Three-dimensional electron microscopy of structurally heterogeneous biological macromolecules

PhD Thesis

in partial fulfilment of the requirements
for the degree "Doctor of Philosophy (PhD)"
in the Molecular Biology Graduate Program
at the Georg August University Göttingen
Faculty of Biology

submitted by

Florian Hauer

born in

Karlsruhe, Germany

2009

Members of Thesis Committee

First Referee: Prof. Dr. Holger Stark

Second Referee: Prof. Dr. Ralf Ficner

Third Referee: Prof. Dr. Reinhard Jahn

Affidavit

I hereby declare in lieu of oath that this thesis has been written independently and with no other sources and aids than quoted.

Göttingen, 31.05.2009

Florian Hauer

List of Publications

1. Sirajuddin, M., Farkasovsky, M.. **Hauer, F**.. Kuhlmann, D., Macara, I. G., Weyand, M.. Stark, H., Wittinghofer, A.., *Structural insight into filament formation by mammalian septins.* Nature, 2007. **449**(7160): p. 311-5.
2. Kastner, B., Fischer, N., Golas, M. M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., **Hauer, F**., Wolf, E., Uchtenhagen, H., Urlaub, H., Herzog, F., Peters, J. M., Poerschke, D., Luhrmann, R., Stark, H., *GraFix: sample preparation for single-particle electron cryomicroscopy.* Nat Methods, 2008. **5**(1): p. 53-5.
3. Schmeisser, M., Heisen, B.C., Luettich, M., Busche, B., **Hauer, F**., Koske, T., Knauber, K.-H., Stark, H, *Parallel, distributed and GPU computing technologies in single-particle electron microscopy.* Acta Crystallogr D Biol Crystallogr, 2009. **65**(Pt 7): p. 659-71.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

•	symbol denoting interactions between biomolecules/biomacromolecular complexes
2D	Two-dimensional
3D	Three-dimensional
Å	Angstrom ($1Å = 1^{-10}$ m)
AMP-PNP	Adenylyl-imidodiphosphate
ATP	Adenosine triphosphate
CCC	Cross Correlation Coefficient
CCD	Charge Coupled Device
CPU	Central Processing Unit
cryo-EM	Cryogenic Electron Microscopy
CTD	C-terminal domain
DDM	Dodecyl-ß-d-maltoside
e$^-$	electron(s)
E. coli	Escherichia coli
EF-G	elongation factor G
EF-Tu	elongation factor Tu
EM	electron microscopy
ES	expansion segment(s)
FFT	Discrete Fast Fourier Transformation
FSC	Fourier Shell Correlation
GHz	Giga Hertz (= 1 000 000 000 (10^9) Hertz)
GPU	Graphical Processing Unit
GTP	Guanosine triphosphate
HAC	Hierarchical Ascending Classification
HBC	homogenous barycentric coordinates
hrs	hours
IF2	initiation factor 2
kDa	kilo Daltons
kV	kilo Volts
Mg	Magnesium
min	minute
MRA	Multi Reference Alignment
MSA	Multivariate Statistical Analysis
NMR	Nuclear magnetic resonance
NTD	N-terminal domain
OTR	Orthogonal Tilt Reconstruction
PCA	Principal Component Analysis
PCTF	Phase-Contrast Transfer Function

PMF	proton motive force
RCT	Random Conical Tilt
RH	relative humidity
rRNA	ribosomal RNA
RSU	ribosomal subunit
S	Svedberg
S. cerevisiae	Saccharomyces cerevisiae
sec	seconds
SNR	Signal-to-noise ratio
T. maritima	Thermotoga maritima
TEM	Transmission Electron Microscopy
tRNA	transfer RNA
σ	sigma, standard deviation

Acknowledgements

I wish to express my gratitude to Prof Holger Stark in whose department this work was carried out for his supervision, guidance and advice and for propagating his enthusiasm for the methodology of three-dimensional electron microscopy.

I would like to thank all members of my thesis committee, Prof Ralf Ficner and Prof Reinhard Jahn for providing guidance and support throughout this thesis.

This work would not have been possible without the expertise and efforts of my collaborators. I would like to thank Dr Ute Kothe and Prof Marina Rodnina from the University of Witten/Herdecke for providing me with samples of the large ribosomal subunit of *Thermotoga maritima*. For providing *Saccharomyces cerevisiae* eIF3 complex, I would like to thank Sohail Khoshnevis and Prof Ralf Ficner from the University of Göttingen. For providing V-ATPase from *Thermus thermophilus*, I would like to thank Dr Christoph Gerle and Prof Yoshinori Fujiyoshi from the University of Kyoto.

Special thanks go to Dr Niels Fischer, without whose outstanding guidance and help, carrying out this work would not have been possible. For assistance and advice as well as for company, I would like to thank Dr Martin Schmeisser, Dr Burkhard Heisen, Boris Busche, Andrius Krasauskas, Tobias Koske, Dr Mario Lüttich and Ilonka Bartoszek from the cryo-EM group. I further wish to thank Dr Monika Golas and Dr Björn Sander for skilful advice and help in any situation. I would like to thank Dr Prakash Dube for his experienced advice throughout the experimental process. I would further like to thank Dr Dietmar Riedel and Dr Dirk Wenzel for assistance and for providing their microscope in times of need. For excellent technical assistance, I would like to thank Gudrun Heim.

I especially want to thank Dr Steffen Burkhardt from the Coordination Office of the International Molecular Biology Master's/PhD program. Thanks to his untiring contribution and efforts, I was provided an outstanding studying environment which really made me enjoy my scientific work.

For unconditional support in any situation ever since, I wish to thank my parents and family, to which this work is dedicated. For her company and support, I would like to thank my girlfriend Kathy.

Finally, I would like to thank my friends Burkhard, Ben, Marc and Christian for sharing all the good and bad times throughout my stay in Göttingen and for always providing discussion and help.

Abstract

Biological macromolecules exert their biological functions in a dynamic network of large assemblies. The biological function of these macromolecular assemblies is closely related to dynamic movement and rearrangement. Three-dimensional Electron microscopy is the only technique by which dynamic states of large macromolecular complexes can potentially be captured. Currently, methodologies are advancing which allow the reconstitution of structural and functional sub-states from a heterogeneous dataset. Using state-of-the-art biochemical and computational methodology, three-dimensional structures of several biochemically and structurally challenging biological macromolecular complexes have been determined in this work.

The structure of the large ribosomal subunit of *Thermotoga maritima* was reconstructed, displaying several domains which could not be visualized in previous structural studies of ribosomes. The ribosomal L7/L12 stalk, a highly flexible domain which promotes recruitment of translation factors to the ribosome and stimulates their activity, could be visualized in full length. Expansion segments within the ribosomal structure could be identified and localized within the 23S rRNA sequence. As another complex involved in translation, the eukaryotic initiation factor 3 (eIF3) of *Saccharomyces cerevisiae* was examined. Utilizing GraFix, a universal sample preparation protocol for electron microscopy, intact complexes of eIF3 could be prepared for electron microscopy analysis. During subsequent three-dimensional reconstruction of eIF3 complexes, a protocol for the validation of heterogeneous datasets was proposed. The general applicability of the GraFix protocol was further confirmed by the preparation of highly intact transmembrane complexes of vacuolar V-type ATPases from *Thermus thermophilus*. Structural heterogeneity of V-ATPases, which has severely impaired previous studies, could be resolved in a set of three-dimensional reconstructions, allowing functional interpretation of domain movements and stochiometry.

In order to resolve structural heterogeneity within a data set, a set of reliable initial references is needed. For the *ab-initio* determination of initial references, the bottleneck of manual correlation of single particle images from different exposure views limits the applicability of the method. To overcome this bottleneck, a software was designed which automatically correlates particles recorded from different views. Thus, initial references can be obtained from large datasets in a semi-automated way, further improving the applicability and reliability of the method.

