RRM (residues 197-274) and eIF3j (residues 45-55) revealed that the unstructured eIF3j loop binds to the  $\alpha$ -helical surface of eIF3b's RRM domain. In some cases, proteins in complexes with their interacting partners could be easier crystallized. Thus the minimal complex composed of *S. cerevisiae* eIF3b-RRM domain (residues 70-170) and *sceIF3j* was reconstituted *in vitro*. The formation of eIF3b<sup>74-170</sup>/eIF3j complex is

salt-dependent as less than 200 mM NaCl resulted in heavily precipitation, but more than 300 mM NaCl caused dissociation. Nevertheless, this complex could be assembled *in vitro* (Figure 3.5).



**Figure 3.5:** Purification of *sceIF3b*-RRM/*sceIF3j* complex. The left panel shows the profile of the SEC, while the right panel shows the corresponding Coomassie-stained SDS gel. Prior to loading onto the Superdex 75 column (26/60), purified *eIF3j* and *eIF3b*<sup>74-170</sup> were mixed with a molar ratio of 1:1 and incubated for 30 min.

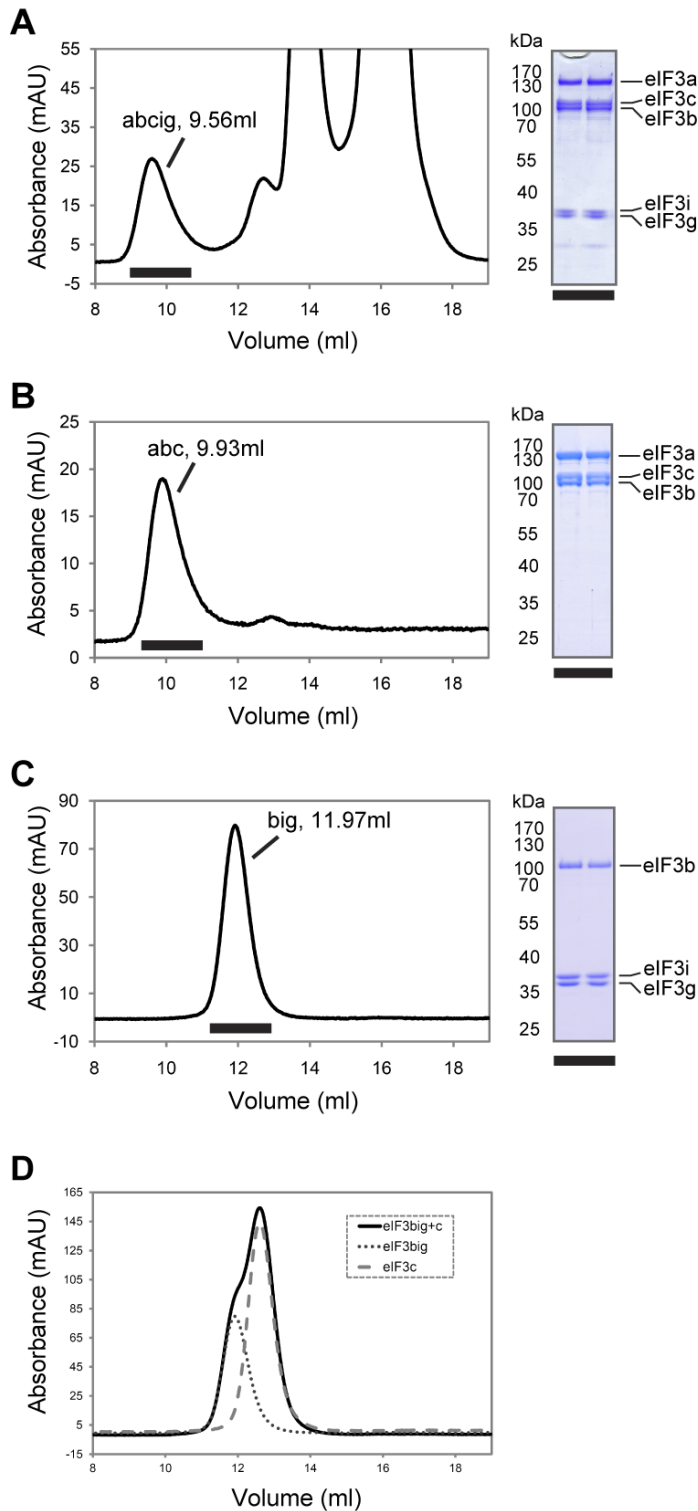
### 3.1.2 Interaction studies and complexes formation

#### 3.1.2.1 Reconstitution of *eIF3* functional core complex

Once proteins of individual *C. thermophilum* *eIF3* subunits were obtained, the formation of *eIF3* core complex and of its subcomplexes were tested. Subunits a, b, c, g and i of *eIF3* form a stable complex as shown by means of analytic SEC (Figure 3.6 A). The five-subunit complex eluted at 9.56 ml from the Superdex 200 (10/300), corresponding to a molecular weight of ~450 kDa. In addition, two distinct *eIF3b*-containing subcomplexes, *eIF3a*-*b*-*c* and *eIF3b*-*g*-*i*, have been assembled *in vitro* (Figure 3.6 B, C), indicating *eIF3b* serves as the major scaffold protein. However, in contrast to yeast *eIF3* [59], a stable *eIF3b*-*c*-*i*-*g* subcomplex could not be reconstituted *in vitro* from *C. thermophilum* proteins (Figure 3.6 D).

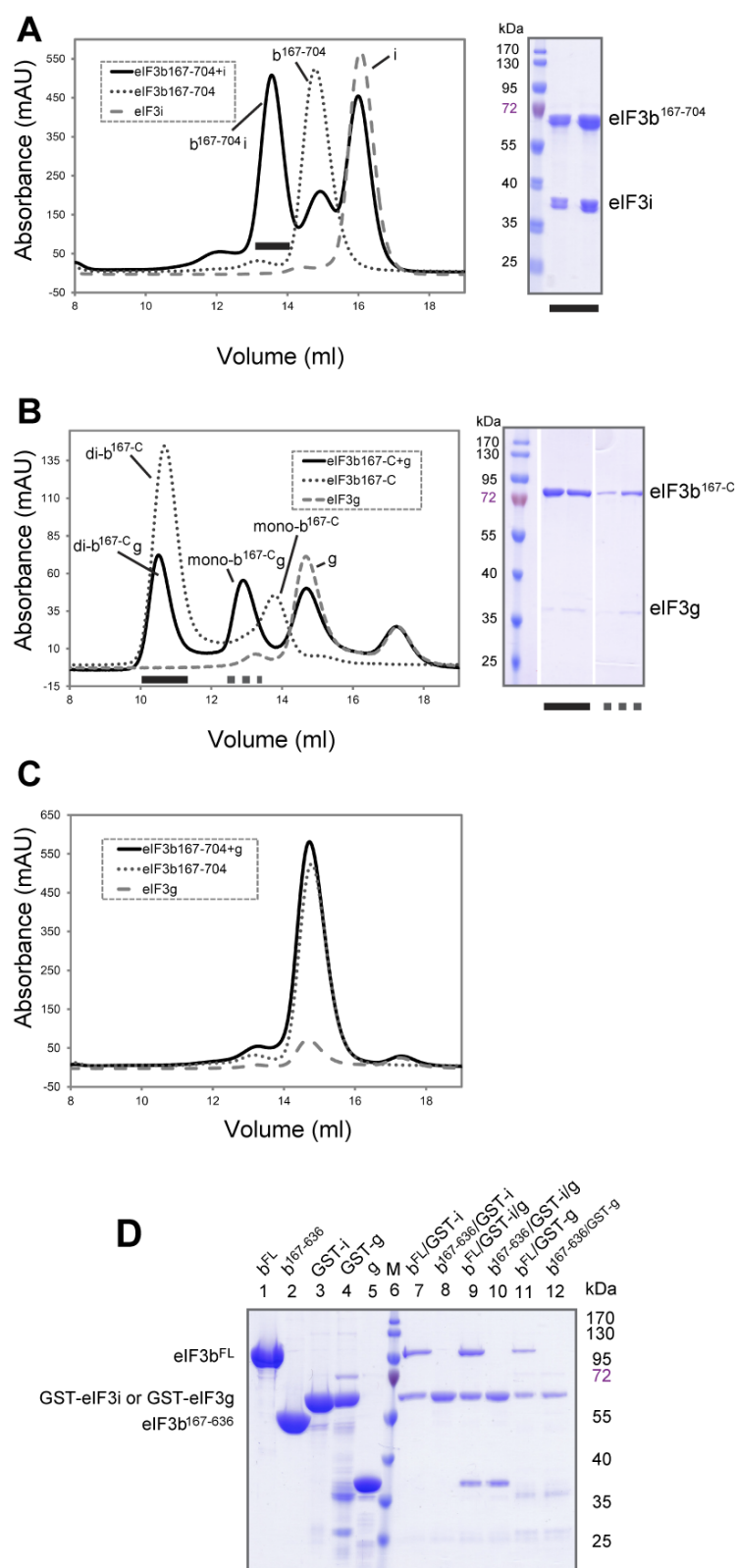
### 3.1.2.2 Mapping the interacting regions in eIF3b-g-i subcomplex

It has been demonstrated that the C-terminal ~100 amino acids of eIF3b are required for the interaction with eIF3g and i [44,48]. Associations with eIF3g and i are mutually beneficial rather than mutually exclusive, which indicates that the binding sites of eIF3b for eIF3g and i are not the same [48,59]. The crystal structure of yeast eIF3i and a C-terminal ~40-residue  $\alpha$ -helix of eIF3b determined the residues for the interaction with eIF3i [48]. However the knowledge of the specific binding region for eIF3g is particularly lacking. Thus, different constructs were designed and subcloned (Figure 3.1). eIF3b<sup>167-704</sup> is able to bind to eIF3i (Figure 3.7 A) but not to eIF3g (Figure 3.7 C) as reflected by the analytical SEC. In order to associate with eIF3g, the last 42 residues of eIF3b (eIF3b<sup>167-C</sup>) are necessary (Figure 3.7 B). However, the interaction between eIF3b<sup>167-C</sup> and eIF3g seems to be not really strong, since only weak eIF3g bands were observed in the SDS gel (Figure 3.7 B). Nevertheless, it is clear that the eIF3g binding site of eIF3b is located sequentially behind the binding site for eIF3i. Additionally, the GST pull-down experiment showed that the WD40 domain of eIF3b (eIF3b<sup>167-C</sup>) could neither interact with eIF3i (Figure 3.7 D, lane 8), eIF3g (Figure 3.7 D, Lane 12) nor eIF3i/eIF3g complex (Figure 3.7 D, lane 10).



**Figure 3.6:** Analytical size exclusion chromatography profiles of the five-component *cteIF3* core (A), the *cteIF3a-b-c* subcomplex (B), and the *cteIF3b-i-g* subcomplex (C), respectively. The corresponding SDS-PAGE gels are shown on the right panel. (D) Overlay of the analytical SEC profiles of the *cteIF3b-i-g* subcomplex (dotted black), *cteIF3c* (dashed gray) and the *cteIF3b-i-g* subcomplex together with subunit c (solid dark). In case of *cteIF3c* + *cteIF3b-i-g*, two overlapping peaks are observed; the main peak at a higher elution volume corresponds to eIF3c alone, whereas the smaller peak at a lower elution volume (forming the shoulder of the eIF3c peak) corresponds to the *cteIF3b-i-g* complex. A shift to a lower elution volume does not occur, indicating that the complex *cteIF3b-i-g-c* is not formed. Complex eIF3a-b-c was assembled by Bernhard Kuhle.





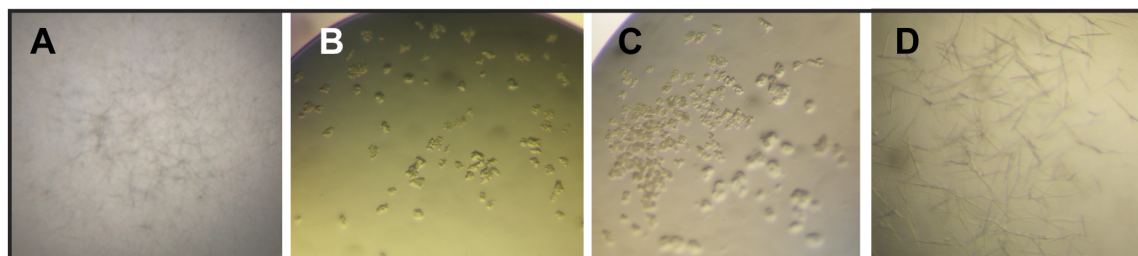
**Figure 3.7:** Mapping the interacting region of eIF3b for eIF3i and eIF3g. **(A)** Overlay of the analytical SEC profiles of *cteIF3b*<sup>167-704</sup> (dotted black), *cteIF3i* (dashed gray) and their mixture (solid dark). Compared with individual proteins, a significantly shifted peak (towards small elution volume) was obtained, suggesting the formation of the complex *eIF3b*<sup>167-704</sup>-*eIF3i*. On the right panel an SDS gel of the complex in the shifted peak is shown. **(B)** Overlay of the analytical SEC profiles of *eIF3b*<sup>167-C</sup> (dotted black), *eIF3g* (dashed gray) and their mixture (solid dark). Instead of one single peak, *eIF3b*<sup>167-C</sup> shows two peaks of which one at ~10.7 mL corresponds to the dimeric protein and another at ~13.8 mL represents the monomeric *eIF3b*<sup>167-C</sup>. In the case of the mixture, the peak shift is obvious for *mono-eIF3b*<sup>167-C</sup>/*g* complex but not for *di-eIF3b*<sup>167-C</sup>/*g* complex due to the unfavorable separation of large molecule by analytical SEC. On the right side, SDS-PAGE reveals the formation of *eIF3b*<sup>167-C</sup>-*g* complex. **(C)** Overlay of the analytical SEC profiles of *cteIF3b*<sup>167-704</sup> (dotted black), *cteIF3g* (dashed gray) and a mixture of both (solid dark) without any considerable shift.

**(D)** GST pull-down assays of full-length eIF3b, eIF3b-WD40, eIF3i, and eIF3g. Lanes 1-5 present the individual proteins as a reference and lane 6 is the marker labeled on the right side. Lanes 6-12 refer to the elution of the protein(s) after extensive washing.

### 3.1.3 Crystallization and structure determination of eIF3 subunits and subcomplexes

#### 3.1.3.1 Crystals of eIF3j

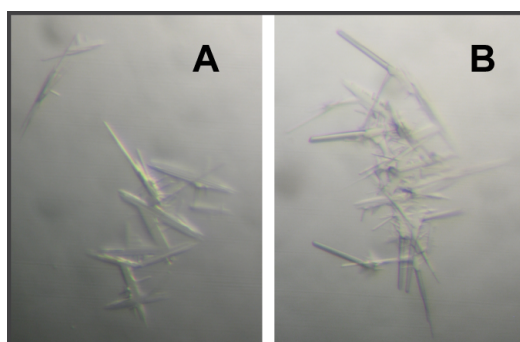
Needle-shaped crystals of full-length *S. cerevisiae* eIF3j appeared in conditions containing 1.7-2.5 M  $(\text{NH}_4)_2\text{SO}_4$ , 0.05 M either MES, Sodium Cacodylate, MOPS or HEPES buffer, pH 5.6-7.5, at 20 °C (Figure 3.8 A). Unfortunately, optimization did not lead to improved crystals that diffracted. Additionally, in the condition that contains 12.5% and 15% PEG 3350, 0.1-0.3 M NaCitrate, pH 7.0, promising crystal-look precipitation came out approximately after two weeks (Figure 3.8 B, C). Similar to those needle shaped crystals, these ‘crystals’ could not be improved. Since, except the middle compact  $\alpha$ -helical domain, eIF3j harbors long flexible tails at both N- and C-terminuses (each consisting of ~50 amino acids), a construct (*sceIF3j*<sup>56-218</sup>) that lacks the N- and C-tails was cloned. Crystals of *sceIF3j*<sup>56-218</sup> were also obtained in the same conditions where the full-length protein was crystallized (Figure 3.8 D). These crystals could not be optimized, neither. Extensive efforts were also made to crystallize *C. thermophilum* eIF3j but were not successful. Great difficulty in getting good crystals of eIF3j is consistent with its behavior on the size exclusive chromatography. Homo-dimeric eIF3j elutes at a volume corresponding to an octamer, indicating large regions of eIF3j are not structured. Extended flexible regions are known usually to impair the crystallization.



**Figure 3.8:** Crystals of full-length *sceIF3j* (A, B, C) and *sceIF3j*<sup>56-218</sup> (D). Crystals in (A) and (D) grew in condition that contained 1.7-2.5 M  $(\text{NH}_4)_2\text{SO}_4$ , 0.05 M either MES, Sodium Cacodylate, MOPS or HEPES buffer, pH 5.6-7.5. Crystals in (B) and (C) grew in condition containing 12.5-15% PEG 3350 and 0.1-0.3 M NaCitrate, pH 7.0

### 3.1.3.2 Crystallization of eIF3b<sup>74-170</sup>/eIF3j subcomplex

Crystallization of *sce*IF3b<sup>74-170</sup>/*sce*IF3j complex yielded crystals in a condition containing 20% polyethylene glycol 3350 and 0.2 M sodium thiocyanate (Figure 3.9). Based on their fragile feature, we can infer these crystals are protein crystals. However, they did not diffract.

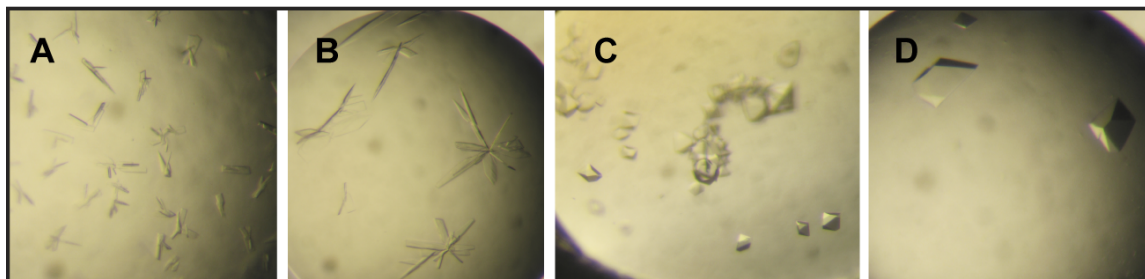


**Figure 3.9:** Crystals of *sce*IF3b<sup>74-170</sup>/*sce*IF3j complex. Crystals grew in a condition containing 20% polyethylene glycol 3350 and 0.2 M sodium thiocyanate.

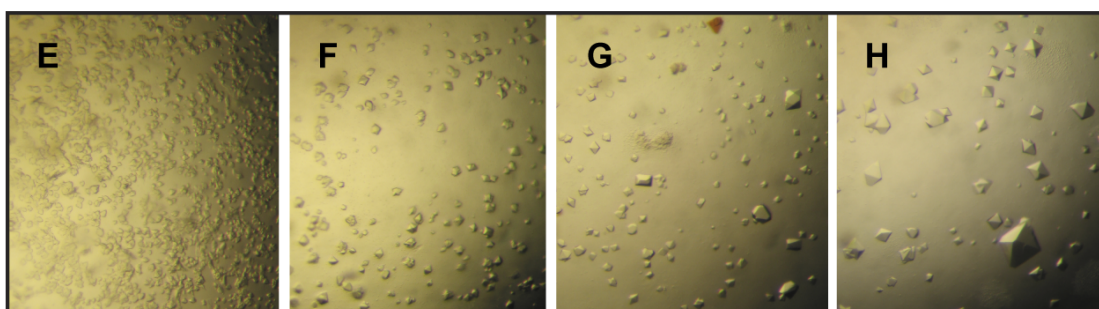
### 3.1.3.3 Crystallization and structure determination of eIF3b

Initial crystals of *cte*IF3b were obtained at 20 °C in a condition containing 15% (w/v) polyethylene glycol (PEG) 20k, 0.08 M MnCl<sub>2</sub>, 0.1 M MES, pH 6.5 after approximately 15 days (Figure 3.10 A). Optimization led to gradually improved crystals (Figure 3.10 B-D). Optimized native crystals diffracted to a resolution of 2.7 Å and belong to the space group *P*<sub>4</sub><sub>3</sub><sub>2</sub><sub>1</sub><sub>2</sub>. Molecular replacement (MR) attempts using 7- and 8-bladed β-propellers and fragments of them, as well as the RRM domains of human and yeast eIF3b did not result in any promising solutions. Therefore selenomethionine (SeMet) derivative crystals were produced (Figure 3.10 E-H). SeMet crystals emerged under the same conditions that native crystals grew and diffracted to a resolution of 3.3 Å. SeMet crystals also belong to space group *P*<sub>4</sub><sub>3</sub><sub>2</sub><sub>1</sub><sub>2</sub>. The phase problem was then solved using diffraction datasets collected at peak, inflection and remote wavelengths. Six out of seven expected selenium atoms were found by SHELXD.

## Native crystals

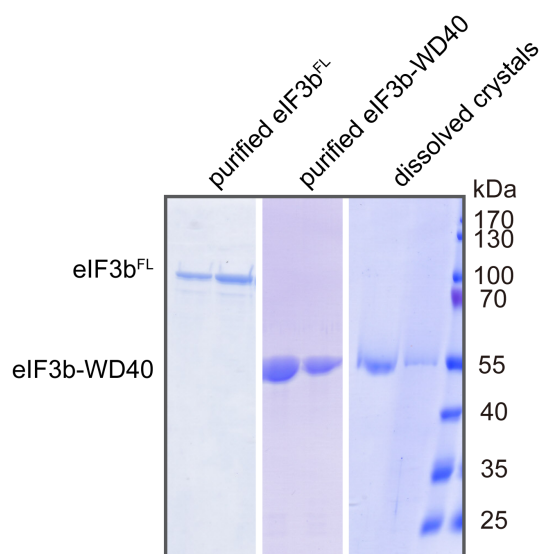


## SelenoMet crystals



**Figure 3.10:** Crystals of native *cteIF3b* (A-D) and SeMet *cteIF3b* (E-H). All crystals grew in a condition that contained 15% (w/v) PEG 20k, 0.08 M  $\text{MnCl}_2$ , 0.1 M MES, pH 6.5.

The final model was refined at a resolution of 2.7 Å with  $R_{\text{factor}} = 20.6\%$ ,  $R_{\text{free}} = 23.0\%$  (Table 3.1). However, it comprises only the WD40 domain of *cteIF3b* (residues 165-638) rather than the full-length protein. Moreover, one loop region (residues 299-311) was not defined in the electron density map. This strongly indicated that specific cleavages by contaminating proteases occurred during the crystallization. This indication has been supported by an SDS-PAGE of the dissolved crystals, which revealed a discrete band corresponding to a molecular weight of ~55 kDa that is consistent with the size of the WD40 domain instead of the expected band for the 86 kDa full-length *cteIF3b* (Figure 3.11).



**Figure 3.11:** SDS-PAGE of full-length *cteIF3b*, *cteIF3b* WD40 domain and dissolved crystals. The molecular weight of full-length *cteIF3b* and its WD40 domain are ~90 kDa and ~55 kDa, respectively. The size of the crystallized fragment is the same as *cteIF3b* WD40 domain.

#### 3.1.3.4 Crystallization of eIF3b-g-i complex

Although the protein of eIF3b-g-i subcomplex was highly pure (Figure 3.6 C), no promising crystallization hits were found. Crystallization trials of eIF3b<sup>167-C</sup>/eIF3i/eIF3g, eIF3b<sup>167-704</sup>/eIF3i and eIF3b<sup>FL</sup>/eIF3i were also performed. However, no crystals of them were obtained.

**Table 3.1:** Data collection and phasing statistics

	SeMet-derivative, MAD			Native
	Peak	Inflection	Remote	
<i>Data collection</i>				
Wavelength (Å)	0.9798	0.9799	0.9775	0.8266
Space group	<i>P</i> 4 <sub>3</sub> 2 <sub>1</sub> 2	<i>P</i> 4 <sub>3</sub> 2 <sub>1</sub> 2	<i>P</i> 4 <sub>3</sub> 2 <sub>1</sub> 2	<i>P</i> 4 <sub>3</sub> 2 <sub>1</sub> 2
Resolution (Å)	50-3.25 (3.35-3.25)	50-3.30 (3.40-3.30)	50-3.31 (3.41-3.31)	50-2.72 (2.82-2.72)
Cell dimensions				
a (Å)	108.32	108.31	108.28	108.38
b (Å)	108.32	108.31	108.28	108.38
c (Å)	186.88	186.90	186.84	172.18
Unique reflections	32904 (2863)	31367 (2719)	30923 (2659)	28259 (2828)
Redundancy	3.1 (2.9)	3.1 (3.0)	2.7 (2.7)	8.5 (8.6)
Completeness (%)	98.2 (98.9)	98.0 (98.7)	97.9 (98.6)	99.9 (100.0)
<i>R</i> <sub>merge</sub> (%)	9.2 (73.8)	9.0 (61.0)	9.5 (60.3)	5.7 (61.5)
<i>I</i> /σ ( <i>I</i> )	11.67(1.76)	12.45 (2.13)	10.71 (1.90)	25.67 (3.48)
<i>CC</i> <sub>1/2</sub>	99.7 (66.5)	99.7 (74.5)	99.5 (67.7)	99.9 (92.0)
<i>Phasing</i>				
Anomalous Corr. <sup>a</sup>	47 (4)	34 (11)	28 (12)	
SigAno <sup>a</sup>	1.33 (0.78)	1.16 (0.83)	1.08 (0.82)	
FOM after <i>DM</i>	0.573			

Values for the data in the highest resolution shell are shown in parentheses.

