

Strategies to stabilize RNP complexes for structural
determination by 3D cryo-electron microscopy

Dissertation

in partial fulfillment of the requirements
for the degree “Doctor rerum naturalium”
in the Molecular Biology Graduate Program
at the Georg August University Göttingen,
Faculty of Biology

submitted by

Wen-ti Liu

born in

Taipei, Taiwan

2013

First Referee: Prof. Dr. Holger Stark

Second Referee: Prof. Dr. Marina Rodnina

Third Referee: Prof. Dr. Kai Tittmann

Date of submission of thesis: 30.09.2013

Date of defense 30.10.2013

Herewith I declare, that I prepared the Doctor's Thesis "Strategies to stabilize RNP complexes for structural determination by 3D cryo-electron microscopy" on my own and with no other sources and aids than quoted.

Göttingen, 30.09.2013

Wen-ti Liu

Contents

Contents	v
List of Figures	ix
List of Tables	xi
Acknowledgements	xiii
Abstract	xv
1 Introduction	1
1.1 Single Particle Electron Microscopy	1
1.1.1 Image formation	3
1.1.2 Image processing	8
1.1.2.1 CTF correction	10
1.1.2.2 Preprocessing of particles	11
1.1.2.3 Alignment	12
1.1.2.4 Principal component analysis and classification	14
1.1.2.5 Angular reconstitution	16
1.1.2.6 Three-dimensional reconstruction	18
1.1.2.7 Refinement	18
1.1.2.8 Resolution determination	20
1.1.2.9 Resolving structural heterogeneity	21
1.2 RNP complexes	25
1.2.1 The ribosome	26

1.2.2	The spliceosome	30
1.3	Aim of the work	35
2	Materials and Methods	37
2.1	Material	37
2.1.1	Chemicals	37
2.1.2	Buffers	39
2.1.3	Laboratory materials	40
2.1.4	Special equipments	41
2.1.5	Softwares	42
2.2	Methods	43
2.2.1	Purification of RNP complexes	43
2.2.1.1	HeLa spliceosomal C complex	43
2.2.1.2	Optimization of crowding agent and selective precipi- tation of human snRNP	43
2.2.1.3	Isolation of the yeast snRNP	44
2.2.1.4	Isolation of the HeLa 80S ribosome under high and low salt conditions	45
2.2.2	GraFix	46
2.2.3	EM sample preparation	49
2.2.3.1	Preparation of continuous carbon film and holey carbon grids	49
2.2.3.2	Preparation of negative stained samples	50
2.2.3.3	Preparation of unstained cryo samples	51
2.2.4	EM analysis	52
2.2.4.1	Image acquisition with TEM	52
2.2.4.2	Image analysis	53
3	Results	55

3.1	Structural Analysis of Human Spliceosomal C Complex	55
3.2	Purification and Structure of Yeast Endogenous snRNP	58
3.3	Crosslink RNP with PMPI: HeLa 80S ribosomes as an example	62
3.4	Structure of HeLa 80S ribosome purified under low salt condition	67
4	Discussion	73
4.1	Structural Analysis of the Human Spliceosomal C Complex	73
4.1.1	Regions with high structural heterogeneity might represent the functional domains	73
4.1.2	Outlook	75
4.2	Purification and Structure of Yeast Endogenous snRNP	76
4.2.1	Sucrose might be a better crowding agent for snRNP purification	76
4.2.2	Purification without affinity-based procedure reduces sample degradation	77
4.3	Crosslink RNP with PMPI: HeLa 80S ribosomes as an example	78
4.3.1	GraFix of 80S ribosome with PMPI	78
4.3.2	The 80S ribosome might be stabilized by polyamines	79
4.4	Structure of HeLa 80S ribosome purified under low salt condition	80
4.4.1	Purification of ribosomes under low salt conditions stabilizes the flexible stalks	80
4.4.2	Tunnel exit bound factor was preserved on the ribosome with low salt preparation procedure	80
5	Conclusions and Perspectives	83
A	Appendix	85
A.1	Mass spectrometry analysis of yeast snRNP sample	85
A.2	Mass spectrometry analysis of low salt purified HeLa 80S ribosome sample	86
	List of Symbols and Abbreviations	91

Bibliography	95
Curriculum Vitae	103

List of Figures

1.1	Overview of single particle cryo-EM	2
1.2	Overview of the image contrast formation	4
1.3	Phase contrast transfer function at different defoci	5
1.4	General procedure of image processing	9
1.5	CTF correction	11
1.6	Preprocessing of particles	12
1.7	Multi-reference alignment	13
1.8	Principal component analysis	15
1.9	Angular reconstitution by the common line theorem	17
1.10	Conformational heterogeneity and compositional heterogeneity	22
1.11	Supervised classification	23
1.12	Resampling and 3D MSA classification	24
1.13	Crystal structure of the yeast ribosome	27
1.14	Splicing cycle of U2-dependent spliceosomes	31
2.1	Preparation of variant GraFix methods	47
2.2	Chemical formulas of glutaraldehyde and PMPI	48
2.3	Preparation of negative stained samples	50
2.4	Preparation of unstained cryo samples	52
3.1	Structure of human spliceosomal C complex	56
3.2	Analysis on structural variance of human spliceosomal C complex	57

3.3	Human snRNA composition of precipitates from different concentration of PEG6000 in different sugar environments	59
3.4	Structural analysis of yeast snRNP	60
3.5	Structures of HeLa 80S ribosomes under the effect of polyamine and PMPI	63
3.6	GraFix with PMPI might alter the Euler angle distribution	64
3.7	ES7L was crosslinked on the ribosome core with PMPI	65
3.8	Three steps of ultracentrifugation gradient purification of the HeLa 80S ribosome	68
3.9	Low salt preparation stabilizes the structure of HeLa 80S ribosome . .	69
3.10	Nascent chain-associated complex (NAC) or ribosome-associated complex (RAC) was copurified under low salt condition and was further stabilized by GraFix	72
4.1	Comparison of heterogeneity domains and the localization of pre-mRNA in human spliceosome C complex	74

List of Tables

1.1	Size and compositional differences between prokaryotic and eukaryotic ribosomes	28
1.2	List of spliceosome and snRNP 3D structures solved by single particle EM	34
2.1	Chemicals used for this thesis	37
2.2	Buffers used for this thesis	39
2.3	Laboratory materials used for this thesis	40
2.4	Special equipments materials used for this thesis	41
2.5	Softwares used for this thesis	42
2.6	Potassium salt concentrations used in HeLa 80S ribosome purification	46
2.7	Ultracentrifugation conditions of GraFix used in this thesis	49
2.8	TEM conditions for the samples in this thesis	53
2.9	Image analysis for the samples in this thesis	54
3.1	Different preparations of HeLa80S ribosome sample for cryo-EM	66
A.1	Mass spectrometry analysis of yeast snRNP sample	85
A.2	Mass spectrometry of low salt purified HeLa 80S ribosome sample	86

Acknowledgements

First of all, I would like to thank my supervisor Prof. Dr. Holger Stark. His excellent supervision and guidance have enriched my knowledge in cryo-EM, and his enthusiasm in science has motivated my interests in this field. I will never forget Holger coming back to the lab in a late night to save the microscope and a helpless student. He is not only a supervisor in the lab, but also a hero for me.

I want to also thank the members of my thesis committee, Prof. Marina Rodnina, Prof. Kai Tittmann, and Prof. Reinhard Lührmann for helpful discussions and advice. I also want to thank my collaborators, Dr. Sergey Bessonov and Dr. Ashwin Chari for providing the samples.

It is very grateful to have Dr. Ashwin Chari in the lab. Without him and his biochemical knowledge, the work in this thesis would never be possible. Discussions with him have also benefited me a lot in the trail of science. I also want to thank Dr. Florian Hauer for suggestions in image processing. I feel thankful for discussions with Dr. Niels Fischer, who is the ribosome expert. I would like to thank David Haselbach, Dr. Stephanie Schell, and Jil Schrader for discussions in the lab as well as sharing the “production line” in sample purification. I also want to thank Dr. Burkhard Heisen and Dr. Boris Busche for kindly sharing their L^AT_EX templates for me to accomplish this thesis. I would like to thank Michael Hons for cheering me up whenever I was frustrated. I also want to thank Dr. Prakash Dube for his deep humor and his warm

Chai in the afternoon. I want to thank the other former lab members: Dr. Ilonka Bartoszek, Dr. Martin Schmeisser, Dr. Florian Platzmann, current lab members: Andrius Krasauskas, Jan-Martin Kirves, Dr. Mario Lüttich, Dr. Florian Brodhun, Tobias Koske, Frank Würriehausen, Karl-Heinz Knauber, and a short-term Taiwanese lab member: Yu-Ju Lin, with whom I shared a great time. I also want to thank the admirable secretary Juliane Moses who helped me with the complicated German bureaucracy. I also want to thank the members of the electron tomography group: Dr. Dietmar Riedel, Dr. Dirk Wenzel, and Gudrun Heim, for sharing their microscope and for sharing a good time in meals or outdoor activities.

I would like to thank Dr. Steffen Burkhardt from the coordination office of the International Molecular Biology MSc/PhD program for organizing an admirable learning environment. I also thank the program assistants Kerstin Grüniger and Ivana Bacakova for helping me settle my life in Göttingen, so that I did not need to worry too much besides my studies.

I want to thank the international friends whom I have met in Göttingen. In the wonderful times hanging around together, I believe that I have broadened my horizons. I also want to thank the Taiwanese community in Göttingen, which has made me feel like home. A special thank goes to Andrius for his company. Besides, I should not forget those sweet long-distance mails from friends in Taiwan. Last but not least, I deeply thank my family. Without their support, nothing like today can ever be possible.

Abstract

The physiological reactions in a cell are generally not performed by single biological macromolecules, but by complexes of several molecules. They can be a complex of several proteins or can be composed of RNA and proteins as a ribonucleoprotein (RNP) complex. The RNP complexes perform their functions either through dynamic assembly and disassembly of components, such as the spliceosome, or through structural dynamics, such as the ribosome. Due to their dynamic nature and large size, single particle cryo-electron microscopy is an ideally suited method to study these RNP complexes. However, due to the high heterogeneity of samples, which can be a result of compositional difference or conformational flexibility, most of the structures thus far could only be obtained with restricted resolution. In this work, biochemical and computational methods were applied to reduce sample heterogeneity and to resolve heterogeneous sub-states of RNP complexes.

In the first approach, the structure of the human spliceosomal C complex was analyzed. Different sub-states representing the heterogeneity were sorted *in silico*, and the regions with high heterogeneity were characterized. Unfortunately, the amount of heterogeneity exceeded the capacity which could be dealt with by image processing alone and has obstructed the improvement of resolution. Therefore, biochemical methods have been developed to stabilize samples and to decrease heterogeneity. Two approaches were pursued: (1) to stabilize the sample during purification, and (2) to prevent the macromolecules from disruption during the sample preparation for transmission electron mi-

croscopy (TEM). As a proof of concept, the optimization of purification was performed on an endogenous small nuclear ribonucleoprotein (snRNP). With the optimization of the crowding agent, as well as replacing chromatography with selective precipitation, the 3D model of the yeast snRNP could be reconstructed for the first time. To prevent macromolecules from disruption during sample preparation, *p*-maleimidophenyl isocyanate (PMPI) was evaluated as an RNA-protein crosslinker. PMPI was tested on the human 80S ribosome. In the reconstructed 3D model, a crosslink between the expansion segment ES7L and the ribosomal protein L7A was observed. Besides, the E-site tRNA showed higher occupancy. To further stabilize the ribosome during purification, the conventional “salt wash” step, which utilizes high salt concentration to remove salt-unstable proteins, was replaced by selective precipitation under low salt conditions. The L1 and P stalks in the reconstructed 3D models were significantly stabilized in the low salt purification procedure. Furthermore, a factor located next to the nascent chain exit tunnel was copurified. These biochemical approaches were shown to stabilize the ribosome and the spliceosome, and can be applicable in the future for cryo-EM studies on all RNP complexes.

Chapter 1

Introduction

1.1 Single Particle Electron Microscopy

The interactions between objects and the light contribute to our world of vision. To explore the subtlety of the world, light microscopes were invented to enlarge the objects in a micrometer scale. However, smaller objects cannot be magnified clearly with the visible light. This is due to the wavelength restricting the resolution limit, which was proposed by Ernst Karl Abbe in 1873:

$$d = \frac{\lambda}{2NA} = \frac{\lambda}{2n \sin \theta} \quad (1.1)$$

d is called the resolution limit, meaning the minimum distance where two neighbouring features can be distinguished. λ is the wavelength. NA stands for *numerical aperture*, which is the refractive index n of the medium filling the space between the cover glass and front lens multiplied by the sine of the half angle θ of the maximum cone of light that can enter or exit the lens. Therefore, to visualize an object at the nanometer range, a light source of nanometer wavelength is required. Since electrons exhibit a duality of wave and particle, as proposed by Louis de Broglie in 1924, and the electron wave has much shorter wavelength comparing to the visible light, the electron beam can be utilized in microscopes for smaller scaled objects.

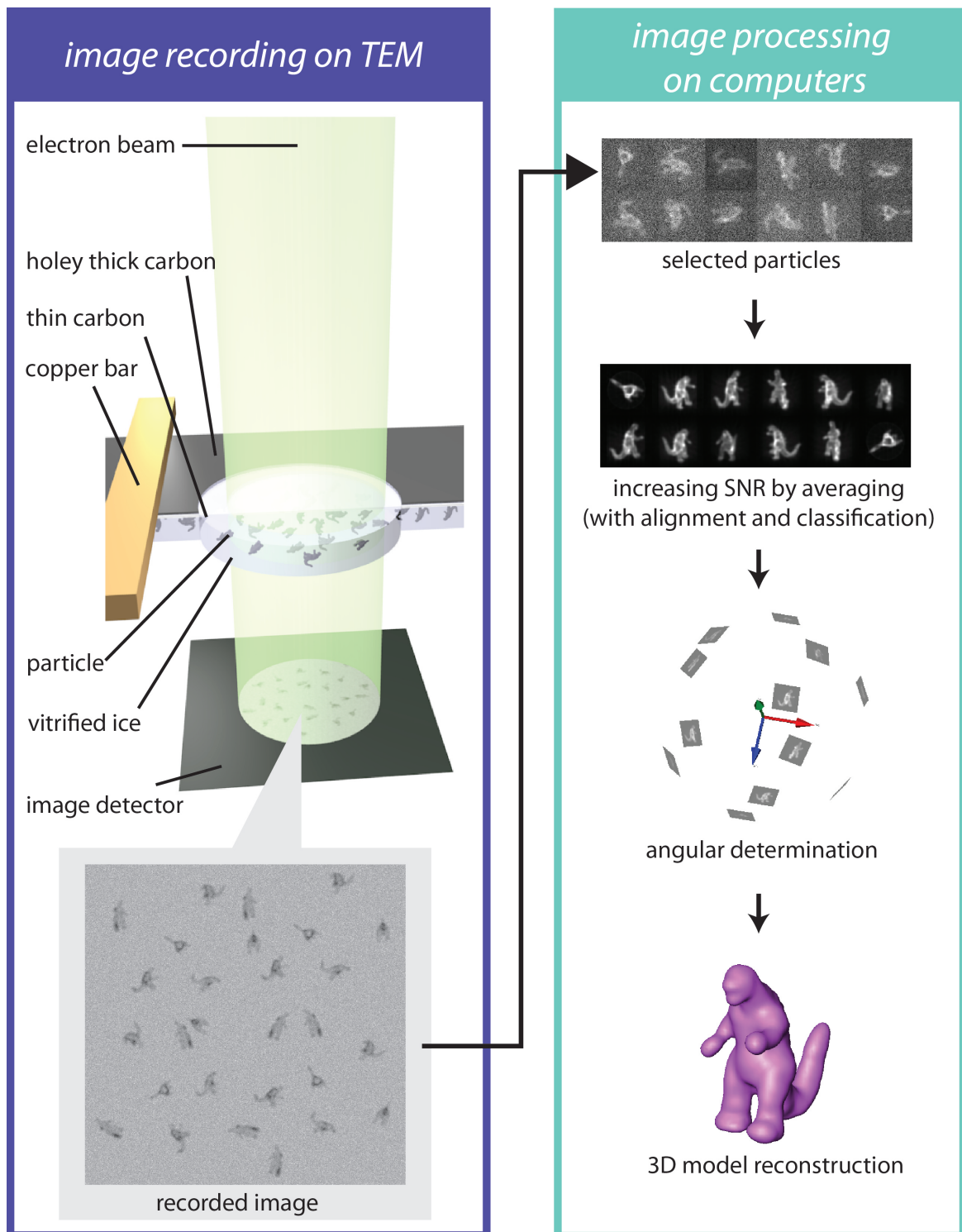


Figure 1.1: Overview of single particle cryo-EM. A sample grid is placed in the TEM, and images are recorded. Image processing aims to reconstruct a 3D model from the 2D images, and the main steps include particle selection, alignment, classification, angular determination, and 3D reconstruction.

The transmission electron microscope (TEM) is nowadays widely used in various biological research, such as single particle electron microscopy and electron tomography. Single particle cryo-electron microscopy (cryo-EM) has become one of the main methods in high-resolution structural biology. In single particle cryo-EM, a sample solution is rapidly frozen to a vitrified ice. Single particles are imaged by TEM. These images are analyzed *in silico*, and the 3D model can be reconstructed (figure 1.1).

Though the average resolution of structures obtained by cryo-EM is still lower than that of X-ray crystallography and NMR structures, cryo-EM possesses significant advantages in studying large, dynamic, flexible, and heterogeneous macromolecules. Compared to NMR, cryo-EM can be used to study much larger molecules. Compared to X-ray crystallography, cryo-EM is not restricted by the necessity of crystals. That means, first, the required amount of the sample is much less. Second, the particles can be analyzed in solution instead of a crystallization state. For particles which are dynamic in solution, particles representing different dynamic states can be trapped by vitrification and further analyzed *in silico*.

1.1.1 Image formation

With a contrast, an object is distinguishable from the other objects or the background in an image. Usually in our visual system, the contrast is determined by the difference in color or illumination. In TEM, the image contrast is determined by the illumination difference. Image formation in the TEM is based on the interaction of the electron beam and the atoms of specimens. As electrons are charged particles, they are influenced by the positively charged atomic nuclei or the negatively charged electrons in the orbitals, therefore the incident electrons can be deflected from their original trajectory. This phenomenon is called *electron scattering*. The scattering can be accompanied

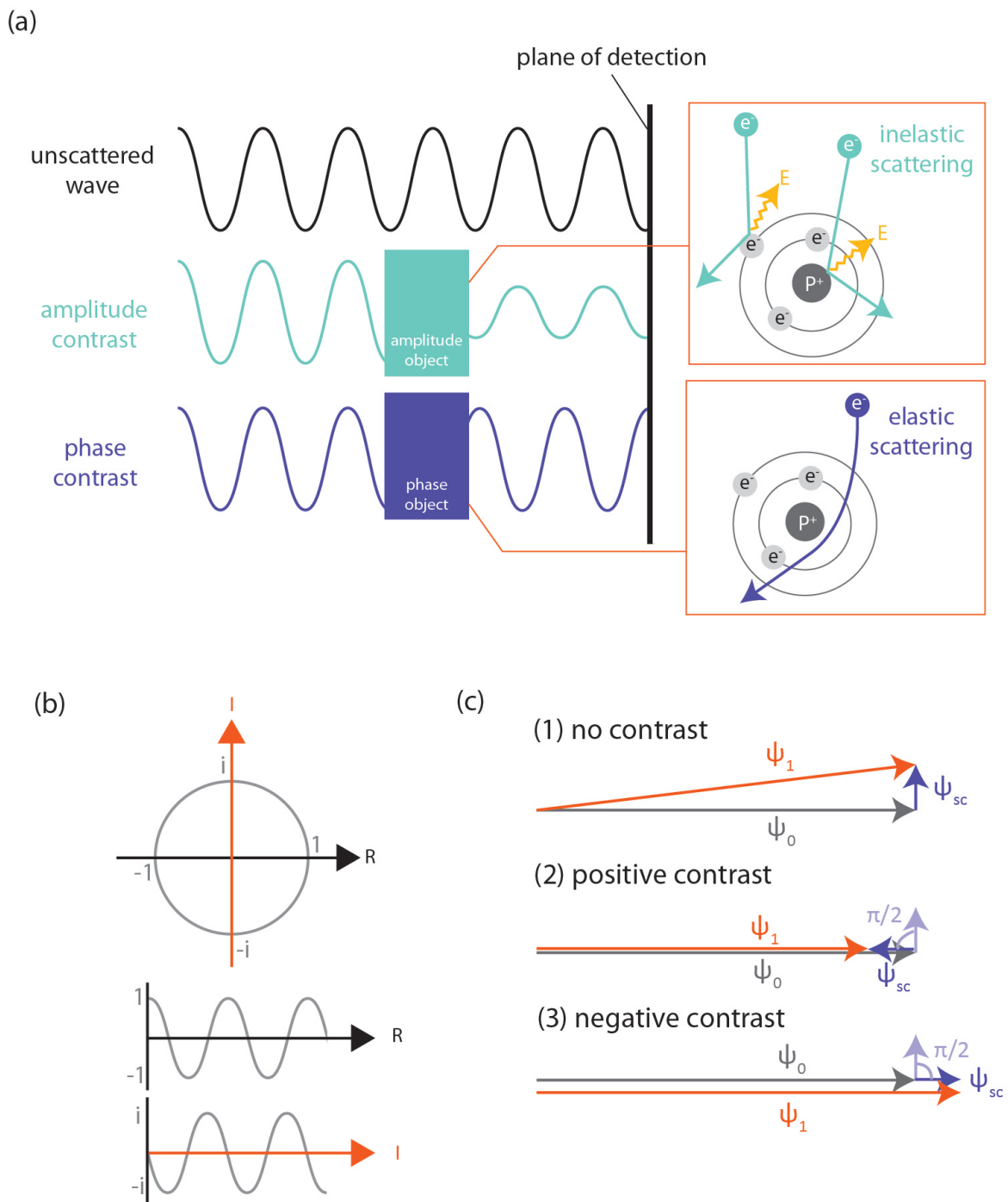


Figure 1.2: Overview of the image contrast formation. (a) Comparison of the amplitude contrast and the phase contrast. (b) Wave can be expressed in a complex plane, and the phase difference between real and imaginary coordinates is $\pi/2$. (c) Exit wave ψ_1 is the vector addition of the incident wave ψ_0 and the scattered wave ψ_{sc} . Depends on the phase shift of ψ_{sc} , different kinds of phase contrast are produced.

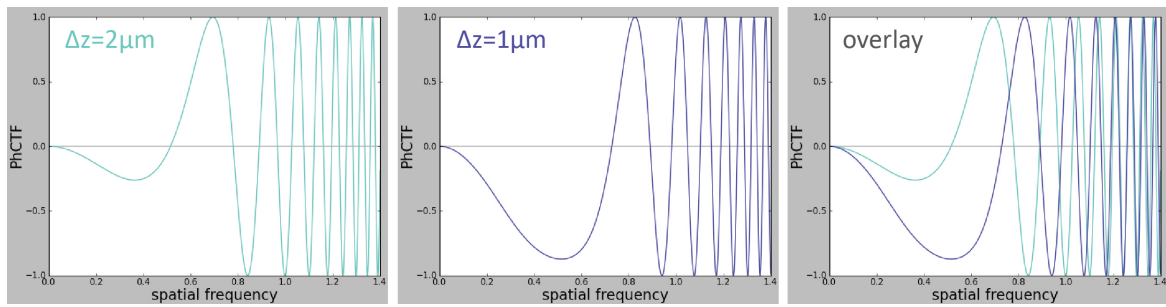


Figure 1.3: Phase contrast transfer function at different defoci. With higher defocus value, the first crossing on the zero line is more close to the original point (the turquoise line). CTF curve was simulated and plotted in matplotlib.

with energy loss, which is called inelastic scattering, where the amplitude of the exit wave is reduced. The amplitude difference between the incident wave and the exit wave produces the *amplitude contrast*. The scattering without energy loss is called elastic scattering, which results in alteration of phase. The phase difference cannot be visualized, but with a phase shift of the scattered wave, *phase contrast* between the incident wave and exit wave can be produced. The difference between amplitude contrast and phase contrast is illustrated in figure 1.2(a).

Because biological specimens are mostly composed of light atoms such as C, H, O, N, S, and P, usually scattering angles are low, and the amplitude remains almost unchanged. Therefore, the main source of contrast in biological samples comes from the phase contrast. Assuming that the wave function of the incident wave is ψ_0 , and the wave function of the exit wave is ψ_1 , the relation can be expressed as:

$$\psi_1 = \psi_0 \exp(i\phi) \quad (1.2)$$

where ϕ is the phase shift at the exit surface. It can also be written according to Euler's formula:

$$\psi_1 = \psi_0 \exp(i\phi) = \psi_0 [\cos \phi + i \sin \phi] \quad (1.3)$$

Since biological specimens for TEM are very thin, it can be taken as a weak-phase object, and phase shift $\phi \ll 1$ can be assumed. Therefore, the exit wave ψ_1 can be approximated as:

$$\psi_1 = \psi_0[\cos \phi + i \sin \phi] \approx \psi_0[1 + i\phi] \quad (1.4)$$

As we can see here, the exit wave can be taken as a superposition of the unscattered and the scattered wave: the term ' $\psi_0 \cdot 1$ ' refers to the unscattered (unmodified) wave, and the term ' $\psi_0 \cdot i\phi$ ' refers to the scattered wave (ψ_{sc}). The imaginary unit of ψ_{sc} refers to a phase shift of $\pi/2$ (figure 1.2(b)). On the plane of detection, only the intensity (I_1) is observed. Intensity is the multiplication of the exit wave (ψ_1) and its complex conjugate (ψ_1^*):

$$I_1 = \psi_1 \cdot \psi_1^* = \psi_0[1 + i\phi] \cdot \psi_0[1 - i\phi] = \psi_0^2[1 + \phi^2] \quad (1.5)$$

Since $\phi \ll 1$ the intensity of exit wave I_1 is similar to the incident wave I_0 :

$$I_1 = \psi_0^2[1 + \phi^2] \approx \psi_0^2 = I_0 \quad (1.6)$$

This means that the object is observed without contrast (figure 1.2(c(1))). However, when the phase of the scattered wave is further shifted by $\pi/2$ (figure 1.2(c(2))), the exit wave ψ_1 is changed as:

$$\psi_1 = \psi_0 - \psi_{sc} \approx \psi_0 - \psi_0 \cdot \phi = \psi_0[1 - \phi] \quad (1.7)$$

So the observed intensity is:

$$I_1 = \psi_1 \cdot \psi_1^* \approx [\psi_0(1 - \phi)]^2 = \psi_0^2[1 - 2\phi + \phi^2] \quad (1.8)$$

Even though $\phi \ll 1$, -2ϕ is much greater than ϕ^2 , and the difference -2ϕ is large enough to be detected as an intensity change, therefore the contrast is observed. The same principle applies for figure 1.2(c(3)). This extra phase shift of -2ϕ can be introduced with a phase plate. Like the light microscope, there are also phase plates for TEM, and the most tested ones so far are the *Zernike phase plate* (Danev and Nagayama, 2001) and the *Hilbert phase plate* (Danev et al., 2002)(Danev and Nagayama,

2004). However, due to technical limitations, such as the requirement of microscope geometry, electron loss, and sample charging, phase plates are not commonly used today (Jensen, 2010). Practically, the extra phase shift of the scattered wave is achieved with *defocus* and the *spherical aberration* of the magnetic lens. The defocus value (Δz) is the distance between the image plane to the focal plane. The spherical aberration (C_s) is caused by an imperfection of the lens, that the waves passing through different distances from the optical axis are not focussed on the same point.

