

Structural Characterization of the Eukaryotic
Translation Initiation by Electron
Cryo-Microscopy

DISSERTATION

for the award of the degree

“Doctor rerum naturalium” (Dr. rer. nat.)
of the GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

within the doctoral program

BIOMOLECULES: STRUCTURE – FUNCTION – DYNAMICS
of the Göttingen Graduate School for Neurosciences,
Biophysics, and Molecular Biosciences (GGNB)

submitted by

Jan Erik Schliep

from Göttingen, Germany

Göttingen 2018

Thesis Committee

Prof. Dr. Holger Stark
Structural Dynamics, Max-Planck-Institute for Biophysical Chemistry

Prof. Dr. Marina Rodnina
Physical Biochemistry, Max-Planck-Institute for Biophysical Chemistry

Prof. Dr. Ralf Ficner
Molecular Structural Biology, Institute for Microbiology & Genetics

Members of the Examination Board

1st Referee: Prof. Dr. Holger Stark
Structural Dynamics, Max-Planck-Institute for Biophysical Chemistry

2nd Referee: Prof. Dr. Marina Rodnina
Physical Biochemistry, Max-Planck-Institute for Biophysical Chemistry

Prof. Dr. Ralf Ficner
Molecular Structural Biology, Institute for Microbiology and Genetics

Prof. Dr. Henning Urlaub
Bioanalytical Mass Spectrometry, Max-Planck-Institute for Biophysical Chemistry

Prof. Dr. Kai Tittmann
Molecular Enzymology, Albrecht-von-Haller-Institute

Prof. Dr. Detlef Doenecke
Molecular Biology, University Medical Center

Date of oral examination: 14 August 2018

AFFIDAVIT

I hereby declare that this dissertation with the title “Structural Characterization of the Eukaryotic Translation Initiation by Electron Cryo-Microscopy” has been written independently and with no other aids or sources than quoted. This thesis (wholly or in part) has not been submitted elsewhere for any academic award or qualification.

Jan Erik Schliep

Abstract

Translation is a highly conserved molecular process during which the genetic information stored in an mRNA molecule is translated into a polypeptide chain that eventually folds into functional three-dimensional proteins. Translation initiation, the first of four steps, is extremely important for the fidelity of the translation process. In eukaryotic translation initiation, the mRNA is brought into contact with the smaller 40S ribosomal subunit and subsequently scanned for the AUG start codon. Once identified, the mRNA gets locked in the P-site of the 40S ribosomal subunit via a constriction of its decoding center and a codon-anticodon-binding between the initiator Met-tRNA_i and the start codon. The entire process is orchestrated and monitored by a large number of highly specialized factors that eventually trigger the joining of the larger 60S ribosomal subunit, followed by elongation of the polypeptide chain.

Translation initiation has been the subject of extensively biochemical studies for many years. However, high-resolution reconstructions of the macromolecular complexes involved have just been obtained during the past decade. Thus, structural knowledge for this highly dynamic process is currently limited to a small number of complexes, depicting only narrow snapshots of what is happening. Additionally, although translation initiation is a highly conserved process, the factors involved vary quite extensively among eukaryotes. As a result, a comprehensive understanding of the dynamics and intermediate steps occurring during translation initiation, especially in higher mammals such as humans, remains missing.

Here, two complexes involved in human translation initiation were structurally investigated: the free eukaryotic translation initiation factor 3 (eIF3) and the 48S initiation complex (48S-IC). eIF3, often described as a scaffold protein for all other initiation factors and the 40S ribosomal subunit, is the largest of the initiation factors and is composed of 13 subunits. The 48S-IC is a late-stage initiation complex, in which the mRNA has already bound to the 40S together with many other initiation factors, including eIF3, to facilitate start codon recognition. The complex marks the final step before joining of the 60S ribosomal subunit and the start of translation elongation.

For the free eIF3, a native purification procedure from HeLa cytoplasmic extract was established and the quality of the purified macromolecules has been validated via tandem mass spectrometry, ProteoPlex complex stability measurements, and protein crystallization. The purified eIF3 complexes were used to obtain a 3D reconstruction at 7–15 Å

resolution that exhibited a great degree of flexibility. Due to the fact that all 13 subunits of the complex have been identified by mass spectrometry and that the complex shows good structural integrity in other biochemical assays in particular ProteoPlex and protein crystallization it can be assumed that the flexibility is not a product of inappropriate sample preparation. Although it has not been possible thus far to stabilize the complexes enough to yield a high-resolution reconstruction for all of its parts, it was possible to apply focused 3D classification to obtain several 3D structures that each show significantly more isotropic density than any other published eIF3 structure before.

In order to study the interaction between the 40S ribosomal subunit and the eIF3 initiation factor, and additionally stabilize the composition of eIF3, the 48S-IC was investigated. Here, a large dataset was acquired which made it possible to extensively classify for structural heterogeneity and eventually yield a 3D reconstruction at 4.5–12 Å resolution. It was possible to identify most of the initiation complexes attached to the surface of the 40S ribosomal subunit and verify their positions via cross-linking mass spectrometry. Especially for eIF1, an initiation factor involved in start codon recognition fidelity, it became possible to follow its track along several positions in the vicinity of the decoding center.

In addition, the overall dynamics of the entire complex were studied using a three-dimensional principle component analysis-based approach. From these results, an energy landscape was calculated that revealed the tilting of the 40S head to be one of the major movements within in the complex with great impact on the stability. This movement occurs to hold the mRNA in the correct position once the start codon is identified. For the closed 40S head, significantly less structural flexibility was detected. Furthermore, movements in the surrounding of the 40S decoding center were also identified as one of the major sources of for structural heterogeneity. These results give a first insight into the important dynamics underlying the processes of eukaryotic translation and provide a promising foundation for future studies.

Acknowledgment

Over the past four years, during which I had the opportunity to work on this thesis and many other exciting scientific projects, I have been extremely fortunate to meet many greatly talented, enthusiastic, and inspiring people. With their help, I was able to expand my scientific knowledge and sharpen my analytical thinking through many fruitful discussions, critical questions, and exciting presentations. I am deeply grateful for this extraordinary experience. Although I may not be able to name everyone here, I would like to highlight some of these people and express my gratitude to them and everyone else.

First of all, I would like to thank Prof. Holger Stark for the opportunity to conduct the research for this thesis as a member of his department. He has not only been the supervisor and first referee for this thesis, but also a constant advisor towards all scientific aspects of this project and a great help in many difficult situations. His knowledge and enthusiasm about electron cryo-microscopy in particular and science in general have deeply impressed me and were one of the reasons why I joined this scientific field.

I would like to thank Prof. Marina Rodnina¹ and Prof. Ralf Ficner², both members of my thesis advisory committee (TAC), for their guidance and help, during our meetings and whenever necessary. Large parts of this work would not have been achievable without their constructive criticism and inspiring ideas. Additionally, I would like to express my gratitude to the people working at the *Göttingen Graduate School for Neurosciences, Biophysics, and Molecular Biosciences* (GGNB), who have enhanced the course of my doctoral studies with many exciting events and helped to make the bureaucratic aspects of it less painful.

A very important part of this thesis would not have been possible without the great work of my collaborators, namely Dr. Akanksha Goyal¹, Dr. Jasmin Corso³ and Sung-Hui Yi¹. They provided high quality biological samples or the results from carefully conducted experiments and were always open for my questions and shared their knowledge during countless discussions. For all this I'm profoundly thankful.

¹Department of Physical Biochemistry, Max-Planck-Institute for Biophysical Chemistry, Göttingen

²Department of Molecular Structural Biology, Georg-August-University of Göttingen

³Research Group Bioanalytical Mass Spectrometry, Max-Planck-Institute for Biophysical Chemistry

Another no less important part of this work was only achievable through the support of and collaboration with Dr. Stephanie Schell, Dr. Ashwin Chari, and Dr. David Haselbach⁴, all three co-members of Prof. Stark's group. They taught me a great deal of what I know today about science, for which I am very grateful. Steffi and Ashwin have mainly contributed to my biochemical education, while I owe David an extremely large part of my understanding of the methodology associated with cryoEM.

Besides the three persons mentioned above, I would also like to thank all current and previous members of the *Department of Structural Dynamics*: Alexey, Alycia, Benjamin, Claudia, Cole (special thanks for proofreading this thesis!), Dietmar, Dirk, Elham, Fabian, Felix, Frank, Gaby, Georg, Gudrun, Jil, Juliane, Kami, Karl, Karl-Heinz, Kashish, Lukas, Mario, Niels, Prakash, Sabrina, Suzan, Thomas, Tobias, Uma, Uwe, Wen-Ti, and Zhenwei. You all have always been greatly supportive, super enthusiastic, and very inspiring. Working with you has helped me get over setbacks and through stressful times, whether in or outside the lab. It is a great pleasure to be a part of this team.

This thesis would not have been possible without my family and friends. I would like to thank my parents for their love and support throughout the last four years and my entire life. They encouraged me to strive for my goals and gave me the confidence to make my way to the point where I am now, write these lines. Furthermore, I am grateful to my friends, especially those I had the pleasure to meet in the basketball division of the *ASC Göttingen von 1846 e.V.*, for supporting me and expanding my life with a lot of fun activities and non-scientific input.

A very special thanks goes to my girlfriend Hannah. She has always supported me, especially when things did not go as planned. She motivated me when I had no motivation and encouraged me when I was discouraged. Thank you for this and so much more!

⁴As of October 2017: Research Institute of Molecular Pathology, Vienna, Austria

Contents

Abstract	II
Acknowledgment	IV
List of Tables	X
List of Figures	XII
1 Introduction	1
1.1 The Central Dogma of Molecular Biology	1
1.2 The Ribosome and the Machinery of Translation	2
1.2.1 The Four Stages of Translation	2
1.3 The Scanning Mechanism of Translation Initiation	4
1.3.1 Start Codon Identification	4
1.3.2 The Structural Context of the 48S Initiation Complex Decoding Center	8
1.3.3 The Human Translation Initiation Factor 3	9
1.4 A Short History of Single-Particle cryoEM	11
1.4.1 The Invention of the Electron Microscope	11
1.4.2 Studying Biological Samples	11
1.4.3 Single-Particle Electron Cryo-Microscopy	12
1.4.4 The <i>Resolution Revolution</i>	13
1.4.5 Investigating Structurally Dynamic Macromolecules	14
1.5 Aim of This Thesis	16
2 Material and Methods	17
2.1 Materials	17
2.1.1 Reagents and Consumables	17
2.1.2 Machines and Equipment	18
2.1.3 Software	19
2.1.4 Buffers	19
2.2 Biochemical Methods	20
2.2.1 Protein Purification	20
2.2.1.1 Purification of Native eIF3 from HeLa Cells	20
2.2.1.2 Purification and Reconstitution of 48S-IC	22
2.2.2 MS-MS Cross-Linking Data Acquisition	22
2.3 Electron Microscopy	23
2.3.1 Fundamentals of Single Particle TEM	23
2.3.1.1 The Electron Microscope	23
2.3.1.2 Contrast Formation in cryoEM	25
2.3.1.3 Image Detection	26
2.3.1.4 Nyquist-Shannon Sampling Theorem	28

2.3.1.5	Image Pre-processing	29
2.3.1.6	2D Alignment and Classification	30
2.3.1.7	3D Model Building	34
2.3.1.8	3D Classification	36
2.3.1.9	3D Refinement	37
2.3.1.10	Resolution Assessment and Validation	38
2.3.2	Sample Preparation for Electron Microscopy	39
2.3.2.1	Stabilization of Macromolecular Complexes (GraFix)	39
2.3.2.2	Negative Stain Grid Preparation	40
2.3.2.3	Cryo-Grid Preparation and Vitrification	40
2.3.3	Image Acquisition	41
2.3.4	Image Processing	41
2.3.4.1	Frame Alignment and Dose Weighting	41
2.3.4.2	CTF Correction and Micrograph Assessment	42
2.3.4.3	Particle Selection and 2D Processing	42
2.3.4.4	Initial Model Building	43
2.3.4.5	3D Conformational Analysis	43
2.3.4.6	Signal subtraction	44
2.3.4.7	3D Refinement and Resolution Determination	45
2.3.4.8	Structure Assessment	45

3 Results 47

3.1	Purification of Native eIF3 for cryoEM	47
3.1.1	Chromatography-Free Isolation from HeLa Cells	47
3.1.2	Streptomycin Treatment	49
3.1.3	Mass Spectrometry Complex Analysis	50
3.2	Structural Analysis of eIF3	51
3.2.1	Chemical Fixation Optimization	52
3.2.2	Buffer-Exchange	54
3.2.2.1	Buffer-Exchange Column Inactivation	55
3.2.2.2	Sucrose Retention	55
3.2.3	Initial Model Reconstruction	56
3.2.3.1	Heterogeneity Analysis	58
3.2.4	CryoEM Structure of Human eIF3	59
3.2.4.1	Structural Heterogeneity Analysis of Native eIF3	60
3.3	eIF3 in Context of the 48S Translation Initiation Complex	63
3.3.1	48S-IC cryoEM Sample Preparation	64
3.3.2	Image Acquisition of the 48S-IC	65
3.3.3	Structural Analysis of the 48S-IC	66
3.3.3.1	48S-IC Starting Model Reconstruction	66
3.3.3.2	Segmenting the 48S-IC Structure	66
3.3.3.3	Hierarchical Classification and Sorting	68
3.3.3.4	Refinement and High-Resolution Structure	68
3.3.4	Model Building and Local Classification	71
3.3.4.1	Rigid Body Fitting for the 48S-IC	71
3.3.4.2	Open- and Closed-conformation of the 40S Head	73
3.3.4.3	Cross-linking Mass Spectrometry for Factor Localization	74
3.3.4.4	Focused Classification of eIF2 and the mRNA Entry Channel	77

3.3.4.5	Focused Classification of mRNA Exit Channel	79
3.3.5	Overall 3D-PCA Analysis	82
4	Discussion	85
4.1	Biochemical Advances	85
4.1.1	eIF3 Purification	85
4.1.2	Sample Stabilization	86
4.2	Interpretation of the eIF3 Structure	87
4.2.1	Flexibility Within the eIF3 Structure	87
4.2.2	eIF3 Composition and Stability	89
4.3	Interpretation of the 48S-IC Structure	90
4.3.1	The eIF3 core Domain	90
4.3.2	Open- and Closed-conformation of the 48S-IC	92
4.3.3	The Vicinity of the 48S-IC Decoding Center	92
4.3.4	Localization of the eIF3big Sub-Complex	94
4.3.5	Structural Heterogeneity at the mRNA Exit Channel	96
4.3.6	Localization of eIF3j	97
4.3.7	Model and Dynamics of Eukaryotic Translation Initiation	97
5	Conclusion and Outlook	101
A	Abbreviations	103
B	Supplementary information	107
B.1	Tables	107
B.2	Figures	115
C	References	119
	Curriculum Vitae	129

List of Tables

1.1	List of eukaryotic translation initiation factors	6
2.1	List of used reagents and consumables	17
2.2	List of used machines and equipment	18
2.3	List of used software	19
2.4	List of used buffers	19
3.1	Mass spectrometry analysis of native human eIF3	51
3.2	Measuring retention potential of buffer-exchange columns	56
3.3	PDB atomic models used for rigid body fitting the 48S-IC	71
S1	Mass spectrometry analysis of the native human eIF3 (addition)	107
S2	List of identified 48S-IC cross-links	114

List of Figures

1.1	Overview of translation and tRNA positions	3
1.2	Schematic view of eukaryotic translation initiation	5
1.3	Schematic representation of translation initiation factor arrangement	8
2.1	Basic components of an electron microscope	24
2.2	Experimentally measured DQE as function of spacial frequency	26
2.3	Fourier filtering	30
2.4	Common line theorem	35
2.5	Resolution overview	38
3.1	eIF3 unfolding monitored with ProteoPlex	48
3.2	Chromatography-free purification strategy for eIF3	50
3.3	Titration for chemical fixation of eIF3	52
3.4	eIF3 particles with and without fixation	53
3.5	Improved buffer-exchange procedure	54
3.6	Initial negative stain reconstruction of eIF3	57
3.7	3D Classification of negative stain eIF3 dataset	58
3.8	CryoEM micrograph and 2D classes of eIF3	60
3.9	3D classification strategy for high-resolution eIF3 structure	61
3.10	Heterogeneity analysis of human eIF3 cryo data	62
3.11	BS3 fixation titration for 48S-IC	64
3.12	Cryo data acquisition and 2D classification of 48S-IC	65
3.13	Segmented 48S-IC structure with factor annotations	67
3.14	Hierarchical sorting strategy for 48S-IC complex	69
3.15	High-resolution reconstruction of the closed 48S-IC	70
3.16	Rigid body fitted models of 48S-IC components	72
3.17	Open- and closed-conformation of the 48S-IC	74
3.18	Additional density in the open-conformation 48S-IC	75
3.19	MS-MS BS3 cross-link analysis of the 48S-IC	76
3.20	Local classification of eIF2 and the decoding center	77
3.21	Factor localization via cross-link mass spectrometry	78
3.22	Local classification of the mRNA exit channel	80
3.23	Cross-links found for proteins at mRNA exit channel	81
3.24	3D-PCA analysis of the 48S-IC	83
4.1	Free eIF3 compared to its 40S-bound conformation	88
4.2	Human eIF3 core compared to yeast eIF3ac	91
4.3	Re-localization of eIF1 after start codon recognition	93
4.4	Positioning of eIF3 subunits b, i, and g	95
4.5	Summarized model of eukaryotic translation initiation	98
S1	SDS-PAGE of eIF3 purification gradients	115
S2	Calibration curve for sucrose concentration	116
S3	Segmented 48S-IC structure with factor annotations — extended	117
S4	eIF3 core components and rigid body fitting	118

1 Introduction

All living organisms share certain characteristics that distinguish them from non-living objects: they are composed of cells, all can reproduce¹, they grow and develop, consume and process energy, respond to their environment, they can regulate their metabolism, and evolve [1]. To accomplish these tasks, they need the ability to store, evaluate, change, and pass on information. This information is stored genetically in the deoxyribonucleic acid (DNA) and determines the specific attributes of every individual living being. Within cells, the genetic information is transmitted and interpreted along a fixed path referred to as the *central dogma of molecular biology*. It is divided into several stages separated by conversion steps of the molecular information carrier. The underlying mechanisms differ considerably between prokaryotes and eukaryotes. For simplification, only the eukaryotic processes will be described here. [2]

1.1 The Central Dogma of Molecular Biology

Transcription and *translation* are the means by which a cell converts and interprets the information stored in DNA. These sequential steps are the framework for the production of enzymes, hormones, or other molecules, but they are simultaneously the target for a variety of regulatory processes. This regulation helps cells to cope with changing external conditions and thus has direct implications on concentrations of individual molecules and the (relative) protein composition of a cell. As a result, studying translation, transcription, and their regulation will contribute to a comprehensive understanding of all inner-cellular processes. [3]

During the first step of gene expression, the nucleic acid sequence of the DNA is transcribed into pre-mature messenger RNA (pre-mRNA) by *RNA-Polymerase II* inside the nucleus of a cell. While transcription is still ongoing, additional enzymes extend the pre-mRNA with a 5'-cap composed of 7-methyl guanosine (m⁷G). Shortly after being fully transcribed, the capped pre-mRNA is subjected to further processing: the sophisticated process of *splicing*. Thereby, non-coding regions of the pre-mRNA, named introns, are removed while the remaining parts, named exons, are fused together. Additionally, *Poly(A)-Polymerase* appends multiple adenosine monophosphates to the 3'-end of the messenger RNA (mRNA) in a process called *polyadenylation* to form mature mRNA.

¹Viruses are the exception, which need their hosts to reproduce.

Both the 5'-cap and the 3'-poly(A)-tail are essential to protect the mRNA from degradation and label it for coding. These modifications distinguish the mature mRNA from other ribonucleic acids (RNAs). Eventually, the mature mRNA gets exported from the nucleus into the cytosol to serve as a template for translation. [3, 4]

1.2 The Ribosome and the Machinery of Translation

The cytoplasmic ribosome² is a large macromolecular complex partly composed of ribosomal RNAs (rRNAs) that synthesizes most of the proteins inside a cell. The RNA within ribosomes forms very conserved and stable secondary structures that provide the backbone for all ribosomal particles. [5]

The eukaryotic 80S ribosome with a mass of approximately 3.2–3.6 MDa is considerably larger than its prokaryotic 70S counterpart (~ 2.5 MDa). It is composed of two subunits, the smaller 40S ribosomal subunit, and the larger 60S ribosomal subunit. The 40S subunit, build from the 18S rRNA and 33 proteins, contains the mRNA reading channel as well as the decoding center (DC). The 60S subunit, build from three rRNAs (5S, 5.8S, and 28S rRNA) and 49 proteins, contains the peptidyl transferase center (PTC) and the exit channel for the nascent polypeptide chain. The PTC is the peptide bond formation site during elongation of the newly synthesized polypeptide chain and therefore considered the main active center of the ribosome. [6–8]

1.2.1 The Four Stages of Translation

Translation has four distinct stages: **initiation**, **elongation**, **termination**, and **recycling**. During initiation, the ribosome comes into contact with the mRNA and searches for the start codon (described in more detail in section 1.3). Once found, the small ribosomal subunit starts elongation. In this second and processive step, it translates the information from the mRNA into a polypeptide chain composed of amino acids. The coding sequence on the mRNA is enclosed by the 5' untranslated region (5'-UTR) upstream of the start codon (AUG) and the 3' untranslated region (3'-UTR) that begins with and continues downstream of the first stop codon (UAG, UGA, or UAA). When reaching a stop codon, the ribosome releases the freshly synthesized polypeptide chain during termination. Depending on external signals during the recycling stage, the ribosome either enters another round of translating its already bound mRNA or disassembles into its subunits. [9]

²Mitochondria have their own but rather different 55S mitoribosomes. For the sake of clarity, they shall be omitted from the discussion here.

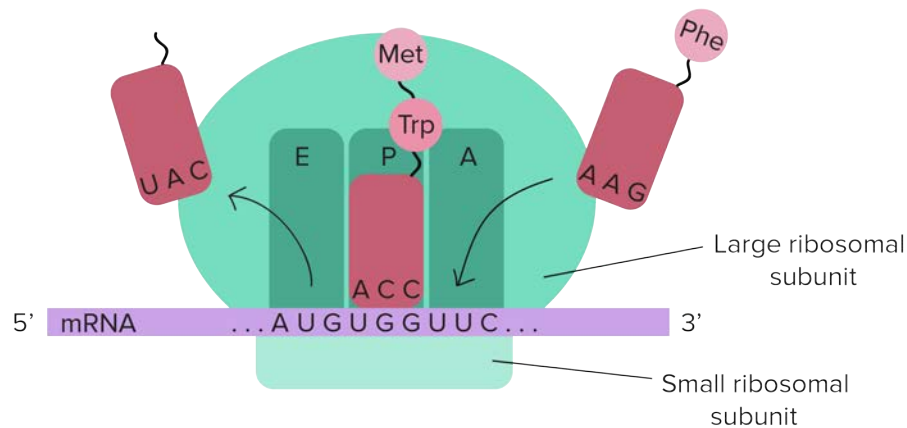


Fig. 1.1: Overview of translation and tRNA positions. The ribosome provides the structural scaffold and facilitates the catalysis of translation. Inside the decoding center of the ribosome, tRNAs adopt one of three possible positions in consecutive order: **A**cetyl-, **P**eptidyl- and **E**xit-site (right to the left in the picture). The mRNA is moving through the ribosome while being read and translated. The growing polypeptide chain, later folded into mature protein, leaves the ribosome through its designated exit channel. (*The figure was reproduced with permission from “Translation: figure 3”, by OpenStax College - cnx.org, Concepts of Biology; CC BY 4.0*)

The amino acids necessary for the synthesis of the nascent polypeptide chain are delivered to the ribosome by transfer RNAs (tRNAs). Each tRNA carries exactly one amino acid, which is defined by the three nucleotide sequence located in the anticodon stem loop (ASL) of a tRNA. Enzymes called aminoacyl-tRNA-synthetases (aaRSs) are responsible for loading each tRNA with its corresponding amino acid by reading their anticodon. A tRNA loaded with an amino acid is called aminoacyl-tRNA (aa-tRNA). [10, 11]

The tRNAs can occupy one of three positions inside the ribosome which are in between the 40S and 60S subunit: the aminoacyl (A), the peptidyl (P), and the exit (E) site (see figure 1.1). Each site corresponds to one codon, a three nucleotide long segment on the mRNA encoding a specific amino acid. The A-site is where new tRNA enter the ribosome, starts to interact with the mRNA, and the first codon-anticodon-interaction takes place. In this process, the tRNA’s anticodon stem-loop enters the decoding center of the 40S subunit, while the aminoacylated end gets oriented towards the PTC of the 60S subunit. The tRNA in the P-site has the nascent polypeptide chain bound while still maintaining the codon-anticodon interaction. As soon as the polypeptide chain gets transferred onto the next tRNA in line, the empty, de-aminoacylated tRNA will advance from P- to E-site. [9, 12]

Two eukaryotic elongation factors (eEFs) are important to ensure correct sequence and direction of the translational steps: eEF-1 and eEF-2, which are both GTPases. Elongation factor 1, a homolog of the prokaryotic EF-Tu, forms a ternary complex (TC) with the aa-tRNA and GTP to deliver a new amino acid to the ribosome. In principle, any TC-

bound aa-tRNA can enter the A-site, but only if the correct codon-anticodon-interaction is detected, the ribosome releases eEF-1 from the aa-tRNA via the hydrolysis of its GTP. As such, the ribosome functions as a GTPase-activating protein (GAP) for eEF-1. The release of eEF-1 stabilizes the aa-tRNA in the A-site and facilitates peptidyl-transferase activity. Furthermore, the geometry of the DC allows the new aa-tRNA in the A-site to only interact with the three nucleotides directly next to those of the P-site. This restriction prevents shifts of the open reading frame (ORF) and is one major fidelity mechanisms of the ribosome. The second factor eEF-2, which is the counterpart of the prokaryotic EF-G, is important during translocation of the tRNA together with the mRNA after one successful round of polypeptide elongation and is therefore partly responsible for the directionality of translation. Through reaching into the intersubunit space between the 40S and 60S, eEF-2 stabilizes the codon-anticodon interaction of the tRNA that is about to move from A- to P-site and eventually leaves the ribosome. [11, 13]

1.3 The Scanning Mechanism of Translation Initiation

Of the many steps of translation, initiation is the one during which it is decided whether an mRNA is translated or not. A plethora of different RNAs and proteins cooperate in a highly orchestrated manner to ensure that the correct start codon is identified and thus the correct sequence on the mRNA is processed. Over the past years, a variety of genetic, biochemical, and structural experiments have provided a model of the mechanisms and interactions underlying translation initiation. For most mRNA molecules, it is facilitated via the so-called *cap-dependent scanning mechanism*: a methionine tRNA is activated and loaded onto the small ribosomal (40S) subunit with the help of eukaryotic translation initiation factors (eIFs) 1, 1A, 2, 3, and 5 (for a full list, see table 1.1). Other eIFs (4A, 4B, 4E, and 4G) bind to the mRNA to bring it into association with the 40S tRNA complex (called 43S pre-initiation complex (PIC) at that stage) to form the 48S initiation complex (IC). This final initiation complex starts scanning the mRNA for the start codon. As soon as it identifies the start codon, it recruits the large ribosomal subunit (60S) to initialize translation elongation (the exact sequence of events is depicted in figure 1.2) [10]. Another way of translation initiation that is utilized less frequently and entirely cap-independent uses internal ribosome entry sites (IRESs). It is not relevant for this thesis and thus omitted (described elsewhere [15]).

1.3.1 Start Codon Identification

As the first prerequisite for translation initiation, eIF2 is activated by the guanosine exchange factor eIF2B via the substitution of an eIF2-bound guanosine-5'-diphosphate

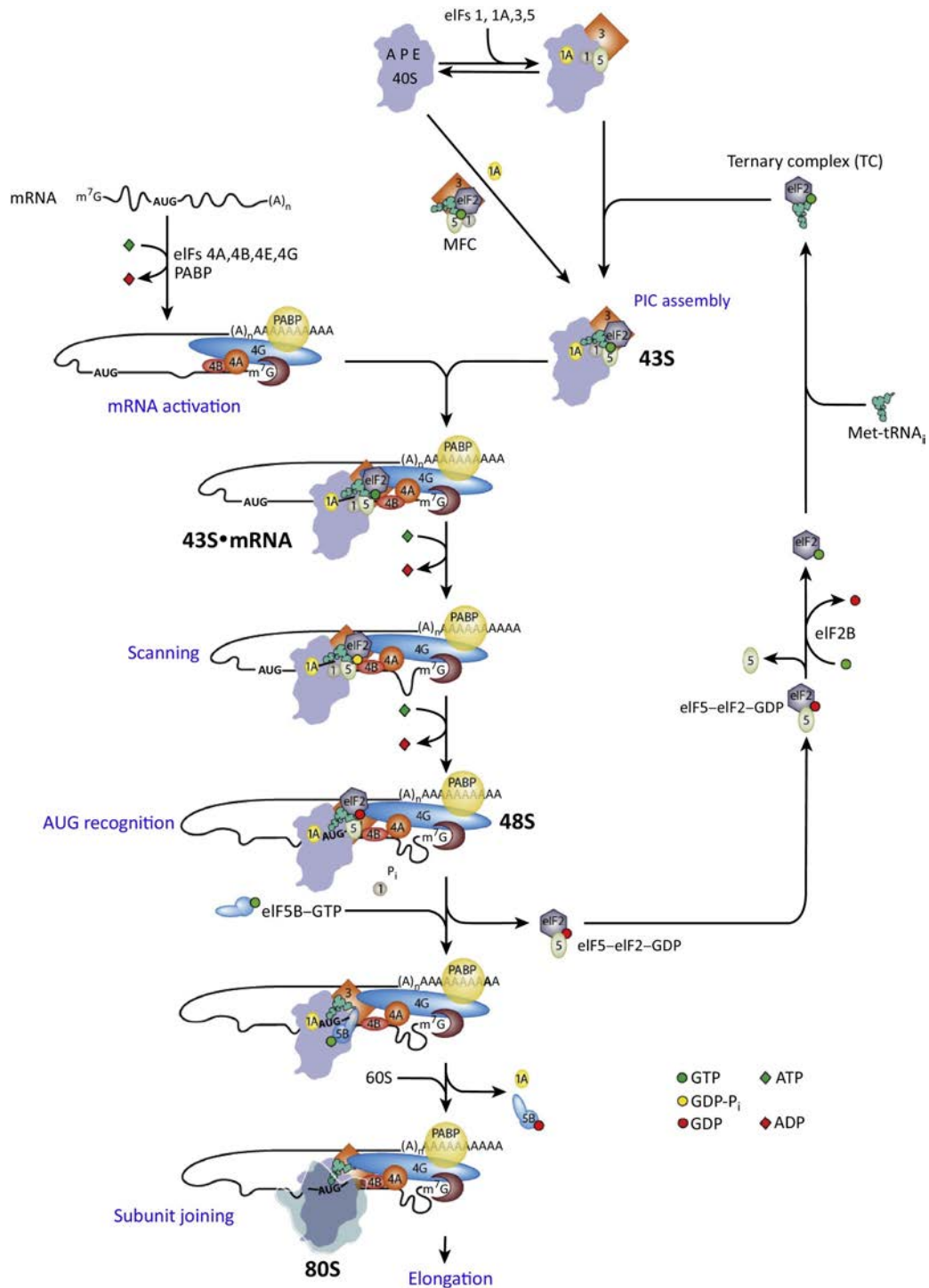


Fig. 1.2: Schematic view of eukaryotic translation initiation. Translation initiation starts with the activation of eIF2 via the guanine exchange factor eIF2B by substitution of GDP with GTP. Subsequently, eIF2-GTP recruits the Met-tRNA_i forming the TC. Now, together with additional factors (eIF1, 1A, 3, 5), eIF2 binds to the 40S ribosomal subunit forming the 43S PIC. Meanwhile, the eIFs 4A, 4B, 4E, and 4G activate an mRNA, which they now bring into contact with the 43S PIC forming the 48S-IC. Afterwards, the ribosomal subunits start scanning the mRNA for an AUG start-codon with *Kozak* characteristics. Once found, eIF5B is recruited to release eIF2 and eIF5, making way for 60S subunit joining. Subsequently, eIF5B and eIF1 leave the complex as well and elongation starts. (The figure was reproduced with permission from [14])

Table 1.1: List of eukaryotic translation initiation factors. Molecular weights for the human variant of each factor according to the polypeptide sequences provided in the UniProt protein database. Functions of factors according to [22]. * *eIF3* has 13 subunits (*a-m*) in mammals, 5 of which (*a, b, c, i, g*) are universally conserved.

Factor	Protein	Bacterial homolog	MW [kDa]	Function (short description)
eIF1	eIF1	IF3-CTD	13	Ensures the fidelity of start-codon recognition. Promotes 48S-IC assembly via TC binding.
eIF1A	eIF1A	IF1	17	Cooperates with eIF1 in start codon recognition and is involved in eIF5B recruitment.
eIF2	eIF2 α	none	36	A GTPase that is part of the TC together with the Met-tRNA _i and GTP. Has a vital role during AUG recognition and prevents premature ribosomal subunit joining.
	eIF2 β	none	38	
	eIF2 γ	none	51	
eIF3	eIF3a-m*	none	~790	Stimulates binding of the TC, eIF1, eIF4G, eIF5, and mRNA to the 40S ribosomal subunit. Functions as scaffold for the formation of the 48S-IC. Facilitates start codon recognition via interaction of its subunits with the 40S and the mRNA.
eIF4A	eIF4A	none	46	DEAD-box ATPase and ATP-dependent RNA helicase.
eIF4B	eIF4B	none	69	An RNA-binding protein that enhances the helicase activity of eIF4A.
eIF4E	eIF4E	none	25	Binds the m ⁷ G-cap of the mRNA 5'-UTR.
eIF4F	eIF4A,E,G	none	246	Synonym for a complex of eIF4A, eIF4E, and eIF4G. Mediates mRNA unwinding and attachment to the 43S complex.
eIF4G	eIF4G	none	176	Binds to eIF4A, eIF4E, eIF3, PABP, SLIP1, and mRNA to enhance helicase activity of eIF4A.
eIF4H	eIF4H	none	27	RNA-binding protein that enhances helicase activity of eIF4A and is homolog to eIF4B.
eIF5	eIF5	none	49	Is required for eIF2-GTP hydrolysis by functioning as a GTPase activating protein upon start codon recognition.
eIF5B	eIF5B	IF2	139	Promotes the release of eIF2 and joining of the 60S ribosomal subunit.

(GDP) with a guanosine-5'-triphosphate (GTP). eIF2-GTP recruits the methionine initiator-tRNA (Met-tRNA_i), a tRNA with a slightly different sequence, specialized on start codon recognition and thus called initiator-tRNA. The newly formed complex of eIF2-GTP and Met-tRNA_i is called ternary complex (TC), similar to the complex of eEF-1, GTP, and aa-tRNA during elongation [16, 17]³. Promoted by the eukaryotic translation initiation factors eIF1, eIF1A, and the multisubunit eIF3, the TC is loaded onto the 40S ribosomal subunit, forming the 43S PIC [18, 19]. eIF1 and eIF1A both are necessary during scanning to identify the correct position of the start codon AUG. Additionally, eIF1A is involved in the recruitment of eIF5B once the AUG is successfully found. Subsequently, eIF5B triggers the release of eIF2 and eIF5 to make way for joining of the 60S large ribosomal subunit. The eukaryotic translation initiation factors eIF2 and eIF3 are each composed of several proteins. eIF2 is composed of an α -, β - and γ -subunit, while eIF3 contains up to 13 different proteins in *H. sapiens*, called eIF3a through eIF3m. *S. cerevisiae*, also eukaryotic, has only eight eIF3 subunits. [20, 21]

In parallel to tRNA recruitment, the mRNA gets activated by another set of eukaryotic translation initiation factors, namely eIF4A, eIF4B, eIF4E, eIF4G together with the poly(A)-binding protein (PABP). eIF4F binds to the m⁷G-cap of the mRNA 5'-UTR, which in return is attached to the scaffolding protein eIF4G. The mRNA forms a loop where the poly(A) tail is also bound to eIF4G via PABP. The DEAD-box helicase eIF4A⁴ and activating factor eIF4B complete the set of mRNA-bound eIFs. [24, 25]

In the next step, the 43S PIC attaches to the 5'-end of the mRNA, forming the 48S-IC, and starts scanning the 5'-UTR for an AUG start codon. The first AUG codon recognized on the mRNA is usually selected as the start codon if the surrounding sequence context complies with the *Kozak* consensus, bearing a guanine at +4 and a purine at -3 (positions relative to the AUG). [26, 27]

As soon as the complementary anticodon of the Met-tRNA_i recognizes the start codon located in the P-site of the 40S ribosomal subunit, the hydrolysis of GTP in the TC is triggered by the release of eIF1 and facilitated with the help of eIF5. Subsequently, eIF2-GDP is released from the complex as well. This irreversible event marks a crucial checkpoint during translation initiation since it commits the ribosome to a particular AUG codon and the corresponding ORF. Eventually, stimulated by eIF5B, the large ribosomal (60S) subunit joins the complex to form the full 80S ribosome and together they enter the elongation phase of translation. [28, 29]

³Since this thesis deals first and foremost with the initiation of translation, the term ternary complex (TC) will always refer to the complex of eIF2-GTP and Met-tRNA, unless stated otherwise.

⁴The name DEAD-box refers to the D-E-A-D (Asp-Glu-Ala-Asp) motif found in these enzymes. [23]

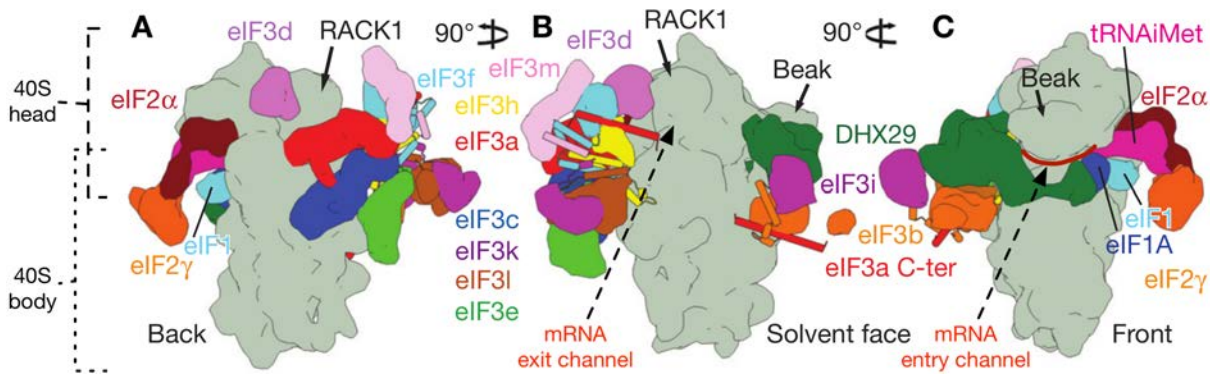


Fig. 1.3: Schematic representation of translation initiation factor arrangement. Here, the arrangement of the translation initiation factors for the mammalian 43S pre-initiation complex are shown. The 40S ribosomal subunit is depicted in gray, all other components are labeled accordingly. (The figure was reproduced and modified with permission from [30])

1.3.2 The Structural Context of the 48S Initiation Complex Decoding Center

As just described, a large set of initiation factors have to work together in an elaborate manner during translation initiation. The high level of complexity in regulation is also eminent when considering the structural composition of the 48S-IC (figure 1.3 shows a schematic arrangement for the closely related 43S pre-initiation complex). The 40S ribosomal subunit, which can be described as the centerpiece of translation initiation, has two particularly noteworthy sites: firstly, the mRNA entry channel (figure 1.3-C) close to the DC with A-, P-, and E-site at the (designated) subunit interface side, and secondly, the mRNA exit channel (figure 1.3-B) on the solvent side. Most of the eIFs bind at either of both positions and interact with the 40S. Furthermore, the whole 40S subunit can be divided (figure 1.3-A) into a slightly larger body and a head. The DC and the mRNA entry channel are positioned directly between these two parts and the 40S head can tilt by $\sim 3^\circ$ compared to the 40S body in order to change the space available in the decoding center. As described before, the TC contains eIF2 and the Met-tRNA_i. It binds on the subunit interface site of the 40S head and positions the tRNA in the P-site of the DC. However, due to the tilting of the 40S head, the tRNA can adopt two positions in the P-site during translation initiation: P_{IN} and P_{OUT} , which differ by $\sim 10 \text{ \AA}$ primarily as to whether or not the tRNA's anticodon interacts directly with a particular codon on the mRNA via hydrogen bond base-pairing. Therefore, these two states of the 40S-TC-complex are referred to as *open-* and *closed-*conformation. [31, 32]

Currently, evidence suggests that the 48S-IC exists in the open-conformation while scanning the mRNA for a start codon that matches all conditions for translation initiation and, once found, adopts the closed-conformation after committing to an AUG codon via GTP-hydrolysis. This process mainly involving the TC is supported by the two factors

eIF1 and eIF1A, which bind near the TC. The eIF1 binding site is close to the P-site of the DC. eIF1 facilitates two main tasks: it promotes binding of the TC to the 40S while at the same time blocking the accommodation of the Met-tRNA_i in the P_{IN} position. The latter step is presumably necessary to keep the 48S-IC open during scanning and allow mRNA movement through the DC. eIF1A binds closer towards the A-site. Its N-terminal tail (NTT) reaches into the DC and interacts with both mRNA and tRNA. This interaction most likely stabilizes the open-conformation and participates in probing for the current codon in the P-site for a suitable start codon. Meanwhile, the γ -subunits of eIF2 stabilize the Met-tRNA_i in the P_{OUT} position and eIF2 α interacts with the -3 residues (relative to AUG) on the mRNA [29, 33]. While this manuscript was in preparation, *Eliseev et al.* [34] published a study where they investigated the human 48S-IC. They were able to prove most of the just mentioned factor positions via cryoEM structure determination, cross-linking mass spectrometry, or both.

For the transition from the open- to the closed-conformation, the components of the 48S-IC must undergo various structural rearrangements: eIF1 has to be released because it would otherwise clash with the 40S head. This steric mechanism is a conserved feature which is not only found in eukaryotes but was also observed in archaea and bacteria [27]. It is assumed to be a crucial fidelity mechanism which detects wrong start codons by preventing 40S head closure. For the Met-tRNA_i to adopt the P_{IN} position and the formation of the final codon-anticodon base pairs, it was observed that the whole TC has to tilt slightly to form new interactions between the mRNA inside the DC and the P-site tRNA. Additional contacts between mRNA and 40S tighten the grip around the mRNA to prevent any unexpected ORF-shifts. However, the structural basis for the hydrolysis of GTP and the closely connected release of eIF1 is not yet fully understood. [27, 32]

1.3.3 The Human Translation Initiation Factor 3

The largest of the eukaryotic translation initiation factors in humans is the 13-subunit translation initiation factor 3. It serves as a scaffold for the interaction of the TC, the 40S, and other eIFs. Additionally, its subunits are supposedly involved in several steps during mRNA scanning and AUG recognition, although the mechanistic and regulatory details remain unclear. The number of eIF3 subunits can vary considerably: in *S. cerevisiae*, for example, eIF3 is composed of only eight subunits and is much lighter in total mass than the human variant (approximately 391 kDa instead of 790 kDa). The thermophilous fungus *C. thermophilum* contains all 13 subunits as in humans, rendering it a suitable model organism although the mass of the individual components can be significantly different from their human counterparts [35, 36]. Knowledge regarding the structure of the human

eIF3 complex provides insights into its role during translation initiation, especially its capability to bind mRNA and role during start codon identification.

Biochemical and structural data indicate that the human eIF3 contains an octameric core structure. Therein, the individual proteins are oriented around a horseshoe-shaped core and show five distinguishable main domains expanding radially. At the center, the C-terminal domains (CTDs) of six eIF3 components (eIF3a, c, e, k, l, and m) form a helical bundle that provides structural stability to the core. In the literature, this characteristic structure is called *proteasome*, *COP9 signalosome*, *translation initiation factor* (PCI) domain, referring to the structural similarities among the three name-giving molecules. The two remaining eIF3 core subunits (f and h) consist of *Mpr1*, *Pad1*, *amino-N-terminal* (MPN) domains that are connected to the CTD helix-bundle as well. In the 48S-IC, the eIF3 core structure binds to the solvent site of the 40S, away from the DC and the entry channel. Mainly two (a and c) of its five proteins are responsible for this connection. [37, 38]

Previous investigations have shown that all eIF3 subunits play a vital role during eukaryotic translation initiation and although the human eIF3 is composed of 13 individual proteins, only eight can be found in the PCI domain as just described. Therefore, the debate about the structure, location, and specific function of the other five eIF3 subunits (b, d, g, i, and j), when bound to the 40S, has been diverse. Cross-linking mass spectrometry experiments have identified direct interactions of eIFa and eIFd with the mRNA at the 48S-IC exit channel. Additionally, it was shown that the presence of the full eIF3 protected the rRNA near the entry channel from chemical or enzymatic modification/cleavage, implying that at least some subunit must be bound here. [39]

The structural characterization of the single eIF3 proteins is, at least for the human variant, still ongoing and it is no surprise that the Protein Data Bank (PDB) [40] does not provide full-coverage entries of all eIF3 subunits. Only for the subunits b, i, k, and parts of d, are individual high-resolution X-ray structures available [41–44]. These studies have shown that subunits b and i contain a nine- and seven-bladed β -propeller, respectively. Furthermore, a variety of partial structures were found in the database, including the RNA recognition motif (RRM) of eIF3g and parts of eIF3j [45, 46]. Despite all these efforts, no high-resolution structure of the free eIF3 with all the 13 subunits has been presented thus far, which could yield important insights into eIF3's role during mRNA recruitment and start codon recognition. Most of the structural information for this versatile complex has been obtained from pre-initiation complexes (PICs) [30, 47].

1.4 A Short History of Single-Particle cryoEM

For the structural investigation of large macromolecular complexes, three major techniques are used to date: nuclear magnetic resonance (NMR), X-ray crystallography, and 3D transmission electron cryo-microscopy (cryoEM). Of the three methods, cryoEM provides the unique capability to study quite large macromolecules (NMR is heavily limited in complex size) in their natural environment (for X-ray crystallography, protein crystals in which the complexes usually adopt a single and uniform orientation are an inevitable prerequisite). However, cryoEM had to overcome a series of technical limitations and conceptual hurdles in order to present the tools and instruments available today. To gain a better understanding about the advantages of cryoEM compared to other structural determination methods, a look into its history is quite instructive.