1 Introduction

In all living organisms, biochemical processes are governed by a countless number of biomolecules which initiate, catalyze and regulate the complex network of reactions necessary to sustain life. In order to achieve selectivity and specificity, biomolecules dynamically assemble into biomacromolecular complexes. The distinct functionality of a biomacromolecular complex is innately linked to its molecular structure. Knowledge of mechanistics of biochemical and biological processes can be gained from knowledge of the structure from involved biomolecules, and gaining this knowledge has been, since many years, the aim of structural biology.

Major techniques used in the field of structural biology comprise, among others, X-ray crystallography, nuclear magnetic resonance (NMR) and electron microscopy (EM). Being the oldest and most established of these techniques, X-ray crystallography has been contributing structural information of proteins and other biomacromolecules since 1950 with atomic and near-atomic resolution. X-ray crystallography requires relatively large amounts of highly pure protein which have to be present in crystalline form. The field of X-ray crystallography has been constantly boosted by progressing development of cloning and protein expression technology, availability of high-throughput crystallization screens, powerful computational methods, novel phasing approaches and synchrotron-based beamlines. Long being recognized as a routine approach for the determination of relatively small, water-soluble proteins and domains, recent successes in the determination of structures from complexes large as up to several megadaltons and a growing number of transmembrane protein structures solved demonstrate a larger applicability of macromolecular X-ray crystallography. The application of nuclear magnetic resonance (NMR) to questions in structural biology is relatively new, yet, the technique has become one of the major players in the field. NMR is not only able to determine folds but also providing insights into dynamic properties of large biomolecules in solution. Currently, the upper size limit of biomacromolecules for NMR lies around 40 kDa, application of NMR to local regions of larger complexes either directly [1] or by selective deuteration [2] allow insights into larger complexes as well.

Single-particle electron microscopy is routinely used for structure determination of large biomacromolecules with a size starting from around 200 kD without the prior need for

crystallization at mid-range resolutions. Higher resolutions of around 4 Å for highly symmetric particles such as viruses have been obtained (for example, see [3]). Deviating from the initial approach to use the whole dataset to obtain a single reconstruction, more recently developed approaches aim for resolving conformational sub-states and assembly intermediates present in the dataset (for review, see [4]). Since most biomacromolecules are present in larger assemblies in which they exert their specific function with high specificity, insight into structure and dynamics of these biomacromolecules will provide key insights into the complex network of reactions and interactions that determine life. Electron microscopy is the only available technique which routinely allows the examination of structural dynamics of very large complexes. Recent advances in both hardware and software will open up new possibilities which anticipate an increasing importance of electron microscopy in the field of structural biology, providing new insights into the structure and function of the network of life.

In the following, the foundations of single particle electron microscopy which are the basis for recent developments aiming at the studies of structural dynamics of biomacromolecular complexes will be introduced. Recent image processing techniques which allow the disentanglement of structural heterogeneity within the sample dataset will be introduced and discussed. As a general biochemical approach to stabilize biomacromolecular complexes for three-dimensional electron microscopy analyses, the GraFix protocol has been developed [5]. The biochemical principles and their implications for the use on structurally heterogeneous particles will be discussed, followed by an introduction to the macromolecular complexes studied in this thesis.

1.1 Transmission electron microscopy

1.1.1 Image formation in transmission electron microscopy

In transmission electron microscopy as in light microscopy, an image is formed by the differential interaction of incident energy waves with the specimen. In electron microscopy, this energy wave is represented by a highly coherent beam of accelerated electrons. To achieve contrast, i.e. variance of intensities in the two-dimensional image, it is required that the interacting energy waves experience changes in their amplitude upon interaction with the specimen. Given the amplitude A, the intensities I giving rise to contrast in the images are related to A by

$$I \sim A^2 \tag{1}$$

Contrast in electron microscopy of thin and biological samples is achieved by two different modes of electron interaction with the specimen. Blocking of electron transmission by the sample results in direct amplitude contrast. Direct amplitude contrast, however, is of minor importance for transmission electron microscopy of biological samples. Stray amplitude contrast accounts for only 5-15% of the total contrast. Phase contrast is a much more important contrast mechanism in thin and biological samples. It occurs when the object does not absorb energy from the incoming electrons but delays incoming waves, leading to a phase shift of the illuminating wave. Phase contrast may occur upon elastic scattering of electrons by the specimen's Coulomb potential or by interaction with the specimen's electromagnetic field. The incoming wave is called the zero wave w_0 and can be described by

$$w_0 = a_0 \sin \varphi_0 \tag{2}$$

where a_0 is the amplitude and φ_0 is the phase of the wave. When w_0 emerges from the specimen, its phase has been shifted by φ_1. The emerging wave, also called the object wave w_1 can therefore be described in the following expression

$$w_1 = a_0 \sin(\varphi_0 + \varphi_1) \tag{3}$$

which can be redefined to

$$w_1 = a_0(\sin \varphi_0 \cos \varphi_1 + \cos \varphi_0 \sin \varphi_1) \tag{4}$$

As for weak phase contrast objects such as biological specimens, phase shifts are very small, therefore $\varphi_1 << \pi/2$ and thus $cos\ \varphi_1 \approx 1$, $sin\ \varphi_1 \approx \varphi_1$ can be presumed. This simplification allows expression of (4) as

$$w_1 = a_0 \sin \varphi_0 + a_0 \varphi_1 \cos \varphi_0 \tag{5}$$

It can be seen that (5) can be divided into two terms, the original term for the zero wave w_0 (2) and the diffraction wave w_2 which is described by

$$w_2 = a_0 \varphi_1 \cos \varphi_0 \tag{6}$$

In total,

$$w_1 = w_0 + w_2 \tag{7}$$

As it can be already seen in (4), phases of w_0 and w_2 are shifted about 90°. Upon addition of two waves with a phase shift of 90°, the amplitude is not changed. Therefore no intensity variations which could contribute to amplitude contrast according to (1) occur. If we define $\varphi_1' = (\varphi_1 + n\ \pi/2,\ n\ in\ Z)$, then