An image can be described as an addition of spatial waves with different frequencies. As the expression of a time progressing wave can be transformed from the time domain to the frequency domain by Fourier transform, a spatial wave can also be transformed from the space domain to the frequency domain, and the Fourier transform of the spatial wave is called “spatial frequency”. Spatial frequencies detected on the image plane represent the scattered angles of the electron wave at the exit plane of the object. Waves with different scattered angles have different phase shifts. This means, except for certain spatial frequencies, where the scattered wave stays at its original $\pi/2$ phase shift relative to the incident wave, other spatial frequencies can be detected. This alteration of phase shift can be written as the *wave aberration* W , a function of the spatial frequency f :

$$W(f) = \frac{\pi}{2}(C_s \lambda^3 f^4 - 2\Delta z \lambda f^2), \quad (1.9)$$

λ is the wavelength of the electron beam. This formula is known as the *Scherzer formula* (Scherzer, 1949). The phase contrast transfer function (phCTF) K , which describes the relative contrast corresponding to different spatial frequencies, can therefore be derived:

$$K(f) = -2 \sin W(f) = -2 \sin\left[\frac{\pi}{2}(C_s \lambda^3 f^4 - 2\Delta z \lambda f^2)\right]. \quad (1.10)$$

As a sine function, phCTF starts from zero, and fluctuates between positive and negative values (figure 1.3). In the range of low spatial frequency, the fluctuation is small,

but increases in the higher spatial frequency range. When the phCTF is equal to zero, there is no contrast, meaning no information for this spatial frequency is transferred. In practice, since phCTF alters with defocus, these gaps of information can be compensated by recording series of images at different defocus values. The first zero-crossing point is determined by the defocus value and the spherical aberration coefficient: the higher the absolute value of defocus, the closer the first zero-crossing point is to the origin point. When the defocus value is equal to *Scherzer focus*, scattered waves ψ_{sc} with a relatively broad range of spatial frequencies undergo a $-\pi/2$ phase shift, therefore the highest theoretical instrumental resolution can be reached.

1.1.2 Image processing

In single particle cryo-EM, each three-dimensional (3D) particle is projected into a two-dimensional (2D) image. The 3D particle in solution has six degrees of freedom: three translational parameters (x , y , and z) and three rotational parameters (Euler angle α , β , and γ). For each recorded 2D projection, the z -translational parameter is stored in the form of different defocus values through projection, whereas the other five parameters still need to be restored. This becomes one important task for image processing.

The other important task for image processing is to improve the low signal-to-noise ratio (SNR) of the images. Theoretically with a higher dose of electrons, the interaction between the beam and the specimen is increased, therefore better signal can be obtained. However, biological samples can only be imaged without damage by low dose of electrons ($< 20e^-/\text{\AA}^2$) (Zeitler, 1982). The solution to increase the SNR is by computationally averaging the single particles which represent the same projection, so that the real information is additively enhanced while the noise is subtractively eliminated.

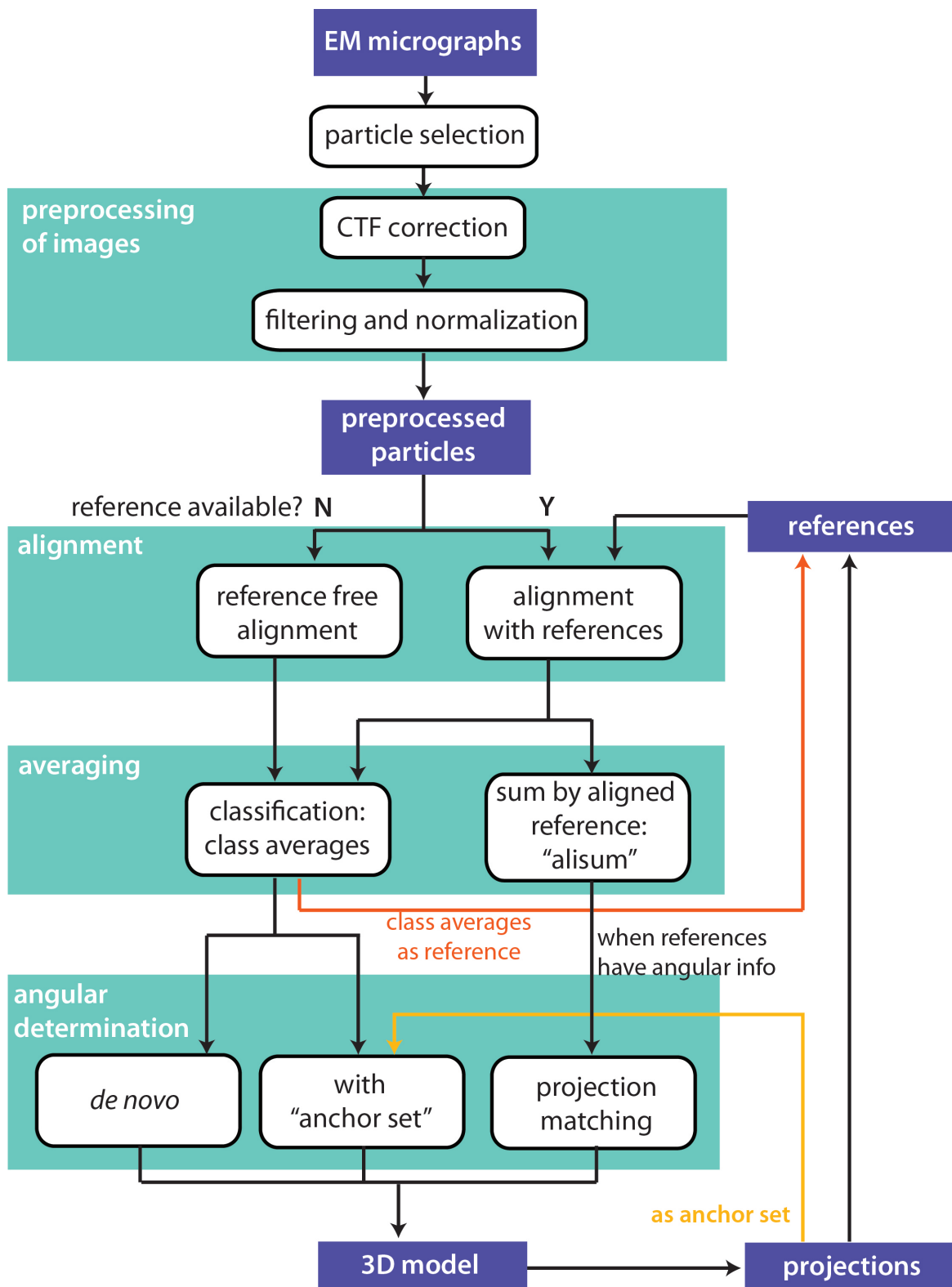


Figure 1.4: General procedure of image processing

To restore the degrees of freedom, first *alignment* is performed, where particles are centered and rotated in the same orientation (see section 1.1.2.3). This means that the translational degrees of freedom x and y are equalized for all particles, and the in-plane rotational parameter α is solved. When the particles are aligned, *classification* can be performed to group particles representing the same projection (see section 1.1.2.4). Images in the same class can hence be averaged, as a “class average”, to obtain better SNR. To restore the last two rotational degrees of freedom: Euler angles β and γ , *angular reconstitution* (van Heel, 1987) is performed. In angular reconstitution, the angular relationship is directly calculated from the 2D images, and can be very dependent on the image quality (see section 1.1.2.5). An alternative is to record the same field twice with different angles, as *random conical tilt*. Since the tilting angle is known, two projections representing the same particle can be correlated with the angular information. Once all translational and rotational parameters are restored, the averaged images, which have satisfactory SNR, can be reprojected into a 3D model. The general procedure of image processing is depicted in figure 1.4, and each step will be explained in detail later.

1.1.2.1 CTF correction

Although the sample can be imaged with TEM, those recorded images are not exactly the same as the original objects. The recorded images are influenced by the CTF and an envelope function. This envelope function describes an intensity decay at higher spatial frequencies, and the decay might come from the imperfection of the device, such as incoherent incident beam or sample damage. Besides, in the PhCTF, for an image recorded at a certain defocus value, information of certain spatial frequencies is missing, and at some spatial frequency ranges, the phase contrast has a negative value. Missing information is compensated by acquiring images at different defoci, where the zero crossings are at different spatial frequencies, and averaging these images. The negative values of the CTF are flipped *in silico*, and this process is called *phase correction*. In

addition, the envelope decay, which reduces detail information, can be corrected with *amplitude correction*. The effect of correction on the CTF is depicted in figure 1.5.

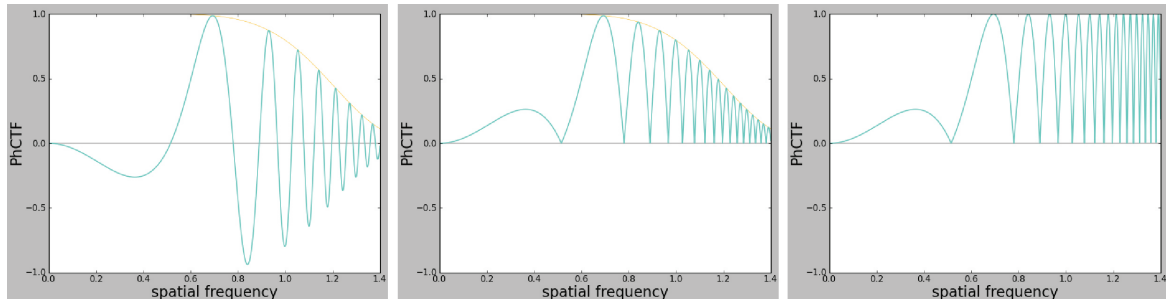


Figure 1.5: CTF correction. Left: uncorrected CTF, which is a product of CTF and an envelope function. The envelope function describing the incoherence of the beam is plotted in yellow. Middle: phase correction, where negative CTF is flipped into positive. Right: amplitude correction. CTF curves were simulated and plotted with matplotlib.

1.1.2.2 Preprocessing of particles

After CTF correction, particles need to go through additional preprocessing procedures before alignment. First, in the early stage, when the overall structure is more important than the details, particles can be *rescaled*, for example, 1/2 or 1/3 of the original size, to speed up the calculation. *Filtering*, which is performed in the Fourier space, keeps only the useful range of spatial frequencies of the images at this stage. The very low spatial frequencies, which represent sharp contrast, are usually from the background gradient of ice, carbon, or stain. The high spatial frequencies can be a mixture of structural details, which is not the main concern in the early stage, and random noise, therefore are also filtered out. Since the particles are spreading on different locations of the carbon film, the micrograph of each particle can have very different gray values. By *normalization*, the mean density of all particles are equalized, and the standard deviation is scaled to the same value. A *circular mask* can be applied on the micro-

graphs of particles to remove unwanted features, such as neighboring particles, on the periphery of the micrograph. The effect of preprocessing is illustrated in figure 1.6.

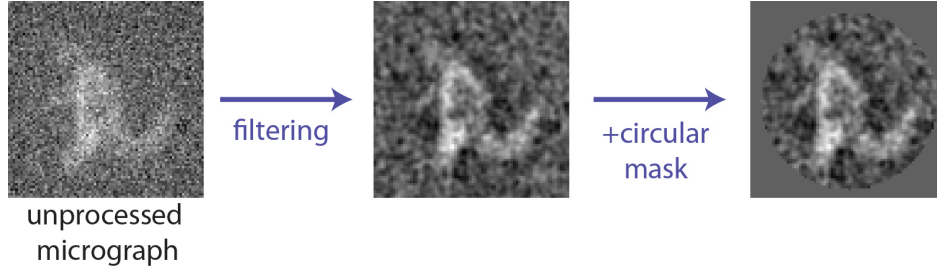


Figure 1.6: Preprocessing of particles, including filtering and application of the circular mask.

1.1.2.3 Alignment

In the alignment procedure, particles are shifted to the center, and particles with similar views are rotated to the same direction according to a set of *reference images*. Commonly used references are 2D class averages or projections from a 3D model. When no reference is available, *reference-free alignment* is performed, where a circular density, which is generated by averaging all images and rotationally self-averaging, is used as the initial reference (Dube et al., 1993). For an alignment process, the image translational shift x and y and in-plane rotation angle θ can be expressed in a transformation matrix T :

$$T = \begin{bmatrix} \cos \theta & -\sin \theta & x \\ \sin \theta & \cos \theta & y \\ 0 & 0 & 1 \end{bmatrix} \quad (1.11)$$

The goal of alignment is to minimize the distance between image f and reference g by applying the transformation matrix T , so that

$$\int |f(u) - g(T(u))|^2 du \rightarrow \min, \quad (1.12)$$

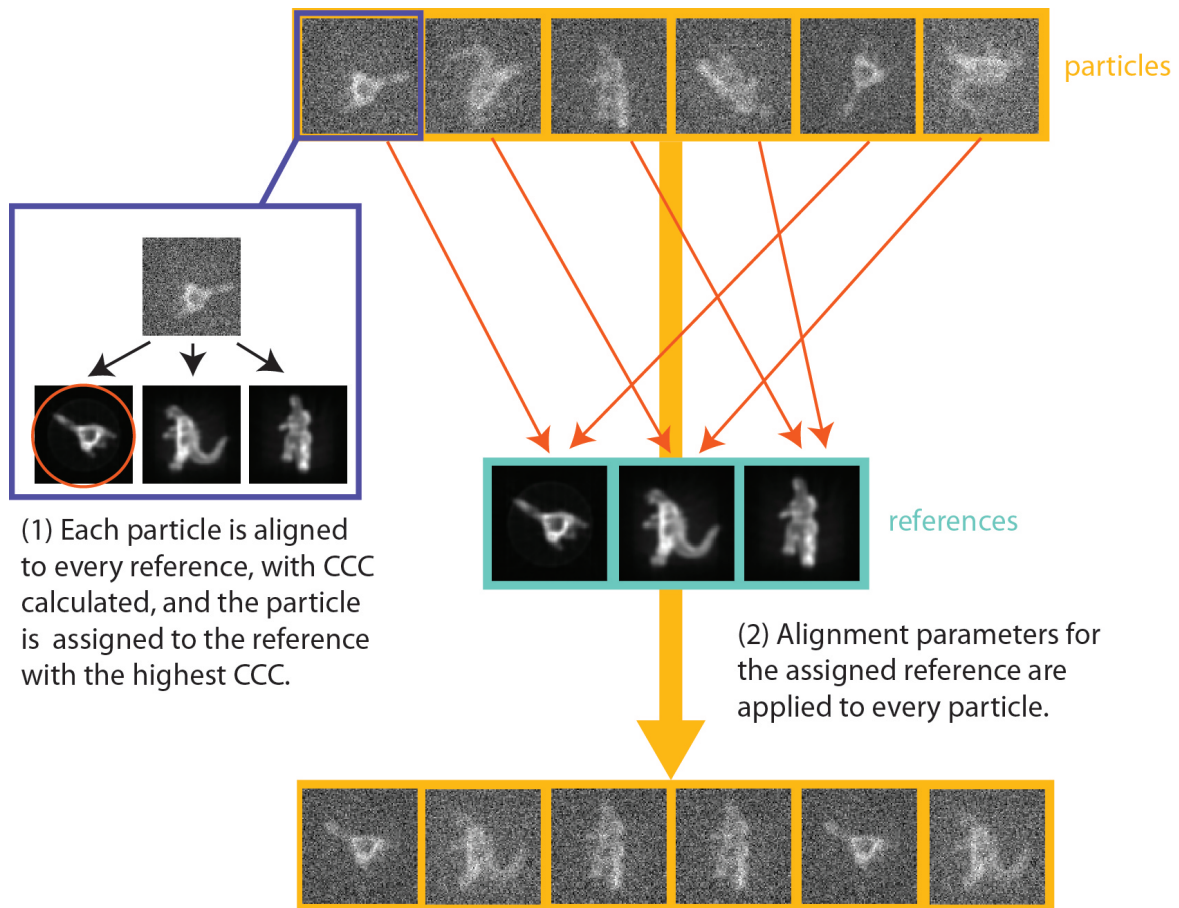


Figure 1.7: Multi-reference alignment

where $u = [u_x \ u_y \ 1]^T$ is a vector of pixel-coordinate.

In the presence of more than one references, every particle is aligned to all the reference images, and the similarities between this particle and each of the reference images can be evaluated. The similarity evaluation is calculated with the *cross-correlation coefficient (CCC)*:

$$CCC = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}, \quad (1.13)$$

where n is the total number of pixels in the image or the reference, and x and y are gray values of pixels in the image or the reference. The CCC of each particle with every reference is calculated, and the alignment parameters, shift and rotation, of the reference which leads to the highest CCC, are applied onto that particle. This process is called *multi-reference alignment* (van Heel et al., 2000), and is illustrated in figure 1.7.

1.1.2.4 Principal component analysis and classification

To group and average particles representing the same projection for better SNR, classification is performed. For a dataset composed of a large amount of particles, the classification process can be computationally demanding. On the other hand, not all the image information is needed for classification, as long as the image features are enough to distinguish the images and classify them. To facilitate calculations, multivariate statistical analysis (MSA) has been used to compress the data (van Heel, 1984).

For electron micrographs, an image of p pixels is described by p gray values, and this image can be expressed as a vector or a point in a p -dimensional coordinate system, with the coordinate of each dimension representing the gray value of a specific pixel. The full image dataset, therefore, can be considered as a “data cloud” in this coordinate system. In MSA, new axes of the component coordinate system are formed, and the new axes, also called *eigenvectors*, describe the variance within the data cloud (i.e. *principle component*). The first eigenvector points to the highest variance, the second

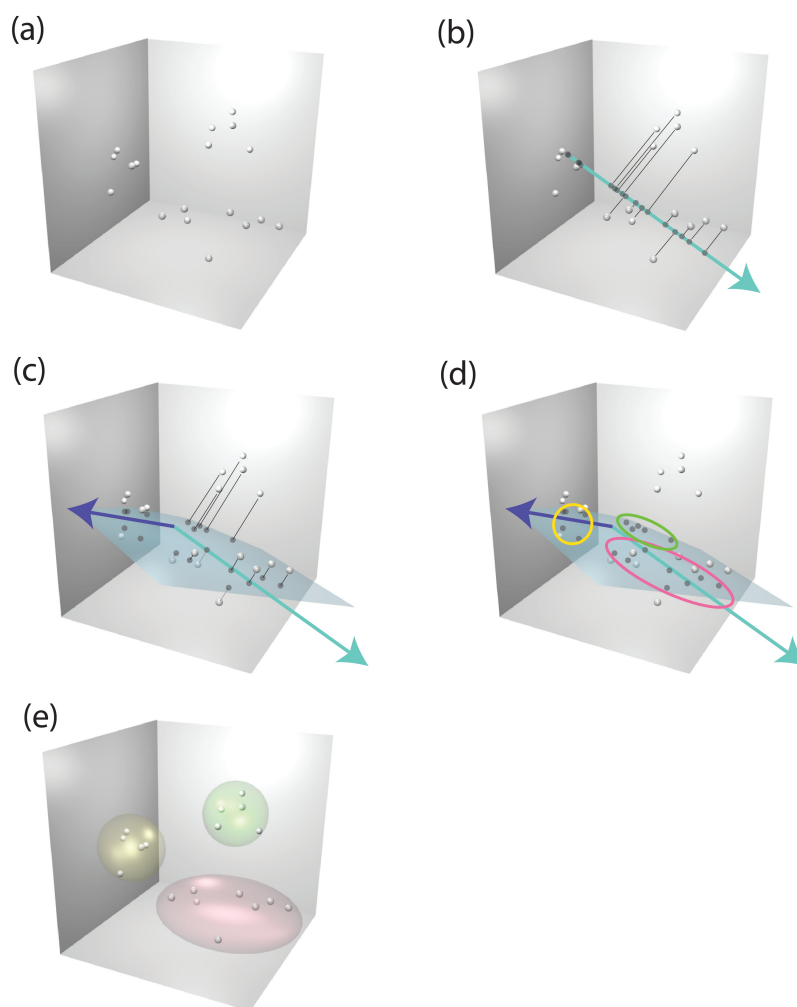


Figure 1.8: Principal component analysis. (a) Images with 3 pixels can be taken as coordinates with 3 dimensions, therefore can be plotted in a 3 dimensional space as a data cloud. (b) In PCA, the first component points towards the direction representing the highest variance (turquoise arrow). The data cloud can be projected on the line so that the representing dimensions reduces from 3 to 1. Note that projected points are not well resolved. (c) To better resolve the projections for more detailed classification, the second axis pointing towards the second highest variance is expanded (indigo arrow). The second axis is perpendicular to the first axis; therefore, a component plane is formed (blue plane). The data cloud can be projected on the plane. (d) Distances between all the projections on the component plane, which represent the variance, are calculated. Also, projections are classified into groups according to the variance. (e) The classification result on the component plane is applied to the data cloud in the data space.

eigenvector, which is perpendicular to the first one, points to the second high variance, and so on. The number n of eigenvectors determines how detailed the variance is taken into consideration for classification. For the classification, distances between all the point projections in the component coordinate system are calculated, and point projections that are near to each other are classified into a class. The maximum distance between two points within a class is determined by the wanted maximum intraclass variance. Instead of the original image coordinate system, classification is performed in the component coordinate system with n axes. Because the very detailed variance might come from random noise, n is much less than p , therefore, the required calculation resources can be significantly reduced. An example of MSA performed on images with three pixels is described in detail in figure 1.8.

1.1.2.5 Angular reconstitution

The key of turning 2D class averages into a 3D model is to reconstitute the relative orientations of projections. There are three rotational degrees of freedom for the particles to lay on the sample grid and be imaged. They are called *Euler angles*: α , β , and γ . The in-plane rotational degree of freedom α is removed by alignment, so what needs to be determined is the two out-of-plane rotations β and γ . Without tilting the sample, the angular reconstitution method (van Heel, 1987) calculates the relative orientations of the projections. Angular reconstitution is based on the *common line projection theorem* (Crowther et al., 1970), which claims that two different 2D projections from the same 3D object have at least an 1D line in common. This can be applied in Fourier space or real space. In this work, the software packages used (IMAGIC and CowEyes, custom-made software) perform angular reconstitution in real space. In real space, to determine the position of this common line, which gives information of the angles, two sets of 1D line projections are generated from the two 2D images. For each 2D image, the set of line projections is with 1° interval, and the stack of a total of 360 line projections is called a *sinogram*. The correlation coefficients of the two sinograms are

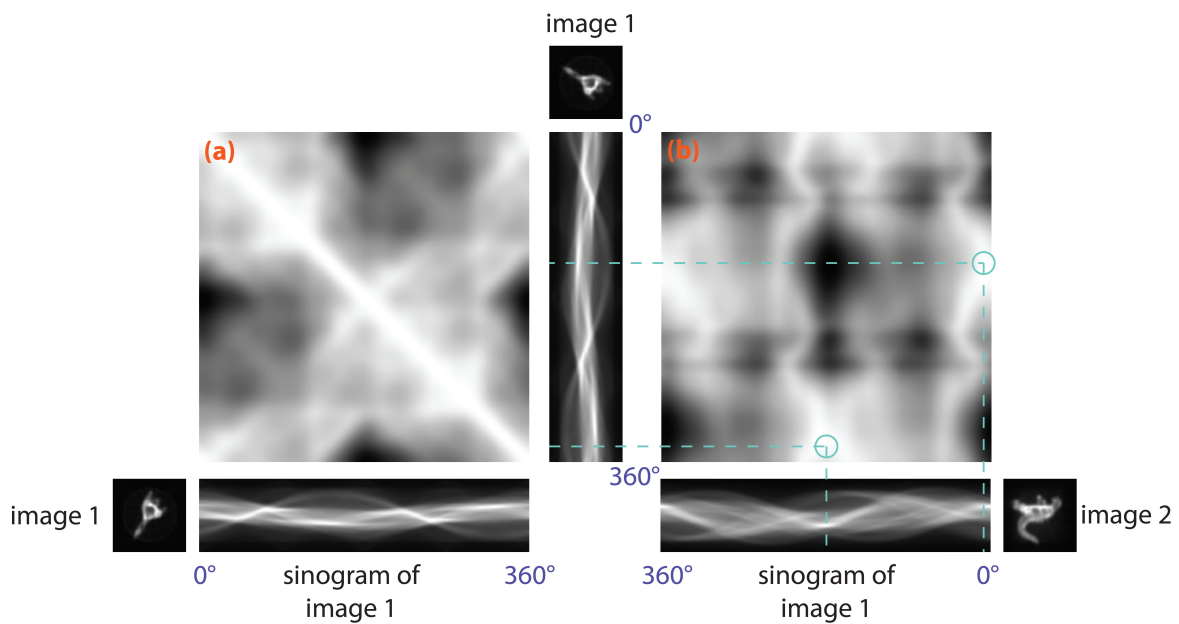


Figure 1.9: Angular reconstitution by the common line theorem. The sinograms of image 1 and 2 were generated, and the sinogram correlation could be calculated: (a) auto-correlation of image 1; (b) correlation between image 1 and 2. The gray value indicates the correlation coefficient: the lighter, the higher correlation. For two projections from an asymmetrical 3D structure, two peaks of highest correlation are observed, with about 180° in between.

calculated line by line (figure 1.9), and the highest correlation coefficient indicates the position of the common line.

In the presence of a known structure, projections from the structure can be used to improve the accuracy for angular assignment. These projections with known projection angles are used as an “anchor set”. Instead of calculating the sinogram correlation between two different class averages, the sinogram correlation of each class average is calculated with every projection from the anchor set.

1.1.2.6 Three-dimensional reconstruction

Once the relative angles of the averaged images are known, the 3D model can be reconstructed. Reconstruction can be done in the real or Fourier space, and the method used in this work is called *exact filter back projection* (Harauz and van Heel, 1986), which is one of the real-space methods. For the conventional back projection method, every image can be taken as a ray of pixels, and each ray comes from a direction which is corresponding to its determined orientation. Where the rays intersect, the 3D model emerges. However, some information of the 3D model might smear out with the ray and lead to a blurry halo around the 3D model. This is because of over-weighting of the low spatial frequency in the central overlapped region. To correct this artifact, a dampening filter, which normalizes the spatial frequencies, is applied individually on every averaged image.

1.1.2.7 Refinement

The reconstructed first 3D model usually does not have a satisfactory resolution, and might even contain artefacts besides real structural features. Therefore, iterative data processing cycles are needed to refine the structure. The overall procedure is similar to the first round, but there are some detailed differences in each step, which can be critical for the refined result. The procedures are also included in the graphical overview (figure 1.4).