1.4.1 The Invention of the Electron Microscope

In 1924, Louis de Broglies published his Ph.D. thesis with the title *Recherches sur la théorie des quanta* [48]. He predicted that the concept of *wave-particle duality* does not exclusively apply to photons but to all components of matter, including electrons, which would show wavelengths well below one nanometer. This idea, known as the *de Broglie hypothesis*, was proven experimentally three years later in 1927 by the *Davisson-Germer experiment*. [49]

In 1931, Ernst Ruska designed, together with Max Knoll (both engineers), the first prototype of an electron microscope although he had no knowledge about the work of de Broglies, as he surprisingly revealed later [50]. This first instrument enabled them to acquire projection images at 400-fold magnification. In 1933, Ruska constructed an electron microscope that already exceeded the magnification of light microscopy. In these machines, the electrons penetrate and traverse the specimen and a projection image is recorded behind it. Therefore, they are called *transmission electron microscopes* [51]. This is the main difference to other methods of electron microscopy, like scanning electron microscopy (SEM) or reflection electron microscopy (REM), which are described in detail elsewhere. [52]

1.4.2 Studying Biological Samples

Although not specifically designed for this purpose in the first place, it turned out that transmission electron microscopes (TEMs) are particularly suitable for imaging biological specimens. However, it required several decades to overcome some fundamental problems arising when working with biological samples, such as the ultra-high vacuum maintained

by the electron microscope inside its column. Any liquid biological sample would not withstand such vacuum, which is why a special specimen preparation technique for biological samples had to be invented. Furthermore, electrons carry a lot of energy and can thus cause severe damage to the specimen molecules when interacting with them. This phenomenon is referred to as *beam damage* and is the main reason why biological specimens are usually imaged at very low electron doses. [53]

During the early years of single particle electron microscopy, the standard method to overcome the problems with radiation sensitive biological material was negative staining. For this method, still being used today, the sample particles are embedded in a thin film of amorphous heavy-metal salt (e.g., uranyl acetate). This creates a negative image of the macromolecules in the electron microscope with the additional advantage of better contrast due to the heavy-metal. However, the resolution is limited by the grain size of the dye.

In 1968, David DeRosier and Aaron Klug were the first to present a successful three-dimensional (3D) reconstruction from EM images showing the tail of the bacteriophage T4 by acquiring two-dimensional (2D) projection images from the negatively stained sample [54]. Klug had earlier proposed that the verification of such 3D structures would only be possible if either imaged from different but known angles by tilting the specimen or by integrating the information from many particles at random orientations [55]. Therefore, in 1970, Klug, DeRosier, and Anthony Crowther presented the method called the *common line approach*, which can be used to determine the relative orientations between different projections of the same particle [56, 57].

1.4.3 Single-Particle Electron Cryo-Microscopy

At the turn of the 1980s, several milestones for the advancement of single particle cryoEM were achieved: on the subject of sample preparation, negative staining of the specimen had been the standard method for many decades. However, in 1981 Jacques Dubochet and Alasdair McDowell published a new and unique method: they showed that thin layers of sample solution applied to an electron microscope (EM) grid could be frozen into vitreous ice when rapidly plunged into liquid ethane. Vitreous ice is a special amorphous form of ice in which water molecules retain their unstructured orientation of the liquid state. This approach, which is the standard method for sample preparation today, has several advantages: the particles were immobilized inside a solid slice, thus protected from the vacuum inside the microscope. Additionally, the lack of any staining dye (like in negative-staining microscopy) and the amorphous nature of the vitreous ice made it possible to directly image the macromolecules themselves. Eventually, the overall cryogenic conditions led to significantly better protection from beam damage. [58]

Meanwhile, obtaining high-resolution results was only possible from thin 2D crystals or highly symmetric particles such as viruses, because here the signal-to-noise ratio (SNR) of the high-resolution information was reasonably high. The signal from single, non-crystalline, asymmetrical, and randomly oriented particles was simply too low for high-resolution reconstructions and due to beam damage, an increase of electron dose was out of the question. The problem of a low SNR for individual images was overcome when Joachim Frank, Marin van Heel, and their colleagues started to think about mathematically processing the data to improve the contrast of cryoEM projection images taken at low exposures. They used cross-correlation functions to identify similar projection images among a larger set of images taken at unknown orientation and averaged these images to improve to SNR. Each image is interpreted as a vector with as many dimensions as it contains pixels. Subsequently, multivariate statistical analysis (MSA) is used to sort the vectors and identify clusters, which in good approximation represent classes of similar 2D projections of a macromolecule [59–62]. Nourished by the increase of computational power at the time, several groups published software packages specifically tailored for the analysis of cryoEM images, such as IMAGIC or SPIDER. [63, 64]

In 1990, the efforts of many distinct scientists, including Fritz Zemlin, Kenneth Downing, and Richard Henderson, made it possible to obtain the first high-resolution structure using cryoEM. This structure of bacteriorhodopsin was obtained via electron crystallography by averaging over many images of the same 2D crystal and subsequent fitting of the amino acid sequence into the reconstructed density [65]. This structure and their work proved that, given a stable specimen-holding stage at cryogenic temperatures, it is possible to limit the beam damage sufficiently to yield the information necessary to see individual amino acid side chains within a complex. [66]

1.4.4 The *Resolution Revolution*

For many years, scientists recorded the projection images created inside an electron microscope on photographic film. Despite great sensitivity and resolution, the use of film has the big disadvantage that it involves a lot of time-consuming steps, like replacement, development, and digitalization. As a result, the acquisition technique limited the number of images that could be taken within a reasonable amount of time and the lack of immediate feedback regarding the specimen quality and optical setup. The introduction of the first direct electronic detectors in the 1990s helped to overcome these problems. In these devices, the electrons are imaged indirectly through a scintillator, like a layer of phosphor, that is tied to a charge coupled device (CCD) image sensor. Since CCD cameras were quite widespread and even found in many consumer products, it was a relatively easy step towards quick and straightforward digital image acquisition. With CCD detectors,

large image datasets could be acquired within a short amount of time and immediately subjected to image processing. [67, 68]

Unfortunately, the CCD detectors lagged behind the properties of film regarding spatial resolution transfer, because the photons could scatter multiple times within the phosphoric layer before being detected. Additionally, the signal from individual incident photons gets considerably reduced, and noise is propagated due to charge-sharing between adjacent pixels. As countermeasures, the layers of scintillator were thinned as much as possible, the surface of the detectors was divided into tiles, and specially designed projection lenses were used [69–71]. Still, the detective quantum efficiency (DQE) for CCD detectors trailed far behind that of film, especially for the high spatial frequencies which are necessary to gain high-resolution information. The DQE is a combined measure of all the effects occurring in an optical system that impact the SNR of the final image. In cryoEM microscopes, it is defined as the ratio of the squared output SNR to the squared input SNR:

$$DQE(u) = \frac{SNR_{out}^2(u)}{SNR_{in}^2(u)}, \quad (1.1)$$

where u is the spatial frequency.

It took another decade for the introduction of the first direct electron detection devices (DDD) to start a technological advancement that is often referred to as the *Resolution Revolution*. These new detectors made the conversion from electrons into photons and onwards into electric signals obsolete. The electrons are detected directly by the chips which, however, require radiation-hard materials. Furthermore, the direct detectors have the advantage of very fast readouts (currently up to 400 Hz) enabling them to operate in the so-called *electron counting mode*: given that the electron dose is below the device-specific threshold, the signal created by an individual electron does not superimpose with that of any other. This non-coinciding makes it possible to match every incident electron to exactly one pixel and even localize the impact positions with sub-pixel accuracy (called *super-resolution*) by computational interpolation if desired. [72, 73]

1.4.5 Investigating Structurally Dynamic Macromolecules

In parallel to the improvements in the field of detector technology, the general progression of computing power and storage capacity led to the development of new and innovative software packages. With these, several groups around the world tried to tackle difficult problems arising for cryoEM image processing. The new recording procedure and speed of direct detectors made it possible to collect movies composed of a frame stack rather than a single image. These frames can be aligned with each other to reduce motion

blurring induced by specimen movement or charging during acquisition [74, 75]. Another very crucial step is the *ab initio* 3D reconstruction of an unknown complex. This step became automated as well, although most cases still require a skilled user to omit false positive results [76, 77]. The development of maximum likelihood algorithms used for 2D alignments against a set of 2D classes or the projections of 3D references, and the sorting of images into more homogeneous classes, was particularly helpful to achieve better resolution using the new detectors. [78–80]

With all these new resources and tools at disposal, it became finally possible to treat large macromolecular complexes as what they are: highly sophisticated molecular machines whose work is frequently associated with a lot of structural dynamics, like the reorientation of whole domains or the binding and dissociation of additional factors. cryoEM proved to be particularly suitable for working with this kind of specimens because the standard sample preparation techniques preserve the full landscape of thermodynamically possible molecular states. This fact, which has been a disadvantage for several decades, finally became an advantage after all and is the reason for cryoEMs leading role in investigating the dynamics of macromolecular machines, like the ribosome, the spliceosome, or the proteasome. Large computer clusters equipped with graphics cards make it possible to analyze and sort for different conformations and refine their structure individually. From this, one does not only achieve a snapshot from the highest populated state but can obtain the full trajectory of enzymatic processes. [81, 82]

1.5 Aim of This Thesis

The eukaryotic translation initiation involves several crucial steps that assure translation fidelity: bringing mRNA in contact with the ribosome and subsequently identifying the correct start-codon. Thus, the aim of this work is to study the complexes and factors involved in this process, namely the free eukaryotic initiation factor 3 and the late-stage 48S initiation complex. As a prerequisite, the establishment of a purification strategy for the free eIF3 is an integral aim of this work. Investigating both complexes as well as their dynamics and different conformations will help to further elucidate the molecular mechanisms involved in translation initiation and the exact order of events. For this, cryoEM has been so far the most suitable method due to its capability to study large macromolecular complexes in their natural, soluble environment. Recent technical developments have made it possible to reconstruct high-resolution structures specifically of conformationally heterogeneous macromolecular complexes.

To further the understanding of eukaryotic translation initiation, the eukaryotic translation initiation factor 3 (eIF3) was investigated in its free form as well as bound to the late-stage 48S initiation complex (48S-IC). In addition, the establishment of a chromatographic-free purification strategy for the free eIF3 was one aim together with studying the localization, dynamics, and purpose of all (additional) factors involved in formation of the 48S-IC.

2 Material and Methods

2.1 Materials

2.1.1 Reagents and Consumables

Table 2.1: List of used reagents and consumables.

Chemicals	Manufacturer
1,4-Dithiothreitol (DTT)	Carl Roth
Adenosine 5'-triphosphate (ATP)	Sigma-Aldrich
Amido Black stain	Sigma-Aldrich
Bradford 1x dye reagent	Bio-Rad Laboratories
Guanosine 5'-triphosphate (GTP)	Sigma-Aldrich
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Sigma-Aldrich
L-Aspartic acid sodium salt monohydrate	Sigma-Aldrich
Lauryl maltose neopentyl glycol (LMNG)	Anatrace
Magnesium chloride (MgCl ₂)	Merck
Octyl glucose neopentyl glycol (OGNG)	Anatrace
Polyethylene glycol (PEG) 400	Sigma-Aldrich
Potassium chloride (KCl)	Merck
Potassium hydroxide (KOH)	Merck
Protein Marker	Bio-Rad Laboratories
SDS-PAGE running buffer	Roth
Streptomycin sulfate	Sigma-Aldrich
Sucrose	Merck
Consumables	
4–15% Criterion TGX protein gel	Bio-Rad Laboratories
Cheesecloth	Local retailer
EM grids Cu200 mesh (R3.5/1, R2/2, R1.2/1.3)	Quantifoil Microtools
Miracloth	Merck Millipore
Nitrocellulose membrane	Merck Millipore
Zeba Spin Desalting Columns, 0.5 mL	Thermo Fisher Scientific

2.1.2 Machines and Equipment

Table 2.2: List of used machines and equipment.

Machine	Manufacturer
Balances	Sartorius
Blender	Waring
Centrifuge LYNX 6000	Thermo Fisher Scientific
Falcon 3EC Direct Electron Detector	FEI Company
GIF Quantum Energy Filter	Gatan
Gradient Master ip	Biocomp
Incubator Multitron Pro	Infors HT
K2 Direct Detection Camera	Gatan
Leica EM GP	Leica
Mortar RM200	Retsch Technology
Multifuge X1R	Thermo Fisher Scientific
Orbital Shaker RS0S20	Pheonix Instrument
Photometer Lambda Bio X	Perkin Elmer
Plate Sealer ALPS 3000	Thermo Fisher Scientific
Pico 17 Heraeus	Thermo Fisher Scientific
Pipetting Robot Microlab Star LET	Hamilton Company
Rt-PCR machine CFX connect	Bio-Rad Laboratories
Thermomixer comfort	Eppendorf
TEM CM200 FEG	Philips
TEM Titan Krios	FEI Company
Vitrobot Mark IV	FEI Company
wX+ Ultra Series ultra centrifuge	Thermo Fisher Scientific
Equipment	
Criterion Vertical Electrophoresis Cell	Bio-Rad Laboratories
Fiberlite rotors (F14, F21, F35L)	Thermo Fisher Scientific
DR201–95 refractometer	A. KRÜSS OPTRONIC
Dounce homogenizer, L	Sartorius
TH-660 rotor	Sorval
Surespin 630 rotor	Thermo Fisher Scientific
SW40Ti rotor	Beckman Coulter

2.1.3 Software

Table 2.3: List of used software.

Software	Source	Citation
Amira 5.2.2	https://www.fei.com/software	[83]
COW EM suite	http://www.cow-em.de	[84–86]
EMAN2	http://blake.bcm.edu/emanwiki/EMAN2	[76]
Gautomatch	https://www.mrc-lmb.cam.ac.uk/kzhang	unpublished
Gctf	https://www.mrc-lmb.cam.ac.uk/kzhang	[87]
ImageJ	http://www.imagej.net	[88]
IMAGIC	https://www.imagescience.de/imagic_em.html	[89]
MotionCor2	http://msg.ucsf.edu/em/software/index.html	[75]
pLink2	http://pfind.ict.ac.cn/index.html	[90]
PyMOL	https://www.pymol.org	
RELION 2.0.3	https://www2.mrc-lmb.cam.ac.uk/relion	[79]
ResMap	http://resmap.sourceforge.net	[91]
SIMPLE	http://www.simplecryoem.com	[77]
UCSF Chimera	https://www.cgl.ucsf.edu/chimera	[92]
xiNET	http://crosslinkviewer.org/index.php	[93]

2.1.4 Buffers

Table 2.4: List of used buffers.

Buffer	Conc. [mmol/L]	Components
1x eIF3 standard buffer	50	HEPES-KOH (pH 7.8)
	100	KCl
	10	MgCl ₂

2.2 Biochemical Methods

2.2.1 Protein Purification

2.2.1.1 Purification of Native eIF3 from HeLa Cells

In order to purify the native eukaryotic translation initiation factor 3, S30 HeLa cytoplasmic extract was used as source material. It was prepared by hypotonic lysis (modifications from previously published protocols [94]): the HeLa cells were harvested from cell culture by centrifugation and subsequently supplemented with hypotonic buffer (containing only a fifteenth of the isotonic salt concentration). Due to this difference in salt concentration between cells and surrounding buffer, an influx of water into the cells driven by osmosis was induced, which led to a swelling of the cells. Afterwards, the cells were lysed with the help of a Dounce homogenizer.

Immediately following hypotonic lysis and centrifugation to pellet the nuclei, the crude cytoplasmic extract was centrifuged again at 30,000x g for 30 minutes at 4 °C. Subsequently, it was decanted and filtered through three layers each of pre-wetted cheese cloth and Miracloth to remove large lipid clumps and other floating aggregates. The flow-through was either used directly or aliquoted into 40 mL fractions, flash frozen in liquid nitrogen, and stored at -80 °C until further use.

Approximately 600 mL of the S30 HeLa cytoplasmic extract were used. If frozen, they were thawed in a 37 °C water bath immediately before use. Afterwards, the extract was supplemented with eIF3 purification buffer from a 10-fold stock to get 1x eIF3 buffer composition (see table 2.4). Additionally, the following compounds were added freshly to complement the buffer: 10 % (w/v) crystalline sucrose, 5 mM dithiothreitol (DTT), 5 mM adenosine triphosphate (ATP), 0.5 mM GTP, 1 mM phenylmethane sulfonyl fluoride (PMSF), and 0.1 % (w/v) octyl glucose neopentyl glycol (OGNG) from a 10 % (w/v) stock.

To remove unspecifically bound or free nucleic acids, a streptomycin treatment was performed by adding 1 % (w/v) of streptomycin sulfate (from 20 % stock) very slowly/dropwise to the extract [95]. This was followed by an incubation period of 30 min at room temperature while gently stirring the extract. Subsequently, the treated extract was centrifuged in an ultra centrifuge at 100.000 x g for 1 hour at 4 °C. Again, the supernatant was filtered through each three layers of cheese cloth and Miracloth (six layers in total), yielding a clarified S100 HeLa cytoplasmic extract.

For the precipitation of the eIF3 protein complexes, the clarified extract was supplemented with 10 % (w/v) of polyethylene glycol (PEG)400 (number specifies the mean molecular

weight), taken from a 80 % (w/v) PEG400 stock also containing the standard concentration eIF3 buffer substances. The 80 % (w/v) PEG400 stock solution was added to the extract in dropwise manner to avoid the local formation of high concentrations of PEG400. The final mix was incubated and slowly stirred for 15 minutes at room temperature.

To separate the precipitated protein from that remaining in solution, the 10 % (w/v) PEG400 solution was centrifuged at 30.000 x g for 30 minutes at 4 °C. Afterwards, the supernatant was carefully decanted and discarded. The remaining pellets containing the 10 % (w/v) PEG400-cut from the original extract were carefully rinsed with PEG-free eIF3 buffer to remove all remaining PEG400 from the centrifugation buckets. Subsequently, the pellets were resuspended with a small amount of resuspension buffer of approximately the same volume as the pellet (composed of eIF3 standard buffer supplemented with 10 % (w/v) sucrose, 5 mM ATP, 1 mM GTP, 5 DTT, 0.1 % (w/v) OGNG, and 0.02 % (w/v) lauryl maltose neopentyl glycol (LMNG)) to keep the concentration of resuspended protein significantly above 1 mg/mL to prevent complex dissociation (known from personal communication with other lab members who work with similar macromolecular complexes). To facilitate a gentle resuspension of the protein pellets, the centrifuge buckets were fixed onto an orbital shaker operating at 250 rpm and 18 °C ambient temperature.

The resuspended material was pooled and loaded onto SW32 sedimentation centrifugation tubes each containing a linear 15–45 % (w/v) sucrose gradient. In addition to the sucrose, the gradient contained the standard eIF3 buffer compounds 1 mM ATP, 0.5 mM GTP, and 5 mM DTT. The amount of material loaded onto each gradient was adjusted to approximately 15–20 mg per tube. Ultracentrifugation was performed for 18 hours at 160,000 x g (avg. speed) and 4 °C.

The gradients were harvested into 400 µL fractions and sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) [96] was used to identify the fractions containing eIF3. These fractions were pooled and the protein therein was precipitated using 30 % (w/v) PEG400 eIF3 buffer similar to the first precipitation step. Again, the PEG400 solution was added slowly, incubated for 30 minutes at room temperature and centrifuged for 30 minutes at 27,500 x g and 4 °C. The supernatant was discarded and the protein pellet was rinsed with resuspension buffer (containing eIF3 standard buffer, 8 % (w/v) sucrose, 5 mM ATP, 1 mM GTP, 5 DTT, and 0.02 % (w/v) LMNG). After removal of PEG400, a small amount of resuspension buffer was added in order to keep the protein concentration above 1 mg/mL and gentle resuspension was performed on an orbital shaker as before.

The resuspended material was again pooled and loaded onto SW40 sedimentation centrifugation linear sucrose gradients containing 10–30 % (w/v) sucrose supplemented with 1 mM ATP, 0.5 mM GTP, and 5 mM DTT. Per gradient, approximately 15–20 mg mate-

rial was loaded and centrifugation was performed for 16 hours at 284,000 x g (avg. speed) and 4 °C.

The gradients were harvested into 400 μ L fractions and once more SDS-PAGE was used to find the fractions that contained eIF3. After pooling them, the same 30 % PEG400 precipitation step was performed as before. After 30 minutes of incubation and gentle steering, centrifugation was performed at 27,500 x g for 30 minutes at 4 °C. The protein pellets were rinsed with resuspension buffer (containing eIF3 standard buffer, 8 % (w/v) sucrose, 5 mM ATP, 1 mM GTP, 5 DTT, and 0.02 % (w/v) LMNG) and subsequently resuspended gently on the orbital shaker in a small amount of the same resuspension buffer as before.

A third round of SW40 sedimentation gradient centrifugation, harvesting, pooling, PEG400 precipitation and resuspension was performed, exactly as before to increase the purity of the complex.

The final protein concentration was determined utilizing the Bradford assay: 1 μ L of resuspended protein solution was added to 1000 μ L of ready-to-use Bradford dye reagent, incubated for 5 minutes at room temperature and subjected to spectrometric absorption analysis at 595 nm wavelength. The measured absorption value was translated into a protein concentration by comparing it to a previously measured calibration curve created from known concentrations of bovine serum albumin (BSA). [97]

2.2.1.2 Purification and Reconstitution of 48S-IC

Expression, purification, and reconstitution of the 48S translation initiation complexes was carried out according to previously published procedures [98]. All work related to this was carried out by Dr. Akanksha Goyal and Sung-Hui Yi, both members of the group of Prof. Marina Rodnina (*Department of Physical Biochemistry, Max-Planck-Institute for Biophysical Chemistry, Göttingen*).

2.2.2 MS-MS Cross-Linking Data Acquisition

When working with cryoEM maps at resolutions worse than 4 Å, the identification of individual amino acids and their side chains becomes difficult if not impossible. Therefore, additional information is necessary to build a *de novo* model or at least to fit already existing structures into cryoEM density maps correctly. A very instructive and comprehensive technique to gain orthogonal information about the orientation of individual proteins as well as the overall architecture of a complex is cross-linking mass spectrometry: the complex is chemically fixed using a chemical agent to connect spatially close

amino acids containing the correct chemical group. Afterwards, the cross-linked complexes are digested, the emerging peptides chromatographically separated, and eventually subjected to tandem mass spectrometry. Inside the instrument, the sequence of the injected peptides, as well as all cross-links between two peptides, can be detected and matched against a reference database containing all amino acid sequences of the proteins within the macromolecular complex at hand.

Here, the sample was incubated with the cross-linker bis(sulfosuccinimidyl)suberate (BS3) at a concentration of 1 mM for 30 min at room temperature. BS3 reacts with primary amino groups and forms chemical links that span a distance of up to 24 Å (measured from C α to C α of the two cross-linked residues). The fixed complexes were digested using trypsin and all links other than between two lysine residues were disregarded. The found cross-links were scored using the software *pLink2* [90]. To reduce the amount of false-positive intermolecular cross-links, an arbitrary score cut-off of 5 was used, which derived from previous experience. Complex digestion, chromatographic peptide separation, mass spectrometry measurement, as well as MS-MS data analysis was carried out by Dr. Jasmin Corso, member of the group of Prof. Henning Urlaub (*Research Group Bioanalytical Mass Spectrometry, Max-Planck-Institute for Biophysical Chemistry, Göttingen*).

2.3 Electron Microscopy

2.3.1 Fundamentals of Single Particle TEM

Biological single particle cryoEM utilizes a series of techniques that are quite unique to the field of biological research and will be explained in more detail on that account: At first, a TEM is used to acquire images, so called micrographs. Subsequently, the individual particles visible on these micrographs are selected, extracted, and stacked. The particle image stacks are then subjected to an array of sophisticated image-processing procedures in order to remove undesired images, improve the contrast and identify groups of similar pictures. Afterwards, another series of methods is applied to recover the information necessary to reconstruct a three-dimensional representation of the originally imaged particles and refine it to high resolution. All these techniques will be described in more detail on the following pages.

2.3.1.1 The Electron Microscope

Invented in 1931 by Ernst Ruska (see section 1.4; [51]), the electron microscope has undergone a long process of technological development and innovation. However, the principal

composition (see figure 2.1) of any electron microscope remains the same and is — to some extent — comparable to that of a light microscope: at the top of the microscope's column sits an electron gun which, just like the light source in a light microscope, generates the illuminating beam. Inside the column, a series of consecutive electromagnetic lenses shapes and directs the electron beam onto a specimen. The interaction between beam and specimen ultimately results in electron scattering events and a system of lenses is used to transform them into an image at the bottom of the column which is detected by a camera.

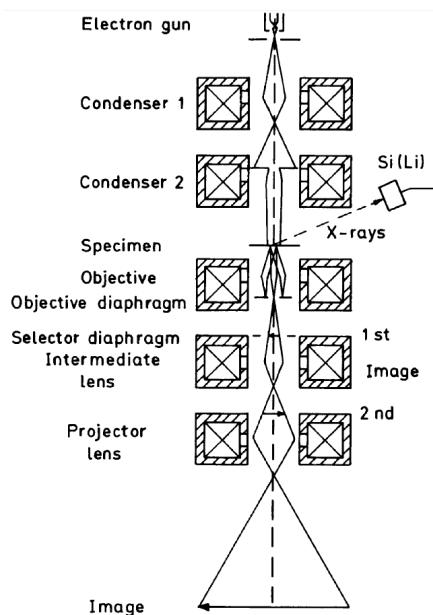


Fig. 2.1: Basic components of an electron microscope. Here, a schematic representation of a typical TEM is shown. The most important parts are labeled on the left. All lenses are depicted as squares, illustrated as a vertical cut through their electro magnetic coils. A principle ray diagram is also shown in the center column with a illustration of the optical influences and the black outline of the beam. (*The figure was reproduced with permission from [99]*)

Electromagnetic Lenses The lenses in an electron microscope are not made from an actual material, but rather formed by the magnetic fields of strong electromagnets. Since electrons are charged particles, they experience a force (the *Loretz force*) when passing through an electromagnetic lens and thus get deflected from their original path. Although electromagnetic lenses usually have considerably worse aberrations than glass lenses (like astigmatism and spherical aberration), one of their advantages is evident: tuning the electrical voltage of the electromagnetic coil within the lenses leads to a change in their focal length. However, round electromagnetic lenses are only capable of condensing an electron beam. Hence, beam diverging is achieved indirectly via strongly converging the beam which leads to a crossover and thereafter a divergent beam.

Field Emission Gun and Acceleration Voltage In contemporary electron microscopes, the most coherent electron beams are generated by field emission guns (FEGs). Inside, a very tiny but sharp tungsten tip, coated with zirconium dioxide, is heated under ultra-high vacuum via the application of several kilovolts of negative potential. At arising temperatures, zirconium oxide has the rare characteristic to provide exceptionally high

electron conductivity. A nearby anode extracts the thus highly mobile electrons from the tip. After passing the anode, the electrons get exposed to an electric field, which accelerates them until their kinetic energy matches that of the field's electrical energy. Thus, the acceleration voltage of the field determines the electron's wavelength λ as

$$\lambda \approx \frac{12.3}{\sqrt{\Phi + 0.000000978\Phi^2}} \quad (2.1)$$

with Φ as the applied electric field. [100]

2.3.1.2 Contrast Formation in cryoEM

In cryoEM, an image is formed as the result of interactions between the incident electrons and the atoms of the specimen. Most of the electrons penetrating the sample will just go through without any interaction. However, due to the electrostatic potential of the positively charged core and the negatively charged shell of the specimen's atoms, incoming electrons can be scattered, both elastically (without energy loss) or inelastically (with energy loss). When an electron collides with a nucleus, it is removed from the beam and thus adding to the amplitude contrast. However, the electron transfers its energy in the process onto the atom causing secondary emissions (e.g., X-rays), ionization, or radicalization. All these effects contribute to the noise in the image and damage the sample, which is the reason for the limited ability of biological samples to withstand high electron doses.

If an electron gets scattered by the Coulomb field of an atom without colliding with its protons or electrons, elastic scattering occurs and a phase shift is introduced. These electrons contribute to the image contrast without damaging the specimen and are therefore the favored event. The phase shift (relative to the unaffected beam) is dependent on the location due to the higher density of the biological specimen (1.33 g/cm^3) compared to the surrounding ice (0.92 g/cm^3) and therefore feasible for contrast formation [101]. The deflection angles of elastically scattered electrons, however, are relatively small, which is why they do not add much to the amplitude contrast under normal conditions. To convert the small phase shift into an amplitude change, images are acquired at a slight defocus of the objective lens. This defocus adds additional phase shift to the scattered electrons and thus stronger amplitude modulation in the resulting diffraction pattern. A stronger amplitude modulation ultimately leads to better contrast in the final image. Mathematically, this effect can be imagined as the convolution of the final image with a spacial point spread function (PSF):

$$\text{defective image} = \text{ideal image} \star \text{PSF} \quad (2.2)$$

Every image taken with an electron microscope represents the convolution of the ideal image with the PSF. Or, in other words, the information from a single point in the specimen is smeared out and distributed over a certain area in the final image. A more convenient way to describe and handle this effect can be achieved by applying Fourier transformation. According to the convolution theorem, a convolution of two images in real space becomes a simple pointwise multiplication in Fourier space. The Fourier transform of the PSF is called contrast transfer function (CTF):

$$\mathcal{F}(\text{defective image}) = \mathcal{F}(\text{ideal image}) \cdot \mathcal{F}(\text{PSF}) = \mathcal{F}(\text{ideal image}) \cdot \text{CTF} \quad (2.3)$$

The CTF represent the sum of the aberrations emerging in an electron microscope, which can influence both the phase and the amplitude. Thus the CTF, which can be affected e.g. via defocussing, is the mathematical representation of the (changes in) contrast transfer induced by aberrations influencing the phase:

$$\text{CTF} \approx -2 \cdot \sin\left[\pi\left(\Delta z \cdot \lambda \cdot u^2 - \frac{C_s \cdot \lambda^3 \cdot u^4}{2}\right)\right], \quad (2.4)$$

where λ is the wavelength of the electron, Δz the defocus, C_s the spherical aberration coefficient of the objective lens, and u the spacial frequency [102]. When increasing the defocus, the amplitude modulation induced by the phase shift is enhanced and so is the contrast for low frequencies. However, this results in negative or even no contrast for higher frequencies at the same time. Additionally, the image defects get intensified and thus correction for the CTF becomes inevitable (see section 2.3.1.5).

2.3.1.3 Image Detection

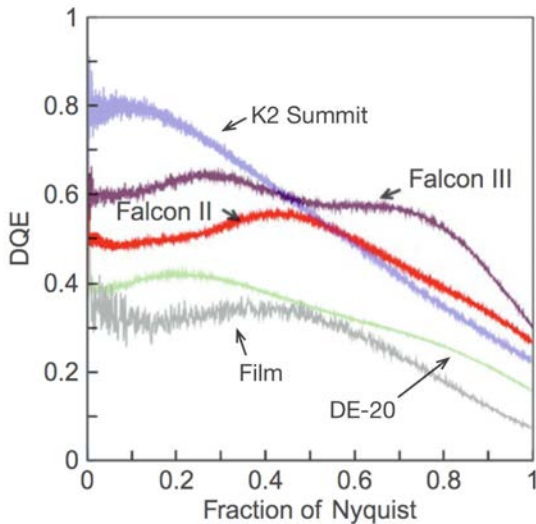


Fig. 2.2: Experimentally measured DQE as function of spacial frequency. The DQE for several cameras and classic film is shown as function of the spacial frequency, which has been normalized to Nyquist frequency (*The figure was reproduced with permission and altered according to (CC BY 3.0) from [103]*)

In the early days of electron cryo-microscopy, images of biological samples were acquired using photographic film. Here, electrons were directly detected by the reduction of silver ions on the surface of the film. Although film has great sensitivity and resolution, its use provokes a considerable obstruction due to film development and scanning between image acquisition and the final digital image. Additionally, the microscope column had to be vented at least partly in order to change films, which resulted in additional delays from vacuum reestablishment. Therefore, the development of CCD detectors decreased the amount of time necessary for acquiring large datasets dramatically [104]. Unfortunately, the early CCD cameras had significantly worse sensitivity compared to film and induced additional noise because of their way of detecting electrons: the top layer of the chip is composed of a scintillator, usually made of phosphor, which transduces the incident electrons into photons. These photons then penetrate the second layer of the detector (the actual CCD layer), where they induce an electrical charge that is detected. These steps of signal conversion come with a considerable amount of noise and localization error, both adding to the above-mentioned PSF (see equation 2.3). The use of low-dose techniques to image biological samples already results in noisy images and thus a poor signal-to-noise ration. In addition, noise is added by using a detector and therefore the efficiency of any such detector can be described via the DQE (see figure 2.2), which is the ratio of the squared output SNR to the squared input SNR:

$$DQE(u) = \frac{SNR_{out}^2(u)}{SNR_{in}^2(u)} \quad (2.5)$$

with u as the spacial frequency, SNR_{in} the input signal-to-noise ratio, and SNR_{out} the output signal-to-noise ratio. In other words: a camera with an average DQE of 0.5 would in theory need twice as much dose to achieve the same SNR in the final image than a camera with a DQE of 1.0. In practice, it is not as simple as this, because the SNR is influenced by more than just the camera (e.g., inelastic scattering events) and thus a simple dose increase will not solve the problem of working with a low-DQE detector.

Especially high-resolution information suffers from noise-amplifying effects, that are present in a CCD detector, because the noise superimposes with fine structural details rendering their differentiation during image processing almost impossible. Fortunately, the shortcomings of the CCD detector technology could be overcome by the development of DDD, which is commonly summarized as the event of *resolution revolution* (see chapter 1.4). These new detectors were able to directly detect and even count individual electrons, which led to notably less noise in the images.

The latest detectors are capable of operating in three different modes: integration, counting, or super-resolution. In integration mode, the signal from all electrons hitting the detector is simply integrated into a single image or a stack of movie frames. Dependent

on the location on the chip hit by an electron, the signal gets distributed over several pixels and therefore create a spread of information. However, integration mode has the advantage of short exposure times and therefore high image-per-hour yields. In counting mode, each electron hitting the detector chip is counted individually and assigned to exactly one pixel. As a prerequisite, the dose rate used for image acquisition has to be significantly lower than in integration mode to prevent coincidence loss. This leads to much longer exposure times and in an overall drop of images being acquired per hour. The third mode, super-resolution, works similar to counting mode with the difference that the impact location of each electron is not only assigned to exactly one pixel but further determined to one of the four quadrants of that pixel by analyzing the spread of signal among all surrounding pixels. By that method, the resolution of the resulting image is two times higher than the physical resolution power of the camera chip, which leads to reduction of noise and thus indirectly increases the DQE of a camera operated in super-resolution mode. [105, 106]

Another way to improve the SNR of an image is the application of energy filters. As described above (see section 2.3.1.2), inelastically scattered electron mainly contribute to noise in an image. To remove them from the beam before image formation on the detector, an energy filter can be used. The FEGs of contemporary electron microscopes are capable of forming a beam with a very narrow energy spread resulting in very small electron wave length variations. Inelastically scattered electrons lose a significant amount of energy in the process. Thus, dependent on the thickness of the specimen, filters are capable of enhancing the SNR tremendously by removing electrons with an energy difference of more than 10 eV from the emergent beam. [107, 108]

2.3.1.4 Nyquist-Shannon Sampling Theorem

The physical resolution limit of an electron microscope is determined by an interplay of different factors and components. One of these components is the camera, which usually is a digital direct electron detector in modern microscopes. When using such detectors, the sampling interval of this digital device is limiting the resolution. The sample interval is determined by the size of each pixel and the set magnification of the optical system. Both these parameters can be used to calculate the so-called pixel size, which refers to the edge length of the sample area that gets projected onto each pixel. Thus, the larger the magnification, the smaller the pixel size. To distinguish two points as individuals within a digital image, at least one empty pixel has to be between them. From this, the *Nyquist frequency* can be determined, which is two times the pixel size. [109]

However, the resolution is not only limited by this Nyquist frequency: as in any optical system, Abbe's law applies [110]. It states, quite similar to the Nyquist frequency, that two

points can only be identified as individual, if they are at least two times the illuminating ray's wave length apart from each other. Fortunately, the wavelength of electrons is usually, due to the high extraction voltages of modern TEMs, significantly smaller than the Nyquist frequency and thus not limiting (see equation 2.1). The aberrations of the optical system, however, can create a much stronger resolution limitation (see section 2.3.1.2).

2.3.1.5 Image Pre-processing

The micrographs acquired during single-particle TEM show noisy projection images of the macromolecular complex in random orientation and distribution. In order to create a 3D model of that complex, a sequence of image-processing steps has to be performed. Image pre-processing includes the handling of movie files, the correction for the PSF introduced during acquisition, as well as particle image selection and extraction. In the subsequent image-processing steps, which are not considered pre-processing steps anymore, the information necessary for three-dimensional reconstruction has to be recovered: the particle center of mass (x and y translational coordinates) as well as their projection direction (the Euler angles α , β and γ). Additionally, images showing the same projection need to be identified and grouped and subsequently averaged to improve the SNR. All steps necessary to go from micrographs to a fully preprocessed stack of particle images are described in more detail in the following sections.

First, the individual frames of each movie file from the detector are aligned with each other and dose weighted using a software package such as *MotionCorr2* [75]. This is helpful to remove the motions of the specimen occurring during acquisition caused by charging or vibrations. The weighting is used to reduce the contribution of high-resolution information from later frames, which have suffered a greater amount of cumulative beam damage. Subsequently, the CTF of each micrograph is determined via cross-correlating its Fourier transform against a database of images with known CTF parameters. This step is usually carried out by software packages like *Gctf* [87]. Subsequently, the particles within each micrograph are identified, selected, and extracted. The challenge of this process is to select only intact particles, while excluding any empty spaces. For small datasets, particles can be selected manually, although this is usually very time consuming. For larger datasets, specialized software packages are available that deploy template cross-correlation, ice- and artifact identification, machine learning, or sophisticated image statistics to find and extract individual macromolecule images. [85, 111–113]

All identified particles get extracted from the micrograph as smaller, quadratic images and stacked into a single file. To improve the processability of these images and to enhance the SNR, they are Fourier filtered (see figure 2.3) to remove background gradients and high-frequency noise, normalized (usually to a mean value of 0 and a sigma of 1), and

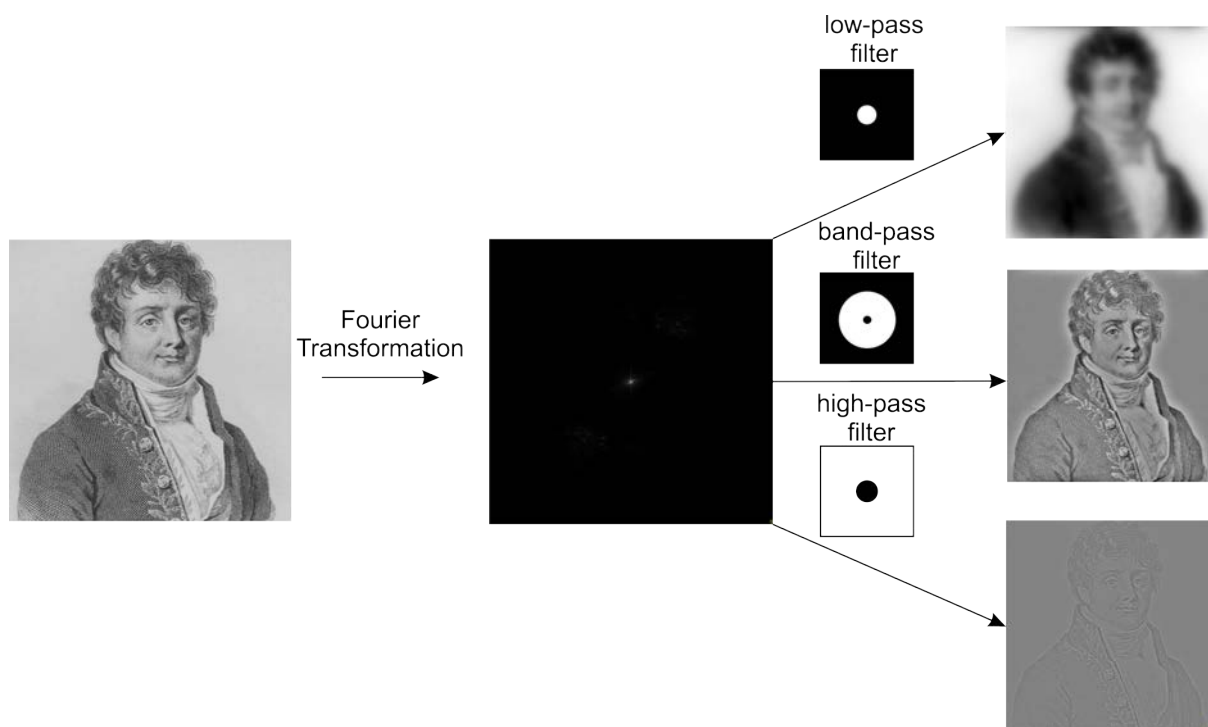


Fig. 2.3: Fourier filtering. This pictures of Jean Baptiste Joseph Fourier first becomes Fourier transformed, which changes the spacial domains into frequency domains. To meet the mathematical prerequisite for Fourier transformation (FT), which is a continuous signal, the image is tiled into all directions indefinitely before being transformed. In the FT image, the pixels further away from the center represent higher frequencies and thus carry higher resolution information. When applying the filters as shown (black areas are omitted while white areas are passed to the final image) and inverting the Fourier transform, the resulting image lacks all the high-resolution information (low-pass filtering), contains only medium-resolution information (bandpass filtering) or consists of nothing but high-resolution information (high-pass filtering), respectively (from top to bottom). (*Courtesy of David Haselbach*)

usually coarsed (reducing the number of pixels via averaging of e.g. 2×2 or 3×3 pixel patches into one pixel) during the early image processing steps.

2.3.1.6 2D Alignment and Classification

The preprocessed stack of particle images is the starting point for all further image processing. At this point, the stack of images usually contains a certain amount of false or empty images (e.g., ice crystals, broken particles, contaminations, artifacts), despite careful particle selection (as described in section 2.3.1.5). Furthermore, the images, also often called *projections*, show the macromolecule from different orientations. Each of these projections is associated with a different set of projection angles, the *Euler angles* (α , β , and γ). In addition, due to imperfect particle center identification by the algorithm, it is very likely that the particles need to be translated along their x- and y-axes to perfectly reach their center of mass. Together, a total of five degrees of freedom have to be determined for each particle, before a correct 3D reconstruction of the macromolecule is possible. This

extrapolation of parameters is performed first at the 2D and later on the 3D level, because a full dataset often contains several thermodynamically possible states or conformations for the same molecule, all of which cause additional heterogeneity.

2D Alignment Before the images can be classified into groups with similar particle views, they need to be aligned and mass centered. During exhaustive alignment, all images are superimposed after applying all permutations of possible rotation (α angle) and translation (Δx and Δy) operation within a fixed range to find the best match with a given reference. Another way to find the best match uses the cross correlation function (CCF), which is a convolution of the particle images with the mirrored reference image:

$$CCF(\vec{s}) = f(\vec{s}) \star g(-\vec{s}) = \mathcal{F}^{-1}(\mathcal{F}(f(\vec{s})) \cdot \mathcal{F}(g(-\vec{s}))) \quad (2.6)$$

Here again the Fourier transformation of the image and reference is quite helpful, because the convolution in real space becomes a multiplication in Fourier space. The resulting function will show a local maximum for coordinates corresponding to the shift with the best fit between image and reference. If several reference images are used, the calculation is performed for each reference image individually. A similar calculation can be used to determine the optimal rotation angle.

Apart from the methods just described, a variety of other approaches are described in the literature [114]. Eventually, the best combination of translation/rotation parameters as well as information about the reference image yielding the highest cross-correlation are assigned to the initial image. By these means, particles become centered and similar images are oriented in the same way. This very time consuming step has been an important topic of software development over the last couple of years. Today, most operations are performed using graphics cards, which tremendously speeds up calculations. [115, 116]

2D Classification The most important measure to improve the SNR during image-processing is averaging of similar particle images. Noise occurs stochastically, but signal has a positive amplitude. Thus, noise only gets amplified by a factor of \sqrt{n} while signal increases by n , with n as the number of averaged particle images. The larger n is, the larger the increase of the SNR. Hence, identifying similar images within the dataset, grouping them into classes, and averaging them dramatically improves the SNR. To identify such images, clustering methods are used (in the field of cryoEM often referred to as *classification*; the result of averaging similar images is usually referred to as *class average* or *class sum*). A rather simple approach would be to average all images that were assigned to the same reference image during alignment. However, this only makes sense

when a very large number of reference images was used, which is most likely not the case during early stages of image processing.

The choice of algorithm is not the only crucial step to gain feasible results from 2D alignment. Additionally, due to the usually poor SNR of the images, the choice of reference images will have a significant impact on the outcome, since the classification will be biased towards the chosen set of reference images. A variety of previous studies has shown that using too few reference images, or reference images that do not fully cover the variance present in the dataset, will lead to imperfect classification and classes with mixed content. When using reference projection images from a 3D model depicting the wrong macromolecular complex, this effect aggravates because eventually the noise in the images will correlate with the given references and thereby reproduce the references to some extent. This effect, often referred to as model bias, is one of the major pitfalls during image processing of very noisy data. Therefore, a thorough assessment of the resulting class averages by a skilled user is mandatory to identify false-positive results. [117–119]

A heavily used method for the identification and clustering of similar images is called *multivariate statistical analysis (MSA)*: each particle image is described as a vector with the dimensionality n as the total number of its pixels. When comparing two vectorized images, their Euclidean distance directly represents their similarity and is therefore easy to calculate. However, when working with a typical dataset consistent of 50,000 images and particle image dimensions of 300 x 300 (90,000 in total) pixels, the comparison and clustering of all these images would be computationally extremely expensive and negatively affected by the full noise within each image. Hence, each image should be described by only a few data points rather than its full set of pixels. This information compression can be achieved by the statistical analysis principle component analysis (PCA).