<sup>a</sup> Anomalous correlation and mean anomalous difference (SigAno), calculated with *XSCALE*.

**Table 3.2:** Refinement statistics

<i>Structure refinement</i>	
Resolution range (Å)	48.47-2.72 (2.82-2.72)
Completeness (%)	100.0 (100.0)
$R_{\text{work}}/R_{\text{free}}$ (%) <sup>b</sup>	20.56/22.98
No. of atoms	
Total	3819
Protein	3764
Solvent	51
Ions	4
RMSD bonds (Å)	0.003
RMSD angles (°)	0.806
B factor (Å <sup>2</sup> )	
Average	74.8
Protein	75.0
Solvent	61.8
Ions	97.0
Wilson B factor (Å <sup>2</sup> ) <sup>c</sup>	66.87
Solvent content (%)	73
Ramachandran plot	
Favored (%)	95.19
Allowed (%)	4.81
Outliers (%)	0.0

Values for the data in the highest resolution shell are shown in parentheses.

<sup>b</sup>  $R_{\text{free}} = \sum_{\text{Test}} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{\text{Test}} |F_{\text{obs}}|$ , where “Test” is a test set of about 5% of the total reflections randomly chosen and set aside prior to refinement for the complex.

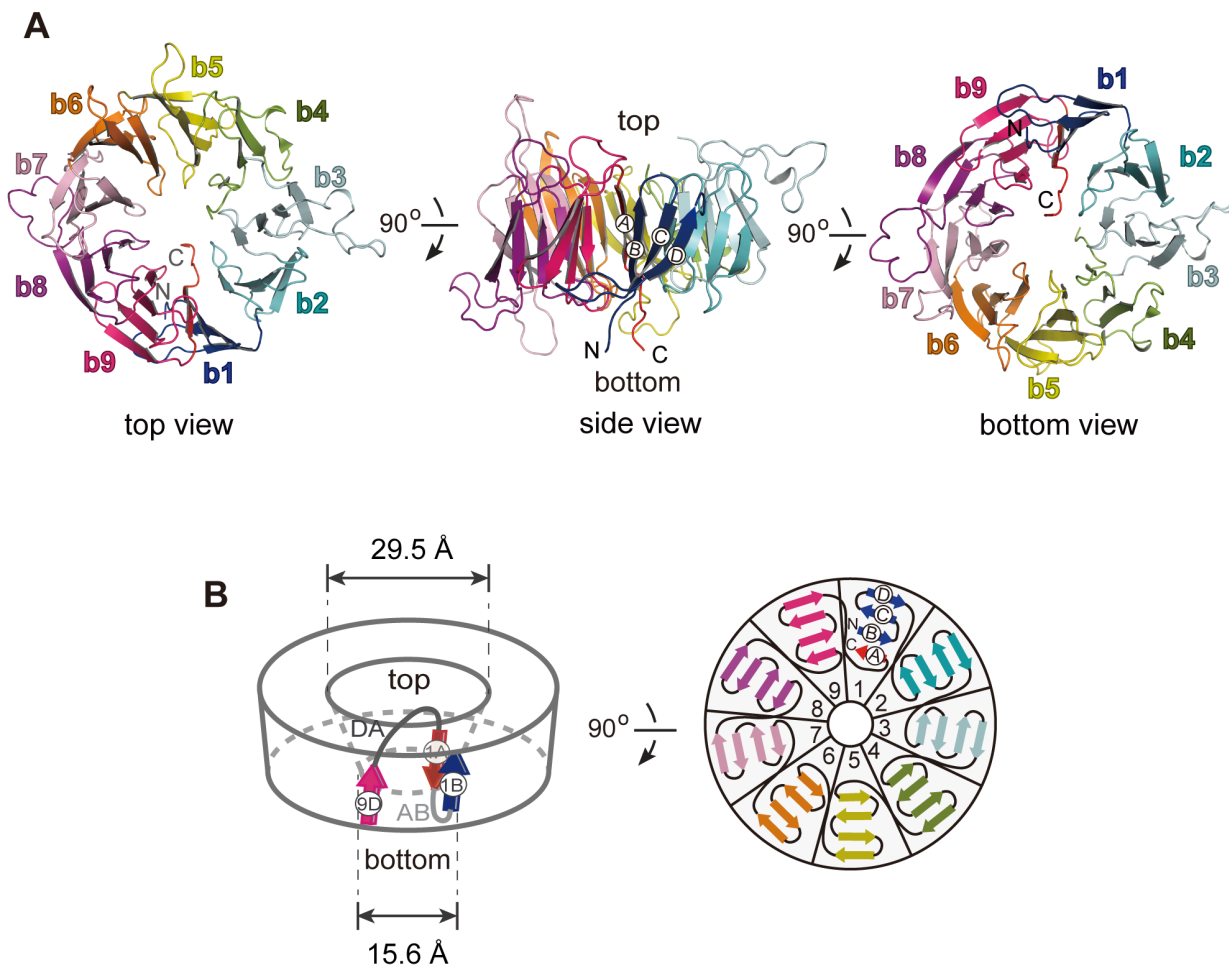
<sup>c</sup> Wilson B factor was estimated with phenix.refine.



### 3.1.4 Structural analysis of eIF3b-WD40

#### 3.1.4.1 Overall structure

The WD40 domain of eIF3b has been predicted to fold into two tandem  $\beta$ -propellers [46]. Surprisingly, in contrast to the prediction, eIF3b-WD40 actually adopts a single  $\beta$ -propeller fold composed of nine blades (Figure 3.12). The nine blades are pseudo-symmetrically arranged in sequential order around a central channel, resulting in a doughnut-like architecture. Each blade of the propeller is formed by a



**Figure 3.12:** Overall structure of *cteIF3b*-WD40. (A) Top, side and bottom view of the overall structure of *cteIF3b*-WD40 in cartoon representation. Blades 1-9 are colored individually. N- (N) and C-termini (C) are labeled. (B) Topology diagram of the *cteIF3b*  $\beta$ -propeller fold. The left and right drawing are shown in orientations similar to the side and bottom view in (A), respectively.

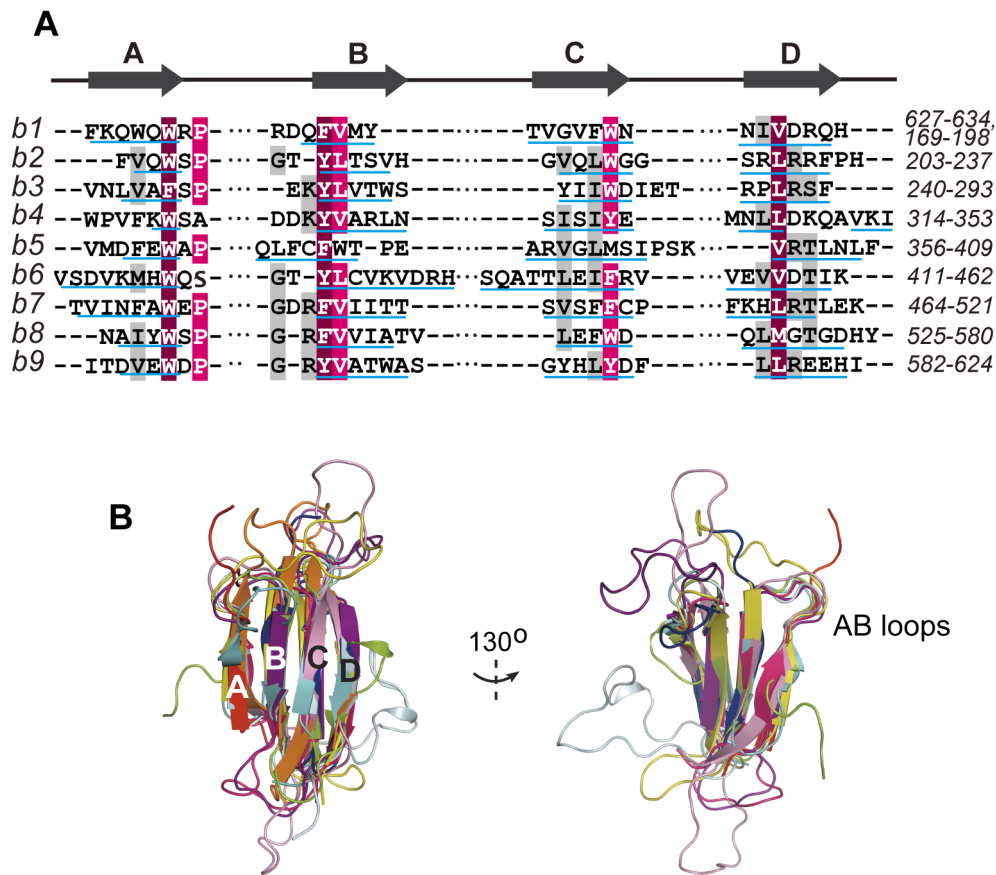


$\beta$ -sheet consisting of four antiparallel strands, which are designated A, B, C and D from the inside to the outside of the propeller. Notably, the first blade of the propeller is composed of three N-terminal strands and a C-terminal strand positioned innermost, creating a “ $3_N + 1_C$  velcro” closure. Interestingly, the central channel of the nine-bladed *cteIF3b* WD40 propeller resembles a circular truncated cone rather than a common cylinder (Figure 3.12). The diameter of the channel decreases from 29.5 Å on the ‘top’ side to ~15.6 Å on the ‘bottom’ side. By contrast, the outermost shape of *cteIF3b* propeller is almost cylindrical. The overall diameter on the top and bottom side is 53.6 Å and 51.5 Å, respectively.

#### 3.1.4.2 Sequence and structural conservation of the WD40 repeats

Structure-based sequence alignments of the WD40 repeats of *cteIF3b* reveal several conserved sequence characteristics, notably the Trp-X-Pro (where X represents any amino acid) motif at the end of strands A, and Phe/Tyr- $\Phi$  (where  $\Phi$  stands for a hydrophobic amino acid) segment in strands B, as well as hydrophobic residues in strands D (Figure 3.13 A). However, the eponymous Trp-Asp (WD) dipeptide located at the ends of strands C is only present in blades 3 and 8. The Trp residue of the WD motif has been replaced by the aromatic amino acid Tyr (in blades 4 and 9), Phe (in blades 6 and 7) and the hydrophobic Met in blade 5, whereas the Asp has been substituted by other polar residues with the only exception of a Gly in blade 2 (Figure 3.13 A).

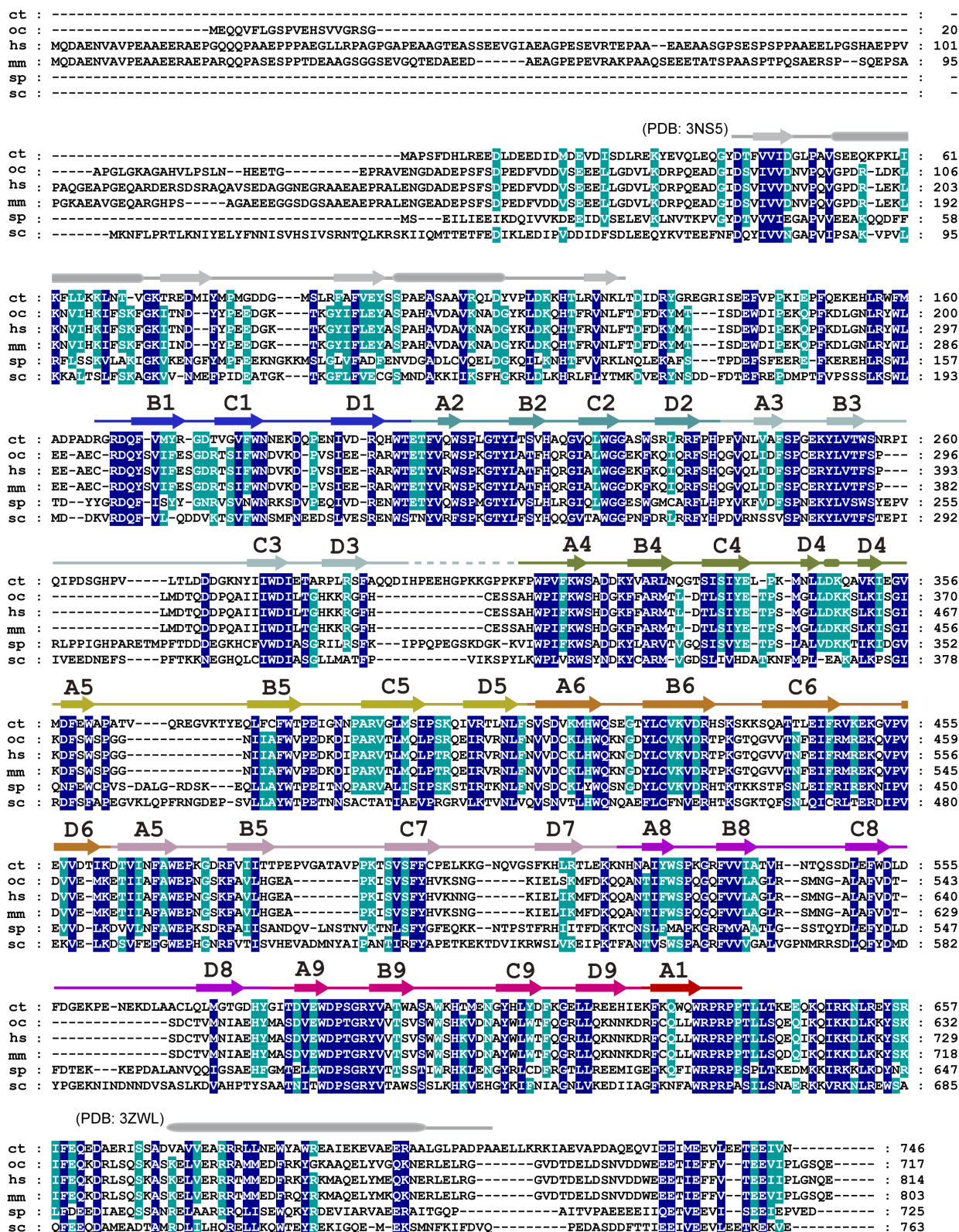
The repeating motifs (single blades) within the *cteIF3b* WD40 domain are structurally well conserved as well, with an root-mean-square deviation (r.m.s.d.) ranging from 0.94 Å to 1.57 Å for all C $\alpha$  atoms, whose positions differed less than 3 Å, when superimposing each blade onto blade 1 (Figure 3.13 B).



**Figure 3.13:** Sequence and structural conservation of eIF3b. **(A)** Sequence alignment of the nine  $\beta$ -sheets of *cteIF3b*-WD40. Residues are colored according to the conservation pattern (highly conserved: dark red; conserved: pink). Residues forming the  $\beta$ -strands are underlined and arrows on the top indicate the positions of strands A-D. **(B)** Superposition of the nine  $\beta$ -propeller blades of *cteIF3b*. The blades are colored as in Fig 1B. In the image on the right, the long  $\beta$ -sheet 6 was omitted for clarity reasons.

### 3.1.4.3 Conservation of eIF3b orthologs

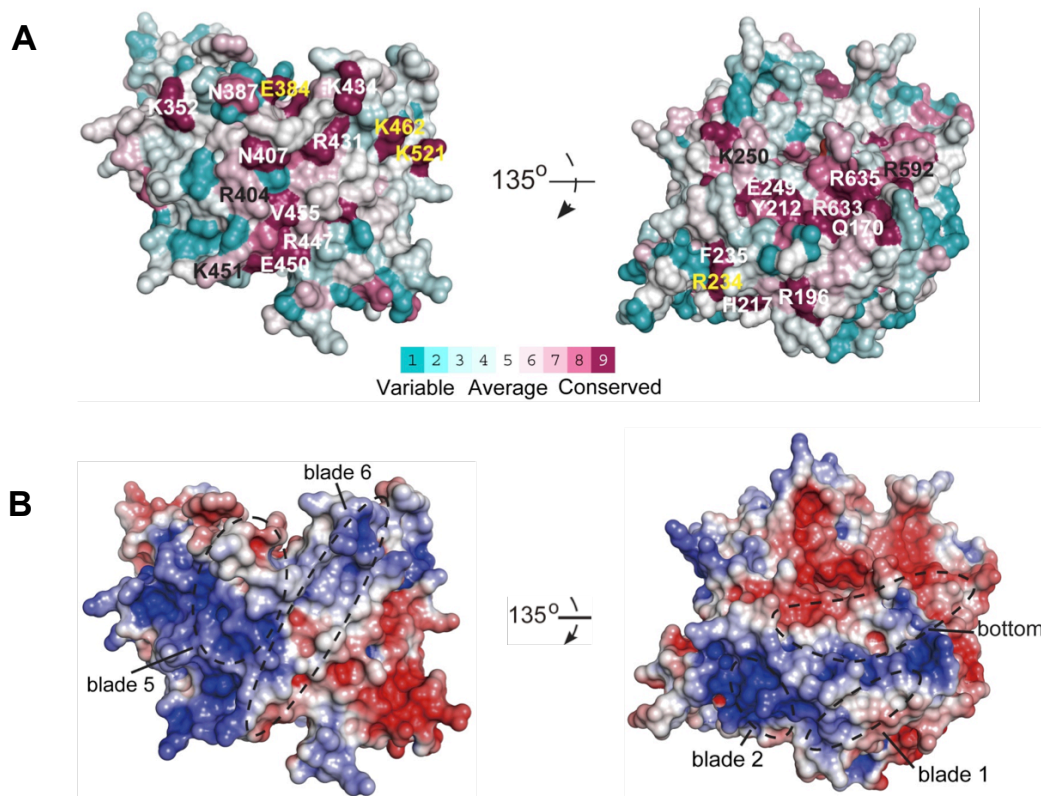
Multiple sequence alignment reveals that eIF3b orthologs from different organisms vary considerably in the N-terminal domain, but are well conserved in the WD40 domain as well as in the C-terminal part of the protein (Figure 3.14). Secondary structure predictions (using the PSIPRED server [158]) of eIF3b orthologs assign 34-36  $\beta$ -strands to their WD40 domains, indicating that the nine-bladed WD40  $\beta$ -propeller fold of *cteIF3b* is not a unique instance, but is employed universally by eIF3b orthologs. Many of the identical residues among the orthologs are hydrophobic and located in the structural core of the WD propeller (Figure 3.14), indicating their role in accurate folding and stability.



**Figure 3.14:** Multiple sequence alignment of eIF3b orthologs. Amino acid sequences of eIF3b from *Chaetomium thermophilum* (ct), *Oryctolagus cuniculus* (oc), *Homo sapiens* (hs), *Mus musculus* (mm), *Schizosaccharomyces pombe* (sp), and *Saccharomyces cerevisiae* (sc) are aligned using ClustalW2

## Results

([www.ebi.ac.uk/Tools/msa/clustalw2/](http://www.ebi.ac.uk/Tools/msa/clustalw2/), [159]). Identical residues are highlighted with *blue* background and conserved residues with *cyan* background. Secondary structure elements ( $\beta$ -strands: arrows;  $\alpha$ -helices: rounded rectangles; loops: lines) are indicated above the alignment (residues that were not built in the structure are represented as dashed lines). The WD40 domain of eIF3b is colored as in Fig 3.12B, while the N-terminal RRM domain and the C-terminal  $\alpha$ -helix, which are not present in our structure, are colored with light gray.



**Figure 3.15:** Surface representation of *cteIF3b*-WD40. (A) Surface representation of *cteIF3b*-WD40 with sequence conservation represented in the color coding bar, ranging from cyan (variable) to purple (conserved). The conservation scores were generated by ConSurf (<http://consurf.tau.ac.il/>) using the sequence alignment of eIF3b-WD40 as shown in Figure 3.14. Highly conserved surface residues are labeled. (B) Electrostatic potential of conserved surface areas of *cteIF3b*-WD40. The electrostatic surface potential of *cteIF3b*-WD40 colored from red (-5kT/e) through white (0kT/e) to blue (+5kT/e). Areas containing conserved surface residues are encircled and labeled based on blade numbers or propeller side. The orientations are similar to (A).

Additionally, two clusters of surface residues are highly conserved from yeast to humans (Figure 3.15 A), suggesting a potential role in binding other proteins and/or the other eIF3b domains. The first conserved surface region is located on the side of the propeller and consists of residues mainly belonging to blades 5 and 6; the second

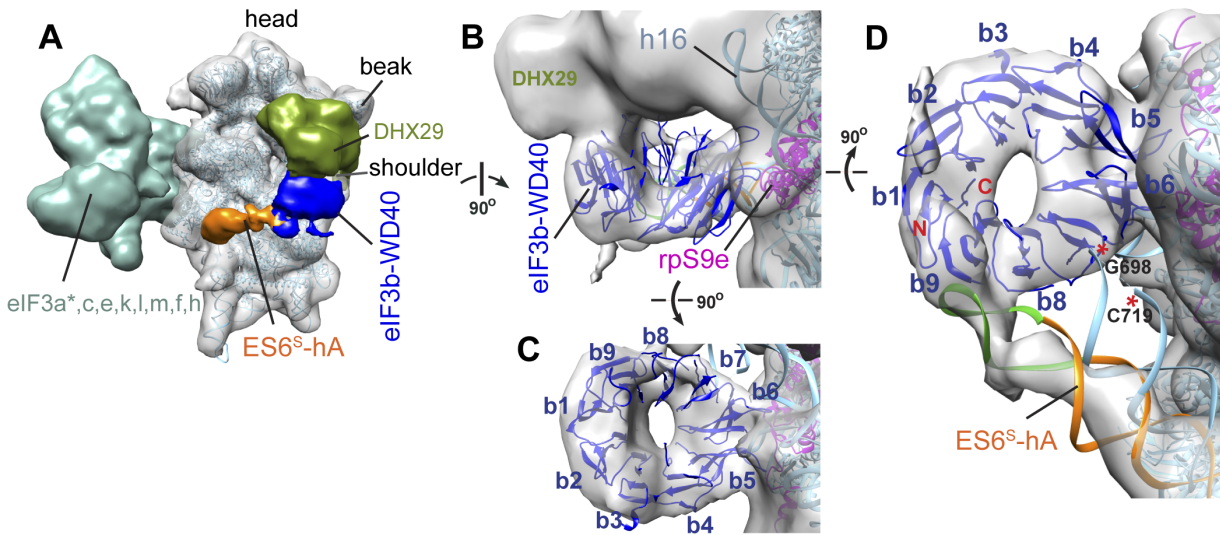
conserved area is situated primarily on the top of the propeller involving residues from the AB loops of blades 9, 1, 2 and 3 and the D strands of blades 1 and 2 (Figure 3.15 A). Interestingly, both conserved surface areas of *cteIF3b*-WD40 mainly contain hydrophobic and positively charged residues, which could be involved in RNA-binding (Figure 3.15 B).