Projection Projections can be produced from the 3D model. They can be used as references for alignment, or as the anchor set for angular reconstitution. Most of the biomolecules do not show a homogeneous angular distribution; therefore, some of the rare views cannot be correctly classified in the first round due to the sparse abundance. Missing angles of image information can lead to distortion of the model, but a 3D model

can still be reconstructed as long as the angular coverage is sufficient. The rare views, which cannot be classified in the first round, can probably be reproduced by projecting the 3D model. When the rare views are in the reference, particles representing the rare views can be better aligned, and it is possible to obtain decent class averages. The interval for the projection angles can be big in the early stage of refinement in order to facilitate the calculation speed, and in the later stage, when reconstructing the structural details becomes more important, the angular interval should be reduced.

Filtering and alignment Due to the damping with increasing spatial frequencies of the CTF, low spatial frequencies are initially dominant. However, overemphasis of low spatial frequencies can influence the accuracy of alignments of structural details at late stages. Therefore, during the refinement, the threshold for the high-pass frequency filter is shifted to be higher, and for the low-pass filter is lower. In addition, for CCC calculation in alignment, *mutual correlation function (MCF)* is used instead, due to *cross correlation functions (CCF)* used in early stage are “squared” correlation functions, which amplify the low spatial frequency domain (van Heel et al., 2000).

Averaging and angular assignment Because images of particles have poor SNR, the alignment becomes model-bias prone. The noise in an image can easily be aligned to a wrong reference. Therefore, alignment and averaging should be decoupled in the early stages, where the averages from classification (class averages) should be used. In the later stages, when the overall structure of the model is validated, an “ali-sum”, which is the average of images aligning to the same reference during multi-reference alignment, can be used instead. “Ali-sums” are used in this step instead of class averages, not only for reducing the imprecision from classification, but also for better angular assignment. Because angular reconstitution is not very precise within a small range, when ali-sums are used, the projected angle for the reference can directly be assigned to the corresponding ali-sum. This method is called *projection matching* (Penczek et al., 1994).

Automated refinement As described above, many parameters are involved in the refinement process, and they are critical for the result of the refinement. Usually input values of the parameters are determined by the user, which requires experience to obtain optimum results. It can also be time consuming for the user to perform several rounds of refinements. Software packages with different approaches have been developed for automated refinement, e.g. **FREALIGN** (Grigorieff, 1998, 2007) and **RELION** (Scheres and Chen, 2012).

1.1.2.8 Resolution determination

For the single particle cryo-EM, resolution determination is in practice an estimation about how detailed real structural features can still be distinguished from the noise. The most commonly used method for resolution determination is called the *Fourier shell correlation (FSC)* (Harauz and van Heel, 1986). The level of detail in this case is described in the Fourier space as spatial frequency. The images utilized for the reconstruction of the final 3D model are split into two sub-sets, and two respective 3D structures are reconstructed independently. Like the Fourier transformation of a 2D image is concentric circles, the Fourier transformation of a 3D structure is concentric shells. The correlation between the Fourier transformations of these two 3D models is calculated shell by shell, and the correlation coefficient in Fourier space is calculated as

$$FSC(R) = \frac{\sum_{R=S} F_1(R)F_2^*(R)}{\sqrt{(\sum_{R=S} |F_1(R)|^2 \sum_{R=S} |F_2(R)|^2)}} \quad (1.14)$$

where R is the spatial frequency, S is the shell radius in Fourier space, F_1 represents the Fourier transformation of the first 3D structure, and F_2^* corresponds to the complex conjugated Fourier transformation of the second 3D structure. The FSC is plotted against the reciprocal of resolution, usually like a decaying curve. A low correlation co-

efficient means two 3D models of sub-datasets have no structural features in common, and the correlation calculated indicates random noise. There are several conventional criteria to determine at which resolution the correlation coefficient still represents the structural information: the 0.5 FSC value, the 0.143 FSC value (Rosenthal and Henderson, 2003), and the intersection with the 3σ threshold curve of the background noise.

In fact, it was originally suggested that the image dataset should be splitted into two halves from the very beginning and refined separately, instead of splitting only at the final refinement stage as described above. Indeed, the latter method reduces the required calculation resources, but might lead to an over-estimation of the resolution and over-fitting. With the improvement of the calculation power, the original method of FSC determination was repropose as the *gold-standard FSC* (Henderson et al., 2012). The structural models determined by gold-standard FSC were compared with crystal structural models, and the gold-standard FSC was proven to be more reliable than the FSC in which data is split only in the end of refinement (Scheres and Chen, 2012).

1.1.2.9 Resolving structural heterogeneity

A significant advantage of single particle cryo-EM over crystallography is the tolerance of sample heterogeneity. Structural heterogeneity can come from the flexibility of molecules (conformational heterogeneity) or different compositions of the molecules (compositional heterogeneity) (figure 1.10). For example, the expansion segments of the human ribosome are very flexible, and the consequence of averaging ribosome particles which have expansion segments pointing towards different directions is that the roots of the expansion segment are blurred out, and the termini of expansion segments are not visualized (Anger et al., 2013). An example of heterogeneous composition is the incomplete occupancy of tRNA in the ribosome. In either case, averaging heterogeneous images results in reduced local resolution. In order to obtain an isotropic resolution,

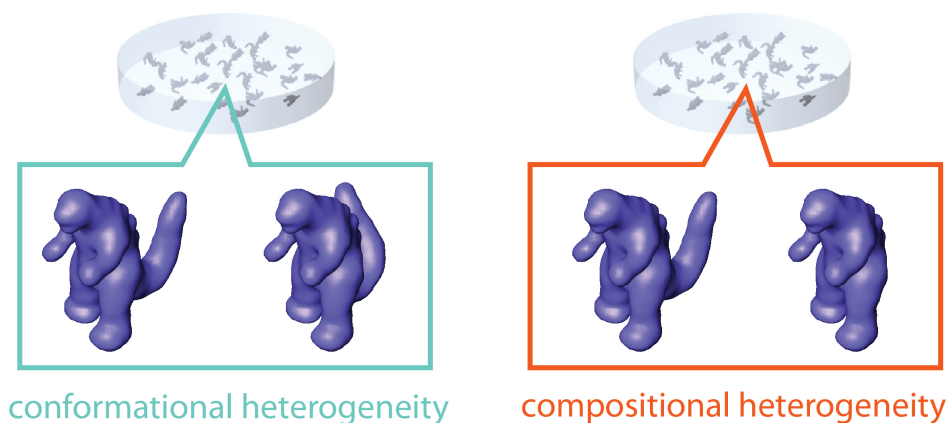


Figure 1.10: Conformational heterogeneity and compositional heterogeneity. Compositional heterogeneity is due to particles in one sample having different composition of components, whereas in a conformational heterogeneous sample, particles have the same composition but different conformations.

resolving the heterogeneity is a critical task. There are several methods for dealing with this problem, and two of them, which have been applied in this work, are reviewed shortly in the following paragraphs. For a more detailed review, see the review from Leschziner and Nogales (2007).

Supervised classification One of the most used *in silico* methods to resolve the heterogeneity is *supervised classification*, first applied by Gao et al. (2004). The brief procedure of supervised classification is illustrated in figure 1.11. If initial structural models representing different states in the image dataset are available, these models can be used as seeds for refinement to analyze the heterogeneity. The projections from all the initial models are combined as a large reference set for alignment. The regular multi-reference alignment (MRA) is performed, where images of particles are assigned and aligned to the individual reference, which yields the highest CCC. So depending on the reference to which the particle is assigned, one can assume to which heterogeneous state that particle belongs. The image dataset can therefore be sorted into different

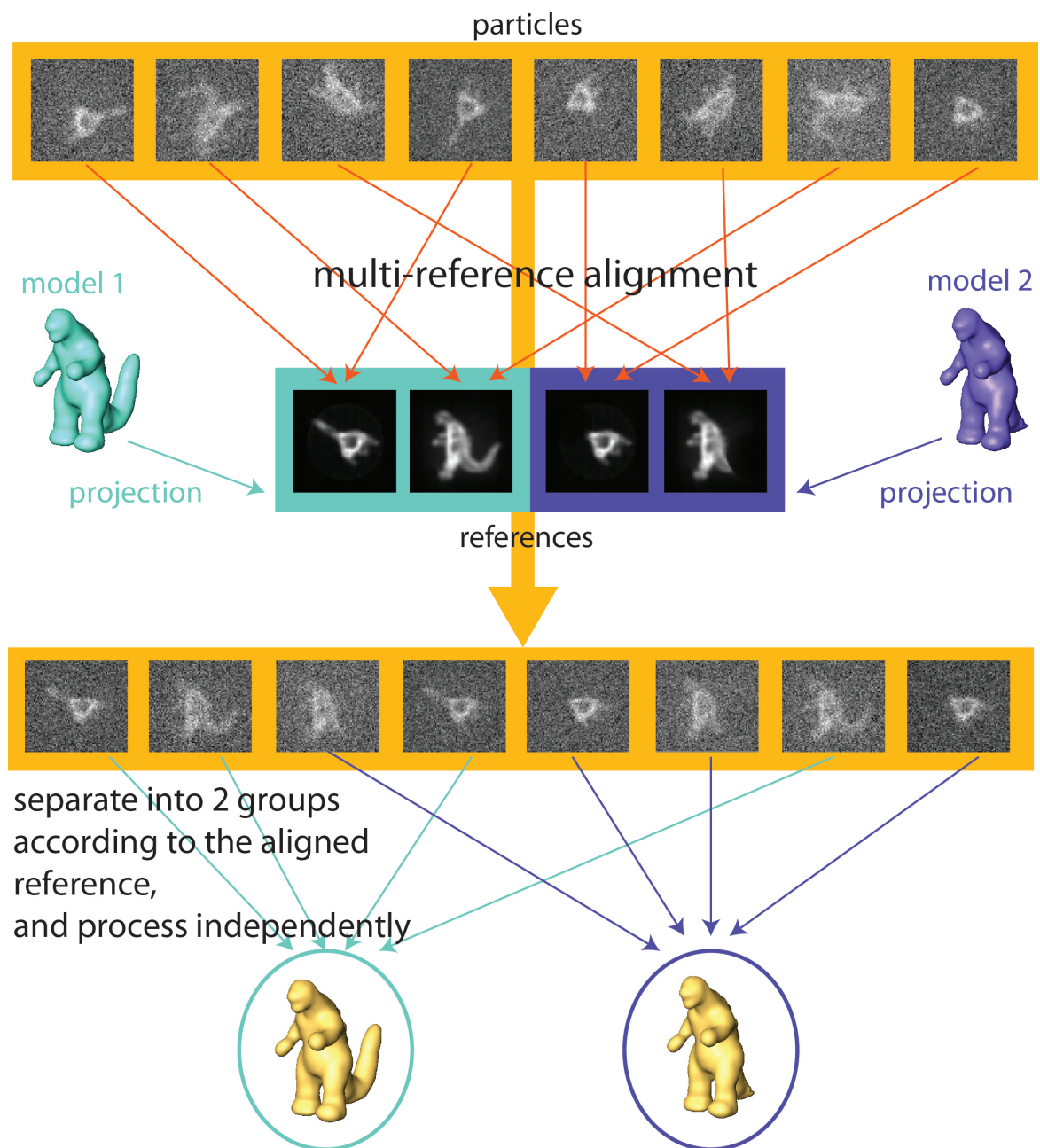


Figure 1.11: Supervised classification: an example of two initial models. Projections from two 3D models are combined as the reference dataset for multi-reference alignment. Each particle is assigned and aligned to the individual reference which yields the highest CCC. The assigned reference decides to which original structure each particle has higher similarity. The particles are sorted into two groups accordingly. Moreover, the 3D models of these two subdatasets, which represents two different states of heterogeneity, are reconstructed independently.

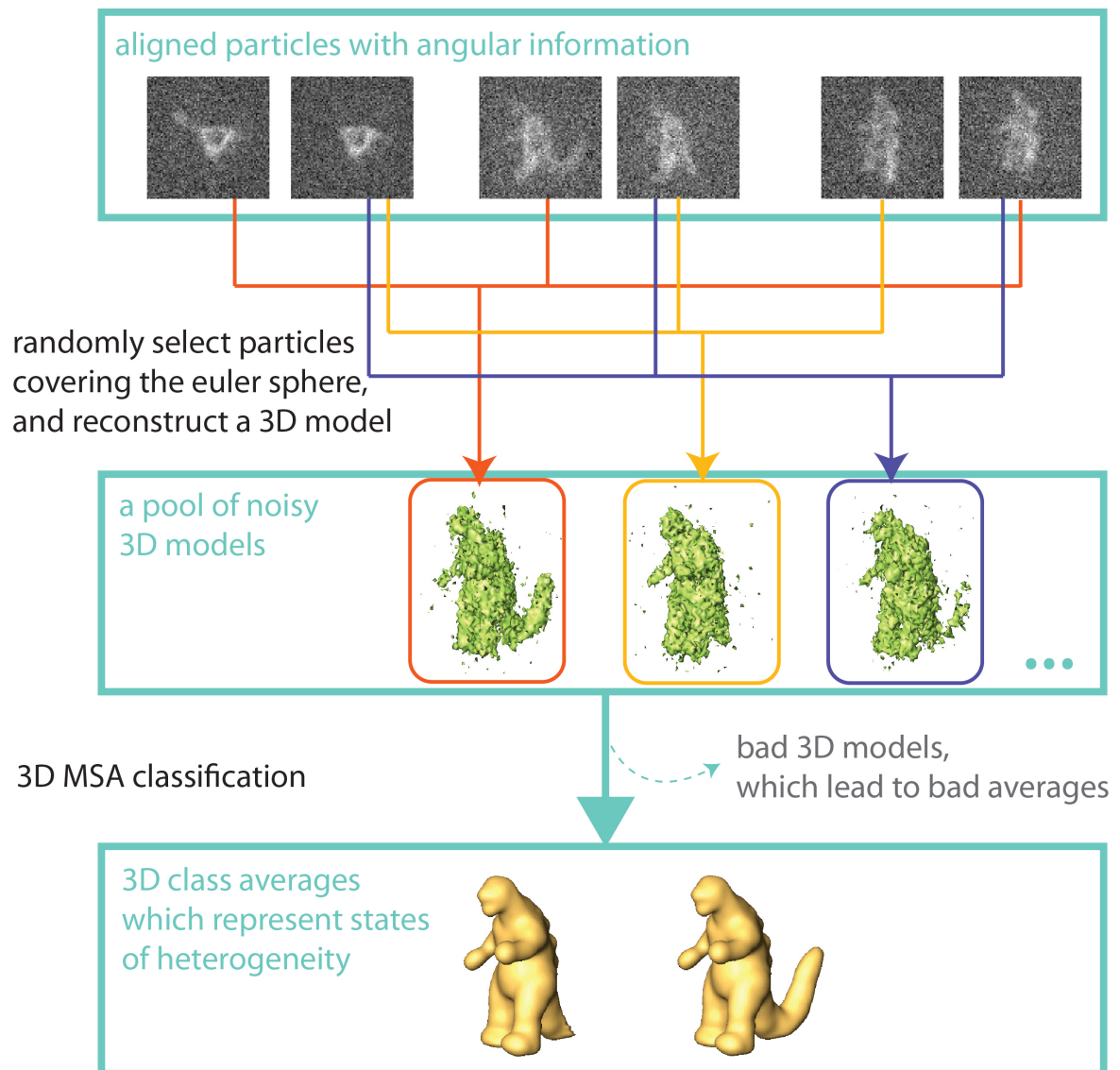


Figure 1.12: Resampling and 3D MSA classification. Aligned particles are randomly selected as groups. Examples are illustrated as orange, yellow, and indigo. Every group of images is used to reconstruct a 3D. These raw 3D structures are classified. Some of the raw 3D structures (orange and yellow) exactly exist in the dataset, and the outcome 3D class averages can represent real heterogeneous states of the sample. However, some of the randomly selected and reconstructed structures (indigo) do not match the existing states (e.g. the dinosaur’s tail does not have full occupancy), and the corresponding class average might look worse defined.

groups and refined separately. However, the features of the initial seeds strongly restrict the heterogeneous states which can be resolved.

Resampling and 3D MSA classification When the overall structure of the particle is available, but preliminary knowledge regarding the individual heterogeneous states is missing, resampling and 3D multivariate statistical analysis classification is useful to obtain the initial seeds for supervised classification (Fischer et al., 2010). The procedure of resampling is illustrated in figure 1.12. First, a preliminary 3D structure, which might not be well resolved, is required as a reference. Particles are aligned, and the Euler angles of the particles are determined by the assigned reference as in projection matching. Several images from the aligned image pool are randomly selected, and they are used to reconstruct a 3D. In order to prevent errors in reconstruction from missing angles, the selected particles should cover the Euler sphere. This step is repeated many times, and a pool of many different 3D models is obtained. Since the input images are not averaged, the SNR of the 3D models is also low. Therefore, MSA and classification are performed at the 3D level. Usually the 3D class averages with the most distinct features are selected as initial seeds for refinement by supervised classification.

1.2 RNP complexes

Single particle cryo-EM is suitable for studying structures of large and dynamic molecules. In a cell, there are many such large complexes utilizing their dynamic nature to accomplish their tasks. Some of them function for maintaining the genetic information flow, such as DNA (deoxyribonucleic acid) polymerases, RNA (ribonucleic acid) polymerases, and spliceosomes. Some of them, such as ribosomes, chaperons, or proteasomes, work for constructing or recycling proteins. These complexes can be entirely composed of proteins, or can be composed of both RNA and proteins as a ribonucleo-protein (RNP) complex. Among all the RNP complexes, structures of ribosomes and

spliceosomes are especially interesting topics for single particle cryo-EM. These two complexes will be further discussed in the next sections.

1.2.1 The ribosome

The ribosome consists of one large subunit and one small subunit, and both subunits are composed of RNA and proteins (figure 1.13). Both subunits work together as a cellular machinery, which is responsible for protein synthesis. The ribosome reads the genetic code (codons) from the mRNA template, and builds up a peptide chain from amino acids according to the codons. Therefore, the protein synthesis process is also called *translation*. Translation can be briefly divided into three steps: initiation, elongation, and termination. The initiation is triggered by recognition of the start codon, which encodes the first amino acid to be translated, and the ribosome is assembled on this location. The amino acid is carried by the transfer RNA (tRNA) which contains the anticodon, the complementary nucleic acid of the codon. At the beginning of the elongation stage, the first aminoacyl-tRNA (fMet-tRNA_f^{Met} in prokaryotes or Met-tRNA_i^{Met} in eukaryotes) moves to the P-site of the ribosome, and the next aminoacyl-tRNA is delivered to the A-site by the elongation factor, EF-Tu in prokaryotes or eEF1a in eukaryotes. The amino acid, or the peptide chain in the next rounds, on the P-site tRNA is transferred to the amino acid on the A-site by the peptidyl transferase. Another elongation factor, EF-G in prokaryotes or eEF2 in eukaryotes, then triggers the translocation, where the A-site and P-site tRNA is moved to the P-site and E-site respectively. This elongation process continues iteratively until the stop codon is reached in most cases. The termination codon is recognized by the release factors, inducing the hydrolysis of the ester bond in peptidyl-tRNA and causes the release of the newly synthesized peptide chain from the ribosome.

Though the roles of the ribosome are the same in prokaryotes and eukaryotes, the size of the eukaryotic ribosomes is significantly larger, with more complicated compo-

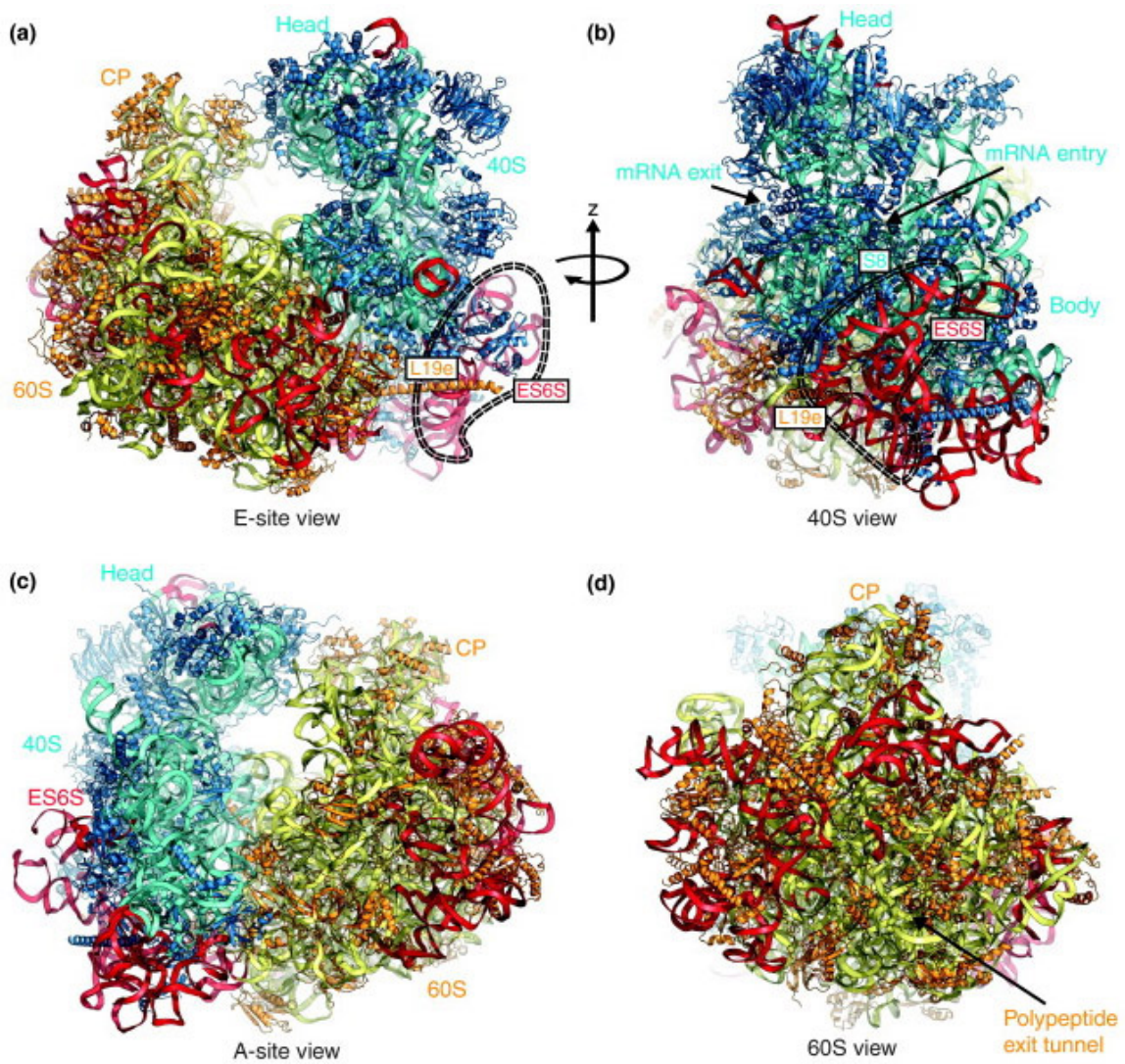


Figure 1.13: Crystal structure of the yeast ribosome. The RNA in the large subunit is shown in yellow, the proteins in the large subunit are shown in orange. The RNA in the small subunit is shown in cyan, the proteins in the small subunit are shown in blue. RNA expansion segments, which are unique in eukaryotes, are shown in red. Reprint from (Jenner et al., 2012) with permission.

sition. The size and compositional differences of the prokaryotic ribosome (*E. coli* as an example) and the eukaryotic ribosome (human as an example) are briefly compared in table 1.1:

Table 1.1:

	<i>E. coli</i>	<i>human</i>
subunits	50S + 30S	60S + 40S
molecular weight	2.3 MDa	4.3 MDa
composition	54 proteins, 3 rRNAs	80 proteins, 4 rRNAs

The size difference comes from the eukaryotic specific proteins and the extra length of rRNAs. While the structures around the active core are relatively conserved, the periphery of the ribosome has altered a lot during the evolution. Most of the eukaryotic specific proteins distribute at the solvent surfaces of the ribosomes. Unlike prokaryotic rRNAs, the eukaryotic rRNAs elongate and form expansion segments (ES), extending dynamically from the ribosome surface (Gerbi, 1996). The expansion segments intertwine with the eukaryotic specific proteins, suggesting the possibility of coevolution (Yokoyama and Suzuki, 2008). The structural complexity has made the structural studies on eukaryotic ribosome much more challenging than the prokaryotic ones. The ribosomal subunits of prokaryotes were resolved by X-ray crystallography in 2000: 30S from *Thermus thermophilus* (Wimberly et al., 2000), 50S from *Haloarcula marismortui* (Ban, 2000) and *Deinococcus radiodurans* (Schluenzen et al., 2000), whereas the yeast ribosome was only resolved by X-ray crystallography one decade later (Ben-Shem et al., 2011). Moreover, due to the significantly longer expansion segments in higher eukaryotes and the accompanying flexibility, high-resolution maps of higher eukaryotes so far could only be obtained by single particle cryo-EM (Armache et al., 2010). However, expansion segments with the highest dynamics have not been able to be visualized with

full length (Anger et al., 2013).

Structural determination of ribosomes provides an insight into how the components work cooperatively to accomplish the translation. It has also been validated that the structural difference between prokaryotic and eukaryotic ribosomes reflects the functional complexity in eukaryotic ribosomes. For example, the “ratcheting” of the ribosome, which is the rotation between the small and the large subunits, is considered to be a driving force to the translocation process. Interestingly, the prokaryotic ribosomes structure seem to prefer the unrotated state (Schuwirth et al., 2005), whereas the structures of eukaryotic ribosomes showed to be at rotated states (Ben-Shem et al., 2011). This might be due to the additional intersubunit connections and interactions at the periphery of the eukaryotic ribosomes (Wilson and Doudna Cate, 2012). The protein-ES clusters mainly distribute on the surface of the 60S subunit, but the surrounding of the nascent peptide chain exit tunnel is left uninterfered. Also, the surrounding of the exit tunnel shows a flat surface, which allows the ribosome to be attached on the endoplasmic reticulum. In addition, the nascent peptide chain emerging from the exit tunnel seems to be more regulated in the eukaryotic system. Besides the chaperone ribosome-associated complex (RAC), which exists in both prokaryotes and eukaryotes, the signal recognition particle (SRP) targets the ribosome-nascent chain complex to the endoplasmic reticulum instead of the plasma membrane. The nascent polypeptide-associated complex (NAC) is only present in eukaryotes, and the role of NAC is to protect the nascent chain from inappropriate interactions with cytosolic proteins, therefore mis-translocation is prevented. The regulation of the factors accessing the exit tunnel is assumed to be related to the ES27 (Beckmann et al., 2001), and the single particle cryo-EM structure of the *in vitro* assembled ribosome-RAC complex in *Chaetomium thermophilum* suggests that ES27 is possible to interact directly to the RAC (Leidig et al., 2013).