During PCA, a set of linearly uncorrelated vectors is identified and ordered by the largest to smallest variance within the dataset. These new vectors are called *Eigenvectors*. Their respective contribution to variance within the dataset is stated by their Eigenvalues. The Eigenvectors can be used to quite accurately describe each image in the dataset via a linear combination of only a few Eigenvectors representing the largest variance:

$$\vec{s}_i = a_1\vec{v}_1 + a_2\vec{v}_2 + \dots + a_n\vec{v}_n \quad (2.7)$$

with a as the linear factor. In practice, usually fewer than 50 Eigenvectors are necessary to describe each original image efficiently. Thus, only the 50 corresponding linear factors are required to describe the differences within the dataset very precisely. For clustering, a new coordinate system is used in which each image is represented by a vector composed of the linear factors after Eigenvector transformation. In addition to drastically reducing

the amount of data, most of the noise gets removed during this kind of transformation. Furthermore, the Eigenvectors have the same dimensionality as the original images and can thus be interpreted as images themselves, showing the biggest variances in the dataset when ordered accordingly. This feature is a valuable tool during EM image analysis and provides important insights (e.g., regarding the symmetry of a particle). [60, 62, 120]

Once PCA has been used for data compression, clustering can be performed. Although several algorithms exist in the field, *hierarchical clustering* and *k-means clustering* are the most heavily used. During hierarchical ascendant clustering, each image is initially assigned to its own cluster (in the field of cryoEM commonly called *class* and thus both terms will be used synonymously). Now, the two classes with the smallest distance are merged to form a new class. This procedure is repeated until an initially defined number of classes is reached. The second algorithm, k-means clustering, spreads a random set of seed points across the multi-dimensional space. Then, each image is assigned to its nearest seed point and the center of gravity is calculated for each class. These now points derived from the center of gravity are the new class seeds and each image is again assigned to its nearest point. By repeating this procedure iteratively, the center of each class will converge over time and so will the classes. While the first procedure (hierarchical clustering) is deterministic and will always reproduce the same result when started with the same set of images and parameters, the latter (k-means clustering) is a stochastic approach and mainly dependent on the initial seeds.

Maximum-Likelihood Classification An entirely different approach is the maximum-likelihood (ML) method: during this *unsupervised classification*, the reference images are created from within the dataset. At first, the user defines a suitable number of classes and each image from the dataset is assigned randomly to one class. Then, the average images for all classes are calculated and used as potential reference images for the next iteration of image alignment. During cross-correlation calculation, each image is not strictly assigned to just one reference image with the best correlation, but rather weighted by the relative matching probability. This way, each image contributes to each class, just with different weight. Furthermore, the images are heavily coarsed and filtered during the initial steps, to diminish the influence of noise and high-resolution information as well as possible. By repeating this iterative procedure of alignment and cross-correlation, and relaxing the filtering parameters, different images will end up in different classes and the homogeneity within each class will increase, leading to mostly unbiased class averages. [79, 116, 121–123]

The results of any clustering and class average calculation method will eventually be used to identify such classes which represent undesired particles (showing artifacts, broken particles, or empty images). These classes will be removed from the dataset and not used

for any further image processing. Additionally, the best 2D class sum images can be used for *ab initio* model building.

2.3.1.7 3D Model Building

One of the most difficult steps in the whole image-processing pipeline of cryoEM is the transition from 2D into 3D space. As described before (see section 2.3.1.6), the reconstitution of all five degrees of freedom for each particle image is crucial to eventually yield a high-resolution 3D reconstruction. How to extrapolate the translational parameters for x- and y-shift is also described above (see section 2.3.1.6). For the determination of the three rotational Euler angles, several methods exist of which angular reconstitution, stochastic gradient descent optimization (SGD), and projection matching are the most commonly used. While SGD and angular reconstitution do not need a 3D model as reference and are therefore capable of *ab initio* model building, projection matching requires at least a valid low resolution 3D reference showing parts of the macromolecule under investigation. An exception is the software package *SIMPLE*, which uses a stochastic neighborhood hill-climbing search combined with projection matching to reconstruct a 3D model *de novo*.

All methods are usually coupled with additional alignment and classification procedures, since any new optimization of the three Euler angles might impact the assessment of the translation parameters and vice versa. Here, only angular reconstitution and projection matching will be described in more detail since they are the only ones used for the work related to this thesis. A possible application of SGD is described elsewhere. [124]

Angular Reconstitution The mathematical groundwork for angular reconstitution is based on the *projection-slice theorem*, which was first used by Ronald Bracewell in 1956 in the field of radio astronomy. This theorem can be summarized as follows:

The projection of $f(x, y)$ in the direction θ is the one-dimensional Fourier transform of the slice through $F(u, v)$ in the corresponding direction. [125]

Derived from that, a real-space equivalent of that theorem, called *common lines theorem*, was developed in the field of cryoEM [127]. It says that when creating 1D Radon projections, or sinograms, from several different 2D projections of the same macromolecule, all these sinograms share at least one common line (see figure 2.4). These common lines can be used to determine the relative rotation angles between each of the initial 2D projections. Although computationally expensive when performed with more than three 2D projection images, and despite the sensitivity to low SNR, the method of angular reconstitution is quite elegant for *ab initio* model building.

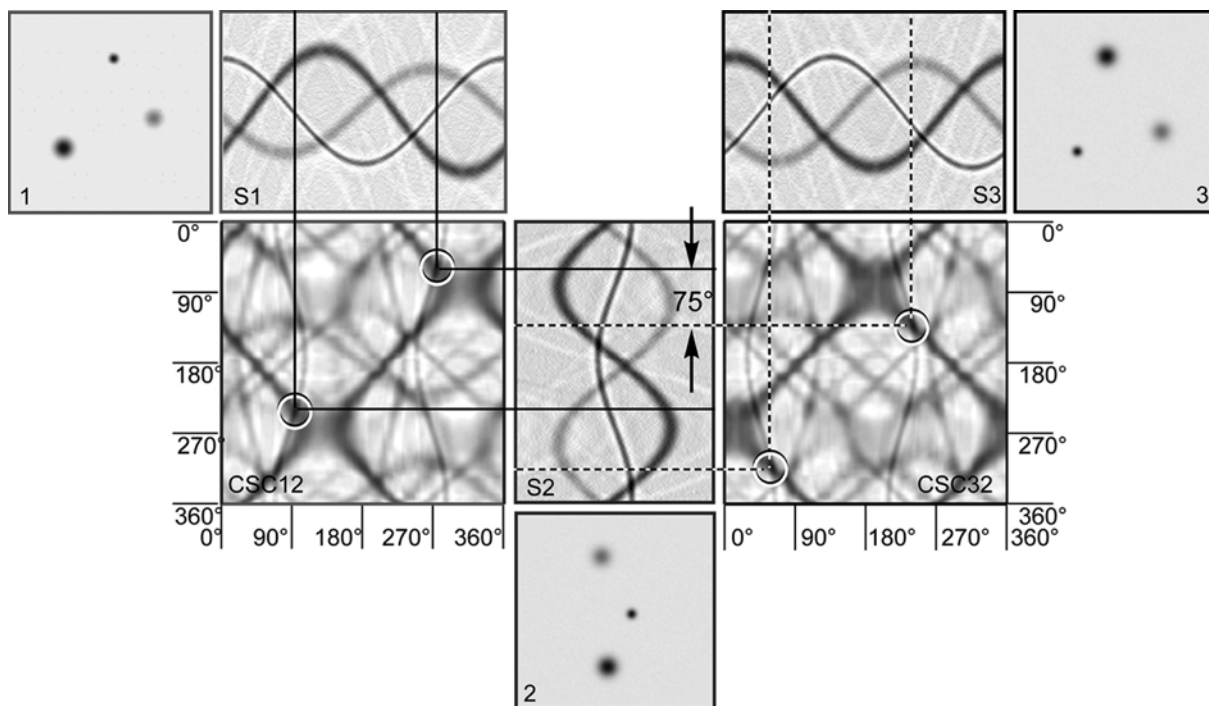


Fig. 2.4: Common line theorem. Every set of 1D projections, called a sinogram (S1, S2, S3), from 2D projection images (1, 2, 3) derived from the same three-dimensional object, share at least one common line (black lines). These common lines can be identified via a correlation map of two sinograms (CSC12, CSC23) and are used to determine the angular relationship between the two originating images. *Figure adapted from (Orlova et al. 2011) [126].*

Projection Matching The simplest way to determine the Euler angles for a given stack of particle images is projection matching. In contrast to the methods described above, a quite accurate 3D starting model is necessary for this approach. The 3D reference has to be selected very carefully, because an imperfect reference leads to very biased and equally false results (see section 2.3.1.6). The 3D reference map is projected in every direction with a suitable angular distance (called angular sampling) to be consistent in terms of resolution limit with the current filtering parameters. For symmetric particles, only the asymmetric unit of the Euler sphere is sampled. The generated 2D projections are used as reference for the alignment step and the Euler angles for all images are adopted from their best matching 2D reference.

SIMPLE PRIME The software package *SIMPLE PRIME* uses a stochastic neighborhood hill-climbing search to reconstruct a 3D model *ab initio*. It provides a quite fast and uncomplicated way of reconstructing an initial model. The starting point is a randomly reconstructed 3D map from the input images. From this first model, projections are generated and used for projection matching. The algorithm does not search the entire set of projections exhaustively, but rather stops after finding a pre-set number of correlation improvements. This procedure is repeated iteratively until convergence. However, this

approach is slightly prone to converging in a local maximum which usually leads to a implausible 3D initial model. [77]

Reconstruction The final step of each 3D model building procedure is the reconstruction of the actual 3D model. As soon as one of the above described methods has been used successfully to recover the angular relationship between all 2D projection images, the information stored within them needs to be transferred back into 3D space. Here, two most commonly used techniques are *back projection* or *Fourier space reconstruction*. For the first, back projection, all 2D images are oriented around a 3D volume of the size $N \times N \times N$ according to their Euler angles, with N as the edge length of the 2D projection images. Subsequently, their pixel gray values are projected along the image's normal vectors into the 3D volume and the gray values of each pixel are added to each intersecting voxel (the 3D version of a pixel). Wherever there is a voxel that receives contributions from many 2D projections, it will end up with a much higher gray value and thus represent density. Unfortunately, this manner of smearing information into the 3D volume will result in a lot of blurring and should be avoided by using appropriate filters. The second approach, Fourier space reconstruction, is based on the projection-slice theorem (see section 2.3.1.7) and works in a similar, but reversed manner, as angular reconstitution: all 2D projections are transformed into Fourier space and oriented according to their angles as slices of a large Fourier space 3D volume. This newly created Fourier space volume is then back-transformed into real space, now showing the 3D real space reconstruction. Although computationally much faster, this approach suffers from interpolation errors in Fourier space. [54, 128, 129]

2.3.1.8 3D Classification

In section 2.3.1.6 it was described how images are aligned and classified on the 2D level. This method is mainly used to identify and discard deficient particle images as well as to find similar images to be averaged for SNR improvement on the other hand. However, macromolecular complexes not only show differences in their 2D projections due to different orientations on the grid, but also tend to show different thermodynamically possible conformational and compositional states. The spectrum of these states can range from very small movements of individual peptide-chains to the rearrangement or dissociation of whole protein subunits or domains.

Therefore, a full dataset contains a mixture of particle images showing several different 3D states of the same macromolecule. If one would simply reconstruct a 3D model using all these images, the result would show low resolutions for the flexible areas as well as only poorly resolved densities for moving/dissociated domains. To overcome this issue,

3D classification can be used to not only classify the images according to their affiliation to a 2D reference (see section 2.3.1.7), but also with respect to one of several 3D states.

The most commonly used 3D classification is based on the maximum-likelihood classification approach, which is similar to its 2D counterpart (see section 2.3.1.6). However, at this step several 3D maps are used as reference. These 3D references can either be provided by the user, or are generated from within the dataset by randomly distributing the image particles to one class in the first iteration. All 3D references are projected to yield 2D projection images for each 3D map, which are all combined together and used as a large set of references for projection matching alignment. Eventually, the matched 2D particle images are assigned to the 2D projection reference and thus 3D reference maps by cross-correlation. Again, a single particle image is not assigned to just one reference, but is rather added to each reference class with a weighting factor according to its matching probability. In this way, a more robust classification of particle images is achieved over several iterations, since smaller differences between classes are not disregarded, which could occur if using hard assignments. [79, 116, 121–123]

2.3.1.9 3D Refinement

Most of the methods described thus far for image processing imply a certain amount of filtering or inaccuracy to either produce more robust results or to speed up their computation. However, once the members of a particular image stack are narrowed down to contain only good particle images (2D classification) which show ideally just one conformational state (3D classification) and a good 3D reference map is available, 3D refinement can be used to improve the resolution of the 3D map as much as possible. Therefore, the above described methods of projection matching and 3D reconstruction are used in an iterative manner: at first, strongly coarsened and filtered 2D images are used together with an equally strongly filtered (usually to as low as 60 Å) 3D reference, which helps to avoid model bias. Then, the 2D particle images are matched to projections of the 3D reference over several rounds during which the sampling distance for the reference projections is slowly decreased to allow for finer and more accurate Euler angle determination. Simultaneously, the low-pass filtering (see figure 2.3) intensity for the particle images is relaxed. The refinement will converge at the best possible resolution after a number of iterations, dependent on the size of the particle, the number of particle images, and their quality (see also section 2.3.1.10).

Although it is possible to perform such 3D refinements manually, several software packages allow a fully automated 3D refinement and are therefore widely used. The most prominent example is *RELION* [79], which employs ML alignment strategies combined with Bayesian estimations. It was also largely used during the preparation of this thesis. However, there

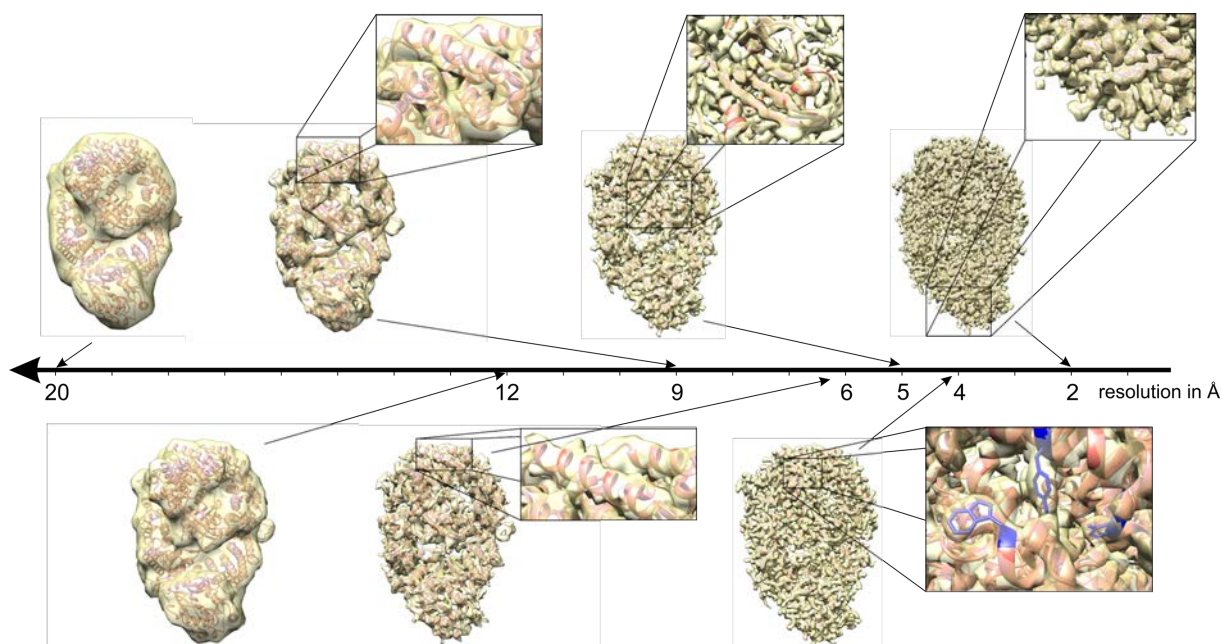


Fig. 2.5: Resolution overview. The better the apparent resolution of a given 3D reconstruction gets, the more details are recognizable. Also, dependent on the kind of recognizable details, a good resolution assessment is possible: At approx. 20 Å only an overall shape of the molecule is visible. At 12 Å first domains and sub-factors are distinguishable. Below 9 Å larger α -helices become visible and form sausage-like structures. At 5 Å resolution β -sheets and the pitch of α -helices becomes resolved. Below 4 Å bulky amino acid are visible whereas only at about 2 Å full amino acid side chains become recognizable. *Courtesy of David Haselbach*

are other software packages available, like *Frealign* [78] or *cryoSPARC* [124] (This list is not intended to be exhaustive).

2.3.1.10 Resolution Assessment and Validation

The assessment and validation of refined 3D reconstructions has been the subject of scientific discussions for a long time and remains so to this day. The resolution value achieved at the end of a refinement indicates how large (or small) the distance between two features in the specimen has to be at least to still be recognized both individually. However, giving an exact number can be difficult due to the highly noisy images from cryoEM convoluted with the microscope aberrations. Therefore, a straightforward way to determine the final resolution of a 3D reconstruction is by searching for biochemically relevant features, such as secondary structures or amino acid side chains. Each of these features is only visible after reaching a certain high-resolution barrier (see figure 2.5).

Fourier Shell Correlation A more mathematical approach is provided by the Fourier shell correlation (FSC) [128]. For this, the dataset is divided into two equal stacks, each of which is reconstructed into a 3D map. Both these structures are transformed into Fourier space and correlated, resulting in an FSC curve showing the correlation at all spatial

frequency shells. For low frequencies, the correlation will be maximal and should be close to 1. With increasing frequency the signal starts to drop and usually shows almost no correlation for Nyquist frequency (see section 2.3.1.4). To minimize (at best) over-fitting, a gold-standard refinement can be used, where the image dataset is split into two equal stacks in the very first refinement iteration [128, 130]. Henceforth, both stacks are refined individually and the FSC calculation is performed with the two resulting 3D structures. The most commonly used resolution is, for the gold-standard refinement, determined as the resolution where the FSC drops below 0.143. [131, 132]

Structure Validation To validate the correctness of a refined 3D structure, the most convenient way is to look for biochemically relevant features, like secondary structure elements (e.g., α -helices or β -sheets) or even individual amino acid side chains. The probability of recovering biochemical characteristics by chance from noise is extremely small. However, if the resolution of the refined structure is not high enough to identify said features, one could use the *cross-reference* or *omit map* method for validation. Here, a refinement is started with either a different map (cross-reference), usually from another 3D class (see section 2.3.1.8), or a smaller map (omit map), where density was arbitrarily removed, as a reference. If the refinement yields a 3D structure, which relaxes back into the original structure, the map can be considered valid. Another method to validate the correctness of a refined 3D structure is called tilt-pair validation. This method was not used during the work for this thesis, but is described elsewhere [133].

2.3.2 Sample Preparation for Electron Microscopy

2.3.2.1 Stabilization of Macromolecular Complexes (GraFix)

In order to chemically stabilize the protein complexes before using them for EM sample preparation, they were subjected to the GraFix method [134, 135]. The principle approach of GraFix is based on a sucrose or glycerol gradient containing an additional gradient of a chemical fixative. Both gradients have their highest concentration at the bottom of the centrifugation tube. This has several advantages: First of all, a heterogeneous sample applied onto such a gradient is separated according to the molecular weight (S-value) of its components, which migrate differently based on their size and shape through the gradient. Thus, a macromolecular complex that is contaminated with unbound sub-complexes can be further purified. Secondly, the gradual increase in fixative concentration leads to a mild cross-linking of the complexes while they migrate through the gradient. Eventually, the large external centrifugal forces applied onto the macromolecular complexes, caused by ultra-centrifugation, reduces their tendency to collide with each other and thus limits

the formation of intermolecular cross-links. This general GraFix protocol was modified in order to meet the individual demands of the respective protein complex.

For the preparation of the eIF3 complexes, they were first incubated at a concentration of 10 mg/mL with 0.5 mM BS3 for 30 minutes at 18 °C (batch fixation). Subsequently, 150 pmol of the fixed complex were loaded onto a SW60 linear 10–40 % (w/v) sucrose gradient, supplemented with standard eIF3 buffer, 5 mM ATP, 1 mM GTP, and 5 mM DTT. Additionally, 0.05 % (w/v) glutaraldehyde (GA) were added to the heavier 40 % sucrose fraction before mixing the gradient, leading to a linear 0–0.05 % GA gradient parallel to that of the sucrose. The gradients were ultracentrifuged for 14 hours at 174,000 x g and 4 °C. At this configuration, the eIF3 complex migrated to about 2/3 of the gradient. Once finished, the gradient was divided into 200 µL fractions and the chemical fixation reaction was quenched by adding 20 mM sodium aspartate (pH 8.0; 100 mM stock). To determine the peak fraction of the protein migration, a dotblot was performed by pipetting 1 µL from each fraction onto a small piece of nitrocellulose membrane, which was then stained with amido black solution (0.1 % (w/v) amido black powder, 25 % (v/v) isopropanol, and 10 % (v/v) acetic acid).

2.3.2.2 Negative Stain Grid Preparation

For the quick and reliable evaluation of the complex stability and homogeneity, as well as their behavior during EM grid preparation, negatively stained specimens were prepared. First, a 2 % (w/v) uranyl formate solution was freshly prepared in water. To yield a homogeneous staining solution, it was vigorously mixed for at least 2 minutes and subsequently centrifuged for 10 minutes at 15,000 x g to remove any precipitate or undissolved salt. For the actual grid preparation, a piece of very thin but continuous film of carbon was floated on the surface of the protein complex solution. The protein complexes adsorbed to the surface of the carbon film, driven by diffusion. After one minute, a holey copper EM grid was used to remove the floating carbon foil from the protein solution. Excessive volume was carefully blotted away using filter paper. Subsequently, the grid was floated with the protein adsorbed site downwards onto the previously prepared uranyl formate solution for a few seconds. Again, after removal of the grid from the staining solution, the excess volume was blotted away leaving only a thin stain film and the grid was set aside for a few minutes to air dry.

2.3.2.3 Cryo-Grid Preparation and Vitrification

In contrast to the preparation of negative stain grids, the remaining sucrose in the protein sample solution, introduced during GraFix, had to be removed entirely. Otherwise,

it would impair the formation of good contrast under cryoEM conditions, due to the increased (optical) density of a sugar solution. The sucrose removal was facilitated using a Zeba Spin Desalting Columns according to the manufacturers protocol, with the exception of not applying more than 60 μL onto a 0.5 mL column to stay below its sucrose saturation threshold (see section 3.2.2). After sucrose removal, a continuous carbon film was floated on the surface of the protein solution to adsorb the protein complexes to it, similar to negative stain grid preparation (see section 2.3.2.2). After one minute, the carbon foil was removed from the sample with a copper grid (Quantifoil) and 4 μL of water were applied onto the protein-carrying side of the grid. A Vitrobot, set to 100 % relative humidity and 4 $^{\circ}\text{C}$, was used to blot away the excessive liquid and plunge freeze the grid into liquid ethane. The high cooling rate of -180 $^{\circ}\text{C}$ liquid ethane led to vitrification of the remaining sample solution on top of the carbon grid. The frozen grids were stored in liquid nitrogen for later use in EM. [58]

2.3.3 Image Acquisition

Negative stain grids were imaged using a Philips CM200 FEG TEM. The images were taken at a defocus of approximately 0.8–1.5 μm . A twice pixel-binned TVIPS CCD camera was used to acquire digital images of 2048 x 2048 pixels at magnifications of 88,000 x. At this magnification, and with two-fold binning, a pixel size of 2.45 \AA was achieved.

Cryo grids were imaged using Titan Krios TEMs equipped either with a post-column GIF quantum energy filter and a K2 direct detection camera or a C_s -corrector and a Falcon III DDD. The dataset for native eIF3 was imaged using the K2 camera operated in counting mode at a magnification of 105,000 x (pixel size of 1.37 $\text{\AA}/\text{px}$), an exposure time of eight seconds, and a dose of 60 electrons per \AA^2 . Images of the 48S-IC were acquired on the Falcon III at 59,000 x magnification, resulting in a pixel size of 1.16 $\text{\AA}/\text{px}$. The specimen was exposed for one second and a movie file, comprising of 20 individual frames with a total dose of 48 electrons per \AA^2 , was recorded.

2.3.4 Image Processing

2.3.4.1 Frame Alignment and Dose Weighting

As a prerequisite for micrograph quality assessment, the individual frames of each movie stack were analyzed and aligned with respect to each other using the software package MotionCor2 [75]. Global alignment was run in competition to local alignment of 5 x 5 tiles and the better result was used. Stop criterion for alignment was a threshold of either ten iterations per alignment step or reaching an alignment error of 0.5 pixels or less. In case of

images taken at the GATAN K2 DDD, the gain reference was subtracted during this step. Eventually, for each stack of aligned frames, two final integration images were calculated and stored: One image with applied dose weighting for all downstream steps from particle selection and one without dose weighting required for CTF parameter determination.

2.3.4.2 CTF Correction and Micrograph Assessment

The unweighted images for each micrograph were subjected to the CTF analysis software *Gctf* [87]. Here, the equiphase average (EPA) algorithm was used to improve the signal of high-frequency information and yield better fitted CTF parameters. Subsequently, all micrographs were loaded into the *Micrograph Quality Checker* program of the *COW software suite* (unpublished) to evaluate the quality of both the micrograph itself and the fitting of CTF parameters. Micrographs showing no particles, too high or too low defocus, insufficient contrast, ice, aggregates or unsatisfactory CTF fits, were discarded via a visual inspection of all micrographs and their corresponding power spectra along with CTF curve fits.

2.3.4.3 Particle Selection and 2D Processing

Micrographs deemed acceptable for further use were subjected to the next step of image-processing: particle selection. Therefore, a combination of both software packages, *Gautomatch* [87] and the *CowPicker* as part of the COW software suite (unpublished), was used. Particles were selected without using a reference, to avoid reference induced bias. The consistency and overall quality of the outcome was again assessed by visual inspection of a subset of micrographs. All selected particles were extracted from the micrographs and stored, along with their CTF parameters, as a new stack of particle images.

Although all steps thus far were meticulously carried out, the stack of extracted particle images inevitably contained a certain number of deficient images, such as broken particles, dirt, ice crystals or empty images. To identify and remove those images, 2D image analysis was applied: All images were down sampled by a factor of 2–4 to reduce data amount significantly leading to much faster processing times and improvement of the SNR. Additionally, all images were filtered and normalized. In the next step, the reference-free 2D maximum likelihood approach was applied using the software package RELION [122]. The resulting 2D class averages were visually evaluated and those showing any of the undesired artifacts described above were removed from further analysis. The 2D classification step was performed several times iteratively to remove as much defective images as possible.

2.3.4.4 Initial Model Building

For the generation of an initial model, a variety of techniques are available and were used synergistically (see section 2.3.1.7). The easiest way to yield a starting model would be the use of the map of a structurally similar but already known molecule, downloaded from a database (such as the Electron Microscopy Data Bank (EMDB) or the PDB). To build a *de novo* starting model the angular reconstitution method was used in combination with a bootstrapping approach. Here, 10–20 2D class average images were drawn randomly from the stack (lowpass-filtered to 15 Å) and subjected to angular reconstitution and reconstruction. This was repeated 100 times and the resulting 3D models were visually inspected. Those showing consistent structural features were selected and averaged. The resulting consensus model was lowpass-filtered again to about 30 Å and used as a starting model. In parallel, a *de novo* model was reconstructed via subjecting a set of 300 manually selected 2D class averages to the software package SIMPLE (see section 2.3.1.7). Eventually, a consensus model from both methods was used for all further procedures.

2.3.4.5 3D Conformational Analysis

The initial models created as described above were used as references for conformational analysis and sorting. The 3D-maximum likelihood classification implemented in RELION was used to search for classes of structurally different 3D maps (e.g., caused by missing components or conformational variations). At first, a set of seed 3Ds was created from the references and all particle images were distributed randomly among these initial classes. Subsequently, by performing several rounds of competitive refinement of all images against these seed classes, a separation of the particle images into clusters of higher internal homogeneity was achieved (see section 2.3.1.8). This approach was also performed for local, very flexible areas of the 3D structure. As a prerequisite, the signal of all surrounding particle density had to be removed from the initial image to diminish its impact on image cross-correlation during classification (see section 2.3.4.6 and [136]).

A different approach (as described in [82]), based on 3D-PCA and Eigenvector sorting (which is similar to its 2D equivalent, see section 2.3.1.6), was used to identify more continuous dynamics within a 3D structure. At first, a large number of 3D classes was calculated either with RELION (as described above) or via iterative bootstrapping of 10–20 particle images from the entire image stack, which were reconstructed into a 3D map. During the next step, all generated 3D classes were subjected to 3D-MSA (similar to [6]). The linear factor of the first Eigenvector was extracted for all 3D maps and used to order them. The resulting list was split into smaller parts, which were used as reference to sort all particle images according to their affiliation to one sub-list of 3D maps. This resulted in particle image sub-stacks each comprising a higher degree of internal homogeneity. The

full cycle was repeated several times to increase the robustness of image classification. Eventually, by using smaller sub-stack cuts, a larger number of image stacks was created and each one was subjected to 3D refinement in RELION.

In order to calculate an energy landscape from 3D-PCA, which was carried out by Felix Lambrecht (co-member of the *Department of Structural Dynamics, Max-Planck-Institute for Biophysical Chemistry, Göttingen*), each available 3D class was projected into a coordinate system, for which the two linear factors representing the two most significant conformational changes are used as x and y coordinates and the particle number per class as z coordinate. The free energy $\Delta\Delta G$ was calculated using

$$\Delta\Delta G = k_B T \ln \frac{p_i}{p_0}, \quad (2.8)$$

with k_B as the Boltzmann constant, T being the absolute temperature, and the particle numbers p_i relative to the particle number in the most populated state p_0 . All this was done using the CowScope algorithm (part of the CowSuite software package; manuscript in preparation).

2.3.4.6 Signal subtraction

When working with large macromolecules, structural homogeneity is rarely found isotropically for the full complex. Instead, many complexes show very rigid areas without much flexibility, which easily reach high-resolution. Other parts may display more flexibility/heterogeneity, resulting in density smearing or vanishing. If these areas are small compared to the size of the whole complex, the standard classification approaches as described before (see section 2.3.1.6) may fail because the ratio between flexible and inflexible density is too small. Then, local classification can be helpful to restrict the area that needs to be explored. To focus the 3D classification entirely onto one distinct part of a structure, two important steps are necessary: first, the 3D reference used during projection matching has to be altered to show only the area of interest. This is usually achieved by applying a 3D mask onto the reference. Secondly, the excluded density information of the 3D reference also has to be subtracted from the individual original images. This is necessary because each 2D projection image always shows an integration of the density information along the imaging path. This can lead to superimposition of density information from the small, flexible area of interest, and the rest of the complex. During cross-correlation in the classification, the superimposed density from the rigid parts of the complex will diminish the variance of the small, flexible part. Signal subtraction is achieved by using the angular information and the 3D map gained during a refinement: A difference-map, showing only those parts of the refined 3D structure that shall be subtracted from the original images,

is used to generate projection images corresponding to each original images. Now, this artificially created projection image is subtracted from the original image, leaving only the density information from the focus region. Unfortunately, this procedure is easily flawed by imperfect Euler angles or noise.

2.3.4.7 3D Refinement and Resolution Determination

The refinement of final 3D structures was performed using RELION [122]. The overall resolution for the produced 3D maps was estimated using the FSC according to the gold-standard approach (see sections 2.3.1.9 and 2.3.1.10). [128, 137, 138]

2.3.4.8 Structure Assessment

3D structures were visualized using *Amira* [83] and *UCSF Chimera* [92]. Atomic models available for individual factors or sub-areas of the 3D maps were fitted into the density with UCSF Chimera. Since secondary structure features were clearly visible for many parts of the 3D map, no further validation was performed (see section 2.3.1.10).

3 Results

3.1 Purification of Native eIF3 for cryoEM

The initiation of eukaryotic translation is a very important mechanism with many implications for other cellular processes, such as the homeostasis and health of the cell (see section 1.3). Many different factors are involved in this elaborate process, whereof eIF3 is the largest and apparently most versatile. It has an important role in the regulation and quality control of translation initiation, as described in section 1.3.3. Thus far, all high-resolution structures of eIF3 depict it in the context of the 40S ribosomal subunit, where only its octameric core is shown. In order to investigate the free, thirteen subunit comprising eIF3, a chromatography-free purification strategy was established in close collaboration with Dr. Stephanie Schell (co-member of the *Department of Structural Dynamics, Max-Planck-Institute for Biophysical Chemistry, Göttingen*). Its subjection to cross-linking mass-spectrometry proved the presence of all thirteen subunits. Furthermore, contamination with unspecifically bound nucleic acids, which led to insufficient contrast in cryoEM, was mitigated via an additional streptomycin treatment.

3.1.1 Chromatography-Free Isolation from HeLa Cells

The most important prerequisite for all structural investigations is a stable, homogeneous, and functional macromolecular complex. Therefore, in order to establish a new purification protocol for the eukaryotic translation initiation factor 3, several constraints were defined in advance to assure the integrity of the complex. Cytosolic extract from HeLa cells was used as source material because it provides a very large endogenous pool for native macromolecular complexes. Furthermore, the use of chromatographic steps was entirely avoided for several reasons: first of all, the substrates of chromatographic components, such as gel filtration columns, are likely to unspecifically interact with the protein complexes, leading to their dissociation. Secondly, when using the binding capabilities of columns for a specific purpose, for example in ion exchange chromatography, the high salt buffers used to elute the protein complexes from the columns, can not only disrupt the interaction between column material and complex, but also between proteins within the complex and thus damage them severely [139]. It is important to emphasize that the forces responsible for holding the components of a macromolecular complex together can

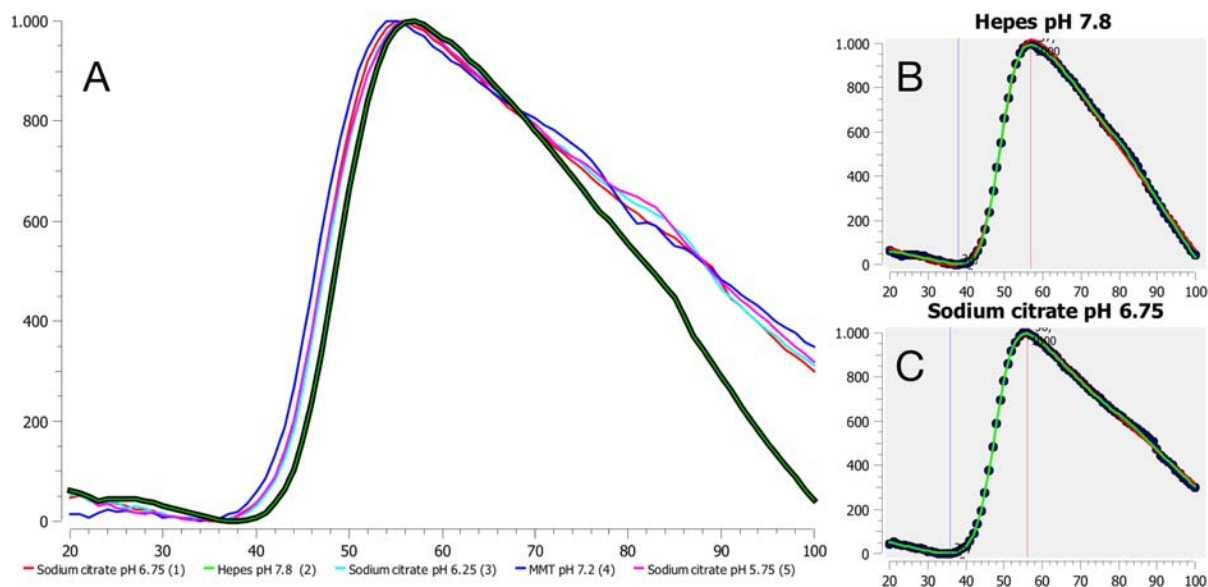


Fig. 3.1: eIF3 unfolding monitored with ProteoPlex. ProteoPlex is a straightforward method to evaluate the stability of macromolecular complexes *in vivo* [142]. Here, the human eIF3 complex was measured and the fluorescent signal, arising from interactions of a dye with the complex's hydrophobic side-chains exposed to the solvent after unfolding, from 88 different buffer substances was compared. The signal curves were fitted, normalized, and scored by a specialized algorithm. On the left (panel **A**), the five best scoring conditions are shown, of which HEPES (pH 7.8) was the best (**black/green** graph). On the right, the best condition is shown again (panel **B**), together with the runner-up result sodium citrate (pH 6.75; panel **C**).

be extremely weak. Thus, the application of any external force onto a complex might easily be strong enough to disrupt the internal connections within it and the choice of appropriate purification measures is of exceptional importance to prevent this from happening. Another important goal for the purification strategy was to prevent dissociation due to dilution, which is why the concentration was kept above 1 mg/mL for the entire purification procedure. This is based on the effect of macromolecular crowding [140], which leads to a change of the dissociation constant of a complex in favor of the assembled macromolecule rather than its disassembled components. With increased complex concentrations, the relative available amount of solvent is reduced which also promotes complex formation. A similar effect is achieved via the supplementation of a buffer with so-called crowding agents, such as PEG, which require a large hydration shell that in turn removes free solvent from the solution and thus supports complex formation. A similar stabilizing effect for macromolecules is observed for sucrose, which is regularly added to buffers for this purpose. [141]

For the identification of the optimal buffer conditions, the ProteoPlex method was used (see figure 3.1; [142]). Therein, the macromolecular complex is supplemented with different buffer substances (e.g., 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), tris(hydroxymethyl)aminomethane (TRIS), or sodium citrate), each at different pH values. Subsequently, all newly mixed conditions are heated up slowly and the complex

integrity is monitored via a fluorescent dye that interacts with the hydrophobic side-chains of the proteins that get exposed to solvent upon unfolding. The fluorescent signal from the dye, indicating complex disintegration, is evaluated with a specialized software called MaxStaf [142], which fits the signal curves and determines several parameters indicating complex stability, such as melting temperature or unfolding rate. By comparing these parameters, the most stabilizing condition can be identified. For eIF3, best complex stability was observed for HEPES buffer at pH 7.8 (see table 2.4; other results not shown). At this condition, the ProteoPlex unfolding curve for eIF3 showed ideal characteristics and lacked any indication for contaminations, unbound sub-complexes, or any other unfavorable artifacts.

The eukaryotic initiation factor 3 (eIF3) was prepared from cytosol from HeLa cytosolic extract via a series of differential precipitation steps combined with sucrose gradient sedimentation. SDS-PAGE was performed to identify the components of eIF3 (see figure 3.2). It was possible to yield approximately 10 mg of pure protein complex from 600 mL of HeLa cytoplasmic extract. A detailed description of the full purification protocol is given in section 2.2.1.1.

3.1.2 Streptomycin Treatment

Initial tests with the purified eIF3 complexes showed exceptional complex quality on negatively stained grids, but lacked significant contrast in cryoEM (data not shown). This suggested that the sample was contaminated with something diminishing the contrast. Given the native source of the complex, and taking into consideration that eIF3 is part of the translation initiation machinery, a contamination with unspecifically bound or free nucleic acids was hypothesized and found to be present via 280/260 nm absorption measurement (data not shown) [143]. To remove these free nucleic acids, an additional precipitation step was introduced into the purification strategy: before a S100 centrifugation step, the extract was additionally incubated with 1 % (w/v) streptomycin sulfate. Streptomycin, which is also commonly used as an antibiotic, binds to nucleic acids and ribosomes (many nucleic acid molecules are bound to ribosomes in the crude cytosolic extract) and precipitates them out of solution [95]. Once precipitated, the nucleic acids were pelleted during subsequent centrifugation at 100,000 x g and thus removed from the sample. 260/280 nm absorption measurements indicated a ratio of 2:98 (RNA:protein of a total of 100) after optimization, compared to 15:85 without streptomycin treatment. The introduction of this step to the purification strategy (see figure 3.2) led to much better contrast in cryoEM, which is why all further experiments were performed using eIF3 purified via the expanded protocol.

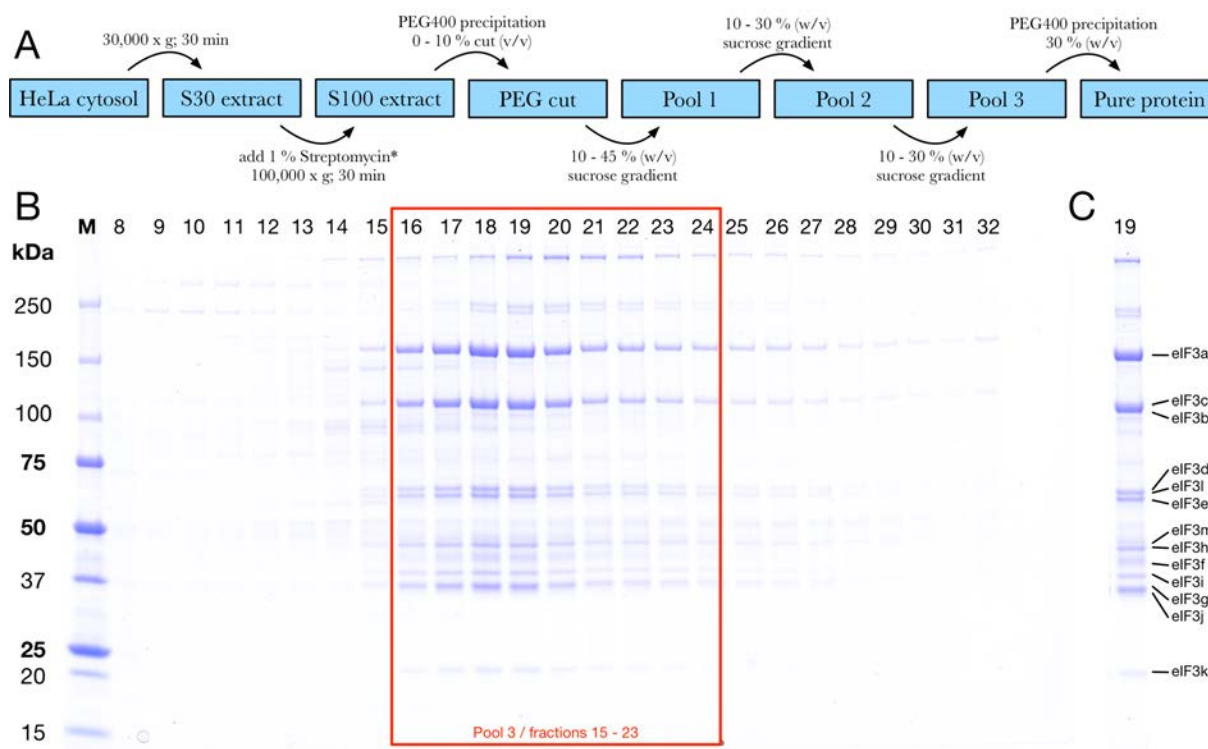


Fig. 3.2: Chromatography-free purification strategy for eIF3. For the preparation of eIF3, a fully chromatography-free purification strategy was established. It consists mainly of protein precipitation steps via the addition of PEG400 and sucrose sedimentation gradient ultracentrifugation (for details see section 2.2.1.1). **A** shows the general workflow of the purification with the most important steps. This strategy keeps the protein concentration very high at all time, reducing the risk of complex dissociation due to low concentrations (* Streptomycin treatment was introduced later, see section 3.1.2). **B** shows an SDS-PAGE of fractions 8 through 32 from the final sucrose gradient. The pooled fractions, containing eIF3, are marked in red. After pooling, the sample has to be precipitated and resuspended one last time, yielding the pure and highly concentrated eIF3 complex. **C** shows only lane 19 from the gel with labels for the individual bands/factors. Allocation is based on size and mass-spectrometry data (see table 1.1 and section 3.1.3). Additional gels from the purification procedure can be found in the supplement (figure S1).

3.1.3 Mass Spectrometry Complex Analysis

After the eIF3 complex was successfully purified, it was subjected to tandem mass spectrometry (MS) to confirm the presence of all thirteen subunits (eIF3a through m; see table 3.1). Additionally, to check the impact of the streptomycin treatment on the factor/protein composition of the purified sample, two samples (with and without streptomycin treatment; see section 3.1.2) were tested and the ratio of the identified peptides for both samples was calculated. The full set of all thirteen eIF3 subunits was identified in both lanes with an identification probability of above 99 %. Thus, the sample contains a compositionally complete eIF3 complex.

Table 3.1: Mass spectrometry analysis of native human eIF3. An SDS-PAGE was performed by loading one eIF3 sample purified without streptomycin treatment (lane 1) and one in which 1 % streptomycin was added (lane 2; see section 3.1.2). These lanes were cut out of the gel and subjected to tandem mass spectrometry analysis. The detected peptides were searched against the entire human genome. The probability is calculated via the PeptideProphet algorithm (according to [144]). All thirteen subunits of eIF3 were identified in both samples with above 99 % probability. All other eukaryotic translation initiation factors found in this experiment are listed in the supplement (see table S1, page 107). Legend: *count* = *total spectrum count*; *stpm* = *streptomycin*; *ratio* = *count(1% stpm) / count(control)*. MS-MS measurements and analysis were carried out by members of the group of Prof. Henning Urlaub (*Research Group Bioanalytical Mass Spectrometry, MPI-BPC, Göttingen*)

protein name	UniProt ID	MW [kDa]	counts (control)	counts (1% stpm)	ratio
eIF3a	EIF3A_HUMAN	167	465	420	0.90
eIF3b	EIF3B_HUMAN	92	259	225	0.87
eIF3c	EIF3C_HUMAN	105	387	335	0.87
eIF3d	EIF3D_HUMAN	64	170	148	0.87
eIF3e	EIF3E_HUMAN	52	198	178	0.90
eIF3f	B3KSH1_HUMAN	39	181	143	0.79
eIF3g	EIF3G_HUMAN	36	132	120	0.91
eIF3h	B3KS98_HUMAN	42	189	152	0.80
eIF3i	EIF3I_HUMAN	37	124	113	0.91
eIF3j	EIF3J_HUMAN	29	27	40	1.48
eIF3k	EIF3K_HUMAN	25	99	72	0.72
eIF3l	B0QY89_HUMAN	71	338	313	0.93
eIF3m	EIF3M_HUMAN	43	119	104	0.87

3.2 Structural Analysis of eIF3

The high yield of purified eIF3 complex derived from the optimized purification protocols allowed for a tandem-method approach, in which X-ray crystallization and cryoEM structural experiments were conducted in parallel. As a consequence, it was possible to obtain protein crystals from eIF3. Unfortunately, although extensive attempts were conducted to improve the crystallization conditions, none of the eIF3 crystals lead to high-resolution diffraction¹. Nevertheless, the fact that protein crystals were obtained in the first place is a clear proof for the quality of the purified complex. Experience from other macromolecular complexes had shown that if a protein has the tendency to crystallize it satisfies the basic requirements to serve as a pure sample for structural biology [139]. These conclusions were bolstered by the findings from the ProteoPlex experiments (see section 3.1.1 and figure 3.1). Thus, we solely focused on using cryoEM as a method for structure determination. Here, a reconstruction of the free eIF3 was obtained, initially with negative

¹X-ray crystallographic experiments were carried out and analyzed by Dr. Stephanie Schell (co-member of the *Department of Structural Dynamics, Max-Planck-Institute for Biophysical Chemistry, Göttingen*) and are thus not discussed in full detail in this thesis.