### 3.1.5 Localization of eIF3b on the ribosome

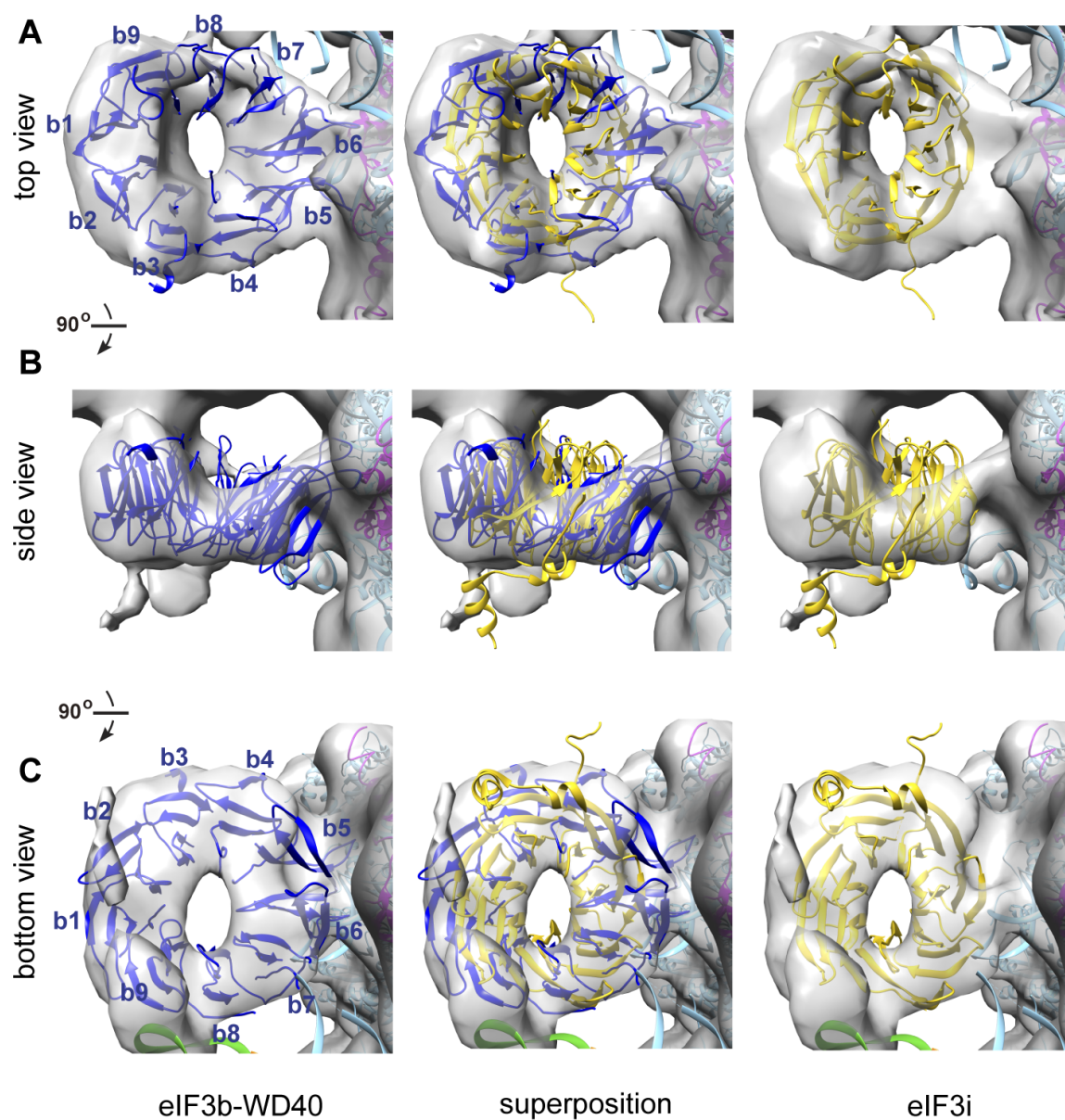
In the cryo-EM map of the rabbit 43S PIC bound to helicase DHX29 (EMDB code: 5658), the interpretation of a doughnut-like density below the shoulder of the 40S subunit could not be verified. The 7-bladed  $\beta$ -propeller protein eIF3i was tried to fit there, however, the possibility that this density originated from the WD40 domain of eIF3b could not be excluded by the authors (Figure 3.16 A) [61]. Notably, this cryo-EM segment definitely possesses a channel with a wide top and narrow bottom. As aforementioned, the WD40 domain of eIF3b harbors a channel with similar geometric parameters, which reaches a consensus with the cryo-EM density on the axle outline (Figure 3.12 and 3.16 B,C). By contrast, eIF3i holds a channel that has a similar width on both sides (PDB code: 3ZWL) (Figure 3.17). Additionally, the overall size of eIF3i is too small to fully explain this EM density (Figure 3.17). Due to the lack of technical details, the fitting of eIF3i to the doughnut-like density could not be repeated and show a similar result as reported [61]. Our fit of eIF3i using Chimera 1.8 [160] resulted in a cross-correlation coefficient (CCC) of 0.52, and 38.5% atom outliers, respectively. The fitting statistics significantly improved when the WD40 propeller (without the non-conserved loops) of *cteIF3b* was positioned in the doughnut-like density (Figure 3.17), with a corresponding CCC of 0.91 and only 17% of the atoms outside the density. On the basis of our fit, blades 5 and 6 of the propeller are oriented towards the ribosomal protein S9e (rpS9e) and the ribosomal RNA helix h16, while the opposite side holding the N- and C-terminuses is solvent exposed (Figure 3.16 B,C). The position of eIF3b directly below the shoulder of the 40S is in good agreement with previous studies. Two eIF3b binding domains (eIF3j-NTD and eIF3a-CTD) have been reported to bind to the 40S mRNA entry



channel that is in close proximity to the shoulder of the 40S subunit [47,157,161,162].



**Figure 3.16:** Interaction of eIF3b with the 40S ribosomal subunit. (A-D) Localization of eIF3b on the 40S subunit. The non-conserved loop insertions of *cteIF3b*-WD40, which are absent in rabbit eIF3b, were removed prior to fitting. In figure C and D, the density assigned to DHX29 was removed for clarity. (A) Overall view of the position of eIF3b-WD40  $\beta$ -propeller on the 40S ribosomal subunit (EMDB ID: 5658). eIF3b-WD40 (blue) is located below the shoulder of the 40S and is disconnected from the eIF3 subcomplex containing the subunits a\*, c, e, f, h, k, l, and m (cyan). The asterisk marks a truncated version of eIF3a. (B) Detail side view of eIF3b-WD40 localization. The rpS9e and the rRNA helix h16 are indicated. (C) Top view of eIF3b-WD40 showing the characteristic conical central channel. (D) Structural model of a possible interaction between the eIF3b WD40 propeller and the 18S rRNA expansion segment ES6<sup>S</sup>-hA. The rRNA helix was manually fitted into the EM density portion using the crystal structure of the 40S ribosome (PDB code: 4KZZ). 19 nucleotides of ES6<sup>S</sup>-hA, which are missing in the crystal structure, have been modeled (colored green) to illustrate their potential extension.

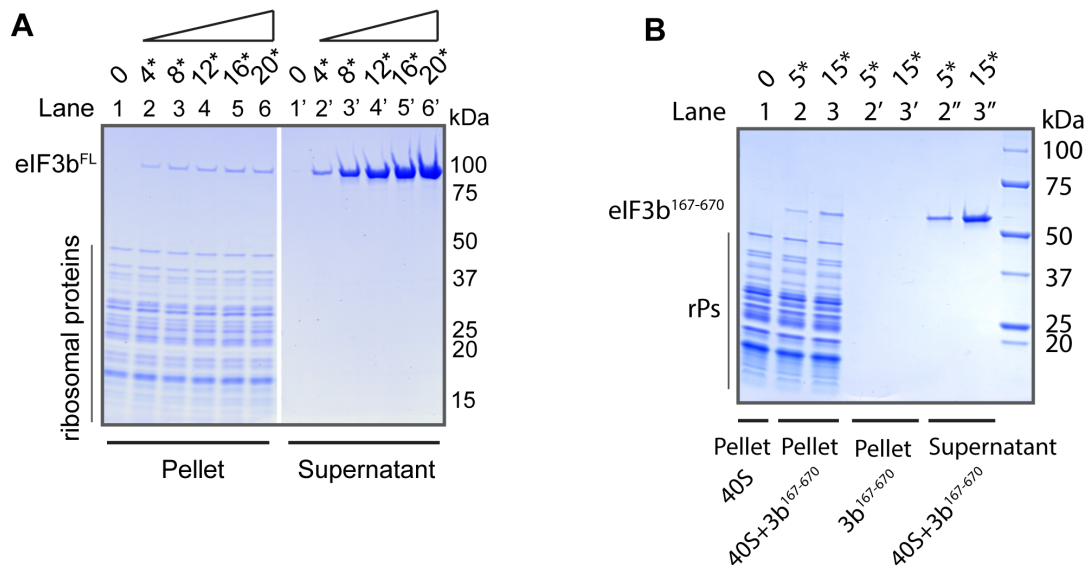


**Figure 3.17:** Comparison of the final fit of eIF3b-WD40 (blue) and eIF3i (yellow, PDB code: 3ZWL). In figure A the density (assigned to DHX29) above the doughnut-like density was removed for clarity.

### 3.1.6 Interaction of eIF3b with the ribosome

#### 3.1.6.1 Interaction with the 40S ribosomal subunit

Since a direct interaction between eIF3b and the 40S ribosomal subunit has so far not been reported, *in vitro* co-sedimentation assays were performed to verify the interpretation of the EM map (in collaboration with S. Schell and A. Chari). Purified *C. thermophilum* 40S ribosomal subunit was incubated with increasing amounts of *cte*eIF3b and the mixtures were applied to sucrose gradients. The appearance of an additional ~86 kDa protein in the pellet, which is absent in sedimented pure 40S subunit, indicates that eIF3b stably associates with the 40S, while unbound eIF3b remained in the supernatant (Figure 3.18 A). Besides full-length eIF3b, a truncated eIF3b (eIF3b<sup>167-670</sup>, containing the WD40 domain plus a ~30-residue C-terminal extension) also binds to the 40S (Figure 3.18 B), which suggests that the N-terminal RRM domain of eIF3b is not necessary for the association of eIF3b and 40S.



**Figure 3.18:** (A) Co-sedimentation of eIF3b and the 40S. eIF3b was co-sedimented with the 40S (Lanes 2-6) and the excess eIF3b remains in supernatant (Lanes 2'-6'). Lanes 1 and 1' represent the 40S alone control. (B) Co-sedimentation of eIF3b<sup>167-670</sup> and the 40S. eIF3b<sup>167-670</sup> was co-sedimented with the 40S (Lanes 2-3) and the excess eIF3b<sup>167-670</sup> remains in the supernatant (Lanes 2''-3''). Lane 1 and Lanes 2'-3' show the 40S alone and the protein alone control, respectively.

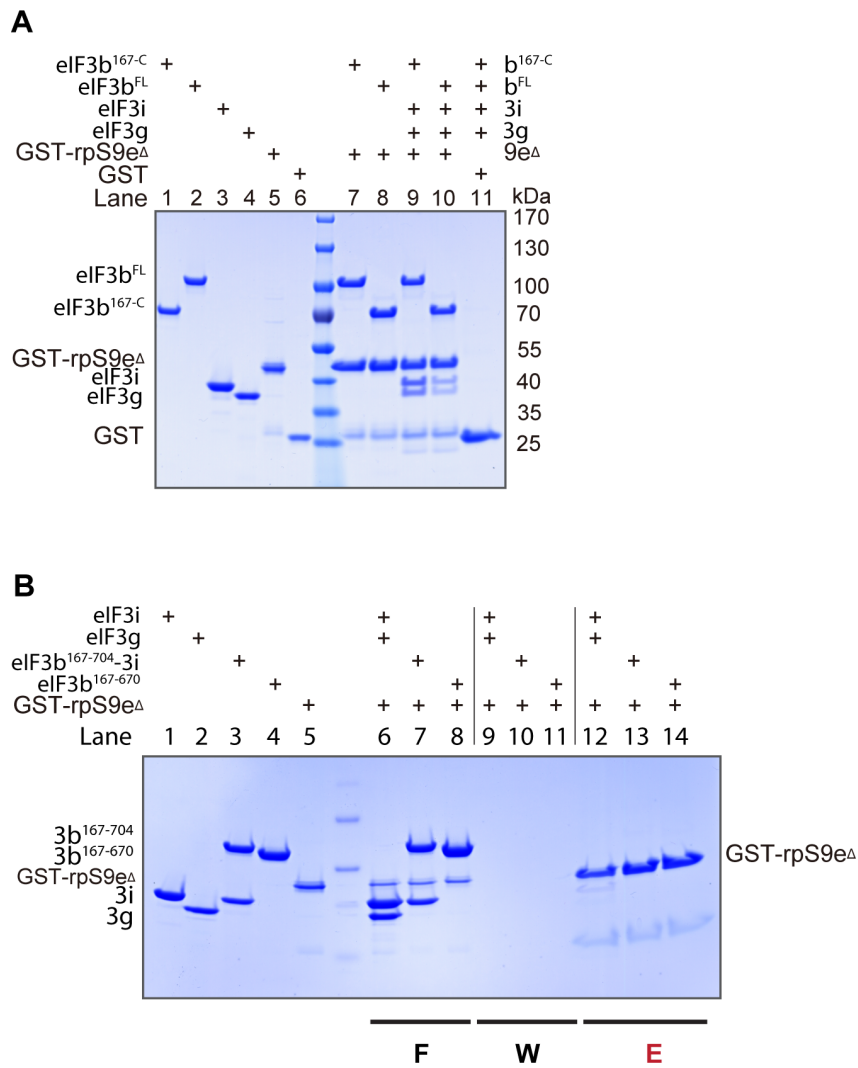


### 3.1.6.2 Interaction with the ribosomal protein S9e

Since our interpretation of the 43S PIC EM-structure places eIF3b next to the ribosomal protein rpS9e, a possible interaction between these two proteins was analyzed by pull-down assays. Both full-length *cteIF3b* and the truncated version *cteIF3b*<sup>167-C</sup> bind to GST-tagged *ctrpS9e* (Figure 3.19 A), while the truncated proteins *cteIF3b*<sup>167-670</sup> and *cteIF3b*<sup>167-704</sup> do not (Figure 3.19 B). This implies that the 42 C-terminal residues of *cteIF3b* are crucial for the interaction with *ctrpS9e*. Importantly, this interaction of the eIF3b-CTD is not a prerequisite for binding of eIF3b to the 40S ribosome, as *cteIF3b*<sup>167-670</sup> is still capable to associate with 40S (Figure 3.18 B). Thus, interactions of the  $\beta$ -propeller with other parts of the 40S ribosome, *e.g.* the rRNA ES6<sup>S</sup>-hA, are sufficient for stable association (see Discussion 4.3).

### 3.1.6.3 Formation of rpS9e-eIF3b-g-i complex

As known, the eIF3b CTD also contains the binding sites for eIF3g and eIF3i. Whether the interaction between eIF3b and rpS9e affects the binding of eIF3g and eIF3i to eIF3b was further tested. The observed quaternary complex (Figure 3.19 A) suggests that these intermolecular interactions are not mutually exclusive. Consistently, the rpS9e binding site of eIF3b is located C-terminally to the  $\alpha$ -helix of eIF3b interacting with the eIF3i  $\beta$ -propeller [48]. Interestingly, the NTD of eIF3b apparently also contributes to the formation of the rpS9e-eIF3b-g-i complex, as the N-terminal truncated eIF3b<sup>167-C</sup> exhibits a lower affinity to eIF3g, i (Figure 3.19 A).



**Figure 3.19:** (A) Interaction study between eIF3b and rpS9e by GST pull-down assay. Lanes 1-6 show the isolated proteins as a reference. Lanes 7-11 show the elution fractions of the GST pull-down after the removal of unbound proteins and extensive washing. rpS9e<sup>Δ</sup> corresponds to rpS9e<sup>1-183</sup>. (B) Interaction study between eIF3b truncations, eIF3i, eIF3g and rpS9e by GST pull-down assay. Lanes 1-5 show the single proteins as a reference. Lanes 12-14 show the elution fractions of the GST pull-down after the removal of unbound proteins (Lanes 6-8) and extensive washing (Lanes 9-11). rpS9e<sup>Δ</sup> corresponds to rpS9e<sup>1-183</sup>. No eIF3b<sup>167-670</sup> band in Lane 14, eIF3b<sup>167-704</sup>-eIF3i complex in Lane 13 indicate that rpS9e could not bind these three polypeptides or the interaction between them is too weak to be observed by GST pull down. The observed weak eIF3i and eIF3g band in lane 12 suggests that rpS9e is able to weakly bind eIF3i/eIF3g complex.

## 3.2 Eukaryotic translation initiation factor 4 (eIF4)

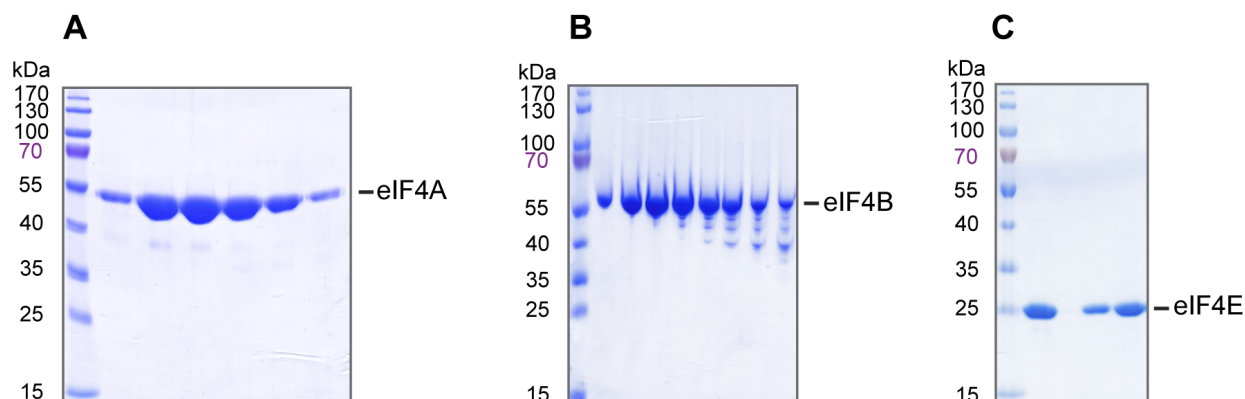
The heterotrimeric complex eIF4F consists of the cap recognition protein eIF4E, the ATP-dependent RNA helicase eIF4A and the large scaffold protein eIF4G. It works in conjunction with eIF4B to activate the mRNA to be translated. eIF4B and eIF4G have been shown to enhance the ATPase and helicase activities of eIF4A (for more details see Introduction 1.4). The aim of this section is to obtain the structural information of the mRNA activation complex and determine the underlying mechanism by which the ATPase and helicase activities of eIF4A are stimulated by eIF4G and eIF4B.

### 3.2.1 Preparation and crystallization of the full-length eIF4F complex

#### 3.2.1.1 Purification of eIF4A, eIF4B and eIF4E

The gene encoding full-length *S. cerevisiae* eIF4A was cloned into a modified pET15b vector that expresses a His<sub>6</sub>-tag at the N-terminus of eIF4A. *S. cerevisiae* eIF4B and *S. cerevisiae* eIF4E were individually fused with an N-terminal GST tag (pGEX-6P-1). All three relatively small proteins (eIF4A: 45kDa; eIF4B: 48kDa; eIF4E: 24kDa) could be overexpressed in *E. coli* cells. Both BL21 (DE3) and Rosetta 2 (DE3) strains were able to produce full-length eIF4A and eIF4E, but eIF4B could only be synthesized in Rosetta 2 (DE3) cells. In addition, overexpressed eIF4A seems to be toxic for the host cells. Thus, it was necessary to harvest the cells five hours after induction with IPTG (rather than overnight expression). All three full-length proteins could be purified to near homogeneity in an RNA free form using Ni-NTA affinity column (for eIF4A) or GST affinity column (for eIF4B and eIF4E) followed by a Source 30Q anion exchange step (eIF4A and eIF4B) (Figure 3.20). All proteins behave as a monomer in solution as revealed by the gel filtration chromatography or multi-angle light scattering. It turned out that eIF4B is prone to degradation (Figure 3.20 B), and thus, it is necessary to add protease inhibitors

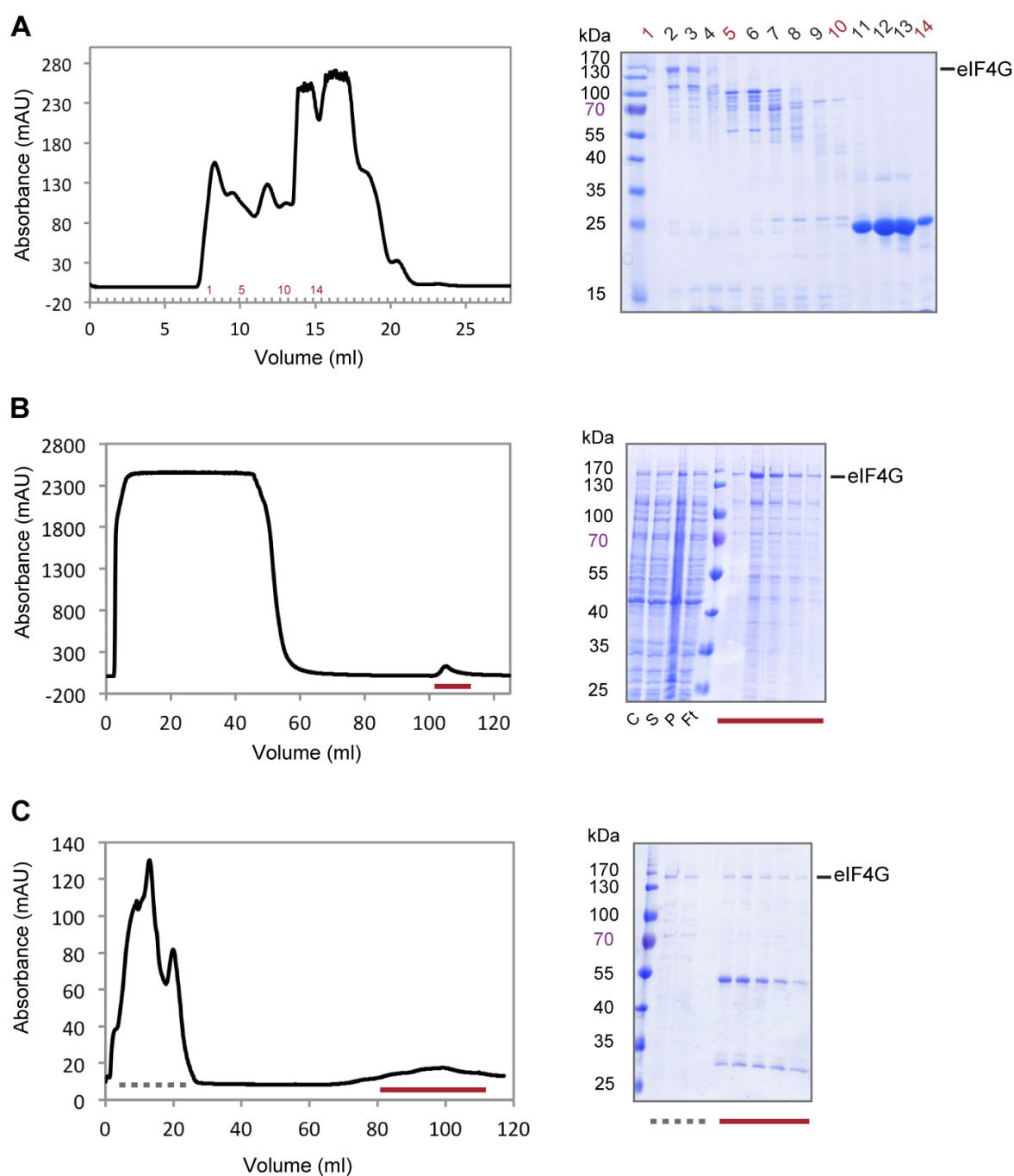
during the purification procedure. The yield for all three proteins was ~8 mg per liter of culture.



**Figure 3.20:** Purification of eIF4A (A), eIF4B (B) and eIF4E (C). Coomassie-stained SDS gels show the samples after the gel filtration as the last purification step. Before gel filtration, eIF4A was purified using a Ni-NTA affinity column and a Source 30Q, while eIF4B and eIF4E were independently passed through a GST affinity column, and a Source 30Q for eIF4B. The tags were cleaved after the first affinity column.

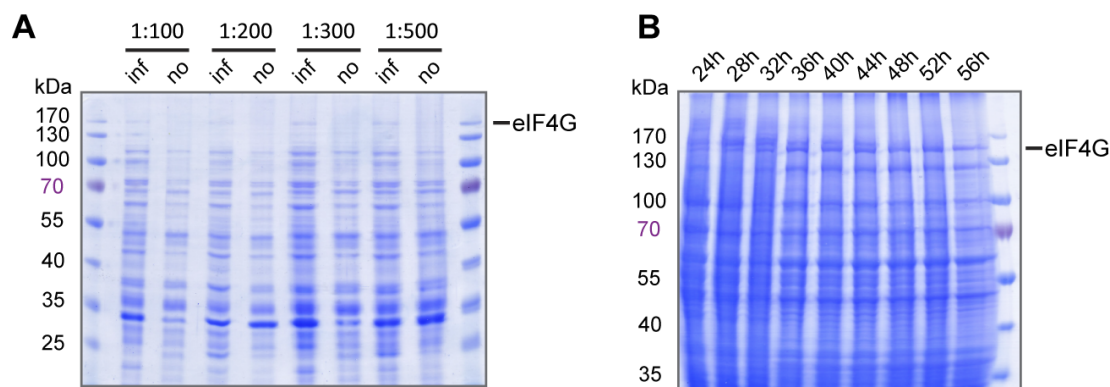
### 3.2.1.2 Expression and purification of full-length eIF4G

**Expression in bacterial cells** *E. coli* cells were first used to express *S. cerevisiae* eIF4G (107 kDa). The N-terminal GST fused full-length eIF4G could be expressed in *E. coli* Rosetta 2 (DE3) cells, albeit the yield was pretty low (~0.25 mg per liter of culture). However, this eIF4G was prone to degradation, and thus purification of the full-length protein failed (Figure 3.21 A). To obtain full-length eIF4G, a construct of eIF4G that possesses an N-terminal Strep tag as well as a C-terminal His<sub>6</sub> tag was designed. This double tagged eIF4G was also expressed in *E. coli* Rosetta 2 (DE3) cells and could be purified to a state that only contain two impurities after the Strep-Tactin affinity and Ni-NTA affinity columns (Figure 3.21 B, C). Despite extensive work, it was not possible to obtain enough protein for an additional purification step (e.g. a size exclusion chromatography), and thus for crystallization trials as well as functional studies.