Unlike crystallography, the dynamic nature of ribosome particles is preserved in single particle cryo-EM sample. On one hand, the resolution of the reconstructed model can be restricted due to the sample heterogeneity. On the other hand, those thermodynamically unpreferred transitional states can be visualized if the heterogeneity can be resolved (Fischer et al., 2010). With the improvement of hardware and software, nowadays a large cryo-EM dataset with high quality can be recorded and analyzed. It is expectable that the dynamic structures of ribosomes can be better understood in the future, and visualization of high eukaryotic ribosomal expansion segments with full length is no longer an impossible task.

1.2.2 The spliceosome

The main components of the spliceosome are small nuclear RNPs (snRNP), which are composed of small nuclear RNAs (snRNA) and proteins. The spliceosome is responsible for catalyzing the *splicing* reaction. In eukaryotes, most genes are expressed in the form of precursor mRNA (pre-mRNA). In a pre-mRNA, the sequence carrying coding information (exon) is segmented with fragments, which will not be used in translation (introns). Therefore, before being exported to the cytosol for translation, pre-mRNA needs to be modified such that the introns are removed and the exons are joined. This modification process is called splicing. Splicing needs two distinct biochemical reaction steps. First, the 2' hydroxyl group of the adenosine at the “branch site” in the intron performs a nucleophilic attack on the 5' splicing site, so the 5' exon is cleaved, and a lariat structure in the intron is formed. In the second step, the 3' hydroxyl group of the 5' exon attacks the 3' splicing site. The exons are joined, and the intron is cleaved and released.

Depending on the targeted sequence at the 5' splicing site and the branch site in the intron, spliceosomes can be classified into two classes. The most common class is called the *U2-dependent spliceosome*, which is mainly composed of U1, U2, U4, U5, and

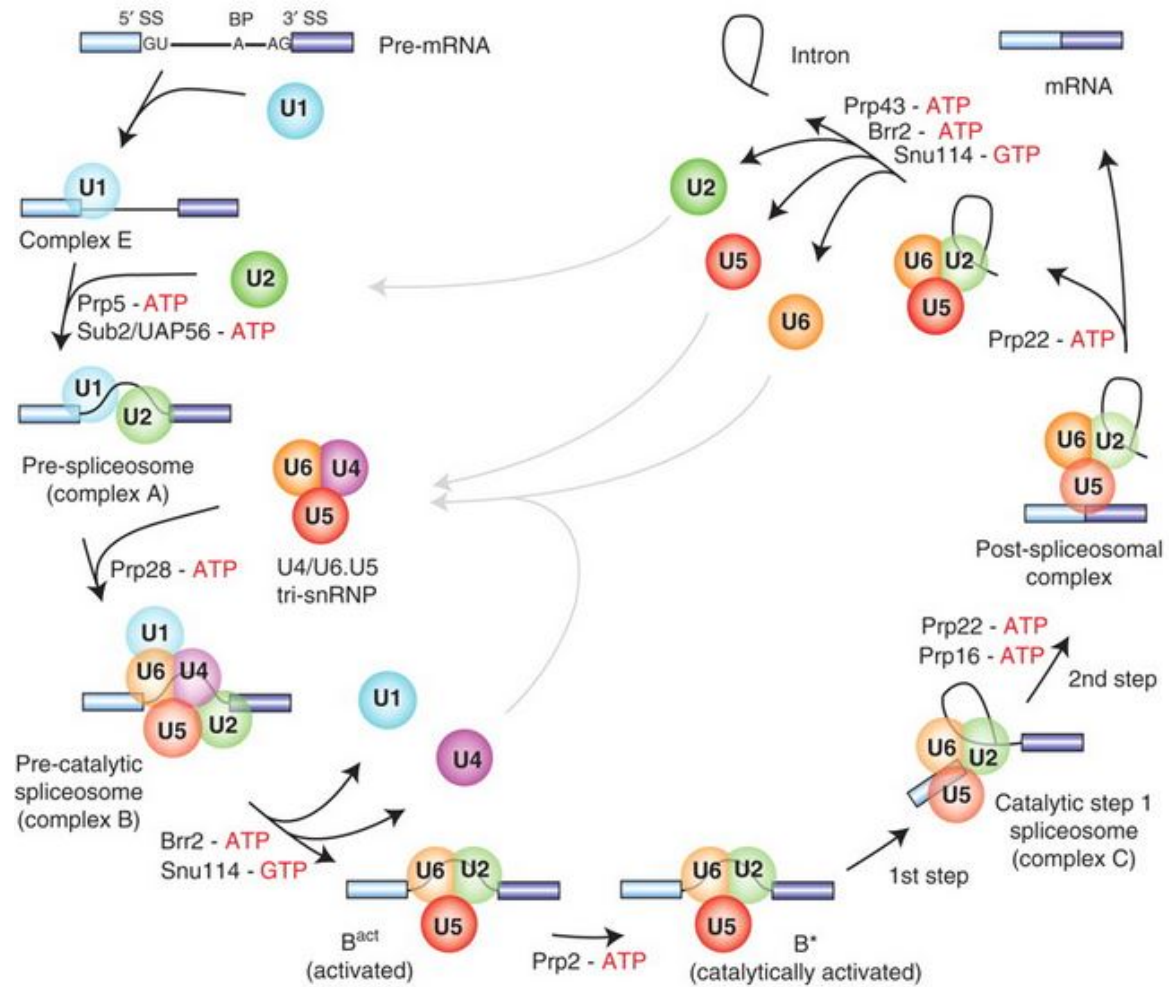


Figure 1.14: Splicing cycle of U2-dependent spliceosomes. Components of spliceosome assemble dynamically and sequentially in time in order to catalyze the splicing reaction. After the splicing reaction, spliceosome is disassembled, and the components are recycled for the next round of splicing cycle. Reprint from (Will and Lührmann, 2011) with permission.

U6 snRNPs. The less abundant class of spliceosome is the *U12-dependent spliceosome*, also known as the minor spliceosome, where the components are U11, U12, U5, and U4atac/U6atac snRNPs (Patel and Steitz, 2003). Different from the ribosomes, which have a rigid assembly and a dedicated catalytic core, components of the spliceosome are thought to assemble in a dynamic and stepwise manner, and the catalytic site is only activated at certain steps of the splicing reaction. After completion of the splicing reaction, components of the spliceosome are recycled and are prepared for the next round of splicing. Therefore, the process of the stepwise splicing reaction with the dynamic spliceosomal assembly is called the *splicing cycle*. The splicing cycle of the U2-dependent spliceosome is discussed in the following as an example (figure 1.14):

Splicing cycle of U2-dependent spliceosomes Before the splicing reaction can be catalyzed by the spliceosome, the positions of the introns and exons must be defined. This is initiated by the U1 snRNP recognizing the 5' splicing site, the non-U snRNP splicing factor 1 (SF1) binding the branch point, and the U2AF binding the pyrimidine tract on the intron at the 3' splicing site. The U2 snRNP interacts with the U2AF and binds the branch point, so that the pre-spliceosomal A complex is formed. The U4/U6·U5 tri-snRNP is next integrated, forming the pre-catalytic B complex. Then a significant compositional remodeling occurs: U1, U4 snRNPs, and most of the U6 proteins are released from the spliceosome, accompanied with a rearrangement of base pairing between RNAs, and the B-complex is activated into the B^{act} complex. The B^{act} complex does not have the catalytic activity, and requires the activation from Prp2 to be transformed into the catalytically active B^* complex (Kim and Lin, 1996; Fabrizio et al., 2009). Though the main components remain unchanged, during the transformation from B^{act} to B^* , the structure is remodeled (Warkocki et al., 2009). The B^* complex catalyzes the first step of splicing, 5' splicing site cleavage, and the C complex with a lariat structured intron is formed. The structure is again remodeled, with association and dissociation of protein components, so that the transesterification

of the 3' splicing site is enabled. After the second step of splicing, the exons are ligated to form the mature mRNA. The mRNA is released from the post-spliceosomal complex and is prepared to be transported to the cytosol. The snRNPs dissociate from the post-spliceosomal complex and can be reused for the next round of splicing cycles, whereas the intron lariat is linearized by the lariat intron debranching enzyme and degraded by RNases.

Interestingly, besides the canonical sequential spliceosomal assembly, it was also proposed that the spliceosome can exist in a form which is already assembled and nearly functional (Stevens et al., 2002). The yeast penta-snRNP, without pre-mRNA, was isolated under low salt condition, and it exhibited catalytic function when pre-mRNA and some additional splicing factors were added.

Structural studies of snRNPs and spliceosomes With years of biochemical research, the big picture of the enigmatic splicing mechanism has become to a great portion well understood. The compositional dynamic nature of spliceosomes has also been explored by mass spectrometry. However, to clarify the functions of all components in this big machinery, and how they interact with each other, structural studies are indispensable. High-resolution structures of important fragments of snRNPs and spliceosomes have been obtained by X-ray crystallography. The U1 snRNP from HeLa cells, determined at 5.5 Å resolution (Pomeranz Krummel et al., 2009) or 4.4 Å resolution (Weber et al., 2010), has been the only structure of a quasi-full snRNP solved by crystallography so far. Due to the compositionally and conformationally dynamic structure, as well as the large size of the particles (e.g. tri-snRNP: ~ 1.7 MDa; spliceosome: ~ 5.5 MDa), single particle electron microscopy (single particle EM) has been used as the main tool for elucidating the structures of full snRNPs and spliceosomes. Three dimensional (3D) structures of snRNPs and snRNP components obtained by single particle EM are listed in table 1.2:

Table 1.2:

<i>Molecule</i>	<i>Publication</i>
<i>snRNPs or snRNP components</i>	
human U1 snRNP	(Stark et al., 2001)
human SF3b (component of U2 snRNP and the U11/U12 di-snRNP)	(Golas et al., 2003)
yeast U11/U12 di-snRNP	(Golas et al., 2005)
human U5 snRNP	(Sander et al., 2006)
U4/U6 di-snRNP	(Sander et al., 2006)
U4/U6·U5 tri-snRNP	(Sander et al., 2006)
<i>spliceosomes</i>	
human A complex	(Behzadnia et al., 2007)
human B Δ U1 complex	(Boehringer et al., 2004)
human C complex	(Golas et al., 2010)
supra spliceosome (penta-snRNP)	(Azubel et al., 2004)

Nevertheless, solving the 3D structures of other snRNPs and spliceosomes have still been obstructed, mainly due to the heterogeneity of the sample. Notably, the transition from B^{act} to C complex involves significant alteration in components and the structure, therefore it becomes an interesting but challenging topic for structural studies. Regardless of the difficulty in 3D reconstruction, image analyses with single particle EM have been performed at the 2D level. For example, the structural remodeling from B^{act} to B* was observed (Warkocki et al., 2009). In addition, antibody labeling on human B complex (Wolf et al., 2009) and yeast tri-snRNP (Häcker et al., 2008) helped us to localize the components and might provide hints of structure related functions. Besides the “missing gaps” of 3D structures in the splicing cycle, resolutions for 3D structures obtained by single particle EM so far have been restricted to max 10

Å (Golas et al., 2003). With the optimization to reduce sample heterogeneity, as well as improvements on image processing technique and calculation power, high-resolution cryo-EM structures of snRNPs and spliceosomes are expected to be available one day, and they will give us an insight into the profound mechanism of splicing.

1.3 Aim of the work

Single particle cryo-EM has been so far the most suitable method to study the structure of large RNP complexes such as the spliceosome and ribosome. Their bulky size and dynamic nature increase the difficulty for crystallography, and the phosphate groups in the RNA provide additional contrast for EM. However, due to the sample heterogeneity, which can be a result of compositional difference or structural flexibility, structures have been obtained with restricted resolution. To reconstruct structures of RNP complexes at higher resolution requires an increase in homogeneity of the sample. To improve the cryo-EM in RNP complexes, three strategies were performed in this thesis:

1. *In silico* purification, as introduced in section 1.1.2.9, was applied to resolve the heterogeneity of the recorded dataset (Chapter 3.1).
2. Due to the high heterogeneity in the sample, computational resolving could not be performed efficiently. Therefore, reducing sample heterogeneity at the biochemical level was a primary goal. One approach is to optimize the sample purification procedure. This thesis was mainly focussed on:
 - (a) Optimization of the crowding agent (Chapter 3.2).
 - (b) Replacing affinity column by selective precipitation (Chapter 3.2).
 - (c) Purification under low salt concentration environment (Chapter 3.4).
3. The other approach to reduce sample heterogeneity is to improve the sample preparation for TEM. In this thesis, crosslinking between protein and RNA

was utilized in the fixation procedure besides the conventional protein-protein crosslinking (Chapter 3.3).

It was expected that the optimization of sample preparation increases the stability of particles, therefore the heterogeneity becomes resolvable by *in silico* sorting. These methods might be used as general procedures for obtaining high-resolution structures of RNP complexes.

Chapter 2

Materials and Methods

2.1 Material

2.1.1 Chemicals

Table 2.1:

<i>Chemical</i>	<i>Supplier</i>
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Sigma-Aldrich
6-aminocaproic acid	Sigma-Aldrich
Acetic acid ester, ultrapure grade	Fluka-Riedel-de-Haen
Acetobutyrate cellulose in ethyl acetate 0.5% (Triafol)	Sigma-Aldrich
Acetone	Merck
Adenosinetriphosphate (ATP)	Amersham Biosciences
Alkylbenzyltrimethylammoniumchlorid 0.5% (Osvan solution)	Sigma-Aldrich
Ammoniumperoxodisulfate (APS)	Carl Roth
Benzamidine hydrochloride hydrate	Sigma-Aldrich

Continued on next page

Table 2.1 – *Continued from previous page*

<i>Chemical</i>	<i>Supplier</i>
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich
di-sodium hydrogen phosphate (Na_2HPO_4)	Merck
Dithiothreitol (DTT)	Carl Roth
Ethanol	Merck
Ethyl acetate	Fluka-Riedel-de-Haen
Ethylenediaminetetraacetic acid (EDTA)	Sigma-Aldrich
Glucose	Merck
Glutaraldehyde	Electron Microscopy Sciences
Glycerol	Merck
Iodoacetamide	Sigma-Aldrich
Lauryl Maltose Neopentyl Glycol (LMNG)	Affymetrix
L-Aspartate	Sigma-Aldrich
L-Cysteine	Sigma-Aldrich
Magnesium chloride	Merck
Millipore water	EMD Millipore
N-Ethylmaleimide	Sigma-Aldrich
Polyethyleneglycol (PEG) 300	Fluka-Riedel-de-Haen
Polyethyleneglycol (PEG) 400	Fluka-Riedel-de-Haen
Polyethyleneglycol (PEG) 6000	EMD Millipore
Potassium chloride	Merck
Potassium dihydrogen phosphate (KH_2PO_4)	Merck
Potassium hydroxide	Sigma-Aldrich
Precision Plus Protein Standards (Unstained)	BIO-RAD
Roti-Aqua-Phenol	Carl Roth
Rotiphorese 10x SDS-PAGE	Carl Roth

Continued on next page

Table 2.1 – *Continued from previous page*

<i>Chemical</i>	<i>Supplier</i>
Rotiphorese Gel 30 (37,5:1) snRNA marker	Carl Roth Department of Cellular Biochemistry, MPIbpc, Göttingen, Germany
Sodium chloride	Merck
Sodium hydroxide	Merck
Spermidine	Fluka-Riedel-de-Haen
Spermine	Fluka-Riedel-de-Haen
Sucrose	Merck
Tetramethylethylenediamine (TEMED)	Carl Roth
Tween 20	Sigma-Aldrich
Uranyl formate	Polyscience Inc.

2.1.2 Buffers

Table 2.2:

<i>Buffer</i>	<i>Composition</i>
HeLa ribosome low salt gradient buffer	50 mM Bis-Tris pH 6.8 50 mM KCl 10 mM MgCl ₂
HeLa ribosome high salt gradient buffer	20 mM HEPES-K pH 7.5 500 mM KCl 5 mM MgCl ₂

Continued on next page

Table 2.2 – *Continued from previous page*

<i>Buffer</i>	<i>Composition</i>
HeLa ribosome sample buffer for low salt prepared sample	10 mM HEPES-K pH 7.5 50 mM KOAc 10 mM NH ₄ Cl 2 mM DTT 5 mM Mg(OAc) ₂
HeLa ribosome sample buffer for high salt prepared sample	50 mM Bis-Tris pH6.8 50 mM KCl 5 mM MgCl ₂ 2 mM DTT 2 mM ATP
HeLa spliceosomal C complex sample buffer	20 mM HEPES pH7.9 180 mM NaCl 1.5 mM MgCl ₂
snRNP sample buffer (Roeder D)	20 mM HEPES pH7.9 100 mM KCl 5 mM MgCl ₂ 0.2 mM EDTA

2.1.3 Laboratory materials

Table 2.3:

<i>Item</i>	<i>Supplier</i>
Carbon rods, highest grade	Ringsdorff Werke GmbH, Bonn, Germany
Copper EM grids, 200 mesh square fine bar	Science Services, Munich, Germany

Continued on next page

Table 2.3 – *Continued from previous page*

<i>Item</i>	<i>Supplier</i>
EM grids Quantifoil R2/2 Cu 200 mesh	Quantifoil Micro Tools GmbH, Jena, Germany
EM grids Quantifoil R3.5/1 Cu 200 mesh	Quantifoil Micro Tools GmbH, Jena, Germany
Ethane (liquid)	Messer, Sulzbach, Germany
Filter paper, 90 mm	Whatman (GE Healthcare)
Mica, 75x25 mm	Plano, Wetzlar, Germany
Nitrocellulose, 0.2 μm	Whatman (GE Healthcare)
Nitrogen (liquid)	Air Liquide, Paris, France
Open-top centrifuge tubes polyclear	SETON scientific, Petaluma, CA, USA
Parafilm “M” laboratory film	Bemis Company, Neenah, WI, USA
PD MiniTrap G25 columns	GE Healthcare, Munich, Germany
Sterile filters, 0.2 μm	Millipore, Billerica, MA, USA
Zeba spin desalting columns, 7K MWCO	Thermo Fisher Scientific Inc., Rockford, IL, USA

2.1.4 Special equipments

Table 2.4:

<i>Item</i>	<i>Source</i>
Cryo electron microscope Titan Krios	FEI, Eindhoven, the Netherlands
Cryo electron microscope CM200 FEG	Philips, Eindhoven, the Netherlands
C_s corrector	CEOS GmbH, Heidelberg, Germany
Eagle 4k CCD Camera	FEI, Eindhoven, The Netherlands
Edwards E12E vacuum coating unit	BOC Edwards, Kirchheim, Germany

Continued on next page

Table 2.4 – *Continued from previous page*

<i>Item</i>	<i>Source</i>
Falcon 2 CMOS Direct Electron Detector camera	FEI, Eindhoven, the Netherlands
Peristaltic Pump LKB-Pump P-1	Amersham Pharmacia Biotech
Room temperature EM specimen holder	Philips, Eindhoven, the Netherlands
TemCam F415 (slow scan 4K x 4K CCD camera)	Tietz Video Systems, Germany
Vitrobot Mark IV	FEI, Eindhoven, the Netherlands

2.1.5 Softwares

Table 2.5:

<i>Program</i>	<i>Source</i>
Amira 5.2.2	Visualization Sciences Group (VSG, part of FEI)
Blender	Stichting Blender Foundation, the Netherlands http://www.blender.org/
CowEyes	Boris Busche, Jan-Martin Kirves, Mario Lüttich (Stark lab)
ImageJ	National Institutes of Health, Bethesda, MD, USA
IMAGIC-5	Image Science, Berlin, Germany
John Henry	Boris Busche (Stark lab)
matplotlib	(Hunter, 2007) http://matplotlib.org/
NumPy	http://www.numpy.org/
Python 2.6.5	Python Software Foundation, Beaverton, OR, USA http://www.python.org
TVIPS EM-MENU	Tietz Video System, Gauting, Germany

Continued on next page

Table 2.5 – *Continued from previous page*

<i>Program</i>	<i>Source</i>
UCSF Chimera	Resource for Biocomputing, Visualization, and Informatics (RBVI) at the University of California, San Francisco (Pettersen et al., 2004)

2.2 Methods

2.2.1 Purification of RNP complexes

2.2.1.1 HeLa spliceosomal C complex

HeLa spliceosomal C complex was purified and kindly provided by Dr. Sergey Bessonov and Dr. Elmar Wolf (Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany) as previously described (Bessonov et al., 2008). In summary, ^{32}P -labeled PM5 RNA, which had a 5'-MS2 stem-loops but lacked a 3' splice site, was incubated in the nuclear extract, so that the splicing reaction was stalled before the exon ligation. Then a sedimentation purification step was performed using a glycerol gradient, and the fractions containing spliceosomes were identified by radioactivity counts. Prior to the splicing reaction, PM5 RNA was pre-incubated with an MS2-MBP fusion protein. After fractionation, affinity purification over an amylose column which binds the MBP tag was used. MS2-MBP-tagged C complexes were eluted with maltose. Eluted sample was prepared for EM using the GraFix methodology.

2.2.1.2 Optimization of crowding agent and selective precipitation of human snRNP

HeLa nuclear extract was kindly provided by Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany. Nuclear extract was

dialyzed against glycerol, glucose, or sucrose. Dialyzed nuclear extract was precipitated with PEG6000, from 0 to 21%. The precipitated pellets were resuspended, and the RNA content was extracted by phenol. Urea polyacrylamide gel (10%) electrophoresis was performed to analyze the snRNA.

2.2.1.3 Isolation of the yeast snRNP

Yeast cells (*Saccharomyces cerevisiae*) were harvested at an OD600 of 2. Harvested cells were washed with Millipore water, and were diluted in the Roeder D buffer to a volume twice as big as the original cell volume, with 20% (w/v) sucrose, 10 mM DTT, 10% PMSF, and 10 mM benzamidine. Cell suspension was frozen in liquid nitrogen as small droplets and fractured with a ZM 200 Ultra Centrifugal Mill (Retsch, Haan, Germany). Cell debris in the thawed extract such as cell walls was pelleted by centrifugation at 30,000 g for 30 minutes, with 10 mM N-ethylmaleimide, 10 mM iodoacetamide, 10 mM benzamidine hydrochloride hydrate added, followed by centrifugation at 100,000 g for 1 hour to obtain an S100 extract. 7% (v/v for the final concentration) of PEG300 was added and incubated with S100 extract for 20 minutes to precipitate impurities, which was later pelleted and removed by centrifugation at 30,000 g for 30 minutes. The snRNP in the supernatant was then precipitated with 20% of PEG300 in total for 1 hour, followed by centrifugation at 15,000 g for 20 minutes. The pellet was resuspended in Roeder D buffer with 5% sucrose containing 5 mM DTT in addition. The sample solution was fractionated on a 10 – 30% sucrose gradient (SW40 rotor (Beckman Coulter, Brea, USA), 33000 rpm, 16 hours), and the snRNP peak was identified by SDS-PAGE analysis. The fractions containing snRNP were pooled and concentrated by precipitation with 15% PEG400 and resuspension. Then the sample solution was fractionated on a second 10 – 30% sucrose gradient (SW60 Ti rotor (Beckman Coulter), 28500 rpm, 16 hours). The peak fractions were collected, precipitated with 20% PEG300, resuspended in Roeder D buffer containing 0.001% (w/v) LMNG (Affymetrix) but without DTT, and ready for the further GraFix pro-

cedure. Every step was performed at 4°C, except for the resuspension of PEG pellets, which was performed in room temperature.

2.2.1.4 Isolation of the HeLa 80S ribosome under high and low salt conditions

HeLa ribosomes prepared under high salt conditions were kindly provided by Dr. Ashwin Chari. HeLa cytosolic extract supplemented with protease inhibitors (PMSF, N-ethylmaleimide, iodoacetamide, benzamidine) was centrifuged at 30,000 g for 30 minutes to yield the S30 supernatant. The S30 was pelleted through a 30% sucrose cushion (500 mM KCl, 5 mM MgCl₂) in a Type 45Ti rotor (Beckman Coulter, Brea, USA) at 40,000 rpm for 5 hours. The pellet was resuspended in the high salt gradient buffer containing 150 mM KCl. The sample solution was loaded onto a 10 – 30% sucrose gradient containing 500 mM KCl. The fractions containing 80S ribosomes are collected and concentrated by pelleting with ultracentrifugation at 40,000 g for 12 hours. The ribosome pellet was resuspended in buffer with 50 mM Bis-Tris pH6.8, 50 mM KCl, 5 mM MgCl₂, 2 mM ATP, 2 mM DTT, and cycloheximide (10 µg/mL).

For the preparation under low salt condition, S30 was prepared and pelleted through a 30% sucrose cushion with 50 mM KCl and 10 mM MgCl₂ instead. To increase the solubility, the buffer to resuspend the pellet was adjusted to 150 mM KCl. Two steps of ultracentrifugation with 10 – 30% sucrose gradient (50 mM KCl, 5 mM MgCl₂) were performed to purify the sample. The fractions containing 80S ribosomes were collected and concentrated by pelleting with 10% PEG6000. Finally, the pellets were resuspended with buffer containing 10 mM HEPES-K pH7.5, 50 mM KOAc, 10 mM NH₄Cl, 2 mM DTT, 5 mM Mg(OAc)₂, and cycloheximide (10 µg/mL).

The potassium salt concentrations in every purification step are listed in table 2.6 and compared with the conventional purification under high salt condition:

Table 2.6:

<i>procedure</i>	<i>high salt preparation</i>	<i>low salt preparation</i>
cushion	500 mM KCl	50 mM KCl
resuspension after cushion	150 mM KCl	150 mM KCl
ultracentrifugation	500 mM KCl	50 mM KCl
resuspension and GraFix	50 mM KCl	50 mM KOAc

2.2.2 GraFix

There are several sources of disruption of particles during cryo-EM sample preparation, which should be taken care. For example, since sugars reduce the contrast in cryo-EM, the sugar, which stabilizes the molecule, needs to be removed before vitrification. In addition, a carbon film is used to adsorb the molecule for the sample grid, but a disrupting force can be created during the interaction. Therefore, fixatives have been commonly used for crosslinking the cryo-EM sample. An optimized fixation procedure named *GraFix* was established in our lab and has become a general procedure (Kastner et al., 2008).

In GraFix, the fixative is prepared as a gradient along with the sugar solution, and ultracentrifugation is performed. Particles do not suddenly encounter high concentration of fixatives, so the intramolecular crosslink can be formed mildly before aggregation occurs. In addition, the centrifugal force can disrupt unwanted intermolecular crosslinking. Even if aggregation unfortunately occurs, the aggregate migrates to the bottom of the gradient due to its higher sedimentation coefficient, thus does not reduce the final sample quality. For highly heterogeneous samples such as the spliceosomal B complex, the GraFix method has shown to significantly stabilize the molecules and to increase the image quality in EM images (Kastner et al., 2008).