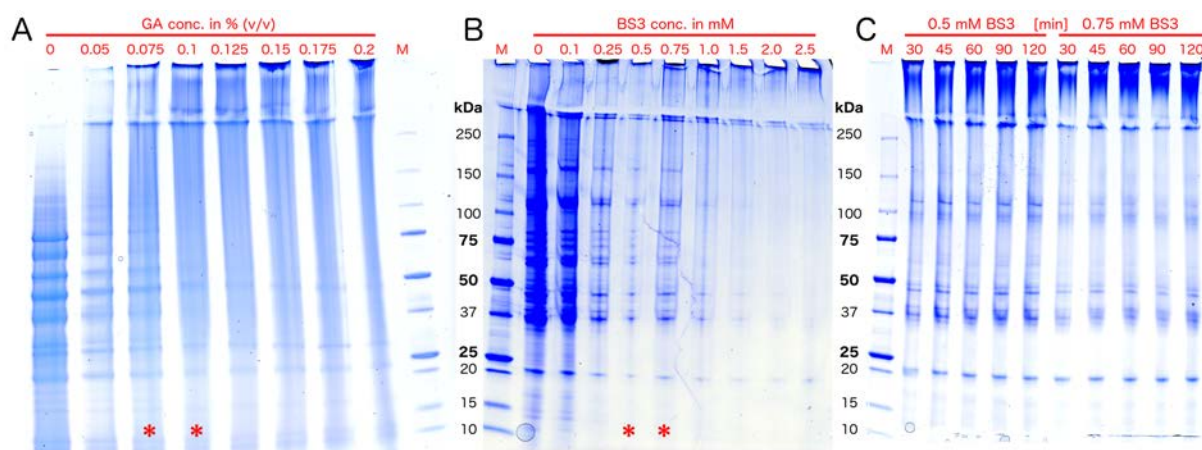


Fig. 3.3: Titration for chemical fixation of eIF3. To find the optimal conditions for chemical fixation, a titration for both GA and BS3 was performed. Fixed complexes are not able to enter the gel and will thus not show up as bands. Gel **A** shows the titration of GA. Optimal fixation is reached for 0.075–0.1 % GA (marked with asterisk). Here, weak bands are still visible but gone in the next higher concentration, indicating stoichiometric fixation. In gel **B**, a similar test is done to find the best BS3 concentration. At 0.5–0.75 mM (marked with asterisk) the last weak bands are visible. For both these concentrations, a kinetic analysis was performed, since it was intended for batch fixation use: in gel **C** the fixative activity over several timespans (30–120 min) is investigated, showing that no significant difference is detectable and therefore no significant time sensitivity is expected for the reaction. The lane containing the molecular marker is labeled with *M* in all three gels.

stain and later with cryoEM to a resolution of 7–15 Å. This structure was significantly larger than the previously published structures of the octameric eIF3 core [30]. This was in agreement with mass spectrometry data presented above, suggesting the presence of all 13 eIF3 subunits. In addition, a tremendous amount of structural flexibility and conformational heterogeneity was observed, impairing a 3D reconstruction at high resolution.

3.2.1 Chemical Fixation Optimization

Chemical fixation is an important tool to reduce the internal flexibility of macromolecular complexes and protect them from damage for example due to unspecific interactions with surfaces, low concentration dissociation, or the removal of stabilizing agents, such as sucrose (described in section 3.2.2). However, if complexes are exposed to excessively high amounts of fixing agent(s), intermolecular cross-links or larger aggregates can emerge. Therefore, a thorough assessment of the cross-linker concentration and kinetics is mandatory to find the optimal conditions for chemical fixation.

Two different substances were used for the fixation of eIF3: BS3 and GA. GA was used as part of the GraFix method (see section 2.3.2.1), while BS3 was applied during batch fixation, prior to GraFix. In contrast to GA, BS3 has significantly slower fixation kinetics, which is why it is most effective when applied to highly concentrated protein solutions rather than added to the gradient, where the protein is already diluted when reaching

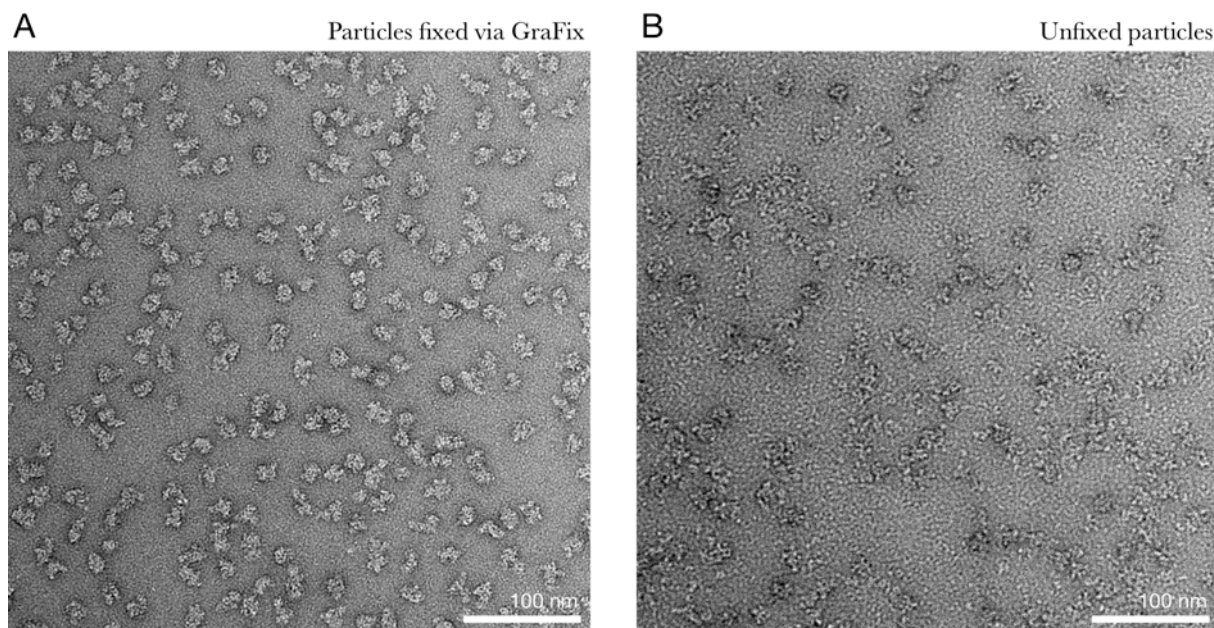


Fig. 3.4: eIF3 particles with and without fixation. In order to chemically stabilize the eIF3 complex and prevent any damage during sample preparation, the sample was cross-linked using the GraFix method (section 2.3.2.1). Here, images taken from negatively stained eIF3 particles are shown, both deriving from the same sample. The only difference during sample preparation was whether (A) or not (B) the particles were chemically fixated. It is clearly visible that the untreated particles suffered severe damage and those remaining only show poor contrast. Compared to that, the fixated sample shows a monodisperse distribution of homogeneous particles and only very few damaged molecules.

the fixative. To find the optimal concentration for both agents, a titration experiment was conducted. A series of GA and BS3 concentrations was individually added to a small amount of a 10 mg/mL eIF3 sample. After 30 minutes of incubation at room temperature, the samples were directly loaded onto an SDS-PAGE and the sample's capability to enter the gel was observed (see figure 3.3).

Best chemical fixation was observed for 0.1 % (v/v) of GA and 0.5 mM BS3 and compared to the untreated sample showing that chemical fixation has a significant impact on the sample stability (see figure 3.4). Experience from previous studies had shown that concentrations giving results after 30 min of fixation at room temperature are also appropriate under GraFix conditions (4 °C and 16+ hours; data not shown). For BS3, a simple and straightforward kinetic study was conducted, where 20 μ L of 10 mg/mL eIF3 sample were incubated with 0.5 mM and 0.75 mM of BS3 for different timespans (30, 45, 60, and 120 min), to study the kinetics of the BS3 fixation reaction. No significant changes in fixation were observed, which verifies that under the given conditions no time sensitivity is expected. Eventually, a combination of both agents was used, where the sample was pre-fixed for 30 minutes in batch with 0.5 mM BS3 and subsequently loaded onto a GraFix gradient, containing 0 — 0.1 % GA.

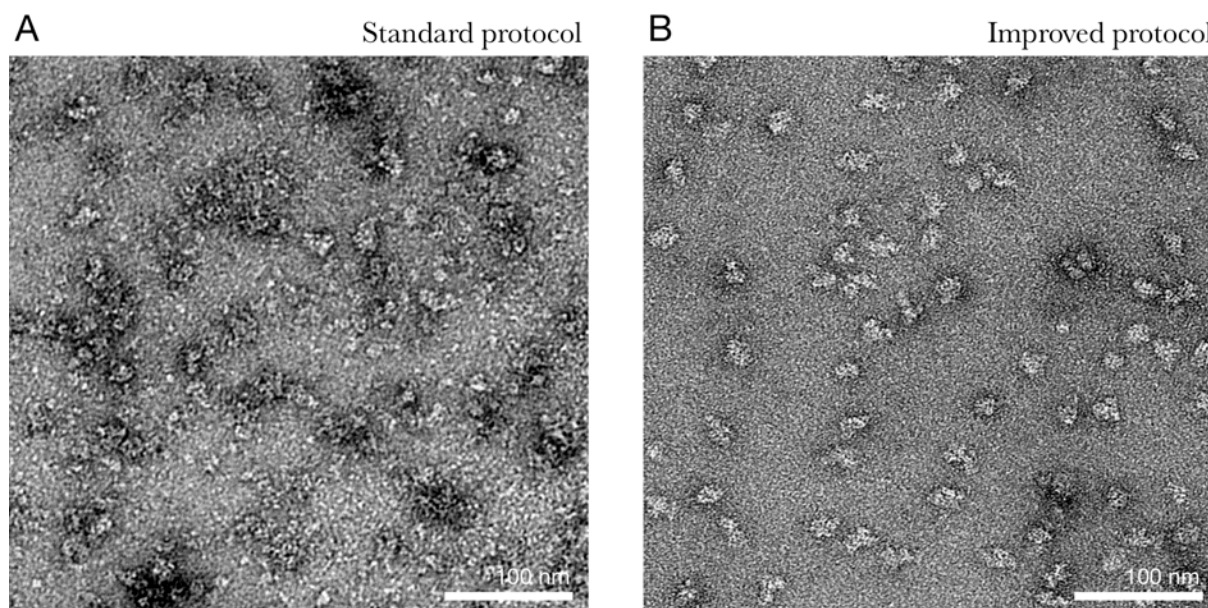


Fig. 3.5: Improved buffer-exchange procedure. To improve the buffer-exchange procedure and reduce the column's resin surface from unspecifically interacting with the protein complex, the columns were inactivated via a 0.1 mg/mL treatment with gelatin. **A** shows the sample after buffer-exchange according to the standard procedure suggested by the manufacturer. **B** shows the same sample after buffer-exchange with a similar column from the same batch, except that this column was inactivated with gelatin prior to buffer-exchange.

3.2.2 Buffer-Exchange

When chemically stabilizing and further purifying a macromolecular complex with the GraFix method (see section 2.3.2.1), the resulting sedimentation gradient fractions contain a significant amount of sucrose (or, in other cases, glycerol). Although not harmful or problematic during negative stain grid preparation and even beneficial for complex stability, sucrose in the buffer significantly impairs contrast formation in cryoEM. The specific density of the buffer converges with that of the macromolecular complex at higher sucrose concentrations and thus no contrast difference between the particles and the surrounding solvent is detectable. Therefore, the sucrose has to be removed from the sample fractions, which is facilitated by the use of size exclusion columns (as described in section 2.3.2.3).

A survey of these columns (Zeba Spin Desalting Columns, 0.5 mL) was conducted to investigate two important parameters: the amount of sucrose retained in one Zeba Spin Desalting Column and the possibility of minimizing their tendency to unspecifically interact with macromolecular complexes. The latter point has led to severe structural disruption of sample in the past and is another indicator for the potential of chromatographic columns to have a negative impact on the integrity of a complex.

3.2.2.1 Buffer-Exchange Column Inactivation

Macromolecular complexes are quite sensitive to a variety of external forces, because their constituent components are usually bound non-covalently (e.g., by hydrogen or ionic bonds) to each other. Therefore, any external force strong enough to overcome the internal forces holding a complex together, can lead to disruption, distortion, or disintegration of the assembly. As already discussed in section 3.1.1, chromatographic columns, as used here for buffer-exchange, have the capacity to impose such forces onto macromolecular complexes and should be used with caution or only under conditions where no superior method is available.

As a result of the GraFix method for protein complex fixation (see section 2.3.2.1), a significant amount of sucrose is introduced to the sample buffer. This sucrose needs to be removed prior to cryo grid preparation, because it would otherwise diminish the contrast. Therefore, the use of size-exclusion buffer-exchange columns is recommended. Unfortunately, these columns can damage or even destroy the complexes, as mentioned above. To compensate for this negative effect, a simple inactivation procedure was employed and three different conditions were tested for the designated columns: before application of the actual sample, the columns were equilibrated with buffer, followed by a treatment with 0.1 mg/mL gelatin, BSA, and a mixture of tRNAs, respectively. Afterwards, the columns were washed extensively (eight to ten equilibration cycles: applying 300 μ L buffer, as suggested by the manufacturer, followed by centrifugation) to remove any remains from the inactivation substances. Eventually, the columns were used to exchange the buffer of the actual protein sample and the complex integrity was investigated with negative stain microscopy. The most stable particles with monodisperse distribution were obtained using a column inactivated with gelatin and the increase in complex stability compared to an untreated column was found to be substantial (see figure 3.5). Furthermore, although not as comprehensive as with gelatin, the results after inactivating the columns with BSA or tRNA were still much better than using uncured columns (data not shown).

3.2.2.2 Sucrose Retention

To analyze the column's capacity to retain sucrose during buffer exchange, a series of increasing volume and sucrose concentration was applied onto a 0.5 mL Zeba spin desalting column after its equilibration. Afterwards, the (remaining) sucrose concentration was measured via the refractive index (RI) of the eluate. For calibration, the RI for a series of known sucrose concentrations was measured and the dependency of RI and sucrose concentration was determined via linear regression (see figure S2):

Table 3.2: Measuring retention potential of buffer-exchange columns. A series of samples with different volumes and containing sucrose concentration was applied onto a 0.5 mL buffer-exchange column. The RI of the eluate was measured and converted into a sucrose concentration according to equation 3.1. Samples marked in green are considered sucrose-free, orange samples are below and red samples are above the threshold of 0.5 % sucrose that is generally accepted for cryoEM samples. [134, 135, 145]

applied vol in μL		Sucrose concentration [w/v]			
		10 %	20 %	30 %	40 %
30	RI	1.3365	1.3366	1.3365	1.3363
	in %	0.0	0.1	0.0	0.0
60	RI	1.3362	1.3363	1.3365	1.3365
	in %	0.0	0.0	0.0	0.0
90	RI	1.3364	1.3367	1.3369	1.3371
	in %	0.0	0.1	0.3	0.4
120	RI	1.3369	1.3374	1.3388	1.3390
	in %	0.3	0.6	1.6	1.7

$$\Phi_{\text{sucrose}} = \frac{RI - 1.3365}{0.1432} \quad (3.1)$$

with Φ as the concentration (w/v) of sucrose in the sample and RI as the measured refractive index.

The results are shown in table 3.2 and color-coded: samples in green were considered sucrose-free, while orange samples contained a small but usually unproblematic amount of sucrose. Samples marked in red, however, exceeded the threshold of 0.5 % sucrose, which is generally regarded as just about acceptable [134, 135, 145]. The data shows that only when applying 60 μL or less onto a 0.5 mL column, full sucrose removal is guaranteed when working with the commonly used sucrose concentrations of up to 40 % sucrose. This is significantly below the amounts suggested (30–130 μL) in the standard protocol by the manufacturer.

3.2.3 Initial Model Reconstruction

The purified and chemically fixated human eIF3 complexes were subjected to negative stain electron microscopy in order to generate a 3D starting model. An exemplary micrograph is shown in figure 3.6-A, showing a monodisperse layer of particles. From this grid, 192 micrographs with a pixel size of 2.5 $\text{\AA}/\text{px}$ were acquired and initially 54,001 particles were selected. Subsequently, reference-free alignment using the COW suite was performed to remove broken or empty particle images as well as images showing contaminations. The

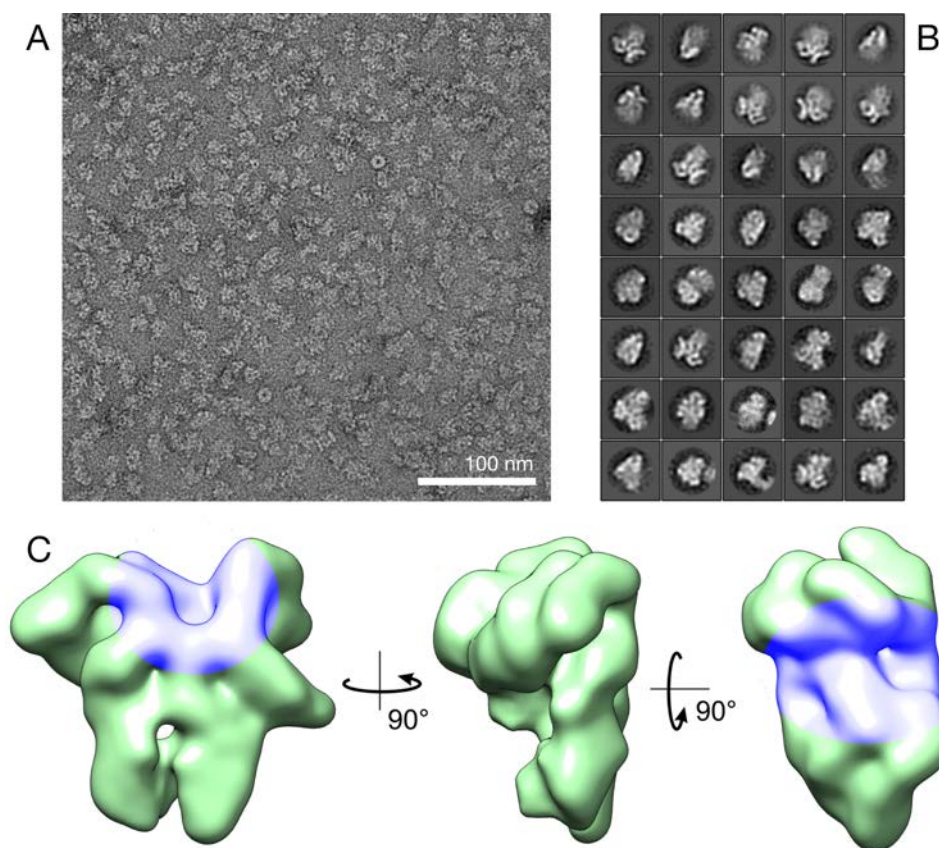


Fig. 3.6: Initial negative stain reconstruction of eIF3. In panel **A**, an exemplary negative stain micrograph is depicted, showing the monodisperse layer of eIF3 particles. **B** shows the sum images of the most populated classes calculated after performing 2D maximum-likelihood classification with RELION. **C** shows the reconstructed and refined 3D structure of h.s. eIF3, using the full dataset. The area of the horseshoe-like PCI core domain is colored in blue.

remaining 35,890 particle projection images were subjected to 2D maximum-likelihood classification using RELION. Again, 2D classes showing broken or empty images were omitted. A selection of the resulting 2D class averages is shown in figure 3.6-B.

Although the EMDB offers various 3D structures for eIF3, the creation of a *de novo* reconstruction was the preferred option. This had the two reasons that the complex at hand was purified natively from the cytosol on the one hand and was not depicted in the context of the 40S on the other. Both these factors led to the speculation that the complex structure may differ significantly from that already in the database. For the reconstruction of the *de novo* 3D model, several methods were used in parallel. Initially, the 2D class averages calculated in the COW suite were used to perform a bootstrapping approach: 100 sets, each consisting of 12 randomly chosen class average images, were subjected to angular reconstitution. The resulting 3D reconstructions were inspected manually for similarity. As a second approach, the PRIME algorithm from the software package SIMPLE was utilized. Both strategies resulted in similar 3D maps. A consensus 3D map was calculated and used as initial model for refinement against the full dataset until convergence, using RELION (see figure 3.6-C). Its overall structure was similar to

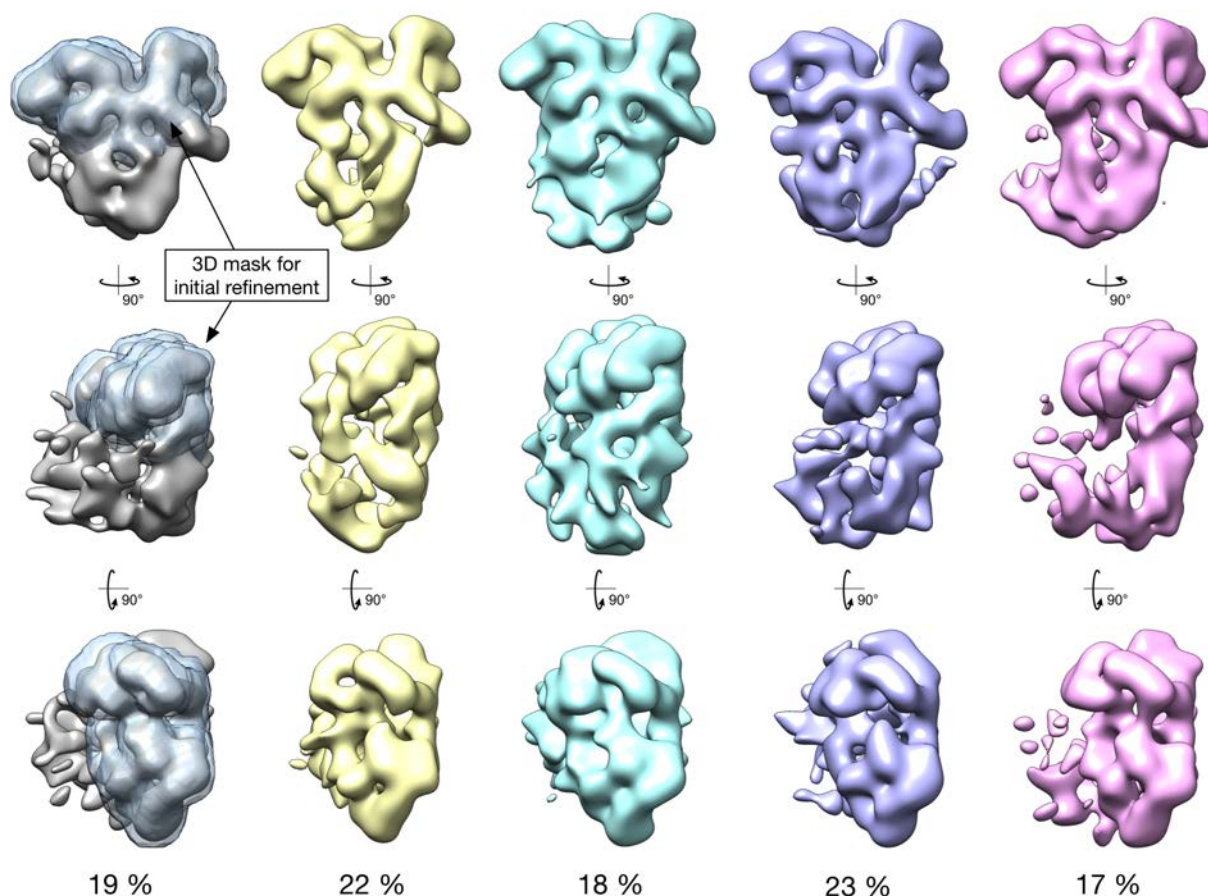


Fig. 3.7: 3D Classification of negative stain eIF3 dataset. Here, the outcome of a 3D classification for the negative stain dataset using RELION is shown. The number of classes was set to five and a two-step approach was carried out, during which at first a 3D mask (shown in blue in the first 3D class) was used to focus the refinement on the rigid area of the eIF3 octameric core and subsequently classify without additional alignment and without a mask for conformationally different states. The percentage values indicate the fraction of projection images allocated to each class after the final iteration.

the eIF3 map by *Des Georges et al.* [30], including the horseshoe like PCI domain (colored in blue) where the eIF3 subunits form a strong connection.

3.2.3.1 Heterogeneity Analysis

While comparing the 2D class averages from the negative stain projection images with the refined 3D map (for both, see figure 3.6), the orientations of the refined structure are easily recognizable. However, when closely inspecting the 2D class averages, only parts of the complex (like the PCI domain; colored in blue) have discernible features, while other areas appear blurry. These fuzzy regions can be allocated to a distinct section of the complex, which is neither well resolved in the 3D map nor shows the necessary density suggested by the 2D images. When decreasing the threshold of the 3D map in a visualization program (e.g., UCSF Chimera), unstructured density becomes visible for the area in question (not shown).

To gain a better understanding of the source for this heterogeneity, 3D classification was performed. To reduce the influence of the heterogeneous part towards the particle alignment, a two-step approach was carried out (adapted from [136]): At first, a 3D auto-refinement utilizing a 3D mask around the rigid PCI domain of the octameric core was performed until convergence was reached. Subsequently, the aligned images were used for a 3D classification without further alignment. The number of classes was set to five. The resulting 3D structures and the fraction of allocated particle projection images per class are shown in figure 3.7.

The structures obtained during 3D classification show more defined density compared to the consensus refinement performed previously (see figure 3.6). Nevertheless, none of the classes seems to represent a stable sub-species with significantly better internal homogeneity. This finding, together with the rather equal distribution of particle images to each class (ranging between 17 and 23 %), indicates an enormous amount of heterogeneity caused by continuous motion within the complex, limiting the resolution. This problem can be tackled either by (biochemically) reducing the internal motion (e.g., via the introduction of an additional binding partner that stabilizes the complex) or by computational analysis and sorting of the internal movements. However, the latter procedure is usually carried out using cryoEM data because the grain size of the negative stain would additionally limit the resolution.

3.2.4 CryoEM Structure of Human eIF3

In order to overcome the resolution limitation of negative staining and to image the eIF3 complex directly, a large cryoEM dataset was acquired. 2,905 micrograph movies, each consisting of 20 frames, were recorded at a pixel size of 1.37 Å/px using a Titan Krios equipped with a GATAN energy filter and a GATAN K2 detector operated in counting mode. The filter's slit size was set to 20 eV during acquisition to filter out inelastically scattered electrons. 410,142 particles were selected and subjected to several rounds of dataset cleaning via 2D classification, resulting in 352,222 particles showing reasonable density. The results of the final 2D classification, alongside a typical micrograph, are shown in figure 3.8.

As a consequence of the findings from the negatively stained particles, the 3D classification strategy (shown in figure 3.9) was designed to first focus on the most rigid part of the PCI domain and only after having sorted for particles showing reasonable stability and integrity there, including additional part of the complex. This was intended to restrict the amount of heterogeneity during 3D classification, most likely making it quite difficult to identify a sufficiently populated state exhibiting high resolution. Therefore, a 3D mask was used, very similar to the mask used for the negative stain data (see figure 3.6), to focus

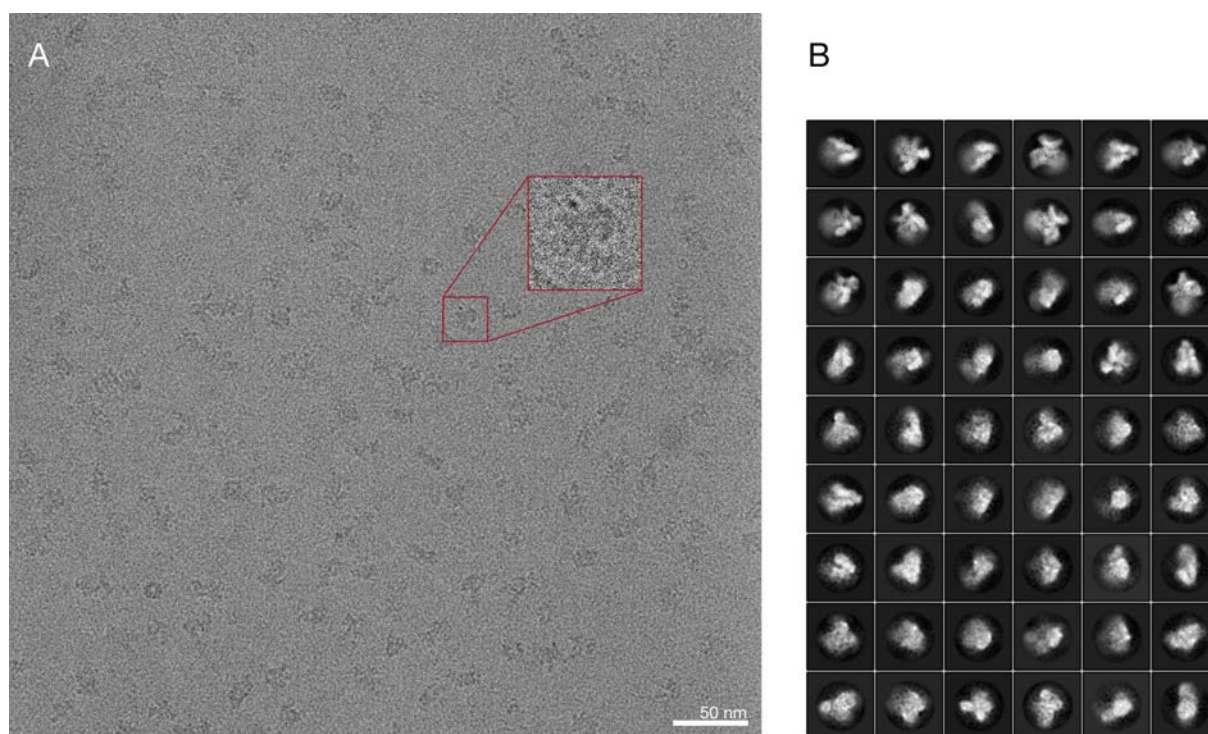


Fig. 3.8: CryoEM micrograph and 2D classes of eIF3. **A** shows a typical micrograph, taken with a GATAN K2 camera at a FEI Titan Krios. The micrographs were taken in counted mode with frame acquisition. These were subsequently aligned and dose-weighted, resulting in a summed micrograph as depicted here. The red square indicates the extraction size for later image processing together with an enlarged version of the selected particle. Panel **B** shows the results of a 2D classification after several rounds of cleaning, ordered by class population.

classification only on the aforementioned region. The negative stain map (see figure 3.6) was scaled to fit the pixel size of the cryo data, filtered to 60 Å resolution and used as reference. Two consecutive rounds of 3D classification were performed with five classes each. At the end of every classification, only those particles belonging to a 3D class showing recognizable features and structural similarity were selected. After the second round of 3D classification, only 81,200 particles still satisfied these requirements. This particle subset was subjected to 3D auto-refinement without the use of a mask, yielding a reconstruction at a resolution of 7.8 Å (following the 0.143 correlation criterion [146]; structure is depicted in green in figure 3.9-B).

3.2.4.1 Structural Heterogeneity Analysis of Native eIF3

In order to analyze the structural heterogeneity of the eIF3 complex using the cryoEM data, a second 3D classification strategy, consisting of a two-step procedure similar to the one used for the negative stain data (first refinement, followed by 3D classification without alignment; see section 3.2.3), was applied. The refined eIF3 structure obtained from cryoEM data (see section 3.2.4) was used as reference. During the first 16 iterations of refinement, no 3D mask was applied. As soon as the refinement had reached convergence,

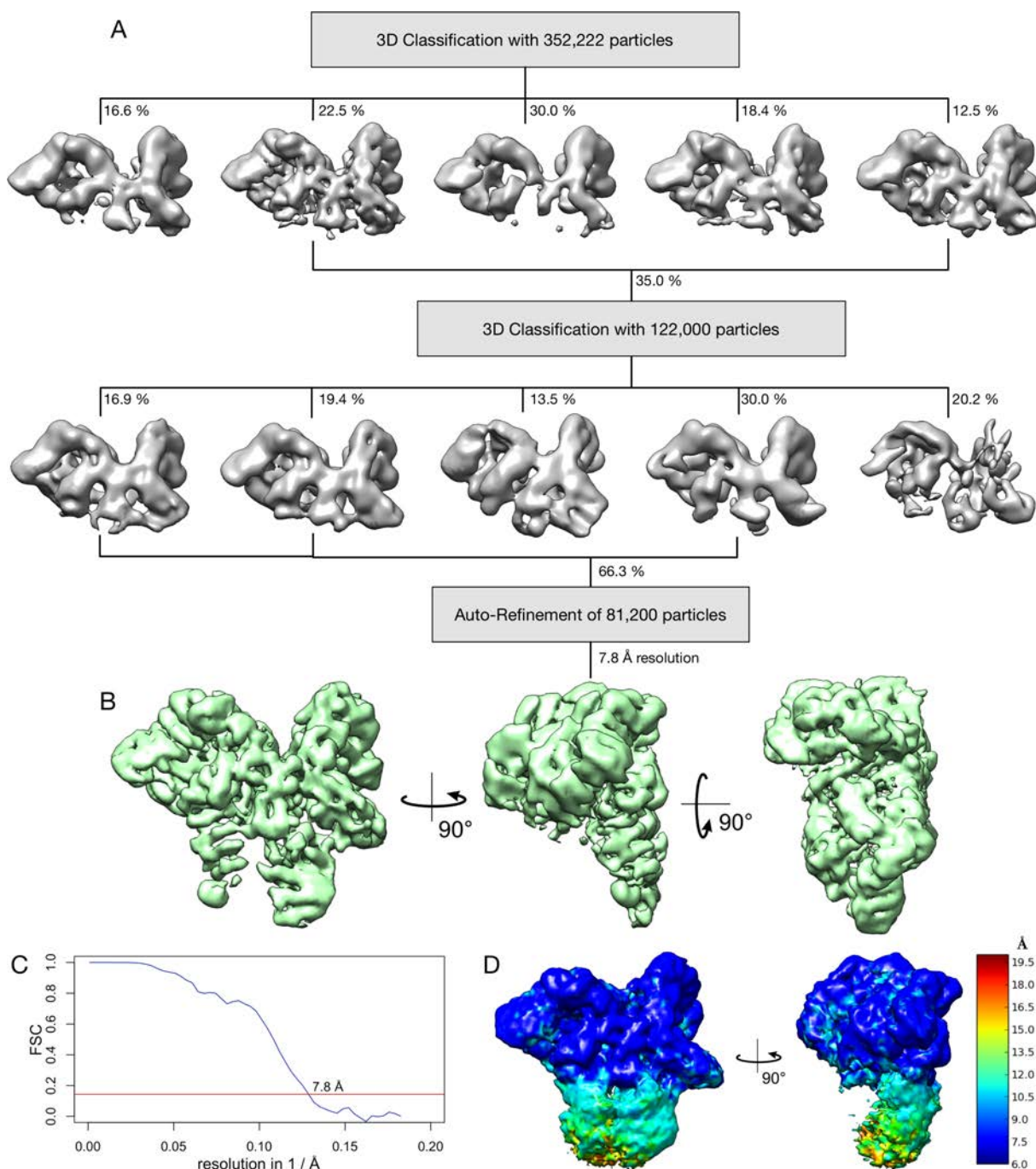


Fig. 3.9: 3D classification strategy for high-resolution eIF3 structure. In order to obtain a high-resolution structure of the human eIF3 complex from cryo data, a hierarchical sorting strategy was applied (panel **A**). In two consecutive rounds of 3D classification, only those classes showing the highest resolution and the most isotropic density throughout the structure were selected for the next round, respectively. Out of the 352,222 particles at the beginning, 81,200 were used to start an auto-refinement, eventually. This resulted in a high-resolution structure with a overall resolution of 7.8 Å (depicted in green; **B**). The gold-standard FSC analysis using the 0.143 criterion is shown in panel **C**. However, when performing a local resolution analysis with *ResMap* (panel **D**; [91]), it is clearly visible that high resolution is only achieved for the PCI domain and is significantly worse for the density farther away from the eIF3 core. The color-scale on the right indicates the resolution in Å.

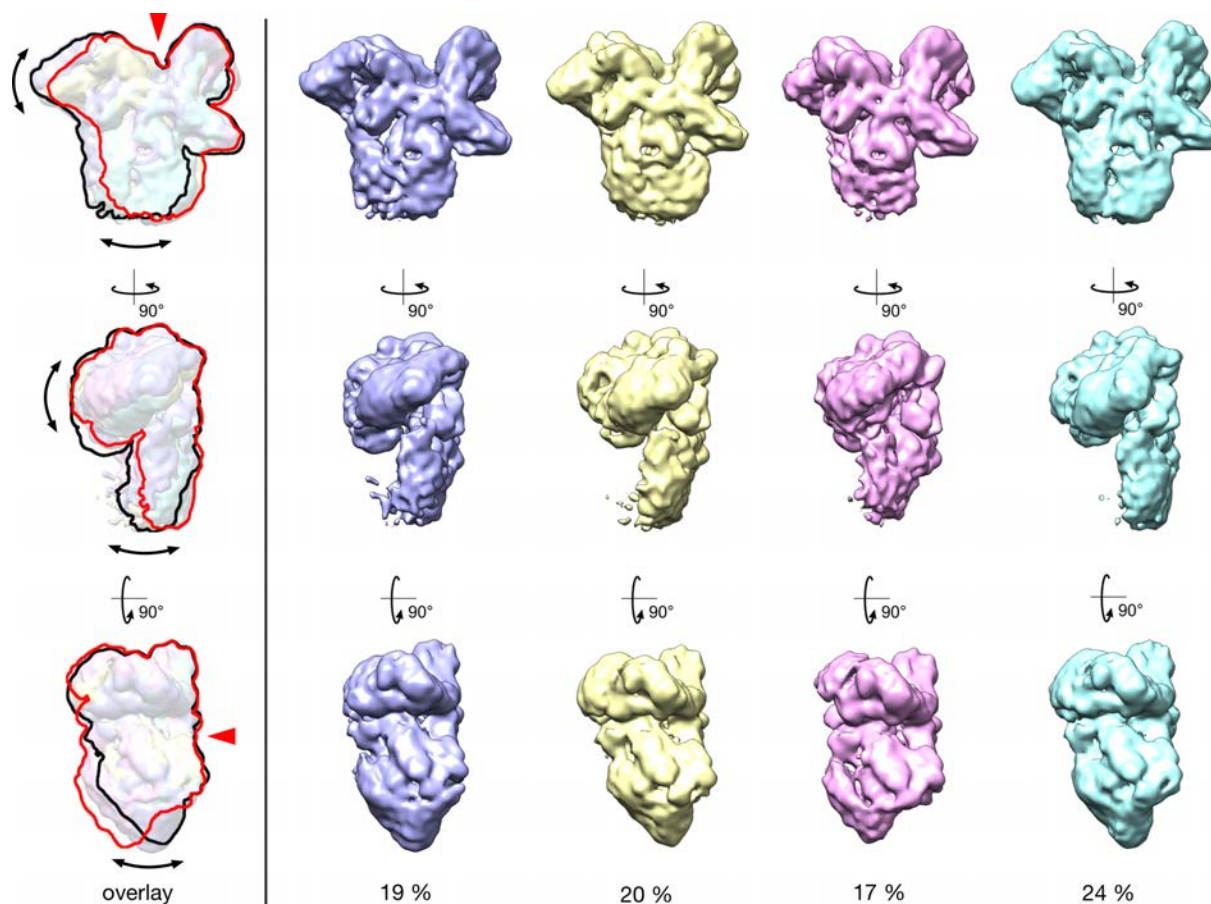


Fig. 3.10: Heterogeneity analysis of human eIF3 cryo data. Once an initial auto-refinement was performed, the alignment was disabled during a subsequent 3D classification. Four of the five resulting classes are shown here individually on the right. The left shows a schematic contour overlay, depicting the amplitudes of motion. All structures are aligned with respect to the PCI domain (indicated by **red arrow** in the overlay). The fifth class gathered mainly broken particles, which led to a poorly structured 3D class and is therefore not shown here. The percentage values given below each set of structures correspond to the amount of particles that went into each class. The black arrows with two arrowheads indicate the areas of conformational differences and direction of domain movement.

a 3D mask (similar to the one used for the negative stain data; see figure 3.7) including only the rigid parts of the PCI domain was used to carry out a focused 3D refinement. After five additional iterations (21 in total), the refinement converged.

Without further alignment, but instead using the rotation- and translation-parameters of the auto-refinement, a 3D classification with five classes was conducted. In order to allow the algorithm to especially search for compositional heterogeneity as well as conformational differences, no mask was used. Four of the resulting five classes are shown in figure 3.10. The fifth class was omitted, because it mainly consisted of broken particles not averaging in a reasonable density. Similar to the results from the negatively stained projection images, significant conformational differences between 3D classes were clearly visible and high resolution was only achieved for the PCI domain (see figure 3.9-D). In addition, the resolution dropped with increasing distance away from this rigid core.

The magnitude of heterogeneity found for the structure of the free human eIF3, especially for most parts far away from the horseshoe-shaped core (see figure 3.7), dramatically limited the reconstruction of a high-resolution structure. Attempts to sort for heterogeneity *in silico* in order to find smaller groups of particles with significantly higher internal homogeneity failed. Thus, the determination of the internal architecture of eIF3, the positions and interactions of the individual subunits, was not possible at resolutions below 7 Å. The most likely causes for the conformational heterogeneity were either an insufficient biochemical purification strategy that imposes damage to the complexes or incomplete chemical stabilization leading to its disruption during sample preparation. Both of these subjects were investigated quite extensively before starting cryoEM experiments (see sections 3.1.1 and 3.2.1) and were considered unproblematic. However, another possible reason was the lack of an additional interaction partner for some of the eIF3 subunits that, if present, would stabilize the molecules and allow better resolution. Therefore, another complex of the eukaryotic translation initiation machinery, the 48S initiation complex, of which eIF3 is an important part, was subsequently investigated.

3.3 eIF3 in Context of the 48S Translation Initiation Complex

The eukaryotic 48S translation initiation complex marks the important step, when the AUG start codon has been recognized after mRNA scanning. It is the last step before the large 60S ribosomal subunits is recruited to start elongation. If errors occur at this stage of initiation, translation can lead to faulty products, imposing threats to the health of the cell. (see section 1.3 and figure 1.2). The eukaryotic translation initiation factor 3 is an integral part of the 48S-IC and its 13 subunits form crucial interactions with all other components of the 48S complex, as previous studies have shown (see [42, 147]). Here, these additional interactions were utilized to restrict the previously observed structural heterogeneity of eIF3 to some extent. As a result, insights into its role for translation initiation were gained: a high-resolution structure of the 48S-IC was obtained, partly fitted with crystal structures of its components to yield a pseudo-atomic model. Furthermore, the dynamics of start codon recognition were analyzed and evidence regarding its mechanism are presented alongside an examination of the measures taken to facilitate 60S subunit joining. The 48S-IC complexes used for this study were purified, reconstituted, and freshly provided by Dr. Akanksha Goyal and Sung-Hui Yi (both members of the *Department of Physical Biochemistry, MPI-BPC, Göttingen*).

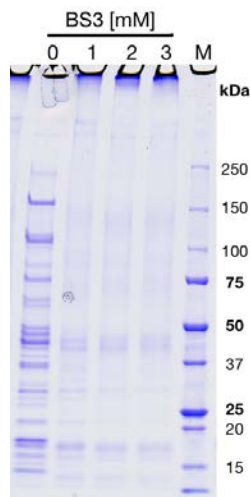


Fig. 3.11: BS3 fixation titration for 48S-IC. To identify the optimal amount of BS3 concentration for batch pre-fixation, a titration at different BS3 concentrations was performed. The fixation with 2 mM BS3, incubated for 30 min at room temperature, showed the best result. Here, a strong smearing of the bands was visible compared to the unfixed sample, while fine bands remain visible. This marks the optimal outcome, where a decent fixation is achieved without over-fixating.

3.3.1 48S-IC cryoEM Sample Preparation

The preparation of the 48S-IC for cryoEM was carried out according to the GraFix method (see section 2.3.2.1) and was very similar to the procedure used for the free eIF3 (see section 3.2.1). However, not just two but three agents were used to chemically stabilize the 48S-IC: pre-fixation with BS3, followed by in-gradient fixation with GA and *p*-maleimidophenyl isocyanate (PMPI). The latter of which was introduced newly to the protocol due to its heterobifunctional capability to cross-link RNA molecules and polypeptides to each other. This modification of the GraFix procedure was initially envisioned for the chemical fixation of spliceosomes and ribonucleoprotein particles (RNPs) which also consist of protein and RNA, similar to the 48S-IC [148]: 1 mg of PMPI was dissolved in 50 μ L of dimethyl sulfoxide (DMSO) and subsequently added to 5 mL of the lighter preparation solution (usually containing 10–15 % sucrose) of the GraFix sucrose gradient, creating an inverse PMPI concentration distribution compared to the one of GA in the final gradient.

The optimal concentrations of BS3 for the pre-fixation was once more determined by titration. Although the vast majority of 48S-IC components was purified with HEPES buffer (pH 7.5), the purification and reconstitution procedure for some factors involved the use of small amounts of buffer substances and additives that contained primary amino-groups (such as spermidine and TRIS), a viable target for the cross-linking reaction with BS3. Hence, a small amount of BS3 was quenched by these buffer components and a significantly higher BS3 concentration optimum was expected (see figure 3.11). The best BS3 fixation was achieved when incubating the 48S-IC at 2 mM BS3 for 30 min at room temperature. This was, compared to the result for eIF3, an approximately four times higher concentration. Any contaminations or aggregates that arose from this cross-reaction were removed during GraFix sucrose gradient centrifugation, since this gradient did not contain any BS3 or GA active substances.