$$w_2' = -a_0 \varphi_1 \sin \varphi_0 \tag{8}$$

$$w_1' = a_0(1 - \varphi_1) \sin \varphi_0 \tag{9}$$

As shown in (9), the object wave becomes an equivalent of the zero wave but with an altered amplitude, thus resulting in image contrast formation, if an additional phase shift of an integer multitude of $\pi/2$ is introduced. Deviation from the Gaussian focal plane in combination with the spherical aberration of the objective lens results in an additional phase shifts in electron microscopy. By altering the objective lens current, deviation from the Gaussian focal plane (defocusing) can be achieved. The initial phase shift φ_1 is dependent on spatial frequencies. The spatial frequency υ is the inverse of distance d of scattering points (being the basic of later image features). The optimal shift of n $\pi/2$ can be achieved only for a distinct υ, allowing amplitude contrast to occur only over a fraction of the whole resolution range.
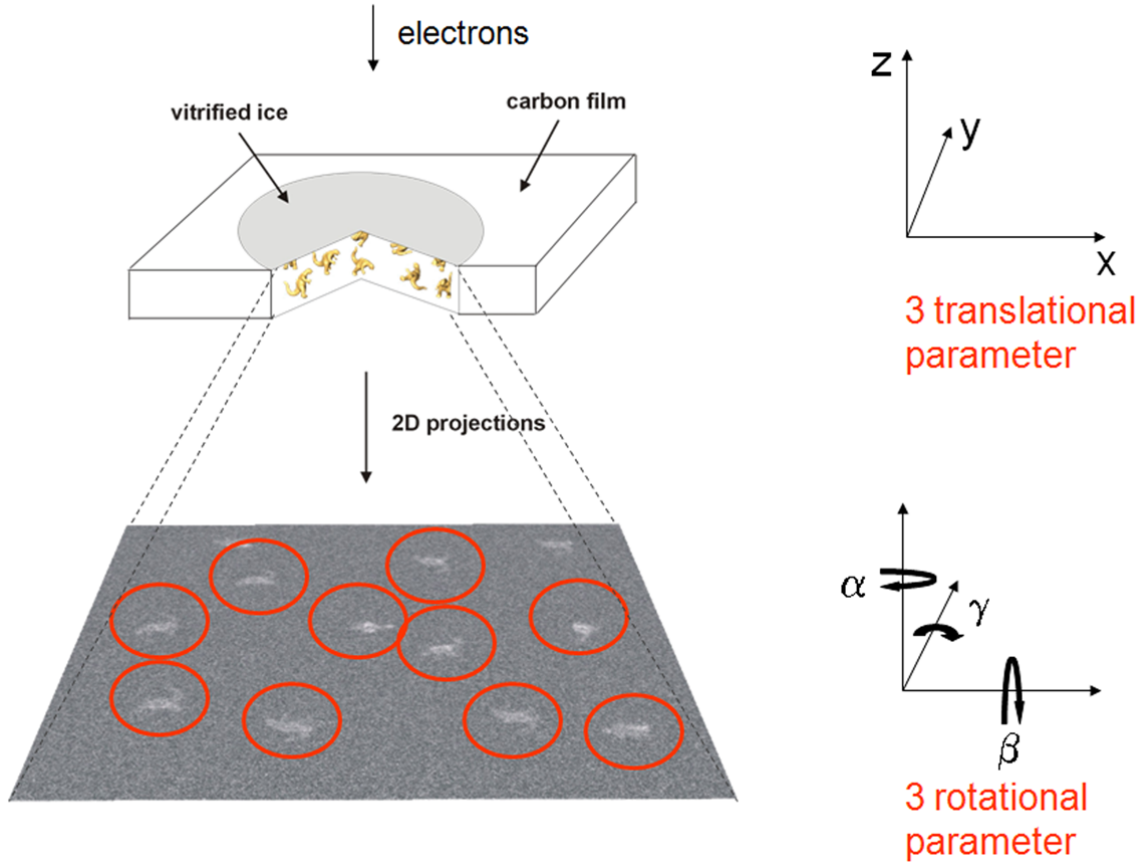
Figure 1 Formation of 2D projection images from a 3D object in electron microscopy. Figure courtesy of Holger Stark.

Generally, transmission electron microscopy of a three-dimensional object will result in a two-dimensional projection image. The orientation of the two-dimensional projection image with respect to the three-dimensional object from which it originates can be

described using three translational parameters (shifts about the x-, y- and z-axis, respectively) and three rotational parameters (euler angles α, β and γ describing rotations around the z-, y- and x-axis, respectively). The relation between the transmission image and its translational and rotational parameters is illustrated in Figure 1.

1.1.2   Phase contrast transfer function

Due to aberrations in the objective lens, the optical pathway in the inner zones of the lens differs from those in the outer zones thus introducing phase shifts to waves passing the lens outside the centre of the optical axis. The objective lens focuses the exit wave emerging from the specimen. The unfocused exit wave represents the diffraction pattern of the specimen, which can mathematically be described as the Fourier transform of individual atomic scattering contributions. Thus, it is possible to describe the differential phase shift introduced by the objective lens in dependency of spatial frequency and defocus [6]:

$$X(R, \Delta z) = 2\pi \left(\frac{R^4}{4} - \frac{\frac{\Delta z}{R^2}}{2} \right) \tag{10}$$

where R is the normalized spatial frequency and Δz is the normalized defocus. As previously described, introduction of phase shifts leads to changes in wave amplitudes and therefore to contrast (see (1)). The contrast in the image changes sinusoidally with varying phase shifts as described in the phase contrast transfer function (*pCTF*) [6]:

$$pCTF = -2 \sin X \tag{11}$$

The *pCTF* at different defoci is shown in Figure 2. It can be seen that for increasing defoci, the first zero value of the *pCTF* is reached at a much lower spatial frequency as compared to low defoci. For spatial frequencies where the *pCTF* equals zero, no specimen information is transmitted.

Spatial frequencies for which the *pCTF* has a negative value have negative phase contrast, i.e. maxima and minima in the image of a periodic structure are inverted. Sometimes, transferred information is not used beyond the first zero value of the *pCTF*, the inverse of the spatial frequency of the first zero is called point resolution limit. However, transfer bands of *pCTF* of higher spatial frequencies can also be used for information transfer and gaps at zero crossings can be filled by merging data recorded at different defoci. In practice, an exponential decay of the *pCTF* can often be observed with increasing spatial frequency. This experimental decay can be due to aberrations in the microscope, incoherence of the electron beam [7, 8], mechanical instability of the sample holder in the
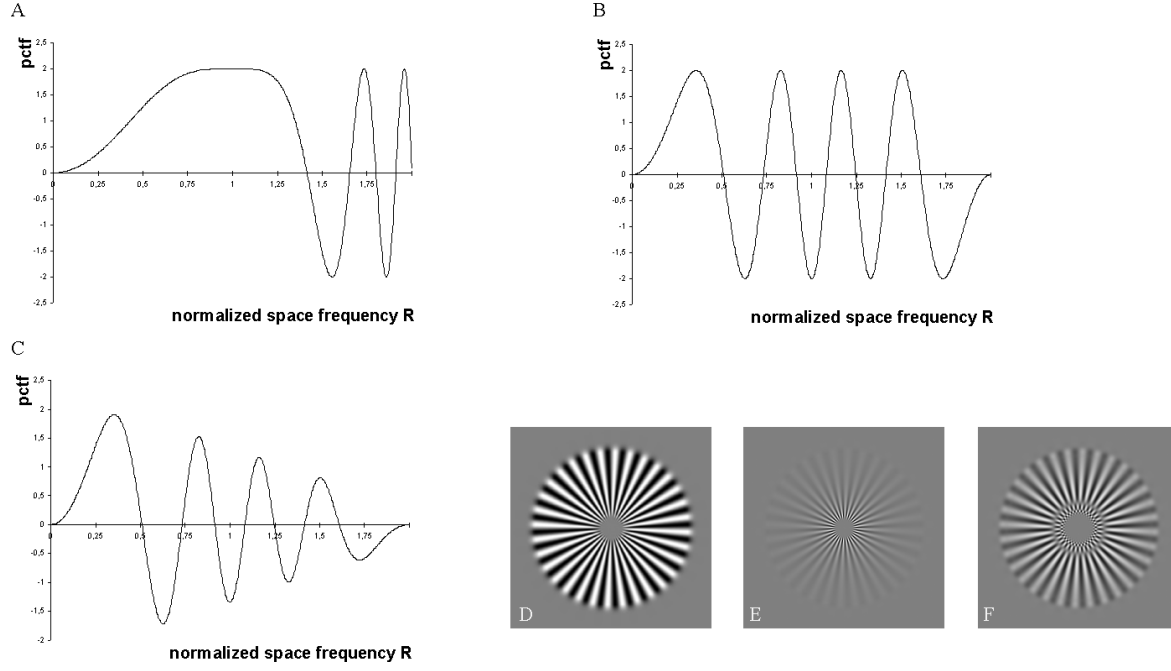
Figure 2 pCTF for different defoci and its effect on image formation. (A) pCTF at 1 Scherzer defocus (B) pCTF at 4 Scherzer. It can be seen that the first zero crossing is reached at lower spatial frequencies for higher defoci. (C) pCTF at 4 Scherzer exponentially decaying (B-factor 0.2). (D) Siemens star. The inner spokes represent high spatial frequencies, spatial frequencies in the picture are decreasing with increasing radius. (E) Siemens star convoluted by the pCTF at 1 Scherzer defocus. High spatial frequencies are transferred without notable intensity losses whereas the transfer of low spatial frequencies is reduced. (F) Siemens star convoluted at high defocus (≈ 75 Scherzer). When the transfer of low spatial frequencies is unhampered, the inversion of phase contrast for higher frequencies can be observed. Inverted areas are separated by grey rings which represent areas around zero crossings were minimum information is transferred.

microscope and modulation transfer functions of film materials and scanners [9]. The exponential decay is described by

$$pCTF' = -2 \sin X \, exp^{-2B(2R^2)} \tag{12}$$

where B is the B-factor describing the characteristic amplitude falloff of the image [10, 11]. Decreasing transfer of higher spatial frequency information is depicted in Figure 2.