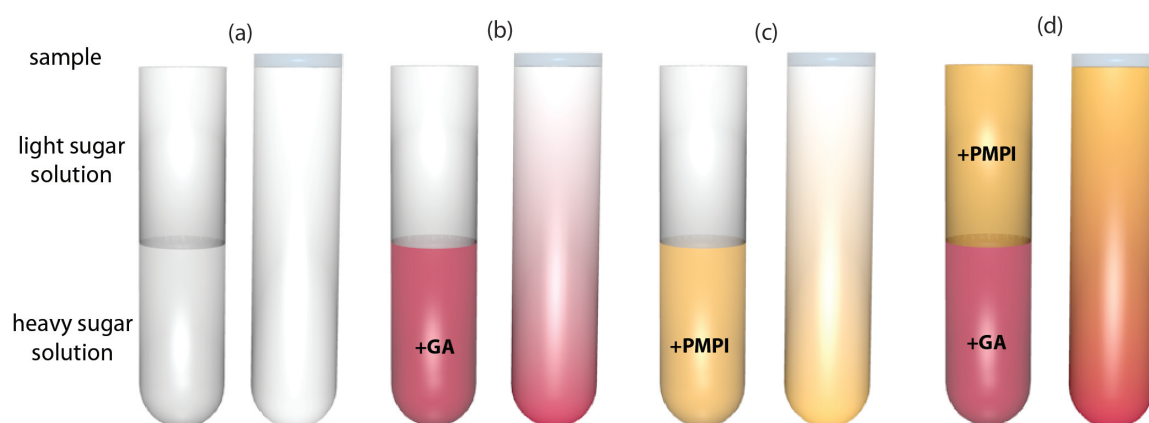


Figure 2.1: Preparation of variant GraFix methods. Fixatives can be added into the light or heavy sugar solutions (left), followed by gradient forming (right). (a) Sugar gradient without any fixatives. (b) Conventional GraFix method with glutaraldehyde as fixative in the heavy sugar solution. (c) PMPI as fixative in the heavy sugar solution. (d) Double fixation with PMPI in the light sugar solution and glutaraldehyde in the heavy sugar solution.

One of the most commonly used fixatives in biological EM is glutaraldehyde. Glutaraldehyde mainly crosslinks the ϵ -amino groups of lysines, but was observed to crosslink tyrosines, histidines, and sulfhydryl residues in some reactions (Habeeb and Hiramoto, 1968). However, for cryo-EM studies on RNP complexes, fixatives for RNA-protein crosslinks are not commonly utilized. *P*-maleimidophenyl isocyanate (PMPI) was used innovatively here as an RNA-protein crosslinker. PMPI crosslinks a thiol group with a hydroxyl group, which in our case would be cysteine and the 2' hydroxyl group on the ribose backbone of the RNA in the ideal situation. However, it can also react with water and hence loses crosslinking activity for the hydroxyl group (Annunziato et al., 1993). Therefore, DMSO was used here as a solvent to predissolve PMPI. The chemical formulas of glutaraldehyde and PMPI are shown in figure 2.2.

The conventional GraFix method uses the glutaraldehyde as a single fixative. Here

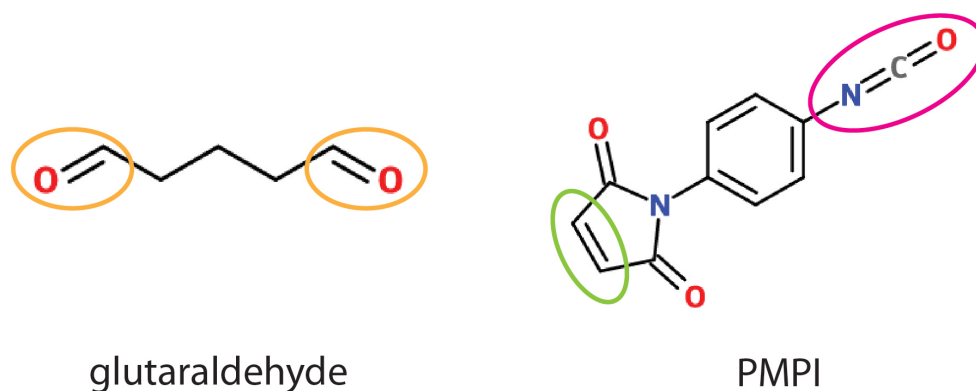


Figure 2.2: Chemical formulas of glutaraldehyde and PMPI. The functional groups of glutaraldehyde which react with amino groups of lysines are circled in orange. In PMPI, the functional group which reacts with the hydroxyl group is circled in pink, and the functional group which reacts with the thiol group is circled in green. Chemical formulas were drawn with eMolecules (<http://www.emolecules.com/>).

PMPI was used as the single fixative, or in a reverse gradient with glutaraldehyde to perform a double fixation (figure 2.1). Fixatives were added into light or heavy sugar solutions, and the gradient was prepared by a gradient former (Gradient Master, Bio-Comp Instruments, Canada). Glutaraldehyde was added into the heavy sugar solution in 0.1% (v/v) final concentration. PMPI was predissolved in DMSO, and then added to the heavy or light solution for a final concentration of 0.2 mg/mL. Samples were loaded on the gradients immediately after the gradients were prepared. The gradients were centrifuged in a Sorvall Centrifuge Evolution RC (Thermo Electron, Langensfeld, Germany) or Optima L-100XP Ultracentrifuge (Beckman Coulter, Brea, USA), with a Sorvall TH660 or Beckman SW60 Ti rotor respectively. The temperature was maintained at 4 °C through the whole process, and the centrifugation conditions are listed in table 2.7:

Gradients were harvested manually into 200 μ L fractions after GraFix. To reduce the inter-molecular reaction from excess fixatives, each fraction was quenched with 25 mM final concentration of aspartate (for glutaraldehyde) or cysteine (for PMPI).

Table 2.7:

<i>Sample</i>	<i>Sugar</i>	<i>Gradient(%_{w/v})</i>	<i>Speed(rpm)</i>	<i>Time(hrs)</i>
Hela ribosome	sucrose	15 – 45	22500	14
HeLa spliceosomal C complex	glycerol	10 – 30	18500	18
yeast snRNP	sucrose	10 – 30	28500	16

2.2.3 EM sample preparation

2.2.3.1 Preparation of continuous carbon film and holey carbon grids

Carbon films and grids were kindly prepared by Frank Würriehausen.

Mica was here used as the coating carrier for continuous carbon films. In order to obtain a smooth surface which does not have any direct air contact before, mica (Plano G250-1, 25x75 mm) was freshly split. With the fresh side exposed, the mica was placed in an Edwards E12E vacuum coating unit (BOC Edwards, Kirchheim, Germany). Electric current was conducted through two carbon rods (Ringsdorff Werke GmbH, Bonn, Germany) as electrodes in the device, and carbon was sputtered from the contact surface of the carbon rods. To prevent heterogeneous distribution of carbon on the mica, carbon was coated on the mica only indirectly. This was achieved by sheltering mica under a piece of metal to prevent direct exposure, and the sputtered carbon was reflected onto the mica.

For evaluating sample quality, low-cost grids with irregular holes were prepared. Self-made triafol film with holes was covered on copper grids. Then carbon was directly sputtered as previously described onto this triafol film. To dissolve and remove the triafol film between the carbon layer and the copper grids, these grids were incubated in ethyl acetate overnight and air-dried.

In order to record images which were used for further analysis and 3D reconstruction, Quantifoil grids (Quantifoil Micro Tools GmbH, Jena, Germany) were used. Triafol

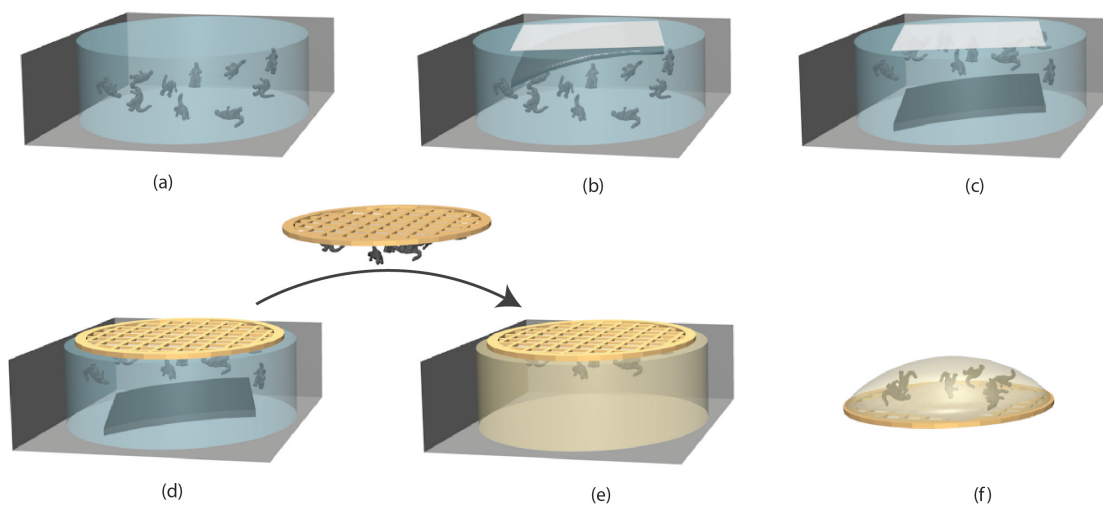


Figure 2.3: Preparation of negative stained samples. (a) Sample solution was loaded into a well of a homemade teflon block. (b) A piece of carbon-coated mica was floated on the top of the well. (c) Floated thin carbon film adsorbed sample molecules over time. (d) (e) The carbon film was transferred to a new well which contained uranyl formate by a holey-carbon coated copper grid. (f) After removing residual stain solution, the sample grid was left air-dried.

films with regular sizes of holes (hole radius= $3.5 \mu\text{m}$ with $1 \mu\text{m}$ spacing or hole radius= $2 \mu\text{m}$ with $2 \mu\text{m}$ spacing) were already coated on the grids. Procedures for carbon coating were as previously described.

2.2.3.2 Preparation of negative stained samples

Negative staining, which provides high contrast with a relatively simple preparation procedure, is best suited for quick sample quality estimation, as well as yielding a starting structure for an unknown molecule. 2% uranyl formate was dissolved in double distilled water. The solution then was centrifuged at 13000 rpm for 20 minutes at 4°C in order to pellet excess and crystallized uranyl formate. $100 \mu\text{L}$ of the stain supernatant was transferred to a well of a home-made teflon block. $25 \mu\text{L}$ of sample

solution was filled in another well. A piece of carbon coated mica was floated on the surface of the sample well to adsorb molecules. Different sample concentrations require different adsorption times. An optimum adsorption time is obtained when particles fully distribute on the carbon but do not attach to each other. After adsorption, a holey-carbon coated copper grid was used to lift out the carbon film, with the carbon side of the grid down. The grid was transferred to the well of stain immediately with forceps, and was incubated on the stain for 1 minute. To maintain a decent thickness of the stain, the residual stain on the grid was removed by blotting with KimWipes (Kimberly-Clark Corporation). The negative stain procedure is illustrated in figure 2.3.

2.2.3.3 Preparation of unstained cryo samples

One of the advantages of cryo EM is that it preserves the native state of biomolecules in aqueous solution. Besides, the unstained cryo sample also offers the possibility to reach higher resolution compared to negatively stained samples. Therefore, samples with a known structure were prepared with unstained cryo procedure, which is illustrated in figure 2.4. The sugars present in the sample solution result in a reduction of the contrast for EM imaging, thus buffer exchange of the sample solution was performed with Zeba spin desalting columns, 7K MWCO (Thermo Fisher Scientific Inc.). Similar to the negative stain procedure, carbon coated mica was floated on the surface of the sample solution in a well, and the thin carbon film was incubated for a certain time to adsorb proper amounts of particles. Then the carbon film was lifted out onto a Quantifoil copper grid, and 3 to 5 μL of distilled water was applied to the grid. The vitrification procedure was executed with a Vitrobot Mark IV (FEI). To maintain the thickness of ice which provides the best contrast, excess sample solution was removed by blotting with filter papers. Then the grid was immediately plunged into liquid ethane cooled by liquid nitrogen to vitrify. The particles should be embedded in a thin layer of amorphous ice when the vitrification procedure is quick enough, so that the

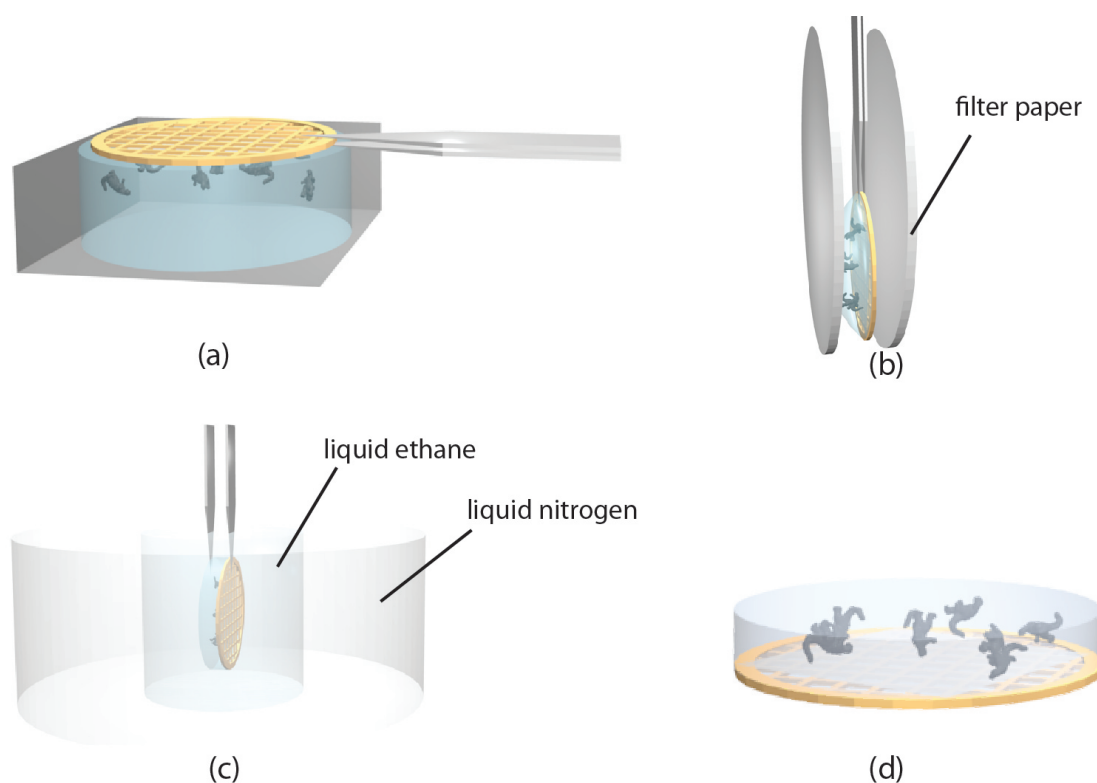


Figure 2.4: Preparation of unstained cryo samples. (a) Floated carbon film with adsorbed molecules was fished with a Quantifoil copper grid by forceps. (b) Residual buffer was blotted with filter papers. (c) The grid was immediately plunged into liquid ethane which was cooled down by liquid nitrogen. (d) Vitrified sample grid was transferred and stored in liquid nitrogen.

water did not have time to form crystals. The cryo-grids were stored in liquid nitrogen before EM imaging.

2.2.4 EM analysis

2.2.4.1 Image acquisition with TEM

Negative stained samples were imaged at room temperature with a Philips CM200 FEG TEM at an acceleration voltage of 160 kV. Images were recorded on a 4K x 4K

Charge Coupled Device (CCD) camera (Tietz Video Systems, Gauting, Germany). Cryo samples were imaged in liquid nitrogen temperature with a Titan Krios TEM (FEI) at an acceleration voltage of 300 kV on a 4K x 4K Eagle 4k CCD camera or a Falcon 2 CMOS Direct Electron Detector camera (FEI). Images were taken at -1.5 to -3 μm defocus. Parameters of image recording for different sample are listed here in table 2.8:

Table 2.8:

<i>Sample</i>	<i>Preparation</i>	<i>Camera</i>	<i>Magnification</i>	<i>Binning</i>	<i>Pixel size</i>
Hela ribosome (high salt)	cryo	Eagle	59000	2	2.6 Å/pix.
Hela ribosome (low salt)	cryo	Falcon 2	37000	1	2 Å/pix.
HeLa spliceosomal C complex	cryo	Eagle	47000	2	3 Å/pix.
yeast snRNP	negative	CCD camera	66000	2	3.3 Å/pix.

2.2.4.2 Image analysis

Particles from the TEM micrographs were selected semi-automatically with JohnHenry (custom-made software, Boris Busche). Image processing was mainly performed with a combination of IMAGIC-5 (Image Science) and CowEyes (custom-made software). The general routine procedure was illustrated in figure 1.4. An overview for image processing of different samples is listed here in table 2.9:

CTF correction The softwares used in this work were *defcorr_inc* (Sander et al., 2003) and *CTF correction module of CowEyes* (Boris Busche, unpublished). In these two softwares, CTF correction works on averaged power spectra of single particles. The class averages of power spectra are approximated by the CTF correction softwares

Table 2.9:

<i>Sample</i>	<i>CTF correction</i>	<i>Methods to obtain the initial structure</i>
Hela ribosome	yes	class averages obtained from reference-free alignment <i>de novo</i> angular reconstitution
HeLa spliceosomal C complex	yes	initial structure from random conical tilt as reference (Golas et al., 2010) class averages obtained from alignment with references angular reconstitution with known angular information
yeast snRNP	no	class averages obtained from reference-free alignment angular reconstitution using angular information from projections of the yeast spliceosome B ^{act} model

with variable parameters. CTF parameters such as astigmatism, defocus, B-factor, and amplitude-contrast proportion were estimated, and the phase correction was performed. Notably, the cryo TEM in our lab is equipped with a Cs corrector, so the spherical aberration Cs value was set close to zero for the CTF approximation here. The envelope decay was only corrected (as amplitude correction) at late stages of refinement directly on the 3D model. Because in the early stage of image processing, without reliable alignment and averaging, noise and structural details might be undistinguishable, it can lead to over-amplification of the noise.

Refinement Refinements of all models were performed manually, except the final model of GraFixed HeLa 80S ribosome purified under low salt condition was automatically refined by RELION (Scheres and Chen, 2012), with the structural heterogeneity resolved.

Chapter 3

Results

3.1 Structural Analysis of Human Spliceosomal C Complex

The human spliceosomal C complex was previously reconstructed by single particle cryo-EM using random conical tilt (RCT) (Golas et al., 2010). However, the final resolution was limited to 20 – 29 Å due to the heterogeneity and flexibility. In order to better resolve the heterogeneous states and further improve the resolution, a larger dataset consisting of about 200,000 particles was recorded. Class averages were calculated without any reference, and projections of the RCT model were used as an anchor set for angular reconstitution. The reconstructed model is shown in figure 3.1. The structure of the model, especially the “head” part was not well defined, and this was due to averaging particles representing different heterogeneous states.

To overcome this heterogeneity and to elucidate different states, resampling and 3D MSA classification were performed. 3D class average models were examined carefully, and those with distinct features were selected as references for supervised classification. After several rounds of refinement, a variance on the “belly” was observed. Moreover, it seemed that in most cases the integrity of the “belly” is related to the unity of the

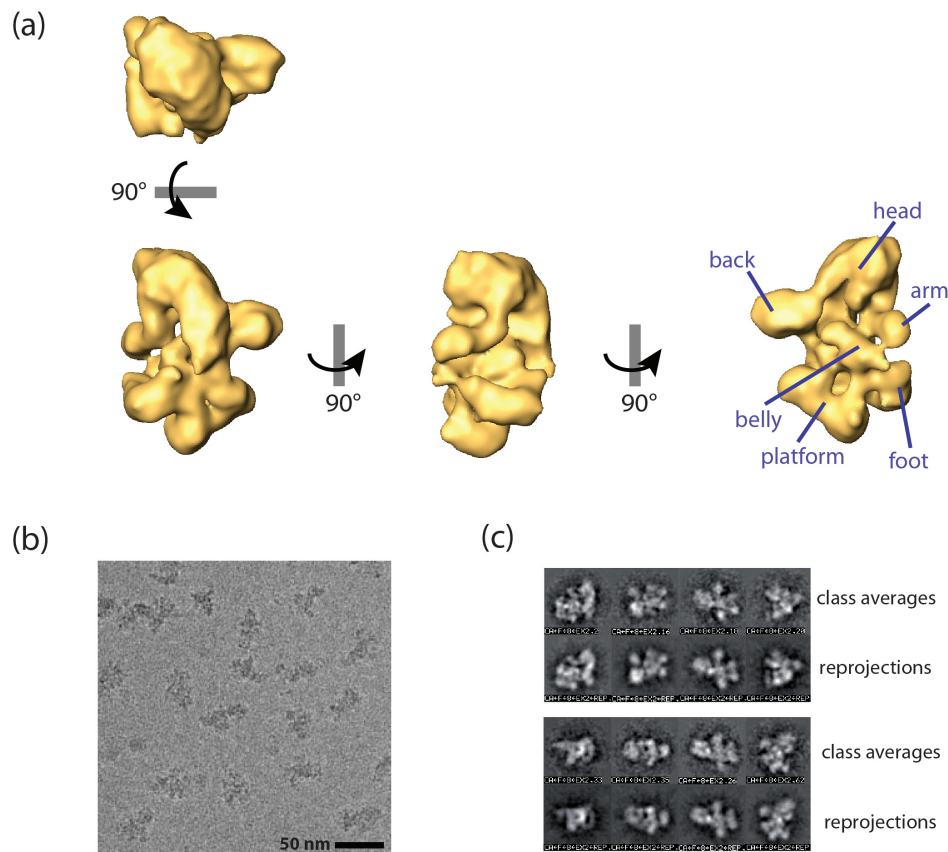


Figure 3.1: Structure of human spliceosomal C complex. (a) The 3D model consists of 33000 particles. (b) An electron micrograph of vitrified C complex. (c) Representative class averages which were used to reconstruct the 3D and their reprojections.

overall structure. With this result, supervised classification was performed, and particles having the “belly” density, which were supposed to be more stable, were selected for further analysis (figure 3.2(a)). As the “head” part displayed the most dominant heterogeneity, 3D MSA classification was performed focusing on this region. The observed degree of heterogeneity was greater than expected: rather than few states of definite variance, it showed many hardly resolvable states without a systematic pattern. In addition, heterogeneity or flexibility was also observed in the “platform” region. In some of the states, the “platform” density had a contact with the “back” region

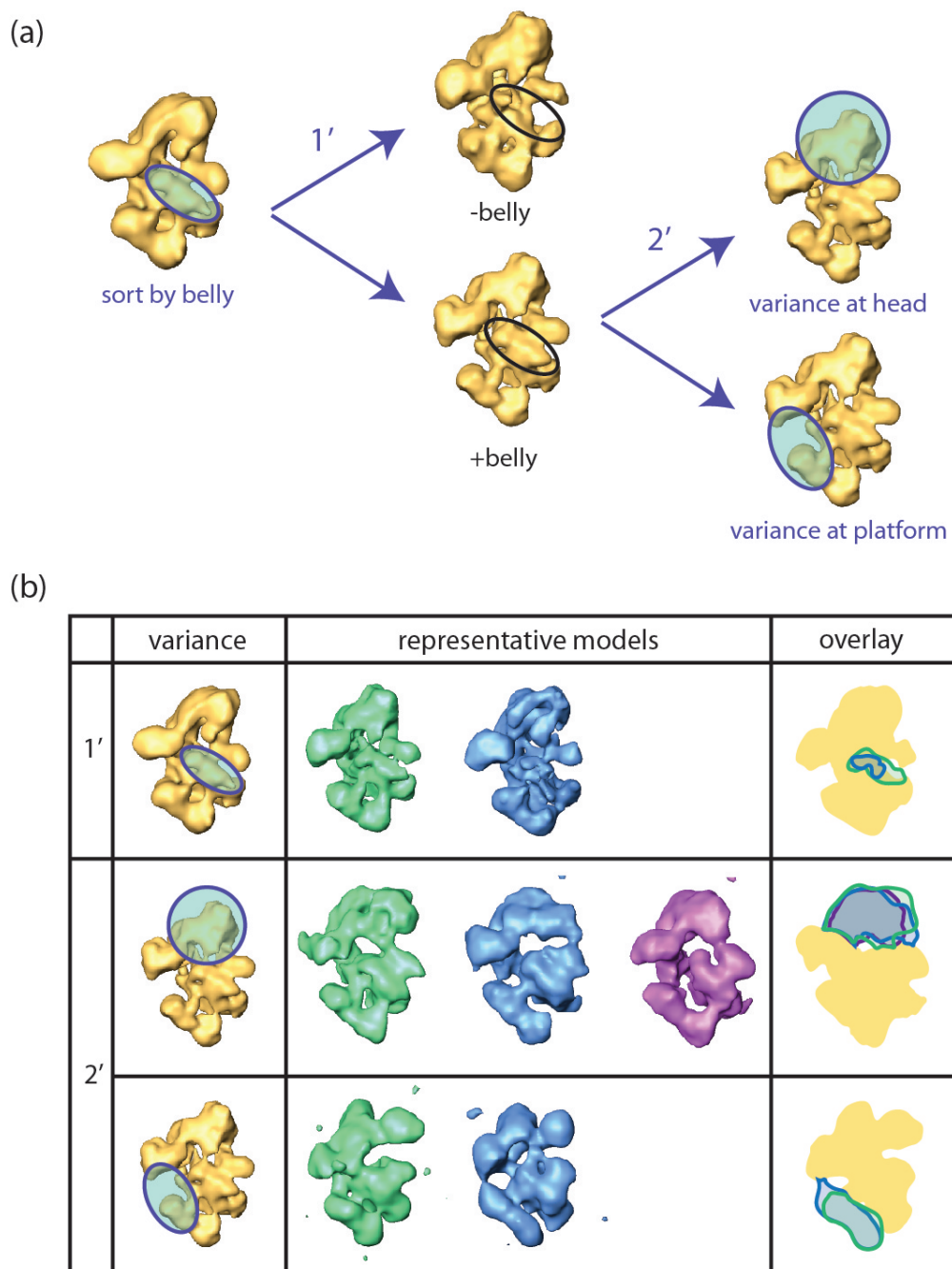


Figure 3.2: Analysis on structure variance of human spliceosomal C complex. (a) Sorting scheme of C complex structures. (b) Structural variance of the C complex. Models were generated by resampling and 3D MSA classification, and refined with supervised classification. Representative refined 3D models are listed here in green, blue, and purple. For comparison purpose, contours of models are overlaid (most right column), and the variance of interest is emphasized with corresponding colors.

(figure 3.2(b)). The *in silico* purification was continued to resolve the heterogeneity further until the amount of particles within one group reached the lowest limit for 3D reconstruction. Unfortunately, the heterogeneity still could not be resolved in this step.

Both conformational and compositional heterogeneity were highly present in the C complex structure, and the heterogeneity could not be sufficiently resolved *in silico* with a limited amount of particles and the current image processing tools available. This might be due to the biochemical instability of the sample, which has exceeded the applicable level for *in silico* sorting. Therefore, reducing the heterogeneity before image acquisition seemed to be an essential task. This can be achieved by two different approaches: stabilizing the biomolecules during purification, or prevent molecules from disruption during the sample preparation for TEM. For particle stabilization during purification, we mainly focused on optimization of the crowding agents (see the next section: 3.2) as well as reduction of the salt concentrations (see section 3.4). Besides, chromatography was replaced by selective precipitation (discussed in the next section: 3.2). For particle stabilization during TEM sample preparation, an innovative RNA-protein crosslinker for TEM sample was applied (discussed in 3.3). To simplify the experiment design, optimization of the purification procedure was performed using endogenous snRNP as the experimental material instead of the C complex. This is because that purification of the C complex involves trapping a time-dependent assembly, while the endogenous snRNP was already assembled *in vivo*.