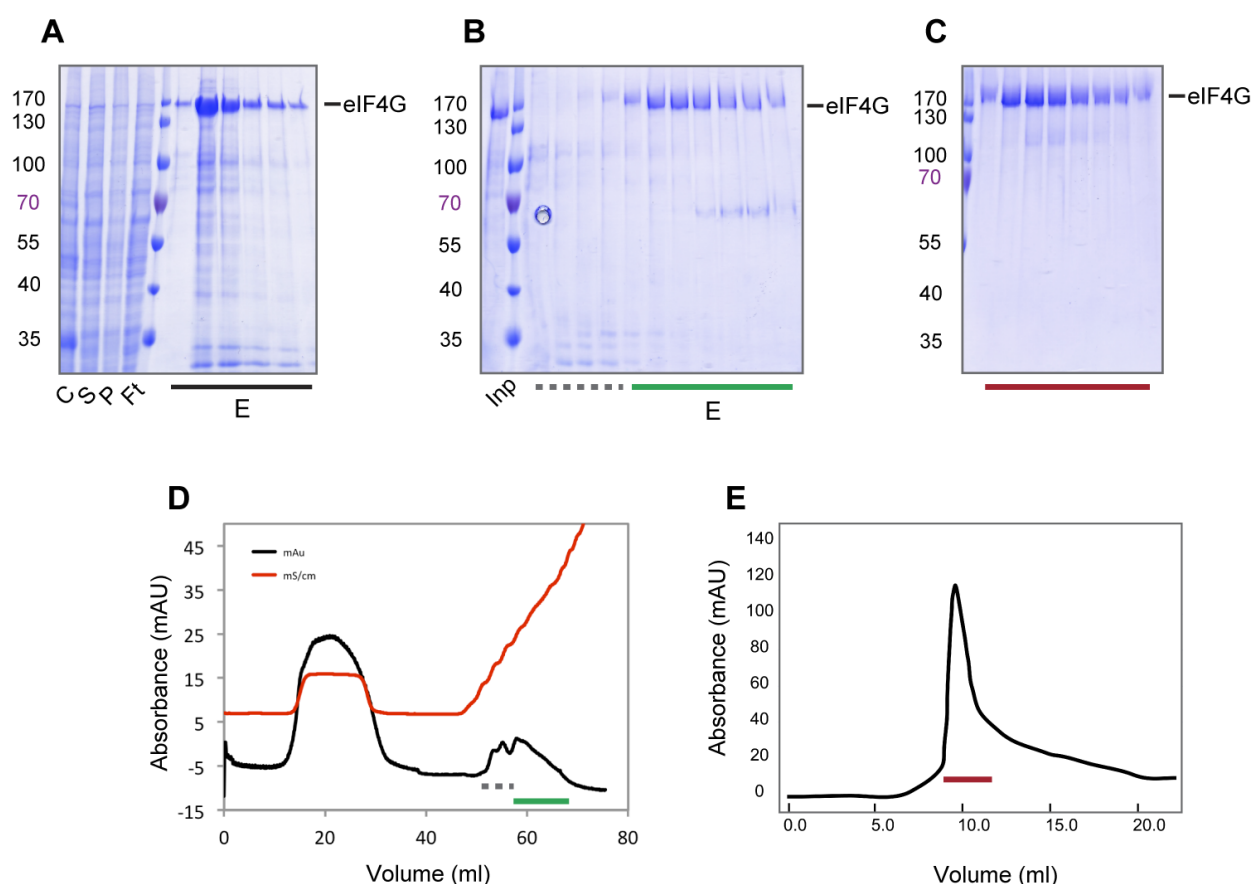
**Figure 3.21:** Purification of full-length eIF4G expressed in bacterial cells. (A) Size exclusion chromatography (SEC) of full-length eIF4G that was expressed with an N-terminal GST tag. The GST tag was removed between the GSTrap and SEC. The protein was severely degraded. (B) The Strep Tactin affinity chromatography of full-length eIF4G that was expressed with an N-terminal Strep tag and a C-terminal His<sub>6</sub> tag. ‘C’: whole cell lysate; ‘S’: supernatant; ‘P’: pellet; ‘Ft’: flowthrough of Strep Tactin chromatography. (C) The HisTrap affinity chromatography after the Strep Tactin affinity chromatography. Full-length eIF4G could be purified to a state that only contains two impurity bands.

**Expression in insect cells** Considering the limitations of bacterial protein expression, insect cells were tried to produce full-length *S. cerevisiae* eIF4G. Thus, The gene encoding eIF4G was cloned into a pFastBac HT B vector (Invitrogen) and a Bacmid DNA containing eIF4G gene was generated. Subsequently, this Bacmid DNA was transfected into Sf9 insect cells for virus generation and amplification (more details see 2.2.2.2). The protein was expressed (with an N-terminal His<sub>6</sub> tag) for various times after infection with different ratios of virus to cells as shown by SDS-PAGE (Figure 3.22). Large-scale expression was done in High Five cells.



**Figure 3.22:** Coomassie-stained SDS gels showing expression of full-length eIF4G in insect cells. **(A)** Test experiment of the amount of virus. ‘inf’: infected; ‘no’: no infection. In this case, the amount of virus did not really affect the level of expression. **(B)** Expression time course of full-length eIF4G. eIF4G starts to be expressed after 24 h of infection (1:300), and the expression increases along with time. Usually, the cells were harvested after 36 h infection, taking into consideration of the total amount of the living cells and stability of expressed full-length eIF4G.

After systematic purification optimization, full-length eIF4G could be prepared with high purity using the HisTrap column followed by SP sepharose and gel filtration chromatography (Figure 3.27). To completely get rid of contaminations (including proteins and RNAs), the sample was extensively washed with high salt (500 mM NaCl and 1 mM LiCl) containing buffer (details see 2.2.2.7).

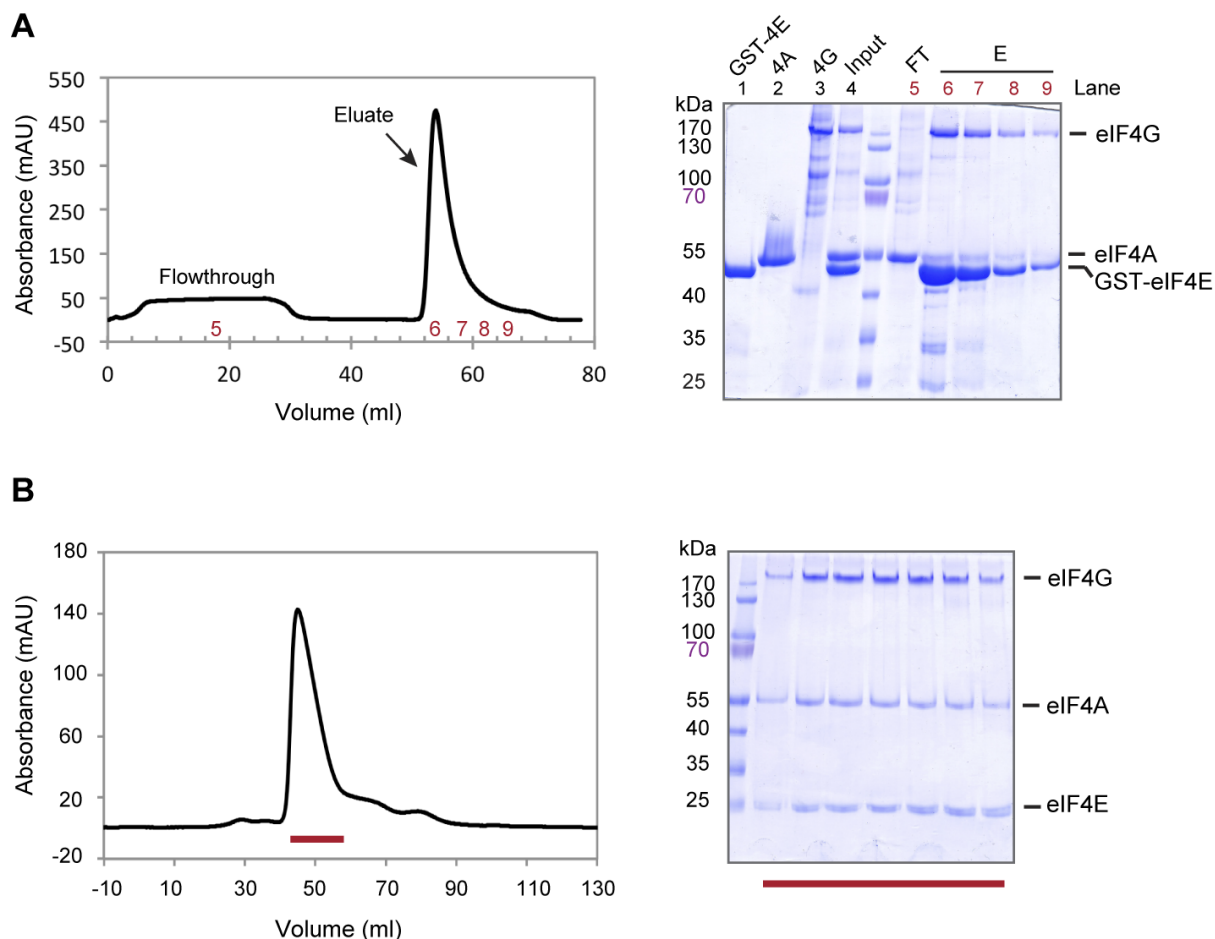


**Figure 3.23:** Purification of full-length eIF4G expressed in insect cells. (A), (B) and (C) Coomassie-stained SDS gels showing samples from HisTrap, SP Sepharose and Size exclusion chromatography (SEC), respectively. ‘C’: whole cell lysate; ‘S’: supernatant; ‘P’: pellet; ‘Ft’: flowthrough; ‘Inp’: input. The profile of HisTrap chromatography is similar to Figure 3.21 B, while the chromatograms of SP Sepharose and SEC are shown in (D) and (E), respectively. Full-length eIF4G elutes from the Superdex 200 (16/300) at ~9.7 ml.

### 3.2.1.3 Preparation of the full-length eIF4F complex

Since full-length eIF4G is difficult to obtain, in the case of preparing eIF4F complex (consisting of eIF4E, eIF4G and eIF4A), we applied a special purification strategy to reduce the loss of the protein during purification. We used purified GST tagged eIF4E as the ‘bait’ to fish eIF4G that was supplied with purified eIF4A. As shown by the SDS gel (Figure 2.24 A), these impurities that are associated with eIF4G and excess eIF4A eluted in the flowthrough, while the GST fused ternary eIF4F complex and GST-eIF4E eluted in buffer containing 35 mM reduced Glutathione. The GST tag was removed by PreScission protease and finally separated by a Superdex 200

(Figure 3.24 B). Thus, eIF4E-eIF4G-eIF4A complex with high purity could be produced.



**Figure 3.24:** Purification of full-length eIF4F complex. **(A)** Affinity purification of full-length eIF4G supplied with purified eIF4A, using purified GST-eIF4E that binds to the GSTrap column. The left panel shows the chromatogram, and the right panel shows corresponding the SDS-PAGE gel. Impurities and excess of eIF4A were eluted in flowthrough (Lane 5), while the GST fused ternary eIF4F complex and excess GST-eIF4E were eluted in elution buffer. **(B)** Size exclusion chromatography of full-length eIF4F complex. The Superdex 200 (16/60) chromatogram is shown on the left panel, and the SDS-PAGE gel is exhibited on the right panel.

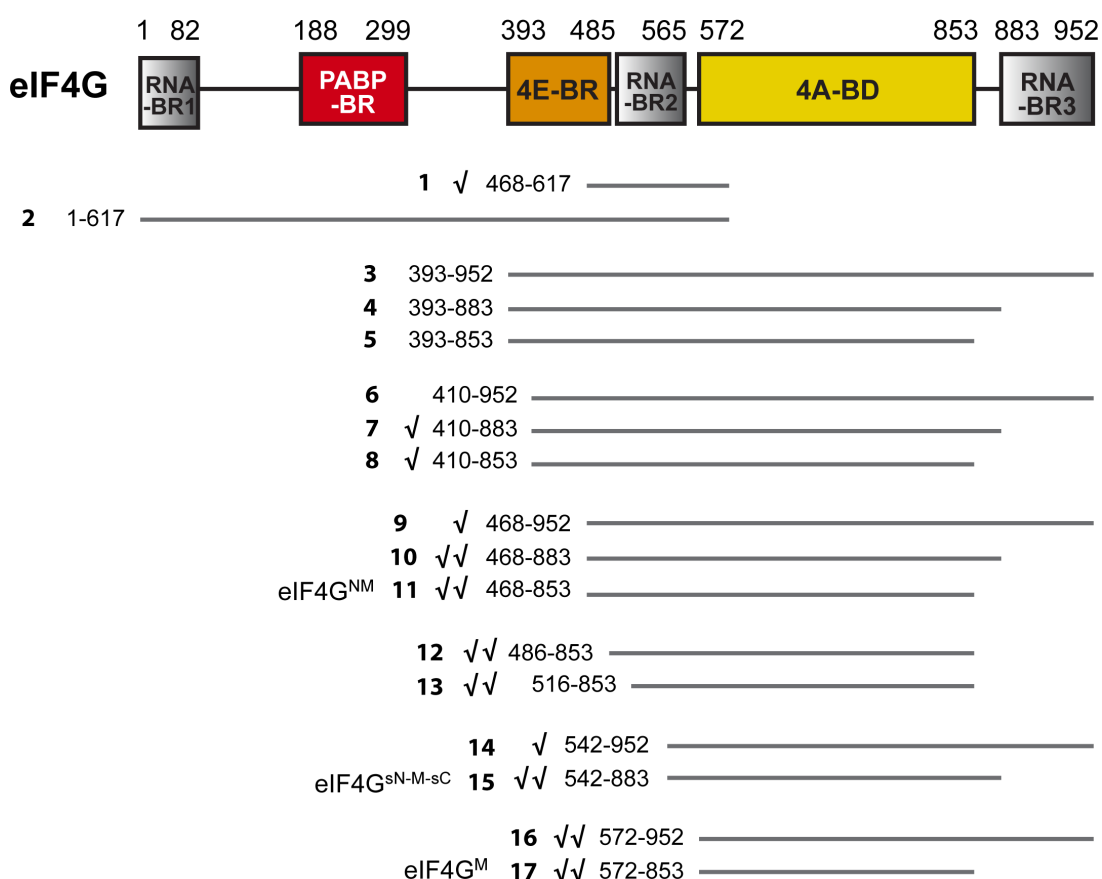
### 3.2.1.4 Crystallization of the eIF4F complex

Although pure full-length eIF4F complex was obtained finally, crystallization trials with 4.5 mg/ml and 8 mg/ml protein at 4 and 20 °C yielded no promising hits. All eleven crystallization screens (Table 2.22) were tested.



### 3.2.2 Preparation of minimal functional eIF4F complex

Secondary structure prediction of eIF4G shows that this protein is largely unstructured, especially two thirds of the N-terminal region. The reason that crystallization of full-length eIF4G and the eIF4F complex was failed might be due to the highly flexible regions of eIF4G. Therefore, systematically screenings for soluble fragments encompassing the eIF4E and/or eIF4A binding regions (Figure 3.25) were performed. Among the 17 different constructs, 12 could be expressed in *E. coli* cells (indicated with ticks), 7 could be purified (labeled with two ticks).



**Figure 3.25:** Diagram illustrating the designed constructs of *sceIF4G*. The boundaries of the constructs are given. Constructs design is based on secondary structure prediction, sequence conservation and functional domains. Among the 17 different constructs, 12 could be expressed in *E. coli* cells (indicated with ticks), and 7 could be purified (labeled with two ticks). These constructs without tick could not be expressed in *E. coli* cells. eIF4G<sup>M</sup>: containing the minimal eIF4A-binding domain only; eIF4G<sup>sN-M-sC</sup>: containing the minimal eIF4A-binding domain and short N- and C-terminal tails; eIF4G<sup>NM</sup>: containing the minimal eIF4A-binding domain and the adjacent basic residue-rich fragment.

### 3.2.2.1 Activities of eIF4A stimulated by eIF4G and eIF4B

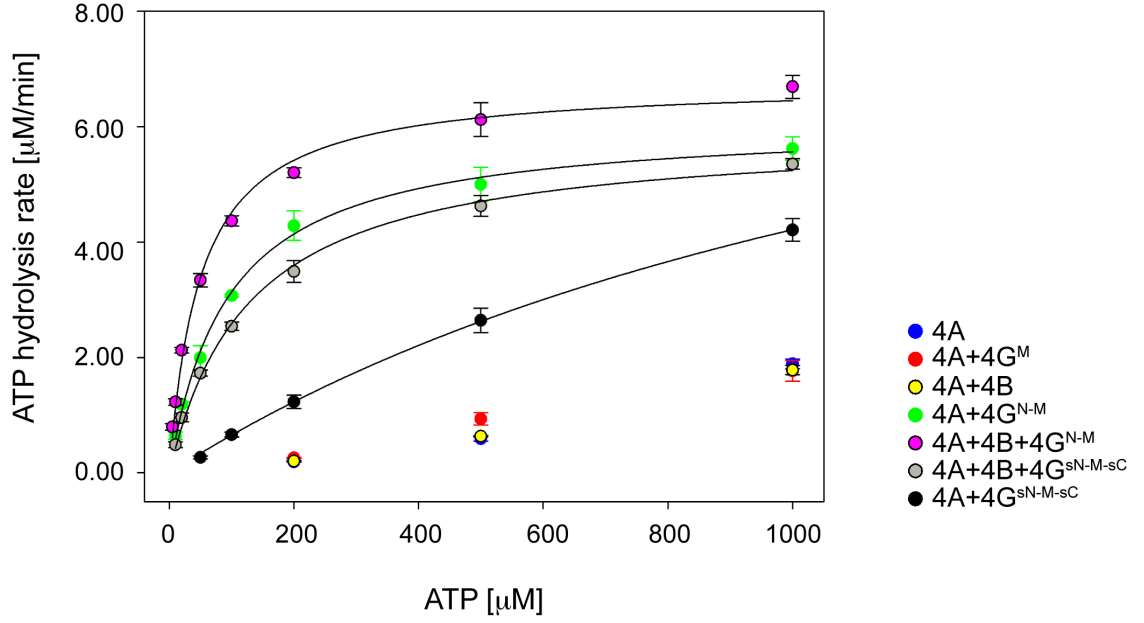
It is known that full-length eIF4G is able to enhance the ATPase and helicase activities of eIF4A. To characterize whether the purified truncations of eIF4G still possess the stimulation ability and to further investigate the underlying mechanism of the stimulation, real-time ATPase activity assays were first performed.

#### ATPase activity

The ATPase activity was carried out using the *EnzCheck*<sup>®</sup> Phosphate Assay kit (Invitrogen) which quantitatively detects the inorganic phosphate (P<sub>i</sub>) that is released upon ATP hydrolysis (see Section 2.3.10). In each reaction, 1 μM protein(s) and 5 μM 22-nucleotide single stranded RNA (ssRNA<sub>15161</sub>, Table 2.21) were used. Consistent with previous studies, eIF4A alone exhibits very low ATP hydrolysis activity (Figure 3.26, blue). Interestingly, eIF4G<sup>572-853</sup> (later called eIF4G<sup>M</sup>: containing the minimal eIF4A-binding domain only) did not stimulate the ATPase activity of eIF4A as reported before [108] (Figure 3.26, red). This might be due to the different methods used (see Discussion 4.5). Because of the extremely low enzymatic activity in the case of *apo* eIF4A and eIF4A/eIF4G<sup>M</sup>, it is impossible to reach the saturation and thus evaluate the kinetic statistics.

Except eIF4G<sup>M</sup>, the remaining 6 purified eIF4G truncations (Figure 3.25, labeled with two ticks) stimulate the ATPase activity of eIF4A to different extents. In the primary round of experiments, the representative eIF4G<sup>M</sup>, eIF4G<sup>468-853</sup> and eIF4G<sup>542-883</sup> were chosen for determining the kinetic parameters, and thus for systematic comparison. eIF4G<sup>468-853</sup> (later called eIF4G<sup>N-M</sup>) contains the minimal eIF4A-binding domain and the adjacent N-terminal fragment, while eIF4G<sup>542-883</sup> (later called eIF4G<sup>sN-M-sC</sup>) contains the minimal eIF4A-binding domain flanked by short fragments at both N- and C-termini. In the presence of eIF4G<sup>N-M</sup> (Figure 3.26, green), the maximal velocity (V<sub>max</sub>) of the ATP hydrolysis by eIF4A is  $6.09 \pm 0.12 \mu\text{M} \cdot \text{min}^{-1}$ , while the  $K_M$  and  $k_{cat}/K_M$  are  $94.13 \pm 6.13 \mu\text{M}$  and  $0.065 \mu\text{M}^{-1} \cdot \text{min}^{-1}$ , respectively (Table 3.3). Compared to eIF4G<sup>N-M</sup>, the stimulation ability of eIF4G<sup>sN-M-sC</sup> (Figure 3.26, black) is weaker. Notably, the estimated kinetic

parameters of eIF4A/eIF4G<sup>sN-M-sC</sup> are questionable because the saturation is not reached.



**Figure 3.26:** The ATPase activity was determined using the *EnzCheck*<sup>®</sup> Phosphate Assay kit (Invitrogen) which quantitatively detects the inorganic phosphate ( $P_i$ ) that is released upon ATP hydrolysis. ATPase activity of eIF4A (blue), eIF4A/eIF4G<sup>M</sup> (red), eIF4A/eIF4B (yellow), eIF4A/ eIF4G<sup>sN-M-sC</sup> (black), eIF4A/eIF4B/eIF4G<sup>sN-M-sC</sup> (gray), eIF4A/eIF4G<sup>N-M</sup> (green), eIF4A/eIF4B/eIF4G<sup>N-M</sup> (magenta). In each reaction, 1  $\mu$ M protein(s) and 5  $\mu$ M 22-nucleotide ssRNA<sub>15161</sub> (see Table 2.21) were used. Every reaction was measured at least three times. The circles represent the mean value, and the error bars indicate the standard deviations.

**Table 3.3:** Kinetic parameters describing the ATPase activity of eIF4A/eIF4B/eIF4G<sup>N-M</sup>, eIF4A/eIF4G<sup>N-M</sup>, eIF4A/eIF4B/eIF4G<sup>sN-M-sC</sup>, eIF4A/eIF4G<sup>sN-M-sC</sup>.

	4A+4B+4G <sup>M</sup>	4A+4G <sup>M</sup>	4A+4B+4G <sup>sN-M-sC</sup>	4A+4G <sup>sN-M-sC</sup> *
$V_{\max}$ ( $\mu\text{M} \cdot \text{min}^{-1}$ )	6.77 $\pm$ 0.10	6.09 $\pm$ 0.12	5.92 $\pm$ 0.10	10.69 $\pm$ 1.10
$K_M$ ( $\mu\text{M}$ )	50.11 $\pm$ 2.84	94.13 $\pm$ 6.13	128.88 $\pm$ 6.94	1535.6 $\pm$ 235.5
$k_{\text{cat}}$ ( $\text{min}^{-1}$ )	6.77 $\pm$ 0.10	6.09 $\pm$ 0.12	5.92 $\pm$ 0.10	10.69 $\pm$ 1.10
$k_{\text{cat}}/K_M$ ( $\mu\text{M}^{-1} \cdot \text{min}^{-1}$ )	0.135	0.065	0.046	0.007

\*: in the case of eIF4A/eIF4G<sup>sN-M-sC</sup>, the evaluated kinetic parameters are questionable due to the limitation of the low activity.

Intriguingly, eIF4B alone showed no effect on the ATPase hydrolysis rate of eIF4A at all, but addition of eIF4B to eIF4A/eIF4G<sup>N-M</sup> and eIF4A/eIF4G<sup>sN-M-sC</sup> induced a further enhancement (Figure 3.26, magenta and gray). The  $V_{max}$  of eIF4A/eIF4G<sup>N-M</sup> increased from  $6.09 \pm 0.12 \mu\text{M} \cdot \text{min}^{-1}$  to  $6.77 \pm 0.10 \mu\text{M} \cdot \text{min}^{-1}$  by eIF4B, while the  $K_M$  decreased from  $94.13 \pm 6.13 \mu\text{M}$  to  $50.11 \pm 1.84 \mu\text{M}$  (Table 3.3). Thus, in the presence of eIF4B, the  $k_{cat}/K_M$  value was elevated additionally ~2-fold relative to that of eIF4A/eIF4G<sup>NM</sup>, indicating higher enzymatic efficiency.

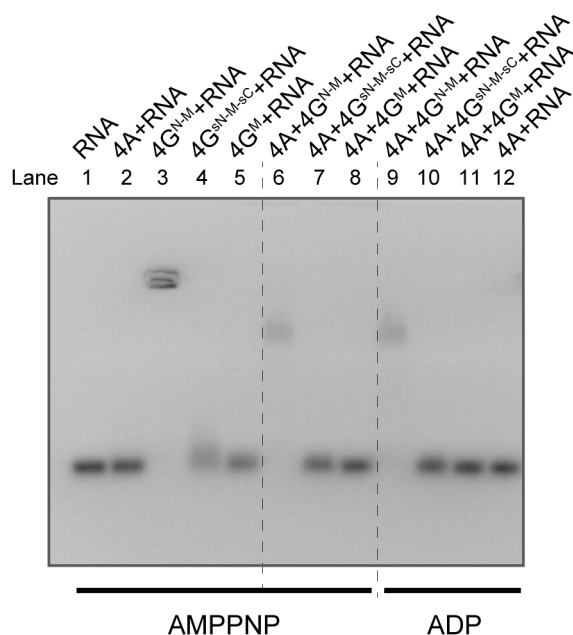
### Helicase activity

The RNA unwinding activity of eIF4A was tested in the absence and presence of eIF4B, eIF4G<sup>N-M</sup> and eIF4G<sup>sN-M-sC</sup>, using a fluorescence-based assay (see Section 2.3.11). However, no helicase activity was detected for all samples (data not shown). This might be due to the RNA substrate (originally designed for the helicase Prp43 [163]), which is not appropriate for eIF4A (see Discussion 4.4).