$pCTF$ parameters can be extracted from power spectra, which are a square of the fourier transform of an image, thus representing distribution and size of amplitudes of the pCTF with respect to spatial frequencies. Rings seen in power spectra are called Thon rings [12]. To extend the resolution of images beyond the first zero of the $pCTF$, CTF-correction of images can be carried out. CTF correction, astigmatism correction and B-factor estimation was done as described in [9]. Here, the $pCTF$ is estimated by the equation

$$pCTF = -2 \sin X \, exp^{-2B(2R^2)} [(1 - F) \sin X + F \cos X] + S \tag{13}$$

where the first term describes the decay of the $pCTF$, the second term describes the combined effect of amplitude and phase contrast with the amplitude contrast proportion

F<1 and *S* describes the non-convoluted background noise arising from noise in the optical media and non-elastically scattered electrons. A theoretical power spectrum is calculated using different factors by which (13) is multiplied for the x- and y axis of the power spectrum to represent astigmatism. For *pCTF* parameter determination, a cross correlation coefficient between estimated power spectrum densities of images and theoretical two-dimensional *pCTFs* (theoretical power spectra) are optimized. Since the signal-to-noise ratio is low in power spectra of single small particle images, power spectra are averaged. Due to astigmatism, true radial symmetry cannot be expected in power spectra of recorded images, rotational averaging will thus lead to a loss of astigmatism information. This information loss can be prevented by averaging similar power spectra. According to previously described methodology [9], multivariate statistical analysis (MSA) of individual power spectra is done, followed by classification of similar power spectra by hierarchical ascendant clustering (HAC). *CTF* correction is then done by applying

$$pCTF_{corr} = \begin{cases} pCTF \ for \ pCTF > 0 \\ (-1)pCTF \ for \ pCTF < 0 \end{cases} \tag{14}$$

after *pCTF* parameters have been determined from averaged power spectra.

1.2 Image processing in single-particle electron microscopy

In the following, standard image processing techniques which are used for the three-dimensional structure determination from single particle datasets are described. The image processing techniques described in this chapter have been widely used for the determination of three-dimensional structures from single particle electron microscopy datasets following the approach of merging all images of a dataset into a single reconstruction. They still form the basis of more recent methodologies which aim for the reconstruction of several 3Ds from a heterogeneous dataset.

1.2.1 General procedures in 3D reconstruction

In the procedure of 3D reconstruction of single particles imaged by electron microscopy, several methodical restraints have to be considered. First, radiation sensitivity of biomolecules allows imaging only at low electron dose (>20 e$^-$/Å) [13]. The low dose of electrons used for illuminating the sample results in poor contrast and a low signal-to-noise ratio (SNR). Improvement of the low signal-to-noise ratio can be achieved by averaging images which have the same three-dimensional orientation on the grid and hence share the same two-dimensional projection. Since particles representing the same view can be

Figure 3 Iterative refinement of single particle electron microscopy data. Filtered single particle electron microscopy images of macromolecular complexes are aligned (using a rotationally averaged total sum of all imeges in the first iteration) and subsequently classified. From class averages, angular orientations are determined either by angular reconstitution in the first iteration or by projection matching in later iterations. Three-dimensional reconstructions are calculated from averaged images using the filtered backprojection algorithm. 2D projections of the obtained 3D model is are used as a reference for a new round of alignment. This process is iterated until the resolution of the 3D model does not increase any further. Figure courtesy of Martin Schmeisser.

positioned on the grid in infinitive translational and rotational variants, an early and essential step in analysis is to align particles to bring them into register to form a smaller subset of different views. This is achieved by alignment and multivariate statistical analysis of particles in which classes of images sharing the same views are averaged. Class averages can then be used for 3D reconstructions after their angular relationships are determined. 3D reconstructions can be used as references for a new round of alignment and subsequent classification which will yield improved results because the new structure gives a better representation of the particle state on the grid. The procedure of alignment, classification and 3D reconstruction is iterated until the resolution of the 3D reconstruction cannot be any further improved.

1.2.2   Alignment of single-particle electron microscopy images

After initial processing, raw images represent a presumably very large number of particle orientations on the grid. For further processing, it is essential to sort the initial dataset into a smaller subset of groups as well as to centre particles [14]. This is achieved by 2D-alignment of particles. 2D-aligment is used for initial mass centering and grouping of images for comparison in later statistical analysis. Alignment can be used in later iterations for the reconstruction of 3D volumes by assigning 3D parameters of a back-projected

reference image to a particle image from which 3D parameters are known. If no references are known to a given particle, alignment against circular structures is used for initial mass-centering (*reference-free alignment*). Alignment to 2D projections of a 3D reconstruction projected on a uniformly spaced angular grid (*multi-reference alignment*) can be done either by searching for translational and rotational parameters separately or by exhaustive search for three translational parameters (*x*-shift, *y*-shift, rotational angle θ). In this work, we used exhaustive search using resampling to polar coordinates [15, 16]. Rotational alignment of two images corresponds to translational alignment of their respective polar coordinate transforms. Resampling to polar coordinates is done using every point which is sampled during translational alignment as origin for polar coordinate transformation. Rotational parameters during alignment are represented by shifts of the resampled images, translation parameters of the images are represented by the origin of resampling to polar coordinates. Finding all transformation parameters between two images *f* and *g* is done by minimizing the least-square discrepancy between two images [17]:

$$\int |f(u) - g(\boldsymbol{T}u)|^2 du \rightarrow min \tag{15}$$

where $u = \begin{bmatrix} u_x & u_y & 1 \end{bmatrix}^T$ is a vector containing the image coordinates and $\boldsymbol{T}$ is the transformation matrix given by [17]:

$$\boldsymbol{T}(\theta, x, y) = \begin{bmatrix} \cos\theta & -\sin\theta & x \\ \sin\theta & \cos\theta & y \\ 0 & 0 & 1 \end{bmatrix} \tag{16}$$

$\boldsymbol{T}$ is dependent on the three translational parameters *x, y* and θ. After application of the optimal transformation $\boldsymbol{T}$ to match an image with a reference, the correlation coefficient *c* between the shifted image and the reference is calculated in real space by

$$C = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \tag{17}$$

where *n* denotes the number of pixels in the images and *x* and *y* represent greyvalues of pixels in the images *f* and *g*, respectively. Shifted images with the highest *c* to a given reference f_m are assigned to that reference after the transformations $\boldsymbol{T}_m$ have been applied.

1.2.3   Multivariate statistical analysis of particle images

Each picture with n pixels can be represented by a vector

$$\vec{p} = \begin{pmatrix} p_1 \\ \dots \\ p_n \end{pmatrix} \tag{18}$$

The information of *m* pictures can be represented in the *m*-dimensional matrix $Z = m \times \vec{p}$. Classification of images in m-dimensional space requires enormous computational effort. To reduce data complexity while preserving relevant information, eigenvector-eigenvalue data compression is applied. For an eigenvector $\vec{v} \neq \vec{0}$ of an image *f*, the condition

$$f(\vec{v}) = \lambda\vec{v} \qquad (19)$$

applies. (19) can be formulated by

$$Z\vec{v} = \lambda\vec{v} \qquad (20)$$

which can be reformed to

$$(Z - \lambda E)\vec{v} = \vec{0} \qquad (21)$$

where *E* is the unity matrix, λ is the eigenvalue of *Z* and $\vec{v}$ is the eigenvector of *Z*. Since $\vec{v} \neq \vec{0}$, (21) can be solved for

$$det(Z - \lambda E) = 0 \qquad (22)$$

For each λ, an eigenvector can be calculated. With descending value, each eigenvector is pointing towards the most significant variance in *Z*. By applying $\vec{v}_n \perp \vec{v}_{n+1}$ with $\vec{v}_1$ being the eigenvector calculated from the highest eigenvalue, eigenvectors represent a new coordinate system in which the axes are orthogonal to the variance in *Z*. By transforming *Z* into the new coordinate system, dimensionality of hyperspace representing variance information of *m* pictures is reduced from *m* to the number of eigenvectors used for building the new coordinate system. Since the number of elements in $\vec{v}$ is equal to the number of elements in $\vec{p}$, $\vec{v}$ can be displayed as an eigenimage in which the most prominent features of all images in *Z* are visible. With descending eigenvalue for which an eigenvector/eigenimage is calculated, less prominent features up to noise are represented. Thus only eigenvectors for which the eigenimages still show significant feature content are taken into account for further analysis.