3.2 Purification and Structure of Yeast Endogenous snRNP

Based on the assumption that the integrity of the complex supports structural stability, in order to obtain a homogeneous and stable structure, the sample purification method could be optimized to minimize the damages during the procedure. Optimization of endogenous snRNP purification was here focussed on two topics: improvement

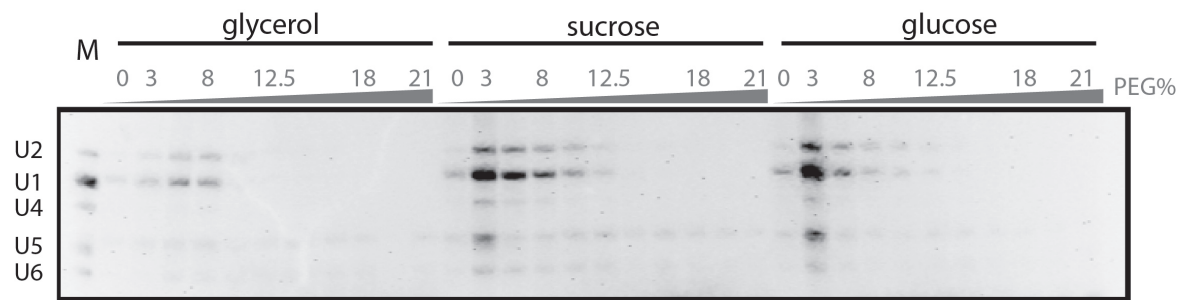


Figure 3.3: Human snRNA composition of precipitates from different concentration of PEG6000 in different sugar environments.

of crowding agent, and establish a chromatography-free purification procedure.

Affinity-based procedures have the advantage of high specificity for purification, and has been commonly applied to purify snRNPs and spliceosomes (Behrens and Lührmann, 1991). However, surface interactions between the antibody and the molecule might compete with the intramolecular interaction, therefore the assembly of the complex might be disrupted. There are two commonly used purification methods in which no chromatography is involved. One is to sediment the target molecules through a sugar cushion by ultracentrifugation. This method has been used for ribosome purification (Belin et al., 2010). However, this method requires high abundance of the molecule in the cell, therefore might not apply for the much less abundant snRNPs. The other method is selective precipitation. Different approaches of protein selective precipitation were described in detail by Lovrien and Matulis (2001). In this work, polyethylene glycol (PEG) was used for selectively precipitating endogenous snRNPs. For more than half century, PEG precipitation has been applied for protein concentration and purification (Stocking, 1956). PEG is inert and does not interact directly with proteins; instead, PEG occupies the solvent so that the proteins have reduced solubility and precipitate (Atha and Ingham, 1981).

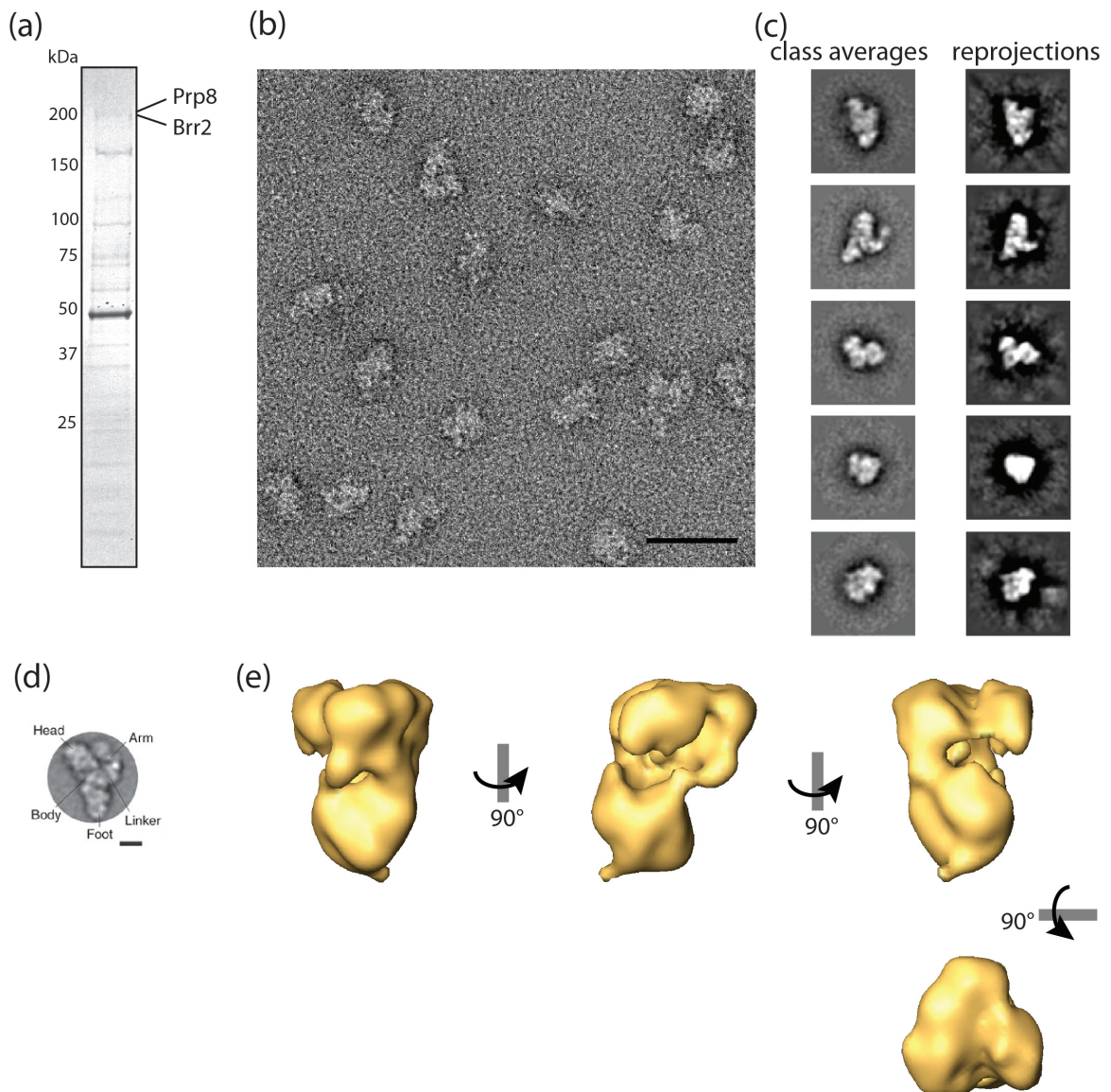


Figure 3.4: Structural analysis of yeast snRNP. (a) Endogenous snRNP was identified with the Prp8/Brr2 double bands on the SDS-PAGE. (b) Raw TEM micrograph of the yeast snRNP. The scale bar represents 30 nm (c) Representative class averages of yeast snRNP and corresponding re-projections from the 3D model. (d) One of the class averages from previously calculated 2D images of yeast tri-snRNP for comparison. The scale bar represents 10 nm. Reprint from (Häcker et al., 2008) with permission. (e) 3D model of the yeast snRNP.

Glycerol has been generally used in our standard snRNP purification method as crowding agent in the buffer and for gradient sedimentation (Häcker et al., 2008). To test other crowding agents, HeLa nuclear extract was dialyzed against glycerol, sucrose, or glucose. Different amounts of PEG6000, from 3% to 21% (w/v), were added to the extract. Most of the snRNP precipitated at 3%, and at 12.5% almost all the snRNP was precipitated. In addition, higher quantity of snRNP was recovered in the extract containing sucrose compare to the standard glycerol (figure 3.3).

Due to the lack of availability of HeLa nuclear extract, the experiments were performed with yeast. During the purification, yeast endogenous snRNP was identified with the Prp8/Brr2 double bands on the SDS-PAGE as well as single particle cryo-EM. For the result of optimization, three steps of sucrose gradient, together with precipitation by combination of PEG300 and PEG400, was able to give a satisfactory yield, in terms of quantity and quality. For cryo-EM analysis, the density of particle distribution like figure 3.4(b) required only 1 minutes for the carbon film adsorption. In comparison, immunoaffinity methods with anti-TAP column and the standard anti-m₃G antibody (Behrens and Lührmann, 1991) have also been tested starting with a comparable amount, but led to a much larger loss of quantity. Even with an overnight carbon film adsorption, the particle distribution never reached the same level. According to the sedimentation in the 10 – 30% sucrose gradient, the size of the purified yeast snRNP is about 26S. The mass spectrometric analysis suggested that the purified snRNP contains protein from U4, U5, and U6 snRNP (appendix A.1).

Around 20,000 negative stained particles of yeast snRNP were imaged with the electron microscope and analyzed (figure 3.4(b, c)). The calculated 2D class averages were structurally very comparable to the known yeast tri-snRNP (Häcker et al., 2008)(figure 3.4(d)). The initial low resolution 3D structure of yeast snRNP was calculated from the class averages(figure 3.4(e)).

3.3 Crosslink RNP with PMPI: HeLa 80S ribosomes as an example

Besides sample purification, the subsequent sample preparation for EM imaging can cause destabilization of the particles. For example, the forces the carbon foil applies on the particles during adsorption are strong enough to disrupt the particles structural integrity. Therefore, crosslinkers have been applied as fixatives to enhance the intramolecular stability. The GraFix method provides good intramolecular fixation, while the intermolecular crosslinks, yielding aggregates, are significantly suppressed (Kastner et al., 2008). Glutaraldehyde has been commonly used in the GraFix procedure as a protein-protein crosslinker. In contrast, even though RNP complexes such as spliceosomes and ribosomes have been popular targets for structural studies in the field of single particle EM, fixatives for RNA-protein crosslinks are not commonly utilized. Here the fixative *p*-maleimidophenyl isocyanate (PMPI) was innovatively used as an RNA-protein crosslinker for EM samples. However, the hydroxyl groups of the sucrose in the gradient solution can also potentially be crosslinked instead of the 2' hydroxyl group of the RNA. Moreover, PMPI reacts with water and loses the activity for hydroxyl crosslinking (Annunziato et al., 1993), therefore it has been only used on small molecules with DMSO as a solvent, mainly for immunochemistry (Fox et al., 2004; Wang et al., 2007). The intramolecular interaction in the RNP macromolecules is much more complicated than the small molecules, so we tried to test if PMPI can serve as a fixative for large RNP complexes. In the following experiments, RNP complexes were dissolved in regular water-based buffers instead of DMSO. To better preserve the reactivity of PMPI, the PMPI powder was initially dissolved in DMSO, and freshly added into the sugar solution just before gradient preparation and centrifugation.

The eukaryotic 80S ribosomes were selected as the model system for proof of concept experiments for testing the crosslinking effect of PMPI. Ribosomes are relatively

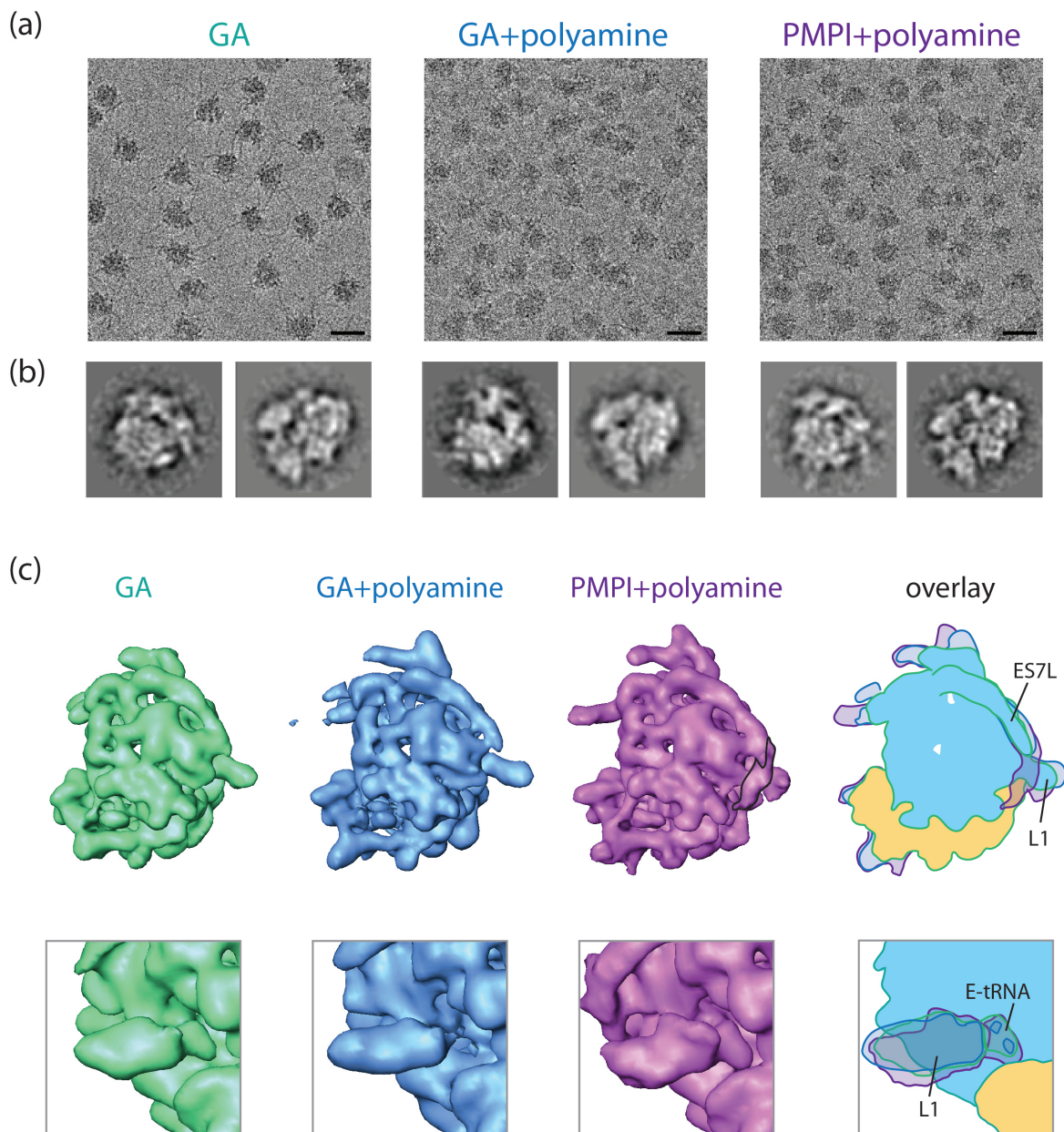


Figure 3.5: Structures of HeLa 80S ribosomes under the effect of polyamine and PMPI. (a) Raw micrographs and (b) class averages of the ribosomes with different GraFix preparation. (c) Upper row: ES7L was stabilized by polyamine and PMPI. Lower row: the position of L1 stalk was altered and the occupancy of E-site tRNA increased in the presence of PMPI. Model for the glutaraldehyde control preparation is shown in green, glutaraldehyde with polyamine preparation in dark blue, and PMPI with polyamine preparation in purple. Overlays of the contours are displayed in the right-most column and illustrated with the corresponding colors. The block of light blue (60S) and yellow (40S) are painted according to the contour of the glutaraldehyde control model. The scale bar represents 30 nm.

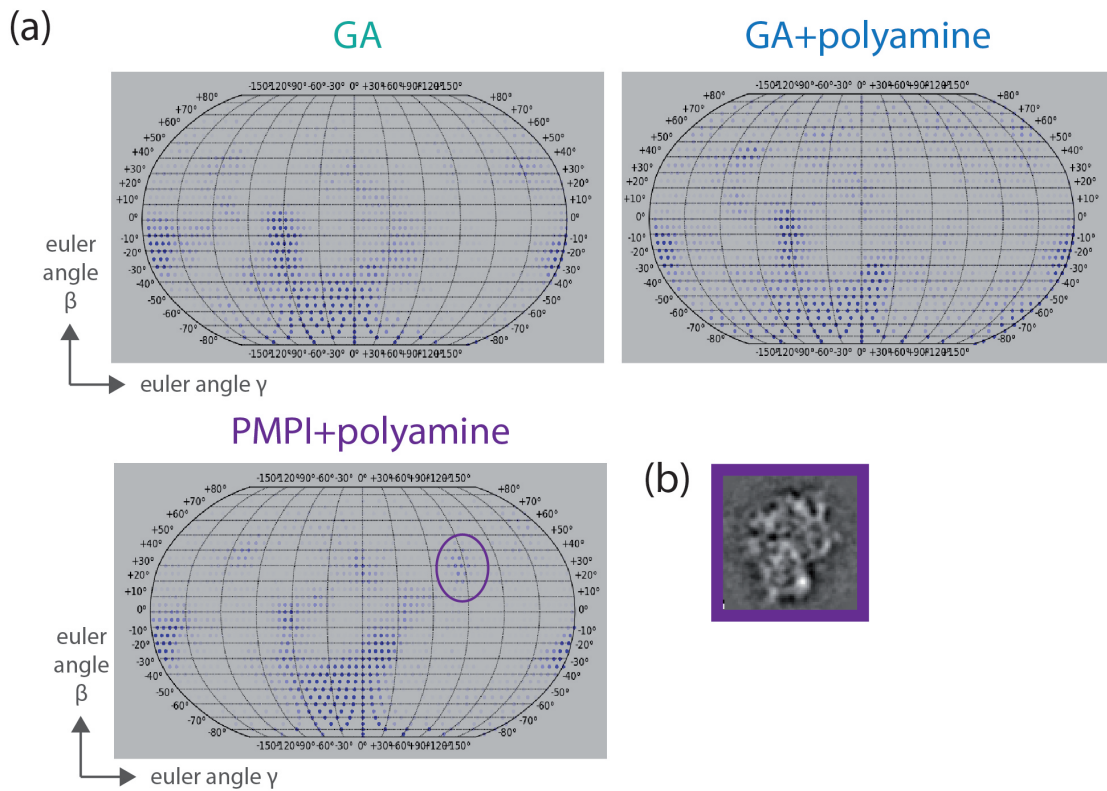


Figure 3.6: GraFix with PMPI might alter the Euler angle distribution. (a) Euler angle distribution of the ribosome models with different GraFix preparation. In the PMPI with polyamine preparation, a new population of the angular distribution was observed, and it is emphasized with a purple circle. A class average representing this orientation is shown in (b). The map of the angular distribution was plotted with matplotlib basemap toolkit.

stable and homogeneous compared to spliceosomes, but also present some degrees of flexibility, which can be further stabilized. Notably, eukaryotic ribosomes possess RNA expansion segments, which are exposed on the surface and known to be very dynamic. We assume that the RNA-protein crosslinking effect can be clearly observed on eukaryotic ribosomes. Besides PMPI, stabilization effects of polyamines were also tested here. Polyamines have been used in prokaryotic 70S ribosome for stabilizing the structure and minimizing motion (Schuwirth et al., 2005). The polyamines used here were a mixture of 0.6 mM spermine and 0.4 mM spermidine. Since the polyamines react

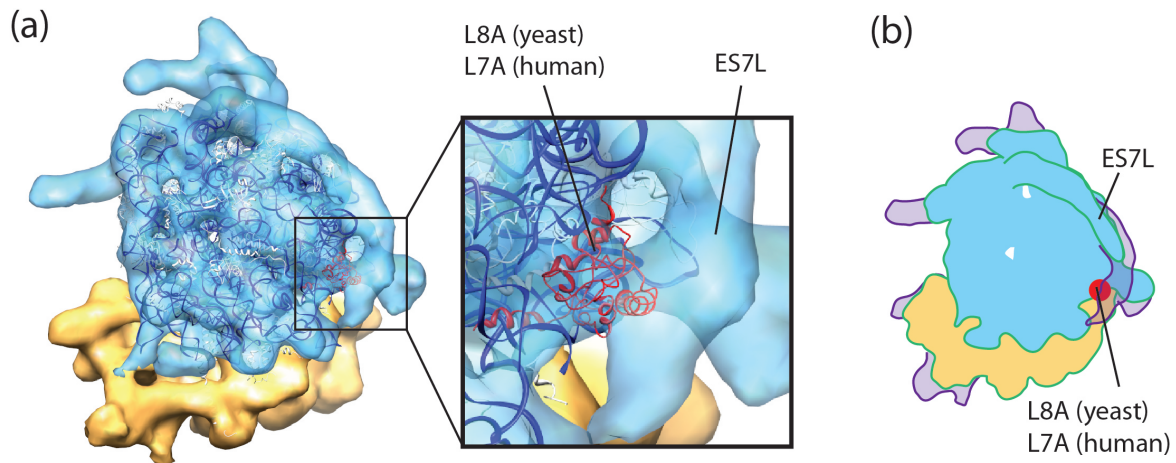


Figure 3.7: ES7L was crosslinked on the ribosome core with PMPI. (a) Crystal structure of yeast 60S ribosomal subunit (PDB: 3O5H) was fitted into 3D map of the HeLa 80S ribosome from the PMPI with polyamine preparation group. RNAs are shown in navy blue, and proteins are shown in white. The possible crosslinked protein L8A (L7A for the human homolog) is emphasized in red. Model fitting and rendering were generated in UCSF Chimera. (b) Contour illustration of the crosslinking effect from PMPI. Green contour with the blue and yellow color block is the contour from the glutaraldehyde control group, and the extra density of PMPI with polyamine preparation is illustrated in purple. The position of L7A is labeled with red.

with the glutaraldehyde and might reduce the efficiency of crosslinking, it is important to avoid co-presence of polyamines and glutaraldehyde. HeLa 80S ribosomes with the same purification method were separated into three groups, with different pre-GraFix, GraFix, and post-GraFix procedures. The differences are listed in table 3.1.

The 3D models of these three groups were reconstructed independently without any starting reference and carefully compared (figure 3.5). Overall, in the presence of the polyamines, longer segments of some RNA expansions could be visualized, indicating the stabilization effect on expansion segments. Moreover, PMPI further reduced the flexibility of the expansion segments, especially ES7L. In addition, in the PMPI sample with polyamines, the position of the flexible L1 stalk was altered, and more

Table 3.1:

<i>procedure</i>	<i>glutaraldehyde control</i>	<i>glutaraldehyde + polyamine</i>	<i>PMPI + polyamine</i>
incubation with polyamine before GraFix for 2 hr	–	+	+
polyamine in gradient solution	–	–	+
fixative	glutaraldehyde	glutaraldehyde	PMPI
after GraFix	–	–	2 hr post-fixation with glutaraldehyde
after buffer exchange	–	+ polyamine	+ polyamine

structural details were observed. The E-site tRNA occupancy was also higher in this group, suggesting less structural heterogeneity in this location due to stabilization of E-site tRNA binding.

For most of the biomolecules, the angular distribution of projections is not homogeneous. During sample grid preparation, particles are adsorbed onto the carbon film with preferred orientations, which are mostly with larger attached surface. The angular distributions of the three groups were plotted and compared in figure 3.6 (a). One population of the angular distribution was observed exclusively in the PMPI with polyamine group, and the class averages corresponding to this orientation is shown in figure 3.6 (b). Adsorption of particles in this orientation was originally not preferred,

unless a new landing surface was created by longer stabilized ES7L. The orientation of the class averages was coherent to this explanation.

The fact that ES7L becomes increasingly stabilized in the PMPI-based GraFix preparation can be explained by a crosslinking effect of PMPI. Since crystal structures of HeLa ribosome are not yet available, the crystal structure of yeast 60S ribosome (Ben-Shem et al., 2011) was fitted into the map (figure 3.7(a)). In spite of significantly different lengths of expansion segments, the core structure of ribosomes is well-conserved through evolution. The protein L8A (L7A for human homolog) presented at the location where the ES7L was crosslinked to the core, might be involved in such a potential RNA-protein crosslink. From the amino acid sequence, there are three cysteines in the human L7A protein. Unfortunately, the resolution for this EM map was not high enough to indicate the exact crosslink-position.

3.4 Structure of HeLa 80S ribosome purified under low salt condition

Electrostatic interaction is considered to be a significant force to maintain the assembly of an RNP complex. The RNA is negatively charged and can directly interact with positively charged surfaces of proteins. Therefore, under high-salt conditions, the intramolecular electrostatic interaction might be destabilized, and this might cause the dissociation of components or even the disassembly of the whole RNP complex.

The common purification procedures of eukaryotic ribosomes usually involve one or more steps of preparation with a high concentration of salt (e.g. 500 mM KCl or KOAc) (Dube et al., 1998). Since ribosomes are more rigid than most of the other

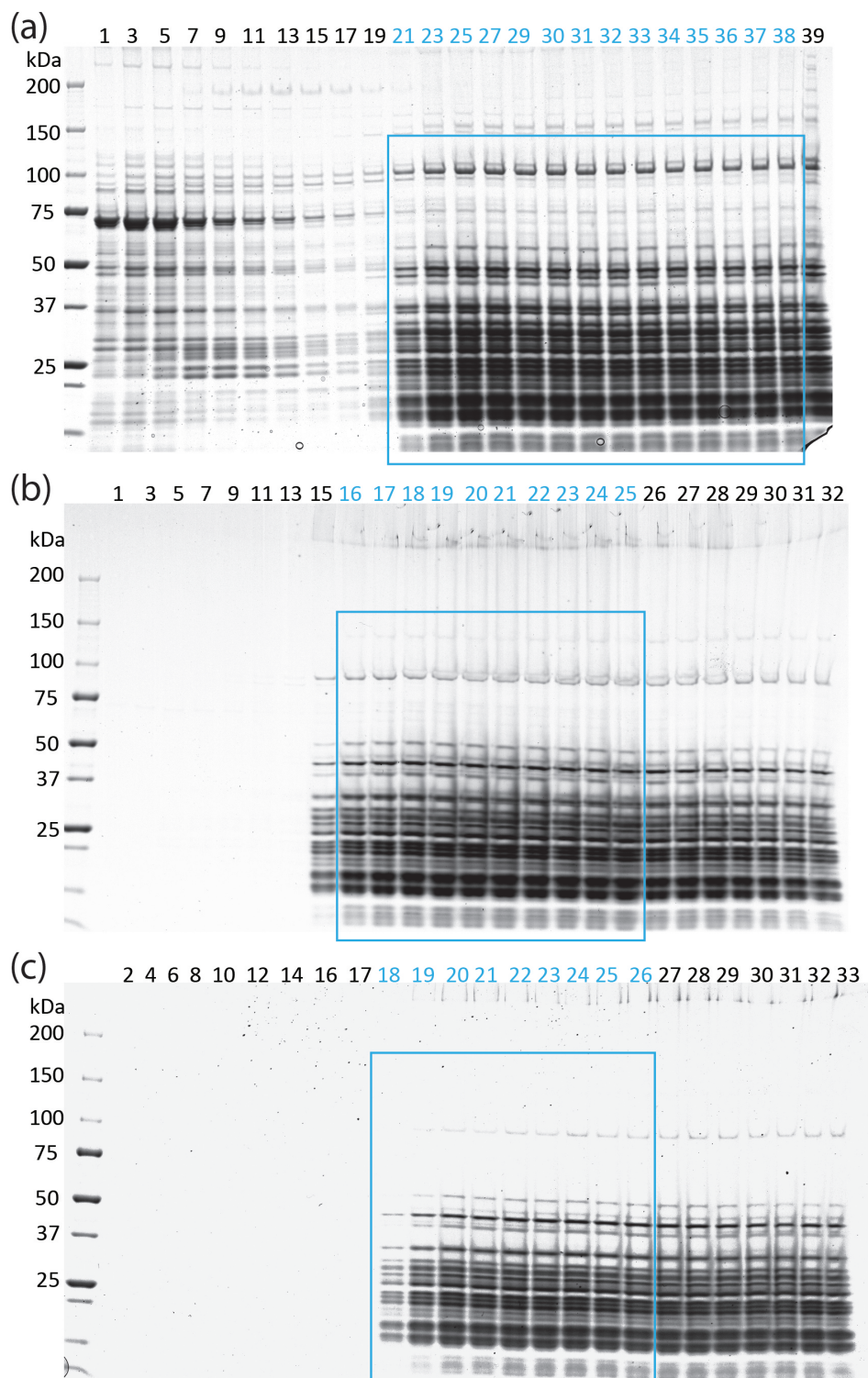


Figure 3.8: Three steps of ultracentrifugation gradient purification of the HeLa 80S ribosome: (a) the first step, (b) the second step, and (c) the third step. Fractions containing 80S ribosomes (labeled in blue) were collected and concentrated for the next step.