#### 3.2.2.2 RNA-binding ability of eIF4A and eIF4G variants

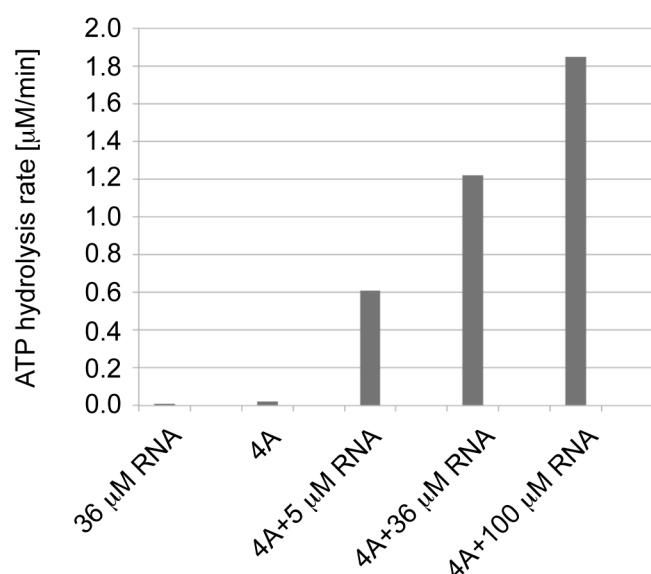
Electrophoretic mobility shift assays (EMSA) were performed to study the potential RNA-binding capability of eIF4G variants, and eIF4A. In the presence of eIF4A, the 22-nucleotide single stranded RNA<sub>15161</sub> (see Table 2.21) migrated slightly slower in the agarose gel (Figure 3.27, Lane 2), consistent with previously reported weak RNA-binding affinity of eIF4A. eIF4G<sup>N-M</sup>, eIF4G<sup>sN-M-sC</sup> and eIF4G<sup>M</sup> alone showed strong, medium and weak effect on the RNA migration, respectively (Figure 3.27, Lane 3-5). Reduced effects on the migration of RNA were observed when eIF4G variants were in complex with eIF4A (Figure 3.27, Lane 6-11). However, this does not mean eIF4A/eIF4G complexes possess lower RNA-binding ability because the total charge state of the complexes also impairs RNA migration. Nevertheless, it is obvious that the RNA-binding capability reflected by the influence on RNA migration corresponds to the enzymatic stimulation ability. Stronger RNA-binding reflects stronger enhancement on ATPase activity, and vice versa (Section 3.2.2.1). In addition, the RNA-binding ability of eIF4A/eIF4G complex is shown to be

nucleotide independent. Almost identical shifts were observed in the cases of addition of AMPPNP and ADP (Figure 3.27)



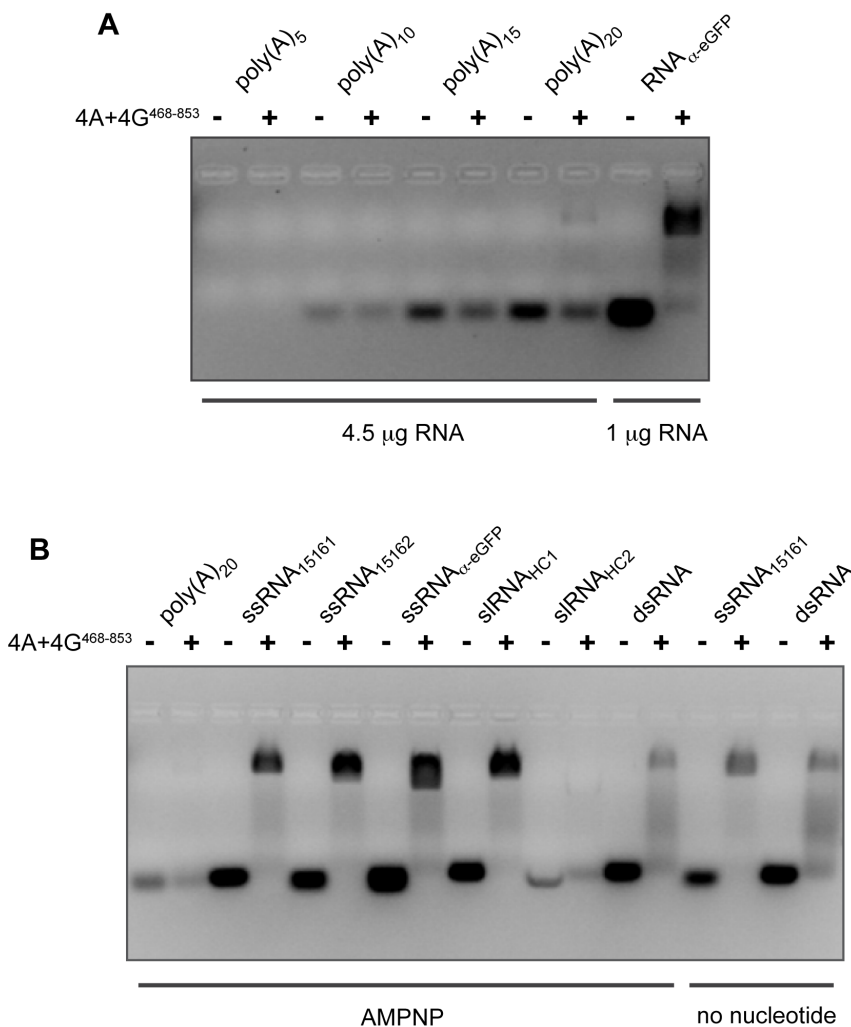
**Figure 3.27:** EMSA of ssRNA<sub>15161</sub> supplied with eIF4A and/or eIF4G variants. Consistent with the ATPase activity stimulation level, eIF4G<sup>N-M</sup>, eIF4G<sup>SN-M-sC</sup> and eIF4G<sup>M</sup> show strong, medium and weak effect on RNA migration, respectively. In each case, 1  $\mu$ g RNA, 5-fold molar excess of protein(s) and 50-fold molar excess of nucleotide were used.

Assuming that the stimulating effect of eIF4G is achieved through providing RNA for eIF4A, the ATP hydrolysis rate of eIF4A alone should increase if eIF4A has higher accessibility to RNA substrates. Therefore, the ATPase activity of eIF4A itself was tested in the presence of ssRNA<sub>15161</sub> with increasing concentrations. The data showed that the ATP hydrolysis rate of eIF4A was elevated 2- and 3-fold as the RNA concentration rose from 5  $\mu$ M to 36  $\mu$ M and 100  $\mu$ M, respectively (Figure 3.28).



**Figure 3.28:** ATP hydrolysis rate by eIF4A supplied with different amount of ssRNA<sub>15161</sub>. The ATPase activity of eIF4A was elevated 2- and 3-fold as the RNA concentration increased from 5  $\mu$ M to 36  $\mu$ M and 100  $\mu$ M, respectively. In each reaction, 1  $\mu$ M eIF4A and 500  $\mu$ M ATP were used.

Then, the shortest and best RNA candidates for crystallization of the active eIF4A-eIF4G<sup>N-M</sup> complex were screened. RNAs poly(A)<sub>5</sub>, poly(A)<sub>10</sub>, poly(A)<sub>15</sub> and poly(A)<sub>20</sub>, as well as single stranded RNA with ~20 nucleotides and ~20-base pair double stranded RNA (Sequences of RNAs see Table 2.21). It turned out that the eIF4A-eIF4G<sup>N-M</sup> complex had the highest ability to those ssRNAs consisting of all the four bases, relatively strong ability to dsRNAs and lowest ability to poly-adenine RNAs (Figure 3.29 A). In the case of ssRNAs containing all the four bases, they had more or less the same shift in the agarose gel upon addition of the eIF4A-eIF4G<sup>N-M</sup> complex (Figure 3.29 B), indicating similar RNA-binding affinity.



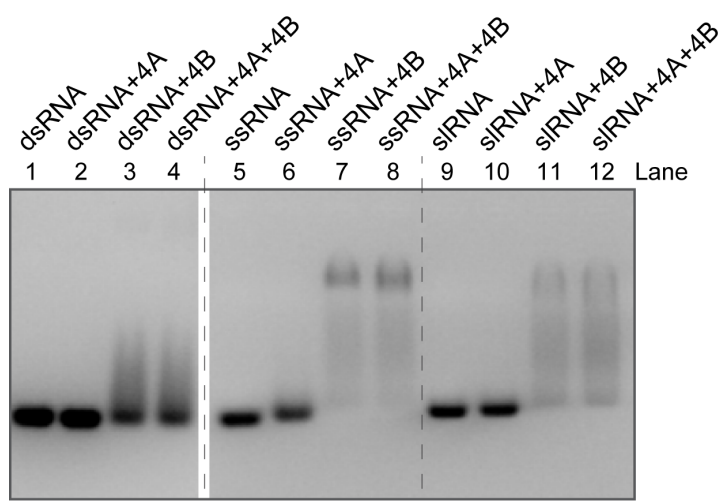
**Figure 3.29:** Selection of RNA substrates for eIF4A/eIF4G<sup>N-M</sup> complex. (A) EMSA of poly(A)<sub>5</sub>-, poly(A)<sub>10</sub>-, poly(A)<sub>15</sub>- and poly(A)<sub>20</sub>-RNA, as well as ssRNAα-eGFP (21 nucleotides) with 5-fold molar excess of eIF4A/eIF4G<sup>468-853</sup>. A clear shifted band is only observed in the case of poly(A)<sub>20</sub>-RNA. Although higher amount of poly(A) RNAs (4.5 μg) were used, compared to ssRNAα-eGFP (1 μg), weaker bands were observed. This is due to low affinity of the GelRed dye to the poly(A) RNAs.

(B) EMSA of ~20 nucleotide ssRNAs and ~20-base pair dsRNA with 5-fold molar excess of eIF4A/eIF4G<sup>N-M</sup>. eIF4A/eIF4G<sup>N-M</sup> prefers ssRNA with all four bases. 50-fold molar excess of AMPPNP was used in the indicated cases. siRNA: RNAs containing a stem loop.

### 3.2.2.3 Interaction test between eIF4A and eIF4B

To study the mechanism by which eIF4B synergistically promotes the activities of eIF4A, series of interaction assays between eIF4A and eIF4B were performed in the absence and presence of RNA fragments, nucleotides and eIF4G variants. Different RNA oligos (double stranded, single stranded and RNA with a stem-loop; the sequences of RNAs see Table 2.21) and different nucleotides (ADP, ATP and AMPPNP) were tested. However, an interaction of eIF4B with eIF4A and/or eIF4G was not observed in all used experiments, including co-purification test, pull-down experiments, electrophoretic mobility shift assays (EMSA), or analytical size exclusion chromatography (SEC). This indicates that eIF4B either transiently or indirectly binds to eIF4A.

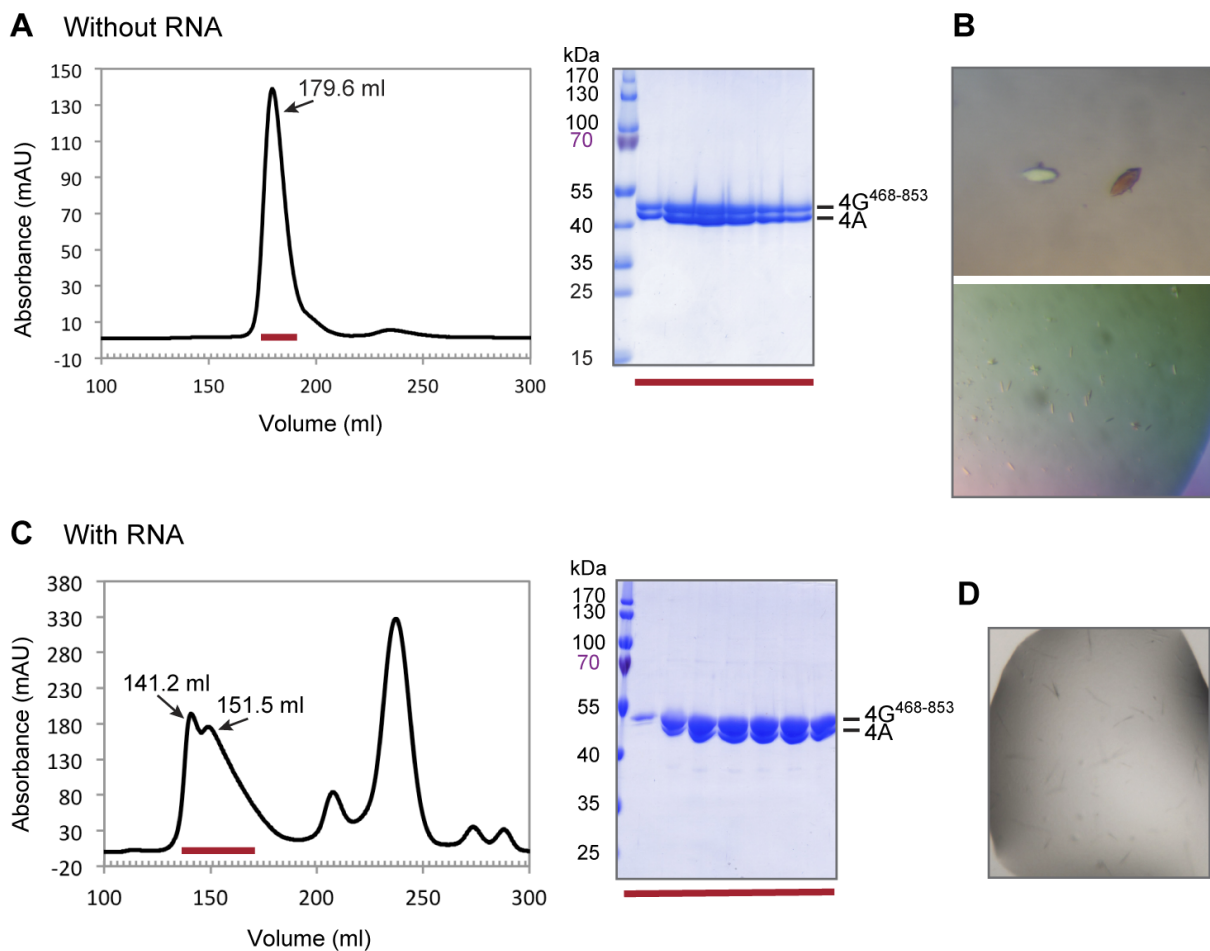
Next, the ability of eIF4B to interact with RNA was tested. In the case of adding eIF4B, both dsRNA and ssRNA migrated slower than free RNA (Figure 3.30), suggesting eIF4B could bind both types of RNA. However, it possesses higher capability to ssRNA than dsRNA since distinct supershifted bands were observed for ssRNA (Figure 3.30, Lane 7, 8), while diffuse higher mobility shifts near the free RNA were noted for dsRNA (Figure 3.30, Lane 3, 4). eIF4B is also able to bind RNA with a stem-loop structure (Figure 3.30, Lane 11, 12). Whether eIF4B enhances the activities of eIF4A through binding and providing RNA substrates for eIF4A or binding to the melted single stranded RNA to prevent the annealing is still elusive.



**Figure 3.30:** EMSA of single stranded (ss) RNA, double stranded (ds) RNA and stem-loop containing (sl) RNA supplied with eIF4A and/or eIF4B. Compared to dsRNA, greater shift is noted for ssRNA. 1  $\mu$ g RNA and 5-fold molar excess of protein(s) were used in each case.

### 3.2.2.4 Preparation and crystallization of the minimal active eIF4A-eIF4G complex

*S. cerevisiae* eIF4A and eIF4G<sup>N-M</sup> form a stable heterodimeric complex (with 1:1 stoichiometry) in low salt (100 mM NaCl) containing buffer (Figure 3.31 A). Crystallization of the eIF4A-eIF4G<sup>N-M</sup> complex produced rice-shaped crystals in a condition containing 1.5 M Li<sub>2</sub>SO<sub>4</sub>, 0.1 M Tris/HCl, pH 8.5 (Figure 3.31 B, upper panel). Optimization by varying the concentration of the salt, the type of the buffer and the pH resulted in more crystals (Figure 3.31 B, lower panel). However, bigger crystals that are suit for X-ray analysis could never be produced.



**Figure 3.31:** Purification and crystallization of the eIF4A/eIF4G<sup>N-M</sup> complex without and with RNA. (A) Gel filtration purification and (B) crystals of the eIF4A/eIF4G<sup>N-M</sup> complex in the absence of RNA. The crystals appeared from high salt containing buffer (1.5 M Li<sub>2</sub>SO<sub>4</sub>). Optimization only produced plenty of tiny crystals. (C) Gel filtration purification and (D) crystals of the eIF4A/eIF4G<sup>N-M</sup> complex supplied with 5-fold molar excess of AMPPNP and 2-fold molar excess of ssRNA<sub>α-eGFP</sub>. ssRNA was added prior to loading onto the gel filtration.



Since eIF4A-eIF4G<sup>N-M</sup> is able to bind RNAs, crystallization trials of the eIF4A-eIF4G<sup>N-M</sup>-ssRNA complex in the presence of 5-fold molar excess of AMPPNP were also performed. Upon addition of 2-fold molar excess of ssRNA <sub>$\alpha$ -eGFP</sub>, the eIF4A-eIF4G<sup>N-M</sup> complex eluted earlier from the Superdex 200 column, shifted from ~180 ml to 140-150 ml (Figure 3.31C). Crystallization trials of the eIF4A-eIF4G<sup>N-M</sup>-ssRNA-AMPPNP complex only yielded long needle crystals from buffer containing 0.2 M NaCl, 20% (w/v) PEG monomethyl ether 2000, 0.1 M MES pH 6.0 (Figure 3.31D). These crystals could not be used for X-ray analysis and improved further.

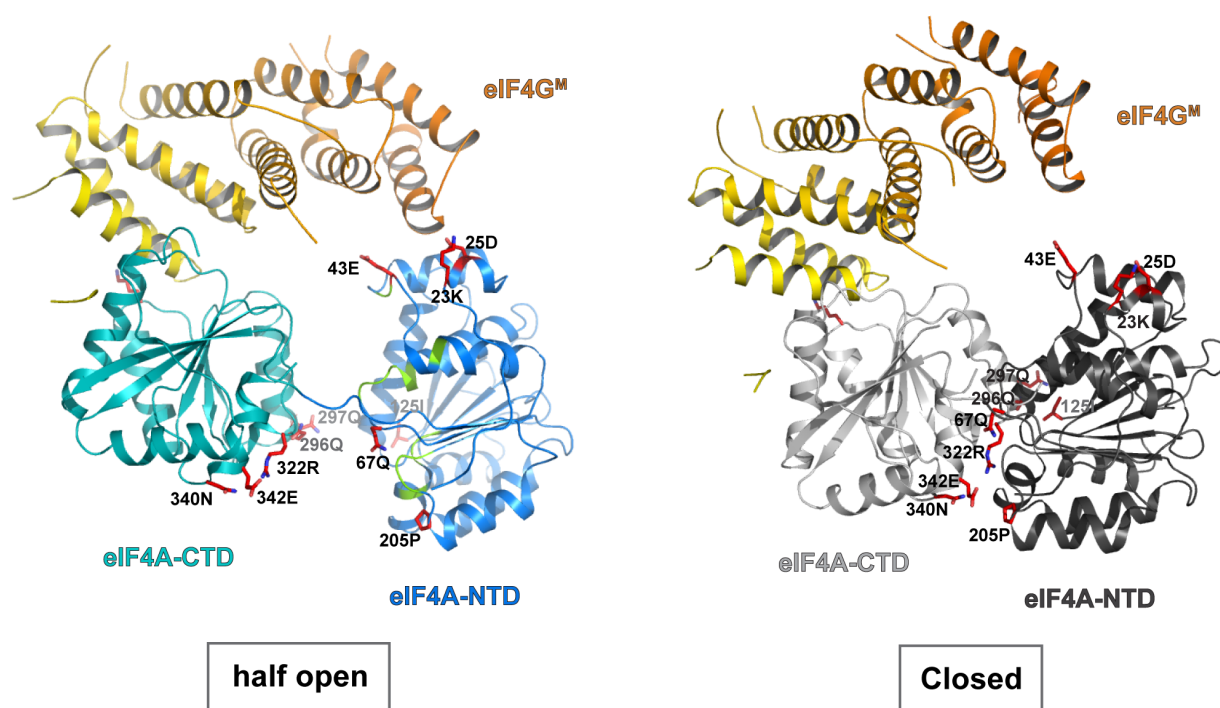
### 3.2.3 Monitoring the conformational change of eIF4A by crosslinking

To explore the structural alterations within eIF4A that are induced by nucleotides, RNA and/or auxiliary proteins, the UV-reactive unnatural amino acid *p*-benzoyl-L-phenylalanine (Bpa) was introduced into the ATP-dependent helicase eIF4A. Eight positions along the cleft of eIF4A two RecA-like domains and three positions in the eIF4G-binding interface of eIF4A-NTD were selected for individual replacement by Bpa (Figure 3.32). All selected residues are not involved in the formation of secondary structures and should not impair the folding and activities of eIF4A. Ten out of eleven Bpa substituted mutants could be expressed and purified (N340B: no expression).

For each Bpa mutant, its conformations in *apo*, ADP-, AMPPNP-, RNA-, RNA and AMPPNP-bound states were studied in the absence or presence of eIF4G<sup>M</sup> or eIF4G<sup>N-M</sup>. In general, two types of crosslink bands (one at high molecular weight, migrating above 100 kDa; another just below nonexposed eIF4A) were observed upon UV irradiation. Since the crosslink bands at high molecular weight appeared only in the presence of eIF4G variants but not in these eIF4A alone experiments, and considering their size, these bands are assumed to represent intermolecular crosslinks between eIF4A and eIF4G. Whereas the faster migrating crosslink bands produced in the absence and presence of eIF4G are most likely due to

intramolecular crosslinking within eIF4A. These intramolecularly crosslinked species possess UV-produced covalent bonds, which cannot be broken by SDS, and thus are more compact than their native counterpart. Therefore, although they possess the same amino acids (except the Bpa point mutation), crosslinked samples migrate faster in the SDS gels than native protein does.

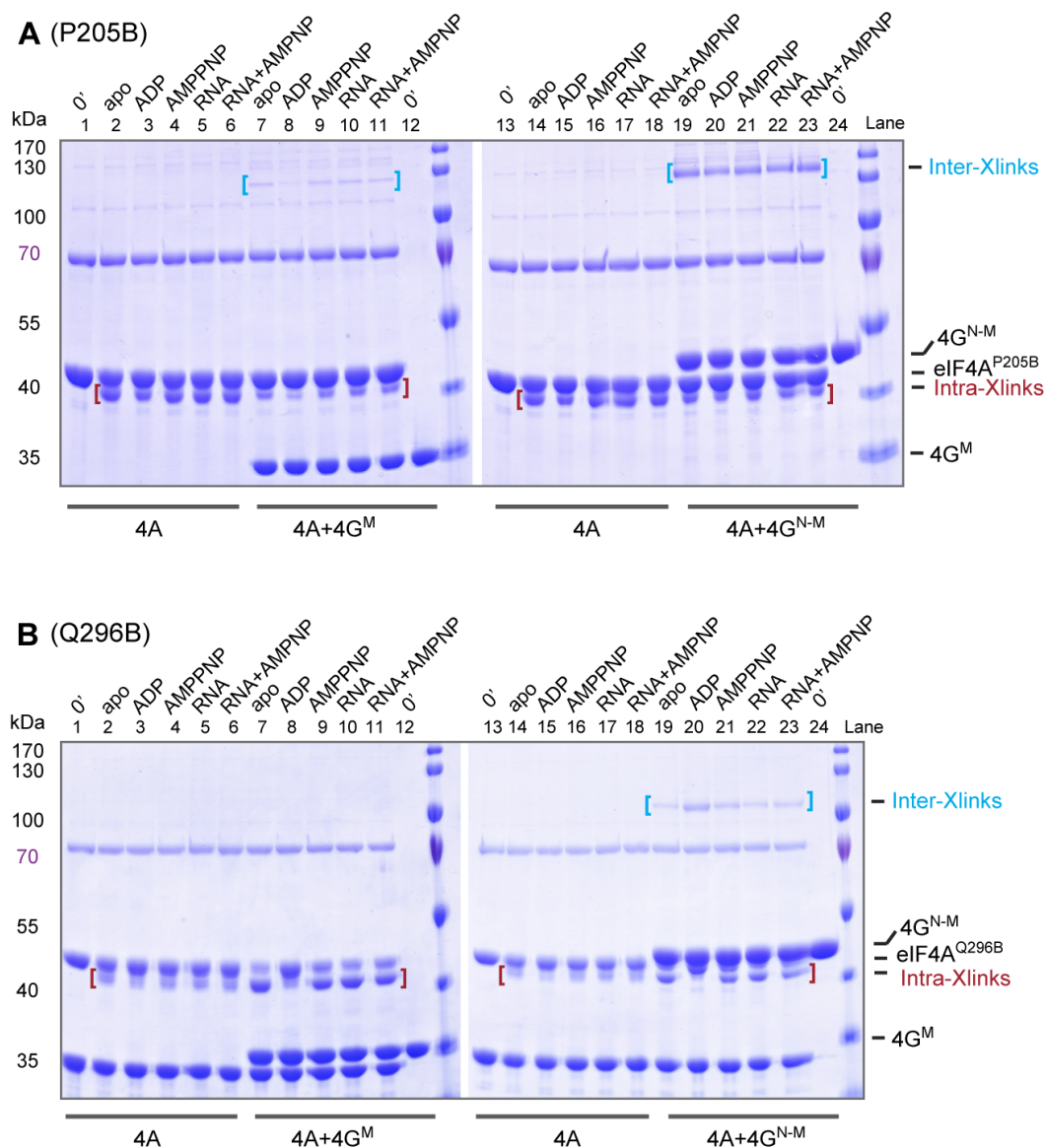
Generally, those intramolecular crosslink bands which would be an indication of a closed conformation were not observed as was expected. However, the data showed that the nucleotide ADP, RNA and eIF4G do trigger conformation changes within eIF4A as revealed by the crosslinks (see below).