For subsequent classification of images, several principal methods can be applied [18]. In the *hierarchical ascendant classification* (HAC) which is used within the Imagic-5 software [19], classification is started with every class containing one population member. Two classes are merged at a time to form larger classes until one total class is formed. Affiliation of a population member to a distinct class dependent on the number of desired classes can be determined within a classification tree. According to the Huygens theorem, optimal classification is achieved if the total variance of a population is equal to the sum of

inter-class variance and intra-class variance. For a predetermined number of classes, the inter-class variance has to be maximal and the intra-class variance has to be minimal to achieve optimal partition [18]. Since the classification of m pictures imposes m possible partitions, testing all possibilities is computationally not feasible. Thus, two classes are added if the increase of the total intra-class variance is minimal upon merging these two classes. This method results in the determination of local minima of intra-class variances which do not necessarily represent the global optimum. Depending on image statistics, weighting of eigenvectors which are used to describe class coordinates can be implied. Either their weighting can be standardized or eigenvectors describing uniform noise (e.g. stain artefacts) can be weighted down.

1.2.4   Angular reconstitution

After having obtained noise-reduced particle views by summing up all particles belonging to the same classes, the angular orientation of these particles with respect to each other has to be determined. To describe the rotational relationship between two orientations of the same three-dimensional object, the angles α, β and γ have to be known. These angles are called Euler angles, γ describes rotation around the z-axis of an object, β describes rotation around the y-axis and α defines rotations around the new z-axis after rotations according to Euler angles β and γ have been applied. Angular reconstitution within the Imagic-5 software suite [19] is accomplished by employing the common line projection theorem. According to this theorem, two 2D projections of a common 3D object share at least one common 1D projection line [20]. In the course of the implemented algorithm, 1D projections of 2D objects are produced, the object is rotated about 1° and a new 1D projection is produced. 1D projections over the full rotational range of 360° are appended one below each other to form an image which is called sinogram (see Figure 4) due to its sinoidal appearance. By calculating the correlation coefficients between all sinogram lines, common 1D projection lines can be found by determining maxima of the correlation coefficient function. For two asymmetrical particles two maxima can be observed, the second maximum is found at the correlation function of two sinograms shifted about 180° with respect to the sinograms where the first peak is observed. This relationship demonstrates the rotational freedom of the 2D projection of two asymmetrical particles around a single common line. This ambiguity can be eliminated by correlating both images with a third image. For symmetrical objects, there is an even number (<2) of common line projections which removes this ambiguity and allows determination of angular
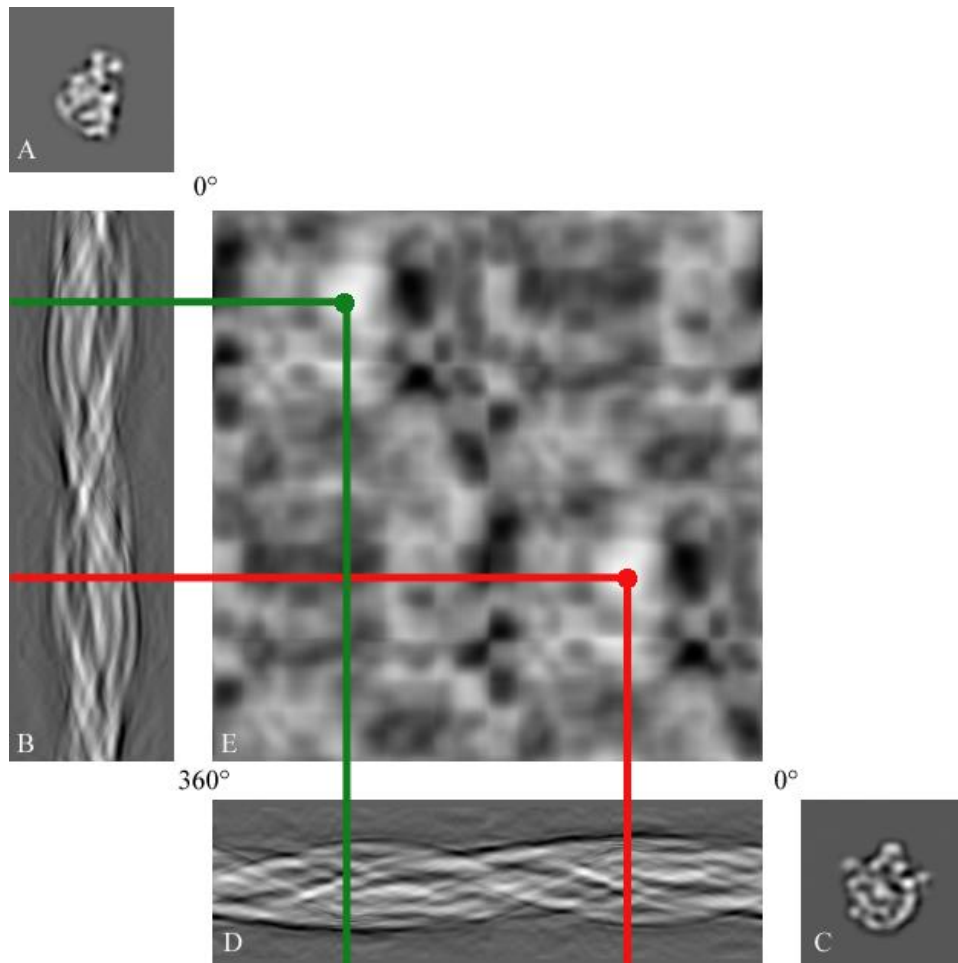
Figure 4 Angular reconstitution by the common line theorem. Two projections of a common 3D object at different Euler angles are represented in (A) and (C). Their sinograms are depicted in (B) and (D), respectively. In (E), correlation coefficients of two 1D projection lines are represented as greyvalues of (E). Note that for two 2D projections of an asymmetrical 3D object, two maxima for the correlation be observed.

relationships between two particles by symmetry-related terms. Once a 3D structure of an object is known, class averages can be determined in their angular orientation by correlation with projections of the previously determined three-dimensional structure (*projection matching*).

1.2.5   Three-dimensional reconstruction

For three-dimensional reconstruction, 2D densities are smeared in three-dimensional space into the direction of their determined Euler angle. Smearing can be described as a projection of each pixel along a vector which is perpendicular to the 2D image plane shifted by its Euler angles. 3D density points are defined by the intersection of two or more vectors. Obtained 3D values can then be projected back and a correlation coefficient of the original 2D and the back-projected 2D can be calculated to determine errors in three-dimensional reconstruction. In the Imagic-5 software package, an *exact filtered back*