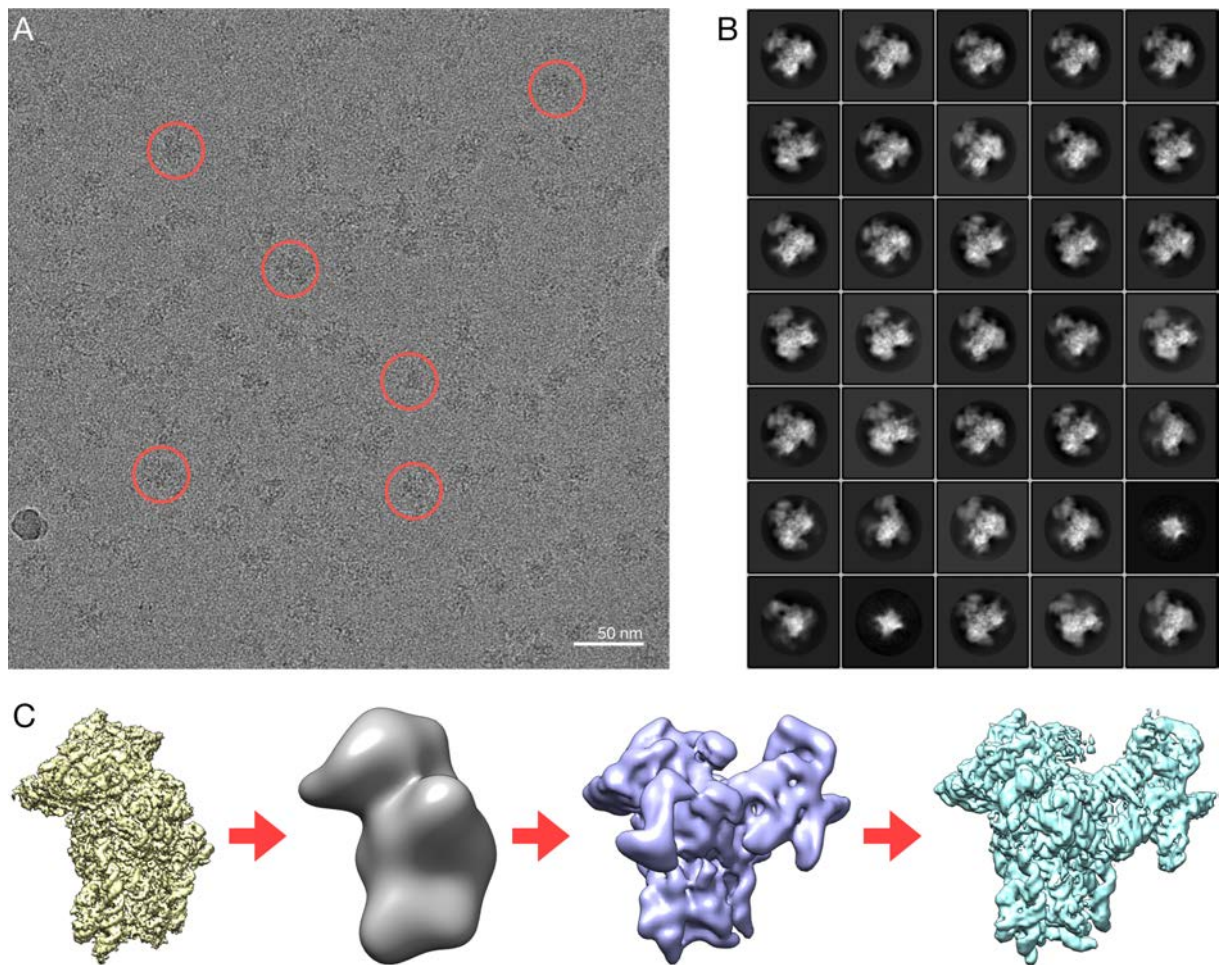


Fig. 3.12: Cryo data acquisition and 2D classification of 48S-IC. In panel **A** a typical micrograph is shown, depicting the uniformly distributed 48S-IC particles. A few exemplary particles are selected and encircled in red. In panel **B** the resulting 2D class averages after a few iterations are depicted, ordered by population of the corresponding class. Panel **C** shows the creation of a starting model for the 48S-IC. The structure of the yeast 40S ribosomal subunit, together with eIF1 and eIF1A, was downloaded from the EMDB (shown in yellow; EMDB ID: 3047; [32]). It was strongly low-pass filtered to approximately 60 Å resolution (shown in gray) and used as reference for several rounds of iterative classification. During this process, the density for the additional factors of the 48S-IC was recovered in the resulting 3D (shown in purple), eventually leading to a consensus 3D class average, showing secondary structure features (shown in cyan).

Before cryo grid preparation, buffer-exchange had to be performed to remove any sucrose from the sample. Here, the same procedure as for the free eIF3 was used, including the column inactivation protocol (see section 3.2.2.1).

3.3.2 Image Acquisition of the 48S-IC

A large cryoEM dataset was acquired for the structural analysis of the 48S-IC: 46,243 micrograph movies, each consisting of 20 frames, were recorded at a pixel size of 1.16 Å/px using a Titan Krios equipped with a C_s -corrector and a Falcon III detector in integration

mode. Subsequently, the movies were subjected to frame alignment and dose weighting using MotionCor2. CTF parameters were calculated using Gctf [87]. Manual inspection of each micrograph was performed with aid of the Micrograph Quality Checker (part of the COW suite software package), leading to a selection of 15,544 micrographs that show decent contrast, Thon rings in the power spectra beyond 4 Å, and no contaminations. An exemplary micrograph is shown in figure 3.12-.

From the micrographs, 990,486 48S-IC particles were selected and subjected to 2D classification. After several rounds of dataset cleaning, 821,651 particles showing reasonable contrast and density remained. Their 2D classes are shown in figure 3.12-B.

3.3.3 Structural Analysis of the 48S-IC

3.3.3.1 48S-IC Starting Model Reconstruction

For the creation of a 3D starting model (see figure 3.12-C), the 3D map of the yeast 40S ribosomal subunit, only complemented with eIF1 and eIF1A, was downloaded from the online EM data base (EMDB ID: 3047; [32]). It was low-pass filtered to approximately 60 Å resolution to prevent model bias and scaled to match the dataset's pixel size. Afterwards, it was subjected to several iterations of 3D classification with one class, during which only 10 iterations were performed, each. This was done to generate a new 3D map adopting the shape of a fully occupied 48S without showing too much detail. Additionally, the iterative approach reduced possible problems arising from scaling errors and missing factors in the original 3D reference.

3.3.3.2 Segmenting the 48S-IC Structure

The 3D reconstruction from the single-class 3D classification, performed for the creation of a 48S-IC starting model, was segmented using USFC Chimera (see figure 3.13) according to the information taken from previously published data of the mammalian 43S-IC [149], the yeast 48S pre-IC [150], and the yeast 40S-eIF1-eIF1A-eIF3-eIF3j complex [151]. Densities for the 40S ribosomal subunit including RACK1, the full eIF3 core domain, the TC composed of eIF2 and the Met-tRNA_i, the WD40 β-propeller of eIF3b, eIF3d, and eIF1A were visible. However, not all factors expected for a fully assembled 48S-IC were present: eIF1, the β-subunit of eIF2, eIF3 subunits g, i, and j, as well as eIF4A, eIF4B, and eIF4G were missing. In addition, some of the visible factors showed significant amounts of heterogeneity and structural flexibility, leading to density smearing and low resolutions in these regions. Therefore, the available information about conformational states of the complex

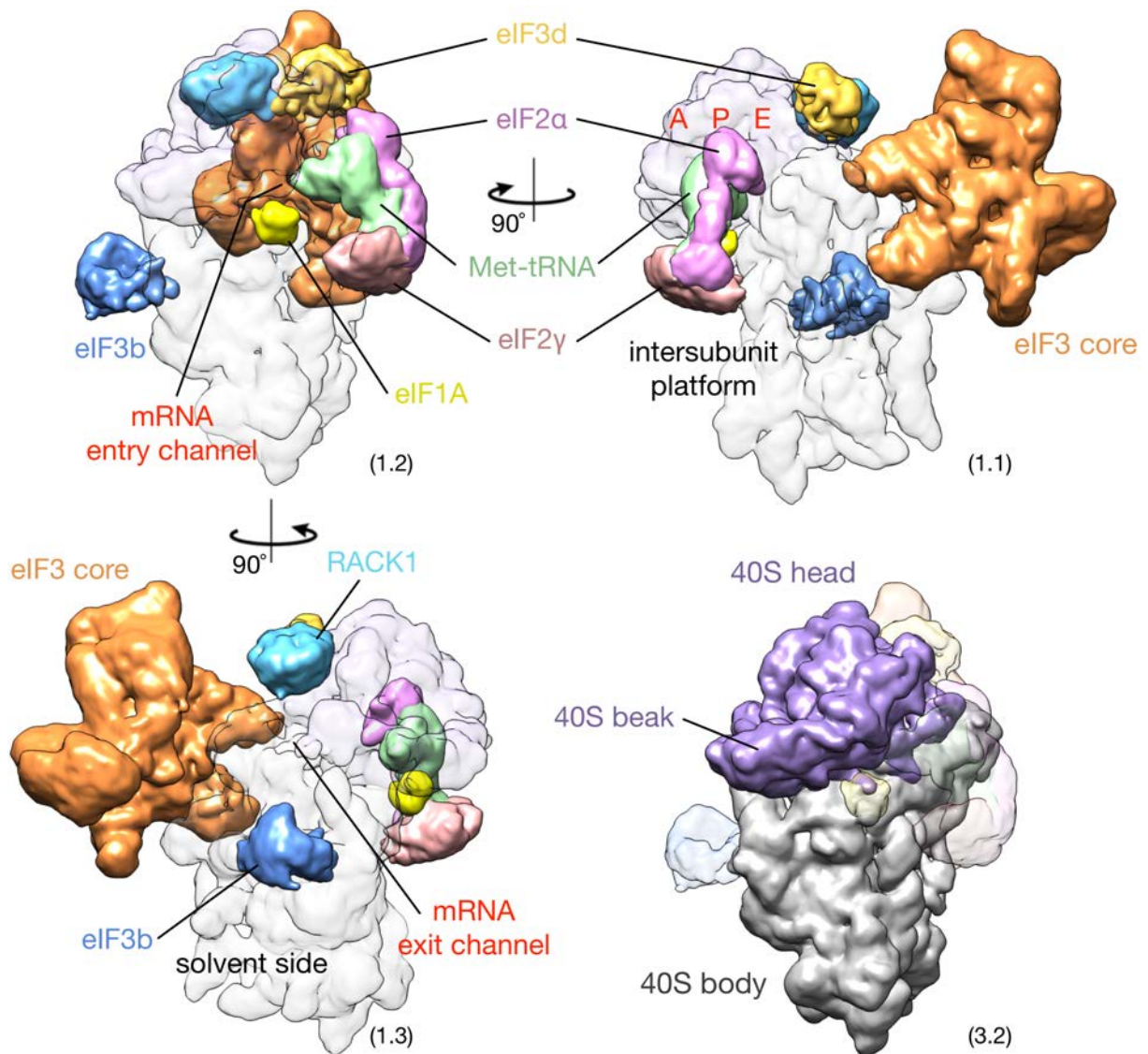


Fig. 3.13: Segmented 48S-IC structure with factor annotations. Here, the result of the final single-class 3D classification from the starting model reconstruction (see section 3.3.3.1) is shown. The structure was segmented according to its components, following the information taken from previously published data of the mammalian 43S pre-IC [149], the yeast 48S pre-IC [150], and the yeast 40S-eIF1-eIF1A-eIF3-eIF3j complex [151]. In structures **1.1**, **1.2**, and **1.3**, the 40S ribosomal subunit is shown transparent and only all other visible factors are depicted in color with annotations. Structure **1.1** shows the intersubunit platform while **1.3** depicts a 180 degree turn and thus the solvent side view. The rough positions of A-, P-, and E-site are indicated. Structure **3.2** shows the same orientations as structure **1.1**, but with inverted transparency. Here, the 40S ribosomal subunit is colored and its head, including the beak, as well as the body are shown. An extended version of the figure is available in the supplement (figure S3).

in combination with the gained data regarding its composition were used to facilitate a hierarchical classification approach.

3.3.3.3 Hierarchical Classification and Sorting

In order to separate the large cryo dataset into sub-sets according to different conformational states of the translation initiation, a hierarchical classification strategy was used (see figure 3.14): at first, an auto-refinement focusing on the body of the 40S smaller ribosomal subunit was conducted. As a result, all particles became aligned with respect to this most rigid part of the whole complex. Subsequently, a 3D classification using a 3D mask only including the 40S head but without further alignment was performed. This made it possible to sort all particles according to the open- and closed-conformation (see section 1.3.2) of the 40S, corresponding to whether or not the AUG start codon had already been identified and the codon-anticodon-interaction was established.

The particles associated with the open- and closed-conformation were each subjected to an individual round of auto-refinement to improve the overall alignment. This time, the entire 40S subunit density was used as reference. Afterwards, differences in the conformation of the main eIF3 core structure for the closed-conformation of the 48S-IC were investigated using an appropriate 3D mask during 3D classification without alignment. Those particles either lacking (parts of) the eIF3 core or showing strong structural heterogeneity were omitted. Two classes emerged from this classification, showing fully assembled eIF3 core structures including secondary structure features. Both these classes were subjected individually to auto-refinement without using a mask (shown in figure 3.15).

3.3.3.4 Refinement and High-Resolution Structure

At the end of the hierarchical classification strategy, three distinct classes of particles were identified: two classes showing the closed-conformation and one class for the open-conformation. The particles from these classes were extracted and each subjected to auto-refinement individually without using any mask. All three resulted in 3D reconstructions with nominal resolutions below 5 Å. However, only one of the two post-processed reconstructions of the closed-conformation showed reasonable structural features, such as α -helices. This 3D map also had the best resolved and less flexible eIF3 core domain. Also, an integrated refinement with images from both closed-conformation classes was performed. However, this resulted in a lower rather than higher resolved structure, which is why only the class with the high-resolution refinement result was used (see figure 3.15) and the other closed-conformation class was omitted.

Unfortunately, the peripheral areas of the 48S-IC complex, as well as flexible parts like eIF2 or most of the non-core eIF3 subunits, were showing smeared out density and thus low resolution (around 10 Å). These findings led to the speculation that the 48S-IC at

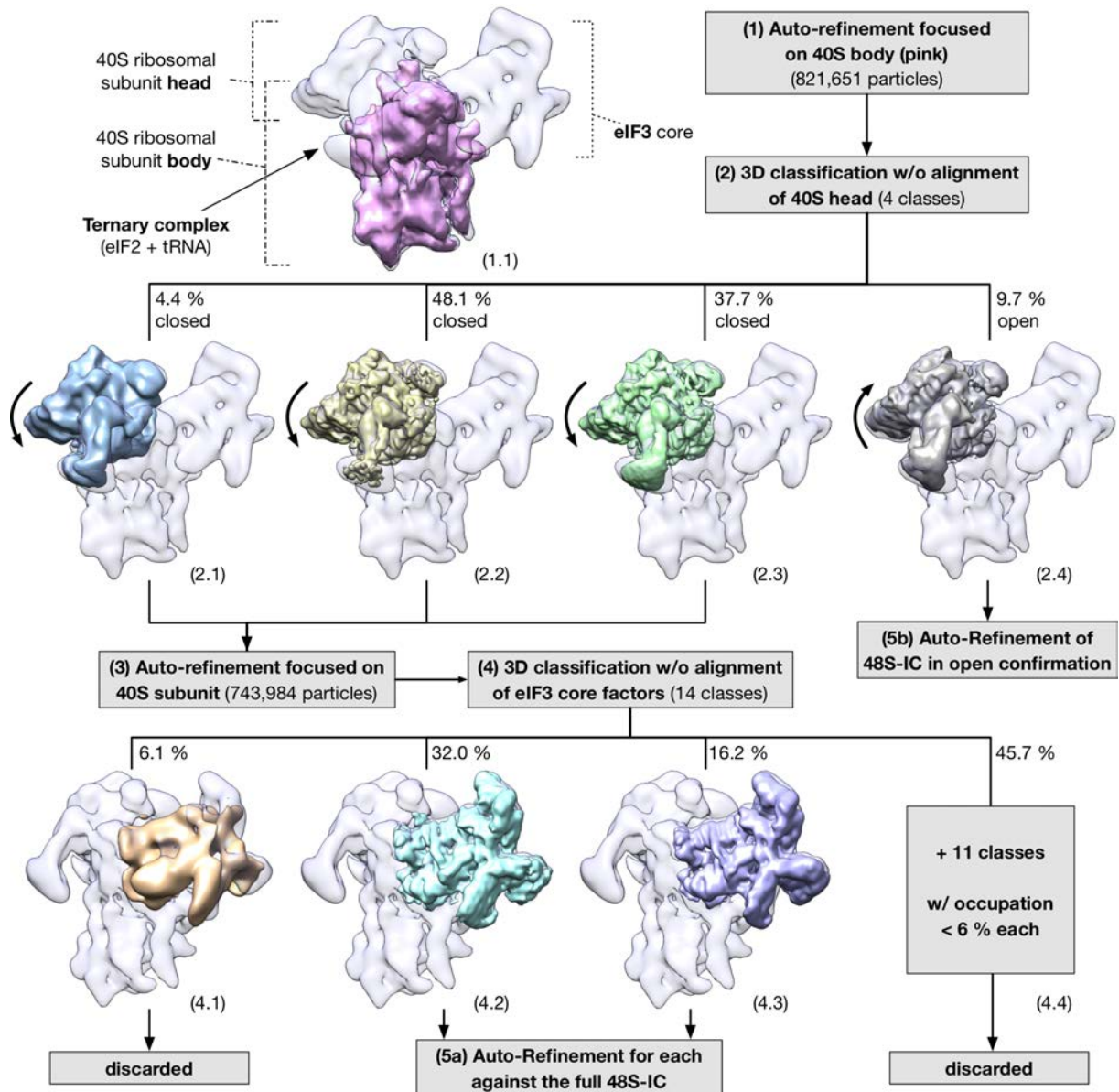


Fig. 3.14: Hierarchical sorting strategy for 48S-IC complex. (1) The full dataset, consisting of 821,651 particle projection images, was at first refined with respect to the body of the 40S ribosomal subunit (structure 1.1; 40S body colored in pink) using a corresponding 3D mask. (2) Afterwards, 3D classification with four classes and without additional alignment was performed, using a 3D mask focusing on the 40S head. The aim was to sort the dataset for the open- and closed-conformation (see section 1.3.2). Three of the four classes (2.1, 2.2, and 2.3) showed the closed-conformation (40S tilting movement indicated with arrows). One class corresponded to the open-conformation (2.4), showing a visible gap between the Met-tRNA_i anticodon and the mRNA. (3) The particles associated with all three closed-conformation classes were selected and subjected to auto-refinement, now focused on the entire 40S subunit. (4) Subsequently, another 3D classification without alignment, this time with 14 classes, was performed focused on the eIF3 core domain using an appropriate 3D mask to sort for different occupation states. Out of the 14 classes, two showed full occupation and already secondary structural features (4.2 and 4.3). The other 12 classes, mostly consisting of less than 6% of the original number of particles, were lacking parts of the eIF3 core domain (e.g. 4.1). These other classes were omitted. (5) Two separate auto-refinements were started, each consisting of the particles associated with one of the highly occupied classes. For each depicted structure, the same low-pass filtered representation of the 48S-IC consensus structure is shown as a transparent reference. This structure resulted from the final step of the initial refinement (step 1), where no 3D mask was applied.

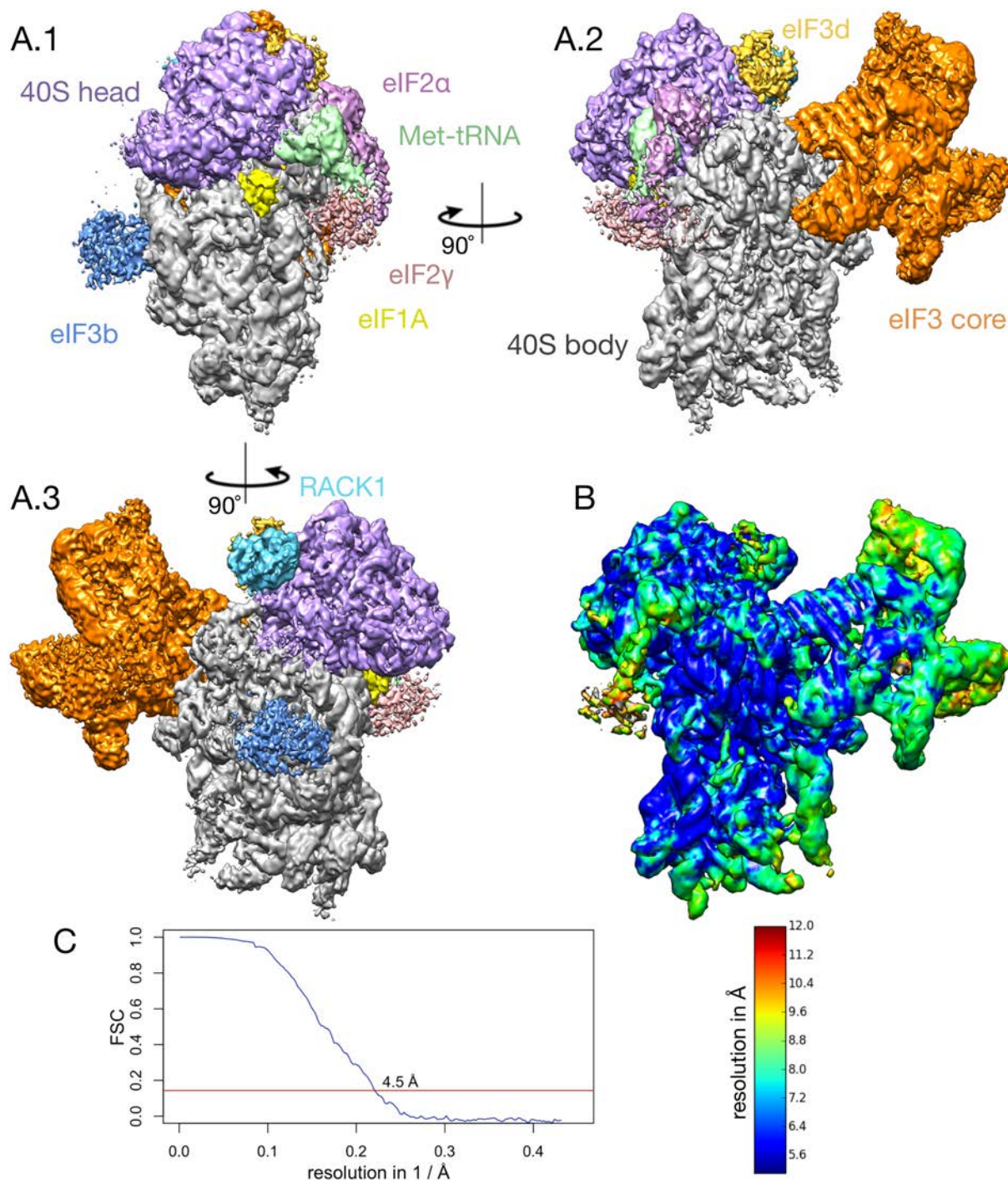


Fig. 3.15: High-resolution reconstruction of the closed 48S-IC. The auto-refined and post-processed reconstruction from all particles assigned to the best class of the 14-classes 3D classification for the closed-conformation 48S-IC is shown (A panels). This class was one of the two best classes emerging from hierarchical classification (see figure 3.14). The 40S ribosomal subunit clearly shows rRNA helices. The segments of the eIF3 core domain, directly attached to the 40S solvent side, show several α -helix packages (Huntingtin, elongation factor 3, PR/A subunit of protein phosphatase 2A and the TOR lipid kinase (HEAT)-repeats). The colored local resolution map (panel B) shows high resolution (dark blue) for the rigid parts of the 40S, while peripheral areas, such as the eIF2, eIF3d, and lateral parts of the eIF3 core domain, exhibit lower resolutions (green to yellow to red). The nominal resolution of 4.5 Å, calculated from the FSC curve (panel C) according to the FSC gold standard with the 0.143 criterion, is mainly achieved for the 40S ribosomal subunit. However, the indicated resolutions are in agreement with the visible structural features (see section 2.3.1.10).

Table 3.3: PDB atomic models used for rigid body fitting the 48S-IC. List of atomic structures used for rigid body fitting. Models from human, X-ray, and with best coverage were favored over others. (*) Structure of the 40S included all 33 ribosomal proteins (one of which was RACK1) as well as the 18S rRNA. (**) The eIF3 core structure included the eight proteins (eIF3a, c, e, f, h, k, l, m) with varying coverage (36, 57, 89, 72, 63, 87, 91, and 90; same order, all in %).

component	PDB ID	species	method	res. [Å]	coverage	reference
40S*	6ek0	human	X-ray	2.9	100 %	[152]
eIF1	4kzy	human	X-ray	7.0	100 %	[27]
eIF1A	4kzy	human	X-ray	7.0	100 %	[27]
eIF2 α	5k0y	rabbit	cryoEM	5.8	84 %	[150]
eIF2 γ	5k0y	rabbit	cryoEM	5.8	86 %	[150]
eIF3 core**	3j8b	human	cryoEM	9.3	(**)	[47]
eIF3b	4nox	fungi	X-ray	2.7	100 %	[41]
eIF3d	5k4b	fly	X-ray	1.4	66 %	[44]
eIF3g	2cq0	human	NMR	n/a	28 %	unpublished
eIF3i	4u1e	yeast	X-ray	2.0	99 %	[47]
Met-tRNA _i	6ek0	human	X-ray	2.9	100 %	[152]

hand was suffering from more conformational heterogeneity than expected and additional particle sorting would be necessary.

3.3.4 Model Building and Local Classification

3.3.4.1 Rigid Body Fitting for the 48S-IC

In order to get an understanding for the localization of the individual components and factors building up the 48S-IC, rigid body fitting for the best-resolved structure was performed. Therefore, all available atomic models of the 48S-IC components were downloaded from the PDB. If present, the human form was favored over other mammalian or eukaryotic homologs (a full list is given in table 3.3). While trying to get as fully covered models as possible, unfortunately not all factors were available at full coverage (no model for eIF2 β and eIF3j; incomplete models for eIF3 core as well as eIF3d and g). All were fitted into the 48S-IC density based on the available knowledge about the factor's localization using UCSF Chimera (see figure 3.16; for detailed eIF3 core fitting, see figure S4) and with help of the previous map segmentation (see section 3.3.3.2).

At first, the 40S ribosomal subunit, including all 33 proteins and the 18S rRNA, was fitted as one structure into the density (figure 3.16-A). Subsequently, the eight components of the eIF3 core as well as eIF3d were each fitted individually (figure 3.16-B.2; orange). Unfortunately, the available models did not cover the entire amino acid sequence, which is why parts of the density attributed to the eIF3 core could not be fitted (for details

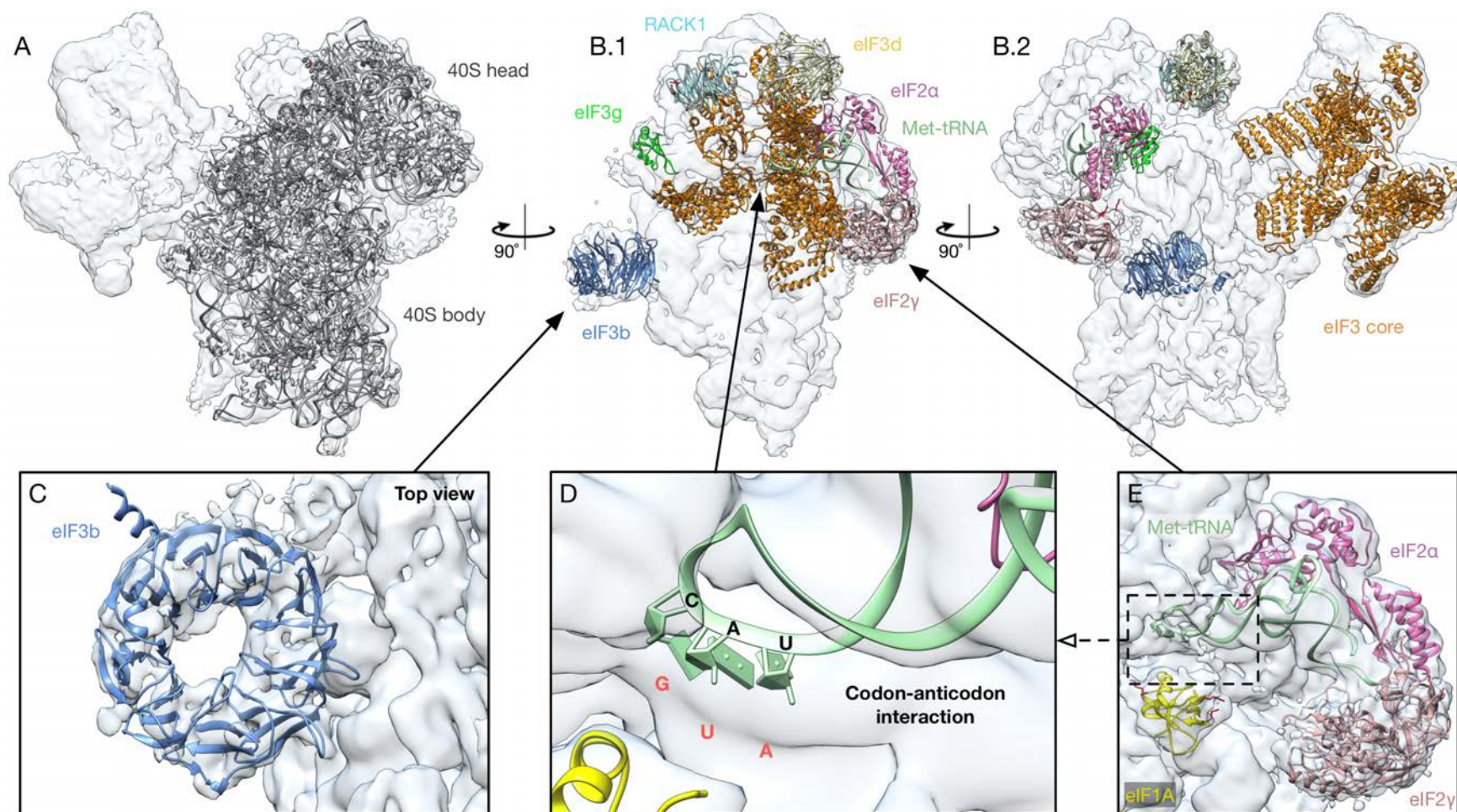


Fig. 3.16: Rigid body fitted models of 48S-IC components. Here, the full 48S-IC complex with either the 40S, panel **A** (all 33 proteins and the 18S rRNA) or, panel **B**, the factors eIF1A (yellow), eIF2 (pink/rose), eIF3 core (orange), eIF3d (pale yellow), eIF3b (blue), eIF3g (green), RACK1 (pale blue), and the tRNA (pale green) fitted into the cryoEM density is shown. Additionally, enlarged versions of **(C)** eIF3b depicted from top, **(D)** the codon-anticodon-region of the DC, and **(E)** the TC are illustrated.

about sequence coverage, see table 3.3). Also, *de novo* model building for this peripheral area was not possible with reasonable confidence due to its low resolution. In the next step, the models for the tRNA, eIF1A, eIF2 α , and eIF2 γ were fitted into their corresponding densities close to the mRNA entry channel and DC (figure 3.16-E). Since the density used for fitting corresponds to the closed-state of the 40S ribosomal subunit, the codon-anticodon interaction between Met-tRNA_{*i*} and mRNA was clearly visible in the P-site of the DC (figure 3.16-D). Eventually the WD40 β -propeller structure of eIF3b was positioned into the density as previously suggested (figure 3.16-C). It was not possible to identify unambiguous densities for the factors eIF1, eIF2 β , eIF3g, eIF3i, and eIF3j at this point. However, lowering the surface threshold for the refined complex revealed several areas of unattributed density that might correspond to the missing factors. Thus, to better localize the missing 48S-IC factors, the differences between open- and closed-conformation were analyzed in the light of the just fitted factors at first and subsequently local classification in combination with cross-linking mass spectrometry was used to improve the resolution for structurally heterogeneous areas.

3.3.4.2 Open- and Closed-conformation of the 40S Head

In addition to the closed-conformation class showing the best resolution, a second class corresponding to the open-conformation of the 40S ribosomal subunit was identified as a result of the hierarchical sorting of the full 48S-IC dataset (see sections 3.3.3.3 and 1.3.2). All factors identified thus far were fitted for both structures individually and compared (see figure 3.17). It became immediately visible how successful identification of the AUG start-codon leads to a tilt of the 40S head, together with the TC, which is attached to it, closing the gap in the decoding center to bring the ASL of the Met-tRNA_{*i*} in contact with the mRNA in the P-site. When superimposing both structures and aligning them with respect to the 40S body, the differences are evident (figure 3.17-B.2): the Met-tRNA_{*i*} was tilted approximately 12 Å downwards and the gap between tRNA and mRNA (figure 3.17-B.1) becomes closed (figure 3.17-B.3).

More importantly, close inspection of the surroundings of the decoding center in both structures (see figure 3.18) revealed additional density present in the open-conformation: next to the P-site and in between the 40S and eIF2 γ , two additional small factors became visible. As established via cross-linking mass spectrometry (described in more detail in the next section 3.3.4.3), one of the two densities belongs to eIF1 which is missing in the previously presented rigid-body fitting approach (see section 3.3.4.1). Therefore, the X-ray structure of eIF1 was fitted into the density (figure 3.18-A/B.open). Even more interesting was the second density in between eIF1 and eIF1A (indicated by a green arrow in figure 3.18-B/C.open), for which no cross-links were detected but which could

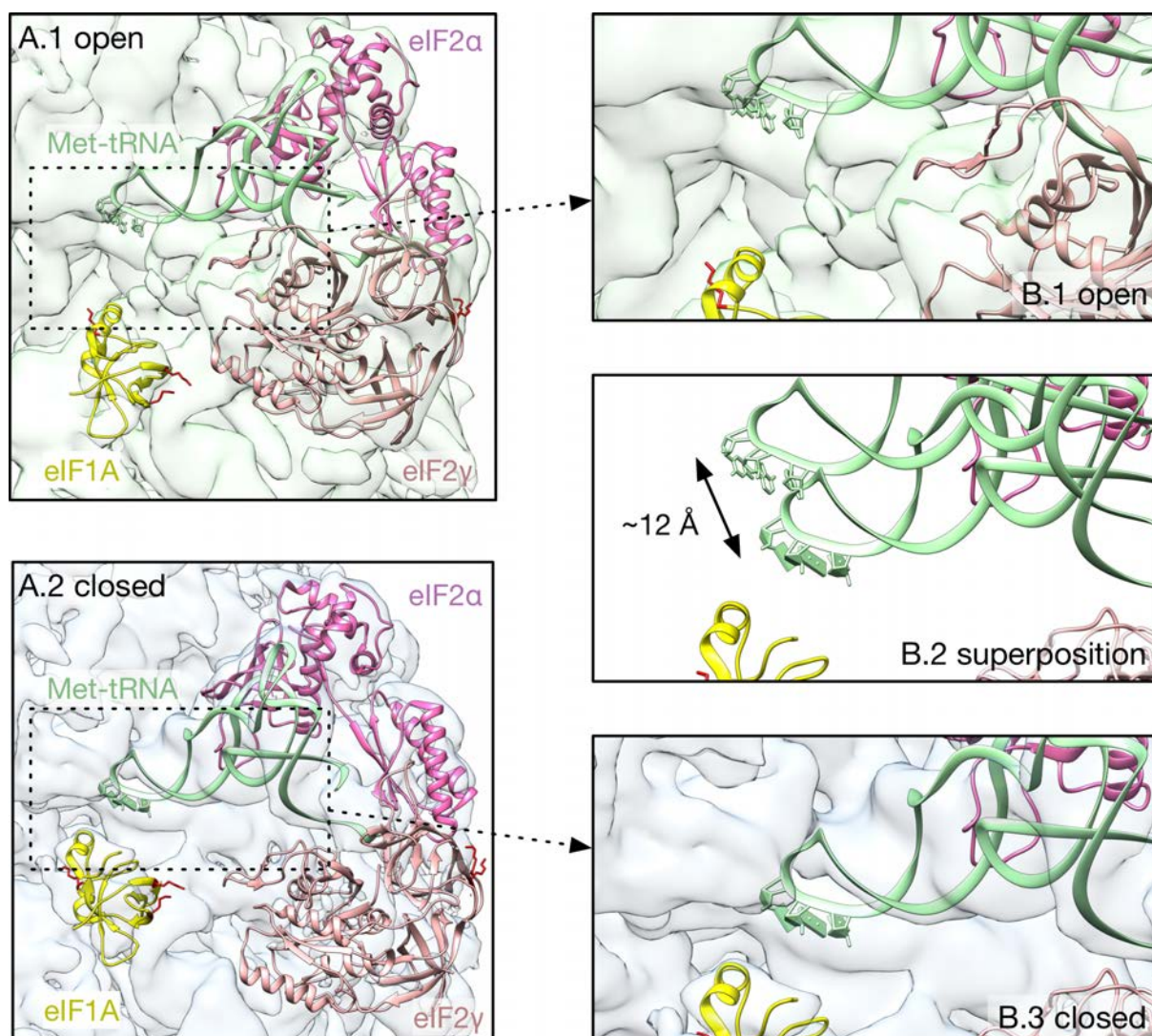


Fig. 3.17: Open- and closed-conformation of the 48S-IC. The 48S-IC can adopt an open- or closed-conformation, depending on whether or not the AUG start-codon has already been identified and locked via direct interaction with the Met-tRNA_i in the P-site of the decoding center. Up on start-codon recognition, the head of the 40S subunits performs a tilt closing the gap of the decoding center and displacing the TC downwards approximately 12 Å with respect to the 40S body. The open-conformation is depicted in panel **A.1** and **B.1**. The closed-conformation is shown in panel **A.2** and **B.3**. A superimposition of the Met-tRNA_i of both structures, aligned according to the 40S body, is depicted in panel **B.2**.

correspond to eIF2 β given its close association with the rest of the eIF2 complex and previous reports expecting it at exactly this position [14].

3.3.4.3 Cross-linking Mass Spectrometry for Factor Localization

Cross-linking mass spectrometry is a veritable tool to identify the locations of individual components of a complex with high accuracy. Here, BS3 was used as cross-linker and only homo-lysine cross-links were included in the analysis. A sophisticated interconnection network was revealed, involving most of the proteins of the 48S-IC. A full map of all found

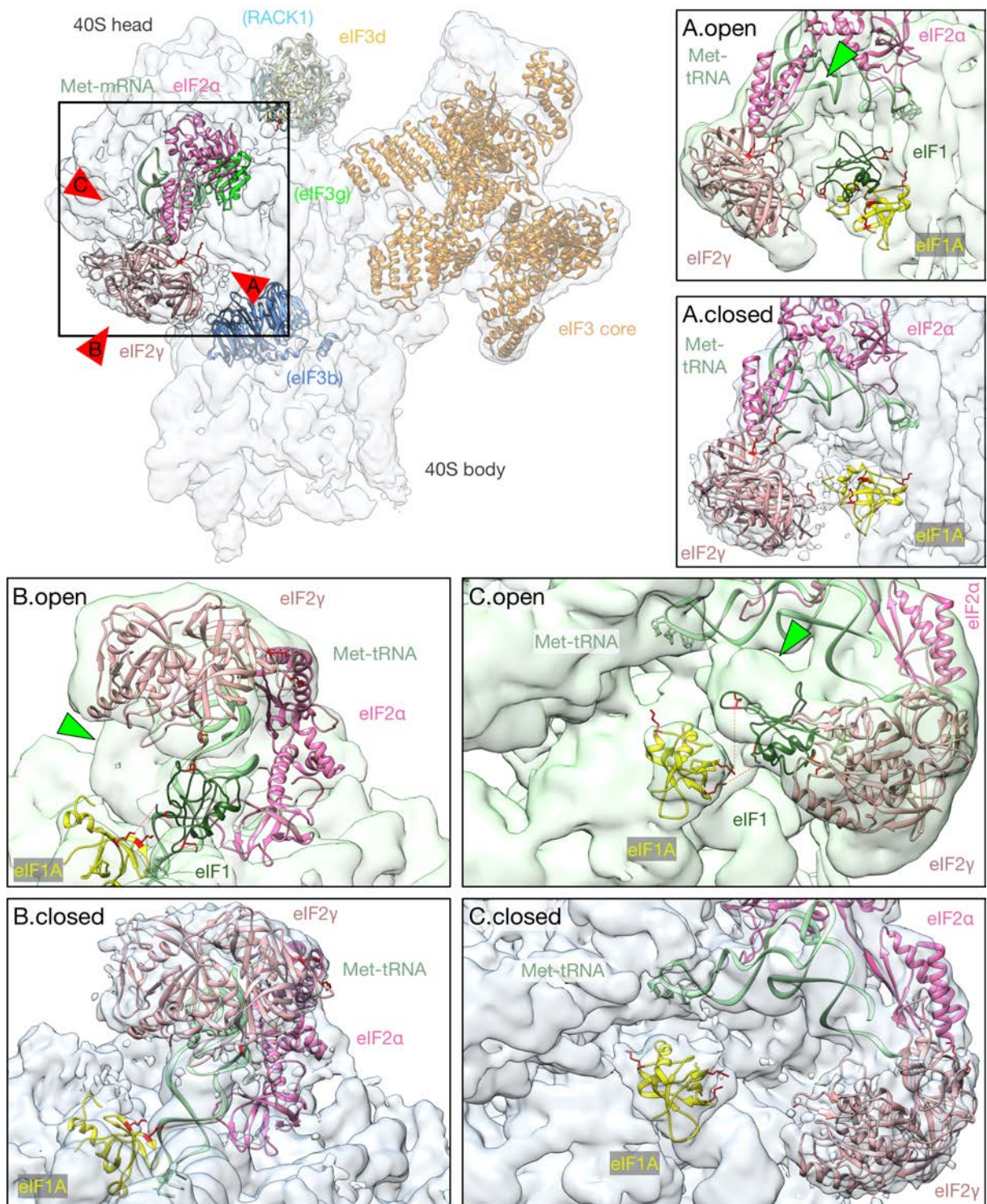


Fig. 3.18: Additional density in the open-conformation 48S-IC. While comparing the proximity of the decoding center in the open- and closed-conformation of the 48S-IC, additional density was identified in the open-conformation. With the help of cross-linking mass-spectrometry, one part of this additional density was attributed to eIF1, which is also known to bind in this particular position facilitating correct codon-anticodon-interaction. A second density was found, indicated by a **green arrows**, which might correspond to eIF2 β . However, no cross-links were found in this region, rendering this attribution highly tentative. Here, the area is depicted from three different angles as indicated by the **red arrows**. Rotation from view **A** to **C** involves a horizontal rotation of approximately 100 degree, while view **B** is from below the ternary complex after tilting the 48S-IC upwards. Labels written in parentheses indicate proteins bound to the back side of the 40S.

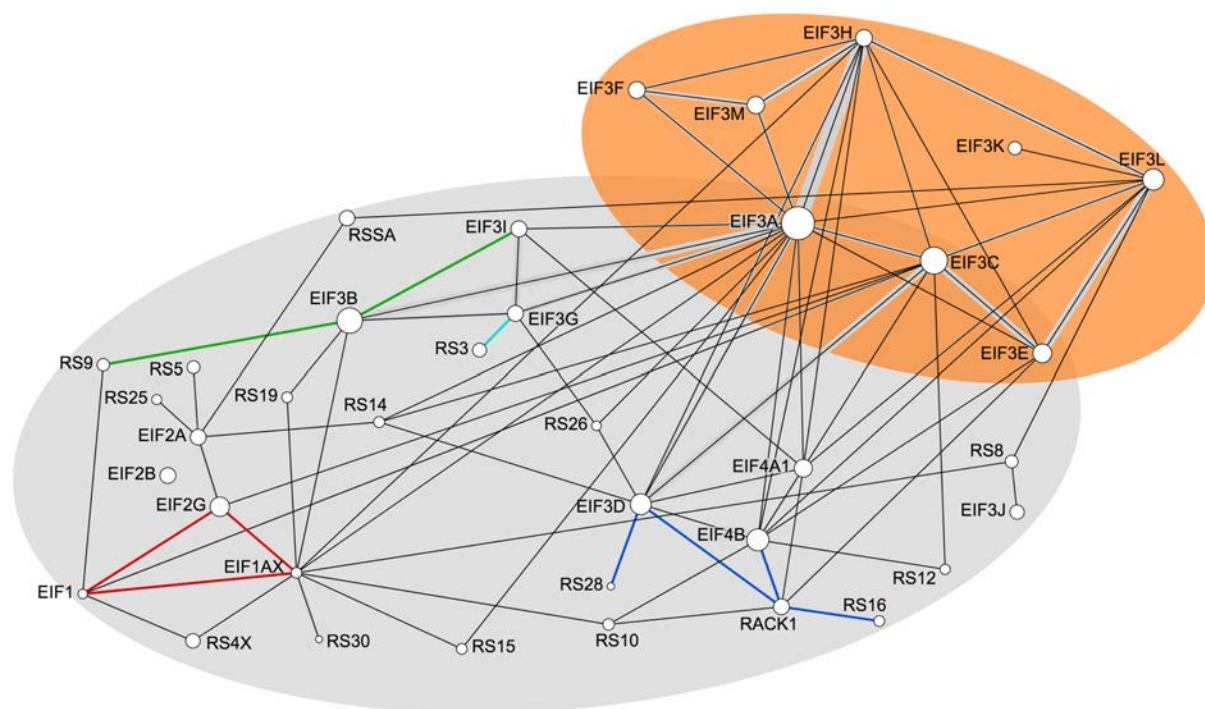


Fig. 3.19: MS-MS BS3 cross-link analysis of the 48S-IC. The network shown here is a schematic representation of the found intermolecular cross-links formed by BS3 within the 48S-IC. The proteins that belong to the eIF3 core structure (also see figure S4) are depicted with orange background while all other proteins, mainly located on the surface of the 40S ribosomal subunit, are depicted on gray background. The most important cross-links, that were particularly interesting for the localization of certain factors are highlighted in green, red, blue, or cyan, respectively (more about this in section 3.3.4.4 and figure 3.20). All cross-links between two proteins of the 40S ribosomal subunit were omitted for clarity. Each line can represent one or more cross-links and their length has no implication nor has the size of the protein-circles. *legend: RSx = 40S ribosomal protein; EIFx = eukaryotic translation initiation factor.*

lysine-lysine cross-links is depicted in figure 3.19. A full list of all cross-links that are not between two 40S ribosomal proteins can be found in the supplement (see table S2).

The findings proved the close interconnection of all eight eIF3 core proteins (on orange background shown in figure 3.19; also see figure S4). Also, all additional eIF3 factors show cross-links, either to each other or to ribosomal protein. The components localized close to the DC and mRNA entry channel during segmentation (section 3.3.3.2) and rigid body fitting (section 3.3.4.1) also show a variety of cross-links among each other, which proves their co-localization. Unfortunately, no cross-links were detected for eIF2 β , which made it impossible to exactly localize it. Eventually, the presence of cross-links of eIF4A and eIF4B with other components of the 48S-IC indicate their presence in the complex although no density was identified for them. During the subsequent local classification of areas showing significant heterogeneity, the cross-link information was used to identify (or improve) the exact position and orientation of the individual factors.