**Figure 3.32:** Positions selected for replacement by Bpa showing in half open (PDB code: 2VSO) and closed conformations. The closed model was made based on the structure of eIF4A III. During modeling, eIF4A-CTD, which is the primary eIF4G-interacting domain, is fixed, while the NTD of eIF4A is moved and superimposed onto eIF4A III's NTD.

### 3.2.3.1 Binding of eIF4G induces conformational changes in eIF4A

Three Bpa point mutants (Q67B, P205B and Q296B) suggest a substantial conformational change in eIF4A upon binding of eIF4G (either eIF4G<sup>M</sup> or eIF4G<sup>N-M</sup>). Compared to eIF4A alone, addition of eIF4G to mutant P205B resulted in a

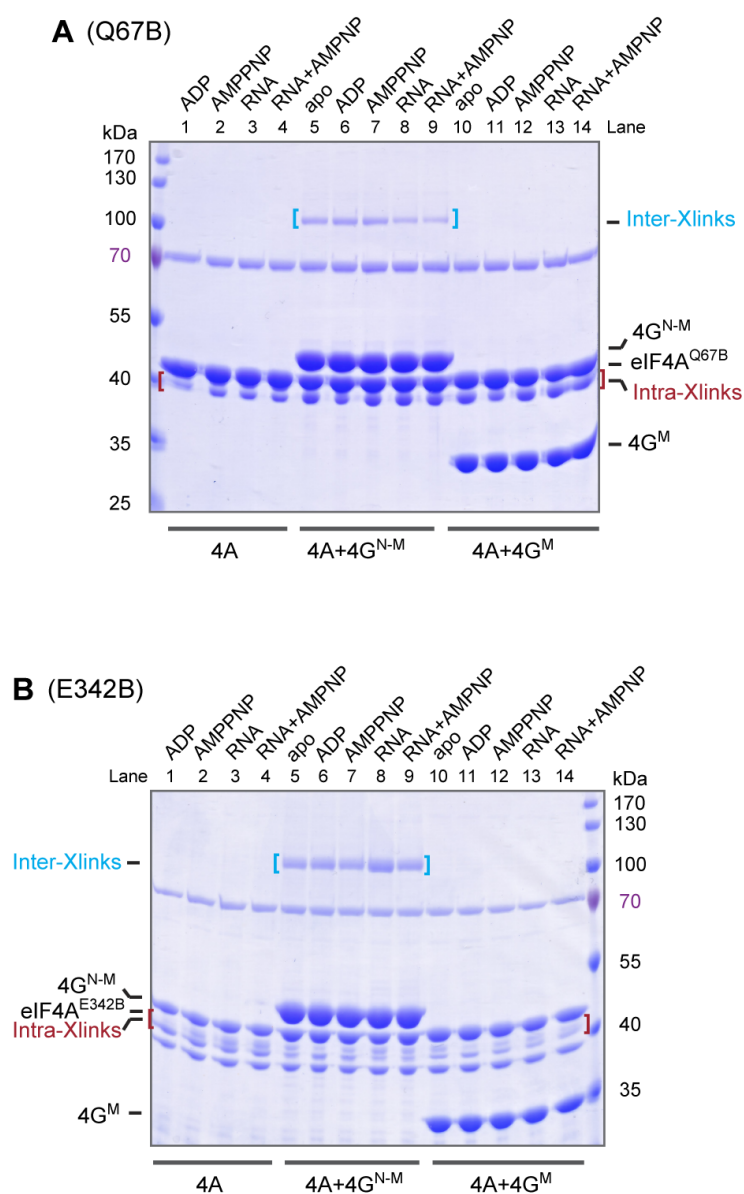


**Figure 3.33:** Crosslinking assays of P205B (A) and Q296B (B) in the absence or presence of eIF4G variants, nucleotides and/or RNA as labeled. Samples were exposed to the UV-light for 30 sec. The crosslink bands are indicated with brackets. 0': no UV exposure; Inter-Xlinks (cyan): intermolecular crosslinks between eIF4G and eIF4A; Intra-Xlinks (red): intramolecular crosslinks within eIF4A. Unlabeled protein bands represent impurities.

decrease of intramolecular crosslink (Figure 3.33A, Intra-Xlinks). Contrarily, in the presence of eIF4G, more accumulation of crosslinked sample was observed in the cases of Q67B and Q296B (except the ADP-bound form) (Figure 3.33 B and 3.34 A, Intra-Xlinks). Since the effect of eIF4G was noted not only for eIF4G<sup>N-M</sup> but also for eIF4G<sup>M</sup>, this conformational change in eIF4A seems to be induced by its association with the minimal eIF4A-binding of eIF4G (region 572-853).

### **3.2.3.2 Conformational changes of eIF4A upon ADP binding**

All the three mutants (Q67B, P205B and Q296B) displayed an intriguing feature. Compared to the *apo* and AMPPNP-bound form, addition of ADP caused a reduction of intramolecular crosslinks (Figure 3.33 and 3.34 A). Notably, in the cases of Q67B and P205B, this effect of ADP is independent of eIF4G. In other words, regardless of presence or absence of eIF4G, a decrease of crosslink bands was observed in the ADP-bound state (Figure 3.33A and 3.34A). By contrast, no reduction of crosslink was observed in ADP-bound form of Q296B if eIF4G was absent (Figure 3.33B). Lower level of intramolecular crosslinks upon addition of ADP was also observed for E342B, which showed two different crosslink bands (Figure 3.34B, Intra-Xlinks). Although the lower crosslink band kept the same in all reactions, the upper one got weak or disappeared when ADP was supplemented. The effect of ADP was also reflected by the intermolecular crosslinks. Upon binding of ADP, P205B generated less intermolecularly crosslinked species (Figure 3.33A), while Q296B produced more (Figure 3.33B). Additionally, in the case of D25B, stronger crosslinking occurred in the ADP-bound state (Figure 3.35B).

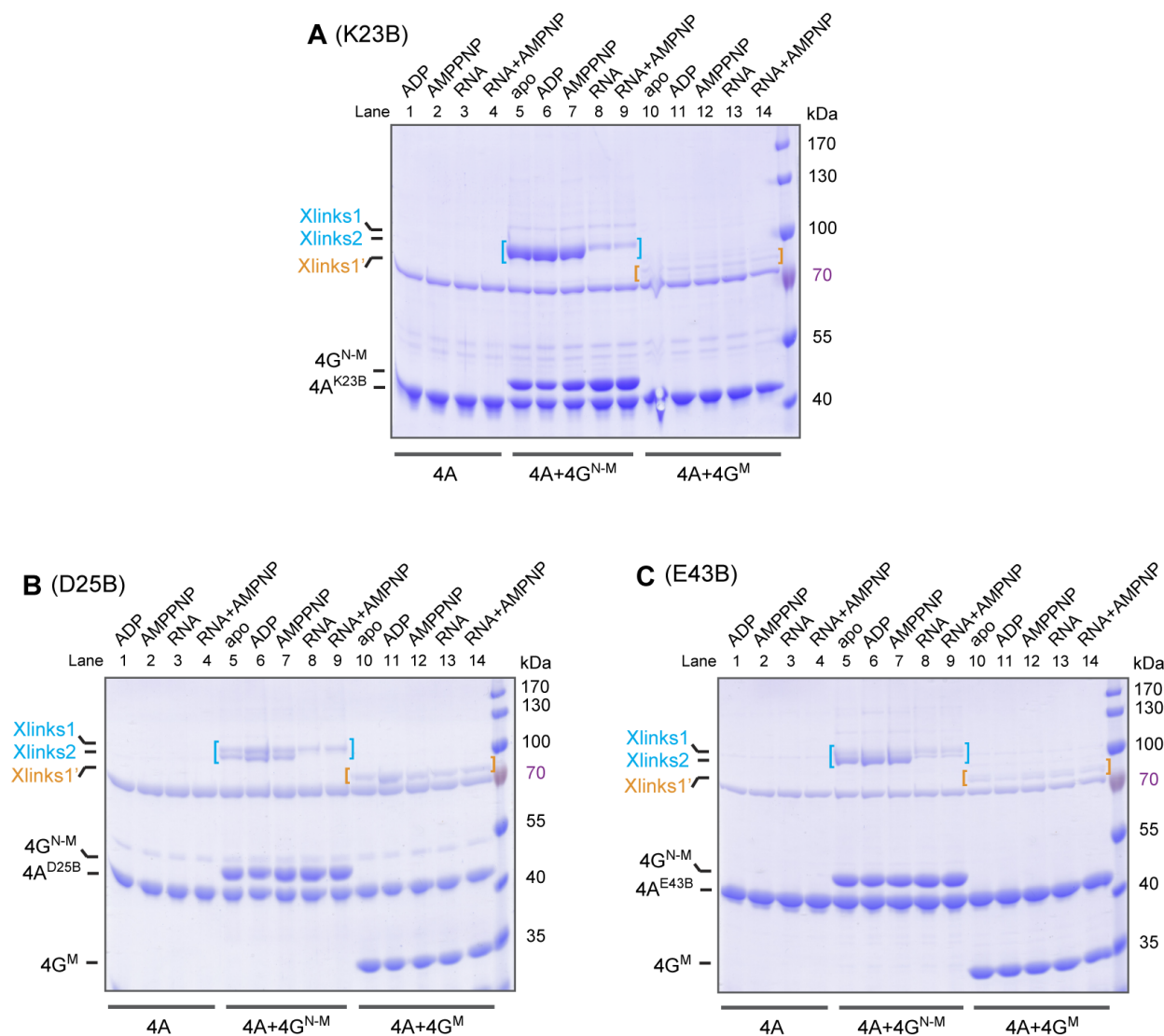


**Figure 3.34:** Crosslinking assays of Q67B (A) and E342B (B) in the absence or presence of eIF4G variants, nucleotides and/or RNA as labeled. Samples were exposed to UV-light for 30 sec. The crosslink bands are indicated with brackets. Inter-Xlinks: intermolecular crosslinks between eIF4G and eIF4A; Intra-Xlinks: intramolecular crosslinks within eIF4A. Unlabeled protein bands represent impurities.

### 3.2.3.3 Conformational changes in eIF4G-RBR2 upon RNA binding

Upon UV exposure, intermolecular crosslink bands were observed for eIF4A mutants K23B, D25B and E43B, which are located in the eIF4G-binding interface of eIF4A-NTD, in the presence of eIF4G<sup>M</sup> or eIF4G<sup>N-M</sup> (Figure 3.35A, B, C). As mentioned above, eIF4G<sup>M</sup> (residues 572-853) contains only the eIF4A-binding domain, while eIF4G<sup>N-M</sup> (residues 468-853) possesses the eIF4A-binding domain and the adjacent N-terminal RNA-binding region (RBR). This RBR of eIF4G is the second RNA-binding region of full-length eIF4G, and thus named RBR2. RBR2 of eIF4G is indispensable for stimulating eIF4A ATPase activity (Figure 3.26).

In the case of eIF4A/eIF4G<sup>M</sup>, one crosslink band migrating at ~75 kDa appeared in each reaction (Figure 3.35 A,B,C, Lane 10-14, labeled with Xlinks 1'), while in general two crosslink bands at ~100 kDa (Figure 3.35A,B,C, Lane 5-9, labeled as Xlinks 1 and 2, respectively) were observed when eIF4G<sup>N-M</sup> was supplied. The upper one of the two adjacent bands (Xlinks 1) kept the same pattern as Xlinks 1' in eIF4G<sup>M</sup>, while the lower one (Xlinks 2) exhibited distinct behaviors in the presence or absence of RNA oligos. When RNA was not supplied, this intermolecular crosslink band (Xlinks 2) clearly emerged, indicating that a considerable amount of eIF4A got crosslinked with eIF4G<sup>N-M</sup>. By contrast, upon addition of RNA, this lower crosslinking band (Xlinks 2) is significant weaker or even disappears. All together, the upper crosslink bands are assumed to result from the crosslinking between eIF4A and the minimal eIF4A-binding domain of eIF4G (residues 572-853), while the lower bands are a product of specific crosslinking between eIF4A and the eIF4G-RBR2 (residues 468-571). RNA-effected intermolecular crosslinks between eIF4A and eIF4G-RBR2 were also observed for eIF4A mutants Q67B and E342B, which showed a decrease and increase of crosslinks upon addition of RNA, respectively (Figure 3.34).



**Figure 3.35:** Crosslinking assays of K23B (A), D25B (B) and E43B (C) in the absence or presence of eIF4G variants, nucleotides and/or RNA as labeled. Samples were exposed to the UV-light for 30 sec. The crosslinked bands are indicated with brackets. Xlinks 1 and 2: intermolecular crosslinking between eIF4G<sup>468-853</sup> and eIF4A; Xlinks 1': intermolecular crosslinking between eIF4G<sup>572-853</sup> and eIF4A. Xlinks 1 and Xlinks 1' should be crosslinks occurred the same position in eIF4G. Unlabeled protein bands represent impurities. No intramolecular crosslinks were observed for the eIF4A K23B, D25B and E43B mutants.

## *Results*



## 4

## Discussion

In this thesis, the functional core complex (containing eIF3a, b, c, g and i) of *C. thermophilum* eIF3 was reconstituted *in vitro*, and the subunit b serves as the major scaffold protein. In addition, the data showed that eIF3b, eIF3i and eIF3g of *C. thermophilum* are able to pairwise interact with each other. The eIF3i- and eIF3g-binding sites in eIF3b are both located in the C-terminal domain of eIF3b. All these results are consistent with studies in yeast [48,59] and present direct proofs that the interacting network of eIF3 is universally conserved.

The crystal structure of the central WD40 domain of the major scaffold protein eIF3b from *C. thermophilum* was determined *de novo* by means of MAD. *cteIF3b*-WD40 adopts a nine-bladed  $\beta$ -propeller architecture. The central channel of the propeller possesses wide top and narrow bottom, resembling an unusual circular truncated cone. Structure based multiple sequence alignment and secondary structure prediction strongly suggested that all eIF3b orthologs employ this nine-bladed  $\beta$ -propeller fold.

By revisiting the cryo-electron microscopy map of the 43S preinitiation complex, eIF3b could be localized on the 40S ribosomal subunit, underneath the shoulder of the 40S subunit. This placement is consistent with previous biochemical data [164], and is supported by our *in vitro* co-sedimentation and GST pull-down experiments, which show direct binding of eIF3b to the 40S subunit and to the isolated ribosomal protein rpS9e.

Furthermore, the expression of full-length *S. cerevisiae* eIF4G has been established using the MultiBac insect cell expression system. Pure full-length eIF4G and the eIF4F complex could be prepared. By testing the ATPase activity of eIF4A in the absence or presence of eIF4G variants and/or eIF4B, the minimal mRNA activation complex was determined. Although it was unable to get crystals of the eIF4F complex, the conformational changes within eIF4A and eIF4G were monitored using site-specific crosslinking experiment.

#### 4.1 Reconstitution of *C. thermophilum* eIF3 core complex

*Chaetomium thermophilum* is a thermophilic fungus living at temperatures up to 60 °C. Compared to proteins from mesophilic organisms, it has been noted that thermophilic proteins seem to be more stable and hence show improved properties for structural and biochemical investigations. Following the reconstitution studies of the *C. thermophilum* inner nuclear pore complex [165], several crystal structures of *C. thermophilum* proteins, of which their homologues from other species have never been crystallized, have been published [41,166,167]. These successful studies led us to identify and characterize subunits of eIF3 in this thermophilic eukaryote. Notably, 13 genes encoding eIF3 subunits were found in *C. thermophilum* genome, suggesting that *cteIF3* resembles mammalian eIF3.

The subunits b, c, g and i of *cteIF3* could be individually expressed in *E. coli* cells and purified through several purification steps (see Section 3.1.1.1). To purify *cteIF3a*, it is required to at least add *cteIF3c* to pass through the gel filtration column since eIF3a itself aggregates and sticks on the column. This reflects that eIF3 subunits have positive effects on each other and thus function cooperatively. In addition, pure *cteIF3* core complex (containing eIF3a, b, c, g and i) was successfully assembled *in vitro* (Figure 3.6), which provides a basis for further structural analysis using either X-ray crystallography or cryo-EM. Unlike its yeast counterpart [59], a stable eIF3b-c-i-g subcomplex could not be reconstituted *in vitro* from *C. thermophilum* proteins (Figure 3.6 D). This indicates that the interaction between eIF3c and the eIF3b-i-g subcomplex is either not universal and therefore

might not be essential for the assembly of the eIF3 complex, or depends on the presence of additional subunits of eIF3 in *C. thermophilum*.

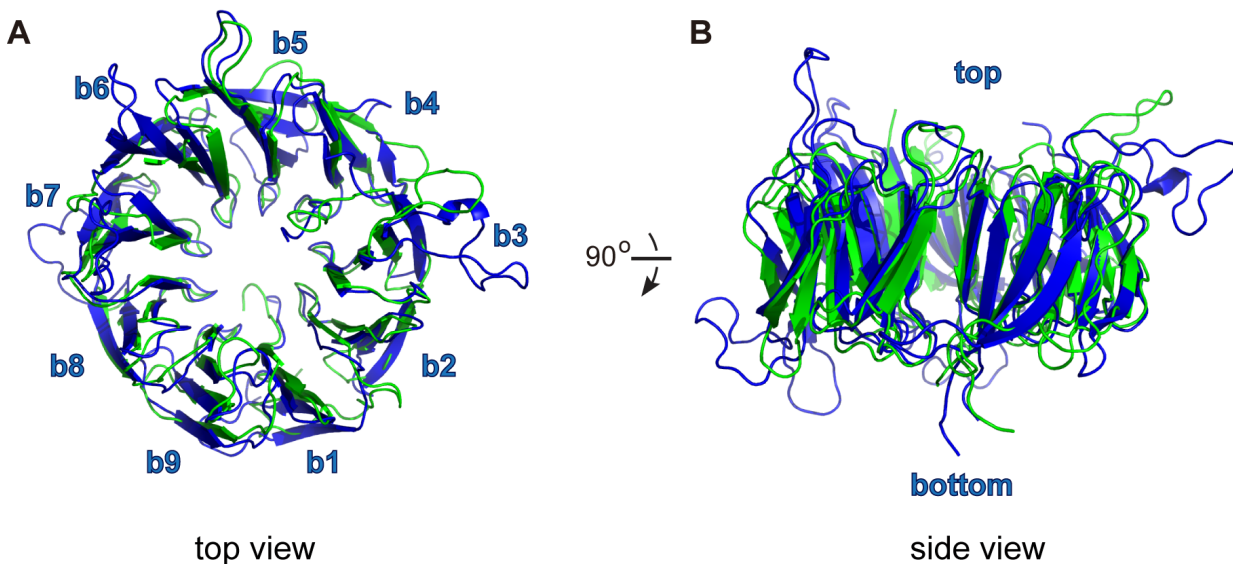
As reported, human eIF3a expressed in bacterial cells missed its C-terminal half, while the N-terminal third of eIF3c is truncated [168]. Rabbit eIF3 complex encounters similar truncation problem as isolated eIF3 from rabbit reticulocyte lysate contains at least half truncated eIF3a as well [61]. Since *C. thermophilum* contains all 13 eIF3 subunits and proteins from it are more stable (revealed by our limited proteolysis experiments), it is good to expand the targets of interest to *C. thermophilum* eIF3 proteins, and further obtain the complete eIF3 complex for dissecting the structure and function.

## 4.2 The nine-bladed $\beta$ -propeller fold of eIF3b-WD40

WD40 propellers differ widely in the composition and length of sequence, resulting in various numbers of blades. So far, four- to ten-bladed WD40  $\beta$ -propellers have been structurally identified, among which the seven-bladed propellers are the most abundant in the Protein Data Bank (PDB), followed by six and eight repeats [138,169]. The WD40 domain of eIF3b adopts a nine-bladed fold, although it has been predicted to fold into two tandem  $\beta$ -propellers [46]. The overall diameter of the propeller is up to 53 Å, which is 7 Å and 11 Å larger than eight- and seven-bladed propellers, respectively. Most of the four- to eight-bladed WD40  $\beta$ -propellers possess cylindrical central channels. Different from this feature, the central channel of eIF3b-WD40 propeller resembles a circular truncated cone shape. However, the reason for this arrangement is currently unclear.

By the time the structure of eIF3b-WD40 was solved, it represented the first nine-bladed  $\beta$ -propeller fold. It was tried to find whether other proteins beyond eIF3b orthologs also employ a nine-bladed fold using bioinformatics methods. By searching the NCBI conserved domain database (CDD) [170] eIF2A (implicated in promotion of initiator tRNA binding to the 40S in internal ribosome entry site-mediated (IRES) translation initiation) was identified as a candidate. Secondary structure prediction of eIF2A assigns 36  $\beta$ -strands, exactly matching the

strand number of a nine-bladed  $\beta$ -propeller. However, other WD40 proteins that adopt a nine-bladed fold were not verified. This might be due to the high degree of sequence degeneracy among WD40 domains and the limitation of current prediction tools, which make the prediction of the correct blade numbers and domain boundaries extremely difficult. Surprisingly, the crystal structure of *S. pombe* eIF2A was reported online recently, revealing a nine-bladed  $\beta$ -propeller architecture as predicted [171]. Structures of eIF2A-WD40 and eIF3b-WD40 superimpose well with an r.m.s.d. of 1.79 Å for 300 common C $\alpha$  atoms (Figure 4.1), although the sequence identity between the WD40 domains of eIF2A and eIF3b is only 20%.



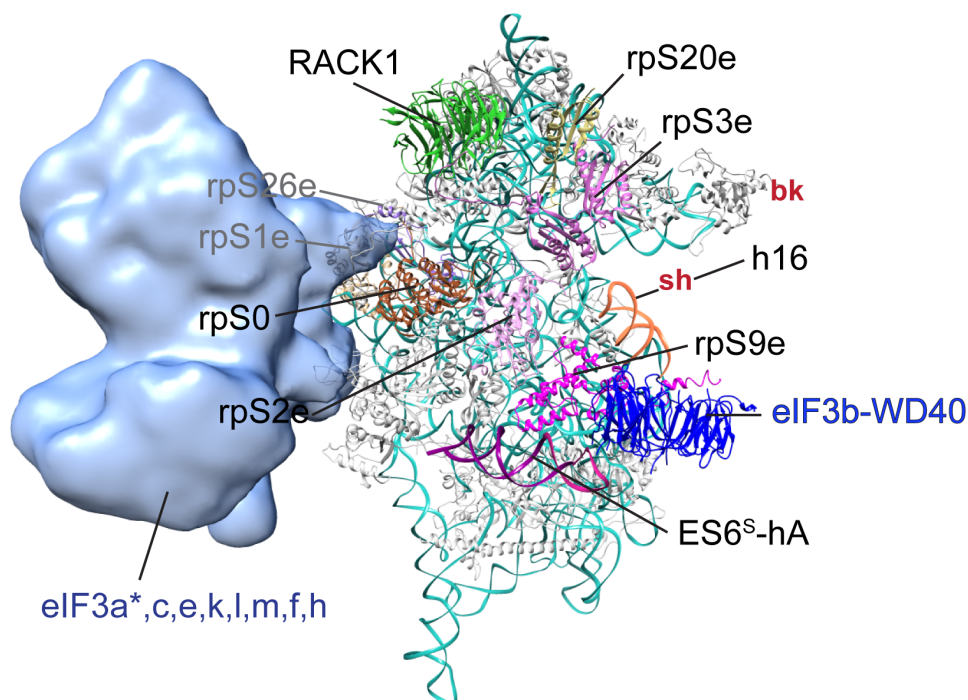
**Figure 4.1:** Superposition of eIF3b-WD40 (blue) and eIF2A-WD40 (green) showing in top view (A) and side view (B). The two structures superimpose well with r.m.s.d. of 1.79 Å for 300 common Ca atoms.

### 4.3 Localization of eIF3 on the 40S ribosomal subunit

Significant efforts have been made to map the interaction sites of eIF3 on the 40S subunit. It has been determined that the three largest subunits (a/Tif32, b/Prt1 and c/Nip1) play the pivotal role in association with the 40S subunit [164]. eIF3a binds to the ribosomal protein rpS0 and rpS1e via its NTD, while eIF3a-CTD interacts

with rpS2e, rpS3e as well as the helices 16-18 of the ribosomal RNA 18S [1,15,55,61,62]. eIF3c also directly binds to ribosome through the interaction with RACK1 [56]. By contrast, the interaction between eIF3b, the major scaffold subunit of eIF3, and the 40S subunit is mediated by another eIF3 subunit, eIF3j, which is not universally conserved [47]. This indirect interaction to some extent would affect the promotion efficiency of the multifactor complex by eIF3 to the ribosome.

Based on the cryo-EM map of the DHX29 bound 43S PIC (EMDB code: 5658) [61], eIF3b was localized underneath the shoulder of 40S and adjacent to rpS9e (Figure 3.16 and 4.2). This position of eIF3b seems to be rather far away from the eIF3a<sup>-</sup> and c-containing octameric complex of eIF3 (including subunits a, c, e, k, l, m, f and h) (Figure 3.16 and 4.2). However, in the 43S PIC, eIF3a is actually lacking the C-terminal ~600 amino acids, which possess the binding site for eIF3b [15]. Hence, the missing subunits (eIF3d, g, i and j) and flexible regions of eIF3 would sufficiently link eIF3b and the octameric complex.



**Figure 4.2:** eIF3 octamer (light blue) and eIF3b-WD40 (blue) bound 40S ribosomal subunit, showing the solvent-exposed side. The model is made based on the cryo-EM map of the DHX29 bound 43S PIC (EMDB code: 5658), and the PDB code of the 40S subunit structure is 4KZZ. The ribosomal proteins and RNA fragments that are involved in the interaction with eIF3 are labeled. sh: shoulder of the 40S subunit; bk: beak of the 40S subunit.