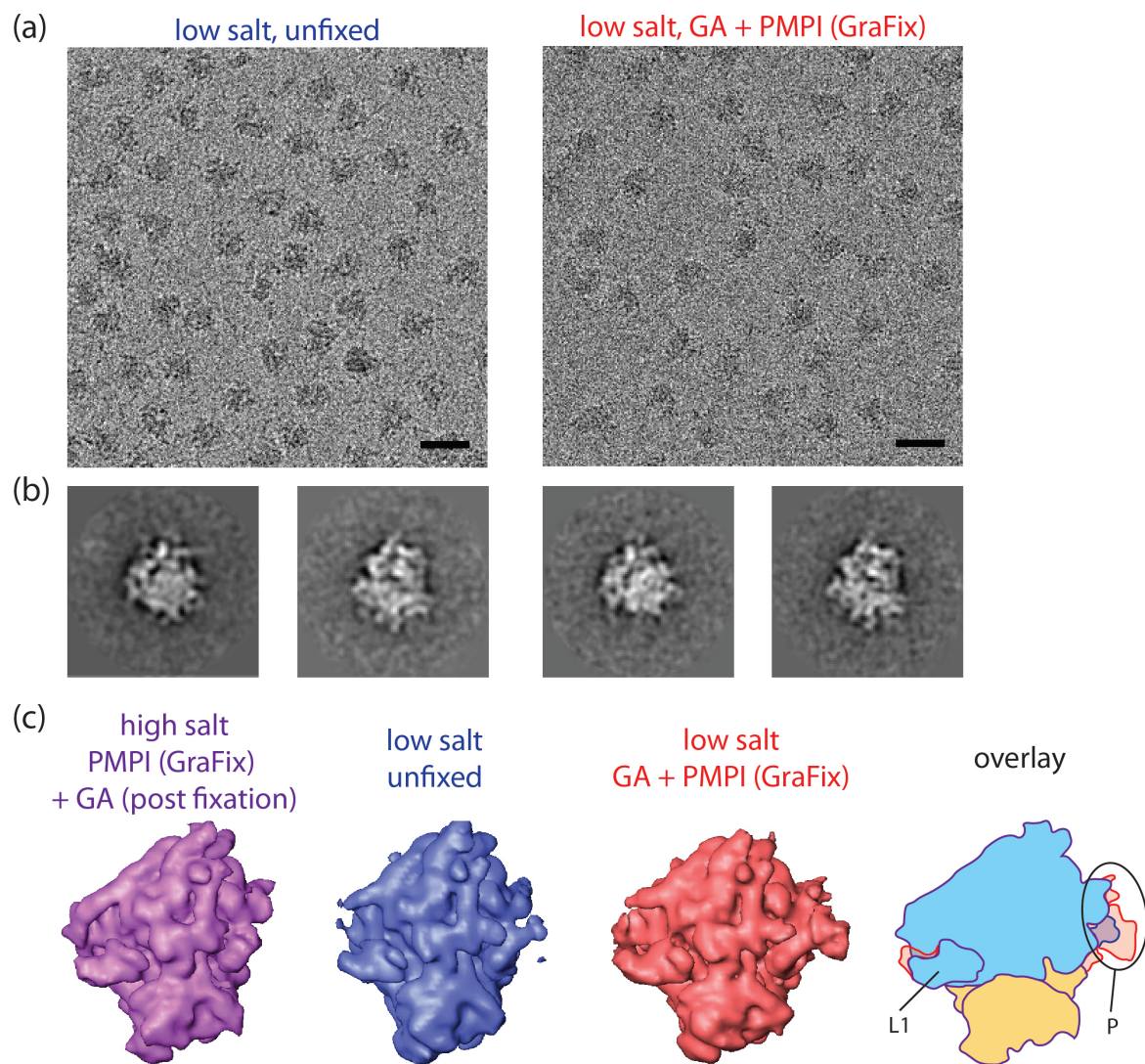


Figure 3.9: Low salt preparation stabilizes the structure of HeLa 80S ribosome. (a) Raw micrographs and (b) class averages of ribosomes prepared under low salt condition. (c) 3D models of low-salt prepared ribosomes are reconstructed and compared with the model of high-salt prepared ribosome. L1 stalk and P stalk are stabilized by low salt preparation and GraFix. The scale bar represents 30 nm.

biomolecules, those unwanted molecules can be efficiently removed by high concentration of salt, and this procedure is the so-called “salt wash”. However, some molecules associated with the ribosome can also be lost during these steps. Comparing to the ribosomes of prokaryotes or lower eukaryotes, the human ribosome is much less understood. Besides the long expansion segments, human ribosomes may also have more interactions with other molecules. Not all of the ribosome associated molecules are clarified, but also for the identified ones, some of them remain mysterious in function (Doudna and Rath, 2002; Dinman, 2009). Assuming that some of the ribosome associated molecules have an influence on structural stability of the ribosome, like some components of the spliceosomal assembly, preventing these molecules from dissociation might preserve a better-stabilized ribosome structure. Therefore, HeLa ribosomes here were purified under a milder condition, where the salt concentration was mostly maintained at 50 mM. Instead of removing other molecules by high salt concentration, three steps of gradient sedimentation with PEG precipitation were performed (figure 3.8). To further confirm the stabilization effect of GraFix with glutaraldehyde and PMPI, the purified ribosomes were either loaded onto a sucrose gradient with double fixatives (as figure 2.1(d)) or a plain sucrose gradient.

3D models of the “GraFixed” and the unfixed samples were reconstructed independently by single particle cryo-EM, and these two models were compared to the previously constructed model of salt-washed ribosome which was “GraFixed” with PMPI and post-fixated with glutaraldehyde (see section 3.3). The L1 stalk and P stalk were structurally better defined in the low salt preparation with GraFix (figure 3.9). In addition, density for eEF2, though not with full occupancy, can be observed in the GraFixed low salt model. Besides, an additional density next to the peptide exit tunnel was observed in both low salt models (figure 3.10). Different factors which bind to the peptide exit tunnel on the eukaryotic ribosomes have been reported, such as nascent chain-associated complex (NAC) (Pech et al., 2010) or Hsp70-based chaperone

system (also known as the ribosome-associated complex (RAC)) (Leidig et al., 2013). From the result of mass spectrometry analysis of the sample, besides the regular ribosomal proteins and the common ribosomal associating factor (e.g. eEF2), both RAC and NAC were identified in our low salt ribosomal preparations (appendix A.2).

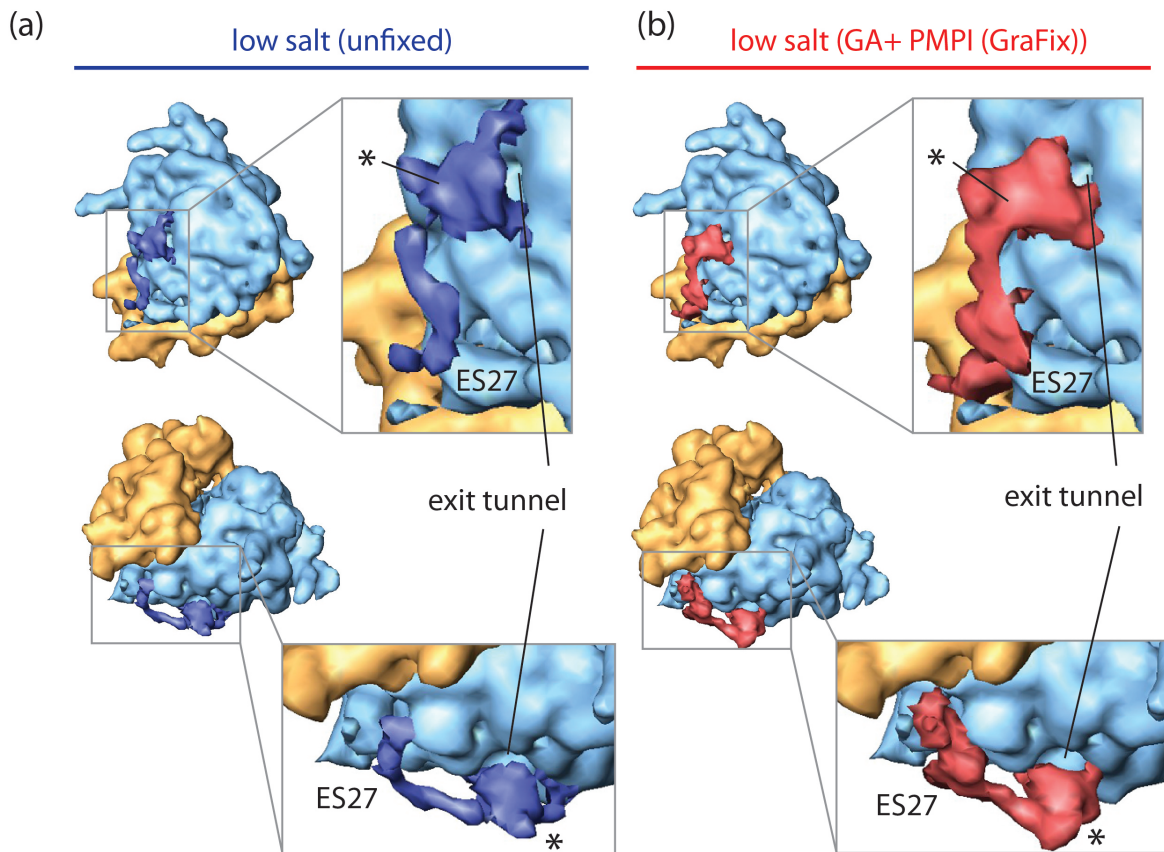


Figure 3.10: Nascent chain-associated complex (NAC) or ribosome-associated complex (RAC) was copurified under low salt condition (a) and was further stabilized by GraFix (b). Model of the ribosome purified under high salt condition was manually segmented into 60S (blue) and 40S (yellow), and the extra density, which represented NAC/RAC (labeled with *) and extension of ES27, was shown in indigo (low salt purification without fixation) or red (low salt purification with GraFix).

Chapter 4

Discussion

4.1 Structural Analysis of the Human Spliceosomal C Complex

4.1.1 Regions with high structural heterogeneity might represent the functional domains

Since the structure of human spliceosomal C complex was previously solved by RCT (Golas et al., 2010), it was aimed here to determine the structure of the C complex at higher resolution by analyzing and separating molecules in different conformational and compositional states. The structure of the C complex was reconstructed by angular reconstitution using the previously solved structure obtained by RCT as the anchor set. With a large amount (200,000) of recorded particles, different heterogeneous states were able to be resolved by resampling and 3D MSA classification. The heterogeneity mainly localized in three regions: “belly”, “head”, and “platform”. The “belly” showed compositional heterogeneity. Interestingly, most structures which lack of the belly region had in other regions less defined density, indicating that deprivation of components might destabilize the full assembly. There are two possibilities for the absence of components in the belly region. One possibility is that those components are not rigidly

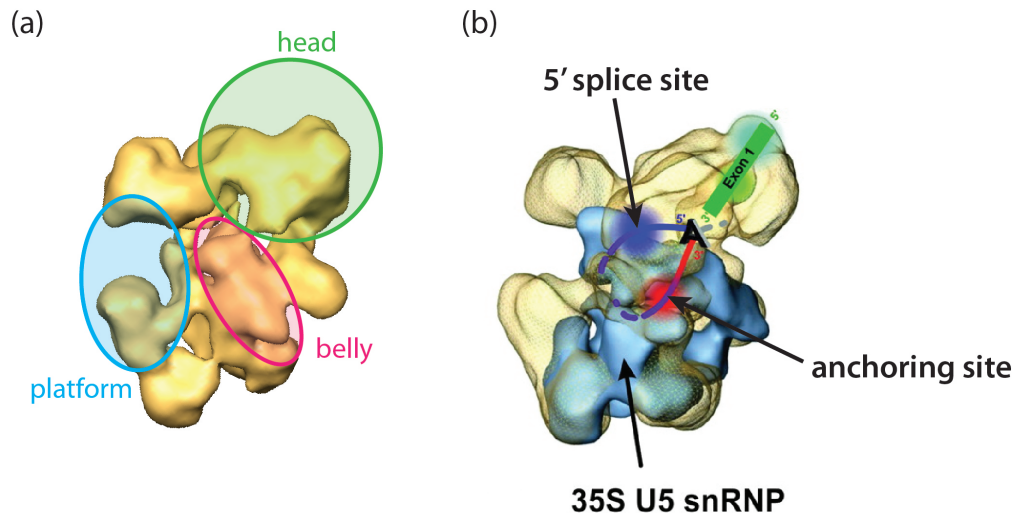


Figure 4.1: Comparison of heterogeneity domains and the localization of pre-mRNA in human spliceosome C complex. (a) Highly heterogeneous domains of the C complex which were analyzed in this work are labeled the circles. (b) Localization of pre-mRNA in the RCT reconstructed C complex. Adapted from (Wolf et al., 2012) with permission.

incorporated in the complex, therefore they can be easily removed during purification. The disruption during purification might also have less dramatic effect on the other regions, and this can explain the instability of the structure. The other possibility is that the belly region might be the active center of the C complex, hence assembly and disassembly occur frequently here. The states which lack of the belly density might be a transition state, with global compositional alterations.

Indeed, these highly heterogeneous regions might correspond to the critical catalytic domains. Recently, antibody labeling on human spliceosomal C complex was performed by Wolf et al. (2012). The possible anchoring site was shown to locate in the belly region (figure 4.1). The anchoring site is located upstream of the branch site, and serves as the anchor for U2 snRNP for assembly (Gozani et al., 1996). Hence, it is very probable that intensive compositional alterations occur in this region. In addi-

tion, the “head” is the region with the highest heterogeneity level, where the 5’ exon is located. The composition and location of the 5’ exon binding components can be very dynamic, so that the 5’ exon can be positioned for ligating to the 3’ exon in the second step of the splicing reaction. Last but not least, the upper part of the “platform” also displayed high heterogeneity in contrast to the relatively stable lower part of the “platform”, which belongs to the 35S U5 snRNP. The heterogeneity in the upper part of the “platform” might indicate structural rearrangements which might be involved in the regulation of the catalytic activity of the core.

4.1.2 Outlook

Due to the unexpected high heterogeneity level, the structural analysis of human spliceosomal C complex was not able to be performed at higher resolution. Resolving all the heterogeneous states *in silico* requires a much larger dataset and new image processing tool that are currently not available. With current hardware and software, image recording and analysis of such a large dataset might be very inefficient. Therefore, it is important to reduce the heterogeneity of the sample.

The heterogeneity of the C complex sample might be a combination of its intrinsic dynamic nature and extrinsic disruption of the molecules. One force causing disruption might be the salt concentration, and this could be proven by previous publication from Golas et al. (2010) showing deprivation of C complex components from high salt treatment. The other disruption force might come from the surface interaction of the affinity column. To minimize the disruption introduced during purification, it is essential to optimize the preparation procedure. Unfortunately, the current purification procedure of the spliceosomal C complex requires high specificity, and no alternative strategy has been developed so far. Besides the stabilization during purification, post-purification stabilization is equally important. GraFix has been performed in our lab to enhance the intramolecular stability (Kastner et al., 2008). Moreover, *p*-maleimidophenyl iso-

cyanate (PMPI) was shown to further stabilize the ribosomes as an additional fixative for GraFix. In fact, same PMPI-based GraFix was also performed on the C complex; however, no significant improvement was observed. It is possible that the heterogeneity of the C complex sample was too high, so that the stabilization effect of PMPI was too subtle to be noticed.

4.2 Purification and Structure of Yeast Endogenous snRNP

4.2.1 Sucrose might be a better crowding agent for snRNP purification

A living cell is full of biomolecules such as proteins or RNP complexes. Therefore, the biomolecules function natively in a crowded environment. To preserve the complexes in their native states during purification, the crowded environment is best simulated with addition of crowding agents. Since the distribution of molecules in the cell is not homogeneous, different molecules might have different behavior in different crowding agents. For spliceosome or snRNP purification, glycerol has been often used as the crowding agent and for gradient sedimentation in our department (Bach, 1989), whereas sucrose has been used instead in some of the other labs (Schrier et al., 1985). In this work, glycerol, sucrose, and glucose were tested as crowding agents for HeLa snRNP purification, and the snRNP molecules were recovered in the sucrose solution with the highest amount (figure 3.4(a)). The environment changing during the cell lysis might be inevitable, but the sucrose could better prevent the snRNP from degradation than glycerol. This finding can be helpful for future purification improvement of snRNP or spliceosomes.

4.2.2 Purification without affinity-based procedure reduces sample degradation

The assembly of a snRNP or spliceosome complex might be based on surface interactions of every component. Therefore, disruption on a single component might cause global destabilization of the complex. The philosophy of the new purification strategy performed in this work is to minimize the disruption force. Besides the optimization of the crowding agent, the high specificity affinity-based method was replaced by a series of PEG precipitation and gradient sedimentation. The new purification procedure has led to a better yield of snRNP compared to the conventional affinity-based purification method, suggesting snRNP particles might be degraded during the affinity purification. During affinity purification, a new surface interaction is created between the antibody and the epitope component. If the component is not tightly bound in the complex, the component can be easily destabilized.

With the new purification strategy, not only the snRNP yield was increased, but also the distribution of particle orientation was improved in single particle EM analysis. The yeast tri-snRNP purified with the conventional method has displayed a highly preferred orientation (Häcker et al., 2008). Though preference in the angular distribution facilitates antibody labelling in the 2D level, lacking different orientations makes 3D reconstruction very difficult. The broader angular distribution of the particles purified with the new strategy has made the 3D reconstruction of yeast snRNP first time possible. The orientation of particles is affected by the interaction between the carbon film and the surface charge distribution of particles. One of the observed views of the particles purified with the new strategy (figure 3.4(c), second row) showed structural similarity to the preferred view of particles from the old preparation (figure 3.4(d)), indicating the overall structure was almost unchanged. Therefore, the improvement in the angular distribution might be due to minor changes on the surface of the particles.

According to the mass spectrometric analysis, the composition of the newly purified snRNP is undistinguishable from the conventionally purified yeast U4/U6·U5 tri-snRNP, so the minor structural difference might come from different intramolecular interactions of the components.

4.3 Crosslink RNP with PMPI: HeLa 80S ribosomes as an example

4.3.1 GraFix of 80S ribosome with PMPI

In this work, *p*-maleimidophenyl isocyanate (PMPI) was innovatively used on ribosome as an RNA-protein crosslinker for single particle EM analysis. Theoretically, PMPI was reported to react rapidly with water and lose the activity for hydroxyl crosslinking (Annunziato et al., 1993). Besides, PMPI might crosslink proteins to sucrose in the gradient solution instead of RNA. Nevertheless, the crosslinking effect of PMPI on the 80S ribosome was observed here. The expansion segment ES7L was tightly linked to the protein L7A, indicating potential RNA-protein crosslink (figure 3.5(c)). In fact, under very low density threshold, the ES7L-L7A linkage could be visualized in the other two models of samples without PMPI fixation, as well as the high-resolution model solved by Anger et al. (2013), suggesting that ES7L might natively have a weak interaction with L7A. This interaction was probably fortified by crosslink with PMPI, and was strong enough to sustain the force introduced by the carbon film adsorption, therefore created an extra landing surface for the particle (figure 3.6). Unfortunately, the ES7L in the high-resolution model (Anger et al., 2013) was not stable enough to provide a reliable atomic model simulation, therefore the exact crosslink could not be clarified.

The L1 stalk is one of the most flexible domains of the ribosome and is usually less

well resolved compared to the global ribosome structure. In the PMPI-fixed sample here, more structural details of the L1 stalk were visualized, suggesting a significant stabilization effect of PMPI. Besides, the L1 stalk position in this model was slightly different from the other models. It was generally believed that the L1 stalk could function as a lateral gate for the E site tRNA (Spahn et al., 2004; Chandramouli et al., 2008). Coherently, the E site tRNA here also showed higher occupancy. With the current model, the resolution was not high enough to tell if the stabilization of E site tRNA came from direct crosslinking or indirect stabilization by the L1 stalk.

4.3.2 The 80S ribosome might be stabilized by polyamines

Many researches in prokaryotic systems have shown that polyamines are important for ribosome stabilization. For example, the polyamines might inhibit the subunits from dissociation (Umekage and Ueda, 2006). The crystallization condition for *E. coli* ribosome also contains polyamines (Schuwirth et al., 2005). Polyamines, which are positively charged, might stabilize ribosomes by neutralizing the negative charge of the RNA. A research from Watanabe et al. (1991) also indicated that polyamines could form complexes with RNA. In contrast, much less has been known about the effect of polyamines on eukaryotic ribosome structures. Our experiments here showed that the polyamine-treated ribosomes seemed to have better-stabilized RNA expansion segments (figure 3.5(c)). This result can be a proof of the electrostatic stabilization.

4.4 Structure of HeLa 80S ribosome purified under low salt condition

4.4.1 Purification of ribosomes under low salt conditions stabilizes the flexible stalks

In an RNP complex, electrostatic interaction between components is probably the most important force to stabilize the full assembly. Therefore, high concentration of salt can cause destabilization of the complex through disrupting the intramolecular electrostatic force. In this work, HeLa 80S ribosomes were purified in a low salt concentration environment, and the reduction of structural instability was demonstrated by the single particle cryo-EM analysis. Even in the absence of any fixatives, the flexible L1 and P stalks were stabilized not worse than those of the “PMPI-GraFixed” ribosome purified with high salt (figure 3.9). Though the L1 and P stalks have been considered to be the most flexible domains of the ribosome, part of the observed flexibility might come from destabilization during the purification. Hence, avoiding electrostatic disruption during the purification might preserve the native stabilization force of the stalks. Furthermore, GraFix with both glutaraldehyde and PMPI has provided extra structural stabilization of the P stalk so that the P stalk could be completely visualized at the same threshold level as the ribosome core.

4.4.2 Tunnel exit bound factor was preserved on the ribosome with low salt preparation procedure

Conventional purification procedures of the ribosome involves the so called “salt wash” step, and this is based on the understanding that the ribosome is relatively salt-stable. This applies to the ribosome core; however, some of the periphery components can be salt sensitive, like the L1 and P stalks as mentioned above. Moreover, the eukaryotic ribosome is a very delicately regulated machinery, and regulation factors might interact

with the ribosome through direct binding. Because surface interaction of the factors and the ribosome is mostly based on electrostatic force, salt wash can potentially cause dissociation of the factors. In the salt-wash free purified ribosomes, an extra density was found next to the tunnel exit of the nascent peptide chain, between the ribosomal protein L19 and L23A, but the exact binding position was not able to be identified. Mass spectrometry analysis indicated this factor can be the nascent chain-associated complex (NAC) or the ribosome-associated complex (RAC). Since these two factors have similar binding location on the ribosome, the observed additional density can also be an averaged structure of both factors.

The tunnel exit bound factor also displayed a connection with the extension of the RNA expansion segment ES27. In the reconstructed structure of yeast ribosome, the terminal ES27 could be visualized in two positions: near the L1 arm or close to the tunnel exit (Beckmann et al., 2001). In addition, *in vitro* assembled *Chaetomium thermophilum* ribosome-RAC complex also showed a possible interaction between ES27 and RAC (Leidig et al., 2013). It is very likely that this tunnel exit bound factor on the human ribosome had direct interaction with ES27 *in vivo*. Since electrostatic force is critical for non-covalent RNA-protein binding, association of ES27 and this factor can easily be broken during the conventional salt wash procedure, and further results in the dissociation of this factor from the ribosome.

Chapter 5

Conclusions and Perspectives

Single particle cryo-EM is so far an ideally suited method to study structures of large dynamic RNP complexes. Sub-states representing the sample heterogeneity can theoretically be resolved *in silico*. However, *in silico* sorting reaches its limit when the heterogeneity level is too high. Therefore, it was aimed to stabilize the sample at the biochemical level. Two approaches were performed in this work: (1) to stabilize the sample during purification, and (2) to prevent macromolecular complexes from disruption during the sample preparation for transmission electron microscopy (TEM). Both methods have shown improvements on the sample quality.

Avoiding purification methods involving strong surface interaction (*e.g.* chromatography) or electrostatic disruption (*e.g.* high salt concentration) keeps the integrity of the particles and further stabilizes the full assembly of the complexes. The RNA-protein crosslinker, *p*-maleimidophenyl isocyanate (PMPI), further strengthens the intramolecular interaction of the particles. In the dataset of the ribosomes purified under low salt condition and crosslinked with PMPI and glutaraldehyde, *in silico* sorting has been performed efficiently so far. Two ratcheting states could be clearly observed, coupled with the occupancy of eEF2 and E-site tRNA. Following these *in silico* sorting steps, calculation of high-resolution ribosome structures is in progress.

These biochemical methodologies do not only facilitate structural studies of RNP complexes by cryo-EM, but might also benefit high-resolution structural studies by X-ray crystallography. Due to the dynamic nature, few of full RNP complexes have been solved by crystallography so far. With the structural stabilization, crystallization of large RNP complexes can be more feasible in the future.