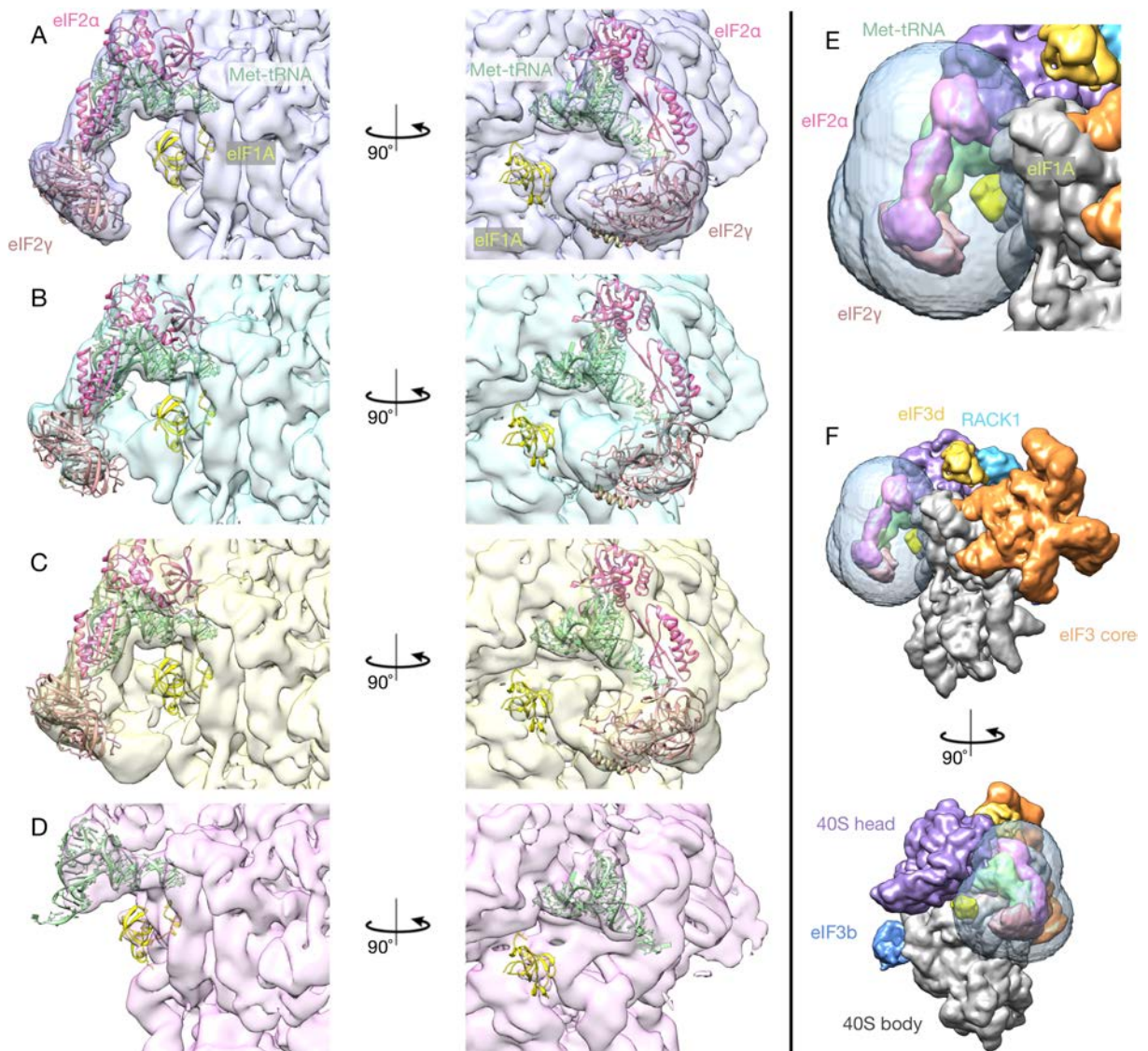


Fig. 3.20: Local classification of eIF2 and the decoding center. After performing a local classification of eIF2, the decoding center, and the surroundings of the mRNA entry channel, refinements were performed for all distinct classes. The results are depicted here. **A** shows a class, which is similar to the consensus 3D structure, lacking most density between eIF2 and the decoding center. **B** shows an additional density connected to eIF2α and γ. **C** depicts the third class that shows even more density, that not just is connected to eIF2, but also ranges towards eIF1A towards the DC. Finally, **D** shows a class, that entirely lacks eIF2 and only the tRNA remains visible. All structures are shown at equal thresholds. In **E**, the 3D mask is shown in transparent blue. It includes the TC, the decoding center, the mRNA entry channel and a reasonable amount of space in between them. Panel **F** shows the 3D mask in context of the fully segmented 48S-IC (color-code similar to figure 3.13).

3.3.4.4 Focused Classification of eIF2 and the mRNA Entry Channel

The mRNA entry channel is located between the 40S ribosomal subunit's head and body (see figure 3.13 or figure 3.16-E). The TC is located next to the entry channel and is responsible for bringing the Met-tRNA_i into position for AUG start codon recognition. The TC is composed of the three-subunit factor eIF2, the tRNA, and GTP. In addition, the factors eIF1 and eIF1A are known to be involved in start codon recognition near the

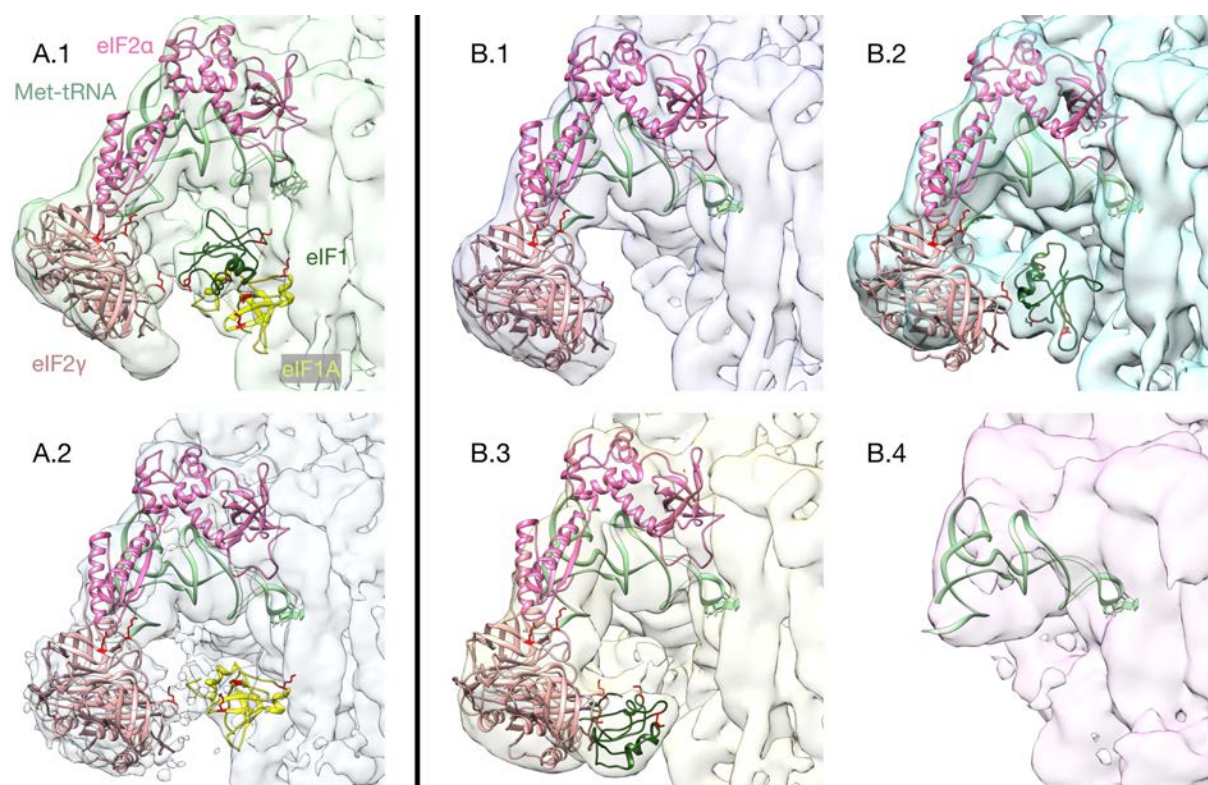


Fig. 3.21: Factor localization via cross-link mass spectrometry. While in the consensus refinement (shown in **A** and **D**) no density was visible close to eIF2 γ (rose) and eIF1A (yellow), local classification revealed two densities for two different classes (densities in pale cyan and yellow; shown in **B**, **C**, and **E**; also see figure 3.20). Cross-linking mass-spectrometry revealed lysines on the surface of both, eIF1A and eIF2 γ pointing towards this density and forming cross-links with eIF1. Therefore, eIF1 was fitted into these densities and arranged to fulfill the restrains given by the cross-links (BS3 cross-link distance is 20 Å). The cross-linking lysines are colored in red.

entry channel. Here, the consensus refinement 3D map showed significant flexible density when increasing the surface threshold.

In order to improve the resolution for this quite important part of the 48S-IC, signal subtraction (see section 2.3.4.6) in combination with focused classification was performed (see figure 3.20). For this, a 3D mask including only density around the mRNA entry channel and eIF2 (figure 3.20-E) was created. A complement mask was used for signal subtraction of the remaining parts of the complex. Afterwards, the projection images were subjected to 3D classification with focus to the area around the mRNA entry channel. Four distinct classes were identified (figure 3.20-A through -D). One class showed, similar to the consensus classification, only densities for eIF2 α and - γ . Another class had some additional density connected to eIF2. In the third class even more density was visible on both sides of eIF2, forming connections towards eIF1A and the decoding center. Unfortunately, both of these additional densities were not resolved well enough to recognize any secondary structures or make a reasonable fit for a specific factor.

However, when looking at the cross-linking data for this area, indicates that the newly emerged density can be assigned to eIF1, interestingly for both classes (see figure 3.21 and red connections in figure 3.19). For both densities, cross-link forming lysines were identified for eIF1A and eIF2 γ towards eIF1. The measured distances between the fitted eIF1 and its counterpart-lysines ranged between 16 to 20 Å, which is exactly the linking distance of BS3. For both areas, several cross-links were found, which restrains the possible number of eIF1 orientations. Therefore, it is highly likely that the newly emerged densities belong to eIF1, which is known to be located in the particular area and forms connections with eIF2 α , γ , and eIF1A.

For the second class (figure 3.20-C), two new densities were found close to eIF2 γ . However, only cross-links between the between other components and eIF1 were found for this area of the molecule. This means that either two eIF1 molecules are bound simultaneously in this class, or another factor is responsible for the second density here, like eIF2 β . Currently, this remains speculative due to the absence of cross-links including eIF2 β .

3.3.4.5 Focused Classification of mRNA Exit Channel

A similar approach as for the mRNA entry channel (see section 3.3.4.4) was performed for the area around the mRNA exit channel. Here, low-resolved and quite heterogeneous density clearly indicated the presence of additional factors or at least factor parts. A smeared out structure was visible connecting the eIF3 core domain around the solvent side of the 40S all the way to eIF3b. Thus, a local mask was created including only this flexible area (figure 3.22-E; also see figure S4 for a detailed representation of the eIF3 core structure). All other parts of the complex — the vast majority of the density — were subtracted from the particle projection images to minimize their influence during 3D classification. Local classification was performed resulting in ten classes of which three showed well resolved density (see figure 3.22-B/C/D).

Especially the first class that was found (figure 3.22-B) showed significantly more density than the consensus refinement. This additional density was not occupied by any of the previously fitted proteins or ribosomal components and therefore must be attributed to something else. Its shape did not show obvious resemblance with any of the thus far unfitted proteins. Furthermore, the search for cross-links in this area, formed by either the ribosomal proteins or any of the other eIF factors already present, did not give any hints about the origin of the additional density. Nevertheless, three parts were showing cross-links: a small density above eIF3b, eIF3b itself, and the already modeled components RACK1 and eIF3d above the mRNA exit channel.

The small density close to eIF3b was attributed to the RNA binding domain of eIFg, based on cross-links found for eIF3g with the adjacent the ribosomal proteins (figure 3.19; cyan

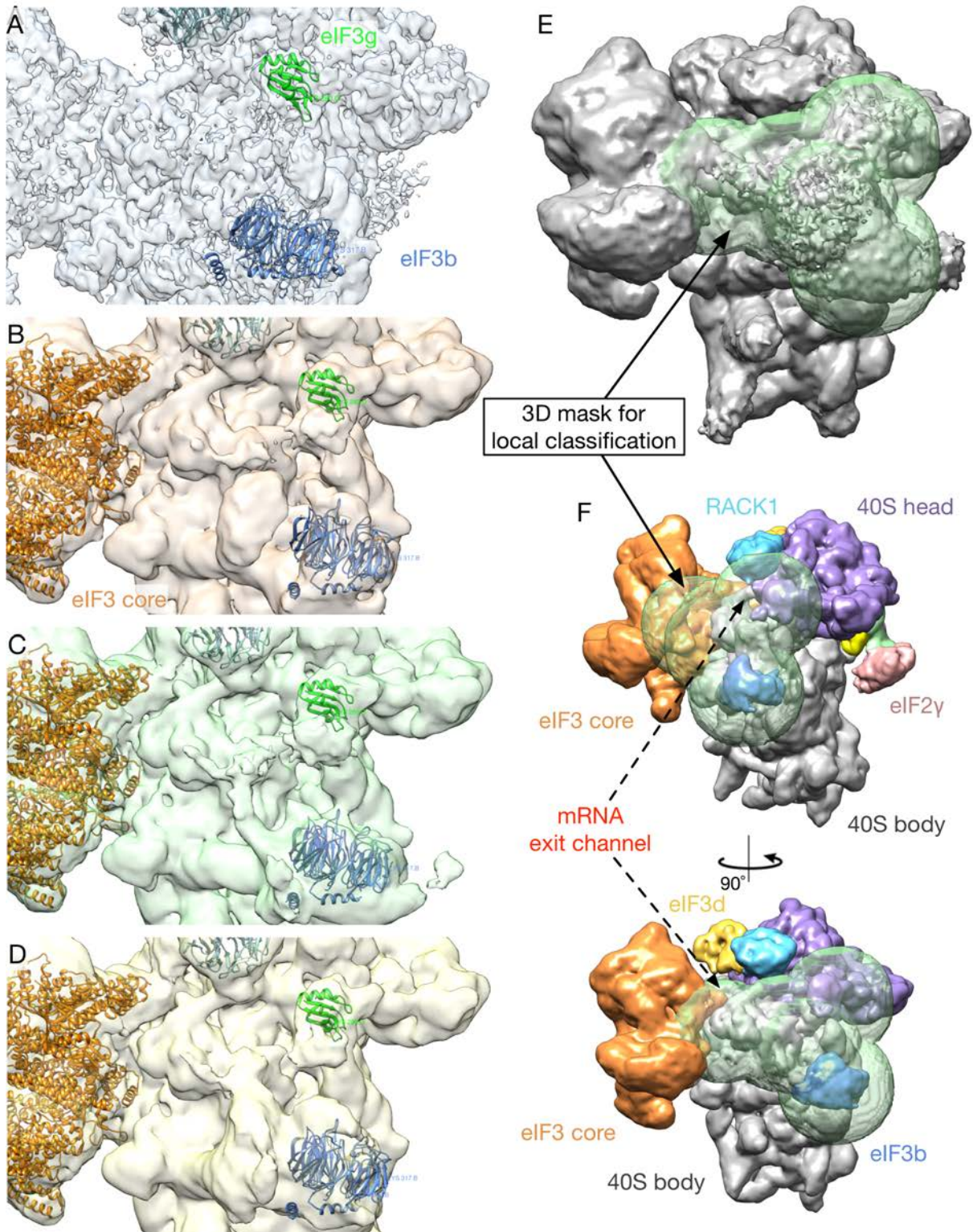


Fig. 3.22: Local classification of the mRNA exit channel. Very little density is visible when investigating the area near the complex's mRNA exit channel (panel A). However, when increasing the threshold, a large amount of spiky, dynamic density shows up from the void (panel E). To get a better understanding for this area enclosed by the green mask (panels E and F), signal subtraction and 3D classification were performed. As a result, three interesting classes were identified (panels B–D), which all show additional, but still quite fuzzy density. Furthermore, a density above the WD40 β -propeller of eIF3b (blue) was found and attributed to eIF3g (green) according to cross-linking mass spectrometry data.

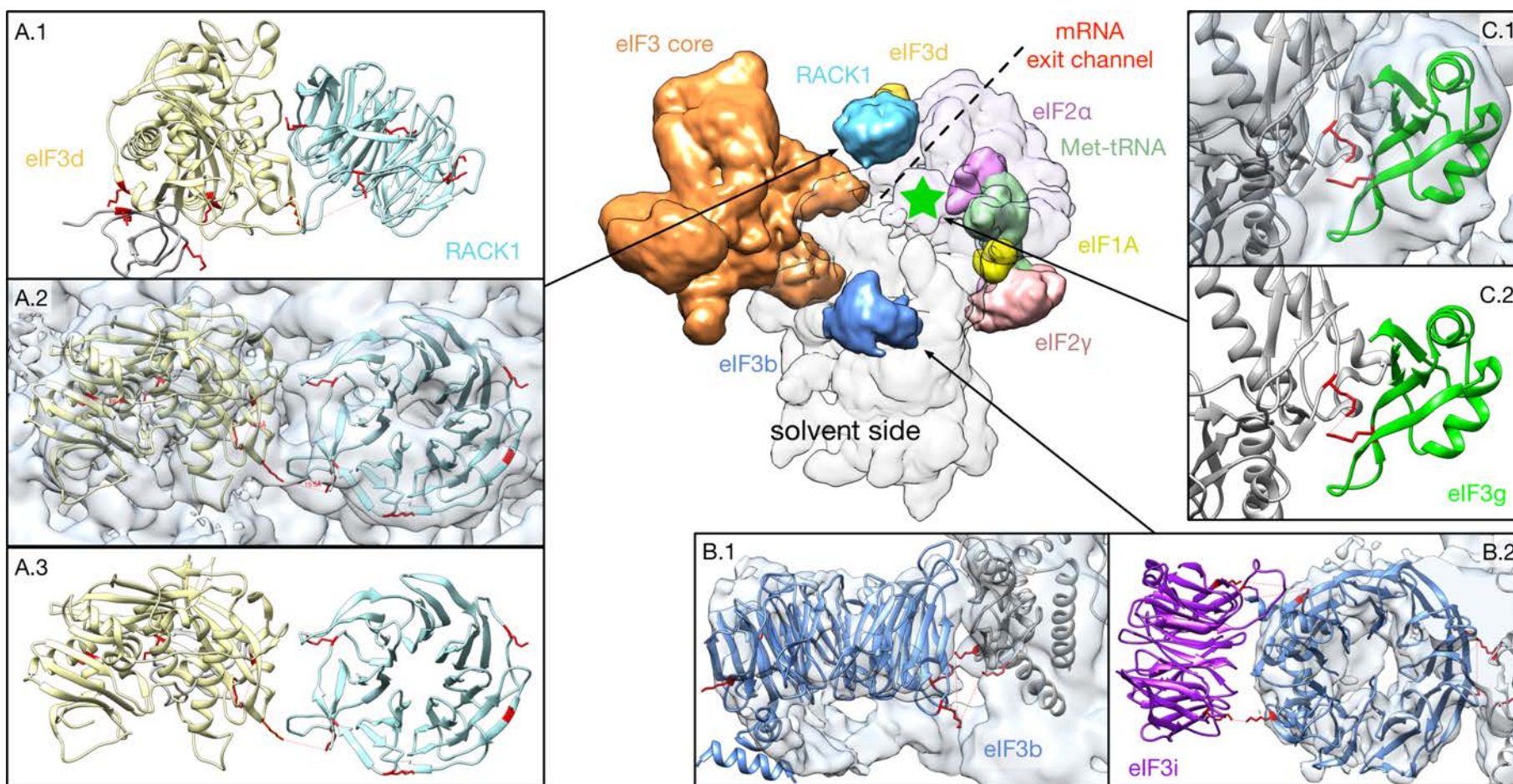


Fig. 3.23: Cross-links found for proteins at mRNA exit channel. At the 48S-IC mRNA exit channel, a heterogeneous, but connected density was identified (figure 3.22-E). Although not stable enough for straightforward model building, the cross-links found for this part of the 48S-IC were helpful to locate the position of several factors. All cross-link forming lysines are shown in red. The **A** panels show eIF3d (pale yellow) and RACK1 (cyan) and their respective cross-links with each other and the 40S ribosomal proteins 28 and 16 (figure 3.19; blue lines). RACK1 shows cross-links on one side with with eIF3d (**A.3**) and on the opposite site with eIF4B, for which no density was visible. For eIF3b, the connection to 40S ribosomal protein 9 is shown (panel **B.1**), as well as a possible connection to eIF3i (panel **B.2**), although no density is visible that could have been attributed to it (figure 3.19; green lines). The **C** panels show the interaction between the RNA binding domain of eIF3g and 40S ribosomal protein 3 (figure 3.19; cyan lines).

connection). Additionally, a part-structure of eIF3g from the PDB was fitted into the density (figure 3.23-C.1).

Although the general position of the nine-bladed eIF3b WD40 β -propeller was already identified using the refined structure, the cross-linking experiment gave more detailed insights (figure 3.19; green connections). Here, cross-links with both ribosomal proteins and eIF3i were found (see figure 3.23-B), helping to identify the exact orientation of the eIF3b and giving an indication about where eIF3i is most likely to be found (see figure 3.23). Judging from the orientation of the cross-links, eIF3i could easily fit into the large circular density of the first class from 3D refinement (figure 3.22-B), left of eIF3b in the picture.

Directly above the mRNA exit channel, the two proteins RACK1 and eIF3d were already fitted into the refined density of the 48S-IC (figure 3.23-A.2). However, an assessment of the cross-links found for both these proteins (figure 3.19; blue connections) not only proved the fitted orientations to be correct, but also revealed evidence regarding plausible locations of eIF4B and eIF4A: lysines on the solvent side of RACK1, pointing towards the unstructured area that was previously enclosed by a mask for 3D classification, apparently form cross-links with eIF4B. Therefore, eIF4B was located close to the mRNA exit channel, being a part of the flexible density within the 3D mask, although not directly visible after refinement. Unfortunately, the local classification did not reveal any additional density that could have been attributed to eIF4A or eIF4B.

Apart from the factors already mentioned, probably a lot of the flexible density surrounding the mRNA exit channel is composed of parts from the eIF3 core subunits. This is suggested by the cross-link mass spectrometry results, since the eIF3 core, particularly eIF3a and eIF3c, shows connections to most of the other complexes bound to the solvent side of the 40S (see supplementary table S2 and figure S4). eIF3a forms cross-links with eIF3b, c, d, f, g, h, m, and also eIF4B. For eIF3c, there are cross-links with eIF1, eIF2 γ , eIF3a, d, e, h, and l listed.

3.3.5 Overall 3D-PCA Analysis

Important steps of the cryoEM image processing workflow are 2D and 3D classification, during which similar images are identified and grouped together in the same class (see sections 2.3.1.6 and 2.3.1.8). This reduction of variance within one class is an extremely helpful technique to improve the resolution of a 2D projection or a 3D reconstruction. Thus far, hierarchical classification was used to focus on several parts or features of the 48S-IC one after the other (see section 3.3.3.3). Although time consuming, this stepwise approach helped to yield a single high-resolution structure. Unfortunately, the

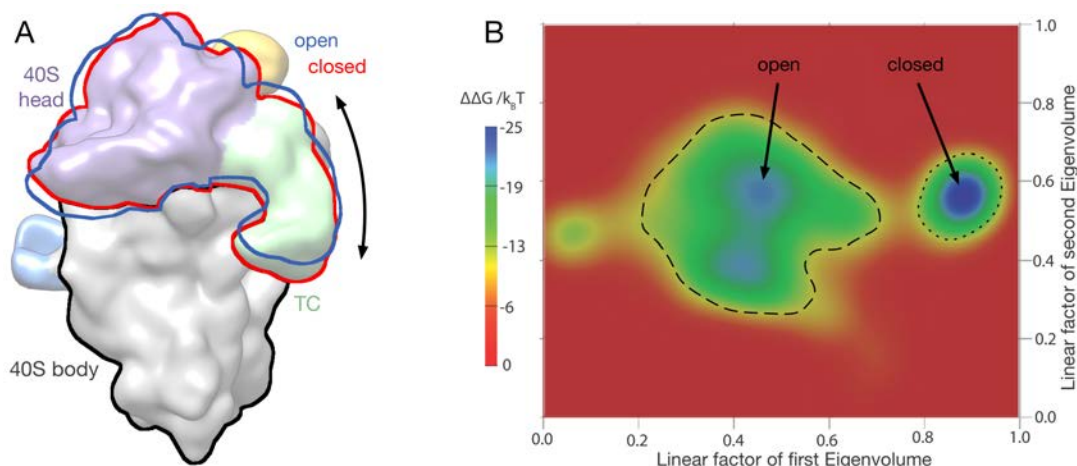


Fig. 3.24: 3D-PCA analysis of the 48S-IC. The full dataset with 821,651 projection images was subjected to extensive 3D classification and 3D-PCA analysis (adopted from previous published methods [82]). The resulting Eigenvolumes were visually inspected and only those representing actual motion within the complex were included. These Eigenvolumes were sorted in descending order according to the magnitude of variance. The largest order of motion was detected for the 40S ribosomal subunit and corresponds to the open- and closed-conformation during start-codon recognition, which involves a tilting movement of the 40S head (schematically depicted in panel **A**; direction of tilt indicated by black line with arrowheads). The second largest variance was detected for the area around the ternary complex (TC). The linear factors from both these Eigenvolumes were used to calculate an energy landscape (panel **B**), with the linear factor of the first Eigenvolume on the x-axis and the linear factor for the second Eigenvolume on the y-axis. The landscape shows two local minima for the open-conformation of the 48S-IC within a relatively large area of possible conformations (enclosed by **dashed line**). For the closed-conformation, a much smaller area around the local minimum is visible (**dotted line**), suggesting that this conformation of the complex exhibits significantly less flexibility.

vast majority of particle projection images was ruled out in the process and little to no insight was gained about any underlying dynamics that might have caused the low resolutions for some parts of the complex.

An entirely different approach, mainly utilizing 3D-PCA aided classification, can be employed to obtain a broader understanding of the different conformational states that exist in the full dataset (described in section 2.3.4.5 and [82]). Here, the full set of particle projection images after cleaning (821,651 projection images in total) was split into eight sets with approximately 100,000 projection images each. For every sub-set, a 20 class 3D-maximum likelihood classification was performed using RELION. The resulting 160 3D reconstructions were aligned, lowpass filtered to 20 Å resolution, and subjected to 3D-PCA analysis in the COW suite, which works similar to its 2D equivalent (see section 2.3.1.6). The 3Ds, and thus classes, were ordered according to their linear factors of the first two Eigenvectors. This ordered list of classes was used to split the dataset into a new set of eight subgroups and the entire procedure was repeated.

Eventually, after three such iterations, the 160 ordered classes were used to analyze the overall dynamic behavior of the 48S-IC complex. Therefore, the ensemble of all 160 3D

classes was lowpass filtered and normalized. A subsequent PCA analysis detected a large variability for the eIF3 core domain, most likely due to sub-stoichiometric occupation, and hence restricted the sensitivity of PCA towards purely conformational motions. Consequently, the eIF3 core was masked out before the ensemble was subjected to analysis via the CowScape algorithm (part of the CowSuite software package; manuscript in preparation). Subsequently, the resulting Eigenvolumes were ordered in descending order by the significance of the variance they describe, including only those which represented actual motion rather than variance caused by dissociation of individual factors. The linear factors of the two Eigenvolumes representing the strongest variance were used for visualization in an energy landscape (see section 2.3.4.5).

This energy landscape (see figure 3.24) describes the first two modes of motion: the x-axis shows the linear factors of the first Eigenvolume, which corresponds to the transition of the 40S ribosomal subunit from open- to closed-conformation. The linear factors for the second Eigenvolume were used for the y-axis, which is derived from motions of the factors bound close to the DC and mRNA entry channel. Several local minima were found for the open-conformation, which are embedded in a relatively large area of low energy and thus possible conformations (dashed line in figure 3.24-B). In contrast, the closed-conformation shows only one local minimum with a very limited surrounding area. Thus, it can be assumed that the open-conformation allows more flexibility than the closed-conformation, which could be a result of the regulatory and fidelity processes that play a role during mRNA scanning and start codon identification. However, it is difficult to determine if the increased heterogeneity of the open-formation is solely a product of dynamics. It could just as well originate from biochemical instability of the complex, like the dissociation of individual factors or their (not yet) stable binding to the entire complex. Nevertheless, the energy landscape provides an elegant way to quantitatively evaluate the amount of heterogeneity present in a large dataset of a macromolecular complex.

4 Discussion

The initiation of the eukaryotic translation, one of the four stages of translation (initiation, elongation, termination, and recycling), is a highly regulated and important step during which mature mRNA is associated with the ribosome and the AUG start-codon of the mRNA is identified in order to locate it correctly in the ribosome's P-site. This process is orchestrated by a large number of different initiation factors that interact with the mRNA, the ribosome, or both simultaneously. The sequence of events as well as the exact composition and structural conformation of the macromolecular complexes involved in translation initiation is of great interest because any errors in this chain of events can result in wrong polypeptide synthesis, imbalance in a cell's protein expression, and ultimately lead to life-threatening diseases like Parkinson's disease or cancer. In this thesis, two complexes purified from human HeLa cells were structurally investigated: the free eukaryotic initiation factor 3 (eIF3) and the 48S initiation complex. With the reconstructed 3D maps it was possible to shed light on the architecture and the tasks of these macromolecular machines during different steps of translation initiation. Furthermore, many of the factors involved were identified and localized precisely within said complexes. In the following section, the sophisticated networks of protein-protein and RNA-protein interaction within these complexes will be discussed and brought in line with previous findings to reveal a better picture of the full translation initiation cycle.

4.1 Biochemical Advances

4.1.1 eIF3 Purification

The outcome of all structural investigations, whether deriving from cryoEM, X-ray crystallography, or NMR, greatly relies on stability, purity, and abundance of the specimen under investigation. Thus, a reproducible and robust purification strategy is of prime importance. Purification procedures for multi-factor complexes often utilize recombinant expression of previously tagged proteins followed by purification and sometimes complicated reconstitution. The use of chromatographic methods, especially when dependent on tagging, usually comes with disadvantages such as high cost, relatively low yield, as well as the lack of easy transferability onto other source materials. In addition, the high-salt

buffers commonly used during chromatography can lead to undesirable dissociation of weakly bound complex components.

Hence, a different protocol was established for eIF3. The protocol employs mainly using differential precipitation with PEG400 and sucrose gradient sedimentation, adopted from previous work on the 20S proteasome [139]. Since chromatographic methods were completely avoided, several potential causes for complex destabilization, like unspecific interactions between chromatographic resins and the macromolecules leading to complex destabilization, were eliminated. Thus, the chromatography-free procedure made it possible to easily purify large amounts of native complexes from crude HeLa cytosolic extract. In addition, high protein concentrations and the use of crowding agents, such as PEG400 and sucrose, resembled the inner-cellular milieu and stabilized the complexes along the purification process. Furthermore, the absence of any high-salt buffer additionally decreased the danger for complex disintegration.

The eIF3 complexes purified according to the newly developed protocol showed great purity and overall compositional homogeneity in negative stain and cryo microscopy. Furthermore, it was possible to yield protein crystals from the purified complex. Although these crystals did not diffract to high resolution, they were another indicator for the quality of the purified complex, which had the eIF3j subunit bound at least sub-stoichiometrically (see table 3.1) while entirely lacking when purified according to previously published protocols [98]. Another difference compared to the cited literature is the fact that here, free eIF3 was purified directly from the cytosol, a large endogenous pool, and not harshly washed off of polysomes (often done by using 500 mM KCl). Selection of the optimal buffer system was guided by ProteoPlex, a method developed by colleagues [142], leading to a pH optimum of the buffer at 7.8 which is higher than in the standard protocol proposing pH 7.5. As pointed out by the authors of the ProteoPlex method, pH 7.5 is often chosen due to the lack of appropriate buffer screening tools and the justification to resemble the cytosolic pH value [153].

4.1.2 Sample Stabilization

The high demands regarding stability and homogeneity that are placed on the purification of macromolecular complexes apply equally to cryoEM sample preparation. However, in contrast to purification protocols, many agents that sustain complex integrity, and are therefore used as buffer components, are disadvantageous for cryoEM because their presence impairs contrast formation in the electron microscope. These compounds, such as sucrose or PEG, need to be removed before cryoEM sample preparation, although the necessary steps of such a protocol may impose damage to the macromolecules.

In order to improve the overall stability of the purified complexes, a modified version of the GraFix protocol was applied here. This led to relatively high amounts of sucrose in the final sample, which had to be removed. For this, size exclusion buffer exchange spin columns were used, which apparently have the tendency to damage the complexes and additionally show a limited potential to retain sucrose. Both these problems were addressed and overcome by pre-inactivation of the columns and the use of smaller elution volumes, respectively. The impact of these improvements was quite evident which is why both are now widely applied in the lab for numerous other complexes.

4.2 Interpretation of the eIF3 Structure

The eukaryotic translation initiation factor 3 plays a major role during translation initiation. It is the largest of the initiation factors and composed of 13 subunits (a through m) in mammals (eIF3 in yeast has eight subunits). It is often described as a scaffold, bringing the mRNA in contact with the small 40S ribosomal subunit. Thus far, all high-resolution structures of the eIF3 have been obtained from exactly that complex, where eIF3 is bound to the 40S. In contrast, the free eIF3 complex has not been resolved to high resolution, although such a structure would most likely provide enlightening insights into the mechanisms of the early translation initiation steps not involving the 40S.

Here, the free eIF3 was analyzed and resolved to a resolution of 7–15 Å. Unfortunately, a great amount of heterogeneity was observed for large parts of the complex rendering pseudo-atomic model building impossible: while the PCI domain, mainly composed of the eIF3 core subunits, exhibited the best resolution, other parts of the structure showed dramatically lower resolution and tremendous heterogeneity. Trials to sort this heterogeneity were performed, resulting in the reconstruction of several conformationally different structures. However, these 3D maps were still too low in resolution to identify any eIF3 subunits other than the quite rigid PCI core domain. Nevertheless, in the following text the structures obtained of the free eIF3 will be compared to what is known about its 40S-bound counterpart. Additionally, possible biochemical reasons for the apparent heterogeneity will be discussed.

4.2.1 Flexibility Within the eIF3 Structure

The structural analysis and 3D classification of the free eIF3 complex revealed several possible structures, which individually showed more isotropic density than the consensus refinement with all particles. This was a clear indicator for the presence of a significant degree of structural heterogeneity. However, none of the obtained 3D maps could be

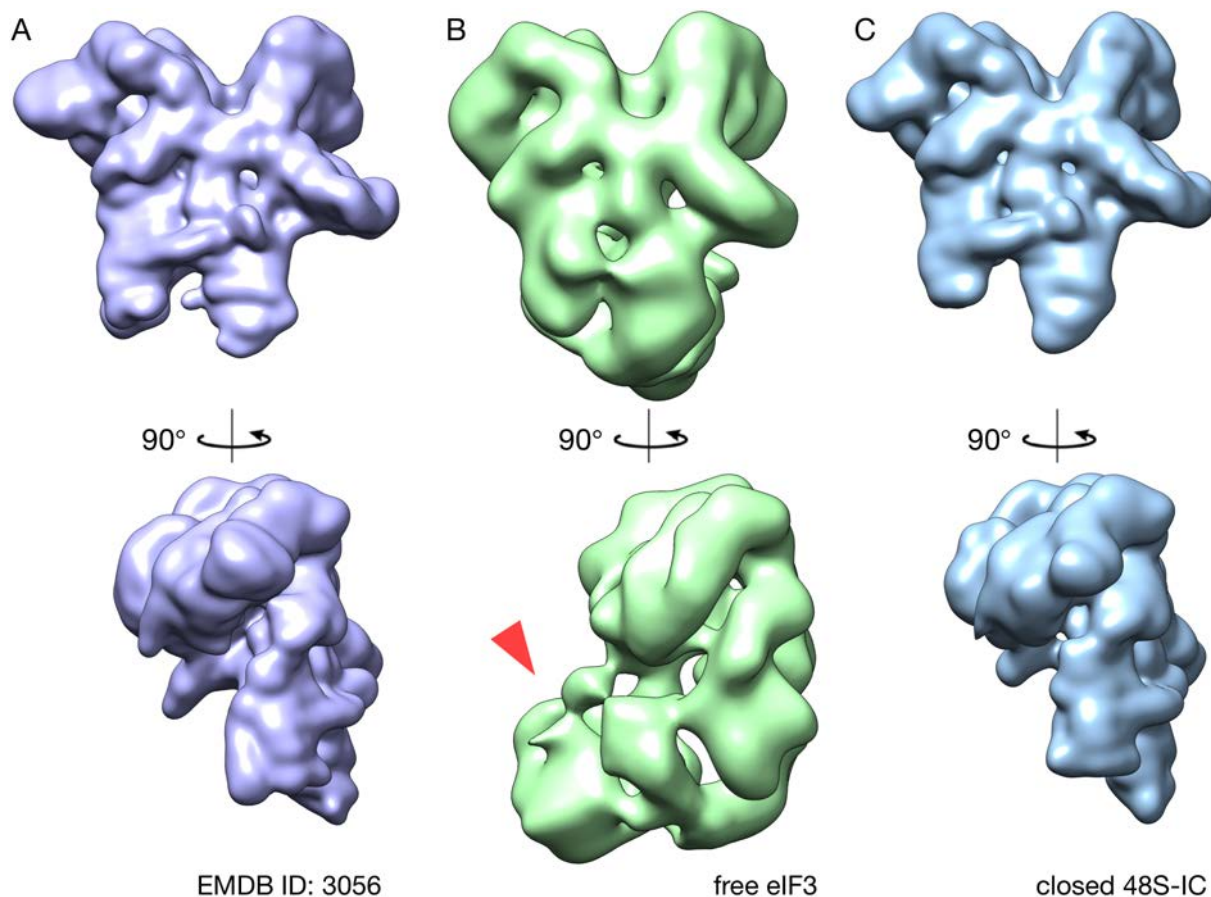


Fig. 4.1: Free eIF3 compared to its 40S-bound conformation. The core domain of the eIF3 map reconstructed here resembles quite accurately the overall structure of the eIF3 core, when bound to the 40S subunit. However, the free eIF3 shows significantly more density (indicated by **red arrow**) which most likely corresponds to the additional eIF3 subunits b, d, g, i, and j. Panel **A** shows the eIF3 core structure from the EMDB database (ID: 3056; [30]), panel **B** is a reconstruction of the free eIF3 complex, and panel **C** shows the segmented and extracted eIF3 core as reconstructed for the closed 48S-IC. All three structures are filtered to similar resolution and depicted in identical orientation.

refined to high resolution, which would have been an important prerequisite for modeling individual subunits into the density. The highest resolution of about 7 Å was achieved for the eIF3 core sub-complex including the PCI domain, composed of subunits a, c, e, f, h, k, l, and m. Its overall shape was quite similar to previously reconstructed maps of the eIF3 core attached to the 40S subunit (see figure 4.1). Thus, it can be assumed that the eight eIF3 core components adopt a conformation that is in line with previous findings for the 43S and 48S complex. [30, 34]

Additional density was clearly visible attached to the eIF3 core structure, showing strong heterogeneity during 3D classification. This density can be attributed with good confidence to the additional eIF3 subunits b, d, g, i, and j based on the fact that all these components were detected in mass spectrometry analysis of the purified complex. The sample for this analysis came from the fraction of a sucrose gradient that could only have contained a larger complex composed of all subunits. Individual and free subunits that

could have “contaminated” this fraction were highly unlikely because they would not have migrated this far in the sucrose gradient. This excludes the possibility of false-positively detecting free subunits.

4.2.2 eIF3 Composition and Stability

When establishing the purification procedure for the free eIF3 complex, great emphasis was placed on its compositional homogeneity and stability. In addition, the ProteoPlex method was not only used to determine the most stabilizing buffer composition, but also to verify the structural integrity of the complex. The ProteoPlex profile of the eIF3 complex (figure 3.1) showed a single peak with quick unfolding and thus no evidence for contaminations, factor pre-unfolding, or other artifacts. The fact that the free eIF3 complex purified in this study formed protein crystals was an additional, strong indicator for its compositional homogeneity. Based on this evidence, it was deduced that the purified sample meets the requirements for cryoEM in terms of stability and integrity.

Based on these findings, it can be assumed that the main reason for the high structural flexibility exhibited by the free eIF3 complex is not caused by compositional, but rather conformational, heterogeneity. Either the additional eIF3 subunits are just loosely bound to the eIF3 core complex, leading to a significant amount of flexibility, or there is an additional binding partner missing that would stabilize the entire construct, if present. A publication by *Smith et al.* [154] points towards the first possibility by reporting a quite “interdependent manner” of eIF3 assembly and stating the existence of stable sub-complexes, which are easily released from the complex, depending on the expression level of each eIF3 component within the cell. This behavior would match that of regulatory proteins, which are often only weakly bound (K_d in the μmol range) and therefore easily dissociate. In addition, weakly attached subunits promote a larger number of transition states. With regard to the second reason for flexibility, mRNA could be a potential stabilizing binding partner missing from the complex. Previous studies have identified a specialized initiation pathway during which not eIF4E but rather eIF3d binds the capped mRNA, as a member of the eIF3 complex [44]. Additionally, many eIF3 subunits are known to exhibit mRNA binding capacities while attached the 40S surface [14]. Also, the mRNA contamination found during cryoEM sample preparation, which led to the introduction of a streptomycin treatment (see section 3.1.2) in the first place, points towards this possibility.

In conclusion, it can be said with high probability that the observed heterogeneity is not a result of an insufficiently optimized purification protocol, but is more likely caused by something else: As already mentioned, the absence of an important ligand could be decisive here. Once identified, it might become possible to add or replace this ligand with

a functionalized derivative in order to induce a stabilizing effect on the complex. As a different countermeasure, it could be possible to further cool down the complex during sample preparation in order to thermodynamically confine it in a more homogeneous conformation. For example, GraFix could be performed at $-10\text{ }^{\circ}\text{C}$ instead of at $4\text{ }^{\circ}\text{C}$.

4.3 Interpretation of the 48S-IC Structure

The 48S translation initiation complex marks the final necessary step in the course of eukaryotic translation initiation during which the AUG start-codon on the mRNA is identified by the Met-tRNA_i and locked in the P-site of the decoding center. When the AUG codon is correctly identified, eIF1 leaves the complex; the eIF2-bound GTP is hydrolyzed with the help of eIF5 leading to the release of eIF2; and the recruitment of the large 60S ribosomal subunit is facilitated by eIF5B. Here, the complex was reconstituted with the small 40S ribosomal subunit; a short, uncapped, and unstructured mRNA; as well as a set of initiation factors (eIF1, 1A, 2, 3, 4A, 4B) from human source material. Thus, to prevent the final steps of initiation and to capture the complex in a conformation just before ribosomal subunit joining, the two factors eIF5 and eIF5B were excluded from this complex. Studying the complex at this stage of translation initiation reveals the mechanisms of start-codon identification quality control, codon-anticodon-binding, and triggering of elongation by subsequent factor release. Thus far, structures of this complex have only been reconstructed from yeast and rabbit reticulocyte lysate (RRL) and there has been some debate about the localization of individual initiation factors, such as the eIF3big sub-complex, and their capacity to adopt different binding positions along the surface of the 40S ribosomal subunit.

For the work of this thesis, high-resolution structures of the open and closed 48S translation initiation complex have been obtained, the closed-conformation structure showed the best resolution of up to 4.5 \AA . Both structures will be used to discuss the 48S-IC architecture, especially in the light of findings derived from different but closely related conformations, such as the 43S pre-initiation complex.

4.3.1 The eIF3 core Domain

The thirteen subunits of the translation initiation factor 3 do not form a single globular complex when bound to the 40S ribosomal subunit, but rather spread out across several positions on the surface of the 40S. In humans, the large eIF3 core sub-complex with eight components (a, c, e, f, h, k, l, and m) is located at the solvent side of the 40S, situated at a very similar position as reported for the 43S mammalian complex [30]. The large

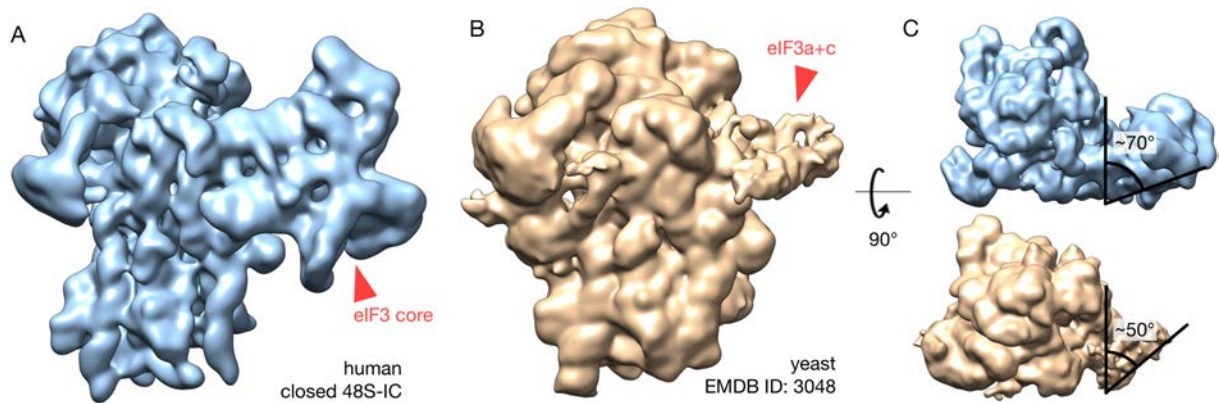


Fig. 4.2: Human eIF3 core compared to yeast eIF3ac. Although both species are quite similar in their execution of translation initiation, one major structural difference between human and budding yeast is the lack of five eIF3 subunits in yeast compared to mammals. Thus, reconstructions of the yeast 48S-IC (panel A; EMD ID: 3048 [32]) show significantly less density at the position where in human the large eIF3 core complex is found (panel B). Furthermore, the additional mass of the the human eIF3 core complex results in an additional lateral tilt by 20° compared to the yeast eIF3ac domain (panel C).

eIF3 core forms interconnections with the 40S mostly via subunits a and c, which is also backed by cross-linking mass-spectrometry data. When compared to the 48S structure from yeast [32], the additional density (for the additional subunits e, f, h, k, l, and m) in the mammalian system is clearly visible lateral of the eIF3ac complex in yeast (see figure 4.2-A) as part of the eIF3 core sub-complex. Furthermore, the mammalian eIF3 core structure is tilted $\sim 20^\circ$ laterally tilted away from the 40S compared to the eIF3ac-complex of yeast (see figure 4.2). The subunits also resemble the overall structure as in the free eIF3 complex (see section 4.2).