Based on the placement, a direct interaction between eIF3b and the 40S subunit has been identified for the first time (in collaboration with S. Schell and A. Chari). Full-length eIF3b and eIF3b<sup>167-670</sup> co-sedimented with the purified 40S subunit (Figure 3.18). In addition, full-length eIF3b and eIF3b<sup>167-C</sup> were pulled down by the GST tagged rpS9e, while eIF3b<sup>167-670</sup> did not (Figure 3.19). These results indicate at least two binding sites within the 40S subunit for eIF3b. (i) The ribosomal protein rpS9e interacts with the last 42 C-terminal amino acids (residues 705 to 746), since this fragment is indispensable for association with the 40S subunit (Figure 3.19). eIF3b<sup>705-746</sup> contains highly conserved residues across yeast to human (Figure 3.14). (ii) The WD40 domain plus the following 30 residues (eIF3b<sup>167-670</sup>) bind to other region(s) of the 40S subunit. eIF3b<sup>167-670</sup> was used because the WD40 domain itself precipitated in the 40S subunit buffer. Based on the structural modeling, it is proposed that the helix 16 and ES6<sup>S</sup>-hA of the ribosomal RNA (18S) could be involved in this interaction. The potential residues of eIF3b that mediate the association with the 40S subunit are these conserved basic and aromatic residues, for example, K434, K462 (with h16); R592, R635 and the cluster 646-659 (K646, R649, K650, R653, Y655, R657, F659) (with ES6<sup>S</sup>-hA). However, to validate the specific intermolecular bridges between eIF3b and the 40S, further work is needed.

The direct interaction between eIF3b and the 40S provides new insight into the eIF3·40S interaction network and suggests an essential role of the central scaffold subunit b of eIF3 in the formation of the 43S PIC. Our studies provide a structural basis for dissecting the mechanisms of the assembly of 43S and 48S PIC.

#### **4.4 RNA substrate of eIF4A**

eIF4A plays an essential role in the cap-dependent translation initiation by melting the secondary structures in the 5'-untranslated region of mRNA to facilitate binding of the 40S subunit and to enable subsequent scanning for the AUG start codon. However, eIF4A has its own 'specific' RNA substrate for unwinding. This specificity does not depend on the length nor the nucleotide sequence, but the overall

stability ( $\Delta G$  value), of the duplex substrate. It has been demonstrated that the initial rate and the amplitude of duplex unwinding by eIF4A is correlated with the stability of the secondary structure element [23]. eIF4A is only able to efficiently separate secondary structure of RNA having relative low stability (with less negative  $\Delta G$  value,  $\sim -24$  kcal/mol) [88]. To translate those mRNAs with highly structured 5' UTR, other helicases are needed. The DEAD/DExH-box mammalian proteins DHX29, DDX3 and RNA helicase A (RHA), as well as yeast Ded1p have been implicated in translation initiation [24,27].

This could explain why it was observed an unwinding by eIF4A/eIF4B/eIF4G<sup>N-M</sup> (which exhibited high ATP hydrolysis rate) when an RNA duplex with  $\Delta G = -65.0$  kcal/mol was used. This RNA substrate was chosen as an initial test because it was readily available. Further experiments will require fluorescence-dye labeled RNA with low stability ( $\Delta G = -19.8$  kcal/mol) according to the requirement of eIF4A. Systematic helicase activity assay will be performed soon.

## 4.5 Regulation of eIF4A activity by eIF4G and eIF4B

DEAD-box (and related DExH-box) helicases melt secondary structures of RNAs in an ATP-dependent manner and function in all aspects of RNA metabolism, *e.g.* transcription, mRNA splicing, translation, ribosome biogenesis, RNA modification and degradation, *etc* [80,172]. The majority of helicases from this family contains N- and/or C-terminal extensions flanking the helicase core. These extensions have been demonstrated to modulate interactions with ATP, RNA substrate or protein partners [81,82,173]. However, the archetypical member of the DEAD-box protein family, eIF4A, consists of the helicase core only. To efficiently melt the duplex in the RNA, eIF4A requires the assistance of auxiliary proteins [88,174].

It has been known that eIF4G and eIF4B (as well as eIF4H in mammalian cells) can promote the ATPase and helicase activities of eIF4A for more than decade [76,79]. Since the time of discovery, extensive studies were performed to understand the mechanism of this stimulation by eIF4G and eIF4B. So far, it is believed that

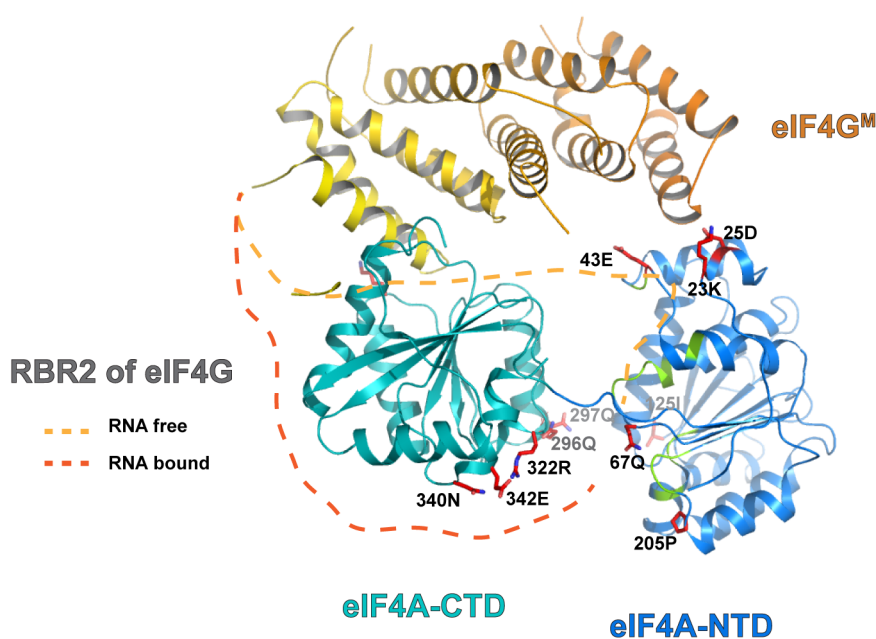
eIF4G enhances the enzymatic activities of eIF4A by inducing a more compact conformation (compared to the *apo* form) of eIF4A, through the direct interaction with eIF4A [108]. In this so-called 'half open' conformation, all the functional motifs of eIF4A are pre-arranged for the activation. Thus, the minimal eIF4A-binding of eIF4G (yeast eIF4G<sup>M</sup> (residues 572-853)) is thought to be the actual effector region. Consistently, it has been reported that yeast eIF4G<sup>M</sup> is able to increase the ATP hydrolysis rate by 3- to 5-fold [86,108]. However, in the ATPase activity assay studied in this thesis, there is no enhancement appearing at all (Figure 3.26). In this thesis, 1  $\mu$ M of proteins and 5-fold molar excess of RNA (i.e. 5  $\mu$ M) were used for the ATPase activity assays, which is reasonable. By contrast, in those assays that noted 3- to 5-fold stimulation, up to 1 mM (1000 molar excess to protein) RNA was used to reach the saturation [86,108]. Basically, this high ratio of RNA to protein is unlikely to represent physiological conditions. It has been reported that high substrate concentration and crowded environment do enhance the ATPase activity of eIF4A [175].

The ATPase activity experiment and electrophoresis mobility shift assay in this thesis revealed that the extent of the stimulation by eIF4G to the ATPase activity of eIF4A correlates with the RNA-binding ability of eIF4G (Figure 3.26 and 3.27). The eIF4G variant that displays the highest enzymatic stimulation (i.e. eIF4G<sup>N-M</sup> (residues 468-853)) also exhibits the strongest RNA-binding ability. By contrast, eIF4G<sup>M</sup> shows lowest RNA-binding affinity and no enhancement to eIF4A activity. Thus, eIF4G is proposed to promote the activities of eIF4A through the interaction with both eIF4A and RNA, thus recruiting the RNA substrate for eIF4A. Hence, as RNA is necessary for efficient ATPase and helicase activities of eIF4A, eIF4G<sup>M</sup> shows no stimulatory effect.

This proposal was further corroborated by our site-specific UV-induced crosslinking experiments. eIF4A mutants K23B, D25B and E43B (located at the eIF4G-binding interface of eIF4A-NTD (Figure 3.32)) substantially crosslinked with eIF4G<sup>N-M</sup> in the absence of RNA, whereas the crosslinks were significantly reduced in the presence of RNA (Figure 3.35). The extent of the crosslink reflects the accessibility of crosslinked region(s) to the Bpa positions. A mutant located at the



upper side of the cleft of the two RecA-like domains of eIF4A (Q67B) showed similar results (Figure 3.34 A). By contrast, upon addition of RNA, more crosslink sample was observed for a mutant carrying Bpa adjacent to the RNA-binding region of eIF4A (E342B) (Figure 3.34 B). As described before (Section 3.2.3), these intermolecular crosslinks most likely occurred between eIF4A and the second RNA-binding region (RBR2, residues 468-571) of eIF4G. Together, the crosslinking studies indicate a substantial conformational change within the RBR2 of eIF4G. In the absence of RNA, the eIF4G-RBR2 is flexible and reaches the eIF4G-binding interface of eIF4A-NTD. Whereas in the presence of RNA, the eIF4G-RBR2 binds the RNA and this RNA-binding places eIF4G-RBR2 to the vicinity of eIF4A RNA-binding region (Figure 4.3). The data implicate a role of the RBR2 of eIF4G in directly providing RNA substrate for the eIF4A and thus enhancing its activities.



**Figure 4.3:** Conformational change of eIF4G-RBR2 upon RNA binding. The structures of eIF4G (orange) and eIF4A (blue and cyan) are shown in cartoon, while the RNA-binding region 2 of eIF4G (RBR2) is indicated with dashed lines. In the absence of RNA, the RBR2 of eIF4G is flexible and reaches the eIF4G-binding interface of eIF4A-NTD. In the presence of RNA, eIF4G-RBR2 binds the RNA and accesses the vicinity of eIF4A RNA-binding region.

eIF4B also showed a stimulation of the ATPase activity of eIF4A when eIF4G was present (Figure 3.26). However, a direct interaction of eIF4B with eIF4A, eIF4G or the eIF4A/4G complex was not observed (Section 3.2.2.3). It has been reported that eIF4B works in conjunction with eIF4G to induce a closed conformation of eIF4A [174]. However, how eIF4B contributes to this action is current unclear. Notably, eIF4B stimulates the ATPase activity of eIF4A only when eIF4G variants that possess RNA-binding ability are present, which indicates a modulatory role of eIF4B in the stimulation.

# References

1. Valasek L, Mathew AA, Shin BS, Nielsen KH, Szamecz B, et al. (2003) The yeast eIF3 subunits TIF32/a, NIP1/c, and eIF5 make critical connections with the 40S ribosome in vivo. *Genes Dev* 17: 786-799.
2. Majumdar R, Bandyopadhyay A, Maitra U (2003) Mammalian translation initiation factor eIF1 functions with eIF1A and eIF3 in the formation of a stable 40 S preinitiation complex. *J Biol Chem* 278: 6580-6587.
3. Asano K, Shalev A, Phan L, Nielsen K, Clayton J, et al. (2001) Multiple roles for the C-terminal domain of eIF5 in translation initiation complex assembly and GTPase activation. *EMBO J* 20: 2326-2337.
4. Valasek L, Nielsen KH, Hinnebusch AG (2002) Direct eIF2-eIF3 contact in the multifactor complex is important for translation initiation in vivo. *Embo J* 21: 5886-5898.
5. Dennis MD, Person MD, Browning KS (2009) Phosphorylation of plant translation initiation factors by CK2 enhances the in vitro interaction of multifactor complex components. *J Biol Chem* 284: 20615-20628.
6. Sokabe M, Fraser CS, Hershey JW (2012) The human translation initiation multi-factor complex promotes methionyl-tRNA<sub>i</sub> binding to the 40S ribosomal subunit. *Nucleic Acids Res* 40: 905-913.
7. Gingras AC, Raught B, Sonenberg N (1999) eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu Rev Biochem* 68: 913-963.
8. Wells SE, Hillner PE, Vale RD, Sachs AB (1998) Circularization of mRNA by eukaryotic translation initiation factors. *Mol Cell* 2: 135-140.
9. Lamphear BJ, Kirchweger R, Skern T, Rhoads RE (1995) Mapping of functional domains in eukaryotic protein synthesis initiation factor 4G (eIF4G) with picornaviral proteases. Implications for cap-dependent and cap-independent translational initiation. *J Biol Chem* 270: 21975-21983.
10. Tarun SZ, Jr., Sachs AB (1996) Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G. *EMBO J* 15: 7168-7177.
11. Imataka H, Sonenberg N (1997) Human eukaryotic translation initiation factor 4G (eIF4G) possesses two separate and independent binding sites for eIF4A. *Mol Cell Biol* 17: 6940-6947.

## References

12. LeFebvre AK, Korneeva NL, Trutschl M, Cvek U, Duzan RD, et al. (2006) Translation initiation factor eIF4G-1 binds to eIF3 through the eIF3e subunit. *J Biol Chem* 281: 22917-22932.
13. Pyronnet S, Imataka H, Gingras AC, Fukunaga R, Hunter T, et al. (1999) Human eukaryotic translation initiation factor 4G (eIF4G) recruits mnk1 to phosphorylate eIF4E. *EMBO J* 18: 270-279.
14. Villa N, Do A, Hershey JW, Fraser CS (2013) Human eukaryotic initiation factor 4G (eIF4G) protein binds to eIF3c, -d, and -e to promote mRNA recruitment to the ribosome. *J Biol Chem* 288: 32932-32940.
15. Chiu WL, Wagner S, Herrmannova A, Burela L, Zhang F, et al. (2010) The C-terminal region of eukaryotic translation initiation factor 3a (eIF3a) promotes mRNA recruitment, scanning, and, together with eIF3j and the eIF3b RNA recognition motif, selection of AUG start codons. *Mol Cell Biol* 30: 4415-4434.
16. Mitchell SF, Walker SE, Algire MA, Park EH, Hinnebusch AG, et al. (2010) The 5'-7-methylguanosine cap on eukaryotic mRNAs serves both to stimulate canonical translation initiation and to block an alternative pathway. *Mol Cell* 39: 950-962.
17. Passmore LA, Schmeing TM, Maag D, Applefield DJ, Acker MG, et al. (2007) The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Mol Cell* 26: 41-50.
18. Pestova TV, Kolupaeva VG (2002) The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev* 16: 2906-2922.
19. Lomakin IB, Kolupaeva VG, Marintchev A, Wagner G, Pestova TV (2003) Position of eukaryotic initiation factor eIF1 on the 40S ribosomal subunit determined by directed hydroxyl radical probing. *Genes Dev* 17: 2786-2797.
20. Rabl J, Leibundgut M, Ataide SF, Haag A, Ban N (2011) Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science* 331: 730-736.
21. Battiste JL, Pestova TV, Hellen CU, Wagner G (2000) The eIF1A solution structure reveals a large RNA-binding surface important for scanning function. *Mol Cell* 5: 109-119.
22. Yu Y, Marintchev A, Kolupaeva VG, Unbehaun A, Veryasova T, et al. (2009) Position of eukaryotic translation initiation factor eIF1A on the 40S ribosomal subunit mapped by directed hydroxyl radical probing. *Nucleic Acids Res* 37: 5167-5182.

23. Svitkin YV, Pause A, Haghighat A, Pyronnet S, Witherell G, et al. (2001) The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. *RNA* 7: 382-394.
24. Pisareva VP, Pisarev AV, Komar AA, Hellen CU, Pestova TV (2008) Translation initiation on mammalian mRNAs with structured 5'UTRs requires DExH-box protein DHX29. *Cell* 135: 1237-1250.
25. Chuang RY, Weaver PL, Liu Z, Chang TH (1997) Requirement of the DEAD-Box protein ded1p for messenger RNA translation. *Science* 275: 1468-1471.
26. de la Cruz J, Iost I, Kressler D, Linder P (1997) The p20 and Ded1 proteins have antagonistic roles in eIF4E-dependent translation in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* 94: 5201-5206.
27. Tarn WY, Chang TH (2009) The current understanding of Ded1p/DDX3 homologs from yeast to human. *RNA Biol* 6: 17-20.
28. Pestova TV, Borukhov SI, Hellen CU (1998) Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature* 394: 854-859.
29. Pisarev AV, Kolupaeva VG, Pisareva VP, Merrick WC, Hellen CU, et al. (2006) Specific functional interactions of nucleotides at key -3 and +4 positions flanking the initiation codon with components of the mammalian 48S translation initiation complex. *Genes Dev* 20: 624-636.
30. Lomakin IB, Steitz TA (2013) The initiation of mammalian protein synthesis and mRNA scanning mechanism. *Nature* 500: 307-311.
31. Jackson RJ, Hellen CU, Pestova TV (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* 11: 113-127.
32. Maag D, Fekete CA, Gryczynski Z, Lorsch JR (2005) A conformational change in the eukaryotic translation preinitiation complex and release of eIF1 signal recognition of the start codon. *Mol Cell* 17: 265-275.
33. Kolitz SE, Takacs JE, Lorsch JR (2009) Kinetic and thermodynamic analysis of the role of start codon/anticodon base pairing during eukaryotic translation initiation. *RNA* 15: 138-152.
34. Saini AK, Nanda JS, Lorsch JR, Hinnebusch AG (2010) Regulatory elements in eIF1A control the fidelity of start codon selection by modulating tRNA(i)(Met) binding to the ribosome. *Genes Dev* 24: 97-110.
35. Paulin FE, Campbell LE, O'Brien K, Loughlin J, Proud CG (2001) Eukaryotic translation initiation factor 5 (eIF5) acts as a classical GTPase-activator protein. *Curr*

- Biol 11: 55-59.
36. Kapp LD, Lorsch JR (2004) GTP-dependent recognition of the methionine moiety on initiator tRNA by translation factor eIF2. *J Mol Biol* 335: 923-936.
  37. Price N, Proud C (1994) The guanine nucleotide-exchange factor, eIF-2B. *Biochimie* 76: 748-760.
  38. Pestova TV, Lomakin IB, Lee JH, Choi SK, Dever TE, et al. (2000) The joining of ribosomal subunits in eukaryotes requires eIF5B. *Nature* 403: 332-335.
  39. Unbehauen A, Borukhov SI, Hellen CU, Pestova TV (2004) Release of initiation factors from 48S complexes during ribosomal subunit joining and the link between establishment of codon-anticodon base-pairing and hydrolysis of eIF2-bound GTP. *Genes Dev* 18: 3078-3093.
  40. Fernandez IS, Bai XC, Hussain T, Kelley AC, Lorsch JR, et al. (2013) Molecular architecture of a eukaryotic translational initiation complex. *Science* 342: 1240-1245.
  41. Kuhle B, Ficner R (2014) eIF5B employs a novel domain release mechanism to catalyze ribosomal subunit joining. *EMBO J* 33: 1177-1191.
  42. Hinnebusch AG (2006) eIF3: a versatile scaffold for translation initiation complexes. *Trends Biochem Sci* 31: 553-562.
  43. Phan L, Zhang X, Asano K, Anderson J, Vornlocher HP, et al. (1998) Identification of a translation initiation factor 3 (eIF3) core complex, conserved in yeast and mammals, that interacts with eIF5. *Mol Cell Biol* 18: 4935-4946.
  44. Asano K, Phan L, Anderson J, Hinnebusch AG (1998) Complex formation by all five homologues of mammalian translation initiation factor 3 subunits from yeast *Saccharomyces cerevisiae*. *J Biol Chem* 273: 18573-18585.
  45. Valasek L, Phan L, Schoenfeld LW, Valaskova V, Hinnebusch AG (2001) Related eIF3 subunits TIF32 and HCR1 interact with an RNA recognition motif in PRT1 required for eIF3 integrity and ribosome binding. *Embo J* 20: 891-904.
  46. Marintchev A, Wagner G (2004) Translation initiation: structures, mechanisms and evolution. *Q Rev Biophys* 37: 197-284.
  47. ElAntak L, Tzakos AG, Locker N, Lukavsky PJ (2007) Structure of eIF3b RNA recognition motif and its interaction with eIF3j: structural insights into the recruitment of eIF3b to the 40 S ribosomal subunit. *J Biol Chem* 282: 8165-8174.
  48. Herrmannova A, Daujotyte D, Yang JC, Cuchalova L, Gorrec F, et al. (2012) Structural analysis of an eIF3 subcomplex reveals conserved interactions required for a stable and

- proper translation pre-initiation complex assembly. *Nucleic Acids Res* 40: 2294-2311.
49. Wang Z, Chen J, Sun J, Cui Z, Wu H (2012) RNA interference-mediated silencing of eukaryotic translation initiation factor 3, subunit B (EIF3B) gene expression inhibits proliferation of colon cancer cells. *World J Surg Oncol* 10: 119.
  50. Liang H, Ding X, Zhou C, Zhang Y, Xu M, et al. (2012) Knockdown of eukaryotic translation initiation factors 3B (EIF3B) inhibits proliferation and promotes apoptosis in glioblastoma cells. *Neurol Sci* 33: 1057-1062.
  51. Wang H, Ru Y, Sanchez-Carbayo M, Wang X, Kieft JS, et al. (2013) Translation initiation factor eIF3b expression in human cancer and its role in tumor growth and lung colonization. *Clin Cancer Res* 19: 2850-2860.
  52. Khoshnevis S, Neumann P, Ficner R (2010) Crystal structure of the RNA recognition motif of yeast translation initiation factor eIF3b reveals differences to human eIF3b. *PLoS One* 5.
  53. Khoshnevis S, Gunisova S, Vlckova V, Kouba T, Neumann P, et al. (2014) Structural integrity of the PCI domain of eIF3a/TIF32 is required for mRNA recruitment to the 43S pre-initiation complexes. *Nucleic Acids Res* 42: 4123-4139.
  54. Pick E, Hofmann K, Glickman MH (2009) PCI complexes: Beyond the proteasome, CSN, and eIF3 Troika. *Mol Cell* 35: 260-264.
  55. Kouba T, Danyi I, Gunisova S, Munzarova V, Vlckova V, et al. (2012) Small ribosomal protein RPS0 stimulates translation initiation by mediating 40S-binding of eIF3 via its direct contact with the eIF3a/TIF32 subunit. *PLoS One* 7: e40464.
  56. Kouba T, Rutkai E, Karaskova M, Valasek L (2012) The eIF3c/NIP1 PCI domain interacts with RNA and RACK1/ASC1 and promotes assembly of translation preinitiation complexes. *Nucleic Acids Res* 40: 2683-2699.
  57. Asano K, Clayton J, Shalev A, Hinnebusch AG (2000) A multifactor complex of eukaryotic initiation factors, eIF1, eIF2, eIF3, eIF5, and initiator tRNA(Met) is an important translation initiation intermediate in vivo. *Genes Dev* 14: 2534-2546.
  58. Cuchalova L, Kouba T, Herrmannova A, Danyi I, Chiu WL, et al. (2010) The RNA recognition motif of eukaryotic translation initiation factor 3g (eIF3g) is required for resumption of scanning of posttermination ribosomes for reinitiation on GCN4 and together with eIF3i stimulates linear scanning. *Mol Cell Biol* 30: 4671-4686.
  59. Khoshnevis S, Hauer F, Milon P, Stark H, Ficner R (2012) Novel insights into the architecture and protein interaction network of yeast eIF3. *RNA* 18: 2306-2319.
  60. Asano K, Krishnamoorthy T, Phan L, Pavitt GD, Hinnebusch AG (1999) Conserved

## References

- bipartite motifs in yeast eIF5 and eIF2Bepsilon, GTPase-activating and GDP-GTP exchange factors in translation initiation, mediate binding to their common substrate eIF2. *EMBO J* 18: 1673-1688.
61. Hashem Y, des Georges A, Dhote V, Langlois R, Liao HY, et al. (2013) Structure of the Mammalian Ribosomal 43S Preinitiation Complex Bound to the Scanning Factor DHX29. *Cell* 153: 1108-1119.
62. Pisarev AV, Kolupaeva VG, Yusupov MM, Hellen CU, Pestova TV (2008) Ribosomal position and contacts of mRNA in eukaryotic translation initiation complexes. *EMBO J* 27: 1609-1621.
63. Siridechadilok B, Fraser CS, Hall RJ, Doudna JA, Nogales E (2005) Structural roles for human translation factor eIF3 in initiation of protein synthesis. *Science* 310: 1513-1515.
64. Phan L, Schoenfeld LW, Valasek L, Nielsen KH, Hinnebusch AG (2001) A subcomplex of three eIF3 subunits binds eIF1 and eIF5 and stimulates ribosome binding of mRNA and tRNA(i)Met. *EMBO J* 20: 2954-2965.
65. Jivotovskaya AV, Valasek L, Hinnebusch AG, Nielsen KH (2006) Eukaryotic translation initiation factor 3 (eIF3) and eIF2 can promote mRNA binding to 40S subunits independently of eIF4G in yeast. *Mol Cell Biol* 26: 1355-1372.
66. Korneeva NL, Lamphear BJ, Hennigan FL, Rhoads RE (2000) Mutually cooperative binding of eukaryotic translation initiation factor (eIF) 3 and eIF4A to human eIF4G-1. *J Biol Chem* 275: 41369-41376.
67. Methot N, Pickett G, Keene JD, Sonenberg N (1996) In vitro RNA selection identifies RNA ligands that specifically bind to eukaryotic translation initiation factor 4B: the role of the RNA motif. *RNA* 2: 38-50.
68. Vornlocher HP, Hanachi P, Ribeiro S, Hershey JW (1999) A 110-kilodalton subunit of translation initiation factor eIF3 and an associated 135-kilodalton protein are encoded by the *Saccharomyces cerevisiae* TIF32 and TIF31 genes. *J Biol Chem* 274: 16802-16812.
69. Valasek L, Nielsen KH, Zhang F, Fekete CA, Hinnebusch AG (2004) Interactions of eukaryotic translation initiation factor 3 (eIF3) subunit NIP1/c with eIF1 and eIF5 promote preinitiation complex assembly and regulate start codon selection. *Mol Cell Biol* 24: 9437-9455.
70. Sonenberg N, Morgan MA, Merrick WC, Shatkin AJ (1978) A polypeptide in eukaryotic initiation factors that crosslinks specifically to the 5'-terminal cap in mRNA. *Proc Natl Acad Sci U S A* 75: 4843-4847.