projection algorithm [21] is implemented. This filter dampens lower spatial frequencies in the images which are projected into three-dimensional space. Back-projection of images corresponds to overlaying the projections' Fourier transforms as central sections.
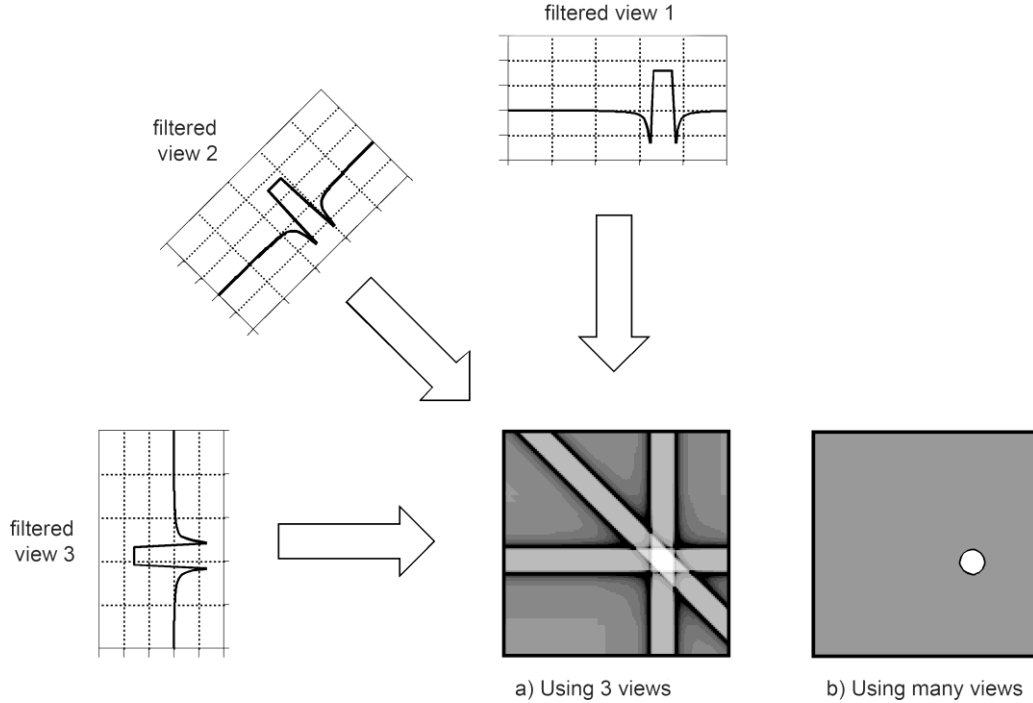
Figure 5 Reconstruction of an image using filtered backprojection. (a) Filtered views of an object are smeared into the direction of their euler angles, indicated by a white arrow. Intersections of projection vectors from each smeared image delineate borders of the reconstructed image. (b) When using many views from different angles, the object is faithfully reconstructed. Reprinted from [22] with author´s permission.

All sections cross in the centre in which lower spatial frequencies are represented, gaps remain in the marginal areas where higher spatial frequencies are represented. This relationship corresponds to an overrepresentation of lower spatial frequencies in real space which is leveled by applying the *exact filtered back projection* algorithm. For an illustration of the filtered backprojection algorithm, see Figure 5.

1.2.6   Resolution assessment

Theoretically, the maximally attainable resolution f_d of cryo electron microscopy 3D reconstructions is limited by the distance of Euler angles δ of 2D images used for three-dimensional reconstitution [23]:

$$f_d \leq \sin \delta \frac{D}{2} \tag{23}$$

In practice, however, the resolution is further limited by several experimental factors such as noise in the images, errors in the determination of euler angles, alignment errors just to

name a few. To calculate the actual resolution of a structure obtained from a certain dataset, the dataset is split arbitrarily into two equal groups. In Fourier space, resolution can be defined as the spatial frequency for which significant correlation between the two images can still be observed. In two-dimensional images, correlation between spatial frequencies of two images is determined by the Fourier ring correlation [24, 25], for three-dimensional image volumes, the Fourier shell correlation is used [21]. Since arbitrarily halved large datasets can be presumed to have the same angular distribution as the original dataset, their maximal resolution is representative for the resolution of the original dataset since parameters for (23) remain nearly unchanged. The correlation between Fourier transforms of two three-dimensional image volumes is calculated by the Fourier shell correlation coefficient (*FSC*) for a spatial frequency R at a shell radius S [21] which is determined by

$$FSC(S) = \frac{\sum_{R=S} F_1(R) \sum_{R=S} F_2^*(R)}{\sqrt{\sum_{R=S} |F_1(R)|^2 \sum_{R=S} |F_2(R)|^{\frac{1}{2}}}} \tag{24}$$

where F_1 is the three-dimensional Fourier transform of the first 3D and F_2 the 3D Fourier transform of the second volume. * denotes the complex conjugated form of the Fourier transform. As final resolution of asymmetrical 3D reconstructions, the inverse of the minimal spatial frequency for which the FSC is three times (3σ criterion) or 5 times (5σ criterion) higher than the FSC of two pictures consisting of Gaussian noise is given (for review, see [26]).

1.3   Analysis of flexibility and structural heterogeneity in single particle electron microscopy

1.3.1   Structural heterogeneity of biomacromolecular complexes

Many proteins are known to exert their biological function in large assemblies, called biomacromolecular complexes [27]. These complexes are stable in some cases like the ribosome, but very often complexes are dynamically assembled. In these cases different functional states of the complex are formed by different assembly states, which, in turn, adopt different conformational states during the exertion of their function. This architecture leads to an enormous heterogeneity of biomacromolecular complexes in the cell. Even for a distinct complex which is designated to perform a distinct function, many different conformational and assembly states can be observed.

During biochemical purification of complexes, it can be expected that those are exposed to a considerable amount of temperature, osmotic, pH and mechanical stress. The purification scheme might lead to a disruption of a part of the complex population, resulting in further heterogeneity of the obtained biomacromolecules.

Further mechanical stress is imposed on the biomacromolecules upon preparation of samples for electron microscopy. This stress is introduced mostly by surface interactions between the biomolecules and the carbon film to which it is adhered or the water-air surface in support film holes.

Standard single particle three-dimensional reconstruction techniques depend heavily on averaging of images of biomacromolecules which share the same orientation. Thus, averaging of a heterogenous particle population strongly impairs the overall process of three-dimensional reconstruction, leading to loss of information and limiting the attainable resolution of the reconstruction. The problem of sample heterogeneity has been addressed both from the side of biochemical sample preparation [5] and image processing techniques (for review, see [4]).

For resolving structural heterogeneity by image processing, a number of techniques have been recently developed. Although differing in methodological detail, all techniques rely on or aim for the reconstruction of an initial set of three-dimensional reference structures in which structural heterogeneity of the dataset is represented at low resolution. Here, two major approaches can be distinguished: Random conical tilt reconstruction (RCT) and sampling with replacements. Both approaches unfold their full potential in context of companion techniques. An overview of established methods for image processing techniques to disentangle structural heterogeneity in electron microscopy datasets is given in the following.

1.3.2 Image processing techniques

1.3.2.1 Supervised classification

A very commonly used method for the analysis of heterogeneity is *supervised classification*, which is among the first described protocols to approach sample heterogeneity [28]. In supervised classification, a dataset is aligned to several sets of references created from projections of different three-dimensional volumes. Notably, pixel intensities of all used volumes should be normalized and the same projection angles should be used for all references. After alignment, each image in the dataset is assigned to the

reference to which it aligned with the best cross-correlation coefficient to any projection. Thus, the dataset is split up in sub-sets according to number of input reference 3D volumes. A new three-dimensional volume is reconstructed from every sub-set and used as a reference to a subsequent competitive alignment step against projections from the other 3D volumes. This procedure is iterated until the resolution of the obtained three-dimensional reconstructions converges.

Obviously, the experimental success of supervised classification is strongly depending on the three-dimensional structures used as input to the first competitive alignment step. In many approaches, the initial model set is obtained via different methods, and supervised classification is used to refine the starting models.

1.3.2.2   Local two-dimensional multivariate statistical analysis (Local 2D MSA)

As an initial approach to estimate structural heterogeneity of the sample, a local classification of the dataset in two-dimensional space can be done. This method does not require prior knowledge of an initial three-dimensional model, thus, model bias from an initial model can be largely excluded. For local multivariate statistical analysis (MSA) in two-dimensional space, images of a dataset have to be passed through several iterations of alignment and classification (which can be done either reference-free or by using an initial model for projection matching). As a result, particles can be averaged into a distinct subset of groups which represent distinct views. In a subsequent step, all single particle images representing a distinct view are submitted to MSA and subsequent classification [29], taking into account only the portion of the image in which structural heterogeneity is to be assessed. Since MSA and classification are done only on a subset of the particle population and, in some cases, only to a fraction of the image, the process is referred to as 'local'. For each particle view, averages are calculated from every class and can be used for interpretation or further analysis.