The eIF3 core structure of the reconstructed 48S shows good rigidity for all parts directly bound to the 40S, yet this rigidity gradually deteriorates with increasing distance from the 40S (see figure 3.15-B). Still, the overall eIF3 sub-complex is quite stable, considering the higher tilt angle of the human form compared to yeast, and the rather minor interaction surface with the 40S. Both of these findings suggest additional interaction partners that stabilize the eIF3 core in its position. Here, cross-linking mass spectrometry yields evidence for additional interactions of eIF3 subunit a and c with RACK1, eIF4A, and eIF4B, all of which are likely to be positioned in the region surrounding the 48S mRNA exit channel (discussed in section 4.3.5). These interactions could provide additional stabilization for the eIF3 core complex, although further experiments will be required to determine whether or not this is the case.

4.3.2 Open- and Closed-conformation of the 48S-IC

When investigating the 48S initiation complex, it is quite important to make one significant conformational distinction: the 48S-IC can exist in the open- and the closed-conformation. The main difference is whether or not the AUG start-codon has already been identified and locked in the decoding center by the Met-tRNA_i via interactions between the tRNA's ASL and the AUG codon. Structurally, the two steps can be distinguished by a $\sim 3^\circ$ tilt of the 40S head compared to the 40S body. This movement closes the ~ 10 Å gap between the mRNA and the Met-tRNA_i. During 3D classification, the number of particles showing the closed-conformation was roughly nine times higher than for the open conformation. Therefore, mRNA binding and start-codon recognition were very efficient and also significantly higher than for previously published studies (72 % and 51 % closed-conformation, respectively [32, 34]). This is especially interesting because eIF5, a canonical member of the 48S complex and previously described as mandatory for start-codon recognition [14], was missing from the complex and thus suggests that it is in fact not required for cognate start-codon identification.

A similar result was retrieved from the 3D-PCA analysis, where the calculated energy landscape (see figure 3.24) showed a clear minimum for the closed-conformation. This concurs with the fact that about 90 % of the particles images corresponded to that conformation. However, it was also found that the conformational flexibility for complexes in the open-conformation is much larger and the local minima are significantly less definitive. Here, it is difficult to specify whether this additional structural heterogeneity is a result of functional dynamics, such as procedures to assure fidelity or the movement of the mRNA through the 48S-IC, or rather just caused by biochemically insufficiently stabilized particles, which suffered from factor dissociation or damage during purification and sample preparation. Still, the application of 3D-PCA and the calculation of an energy landscape make it possible to quantify the structural heterogeneity found within a single sample and create the possibility to evaluate the impact of future biochemical improvements.

4.3.3 The Vicinity of the 48S-IC Decoding Center

At the decoding center of the 48S-IC, which is localized at the intersubunit side, several initiation factors bind to the surface of the 40S: eIF1, eIF1A, and eIF2. While eIF2 is responsible for bringing the Met-tRNA_i in contact with the P-site of the 40S during an earlier step of translation initiation (see section 1.3.2), eIF1 (binding close to the P-site) is very important to prevent initiation in non-Kozak context or at non-cognate codons. It functions as a figurative doorstop, keeping the DC gap open until correct AUG-tRNA-interaction in the P-site was detected. Afterwards, it is supposed to leave

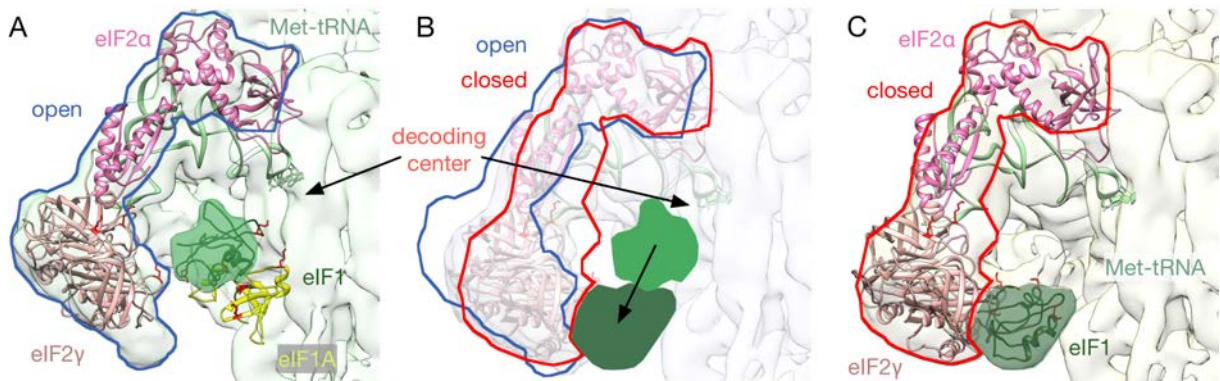


Fig. 4.3: Re-localization of eIF1 after start codon recognition. The initiation factor eIF1 was found to adopt different positions in the open-conformation (panel **A**) compared to the closed-conformation (one of several positions depicted in panel **C**). This is most likely due to the fact that eIF1 is involved in start codon recognition and responsible for fidelity. Furthermore, it has to leave its position in the open-conformation in order to make way for the tilt of the 40S head. Local classification has revealed several new positions for eIF1 in the closed-conformation, one of which is depicted here schematically (panel **B**).

the complex making way for the 40S to enter the closed-conformation. Meanwhile, eIF2 prevents joining of the 60S large ribosomal subunit. eIF1A (binding close to the A-site) participates in eIF5B recruitment, which in turn triggers eIF2 release and 60S joining.

When comparing the reconstructions of the open-conformation with the best closed-conformation structure, the difference in eIF1 localization is quite obvious and in accordance with the predicted order of events (see figure 4.3). However, eIF1 apparently does not entirely leave the 48S-IC, because cross-links show its association to eIF2 in two different positions for the closed-conformation (one of which is shown in figure 4.3; also see section 3.3.4.4 and figure 3.21). The identification of these two new positions were only possible after thorough local 3D classification and are in agreement with cross-link mass spectrometry results. The new positions in the proximity of the decoding center appear to be functionally relevant, because eIF1 can not prevent tilting of the 40S head anymore, but could still be in place to recruit eIF5. This would be important, since eIF5 has GAP capabilities necessary to hydrolyze the eIF2-bound GTP that triggers the release of eIF2. The repositioning of eIF1 has not been observed thus far, most likely because previous 48S-IC structures were reconstituted in the presence of eIF5. Another reason could be that eIF1 can apparently adopt several different positions which could lead to averaging out of its density when not tightening the focus of classification as was done here.

In addition to the newly identified positions of eIF1, another slightly larger density has been observed for both, the open- and the closed-conformation (see figure 3.20-B/C). Due to the lack of cross-linking information for this density, it was not possible to unambiguously attribute it to a specific factor, although its position and size would be in alignment with knowledge about the eIF2 β factor. A recently publication by *Eliseev et al.* [34] has detected a comparable density at the same position and also attributes it to eIF2 β . Sim-

ilar to the eIF1 factor, the density disappears after closure of the 40S head, but 3D local classification recovers a similar sized density at a position close but slightly tilted away from the decoding center (see figure 3.20-B/C), still attached to eIF2. This unknown density could facilitate a similar duty as does the IF3-CTD in the bacterial system (usually compared to eIF1), because its localization and rearrangement is somewhat similar with published results [155]. However, without conclusive cross-linking results, this attribution has to be considered highly tentative.

The energy landscape generated from 3D-PCA offers a possible explanation for the lack of clearly defined densities: the second of the two Eigenvolumes used for its creation reflects the heterogeneity in the environment of the TC. The relatively large area of possible conformations for the open-conformation of the 48S-IC is an indication of large amounts of structural flexibility at this point. As already mentioned above (see section 4.3.2), it is however not possible to unambiguously interpret this flexibility as functional dynamics. In order to clarify whether there additional, possibly thus far unknown, transition states exist, a dataset with larger numbers of open-conformation 48S-IC complexes would be required. Then, the calculation of an energy landscape could make it possible to identify the particles corresponding to such intermediate states and refining them selectively.

4.3.4 Localization of the eIF3big Sub-Complex

The translation initiation factor 3 tends to, although composed of thirteen subunits, not to form a large, single complex when attached to the 40S, but rather splits up into several smaller sub-complexes that attach to different positions on the surface of the 40S. One of these subcomplexes consists of the subunits b, i, and g. For this eIF3big complex, several RRM s have been reported as well as its function as a scaffold for additional factors, such as DHX29 [30, 41, 42, 147]. These publications report that eIF3big is involved in helicase activities for the unfolding of mRNA secondary structures, or at least aiding other molecules to do so. However, the exact positioning of eIF3big and its capabilities to switch positions along the surface of the 40S has been subject to debate.

With a combination of cross-linking mass spectrometry and rigid body fitting, X-ray and NMR (partial) structures could be positioned into density at the platform and solvent exposed site of the 40S. For subunits b and g, at least partial density is visible and resembling their shape. Especially, the quite characteristic shape of eIF3b, a nine-bladed WD40 β -propeller, is clearly visible. The position suggested for eIF3i by cross-links show a smeared out and low abundance density, which could thus not be attributed with full confidence. Still, all reconstructions calculated for different classes during 3D classification either show eIF3big at the positions just described or not to be present in the 48S-IC at all. This is corroborated by the cross-links, which agree only with these specific

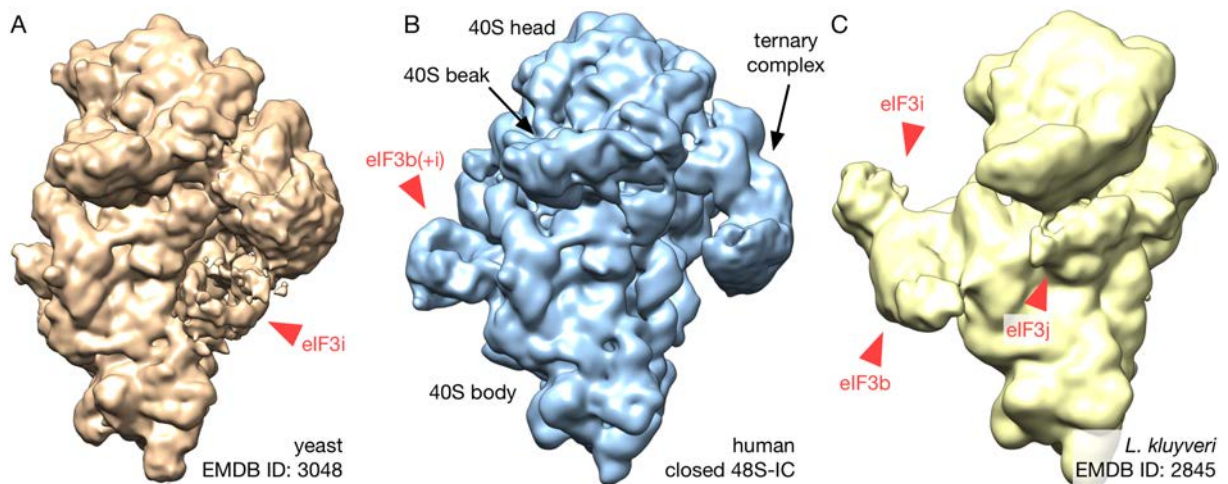


Fig. 4.4: Positioning of eIF3 subunits b, i, and g. The exact position and re-localization of the eIF3 subunits b, i, g, and j has been subject to debate. Here, different postulated positions along the platform of the 40S are compared. Panel **A** shows the closed 48S-IC published by *Hussain et al.* (EMDB ID: 3048; [31]), where the eIF3i subunit seems to adopt a position next to the P-site of the decoding center interacting with eIF1A and eIF2. Panel **B** shows the human 48S-IC as reconstructed during the work of the thesis, where eIF3b is clearly visible at the solvent side of the 40S and eIF3i was associated via cross-links directly next to it. In panel **C**, the fungi 43S is shown (EMDB OD: 2845; [151]), where subunit eIF3j was detected between eIF1A and eIF3b. Furthermore, clear densities for the eIF3b+i subcomplex are visible at the solvent side of the 40S, placing both at the same position as they were found here. All three structures were filtered to similar resolution and depicted in identical orientation.

locations. Additionally, the position of the eIF3big complex is similar to that reported by others [34, 47, 149, 151]. Apart from internal cross-links that are formed among all three subunits, additional cross-links were identified for the CTD of eIFa. This supports the previously elsewhere described model of eIF3a forming a connection with the eIF3big complex via a rather flexible connection [14]. In the closed-conformation reconstruction, this flexible linker could not be identified with full confidence, although a low abundant density leaving eIFa and pointing towards eIF3big is visible.

However, eIF3 subunits b, i, and g have also been associated to other positions within the 48S-IC in the past (see figure 4.4). For the first time, *Llácer et al.* [32] have presented a high-resolution cryoEM structure where they modeled eIF3i to bind directly next to eIF1A and eIF2 and thus in the direct proximity of the decoding center of the 40S. In contrast to the human complex used for this thesis, their 48S-IC was reconstructed from yeast proteins in the presence of eIF5, but lacking all eIF4 components. Accordingly, these compositional differences could have induced the alternative arrangement of subunits. Another possible explanation for the different position of eIF3i given in the paper is its putative function as locking mechanism to keep the mRNA in place. The two structures also differ slightly in their eIF2 positioning: the ternary complex in the 48S-IC from *Llácer et al.* adopts a orientation, where eIF2 γ is tilted upwards and towards the beak of the 40S head. However, this tilt angle is more comparable with the 48S-IC open-conformation presented

here rather than the closed-conformation. Unfortunately, no additional information (such as cross-linking or directed hydroxyl radical cleavage) were presented by *Llácer et al.* [32] to back the rigid body fitting done for this particular complex. The second noteworthy publication showing a cryoEM structure with a quite different orientation of the eIF3 subunits i and g was published in 2016 by *Simonetti et al.* [150]. Here, the subunits were positioned again very close to eIF1A, eIF2, and the decoding center of the yeast 48S-IC. However, this interpretation was revised by the authors in a second paper in 2017 [156], stating that the density previously attributed to eIF3i and g has to be attributed to the NTPase ABCE1 instead, which was also bound to the complex.

4.3.5 Structural Heterogeneity at the mRNA Exit Channel

Besides the decoding center, the mRNA entry channel, and the eIF3big sub-complex, the 48S-IC has another quite interesting area where factors bind: the mRNA exit channel. It is located close to RACK1 on the solvent side of the 40S subunit, located between the eIF3big binding position and the eIF3 core domain. It has been shown that several proteins involved in binding of the mRNA 5'-cap are located here, to stabilize the 5'-UTR during translation initiation [14]. Elucidating the specific interactions at this part of the 48S-IC would foster a better understanding of how the mRNA-associated factors, like the components of eIF4F, help in start codon recognition.

It was possible to show that eIF3d is indeed bound to the 40S at the mRNA exit channel, which is backed by the formation of cross-links with both the 40S and RACK1. Furthermore, the density attributed to eIF3d clearly resembles the shape of its X-ray derived model. The positioning of eIF3d at the mRNA exit channel, where it has also been reported previously [34], would be explainable by its cap-binding activity and thus mRNA stabilization capabilities in earlier studies [44]. Additionally, for eIF3d as well as for RACK1, which is positioned right next to eIF3d, cross-linking mass spectrometry has found links with eIF4A and eIF4B, both were added during reconstitution of the 48S-IC and are known for their mRNA binding and helicase activity. This position is similar to the one reported by *Eliseev et al.* for eIF4B [34]. Unfortunately, no density was identified that could have been attributed to these two factors. Therefore, their positioning just via the cross-links has to be considered very cautiously. Two possible reasons for the lack of any eIF4A and eIF4B density are first the fact that an unstructured mRNA has been used to reconstitute the 48S-IC, which might make the presence of mRNA helicase obsolete, and secondly, both factors are known to bind directly to the mRNA and only indirectly, via the mRNA, to the 40S surface. Thus, their connection to the 48S-IC via mRNA might have been simply too flexible in order to show up as rigid density in the reconstruction.

4.3.6 Localization of eIF3j

One of the thirteen eIF3 subunits, eIF3j, seems to behave a little differently than the other twelve subunits. It apparently tends to bind to the eIF3 complex just loosely and is thus in some publications not counted as a eIF3 subunit at all [154]. Unfortunately, it was not possible to identify any density or instructive cross-links for eIF3j either in the open- or in the closed-conformation of the 48S-IC. At the position where it has been shown previously, in between eIF1A and eIF3b at the 40S surface, no additional density was visible (see figure 4.4-C; [151]). And for the sole cross-link that showed up for eIF3j in the mass spectrometry data, no unattributed density could be found. It thus remains unclear, whether eIF3j was bound to the 48S-IC or not. The lack of density and cross-links would point towards this direction, but the absence of density can also be a result of conformational heterogeneity or a sub-stoichiometric occupation. Following this line of argument, *Sokabe et al.* [157] show that the association of eIF1, eIF1A, and eIF3j to the 43S PIC is thermodynamically coupled, which contradicts the lack of eIF3j. In other publications, the importance of eIF3j for initiation fidelity was postulated, because point mutations within eIF3j would result in leaky scanning [47]. For our 48S-IC, no such leaky scanning was detected and *in vitro* translation resulted in poly-peptides corresponding to only one (the first) AUG codon on the translated mRNA (personal communication with Dr. Akanksha Goyal).

4.3.7 Model and Dynamics of Eukaryotic Translation Initiation

When merging all the information gathered about the 48S-IC, it becomes possible to assemble a full model (shown in figure 4.5) of this late-stage translation initiation complex. Furthermore, from the local classifications performed for several parts of the complex, conclusions regarding the dynamics of the complex can be drawn and the re-localization of individual factors was observed. However, this model represents only a snapshot of a single, or at least very narrow, moment in the full cycle of translation initiation.

In order to study the dynamics of the 48S-IC in more details, all particles images were subjected to an approach based on 3D-PCA. It revealed in its first iteration a large amount of heterogeneity for the eIF3 core complex which was hence excluded from the analysis. The heterogeneity was most likely a result of the (incomplete) dissociation of the eIF3 core from the full complex during cryoEM sample preparation. Given the exposed binding position of the eIF3 core on the surface of the 40S and its relatively small area of interaction with the rest of the complex, its dissociation appears easily possible. Still, this artifact of the sample preparation procedure might have concealed any relevant dynamics of the eIF3 core.

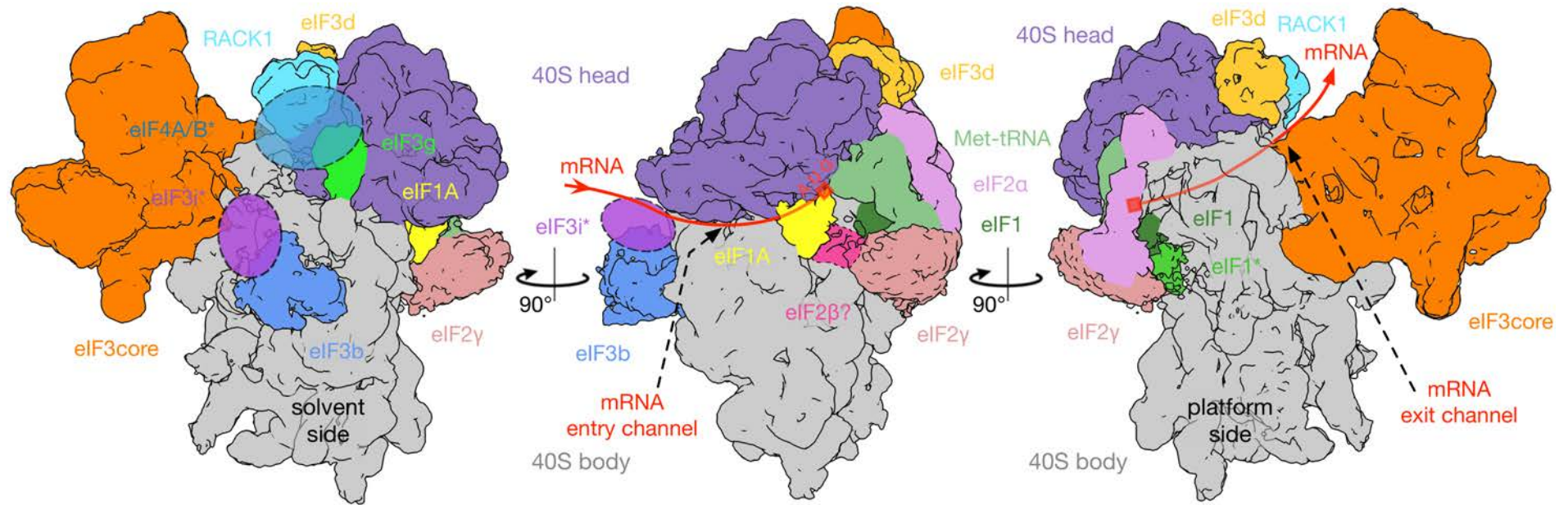


Fig. 4.5: Summarized model of eukaryotic translation initiation. Here, a schematic view of the 48S-IC with the positions for all identified factors and components involved in eukaryotic translation initiation is depicted. The main structural scaffold is the 40S ribosomal subunit, composed of the 40S body (gray) and the 40S head (purple). Attached to it on the solvent side is the eight subunit eIF3 core (eIF3a, c, e, f, h, k, l, and m; orange). In the vicinity of the decoding center (marked with the AUG and a red square), the ternary complex is bound, composed of eIF2-GTP (eIF2 α and eIF2 γ ; light and pale rose) and the Met-tRNA_i (light green). The path of the mRNA is shown in red, although density for the mRNA was only obtained in the P-site of the decoding center. Directly next this position, eIF1 (dark green) is bound, which was also found for an additional position (asterisk) in the closed-conformation, further away from the P-site and bound to eIF2 γ . Next to the A-site, eIF1A (light yellow) is bound. During focused classification, an additional density was visible in the open-conformation forming a stabilizing connection between eIF1A and eIF2. This might be attributed to eIF2 β (light pink), although no cross-links were found to support this assumption. On the back of the 40S, exposed to the solvent, the nine β -propeller factor eIF3b was identified (blue). Attached to it, cross-links indicated the presence of eIF3i (transparent purple), although no density could be recovered during refinement. Next to eIF3b, cross-links and density were found and attributed to parts of eIF3g (light green). An top of the 40S head, next to RACK1 (cyan), density for eIF3d (dark yellow) was identified and proved via cross-links close to the mRNA exit channel. Attached to these two proteins, cross-links for eIF4A and eIF4B were found (transparent blue). However, no visible density could be reconstructed for both.

After excluding the eIF3 core from the 3D-PCA, the conformational changes of the 40S head between the open and closed state and heterogeneity in the vicinity of the decoding center were found to represent the most significant movements within the complex. This seems logical because they are the most important to facilitate the task of the 48S-IC: start codon identification via the codon-anticodon-interaction. In addition, the open-conformation was found to be more heterogeneous and to be capable of adopting a larger number of different intermediate states. However, due to the fact that only 10 % of the particle images from cryoEM represented that state, it was difficult to sort for the heterogeneity and recover high-resolution structures.

Although it was not possible to recover any additional high-resolution structures from the particle sorting with the help of the 3D-PCA, it still provided a more comprehensive picture of the dynamics underlying start codon recognition. Thus, 3D-PCA proved to be an invaluable tool for the investigation of structurally very heterogeneous processes such as eukaryotic translation initiation.

5 Conclusion and Outlook

Eukaryotic translation initiation involves several steps that are extremely important for the fidelity of translation: correct mRNA placement within the decoding center of the ribosome, correct Met-tRNA_i recruitment, and most importantly identification of the correct start codon. These duties are orchestrated by a broad range of different initiation factors that come and go during different sub-steps of translation initiation to form functional macromolecular complexes.

For this doctoral thesis, the human translation initiation factor 3 has been studied in its free form as well as associated with the 48S initiation complex. For both of these macromolecules, high-resolution structures have been presented, which made it possible to analyze previously unknown compositional states based on newly identified factor cooperations and localizations.

As a prerequisite to study the free eIF3 complex, an entirely chromatography-free purification strategy was established with the potential to isolate it from different source material. The structural integrity of the complex was verified via the use of the ProteoPlex assay as well as the fact that protein crystals were obtained. It could be shown that parts of the free eIF3 complex adopt a rigid and similar 3D structure compared to its equivalent bound to the 40S ribosomal subunit, while other parts presented strong conformational heterogeneity at the same time. Due to the results from structural integrity investigations, it was assumed that the heterogeneity was a result of conformational flexibility rather than compositional differences or even loss of structural stability. Extensive 3D classification made it possible to recover several 3D reconstructions with significantly more isotropic density than for the eIF3 core (the rigid part bound to the 40S ribosomal subunit) alone.

The 48S-IC was resolved to a resolution of 4.5–12 Å and it was thus possible to identify 12 of the 13 eIF3 subunits (all but eIF3j) via rigid body fitting and cross-linking mass spectrometry. In addition, local 3D classification made it possible to detect new positions adopted by eIF1 after start-codon recognition and also verify them with cross-links. Additional cross-links gave at least hints for the positioning of the initiation factors eIF4A and eIF4B. The integration of all these findings made it possible to gain a better understanding of the steps involved in eukaryotic translation initiation and construct a model including most of the involved factors. In addition, a more integral analysis of the heterogeneity present within the sum of all particle images was performed, based on

3D-PCA. Here, the aim was to identify the dynamics underlying the structural flexibility of the complex and sort for it. The opening and closing of the 40S head were identified to be the most significant reason for the heterogeneity together with movements of the components bound to the surroundings of the decoding center.

However, some questions remain unresolved and new arose during this study: The most prominent regards the exact order of events during mRNA feeding into the decoding center, the mechanism of scanning, and the factor rearrangements occurring in the process. This also involves the question of the exact binding positions of eIF3j, the different eIF4s (A, B, E, G, and H), eIF5, and eIF5B. Furthermore, the understanding of the exact mechanism of eIF2-GTP hydrolysis, as well as the involved factors, would greatly benefit from structural examination.

In order to answer these questions, two main topics are crucial: biochemical measures to further stabilize the investigated complexes in conformations that are the most interesting to study and computational methods for a better separation of compositional and conformational heterogeneity. For the biochemical stabilization, the addition of agents that reduce the macromolecule's degrees of freedom and confine it in a more homogeneous, single conformation, such as drugs or specialized nanobodies, should be beneficial for later high-resolution refinement. This might be especially advantageous for the thus far elusive factors that are bound to the mRNA. The sorting for structural heterogeneity is a task that has undergone great improvements in the last couple of years due to the development of new, specialized algorithms and the availability of high-performance computational clusters. Still, it remains difficult to separate compositional heterogeneity (e.g. damaged or broken particles) from conformational heterogeneity (e.g. dynamics involved in functionality of a complex) and identify all macromolecular intermediate states. To overcome this, the already utilized 3D-PCA method is most promising because it provides the advantage of separating the structural heterogeneity within a dataset without previous knowledge and yields insights into the dynamics of a complex at the same time.

Through a combination of refined biochemical stabilization and innovative computational sorting, it should become possible not only to yield structures depicting more distinct intermediate states, but also improve the resolution significantly.

A Abbreviations

2D	two-dimensional
3D	three-dimensional
3'-UTR	3' untranslated region
5'-UTR	5' untranslated region
aaRS	aminoacyl-tRNA-synthetase
aa-tRNA	aminoacyl-tRNA
ASL	anticodon stem loop
ATP	adenosine triphosphate
BS3	bis(sulfosuccinimidyl)suberate
BSA	bovine serum albumin
CCD	charge coupled device
CCF	cross correlation function
cryoEM	3D transmission electron cryo-microscopy
CTD	C-terminal domain
CTF	contrast transfer function
DC	decoding center
DDD	direct electron detection device
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DQE	detective quantum efficiency
DTT	dithiothreitol
eEF	eukaryotic elongation factor
eIF	eukaryotic translation initiation factor
EM	electron microscope
EMDB	Electron Microscopy Data Bank
FEG	field emission gun
FSC	Fourier shell correlation
FT	Fourier transformation
GA	glutaraldehyde

GAP	GTPase-activating protein
GDP	guanosine-5'-diphosphate
GTP	guanosine-5'-triphosphate
HEAT	H untingtin, e longation factor 3, PR/A subunit of protein phosphatase 2A and the TOR lipid kinase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IC	initiation complex
IRES	internal ribosome entry site
LMNG	lauryl maltose neopentyl glycol
m⁷G	7-methyl guanosine
Met-tRNA_i	methionine initiator-tRNA
ML	maximum-likelihood
MPN	<i>Mpr1, Pad1, amino-N-terminal</i>
mRNA	messenger RNA
MS	mass spectrometry
MSA	multivariate statistical analysis
MW	molecular weight
NMR	nuclear magnetic resonance
NTT	N-terminal tail
OGNG	octyl glucose neopentyl glycol
ORF	open reading frame
PABP	poly(A)-binding protein
PCA	principle component analysis
PCI	<i>proteasome, COP9 signalosome, translation initiation factor</i>
PDB	Protein Data Bank
PEG	polyethylene glycol
PIC	pre-initiation complex
PMPI	p-maleimidophenyl isocyanate
PMSF	phenylmethane sulfonyl fluoride
pre-mRNA	pre-mature messenger RNA
PSF	point spread function
PTC	peptidyl transferase center
REM	reflection electron microscopy
RI	refractive index

RNA	ribonucleic acid
RNP	ribonucleoprotein particle
rRNA	ribosomal RNA
RRM	RNA recognition motif
RRL	rabbit reticulocyte lysate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM	scanning electron microscopy
SGD	stochastic gradient descent optimization
SNR	signal-to-noise ratio
TC	ternary complex
TEM	transmission electron microscope
TRIS	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA

B Supplementary information

B.1 Tables

protein name	UniProt ID	MW [kDa]	lane 1 (control)		lane 2 (1 % stpm)	
			count	prob	count	prob
<i>h.s.</i> eIF2 α	IF2A_HUMAN	36	26	100	5	100
<i>h.s.</i> eIF2 β	IF2B_HUMAN	38	5	100	5	100
<i>h.s.</i> eIF2 γ	IF2G_HUMAN	51	18	100	6	100
<i>h.s.</i> eIF2B- α	EI2BA_HUMAN	34	5	100	0	—
<i>h.s.</i> eIF2B- β	EI2BB_HUMAN	39	6	100	0	—
<i>h.s.</i> eIF2B- γ	EI2BG_HUMAN	50	8	100	0	—
<i>h.s.</i> eIF2B- δ	E7ERK9_HUMAN	60	42	100	2	100
<i>h.s.</i> eIF2B- ϵ	E9PC74_HUMAN	78	13	100	1	100
<i>h.s.</i> eIF4G	E7EUV4_HUMAN	172	97	93	97	100
<i>h.s.</i> eIF5	IF5_HUMAN	49	11	100	3	100
<i>h.s.</i> eIF5B	IF2P_HUMAN	139	87	100	18	100

Table S1: Mass spectrometry analysis of the native human eIF3 (addition). Here, additional eukaryotic translation initiation factors, other than eIF3's components, are shown. They were found in the same run as eIF3 (see section 3.1.3 and table 3.1, page 51). The probability is calculated via the PeptideProphet algorithm (according to [144]). Legend: *count* = total spectrum count; *prob* = probability; *stpm* = streptomycin. MS-MS measurements were carried out by members of the group of Prof. Henning Urlaub (*Research Group Bioanalytical Mass Spectrometry, Max-Planck-Institute for Biophysical Chemistry, Göttingen*)

protein 1	protein 2	residue 1	residue 2	max of score	
EIF1_HUMAN	EIF1AX_HUMAN	42	40	9.28	
		91	56	5.83	
	EIF2G_HUMAN	2	303	8.04	
		2	312	5.06	
	EIF3C_HUMAN	58	267	7.20	
EIF1AX_HUMAN	EIF1_HUMAN	40	42	9.28	
		56	91	5.83	
	EIF3H_HUMAN	68	189	11.57	
	RS10_HUMAN	88	31	5.08	
	RS30_HUMAN	16	1	9.38	
		23	1	13.63	
		68	1	8.23	
		88	1	5.75	
EIF2A_HUMAN	EIF2G_HUMAN	61	312	5.57	
		312	303	11.77	
		312	449	11.36	
	RS14_HUMAN	97	63	5.15	
EIF2G_HUMAN	EIF1_HUMAN	303	2	8.04	
		312	2	5.06	
	EIF2A_HUMAN	303	312	11.77	
		312	61	5.57	
		449	312	11.36	
	EIF3C_HUMAN	303	259	5.50	
EIF3A_HUMAN	EIF3B_HUMAN	694	718	12.39	
		694	722	7.58	
		694	729	8.44	
		720	345	5.46	
		742	459	7.21	
	EIF3C_HUMAN	191	760	5.43	
		191	764	5.64	
		196	643	10.54	
		351	742	8.85	
		EIF3D_HUMAN	45	514	6.23
			63	90	5.62
	68		90	7.61	
	196		90	8.71	
	563		458	8.42	
	EIF3F_HUMAN	308	126	11.37	
		388	126	5.81	
		532	238	7.28	
		538	297	5.10	
	EIF3G_HUMAN	632	209	7.39	
		689	148	6.02	
	EIF3H_HUMAN	532	221	9.97	
		538	221	8.88	
		559	5	7.13	
		563	5	16.70	
		563	24	22.07	
		563	28	8.62	
		591	5	6.58	

protein 1	protein 2	residue 1	residue 2	max of score
		591	24	14.82
		591	28	5.97
		591	178	9.20
		591	180	7.00
		603	5	5.54
		603	24	5.57
		606	5	8.45
		606	24	12.28
		606	306	6.84
	EIF3M_HUMAN	532	183	12.53
		532	319	7.34
	EIF4B_HUMAN	563	577	8.67
	RS2_HUMAN	632	275	12.66
		645	76	6.01
		645	176	8.31
		645	275	6.25
	RS21_HUMAN	632	41	9.67
	RS3A_HUMAN	2	144	13.08
		63	199	6.11
EIF3B_HUMAN	EIF3A_HUMAN	345	720	5.46
		459	742	7.21
		718	694	12.39
		722	694	7.58
		729	694	8.44
	EIF3G_HUMAN	693	148	9.05
		693	153	8.95
	EIF3I_HUMAN	366	185	5.25
		693	264	5.49
		729	282	8.27
	RS9_HUMAN	552	139	12.69
		552	155	9.99
		599	155	7.47
EIF3C_HUMAN	EIF1_HUMAN	267	58	7.20
	EIF2G_HUMAN	259	303	5.50
	EIF3A_HUMAN	643	196	10.54
		742	351	8.85
		760	191	5.43
		764	191	5.64
	EIF3D_HUMAN	627	58	27.37
		643	90	5.35
		664	90	6.87
		712	53	10.81
		712	58	14.92
		712	90	5.12
		889	58	5.65
		894	58	13.58
	EIF3E_HUMAN	627	275	6.89
		627	279	10.73
		627	289	6.65
		889	409	5.52
		889	424	7.90

protein 1	protein 2	residue 1	residue 2	max of score
		889	425	9.90
		894	424	15.56
		894	425	10.84
	EIF3H_HUMAN	889	274	8.92
		894	274	8.86
	EIF3L_HUMAN	894	347	5.24
		894	549	8.37
	RS13_HUMAN	312	100	7.93
		331	93	6.18
		331	100	6.38
		373	43	11.33
	RS3A_HUMAN	312	128	5.58
	RS7_HUMAN	267	115	7.04
EIF3D_HUMAN	EIF3A_HUMAN	90	63	5.62
		90	68	7.61
		90	196	8.71
		458	563	8.42
		514	45	6.23
	EIF3C_HUMAN	53	712	10.81
		58	627	27.37
		58	712	14.92
		58	889	5.65
		58	894	13.58
		90	643	5.35
		90	664	6.87
		90	712	5.12
	EIF3H_HUMAN	53	269	7.72
		58	269	13.04
		58	306	16.37
		90	306	5.92
	EIF4B_HUMAN	58	223	6.19
		58	365	8.46
	RS14_HUMAN	181	86	10.14
	RS26_HUMAN	142	38	5.59
	RS28_HUMAN	142	16	9.00
		181	10	6.09
EIF3E_HUMAN	EIF3C_HUMAN	275	627	6.89
		279	627	10.73
		289	627	6.65
		409	889	5.52
		424	889	7.90
		424	894	15.56
		425	889	9.90
		425	894	10.84
	EIF3H_HUMAN	425	188	12.07
	EIF3L_HUMAN	265	403	7.11
		387	474	8.25
		409	302	5.20
		409	347	11.58
		409	393	8.53
		409	549	11.17

protein 1	protein 2	residue 1	residue 2	max of score
		424	101	9.24
		424	302	9.67
		424	347	15.61
		425	101	8.66
	RACK1_HUMAN	82	139	9.00
EIF3F_HUMAN	EIF3A_HUMAN	126	308	11.37
		126	388	5.81
		238	532	7.28
		297	538	5.10
	EIF3H_HUMAN	214	221	14.64
		238	221	13.82
	EIF3M_HUMAN	214	115	10.38
		238	149	5.95
		238	177	5.39
		297	344	5.64
EIF3G_HUMAN	EIF3A_HUMAN	148	689	6.02
		209	632	7.39
	EIF3B_HUMAN	148	693	9.05
		153	693	8.95
	EIF3I_HUMAN	71	2	7.41
		134	282	5.69
		148	264	8.87
		161	264	6.06
		161	282	5.30
	RS3_HUMAN	280	90	9.29
EIF3H_HUMAN	EIF1AX_HUMAN	189	68	11.57
	EIF3A_HUMAN	5	559	7.13
		5	563	16.70
		5	591	6.58
		5	603	5.54
		5	606	8.45
		24	563	22.07
		24	591	14.82
		24	603	5.57
		24	606	12.28
		28	563	8.62
		28	591	5.97
		178	591	9.20
		180	591	7.00
		221	532	9.97
		221	538	8.88
		306	606	6.84
	EIF3C_HUMAN	274	889	8.92
		274	894	8.86
	EIF3D_HUMAN	269	53	7.72
		269	58	13.04
		306	58	16.37
		306	90	5.92
	EIF3E_HUMAN	188	425	12.07
	EIF3F_HUMAN	221	214	14.64
		221	238	13.82

protein 1	protein 2	residue 1	residue 2	max of score
	EIF3L_HUMAN	5	101	5.81
		5	302	11.87
		5	559	10.71
		24	559	13.31
		28	559	9.81
		172	559	10.98
	EIF3M_HUMAN	220	149	6.66
		221	115	10.18
		221	149	12.66
		221	177	9.10
	EIF4A1_HUMAN	5	2	12.54
	RS17_HUMAN	303	103	5.45
EIF3I_HUMAN	EIF3B_HUMAN	185	366	5.25
		264	693	5.49
		282	729	8.27
	EIF3G_HUMAN	2	71	7.41
		264	148	8.87
		264	161	6.06
		282	134	5.69
		282	161	5.30
EIF3K_HUMAN	EIF3L_HUMAN	16	237	5.16
EIF3L_HUMAN	EIF3C_HUMAN	347	894	5.24
		549	894	8.37
	EIF3E_HUMAN	101	424	9.24
		101	425	8.66
		302	409	5.20
		302	424	9.67
		347	409	11.58
		347	424	15.61
		393	409	8.53
		403	265	7.11
		474	387	8.25
		549	409	11.17
	EIF3H_HUMAN	101	5	5.81
		302	5	11.87
		559	5	10.71
		559	24	13.31
		559	28	9.81
		559	172	10.98
	EIF3K_HUMAN	237	16	5.16
	EIF4B_HUMAN	549	223	5.87
EIF3M_HUMAN	EIF3A_HUMAN	183	532	12.53
		319	532	7.34
	EIF3F_HUMAN	115	214	10.38
		149	238	5.95
		177	238	5.39
		344	297	5.64
	EIF3H_HUMAN	115	221	10.18
		149	220	6.66
		149	221	12.66

protein 1	protein 2	residue 1	residue 2	max of score
		177	221	9.10
EIF4A1_HUMAN	EIF3H_HUMAN	2	5	12.54
	EIF4B_HUMAN	2	223	7.29
		54	365	6.52
		369	343	8.95
		369	365	8.31
EIF4B_HUMAN	EIF3A_HUMAN	577	563	8.67
	EIF3D_HUMAN	223	58	6.19
		365	58	8.46
	EIF3L_HUMAN	223	549	5.87
	EIF4A1_HUMAN	223	2	7.29
		343	369	8.95
		365	54	6.52
		365	369	8.31
	RACK1_HUMAN	365	139	5.94
		511	139	15.64
		511	185	13.17
		511	264	12.01
	RS12_HUMAN	177	116	6.75
RACK1_HUMAN	EIF3E_HUMAN	139	82	9.00
	EIF4B_HUMAN	139	365	5.94
		139	511	15.64
		185	511	13.17
		264	511	12.01
RS10_HUMAN	EIF1AX_HUMAN	31	88	5.08
RS12_HUMAN	EIF4B_HUMAN	116	177	6.75
RS13_HUMAN	EIF3C_HUMAN	43	373	11.33
		93	331	6.18
		100	312	7.93
		100	331	6.38
RS14_HUMAN	EIF2A_HUMAN	63	97	5.15
	EIF3D_HUMAN	86	181	10.14
RS17_HUMAN	EIF3H_HUMAN	103	303	5.45
		129	40	5.49
RS2_HUMAN	EIF3A_HUMAN	76	645	6.01
		176	645	8.31
		275	632	12.66
		275	645	6.25
RS21_HUMAN	EIF3A_HUMAN	41	632	9.67
RS26_HUMAN	EIF3D_HUMAN	38	142	5.59
RS28_HUMAN	EIF3D_HUMAN	10	181	6.09
		16	142	9.00
RS3_HUMAN	EIF3G_HUMAN	90	280	9.29
RS30_HUMAN	EIF1AX_HUMAN	1	16	9.38
		1	23	13.63
		1	68	8.23
		1	88	5.75
RS3A_HUMAN	EIF3A_HUMAN	144	2	13.08
		199	63	6.11
	EIF3C_HUMAN	128	312	5.58

protein 1	protein 2	residue 1	residue 2	max of score
RS7_HUMAN	EIF3C_HUMAN	115	267	7.04
RS9_HUMAN	EIF3B_HUMAN	139	552	12.69
		155	552	9.99
		155	599	7.47

Table S2: List of identified 48S-IC cross-links. This table contains a filtered list of cross-links found within the 48S-IC when using the crosslinker BS3. From the original list, all cross-links with a score of less than 5 as well as cross-links between two 40S ribosomal proteins were omitted for clarity. Protein 1 and 2 are the proteins between which the cross-link was found, while residue 1 is the sequential position of protein 1, residue 2 on protein 2.

B.2 Figures

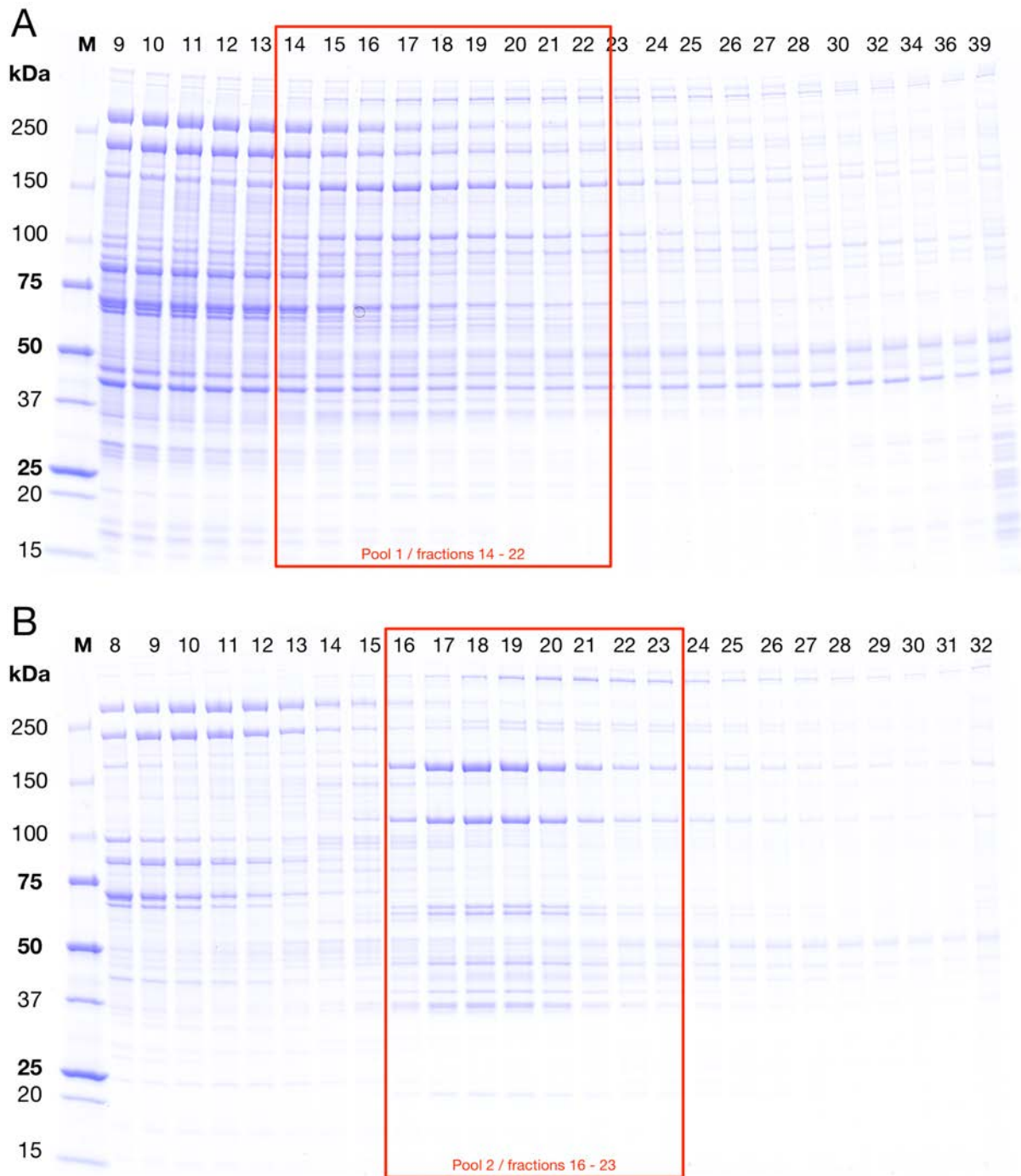


Fig. S1: SDS-PAGE of eIF3 purification gradients. After fractionation of each sedimentation gradient, an SDS-PAGE was performed. The most interesting fractions were loaded onto gels, which are shown here. **A** shows the gel after the first gradient from an SW32 rotor. The pooled fraction used for further precipitation, containing eIF3, is marked in red. **B** shows the gel from the second gradient, made with an SW40 rotor. Again, the pooled fractions are shown in red. Further information can be found in section 3.1.1 and on figure 3.2, page 50.