71. Sonenberg N, Rupprecht KM, Hecht SM, Shatkin AJ (1979) Eukaryotic mRNA cap binding protein: purification by affinity chromatography on sepharose-coupled m7GDP. *Proc Natl Acad Sci U S A* 76: 4345-4349.
72. Svitkin YV, Ovchinnikov LP, Dreyfuss G, Sonenberg N (1996) General RNA binding proteins render translation cap dependent. *EMBO J* 15: 7147-7155.
73. Altmann M, Muller PP, Pelletier J, Sonenberg N, Trachsel H (1989) A mammalian translation initiation factor can substitute for its yeast homologue in vivo. *J Biol Chem* 264: 12145-12147.
74. Marcotrigiano J, Gingras AC, Sonenberg N, Burley SK (1997) Cocystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP. *Cell* 89: 951-961.
75. Gross JD, Moerke NJ, von der Haar T, Lugovskoy AA, Sachs AB, et al. (2003) Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. *Cell* 115: 739-750.
76. Ray BK, Lawson TG, Kramer JC, Cladaras MH, Grifo JA, et al. (1985) ATP-dependent unwinding of messenger RNA structure by eukaryotic initiation factors. *J Biol Chem* 260: 7651-7658.
77. Lorsch JR, Herschlag D (1998) The DEAD box protein eIF4A. 1. A minimal kinetic and thermodynamic framework reveals coupled binding of RNA and nucleotide. *Biochemistry* 37: 2180-2193.
78. Linder P, Lasko PF, Ashburner M, Leroy P, Nielsen PJ, et al. (1989) Birth of the D-E-A-D box. *Nature* 337: 121-122.
79. Grifo JA, Abramson RD, Satler CA, Merrick WC (1984) RNA-stimulated ATPase activity of eukaryotic initiation factors. *J Biol Chem* 259: 8648-8654.
80. Linder P, Jankowsky E (2011) From unwinding to clamping - the DEAD box RNA helicase family. *Nat Rev Mol Cell Biol* 12: 505-516.
81. Hardin JW, Hu YX, McKay DB (2010) Structure of the RNA binding domain of a DEAD-box helicase bound to its ribosomal RNA target reveals a novel mode of recognition by an RNA recognition motif. *J Mol Biol* 402: 412-427.
82. Klostermeier D, Rudolph MG (2009) A novel dimerization motif in the C-terminal domain of the *Thermus thermophilus* DEAD box helicase Hera confers substantial flexibility. *Nucleic Acids Res* 37: 421-430.
83. Bono F, Ebert J, Lorentzen E, Conti E (2006) The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell* 126: 713-725.

## References

84. Andersen CB, Ballut L, Johansen JS, Chamieh H, Nielsen KH, et al. (2006) Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science* 313: 1968-1972.
85. Caruthers JM, Johnson ER, McKay DB (2000) Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. *Proc Natl Acad Sci U S A* 97: 13080-13085.
86. Schutz P, Bumann M, Oberholzer AE, Bieniossek C, Trachsel H, et al. (2008) Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions. *Proc Natl Acad Sci U S A* 105: 9564-9569.
87. Altmann M, Muller PP, Wittmer B, Ruchti F, Lanker S, et al. (1993) A *Saccharomyces cerevisiae* homologue of mammalian translation initiation factor 4B contributes to RNA helicase activity. *EMBO J* 12: 3997-4003.
88. Ozes AR, Feoktistova K, Avanzino BC, Fraser CS (2011) Duplex unwinding and ATPase activities of the DEAD-box helicase eIF4A are coupled by eIF4G and eIF4B. *J Mol Biol* 412: 674-687.
89. Pestova TV, Hellen CU, Shatsky IN (1996) Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry. *Mol Cell Biol* 16: 6859-6869.
90. Coppolecchia R, Buser P, Stotz A, Linder P (1993) A new yeast translation initiation factor suppresses a mutation in the eIF-4A RNA helicase. *EMBO J* 12: 4005-4011.
91. Methot N, Song MS, Sonenberg N (1996) A region rich in aspartic acid, arginine, tyrosine, and glycine (DRYG) mediates eukaryotic initiation factor 4B (eIF4B) self-association and interaction with eIF3. *Mol Cell Biol* 16: 5328-5334.
92. Methot N, Rom E, Olsen H, Sonenberg N (1997) The human homologue of the yeast Prt1 protein is an integral part of the eukaryotic initiation factor 3 complex and interacts with p170. *J Biol Chem* 272: 1110-1116.
93. Methot N, Pause A, Hershey JW, Sonenberg N (1994) The translation initiation factor eIF-4B contains an RNA-binding region that is distinct and independent from its ribonucleoprotein consensus sequence. *Mol Cell Biol* 14: 2307-2316.
94. Naranda T, Strong WB, Menaya J, Fabbri BJ, Hershey JW (1994) Two structural domains of initiation factor eIF-4B are involved in binding to RNA. *J Biol Chem* 269: 14465-14472.
95. Fleming K, Ghuman J, Yuan X, Simpson P, Szendroi A, et al. (2003) Solution structure and RNA interactions of the RNA recognition motif from eukaryotic translation initiation factor 4B. *Biochemistry* 42: 8966-8975.

96. Richter-Cook NJ, Dever TE, Hensold JO, Merrick WC (1998) Purification and characterization of a new eukaryotic protein translation factor. Eukaryotic initiation factor 4H. *J Biol Chem* 273: 7579-7587.
97. Imataka H, Gradi A, Sonenberg N (1998) A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. *EMBO J* 17: 7480-7489.
98. Safaee N, Kozlov G, Noronha AM, Xie J, Wilds CJ, et al. (2012) Interdomain allostery promotes assembly of the poly(A) mRNA complex with PABP and eIF4G. *Mol Cell* 48: 375-386.
99. Marintchev A, Edmonds KA, Marintcheva B, Hendrickson E, Oberer M, et al. (2009) Topology and regulation of the human eIF4A/4G/4H helicase complex in translation initiation. *Cell* 136: 447-460.
100. Morino S, Imataka H, Svitkin YV, Pestova TV, Sonenberg N (2000) Eukaryotic translation initiation factor 4E (eIF4E) binding site and the middle one-third of eIF4GI constitute the core domain for cap-dependent translation, and the C-terminal one-third functions as a modulatory region. *Mol Cell Biol* 20: 468-477.
101. Goyer C, Altmann M, Lee HS, Blanc A, Deshmukh M, et al. (1993) TIF4631 and TIF4632: two yeast genes encoding the high-molecular-weight subunits of the cap-binding protein complex (eukaryotic initiation factor 4F) contain an RNA recognition motif-like sequence and carry out an essential function. *Mol Cell Biol* 13: 4860-4874.
102. Tarun SZ, Jr., Wells SE, Deardorff JA, Sachs AB (1997) Translation initiation factor eIF4G mediates in vitro poly(A) tail-dependent translation. *Proc Natl Acad Sci U S A* 94: 9046-9051.
103. Marcotrigiano J, Gingras AC, Sonenberg N, Burley SK (1999) Cap-dependent translation initiation in eukaryotes is regulated by a molecular mimic of eIF4G. *Mol Cell* 3: 707-716.
104. Ptushkina M, von der Haar T, Vasilescu S, Frank R, Birkenhager R, et al. (1998) Cooperative modulation by eIF4G of eIF4E-binding to the mRNA 5' cap in yeast involves a site partially shared by p20. *EMBO J* 17: 4798-4808.
105. Etchison D, Milburn S (1987) Separation of protein synthesis initiation factor eIF4A from a p220-associated cap binding complex activity. *Mol Cell Biochem* 76: 15-25.
106. Goyer C, Altmann M, Trachsel H, Sonenberg N (1989) Identification and characterization of cap-binding proteins from yeast. *J Biol Chem* 264: 7603-7610.

## References

107. Zapata JM, Martinez MA, Sierra JM (1994) Purification and characterization of eukaryotic polypeptide chain initiation factor 4F from *Drosophila melanogaster* embryos. *J Biol Chem* 269: 18047-18052.
108. Hilbert M, Kebbel F, Gubaev A, Klostermeier D (2011) eIF4G stimulates the activity of the DEAD box protein eIF4A by a conformational guidance mechanism. *Nucleic Acids Res* 39: 2260-2270.
109. Pestova TV, Hellen CU (2003) Coupled folding during translation initiation. *Cell* 115: 650-652.
110. Rajagopal V, Park EH, Hinnebusch AG, Lorsch JR (2012) Specific domains in yeast translation initiation factor eIF4G strongly bias RNA unwinding activity of the eIF4F complex toward duplexes with 5'-overhangs. *J Biol Chem* 287: 20301-20312.
111. Bonneau AM, Sonenberg N (1987) Involvement of the 24-kDa cap-binding protein in regulation of protein synthesis in mitosis. *J Biol Chem* 262: 11134-11139.
112. Waskiewicz AJ, Johnson JC, Penn B, Mahalingam M, Kimball SR, et al. (1999) Phosphorylation of the cap-binding protein eukaryotic translation initiation factor 4E by protein kinase Mnk1 in vivo. *Mol Cell Biol* 19: 1871-1880.
113. Minich WB, Balasta ML, Goss DJ, Rhoads RE (1994) Chromatographic resolution of in vivo phosphorylated and nonphosphorylated eukaryotic translation initiation factor eIF-4E: increased cap affinity of the phosphorylated form. *Proc Natl Acad Sci U S A* 91: 7668-7672.
114. Poulin F, Gingras AC, Olsen H, Chevalier S, Sonenberg N (1998) 4E-BP3, a new member of the eukaryotic initiation factor 4E-binding protein family. *J Biol Chem* 273: 14002-14007.
115. Yang HS, Jansen AP, Komar AA, Zheng X, Merrick WC, et al. (2003) The transformation suppressor Pdc4 is a novel eukaryotic translation initiation factor 4A binding protein that inhibits translation. *Mol Cell Biol* 23: 26-37.
116. Chang JH, Cho YH, Sohn SY, Choi JM, Kim A, et al. (2009) Crystal structure of the eIF4A-PDCD4 complex. *Proc Natl Acad Sci U S A* 106: 3148-3153.
117. Loh PG, Yang HS, Walsh MA, Wang Q, Wang X, et al. (2009) Structural basis for translational inhibition by the tumour suppressor Pdc4. *EMBO J* 28: 274-285.
118. Lindqvist L, Oberer M, Reibarkh M, Cencic R, Bordeleau ME, et al. (2008) Selective pharmacological targeting of a DEAD box RNA helicase. *PLoS One* 3: e1583.
119. Low WK, Dang Y, Schneider-Poetsch T, Shi Z, Choi NS, et al. (2005) Inhibition of eukaryotic translation initiation by the marine natural product pateamine A. *Mol Cell*

- 20: 709-722.
120. Bordeleau ME, Cencic R, Lindqvist L, Oberer M, Northcote P, et al. (2006) RNA-mediated sequestration of the RNA helicase eIF4A by Patamine A inhibits translation initiation. *Chem Biol* 13: 1287-1295.
  121. Lin D, Pestova TV, Hellen CU, Tiedge H (2008) Translational control by a small RNA: dendritic BC1 RNA targets the eukaryotic initiation factor 4A helicase mechanism. *Mol Cell Biol* 28: 3008-3019.
  122. Morley SJ, Pain VM (1995) Hormone-induced meiotic maturation in *Xenopus* oocytes occurs independently of p70s6k activation and is associated with enhanced initiation factor (eIF)-4F phosphorylation and complex formation. *J Cell Sci* 108 ( Pt 4): 1751-1760.
  123. Morley SJ, Pain VM (1995) Translational regulation during activation of porcine peripheral blood lymphocytes: association and phosphorylation of the alpha and gamma subunits of the initiation factor complex eIF-4F. *Biochem J* 312 ( Pt 2): 627-635.
  124. Morley SJ, Traugh JA (1993) Stimulation of translation in 3T3-L1 cells in response to insulin and phorbol ester is directly correlated with increased phosphate labelling of initiation factor (eIF-) 4F and ribosomal protein S6. *Biochimie* 75: 985-989.
  125. Morley SJ, Traugh JA (1990) Differential stimulation of phosphorylation of initiation factors eIF-4F, eIF-4B, eIF-3, and ribosomal protein S6 by insulin and phorbol esters. *J Biol Chem* 265: 10611-10616.
  126. Bu X, Haas DW, Hagedorn CH (1993) Novel phosphorylation sites of eukaryotic initiation factor-4F and evidence that phosphorylation stabilizes interactions of the p25 and p220 subunits. *J Biol Chem* 268: 4975-4978.
  127. Henis-Korenblit S, Shani G, Sines T, Marash L, Shohat G, et al. (2002) The caspase-cleaved DAP5 protein supports internal ribosome entry site-mediated translation of death proteins. *Proc Natl Acad Sci U S A* 99: 5400-5405.
  128. Lewis SM, Cerquozzi S, Graber TE, Ungureanu NH, Andrews M, et al. (2008) The eIF4G homolog DAP5/p97 supports the translation of select mRNAs during endoplasmic reticulum stress. *Nucleic Acids Res* 36: 168-178.
  129. Nevins TA, Harder ZM, Korneluk RG, Holcik M (2003) Distinct regulation of internal ribosome entry site-mediated translation following cellular stress is mediated by apoptotic fragments of eIF4G translation initiation factor family members eIF4GI and p97/DAP5/NAT1. *J Biol Chem* 278: 3572-3579.
  130. Bushell M, Poncet D, Marissen WE, Flotow H, Lloyd RE, et al. (2000) Cleavage of

## References

- polypeptide chain initiation factor eIF4GI during apoptosis in lymphoma cells: characterisation of an internal fragment generated by caspase-3-mediated cleavage. *Cell Death Differ* 7: 628-636.
131. Wei Z, Zhang P, Zhou Z, Cheng Z, Wan M, et al. (2004) Crystal structure of human eIF3k, the first structure of eIF3 subunits. *J Biol Chem* 279: 34983-34990.
132. Sambrook J, Fritsch, E. F., and Maniatis, T. (1989) *Molecular cloning: a laboratory manual*.
133. Cline J, Braman JC, Hogrefe HH (1996) PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res* 24: 3546-3551.
134. Norgard MV, Keem K, Monahan JJ (1978) Factors affecting the transformation of *Escherichia coli* strain chi1776 by pBR322 plasmid DNA. *Gene* 3: 279-292.
135. Studier FW (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41: 207-234.
136. Forne I, Ludwigsen J, Imhof A, Becker PB, Mueller-Planitz F (2012) Probing the conformation of the ISWI ATPase domain with genetically encoded photoreactive crosslinkers and mass spectrometry. *Mol Cell Proteomics* 11: M111 012088.
137. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72: 248-254.
138. Liu Y, Neumann P, Kuhle B, Monecke T, Schell S, et al. (2014) Translation Initiation Factor eIF3b Contains a Nine-Bladed beta-Propeller and Interacts with the 40S Ribosomal Subunit. *Structure*.
139. Tanaka Y, Bond MR, Kohler JJ (2008) Photocrosslinkers illuminate interactions in living cells. *Mol Biosyst* 4: 473-480.
140. Webb MR (1992) A continuous spectrophotometric assay for inorganic phosphate and for measuring phosphate release kinetics in biological systems. *Proc Natl Acad Sci U S A* 89: 4884-4887.
141. Belon CA, Frick DN (2008) Monitoring helicase activity with molecular beacons. *Biotechniques* 45: 433-440, 442.
142. Newman J, Egan D, Walter TS, Meged R, Berry I, et al. (2005) Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy. *Acta Crystallogr D Biol Crystallogr* 61: 1426-1431.
143. Radaev S, Li S, Sun PD (2006) A survey of protein-protein complex crystallizations.

- Acta Crystallogr D Biol Crystallogr 62: 605-612.
144. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* 66: 125-132.
145. Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr A* 64: 112-122.
146. Vonrhein C, Blanc E, Roversi P, Bricogne G (2007) Automated structure solution with autoSHARP. *Methods Mol Biol* 364: 215-230.
147. Abrahams JP, Leslie AG (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr D Biol Crystallogr* 52: 30-42.
148. Langer G, Cohen SX, Lamzin VS, Perrakis A (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* 3: 1171-1179.
149. Perrakis A, Morris R, Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 6: 458-463.
150. Morris RJ, Zwart PH, Cohen S, Fernandez FJ, Kakaris M, et al. (2004) Breaking good resolutions with ARP/wARP. *J Synchrotron Radiat* 11: 56-59.
151. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66: 486-501.
152. Brunger AT (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355: 472-475.
153. Brunger AT (1993) Assessment of phase accuracy by cross validation: the free R value. Methods and applications. *Acta Crystallogr D Biol Crystallogr* 49: 24-36.
154. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66: 213-221.
155. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, et al. (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res* 35: W522-525.
156. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* 98: 10037-10041.
157. Nielsen KH, Valasek L, Sykes C, Jivotovskaya A, Hinnebusch AG (2006) Interaction of the RNP1 motif in PRT1 with HCR1 promotes 40S binding of eukaryotic initiation factor 3 in yeast. *Mol Cell Biol* 26: 2984-2998.

## References

158. Buchan DW, Minneci F, Nugent TC, Bryson K, Jones DT (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res* 41: W349-357.
159. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, et al. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31: 3497-3500.
160. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605-1612.
161. Fraser CS, Berry KE, Hershey JW, Doudna JA (2007) eIF3j is located in the decoding center of the human 40S ribosomal subunit. *Mol Cell* 26: 811-819.
162. Elantak L, Wagner S, Herrmannova A, Karaskova M, Rutkai E, et al. (2010) The indispensable N-terminal half of eIF3j/HCR1 cooperates with its structurally conserved binding partner eIF3b/PRT1-RRM and with eIF1A in stringent AUG selection. *J Mol Biol* 396: 1097-1116.
163. Christian H, Hofele RV, Urlaub H, Ficner R (2014) Insights into the activation of the helicase Prp43 by biochemical studies and structural mass spectrometry. *Nucleic Acids Res* 42: 1162-1179.
164. Valasek LS (2012) 'Ribozoomin'--translation initiation from the perspective of the ribosome-bound eukaryotic initiation factors (eIFs). *Curr Protein Pept Sci* 13: 305-330.
165. Amlacher S, Sarges P, Flemming D, van Noort V, Kunze R, et al. (2011) Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile. *Cell* 146: 277-289.
166. Monecke T, Haselbach D, Voss B, Russek A, Neumann P, et al. (2013) Structural basis for cooperativity of CRM1 export complex formation. *Proc Natl Acad Sci U S A* 110: 960-965.
167. Leidig C, Bange G, Kopp J, Amlacher S, Aravind A, et al. (2013) Structural characterization of a eukaryotic chaperone--the ribosome-associated complex. *Nat Struct Mol Biol* 20: 23-28.
168. Sun C, Todorovic A, Querol-Audi J, Bai Y, Villa N, et al. (2011) Functional reconstitution of human eukaryotic translation initiation factor 3 (eIF3). *Proc Natl Acad Sci U S A* 108: 20473-20478.
169. Chen CK, Chan NL, Wang AH (2011) The many blades of the beta-propeller proteins: conserved but versatile. *Trends Biochem Sci* 36: 553-561.
170. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, et al. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids*



Res 39: D225-229.

171. Kashiwagi K, Ito T, Yokoyama S (2014) Crystal structure of the eukaryotic translation initiation factor 2A from *Schizosaccharomyces pombe*. *J Struct Funct Genomics*.
172. Cordin O, Banroques J, Tanner NK, Linder P (2006) The DEAD-box protein family of RNA helicases. *Gene* 367: 17-37.
173. Andreou AZ, Klostermeier D (2013) The DEAD-box helicase eIF4A: paradigm or the odd one out? *RNA Biol* 10: 19-32.
174. Andreou AZ, Klostermeier D (2014) eIF4B and eIF4G jointly stimulate eIF4A ATPase and unwinding activities by modulation of the eIF4A conformational cycle. *J Mol Biol* 426: 51-61.
175. Akabayov SR, Akabayov B, Richardson CC, Wagner G (2013) Molecular crowding enhanced ATPase activity of the RNA helicase eIF4A correlates with compaction of its quaternary structure and association with eIF4G. *J Am Chem Soc* 135: 10040-10047.

## Abbreviations

Abbreviation	Expanded
Å	Angstrom ( $1 \text{ Å} = 10^{-10} \text{ m}$ )
APS	ammoniumpersulfate
ATP	adenosine triphosphate
ADP	adenosine diphosphate
BisTris	Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane
Bpa	<i>p</i> -benzoyl-L-phenylalanine
CTD	C-terminal domain
cryo-EM	Cryo-electron microscopy
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dsRNA	double-stranded RNA
DTT	1,4-Dithiothreitol
EDTA	ethylenediaminetetraacetic acid
<i>e.g.</i>	<i>exempli gratia</i>
eIF	eukaryotic translation initiation factor
EMSA	electrophoretic mobility shift assay
GST	Glutathione <i>S</i> -transferases
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
<i>i.e.</i>	<i>id est</i>
IRES	Internal ribosome entry site
IPTG	isopropyl-β-D-1-thiogalactopyranoside
kDa	kilodalton (unit of molecular weight)
<i>k</i> <sub>cat</sub>	catalytic constant
<i>K</i> <sub>M</sub>	Michaelis constant

Abbreviation	Expanded
m <sup>7</sup> G	7-methylguanosine
MAD	multi-wavelength anomalous dispersion
MR	molecular replacement
mRNA	messenger RNA
NMR	nuclear magnetic resonance
NTD	N-terminal domain
PCR	polymerase chain reaction
PIC	preinitiation complex
poly(A)	poly-adenine
r.m.s.d.	root mean square deviation
RNA	ribonucleic acid
SEC	size exclusion chromatography
SDS	sodium dodecylsulfate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
TEMED	N,N,N',N'-tetraethylenediamide
TRIS	tris-(hydroxymethyl)-aminomethane
tRNA	transfer RNA
UTR	untranslated region
UV	ultra-violet
β-ME	2-Mercaptoethanol



# Acknowledgements

I am truly grateful to my supervisor Prof. Dr. Ralf Ficner. I sincerely appreciate all his support, encouragement and great guidance during my 3.5-year Ph.D study. I would like to acknowledge Prof. Dr. Wolfgang Wintermeyer and Prof. Dr. Holger Stark for the fruitful discussions and helpful suggestions during the thesis committee meetings. I am truly thankful to Prof. Dr. Reinhard Lührmann, Prof. Dr. Jörg Stülke and Prof. Dr. Kai Tittmann for being members of my extended thesis committee.

In addition, I appreciate that Prof. Dr. Holger Stark offered me an opportunity to work in his lab for an important experiment. I would also like to thank all people from his group for their kindness and friendship. I am very thankful to Dr. Stephanie Schell and Dr. Ashwin Chari for the kind help and brilliant advises during the period that I was working with them.

I express my sincere thanks to all former and current members of the Department of Molecular Structural Biology for their friendship and great working atmosphere. I am special grateful to Bernhard Kuhle for extensive discussions, brilliant suggestions and great company. I appreciate Dr. Piotr Neumanm for his help on diffraction data collection and structure determination.

I want to thank Bernhard Kuhle, Dr. Yasar Luqman Ahmed, Dr. Thomas Monecke and Dr. Po-chia Chen for critically reading and correcting parts or whole of this thesis.

I am forever indebted to my parents and Fengbiao for their unconditional support and endless love. I would also like to express my deepest gratitude to all my friends for their great company, encouragement and love.



# CURRICULUM VITAE OF YI LIU

## PERSONAL INFORMATION

---

- ♦ Given Name, Surname: Yi, Liu
- ♦ Gender: Female
- ♦ Nationality: Chinese
- ♦ Current Position: Ph.D student in Department of Molecular Structural Biology, Goettingen University, Germany
- ♦ E-mail: yi.liuzhou@gmail.com



## UNIVERSITY EDUCATION

---

- |                   |   |
|-------------------|---|
| 2010.12 – Now     | Ph.D student in Department of Molecular Structural Biology, Goettingen University<br><br>Supervisor: Prof. Dr. Ralf Ficner<br><br>Ph.D Thesis Project: Structural and functional studies on proteins involved in eukaryotic translation initiation                          |
| 2007.09 – 2010.07 | Master student in Biochemistry and Molecular Biology, Institute of Biomedical Science (IBS), Fudan University, Shanghai, China<br><br>Supervisor: Prof. Dr. Yanhui Xu<br><br>Master Thesis Project : Structural and functional studies of a complex in <i>mTOMC pathway</i> |
| 2003.09 – 2007.07 | BSc ( <i>first class</i> ) in zoology, Department of Food Science, Jiangnan University, Wu Xi, China  |