Appendix A

Appendix

A.1 Mass spectrometry analysis of yeast snRNP sample

Table A.1:

Protein name	GenBank accession no.	MW (kDa)	Number of peptides sequenced	Complex
Prp3	gi—6320681	56 kDa	24	U4
Prp31	gi—51830339	56 kDa	29	U4
Prp4	gi—207340214 (+1)	52 kDa	32	U4
Brr2	gi—6321020	246 kDa	125	U5
Prp6	gi—151946445 (+1)	104 kDa	40	U5
Prp8	gi—151944111 [4]	280 kDa	110	U5
Snu114	gi—151941746 (+1)	114 kDa	63	U5
Prp38	gi—256269086 (+1)	28 kDa	8	U4/U6·U5 tri-snRNP
Snu23	gi—6320105	23 kDa	12	U4/U6·U5 tri-snRNP
Sm B	gi—256271132 (+1)	22 kDa	13	U1, U2, U4, U5

Continued on next page

Table A.1 – *Continued from previous page*

Protein name	GenBank accession no.	MW (kDa)	Number of peptides sequenced	Complex
SmD1	gi—6321510	16 kDa	4	U1, U2, U4, U5

A.2 Mass spectrometry analysis of low salt purified HeLa 80S ribosome sample

Table A.2:

Protein name	GenBank accession no.	MW (kDa)	Number of peptides sequenced
--------------	-----------------------	----------	------------------------------

40S proteins

S2	gi—15055539	31 kDa	231
S3a	gi—4506723 (+1)	30 kDa	219
S4	gi—119592221 (+1)	43 kDa	256
S5	gi—119592989 (+2)	22 kDa	68
S6	gi—337514	29 kDa	135
S7	gi—4506741	22 kDa	95
S8	gi—119627428 (+1)	27 kDa	204
S9	gi—14141193	23 kDa	213
S10	gi—4506679	19 kDa	89
S11	gi—4506681	18 kDa	107
S12	gi—14277700	15 kDa	52
S13	gi—4506685	17 kDa	89

Continued on next page

Table A.2 – *Continued from previous page*

Protein name	GenBank accession no.	MW (kDa)	Number of peptides sequenced
S14	gi—5032051	16 kDa	101
S15	gi—119589905 (+2)	14 kDa	10
S15A	gi—119570641 (+2)	29 kDa	142
S16	gi—119577297 (+1)	16 kDa	185
S17	gi—4506693	16 kDa	79
S19	gi—4506695	16 kDa	58
S20	gi—4506697	13 kDa	59
S21	gi—4506699	9 kDa	39
S23	gi—4506701	16 kDa	76
S24	gi—119575003 (+8)	15 kDa	44
S25	gi—119569329 (+1)	14 kDa	73
S27	gi—119611832 (+1)	10 kDa	96
S27-like	gi—13277528 (+1)	9 kDa	61
S28	gi—4506715	8 kDa	29
S29	gi—4506717	7 kDa	19

60S proteins

P1	gi—49457412	12 kDa	29
P2	gi—4506671	12 kDa	87
L3	gi—119580717	49 kDa	357
L4	gi—62087534	49 kDa	392
L5	gi—14591909 (+1)	34 kDa	318
L6	gi—16753227 (+2)	33 kDa	292

Continued on next page

Table A.2 – *Continued from previous page*

Protein name	GenBank accession no.	MW (kDa)	Number of peptides sequenced
L7	gi—119582155 (+2)	29 kDa	240
L7a	gi—119608470	35 kDa	340
L8	gi—4506663	28 kDa	218
L9	gi—15431303	22 kDa	244
L10	gi—119593144 (+4)	25 kDa	110
L11	gi—59896138	20 kDa	82
L12	gi—4506597	18 kDa	183
L13	gi—15431295 (+1)	24 kDa	127
L14	gi—12653649	24 kDa	166
L18	gi—4506607	22 kDa	120
L18a	gi—11415026	21 kDa	150
L19	gi—119580972 (+3)	26 kDa	69
L21	gi—18104948 (+1)	19 kDa	82
L22	gi—119583342	32 kDa	124
L22e	gi—33150766 (+1)	15 kDa	86
L23a	gi—119571516 (+3)	22 kDa	118
L24	gi—119600188 (+3)	13 kDa	136
L26	gi—292435 (+2)	17 kDa	83
L27	gi—4506623	16 kDa	103
L27a	gi—119568094 (+3)	17 kDa	88
L28	gi—13904866 (+1)	16 kDa	38
L29	gi—119585586 (+4)	18 kDa	34
L30	gi—4506631	13 kDa	139

Continued on next page

Table A.2 – *Continued from previous page*

Protein name	GenBank accession no.	MW (kDa)	Number of peptides sequenced
L32	gi—4506635	16 kDa	116
L32	gi—16117787	13 kDa	37
L35	gi—119608008 (+2)	22 kDa	62
L35a	gi—119612656 (+1)	16 kDa	26
L36	gi—16117794 (+1)	12 kDa	43
L36a	gi—119623264 (+1)	14 kDa	33
L38	gi—4506645	8 kDa	149

Translation factors

eIF1A	gi—33356163 (+2)	16 kDa	3
eIF2 subunit 1	gi—4758256	36 kDa	4
eIF3 subunit 6	gi—62896617	67 kDa	24
eIF3 subunit A	gi—4503509 (+1)	167 kDa	100
eIF3 subunit I	gi—4503513	37 kDa	12
eIF3 subunit K	gi—10801345 (+2)	25 kDa	6
eIF3 subunit M	gi—23397429 (+1)	43 kDa	6
eIF3 subunit 7 zeta	gi—119580514 (+3)	64 kDa	16
eIF3 p35 subunit	gi—3264861 (+1)	29 kDa	7
eIF3 p44 subunit	gi—3264859 (+2)	36 kDa	2
eIF beta	gi—182067 (+2)	38 kDa	5
eEF2	gi—4503483	95 kDa	553
eRF1/eRF3	gi—237823782	51 kDa	71

Continued on next page

Table A.2 – *Continued from previous page*

Protein name	GenBank accession no.	MW (kDa)	Number of peptides sequenced
<i>Tunnel exit binding factors</i>			
heat shock 70 kDa protein 1A/1B	gi—167466173 (+2)	70 kDa	25
heat shock cognate 71 kDa protein isoform 1	gi—5729877	71 kDa	38
nascent polypeptide-associated complex subunit alpha isoform b	gi—5031931 (+1)	23 kDa	106

List of Symbols and Abbreviations

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
Å	Ångstrom ($1 \text{ Å} = 10^{-10} \text{ m}$)
B ^{act}	activated spliceosomal B-complex
B*	catalytically activated spliceosomal B-complex
CCC	cross correlation coefficient
CCD	charge coupled device
cryo-EM	cryogenic electron microscopy
C _s	spherical aberration constant
CTF	contrast transfer function
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
e ⁻	electron(s)
eEF1a	eukaryotic elongation factor 1A
eEF2	eukaryotic elongation factor 2
<i>E. coli</i>	<i>Escherichia coli</i>
EF-G	elongation factor G
EF-Tu	elongation factor Tu
<i>e.g.</i>	<i>exemplo gratia</i>

EM	electron microscopy
ES	expansion segment
FFT	fast Fourier transformation
FSC	Fourier shell correlation
GraFix	gradient fixation
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
kDa	kilo Daltons
kV	kilo Volts
MBP	maltose binding protein
MRA	multi-reference alignment
mRNA	messenger-RNA
MSA	multivariate statistical analysis
NAC	nascent chain-associated complex
NMR	nuclear magnetic resonance
OD	optical density
PCA	principal component analysis
PDB	Protein Data Bank
PEG	polyethylene glycol
PhCTF	phase-contrast transfer function
PMPI	<i>p</i> -maleimidophenyl isocyanate
pre-mRNA	precursor messenger-RNA
RAC	ribosome-associated complex
RCT	random conical tilt
RNA	ribonucleic acid
RNP	ribonucleoprotein
rpm	rotations per minute
rRNA	ribosomal RNA
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis

SF	splicing factor
SNR	signal-to-noise ratio
snRNA	small nuclear ribonucleic acid
snRNP	small nuclear ribonucleoproteins
TEM	transmission electron microscopy
tri-snRNP	triple small nuclear ribonucleoprotein
tRNA	transfer-RNA
U2AF	U2 auxiliary factor
v/v	volume per volume
w/v	weight per volume

Bibliography

- A. M. Anger, J.-P. Armache, O. Berninghausen, M. Habeck, M. Subklewe, D. N. Wilson, and R. Beckmann. Structures of the human and Drosophila 80S ribosome. *Nature*, 497(7447): 80–5, May 2013.
- M. E. Annunziato, U. S. Patel, M. Ranade, and P. S. Palumbo. p-Maleimidophenyl isocyanate: a novel heterobifunctional linker for hydroxyl to thiol coupling. *Bioconjugate Chemistry*, 4(3):212–218, May 1993.
- J.-P. Armache, A. Jarasch, A. M. Anger, E. Villa, T. Becker, S. Bhushan, F. Jossinet, M. Habeck, G. Dindar, S. Franckenberg, V. Marquez, T. Mielke, M. Thomm, O. Berninghausen, B. Beatrix, J. Söding, E. Westhof, D. N. Wilson, and R. Beckmann. Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5 Å resolution. *Proceedings of the National Academy of Sciences of the United States of America*, 107(46): 19748–53, Nov. 2010.
- D. H. Atha and K. C. Ingham. Mechanism of precipitation of proteins by polyethylene glycols. Analysis in terms of excluded volume. *The Journal of biological chemistry*, 256(23):12108–17, Dec. 1981.
- M. Azubel, S. G. Wolf, J. Sperling, and R. Sperling. Three-dimensional structure of the native spliceosome by cryo-electron microscopy. *Molecular cell*, 15(5):833–9, Sept. 2004.
- M. Bach. 20S Small Nuclear Ribonucleoprotein U5 Shows a Surprisingly Complex Protein Composition. *Proceedings of the National Academy of Sciences*, 86(16):6038–6042, Aug. 1989.
- N. Ban. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. *Science*, 289(5481):905–920, Aug. 2000.
- R. Beckmann, C. M. Spahn, N. Eswar, J. Helmers, P. A. Penczek, A. Sali, J. Frank, and G. Blobel. Architecture of the Protein-Conducting Channel Associated with the Translating 80S Ribosome. *Cell*, 107(3):361–372, Nov. 2001.

- S. E. Behrens and R. Lührmann. Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells. *Genes & Development*, 5(8):1439–1452, Aug. 1991.
- N. Behzadnia, M. M. Golas, K. Hartmuth, B. Sander, B. Kastner, J. Deckert, P. Dube, C. L. Will, H. Urlaub, H. Stark, and R. Lührmann. Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *The EMBO journal*, 26(6):1737–48, Mar. 2007.
- S. Belin, S. Hacot, L. Daudignon, G. Therizols, S. Pourpe, H. C. Mertani, M. Rosa-Calatrava, and J.-J. Diaz. Purification of ribosomes from human cell lines. *Current protocols in cell biology*, Chapter 3:Unit 3.40, Dec. 2010.
- A. Ben-Shem, N. Garreau de Loubresse, S. Melnikov, L. Jenner, G. Yusupova, and M. Yusupov. The structure of the eukaryotic ribosome at 3.0Å resolution. *Science (New York, N.Y.)*, 334(6062):1524–9, Dec. 2011.
- S. Bessonov, M. Anokhina, C. L. Will, H. Urlaub, and R. Lührmann. Isolation of an active step I spliceosome and composition of its RNP core. *Nature*, 452(7189):846–850, Apr. 2008.
- D. Boehringer, E. M. Makarov, B. Sander, O. V. Makarova, B. Kastner, R. Lührmann, and H. Stark. Three-dimensional structure of a pre-catalytic human spliceosomal complex B. *Nature structural & molecular biology*, 11(5):463–8, May 2004.
- P. Chandramouli, M. Topf, J.-F. Mntret, N. Eswar, J. J. Cannone, R. R. Gutell, A. Sali, and C. W. Akey. Structure of the Mammalian 80S Ribosome at 8.7 Å Resolution. *Structure*, 16(4):535–548, Apr. 2008.
- R. A. Crowther, D. J. DeRosier, and A. Klug. The Reconstruction of a Three-Dimensional Structure from Projections and its Application to Electron Microscopy. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 317(1530):319–340, June 1970.
- R. Danev and K. Nagayama. Transmission electron microscopy with Zernike phase plate. *Ultramicroscopy*, 88(4):243–252, Sept. 2001.
- R. Danev and K. Nagayama. Complex Observation in Electron Microscopy: IV. Reconstruction of Complex Object Wave from Conventional and Half Plane Phase Plate Image Pair. *Journal of the Physics Society Japan*, 73(10):2718–2724, Oct. 2004.
- R. Danev, H. Okawara, N. Usuda, K. Kametani, and K. Nagayama. A Novel Phase-contrast Transmission Electron Microscopy Producing High-contrast Topographic Images of Weak objects. *Journal of biological physics*, 28(4):627–35, Dec. 2002.

- J. D. Dinman. The eukaryotic ribosome: current status and challenges. *The Journal of biological chemistry*, 284(18):11761–5, May 2009.
- J. A. Doudna and V. L. Rath. Structure and Function of the Eukaryotic Ribosome: The Next Frontier. *Cell*, 109(2):153–156, Apr. 2002.
- P. Dube, P. Tavares, R. Lurz, and M. van Heel. The portal protein of bacteriophage SPP1 a DNA pump with 13-fold symmetry. *The EMBO Journal*, 12(4):1303–1309, Apr. 1993.
- P. Dube, G. Bacher, H. Stark, F. Mueller, F. Zemlin, M. van Heel, and R. Brimacombe. Correlation of the expansion segments in mammalian rRNA with the fine structure of the 80S ribosome; a cryoelectron microscopic reconstruction of the rabbit reticulocyte ribosome at 21 Å resolution. *Journal of Molecular Biology*, 279(2):403–421, June 1998.
- P. Fabrizio, J. Dannenberg, P. Dube, B. Kastner, H. Stark, H. Urlaub, and R. Lührmann. The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Molecular cell*, 36(4):593–608, Nov. 2009.
- N. Fischer, A. L. Konevega, W. Wintermeyer, M. V. Rodnina, and H. Stark. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature*, 466(7304):329–333, July 2010.
- M. Fox, G. Gray, K. Kavanagh, C. Lewis, and S. Doyle. Detection of *Aspergillus fumigatus* mycotoxins: immunogen synthesis and immunoassay development. *Journal of Microbiological Methods*, 56(2):221–230, Feb. 2004.
- H. Gao, M. Valle, M. n. Ehrenberg, and J. Frank. Dynamics of EF-G interaction with the ribosome explored by classification of a heterogeneous cryo-EM dataset. *Journal of Structural Biology*, 147(3):283–290, Sept. 2004.
- S. A. Gerbi. Expansion segments: Regions of variable size that interrupt the universal core secondary structure of ribosomal RNA. *Ribosomal RNA: Structure, evolution, processing, and function in protein biosynthesis*, pages 71–87, 1996.
- M. M. Golas, B. Sander, C. L. Will, R. Lührmann, and H. Stark. Molecular architecture of the multiprotein splicing factor SF3b. *Science (New York, N.Y.)*, 300(5621):980–4, May 2003.
- M. M. Golas, B. Sander, C. L. Will, R. Lührmann, and H. Stark. Major conformational change in the complex SF3b upon integration into the spliceosomal U11/U12 di-snRNP as revealed by electron cryomicroscopy. *Molecular cell*, 17(6):869–83, Mar. 2005.

- M. M. Golas, B. Sander, S. Bessonov, M. Grote, E. Wolf, B. Kastner, H. Stark, and R. Lührmann. 3D cryo-EM structure of an active step I spliceosome and localization of its catalytic core. *Molecular cell*, 40(6):927–38, Dec. 2010.
- O. Gozani, R. Feld, and R. Reed. Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes & Development*, 10(2):233–243, Jan. 1996.
- N. Grigorieff. Three-dimensional structure of bovine NADH:ubiquinone oxidoreductase (complex I) at 2.2 Å in ice. *Journal of molecular biology*, 277(5):1033–46, Apr. 1998.
- N. Grigorieff. FREALIGN: high-resolution refinement of single particle structures. *Journal of structural biology*, 157(1):117–25, Jan. 2007.
- A. Habeeb and R. Hiramoto. Reaction of proteins with glutaraldehyde. *Archives of Biochemistry and Biophysics*, 126(1):16–26, July 1968.
- I. Häcker, B. Sander, M. M. Golas, E. Wolf, E. Karagöz, B. Kastner, H. Stark, P. Fabrizio, and R. Lührmann. Localization of Prp8, Brr2, Snu114 and U4/U6 proteins in the yeast tri-snRNP by electron microscopy. *Nature structural & molecular biology*, 15(11):1206–12, Nov. 2008.
- G. Harauz and M. van Heel. Exact filters for general geometry three dimensional reconstruction. *OPTIK.*, 73(4):146–156, 1986.
- R. Henderson, A. Sali, M. L. Baker, B. Carragher, B. Devkota, K. H. Downing, E. H. Egelman, Z. Feng, J. Frank, N. Grigorieff, W. Jiang, S. J. Ludtke, O. Medalia, P. A. Penczek, P. B. Rosenthal, M. G. Rossmann, M. F. Schmid, G. F. Schröder, A. C. Steven, D. L. Stokes, J. D. Westbrook, W. Wriggers, H. Yang, J. Young, H. M. Berman, W. Chiu, G. J. Kleywegt, and C. L. Lawson. Outcome of the first electron microscopy validation task force meeting. *Structure (London, England : 1993)*, 20(2):205–14, Feb. 2012.
- J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing In Science & Engineering*, 9(3):90–95, 2007.
- L. Jenner, S. Melnikov, N. Garreau de Loubresse, A. Ben-Shem, M. Iskakova, A. Urzhumtsev, A. Meskauskas, J. Dinman, G. Yusupova, and M. Yusupov. Crystal structure of the 80S yeast ribosome. *Current opinion in structural biology*, 22(6):759–67, Dec. 2012.
- E. b. G. J. Jensen. *Cryo-EM Part A: Sample Preparation and Data Collection: Sample Preparation and Data Collection, Teil 1*. 2010. ISBN 0080956955.

- B. Kastner, N. Fischer, M. M. Golas, B. Sander, P. Dube, D. Boehringer, K. Hartmuth, J. Deckert, F. Hauer, E. Wolf, H. Uchtenhagen, H. Urlaub, F. Herzog, J. M. Peters, D. Poerschke, R. Lührmann, and H. Stark. GraFix sample preparation for single-particle electron cryomicroscopy. *Nat Meth*, 5(1):53–55, 2008.
- S. Kim and R. Lin. Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing. *Mol. Cell. Biol.*, 16(12):6810–6819, Dec. 1996.
- C. Leidig, G. Bange, J. Kopp, S. Amlacher, A. Aravind, S. Wickles, G. Witte, E. Hurt, R. Beckmann, and I. Sinning. Structural characterization of a eukaryotic chaperone–the ribosome-associated complex. *Nature structural & molecular biology*, 20(1):23–8, Jan. 2013.
- A. E. Leschziner and E. Nogales. Visualizing flexibility at molecular resolution: analysis of heterogeneity in single-particle electron microscopy reconstructions. *Annual review of biophysics and biomolecular structure*, 36:43–62, Jan. 2007.
- R. E. Lovrien and D. Matulis. Selective precipitation of proteins. *Current protocols in protein science*, Chapter 4:Unit 4.5, May 2001.
- A. A. Patel and J. A. Steitz. Splicing double: insights from the second spliceosome. *Nature reviews. Molecular cell biology*, 4(12):960–70, Dec. 2003.
- M. Pech, T. Spreter, R. Beckmann, and B. Beatrix. Dual binding mode of the nascent polypeptide-associated complex reveals a novel universal adapter site on the ribosome. *The Journal of biological chemistry*, 285(25):19679–87, June 2010.
- P. A. Penczek, R. A. Grassucci, and J. Frank. The ribosome at improved resolution: New techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy*, 53(3):251–270, Mar. 1994.
- E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004.
- D. A. Pomeranz Krummel, C. Oubridge, A. K. W. Leung, J. Li, and K. Nagai. Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature*, 458(7237):475–80, Mar. 2009.
- P. B. Rosenthal and R. Henderson. Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. *Journal of Molecular Biology*, 333(4):721–745, Oct. 2003.

- B. Sander, M. M. Golas, and H. Stark. Automatic CTF correction for single particles based upon multivariate statistical analysis of individual power spectra. *Journal of Structural Biology*, 142(3):392–401, June 2003.
- B. Sander, M. M. Golas, E. M. Makarov, H. Brahm, B. Kastner, R. Lührmann, and H. Stark. Organization of core spliceosomal components U5 snRNA loop I and U4/U6 Di-snRNP within U4/U6.U5 Tri-snRNP as revealed by electron cryomicroscopy. *Molecular cell*, 24(2):267–78, Oct. 2006.
- S. H. W. Scheres and S. Chen. Prevention of overfitting in cryo-EM structure determination. *Nature methods*, 9(9):853–4, Sept. 2012.
- O. Scherzer. The theoretical resolution limit of the electron microscope. *J. Appl. Phys.*, 20:20, 1949.
- F. Schluenzen, A. Tocilj, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon, F. Franceschi, and A. Yonath. Structure of Functionally Activated Small Ribosomal Subunit at 3.3 Å Resolution. *Cell*, 102(5):615–623, Sept. 2000.
- W. H. Schrier, R. Feinbaum, and T. B. Okarma. Isofocusing of antigenic small nuclear ribonucleoproteins. *Analytical Biochemistry*, 147(1):38–48, May 1985.
- B. S. Schuwirth, M. A. Borovinskaya, C. W. Hau, W. Zhang, A. Vila-Sanjurjo, J. M. Holton, and J. H. D. Cate. Structures of the bacterial ribosome at 3.5 Å resolution. *Science (New York, N.Y.)*, 310(5749):827–34, Nov. 2005.
- C. M. T. Spahn, M. G. Gomez-Lorenzo, R. A. Grassucci, R. Jorgensen, G. R. Andersen, R. Beckmann, P. A. Penczek, J. P. G. Ballesta, and J. Frank. Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. *{EMBO} J*, 23(5):1008–1019, Mar. 2004.
- H. Stark, P. Dube, R. Lührmann, and B. Kastner. Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature*, 409(6819):539–42, Jan. 2001.
- S. W. Stevens, D. E. Ryan, H. Y. Ge, R. E. Moore, M. K. Young, T. D. Lee, and J. Abelson. Composition and Functional Characterization of the Yeast Spliceosomal Penta-snRNP. *Molecular Cell*, 9(1):31–44, Jan. 2002.
- C. R. Stocking. Precipitation of Enzymes during Isolation of Chloroplasts in Carbowax. *Science*, 123(3206):1032–1033, June 1956.
- S. Umekage and T. Ueda. Spermidine inhibits transient and stable ribosome subunit dissociation. *FEBS Letters*, 580(5):1222–1226, Feb. 2006.

- M. van Heel. Multivariate statistical classification of noisy images (randomly oriented biological macromolecules). *Ultramicroscopy*, 13(1-2):165–183, 1984.
- M. van Heel. Angular reconstitution: A posteriori assignment of projection directions for {3D} reconstruction. *Ultramicroscopy*, 21(2):111–123, 1987.
- M. van Heel, B. Gowen, R. Matadeen, E. V. Orlova, R. Finn, T. Pape, D. Cohen, H. Stark, R. Schmidt, M. Schatz, and a. Patwardhan. Single-particle electron cryo-microscopy: towards atomic resolution. *Quarterly reviews of biophysics*, 33(4):307–69, Nov. 2000.
- S.-H. Wang, X.-Y. Du, Y.-M. Huang, D.-S. Lin, P. L. Hart, and Z.-H. Wang. Detection of deoxyvalenol based on a single-chain fragment variable of the antideoxyvalenol antibody. *FEMS microbiology letters*, 272(2):214–9, July 2007.
- Z. Warkocki, P. Odenwalder, J. Schmitzova, F. Platzmann, H. Stark, H. Urlaub, R. Ficner, P. Fabrizio, and R. Luhrmann. Reconstitution of both steps of *Saccharomyces cerevisiae* splicing with purified spliceosomal components. *Nature structural & molecular biology*, 16(12):1237–43, Dec. 2009.
- S. Watanabe, K. Kusama-Eguchi, H. Kobayashi, and K. Igarashi. Estimation of polyamine binding to macromolecules and ATP in bovine lymphocytes and rat liver. *J. Biol. Chem.*, 266(31):20803–20809, Nov. 1991.
- G. Weber, S. Trowitzsch, B. Kastner, R. Luhrmann, and M. C. Wahl. Functional organization of the Sm core in the crystal structure of human U1 snRNP. *The EMBO journal*, 29(24):4172–84, Dec. 2010.
- C. L. Will and R. Luhrmann. Spliceosome structure and function. *Cold Spring Harbor perspectives in biology*, 3(7), July 2011.
- D. N. Wilson and J. H. Doudna Cate. The structure and function of the eukaryotic ribosome. *Cold Spring Harbor perspectives in biology*, 4(5):a011536–, May 2012.
- B. T. Wimberly, D. E. Brodersen, W. M. Clemons, R. J. Morgan-Warren, A. P. Carter, C. Vornrhein, T. Hartsch, and V. Ramakrishnan. Structure of the 30S ribosomal subunit. *Nature*, 407(6802):327–39, Sept. 2000.
- E. Wolf, B. Kastner, J. Deckert, C. Merz, H. Stark, and R. Luhrmann. Exon, intron and splice site locations in the spliceosomal B complex. *The EMBO journal*, 28(15):2283–92, Aug. 2009.
- E. Wolf, B. Kastner, and R. Luhrmann. Antisense-targeted immuno-EM localization of the pre-mRNA path in the spliceosomal C complex. *RNA (New York, N. Y.)*, 18(7):1347–57, July 2012.

- T. Yokoyama and T. Suzuki. Ribosomal RNAs are tolerant toward genetic insertions: evolutionary origin of the expansion segments. *Nucleic acids research*, 36(11):3539–51, June 2008.
- E. Zeitler. *Cryomicroscopy and radiation damage : International Study Group for Cryo Electron Microscopy Workshop, Schloss Ringberg am Tegernsee, 9-14 November 1981*. North-Holland Publ, Amsterdam, 1982.

Curriculum Vitae

Wen-ti Liu

born 6th June 1985 in Taipei, Taiwan

Phone +49 551 201-1373

E-Mail wliu@gwdg.de

Education

- 2010 - present Max Planck Institute for Biophysical Chemistry, Göttingen, Germany: Three-dimensional Electron Cryomicroscopy. PhD thesis: *Strategies to stabilize RNP complexes for structural determination by 3D cryo-electron microscopy*
- 2009 - 2010 Max Planck Institute for Biophysical Chemistry, Göttingen, Germany: Three-dimensional Electron Cryomicroscopy. Master thesis: *Structural analysis of IGHMBP2 bound human 80S ribosome by single-particle 3D cryo-electron microscopy*
- 2008 - 2009 Georg-August-University, Göttingen, Germany: International MSc/PhD program
- 2007 - 2008 Institute of Biomedical Science, Academia Sinica, Taipei, Taiwan. Research assistant.
- 2003 - 2007 National Taiwan University, Taiwan. Biochemical Science and Technology (BSc).
- 2000 - 2003 Taipei First Girls High School, Taipei, Taiwan.

Scholarships

- 2008 - 2009 Stipend of the Excellence Foundation for the Promotion of the Max Planck Society
- 2005 Fellowship for college students research project, National Science Council, Taiwan