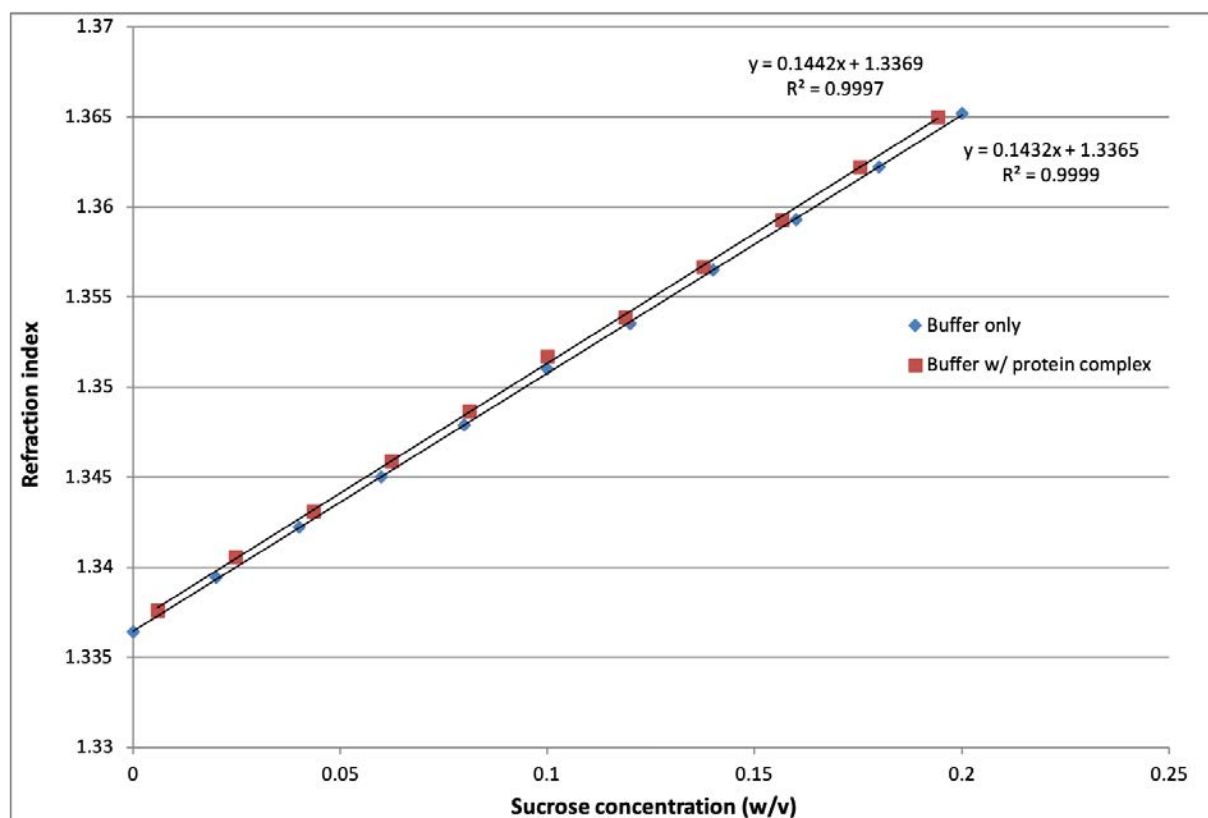


Fig. S2: Calibration curve for sucrose concentration. To investigate the dependency of sucrose concentration (w/v) and RI for both, a sample with (in red) and without (in blue) protein complex (here, glutamate dehydrogenase was used), a series of standards with known sucrose concentrations was measured. The protein containing sample had the same amount of protein in all samples (0.1 mg/mL). Linear regression was used for interpolation. The refractometer used (DR201-95, Krüss) has an accuracy of measurement of ± 0.0002 RI. Mixing of the protein stock solution with the sucrose containing standard buffers led to a slightly different final sucrose concentration for the protein containing testing samples. This did not impact the slope of the curve, but rather induced a slight increase in basal signal.

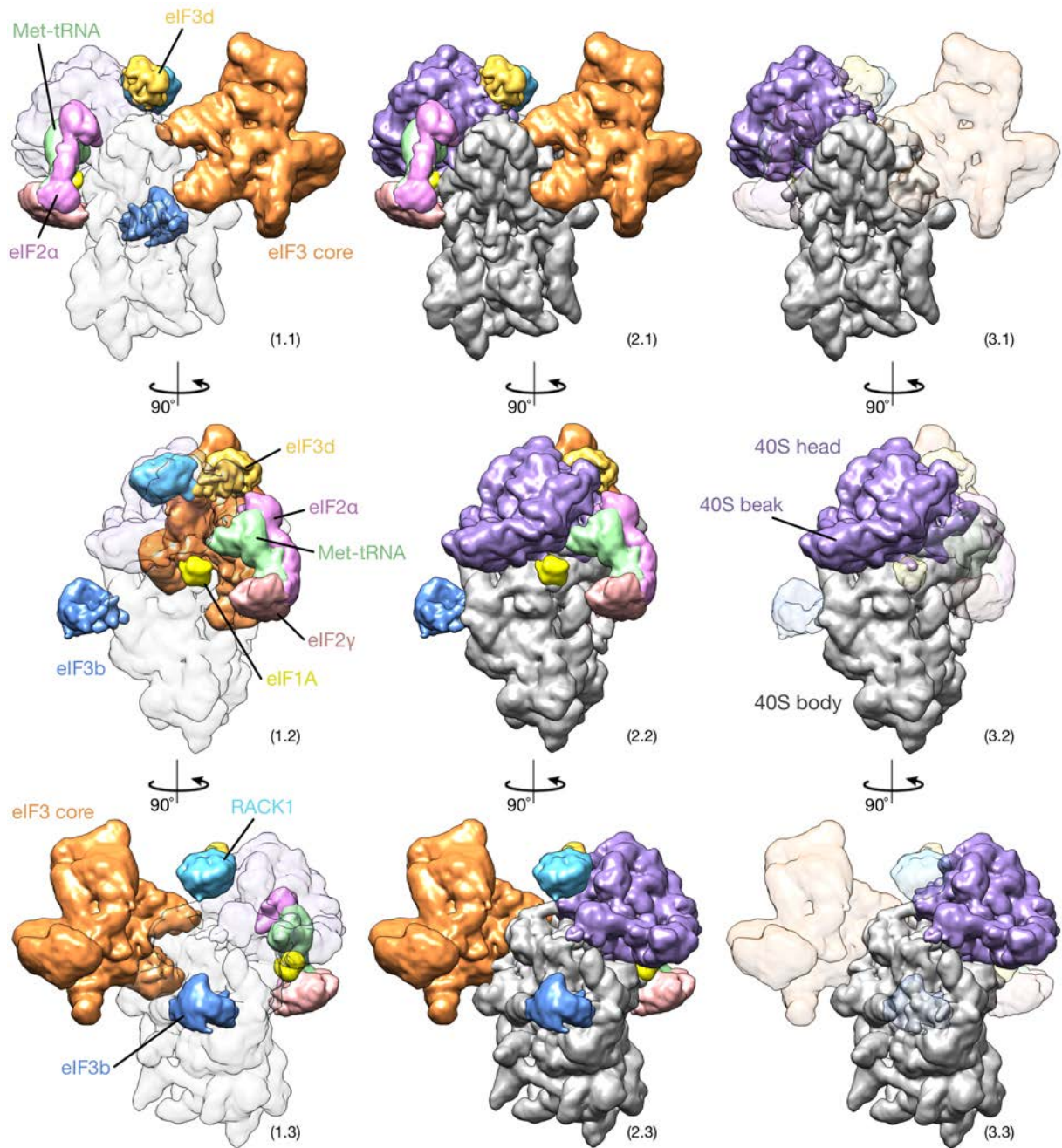


Fig. S3: Segmented 48S-IC structure with factor annotations — extended. Here, the result of the final single-class 3D classification from the starting model reconstruction (see section 3.3.3.1) is shown. The structure was segmented according to its components, following the information taken from previously published data of the mammalian 43S pre-IC [149], the yeast 48S pre-IC [150], and the yeast 40S-eIF1-eIF1A-eIF3-eIF3j complex [151]. In structures 1.1, 1.2, and 1.3, the 40S ribosomal subunit is shown transparent and only all other visible factors are depicted in color with annotations. Structures 3.1, 3.2, and 3.3 show the same orientations, but with inverted transparency. Here, the 40S ribosomal subunit is colored and its head, including the beak, as well as the body are shown. In structures 2.1, 2.2, and 2.3, all components are colored.

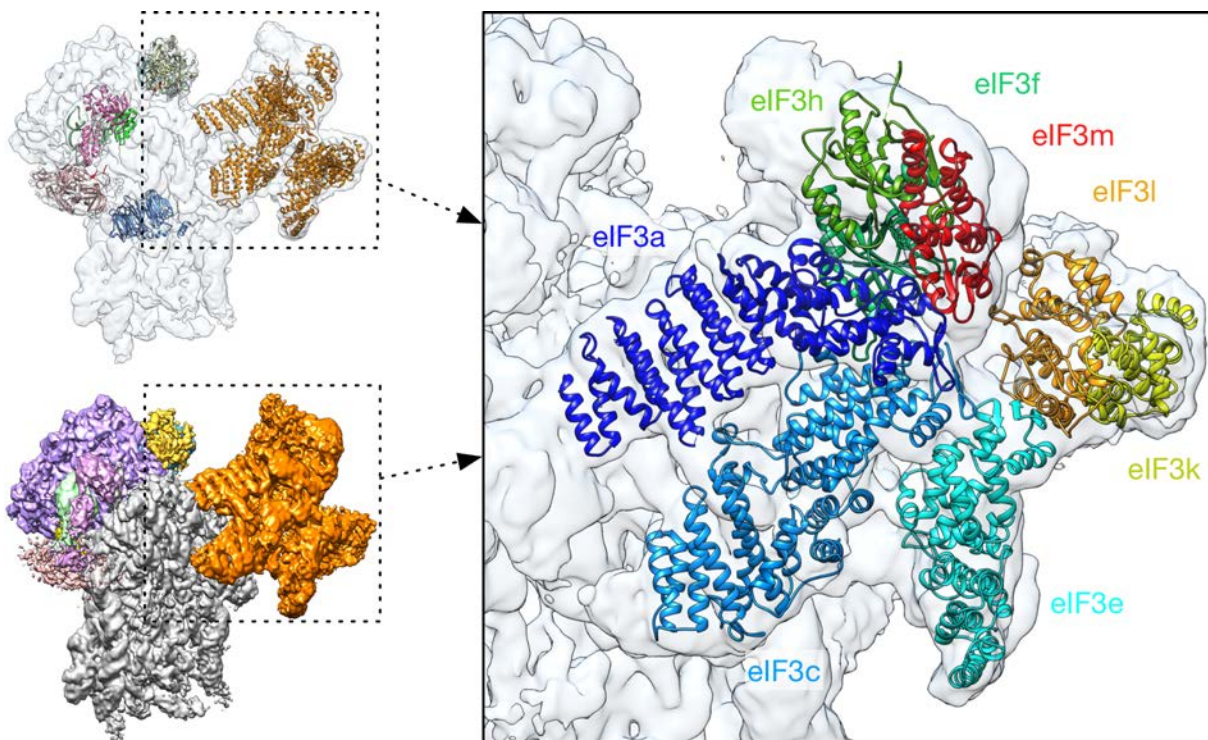


Fig. S4: eIF3 core components and rigid body fitting. Here, the eIF3 core domain is shown in more details with its eight components labeled in different colors to distinguish them. In all other figures, the eIF3 core will be shown in one color for the sake of clarity although it actually is always composed of these eight subunits.

C References

- [1] Reece, J. B. *et al.* *Campbell biology* (Pearson, 2016).
- [2] Crick, F. Central dogma of molecular biology. *Nature* 227, 561–3 (1970). <http://doi.org/10.1038/227561a0>.
- [3] Alberts, B. *Molecular biology of the cell* (2014).
- [4] Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701–18 (2009). <http://doi.org/10.1016/j.cell.2009.02.009>.
- [5] Ramakrishnan, V. Ribosome structure and the mechanism of translation. *Cell* 108, 557–72 (2002). [http://doi.org/10.1016/S0092-8674\(02\)00619-0](http://doi.org/10.1016/S0092-8674(02)00619-0).
- [6] Fischer, N., Konevega, A. L., Wintermeyer, W., Rodnina, M. V. & Stark, H. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature* 466, 329–33 (2010). <http://doi.org/10.1038/nature09206>.
- [7] Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289, 905–20 (2000). <http://doi.org/10.1126/science.289.5481.905>.
- [8] Wimberly, B. T. *et al.* Structure of the 30S ribosomal subunit. *Nature* 407, 327–39 (2000). <http://doi.org/10.1038/35030006>.
- [9] Schmeing, T. M. & Ramakrishnan, V. What recent ribosome structures have revealed about the mechanism of translation. *Nature* 461, 1234–42 (2009). <http://doi.org/10.1038/nature08403>.
- [10] Hinnebusch, A. G. The scanning mechanism of eukaryotic translation initiation. *Annu. Rev. Biochem.* 83, 779–812 (2014). <http://doi.org/10.1146/annurev-biochem-060713-035802>.
- [11] Voorhees, R. M. & Ramakrishnan, V. Structural basis of the translational elongation cycle. *Annu. Rev. Biochem.* 82, 203–36 (2013). <http://doi.org/10.1146/annurev-biochem-113009-092313>.
- [12] Valášek, L. S. 'Ribozoomin'–translation initiation from the perspective of the ribosome-bound eukaryotic initiation factors (eIFs). *Curr. Protein Pept. Sci.* 13, 305–30 (2012). <http://doi.org/10.2174/138920312801619385>.
- [13] Rodnina, M. V. & Wintermeyer, W. Recent mechanistic insights into eukaryotic ribosomes. *Curr. Opin. Cell Biol.* 21, 435–43 (2009). <http://doi.org/10.1016/j.ceb.2009.01.023>.
- [14] Hinnebusch, A. G. Structural Insights into the Mechanism of Scanning and Start Codon Recognition in Eukaryotic Translation Initiation. *Trends Biochem. Sci.* xx, 1–23 (2017). <http://doi.org/10.1016/j.tibs.2017.03.004>.
- [15] Pelletier, J. & Sonenberg, N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* 334, 320–5 (1988). <http://doi.org/10.1038/334320a0>.

- [16] Erickson, F. L. & Hannig, E. M. Ligand interactions with eukaryotic translation initiation factor 2: role of the gamma-subunit. *EMBO J.* 15, 6311–20 (1996).
- [17] Kapp, L. D. & Lorsch, J. R. GTP-dependent recognition of the methionine moiety on initiator tRNA by translation factor eIF2. *J. Mol. Biol.* 335, 923–36 (2004). <http://doi.org/10.1016/j.jmb.2003.11.025>.
- [18] Trachsel, H., Erni, B., Schreier, M. H. & Staehelin, T. Initiation of mammalian protein synthesis. II. The assembly of the initiation complex with purified initiation factors. *J. Mol. Biol.* 116, 755–67 (1977). [http://doi.org/10.1016/0022-2836\(77\)90269-8](http://doi.org/10.1016/0022-2836(77)90269-8).
- [19] Majumdar, R., Bandyopadhyay, A. & Maitra, U. Mammalian translation initiation factor eIF1 functions with eIF1A and eIF3 in the formation of a stable 40 S preinitiation complex. *J. Biol. Chem.* 278, 6580–7 (2003). <http://doi.org/10.1074/jbc.M210357200>.
- [20] Maag, D. & Lorsch, J. R. Communication between eukaryotic translation initiation factors 1 and 1A on the yeast small ribosomal subunit. *J. Mol. Biol.* 330, 917–24 (2003). [http://doi.org/10.1016/S0022-2836\(03\)00665-X](http://doi.org/10.1016/S0022-2836(03)00665-X).
- [21] Sokabe, M., Fraser, C. S. & Hershey, J. W. B. The human translation initiation multi-factor complex promotes methionyl-tRNA_i binding to the 40S ribosomal subunit. *Nucleic Acids Res.* 40, 905–13 (2012). <http://doi.org/10.1093/nar/gkr772>.
- [22] Jackson, R. J., Hellen, C. U. T. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* 11, 113–27 (2010). <http://doi.org/10.1038/nrm2838>.
- [23] Linder, P. *et al.* Birth of the D-E-A-D box. *Nature* 337, 121–2 (1989). <http://doi.org/10.1038/337121a0>.
- [24] Imataka, H. & Sonenberg, N. Human eukaryotic translation initiation factor 4G (eIF4G) possesses two separate and independent binding sites for eIF4A. *Mol. Cell. Biol.* 17, 6940–7 (1997). <http://doi.org/10.1128/MCB.17.12.6940>.
- [25] Schütz, P. *et al.* Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* 105, 9564–9 (2008). <http://doi.org/10.1073/pnas.0800418105>.
- [26] Kozak, M. Effects of long 5' leader sequences on initiation by eukaryotic ribosomes in vitro. *Gene Expr.* 1, 117–25 (1991).
- [27] Lomakin, I. B. & Steitz, T. A. The initiation of mammalian protein synthesis and mRNA scanning mechanism. *Nature* 500, 307–11 (2013). <http://doi.org/10.1038/NATURE12355>.
- [28] Lomakin, I. B., Shirokikh, N. E., Yusupov, M. M., Hellen, C. U. T. & Pestova, T. V. The fidelity of translation initiation: reciprocal activities of eIF1, IF3 and YciH. *EMBO J.* 25, 196–210 (2006). <http://doi.org/10.1038/sj.emboj.7600904>.
- [29] Nanda, J. S., Saini, A. K., Muñoz, A. M., Hinnebusch, A. G. & Lorsch, J. R. Co-ordinated movements of eukaryotic translation initiation factors eIF1, eIF1A, and eIF5 trigger phosphate release from eIF2 in response to start codon recognition by the ribosomal preinitiation complex. *J. Biol. Chem.* 288, 5316–29 (2013). <http://doi.org/10.1074/jbc.M112.440693>.
- [30] des Georges, A. *et al.* Structure of mammalian eIF3 in the context of the 43S preinitiation complex. *Nature* 525, 491–5 (2015). <http://doi.org/10.1038/nature14891>.

-
- [31] Hussain, T. *et al.* Structural changes enable start codon recognition by the eukaryotic translation initiation complex. *Cell* 159, 597–607 (2014). <http://doi.org/10.1016/j.cell.2014.10.001>.
- [32] Llácer, J. L. *et al.* Conformational Differences between Open and Closed States of the Eukaryotic Translation Initiation Complex. *Mol. Cell* 59, 399–412 (2015). <http://doi.org/10.1016/j.molcel.2015.06.033>.
- [33] Algire, M. A., Maag, D. & Lorsch, J. R. Pi release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation. *Mol. Cell* 20, 251–62 (2005). <http://doi.org/10.1016/j.molcel.2005.09.008>.
- [34] Eliseev, B. *et al.* Structure of a human cap-dependent 48S translation pre-initiation complex. *Nucleic Acids Res.* 46, 2678–2689 (2018). <http://doi.org/10.1093/nar/gky054>.
- [35] Hinnebusch, A. G. eIF3: a versatile scaffold for translation initiation complexes. *Trends Biochem. Sci.* 31, 553–62 (2006). <http://doi.org/10.1016/j.tibs.2006.08.005>.
- [36] Wagner, S., Herrmannová, A., Malík, R., Peclinovská, L. & Valášek, L. S. Functional and biochemical characterization of human eukaryotic translation initiation factor 3 in living cells. *Mol. Cell. Biol.* 34, 3041–52 (2014). <http://doi.org/10.1128/MCB.00663-14>.
- [37] Querol-Audi, J. *et al.* Architecture of human translation initiation factor 3. *Structure* 21, 920–8 (2013). <http://doi.org/10.1016/j.str.2013.04.002>.
- [38] Wagner, S., Herrmannová, A., Šikrová, D. & Valášek, L. S. Human eIF3b and eIF3a serve as the nucleation core for the assembly of eIF3 into two interconnected modules: the yeast-like core and the octamer. *Nucleic Acids Res.* 44, 10772–10788 (2016). <http://doi.org/10.1093/nar/gkw972>.
- [39] Pisarev, A. V., Kolupaeva, V. G., Yusupov, M. M., Hellen, C. U. T. & Pestova, T. V. Ribosomal position and contacts of mRNA in eukaryotic translation initiation complexes. *EMBO J.* 27, 1609–1621 (2008). <http://doi.org/10.1038/emboj.2008.90>.
- [40] Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* 10, 980 (2003). <http://doi.org/10.1038/nsb1203-980>.
- [41] Liu, Y. *et al.* Translation initiation factor eIF3b contains a nine-bladed β -propeller and interacts with the 40S ribosomal subunit. *Structure* 22, 923–30 (2014). <http://doi.org/10.1016/j.str.2014.03.010>.
- [42] Herrmannová, A. *et al.* Structural analysis of an eIF3 subcomplex reveals conserved interactions required for a stable and proper translation pre-initiation complex assembly. *Nucleic Acids Res.* 40, 2294–311 (2012). <http://doi.org/10.1093/nar/gkr765>.
- [43] Wei, Z. *et al.* Crystal structure of human eIF3k, the first structure of eIF3 subunits. *J. Biol. Chem.* 279, 34983–90 (2004). <http://doi.org/10.1074/jbc.M405158200>.
- [44] Lee, A. S. Y., Kranzusch, P. J., Doudna, J. A. & Cate, J. H. D. eIF3d is an mRNA cap-binding protein that is required for specialized translation initiation. *Nature* 536, 96–9 (2016). <http://doi.org/10.1038/nature18954>.
- [45] Sinz, A., Arlt, C., Chorev, D. & Sharon, M. Chemical cross-linking and native mass spectrometry: A fruitful combination for structural biology. *Protein Sci.* 24, 1193–209 (2015). <http://doi.org/10.1002/pro.2696>.

- [46] Elantak, L. *et al.* 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, 1097–116 (2010). <http://doi.org/10.1016/j.jmb.2009.12.047>.
- [47] Erzberger, J. P. *et al.* Molecular architecture of the 40S-eIF1-eIF3 translation initiation complex. *Cell* 158, 1123–1135 (2014). <http://doi.org/10.1016/j.cell.2014.07.044>.
- [48] De Broglie, L. Recherches sur la théorie des Quanta. *Ann. Phys. (Paris)*. 10, 22–128 (1925). <http://doi.org/10.1051/anphys/192510030022>.
- [49] Sievers, H. Louis de Broglie und die Quantenmechanik (1998).
- [50] Williams, D. B. & Carter, C. B. *Transmission Electron Microscopy: A Textbook for Materials Science* (Springer, 1996).
- [51] Hawkes, P. W. Ernst Ruska. *Phys. Today* 43, 84–85 (1990). <http://doi.org/10.1063/1.2810640>.
- [52] Stokes, D. & Royal Microscopical Society (Great Britain). *Principles and practice of variable pressure/environmental scanning electron microscopy (VP-ESEM)* (Wiley, 2008).
- [53] Baker, L. A. & Rubinstein, J. L. Radiation damage in electron cryomicroscopy. In *Methods Enzymol.*, vol. 481, 371–388 (Elsevier Inc., 2010). [http://doi.org/10.1016/S0076-6879\(10\)81015-8](http://doi.org/10.1016/S0076-6879(10)81015-8).
- [54] De Rosier, D. J. & Klug, A. Reconstruction of three dimensional structures from electron micrographs. *Nature* 217, 130–4 (1968). <http://doi.org/10.1038/217130a0>.
- [55] Klug, A. & Finch, J. T. Structure of viruses of the papilloma-polyoma type. *J. Mol. Biol.* 11, 424–IN45 (1965). [http://doi.org/10.1016/S0022-2836\(65\)80067-5](http://doi.org/10.1016/S0022-2836(65)80067-5).
- [56] Crowther, R. A., Amos, L. A., Finch, J. T., De Rosier, D. J. & Klug, A. Three dimensional reconstructions of spherical viruses by fourier synthesis from electron micrographs. *Nature* 226, 421–5 (1970). <http://doi.org/10.1038/226421a0>.
- [57] Crowther, R. A., DeRosier, D. J. & Klug, A. The Reconstruction of a Three-Dimensional Structure from Projections and its Application to Electron Microscopy. *Proc. R. Soc. A Math. Phys. Eng. Sci.* 317, 319–340 (1970). <http://doi.org/10.1098/rspa.1970.0119>.
- [58] Dubochet, J. & McDowell, A. VITRIFICATION OF PURE WATER FOR ELECTRON MICROSCOPY. *J. Microsc.* 124, 3–4 (1981). <http://doi.org/10.1111/j.1365-2818.1981.tb02483.x>.
- [59] Frank, J. Averaging of low exposure electron micrographs of non-periodic objects. *Ultramicroscopy* 1, 159–62 (1975). [http://doi.org/10.1016/S0304-3991\(75\)80020-9](http://doi.org/10.1016/S0304-3991(75)80020-9).
- [60] Van Heel, M. & Frank, J. Use of multivariate statistics in analysing the images of biological macromolecules. *Ultramicroscopy* 6, 187–194 (1981). [http://doi.org/10.1016/0304-3991\(81\)90059-0](http://doi.org/10.1016/0304-3991(81)90059-0).
- [61] Frank, J. & van Heel, M. Correspondence analysis of aligned images of biological particles. *J. Mol. Biol.* 161, 134–7 (1982). [http://doi.org/10.1016/0022-2836\(82\)90282-0](http://doi.org/10.1016/0022-2836(82)90282-0).
- [62] van Heel, M. Multivariate statistical classification of noisy images (randomly oriented biological macromolecules). *Ultramicroscopy* 13, 165–183 (1984). [http://doi.org/10.1016/0304-3991\(84\)90066-4](http://doi.org/10.1016/0304-3991(84)90066-4).

-
- [63] van Heel, M. Imagic and its results. *Ultramicroscopy* 4, 117 (1979). [http://doi.org/10.1016/0304-3991\(79\)90026-3](http://doi.org/10.1016/0304-3991(79)90026-3).
- [64] Frank, J., Shimkin, B. & Dowse, H. SPIDER - A modular software system for electron image processing. *Ultramicroscopy* 6, 343–357 (1981). [http://doi.org/10.1016/S0304-3991\(81\)80221-5](http://doi.org/10.1016/S0304-3991(81)80221-5).
- [65] Henderson, R. *et al.* Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* 213, 899–929 (1990). [http://doi.org/10.1016/S0022-2836\(05\)80271-2](http://doi.org/10.1016/S0022-2836(05)80271-2).
- [66] Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* 28, 171 (1995). <http://doi.org/10.1017/S003358350000305X>.
- [67] Faruqi, A. R. & Subramaniam, S. CCD detectors in high-resolution biological electron microscopy. *Q. Rev. Biophys.* 33, 1–27 (2000). <http://doi.org/10.1017/S0033583500003577>.
- [68] Faruqi, A. & Henderson, R. Electronic detectors for electron microscopy. *Curr. Opin. Struct. Biol.* 17, 549–555 (2007). <http://doi.org/10.1016/j.sbi.2007.08.014>.
- [69] Faruqi, A. R., Andrews, H., Cattermole, D. & Stubbings, S. A tiled CCD detector with 2CE2 array and tapered fibre optics for electron microscopy. *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.* 477, 137–142 (2002). [http://doi.org/10.1016/S0168-9002\(01\)01883-6](http://doi.org/10.1016/S0168-9002(01)01883-6).
- [70] Xuong, N. *et al.* Future directions for camera systems in electron microscopy. *Methods Cell Biol.* 79, 721–39 (2007). [http://doi.org/10.1016/S0091-679X\(06\)79028-8](http://doi.org/10.1016/S0091-679X(06)79028-8).
- [71] Mooney, P. Optimization of image collection for cellular electron microscopy. *Methods Cell Biol.* 79, 661–719 (2007). [http://doi.org/10.1016/S0091-679X\(06\)79027-6](http://doi.org/10.1016/S0091-679X(06)79027-6).
- [72] Kühlbrandt, W. The Resolution Revolution. *Science (80-.)*. 343, 1443–1444 (2014). <http://doi.org/10.1126/science.1251652>.
- [73] Faruqi, A. R. & McMullan, G. Electronic detectors for electron microscopy. *Q. Rev. Biophys.* 44, 357–390 (2011). <http://doi.org/10.1017/S0033583511000035>.
- [74] Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* 10, 584–590 (2013). <http://doi.org/10.1038/nmeth.2472>.
- [75] Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* 14, 331–332 (2017). <http://doi.org/10.1038/nmeth.4193>.
- [76] Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* 157, 38–46 (2007). <http://doi.org/10.1016/j.jsb.2006.05.009>.
- [77] Elmlund, D. & Elmlund, H. SIMPLE: Software for ab initio reconstruction of heterogeneous single-particles. *J. Struct. Biol.* 180, 420–7 (2012). <http://doi.org/10.1016/j.jsb.2012.07.010>.
- [78] Grigorieff, N. FREALIGN: high-resolution refinement of single particle structures. *J. Struct. Biol.* 157, 117–25 (2007). <http://doi.org/10.1016/j.jsb.2006.05.004>.

- [79] Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* 180, 519–30 (2012). <http://doi.org/10.1016/j.jsb.2012.09.006>.
- [80] Scheres, S. H. W. Semi-automated selection of cryo-EM particles in RELION-1.3. *J. Struct. Biol.* (2014). <http://doi.org/10.1016/j.jsb.2014.11.010>.
- [81] Haselbach, D. *et al.* Long-range allosteric regulation of the human 26S proteasome by 20S proteasome-targeting cancer drugs. *Nat. Commun.* 8, 15578 (2017). <http://doi.org/10.1038/ncomms15578>.
- [82] Haselbach, D. *et al.* Structure and Conformational Dynamics of the Human Spliceosomal B act Complex. *Cell* 1–11 (2018). <http://doi.org/10.1016/j.cell.2018.01.010>.
- [83] Stalling, D., Westerhoff, M. & Hege, H.-c. Amira: A Highly Interactive System for Visual Data Analysis. *Vis. Handb.* 1, 749–767 (2005). <http://doi.org/10.1.1.129.6785>.
- [84] Schmeisser, M. *et al.* Parallel, distributed and GPU computing technologies in single-particle electron microscopy. *Acta Crystallogr.* 65, 659–71 (2009). <http://doi.org/10.1107/S0907444909011433>.
- [85] Busche, B. *New Algorithms for Automated Processing of Electronmicroscopic Images*. Ph.d. thesis, Georg-August-University, Göttingen (2013).
- [86] Kirves, J.-M. *Algorithms for Single Particle Cryo Electron Microscopic Image Processing*. Ph.d. thesis, Georg-August-University, Göttingen (2014).
- [87] Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* 193, 1–12 (2016). <http://doi.org/10.1016/j.jsb.2015.11.003>.
- [88] Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–5 (2012). <http://doi.org/10.1038/nmeth.2089>.
- [89] van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* 116, 17–24 (1996). <http://doi.org/10.1006/jsbi.1996.0004>.
- [90] Yang, B. *et al.* Identification of cross-linked peptides from complex samples. *Nat. Methods* 9, 904–6 (2012). <http://doi.org/10.1038/nmeth.2099>.
- [91] Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* 11, 63–5 (2014). <http://doi.org/10.1038/nmeth.2727>.
- [92] Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–12 (2004). <http://doi.org/10.1002/jcc.20084>.
- [93] Combe, C. W., Fischer, L. & Rappsilber, J. xiNET: cross-link network maps with residue resolution. *Mol. Cell. Proteomics* 14, 1137–47 (2015). <http://doi.org/10.1074/mcp.0114.042259>.
- [94] Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* 11, 1475–89 (1983). <http://doi.org/10.1093/nar/11.5.1475>.
- [95] Oxenburgh, M. S. & Snoswell, A. M. Use of streptomycin in the separation of nucleic acids from protein in a bacterial extract. *Nature* 207, 1416–7 (1965). <http://doi.org/10.1038/2071416a0>.

-
- [96] Weber, K. & Osborn, M. The reliability of molecular weight determinations by dodecyl sulfate-polyacrylamide gel electrophoresis. *J. Biol. Chem.* 244, 4406–12 (1969).
- [97] Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–54 (1976). [http://doi.org/10.1016/0003-2697\(76\)90527-3](http://doi.org/10.1016/0003-2697(76)90527-3).
- [98] Pisarev, A. V., Unbehaun, A., Hellen, C. U. T. & Pestova, T. V. Assembly and analysis of eukaryotic translation initiation complexes. *Methods Enzymol.* 430, 147–77 (2007). [http://doi.org/10.1016/S0076-6879\(07\)30007-4](http://doi.org/10.1016/S0076-6879(07)30007-4).
- [99] Reimer, L. & Kohl, H. *Transmission Electron Microscopy: Physics of Image Formation*. Springer Series in Optical Sciences (Springer, 2008).
- [100] Baker, T. S. & Henderson, R. Electron cryomicroscopy of biological macromolecules. In *Int. Tables Crystallogr.*, vol. F, 593–614 (International Union of Crystallography, Chester, England, 2012). <http://doi.org/10.1107/97809553602060000872>.
- [101] Moody, M. F. *Structural Biology - Using Electrons and X-Rays: An Introduction for Biologists* (Academic Pr Inc, 2011).
- [102] Wade, R. A brief look at imaging and contrast transfer. *Ultramicroscopy* 46, 145–156 (1992). [http://doi.org/10.1016/0304-3991\(92\)90011-8](http://doi.org/10.1016/0304-3991(92)90011-8).
- [103] McMullan, G., Faruqi, A. R., Clare, D. & Henderson, R. Comparison of optimal performance at 300keV of three direct electron detectors for use in low dose electron microscopy. *Ultramicroscopy* 147, 156–63 (2014). <http://doi.org/10.1016/j.ultramic.2014.08.002>.
- [104] Sander, B., Golas, M. M. & Stark, H. Advantages of CCD detectors for de novo three-dimensional structure determination in single-particle electron microscopy. *J. Struct. Biol.* 151, 92–105 (2005). <http://doi.org/10.1016/j.jsb.2005.04.004>.
- [105] Jha, P., Wang, X. & Auwerx, J. Analysis of Mitochondrial Respiratory Chain Supercomplexes Using Blue Native Polyacrylamide Gel Electrophoresis (BN-PAGE). *Curr. Protoc. Mouse Biol.* 6, 1–14 (2016). <http://doi.org/10.1002/9780470942390.mo150182>.
- [106] Li, X., Zheng, S. Q., Egami, K., Agard, D. A. & Cheng, Y. Influence of electron dose rate on electron counting images recorded with the K2 camera. *J. Struct. Biol.* 184, 251–60 (2013). <http://doi.org/10.1016/j.jsb.2013.08.005>.
- [107] Bartesaghi, A. *et al.* 2.2 Å resolution cryo-EM structure of β -galactosidase in complex with a cell-permeant inhibitor. *Science (80-)*. 48, 826–840 (2015). <http://doi.org/10.1126/science.aab1576>.
- [108] Lu, P. *et al.* Three-dimensional structure of human γ -secretase. *Nature* 512, 166–170 (2014). <http://doi.org/10.1038/nature13567>.
- [109] Bammes, B. E., Rochat, R. H., Jakana, J., Chen, D.-H. & Chiu, W. Direct electron detection yields cryo-EM reconstructions at resolutions beyond 3/4 Nyquist frequency. *J. Struct. Biol.* 177, 589–601 (2012). <http://doi.org/10.1016/j.jsb.2012.01.008>.
- [110] Abbe, E. Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. *Arch. für Mikroskopische Anat.* 9, 413–418 (1873). <http://doi.org/10.1007/BF02956173>.

- [111] Chen, J. Z. & Grigorieff, N. SIGNATURE: a single-particle selection system for molecular electron microscopy. *J. Struct. Biol.* 157, 168–73 (2007). <http://doi.org/10.1016/j.jsb.2006.06.001>.
- [112] Voss, N. R., Yoshioka, C. K., Radermacher, M., Potter, C. S. & Carragher, B. DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. *J. Struct. Biol.* 166, 205–13 (2009). <http://doi.org/10.1016/j.jsb.2009.01.004>.
- [113] Langlois, R., Pallesen, J. & Frank, J. Reference-free particle selection enhanced with semi-supervised machine learning for cryo-electron microscopy. *J. Struct. Biol.* 175, 353–61 (2011). <http://doi.org/10.1016/j.jsb.2011.06.004>.
- [114] Joyeux, L. & Penczek, P. A. Efficiency of 2D alignment methods. *Ultramicroscopy* 92, 33–46 (2002). [http://doi.org/10.1016/S0304-3991\(01\)00154-1](http://doi.org/10.1016/S0304-3991(01)00154-1).
- [115] Sander, B., Golas, M. M. & Stark, H. Corrim-based alignment for improved speed in single-particle image processing 143, 219–228 (2003). <http://doi.org/10.1016/j.jsb.2003.08.001>.
- [116] Sigworth, F. J., Doerschuk, P. C., Carazo, J. M. & Scheres, S. H. W. An introduction to maximum-likelihood methods in cryo-EM. In *Methods Enzymol.*, vol. 482, 263–294 (Elsevier Inc., 2010), 1 edn. [http://doi.org/10.1016/S0076-6879\(10\)82011-7](http://doi.org/10.1016/S0076-6879(10)82011-7).
- [117] Henderson, R. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proc. Natl. Acad. Sci. U. S. A.* 110, 18037–41 (2013). <http://doi.org/10.1073/pnas.1314449110>.
- [118] Stewart, A. & Grigorieff, N. Noise bias in the refinement of structures derived from single particles. *Ultramicroscopy* 102, 67–84 (2004). <http://doi.org/10.1016/j.ultramic.2004.08.008>.
- [119] van Heel, M. Finding trimeric HIV-1 envelope glycoproteins in random noise. *Proc. Natl. Acad. Sci. U. S. A.* 110, E4175–7 (2013). <http://doi.org/10.1073/pnas.1314353110>.
- [120] Borland, L. & van Heel, M. Classification of image data in conjugate representation spaces. *J. Opt. Soc. Am. A* 7, 601 (1990). <http://doi.org/10.1364/JOSAA.7.000601>.
- [121] Sigworth, F. J. A maximum-likelihood approach to single-particle image refinement. *J. Struct. Biol.* 122, 328–39 (1998). <http://doi.org/10.1006/jsbi.1998.4014>.
- [122] Scheres, S. H. W. *et al.* Maximum-likelihood multi-reference refinement for electron microscopy images. *J. Mol. Biol.* 348, 139–49 (2005). <http://doi.org/10.1016/j.jmb.2005.02.031>.
- [123] Scheres, S. H. W. Classification of structural heterogeneity by maximum-likelihood methods. In *Methods Enzymol.*, vol. 482, 295–320 (Elsevier Inc., 2010), 1 edn. [http://doi.org/10.1016/S0076-6879\(10\)82012-9](http://doi.org/10.1016/S0076-6879(10)82012-9).
- [124] Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* 14, 290–296 (2017). <http://doi.org/10.1038/nmeth.4169>.
- [125] Garces, D. H., Rhodes, W. T. & Peña, N. M. Projection-slice theorem: a compact notation. *J. Opt. Soc. Am. A. Opt. Image Sci. Vis.* 28, 766–9 (2011). <http://doi.org/10.1364/JOSAA.28.000766>.

-
- [126] Orlova, E. V. & Saibil, H. R. Structural analysis of macromolecular assemblies by electron microscopy. *Chem. Rev.* 111, 7710–48 (2011). <http://doi.org/10.1021/cr100353t>.
- [127] van Heel, M. Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy* 21, 111–23 (1987).
- [128] Harauz, G. & van Heel, M. Exact filters for general geometry three dimensional reconstruction (1986).
- [129] Penczek, P. A. Fundamentals of Three-Dimensional reconstruction from projections. In *Methods Enzymol.*, vol. 482, 1–33 (Elsevier Inc., 2010), 1 edn. [http://doi.org/10.1016/S0076-6879\(10\)82001-4](http://doi.org/10.1016/S0076-6879(10)82001-4).
- [130] Henderson, R. *et al.* Outcome of the first electron microscopy validation task force meeting. *Structure* 20, 205–14 (2012). <http://doi.org/10.1016/j.str.2011.12.014>.
- [131] Penczek, P. A. Resolution measures in molecular electron microscopy. In *Methods Enzymol.*, vol. 482, 73–100 (Elsevier Inc., 2010), 1 edn. [http://doi.org/10.1016/S0076-6879\(10\)82003-8](http://doi.org/10.1016/S0076-6879(10)82003-8).
- [132] Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* 9, 853–4 (2012). <http://doi.org/10.1038/nmeth.2115>.
- [133] Russo, C. J. & Passmore, L. a. Robust evaluation of 3D electron cryomicroscopy data using tilt-pairs. *J. Struct. Biol.* 187, 112–118 (2014). <http://doi.org/10.1016/j.jsb.2014.06.006>.
- [134] Kastner, B. *et al.* GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* 5, 53–5 (2008). <http://doi.org/10.1038/nmeth1139>.
- [135] Stark, H. GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. In *Methods Enzymol.*, vol. 481, 109–26 (Elsevier Inc., 2010). [http://doi.org/10.1016/S0076-6879\(10\)81005-5](http://doi.org/10.1016/S0076-6879(10)81005-5).
- [136] Scheres, S. H. W. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. In Crowther, R. A. (ed.) *Methods Enzymol.*, vol. 579, chap. 6, 125–157 (Elsevier Inc., 2016). <http://doi.org/10.1016/bs.mie.2016.04.012>.
- [137] Scheres, S. H. W. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* 415, 406–18 (2012). <http://doi.org/10.1016/j.jmb.2011.11.010>.
- [138] van Heel, M. & Schatz, M. Fourier shell correlation threshold criteria. *J. Struct. Biol.* 151, 250–62 (2005). <http://doi.org/10.1016/j.jsb.2005.05.009>.
- [139] Schrader, J. *et al.* The inhibition mechanism of human 20S proteasomes enables next-generation inhibitor design. *Science (80-.)*. 353, 594–598 (2016). <http://doi.org/10.1126/science.aaf8993>.
- [140] Zhou, H.-X., Rivas, G. & Minton, A. P. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* 37, 375–97 (2008). <http://doi.org/10.1146/annurev.biophys.37.032807.125817>.
- [141] Lee, J. C. & Timasheff, S. N. The stabilization of proteins by sucrose. *J. Biol. Chem.* 256, 7193–201 (1981). [http://doi.org/10.1016/S0006-3495\(85\)83932-1](http://doi.org/10.1016/S0006-3495(85)83932-1).
- [142] Chari, A. *et al.* ProteoPlex: stability optimization of macromolecular complexes by sparse-matrix screening of chemical space. *Nat. Methods* 12, 859–65 (2015). <http://doi.org/10.1038/nmeth.3493>.

- [143] Sambrook, J. & Russell, D. W. D. W. *Molecular cloning : a laboratory manual* (Cold Spring Harbor Laboratory Press, 2001).
- [144] Keller, A., Nesvizhskii, A. I., Kolker, E. & Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 74, 5383–92 (2002). <http://doi.org/10.1021/ac025747h>.
- [145] Grassucci, R. A., Taylor, D. J. & Frank, J. Preparation of macromolecular complexes for cryo-electron microscopy. *Nat. Protoc.* 2, 3239–46 (2007). <http://doi.org/10.1038/nprot.2007.452>.
- [146] Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* 333, 721–45 (2003). <http://doi.org/10.1016/j.jmb.2003.07.013>.
- [147] Aitken, C. E. *et al.* Eukaryotic translation initiation factor 3 plays distinct roles at the mRNA entry and exit channels of the ribosomal preinitiation complex. *Elife* 5, e20934 (2016). <http://doi.org/10.7554/eLife.20934>.
- [148] Liu, W.-T. *Strategies to stabilize RNP complexes for structural determination by 3D cryo-electron microscopy*. Ph.d. thesis, Georg-August-University, Göttingen (2013).
- [149] Hashem, Y. *et al.* Structure of the mammalian ribosomal 43S preinitiation complex bound to the scanning factor DHX29. *Cell* 153, 1108–19 (2013). <http://doi.org/10.1016/j.cell.2013.04.036>.
- [150] Simonetti, A. *et al.* eIF3 Peripheral Subunits Rearrangement after mRNA Binding and Start-Codon Recognition. *Mol. Cell* 63, 206–17 (2016). <http://doi.org/10.1016/j.molcel.2016.05.033>.
- [151] Aylett, C. H. S., Boehringer, D., Erzberger, J. P., Schaefer, T. & Ban, N. Structure of a Yeast 40S-eIF1-eIF1A-eIF3-eIF3j initiation complex. *Nat. Struct. Mol. Biol.* 22, 269–271 (2015). <http://doi.org/10.1038/nsmb.2963>.
- [152] Natchiar, S. K., Myasnikov, A. G., Kratzat, H., Hazemann, I. & Klaholz, B. P. Visualization of chemical modifications in the human 80S ribosome structure. *Nature* 551, 472–477 (2017). <http://doi.org/10.1038/nature24482>.
- [153] Casey, J. R., Grinstein, S. & Orlowski, J. Sensors and regulators of intracellular pH. *Nat. Rev. Mol. Cell Biol.* 11, 50–61 (2010). <http://doi.org/10.1038/nrm2820>.
- [154] Smith, M. D. *et al.* Assembly of eIF3 Mediated by Mutually Dependent Subunit Insertion. *Structure* 1–11 (2016). <http://doi.org/10.1016/j.str.2016.02.024>.
- [155] Hussain, T., Llácer, J. L., Wimberly, B. T., Kieft, J. S. & Ramakrishnan, V. Large-Scale Movements of IF3 and tRNA during Bacterial Translation Initiation. *Cell* 167, 133–144.e13 (2016). <http://doi.org/10.1016/j.cell.2016.08.074>.
- [156] Mancera-Martínez, E., Brito Querido, J., Valášek, L. S., Simonetti, A. & Hashem, Y. ABCE1: A special factor that orchestrates translation at the crossroad between recycling and initiation. *RNA Biol.* 1–7 (2017). <http://doi.org/10.1080/15476286.2016.1269993>.
- [157] Sokabe, M. & Fraser, C. S. Human eukaryotic initiation factor 2 (eIF2)-GTP-Met-tRNA_i ternary complex and eIF3 stabilize the 43 S preinitiation complex. *J. Biol. Chem.* 289, 31827–36 (2014). <http://doi.org/10.1074/jbc.M114.602870>.