In this approach, the user has to define the number of classes that have to be calculated from every local dataset. Owing to the fact that a too small number of classes may not resolve differences in the local features and a too high number of classes may result in class averages with low signal-to-noise ratio (SNR), this process is subject to experimental optimization.

The user further has to decide which area of the particle will be analysed for structural heterogeneity. If flexible domains are known from previous biochemical or structural

studies, this decision might be straightforward. It becomes more challenging, however, in the cases where no previous knowledge of the complex exists.

1.3.2.3 Estimation of variance in three-dimensional reconstructions using the bootstrap technique

In the single-particle three-dimensional reconstruction macromolecular complexes, flexible domains significantly impair resolution. Structural heterogeneity affecting only certain domains leads to a loss in reconstruction accuracy (and thus resolution) in affected regions within the macromolecule. Theoretically, local structural flexibility of a three-dimensional reconstruction which averages all conformers could be described as the variance of the set of reconstructions from each reconstruction of each conformer. An estimation of variance in three-dimensional reconstructions of single-particle cryomicroscopy images has been proposed [30]. In this method, three-dimensional *bootstrap volumes* are calculated from projections P with k elements. Using selections with replacements, indices (i_1, i_2,.., i_k) are generated for each $p \in P$ elements from all available projections k. A three-dimensional volume is calculated where each projection is weighted by the factor i corresponding to its index determined by selection with replacements. This approach is iterated until n bootstrap volumes V_n are calculated. Considering non-weighted three-dimensional reconstructions of volumes V_m from projections k, a three-dimensional bootstrap variance map is calculated using

$$\sigma_k^2 = \frac{1}{k-1} \sum_{n=1}^{k} \left(V_n - \frac{1}{k} \sum_{m=1}^{k} V_m \right)^2 \tag{25}$$

Practically, sources of variability are diverse in electron microscopy. For the estimation of structural variance as proposed in [30], the overall structural variance in a cryo-electron microscopy 3D reconstruction is regarded to be composed of conformational variance σ_{Conf}^2 due to alignment errors σ_{Ali}^2, variance introduced from three-dimensional reconstruction errors σ_{Rec}^2 and variance originating from background noise σ_{Back}^2 [30]. The structural variance of the reconstruction σ_{SV} can thus be formulated as a compound variance [30]:

$$\sigma_{SV}^2 = \sigma_{Conf}^2 + \sigma_{Ali}^2 + \sigma_{Rec}^2 + \sigma_{Back}^2 \tag{26}$$

σ_{Rec}^2 can be estimated by applying the bootstrap variance analysis to 2D noise-free projection images and σ_{Back}^2 can be estimated by performing bootstrap variance analysis on

projection images containing samples of background noise from the electron micrograph of the sample [30]. Presuming that the same reconstruction algorithms have been used for the estimation of σ^2_{SV} and σ^2_{Back}, the latter also contains the information on reconstruction variance. Since σ^2_{Ali} has the same statistical properties as σ^2_{Conf} and the two components cannot be distinguished. The structural variance in three-dimensional reconstructions σ^2_{Struct} which is due to conformational variability can be estimated by

$$\sigma^2_{Struct} = K(\sigma^2_{SV} - \bar{\sigma}^2_{Back})$$
(27)

where K is the number of projections and $\bar{\sigma}^2_{Back}$ is the average background variance [30]. The obtained three-dimensional bootstrap variance map can be used to localize flexible parts in the macromolecular complexes which can, as one example, be submitted to local 2D MSA analysis as described in chapter 1.3.2.2.

1.3.2.4   Reconstruction of structures with flexible domains by cross-correlation of common lines

For the reconstruction of three-dimensional volumes from local class averages which have been obtained from images representing all available angular orientations of the particle, the problem remains of assigning two-dimensional projection averages to a three-dimensional volume from which it originates. An approach utilizing the Fourier slice theorem was proposed in which the cross-correlation of common lines is used to sort 2D projections with respect to their underlying 3D model [31]. The approach was demonstrated to perform fairly well on synthetic and negative stain data when the flexible domain in the complex contributed to between one third and one fifth of the overall complex [31]. When applying this approach to unstained cryo electron microscopy data of the ribosomal L7/L12 stalk from *Thermotoga maritima*, output models had no significantly discernible features in the examined stalk region (data not shown). This might be due to limits of this method at low SNR (as envisaged in unstained cryo data) or when the examined region is sufficiently small and flexible (as observed for the ribosomal L7/L12 stalk).

1.3.2.5   Maximum-likelihood refinement of electron microscopy data

In order to account for structural heterogeneity in cryo-electron microscopy at a low SNR, a maximum-likelihood refinement has been proposed [32]. In this approach, each image Xi of dataset X is presumed to be a representation A of k out of K model structures to which

translational and rotational parameters φ are applied, and an independent Gaussian noise with the standard deviation σ is added as image R_i. Thus, each image will further be represented by

$$X_i(\varphi_i) = A_{k_i} + R_i \tag{28}$$

this representation allows a statistical description of the noise portion of each image X_i. The log-likelihood of observing dataset X given a model with parameter set Θ $L(X|\Theta)$ is formulated as the sum of log-likelihood for all images $X_i \in X$:

$$L(X|\Theta) = \sum_{i=1}^{N} \log P(X_i|\Theta) = \sum_{i=1}^{N} log \sum_{k=1}^{K} \int_{\varphi} P(X_i|k,\varphi,\Theta)P(k,\varphi|\Theta)d\varphi \tag{29}$$

where $P(X_i|k,\varphi,\Theta)$ is the conditional probability to observe image Xi at a given k, φ and Θ [32]. The value of $P(X_i|k,\varphi,\Theta)$ is modelled for each pixel j from each translated and rotated image $X_i(\varphi)$ with respect to the corresponding pixel in A_k. Since the probability density function for each pixel value is assumed to have a Gaussian distribution and independence is assumed for the probability density functions of all individual pixels, the conditional probability of observing the entire image given k and φ ban be expressed as a multiplication of $M = (j_1, . j_2, ... j_M)$ Gaussians:

$$P(X_i|k,\varphi,\Theta) = \prod_{j=1}^{M} \frac{1}{\sqrt{2\pi\sigma}} exp\left(\frac{\left([A_k]_i - [X_i(\varphi)]_j\right)^2}{2\sigma^2}\right) \tag{30}$$

The probability density function of k and φ is given by

$$P(k,\varphi|\Theta)d\varphi = \alpha_k \frac{1}{2\pi\varepsilon^2} exp\left(-\frac{q_x^2 q_y^2}{\varepsilon^2}\right)\frac{dq_\alpha}{2\pi}dq_x dq_y \tag{31}$$

where the integral over φ is represented by the summation over $p = 1,...,P$ distinct combinations of projection euler angles β and γ and a summation over $q = 1,...,Q$ in-plane transformations (translations and rotations) [33]. ε is the estimated standard deviation of the assumed Gaussian distribution of the offsets resulting from centering errors in the particle selection process [32]. Since parameters k_i and φ_i for all images are unknown, an expectation-maximization algorithm is used for maximizing the log-likelihood in (29). After each iteration n, estimates of the references A_k are updated by averaging projections X_i with projection vector p and weighting factor w_{kp} for each class k where weighting is applied according to the probability of observing X_i, given class k, parameter set φ and a model with parameter set $\Theta^{(n)}$:

$$w_{kp} = \sum_{i=1}^{I} \sum_{q=1}^{Q} P\left(k, \varphi \middle| X_i, \Theta^{(\mathrm{n})}\right) \qquad (32)$$

The new models are used as references for the next iteration until a user-defined maximum of iterations has been reached.

Maximum-likelihood refinement of electron microscopy by iterative alignment to a weighted average determined in the iteration has been implemented both for the refinement of 2D *projection matching* [32] and 3D structural refinement [33]. Maximum-likelihood-based approaches have been demonstrated to perform significantly better than approaches based on the best cross-correlation coefficient for each images, especially at low SNR [32, 34], since the noise portion of each image is statistically described (see (28)) and can be included in the determination of the weighting factor. By utilizing weighted averages as new references, initial model bias could be significantly reduced to a negligible level [32].

For initial references in the refinement of 3D structures from a dataset, K low-resolution three-dimensional reconstructions from two-dimensional projections which are randomly sampled with replacements are obtained [33]. The number K of initial reference seed volumes V_k (representing three-dimensional cases of references A_k as introduced in (30)) has to be user-defined [33].

The method has been used on several biomacromolecular complexes, including ribosomal samples, demonstrating its applicability to disentangle conformational flexibilities which comprise nonstoichiometric ligand binding and flexible domain movements [33]. For obtaining initial reference volumes, angular orientations of projection images of the dataset have to be known, thus, knowledge of at least a low-resolution map of the structure is required to obtain initial seeds. The number of seeds is determined by the user, yet, if no prior knowledge about the degree of structural flexibility exists, undersampling of flexible sub-states may occur. Seed volumes must contain at least minimal structural difference in order for the maximum-likelihood approach to identify classes of a structurally different sub-state of the biomacromolecule. Since prior knowledge of a consensus structure used to generate initial seeds must be known, an initial bias is introduced which might lead to inaccessibility of structures in which conformational differences comprise major parts of the complex and/or originate from differential assembly states.

1.3.2.6   Three-dimensional multivariate statistical analysis (3D-MSA)

If prior knowledge of singular orientation of projections from a dataset exists, 3D-MSA can be applied as a computationally inexpensive, unsupervised method to visualize conformational and assembly sub-states of a population of biomacromolecules. In this approach, images representing the same orientation with respect to euler angles β and γ are grouped together, either by classification [29] after alignment or by *projection matching*. Images belonging to the same group are submitted to multivariate statistical analysis and classification [29] either over a defined area (local 2D MSA, see chapter 1.3.2.2) or over the complete image. Class averages from all classified sub-groups are randomly selected with replacements and a large set of three-dimensional volumes is reconstructed. Alternatively, random selection with replacements for the reconstruction of three-dimensional volumes can be performed on non-averaged single projection images.
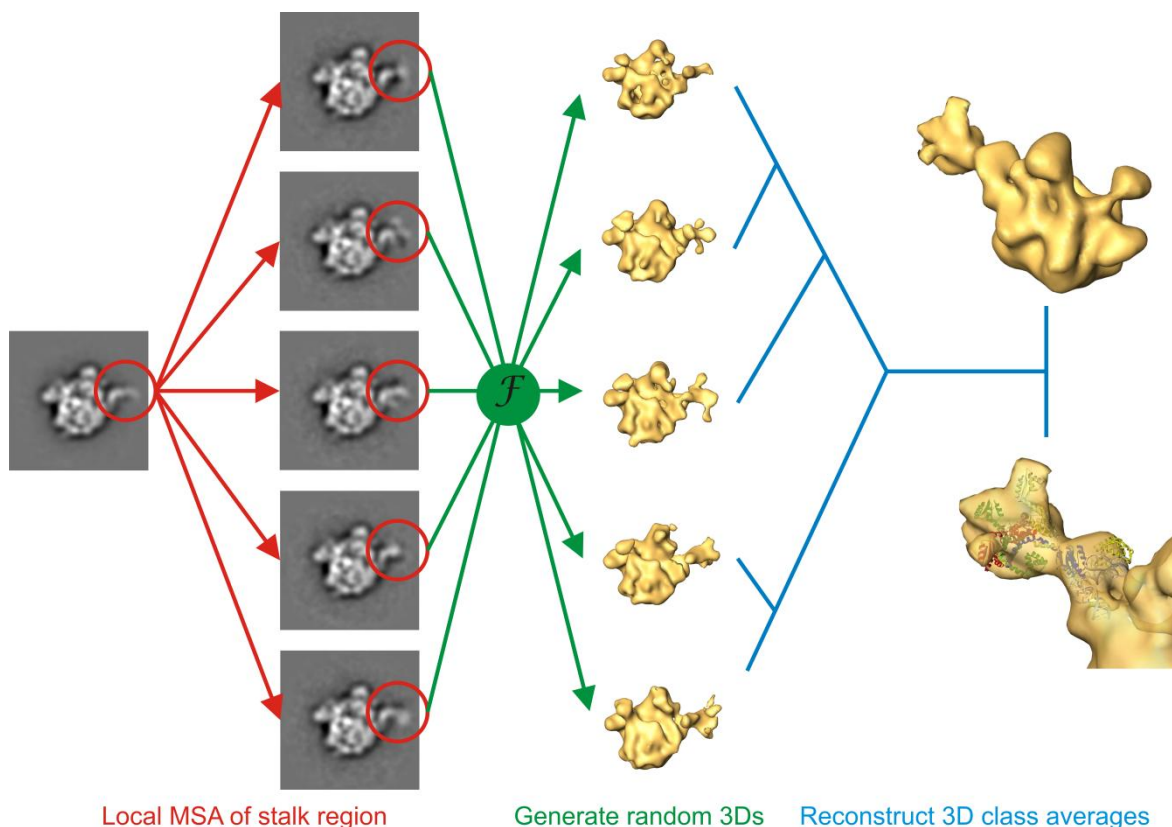
Figure 6 3D MSA of random 3Ds generated from locally classified projection averages. In order to resolve conformational sub-states of distinct parts, images from the same angular orientation were classified according to variance in a distinct part of the projection, here shown for the L7/L12 stalk (left column). Class averages from the same view which are locally classified are shown in the middle left column. Class averages from different angular orientations are randomly mixed in order to obtain a high number of random 3Ds (middle column). In a subsequent step, classification in 3D is done and 3D class averages are formed to resolve conformational sub-states of the particle population (right column).

The large set of obtained three-dimensional volumes is then submitted to multivariate statistical analysis in three-dimensional space [35]. 3D class averages which represent conformational and/or assembly sub-states and in which reconstruction errors in the 3D reconstruction process are suppressed are selected for further refinement by supervised classification (see chapter 1.3.2.1). The overall process of 3D-MSA is illustrated for the example of the ribosomal L7/L12 stalk of *Thermotoga maritima* in Figure 6.

Methodological limitations as mentioned for obtaining initial seeds for the maximum-likelihood refinement of electron microscopy (see chapter 1.3.2.5) apply to the reconstruction of random 3Ds as well, since the principal method is the same: Prior knowledge of the low-resolution consensus structure, i.e. the angular orientation of all projections used for the reconstruction of random volumes has to be known. Furthermore, visual inspection of the obtained 3D class averages which are used for further refinement introduces additional bias into selection of features that are further refined in the subsequent analysis.

1.3.2.7 Random Conical Tilt (RCT)

The random conical tilt (RCT) methodology was introduced in 1987 [36] as a method to obtain initial unbiased three-dimensional reconstructions of biomacromolecules in electron cryomicroscopy. For this method, each particle used for analysis is recorded twice, once in the normal, *untilted* view, and a second time with the specimen holder being *tilted* about a fixed angle, the experimental *tilt angle*. Images collected in the untilted view are iteratively aligned and classified in a reference-free approach [37] until the quality of class averages converges. As a result of the last alignment step, images are rotated by an in-plane angle, the euler angle α. The tilted correlates of all untilted images being grouped into the same class will form a right circular cone with an opening angle that is equivalent to the experimentally fixed tilt angle and a random azimuth rotation around the cone base. The angle of the random rotation corresponds to the euler angle α determined in the previous alignment step (Figure 3).

Since only a few projections have been used for the reconstruction of the RCT 3D volume, the SNR of the obtained three-dimensional volume is very low. Furthermore, the reconstruction will suffer from missing information outside the cone, an experimental parameter referred to as the *missing cone*.
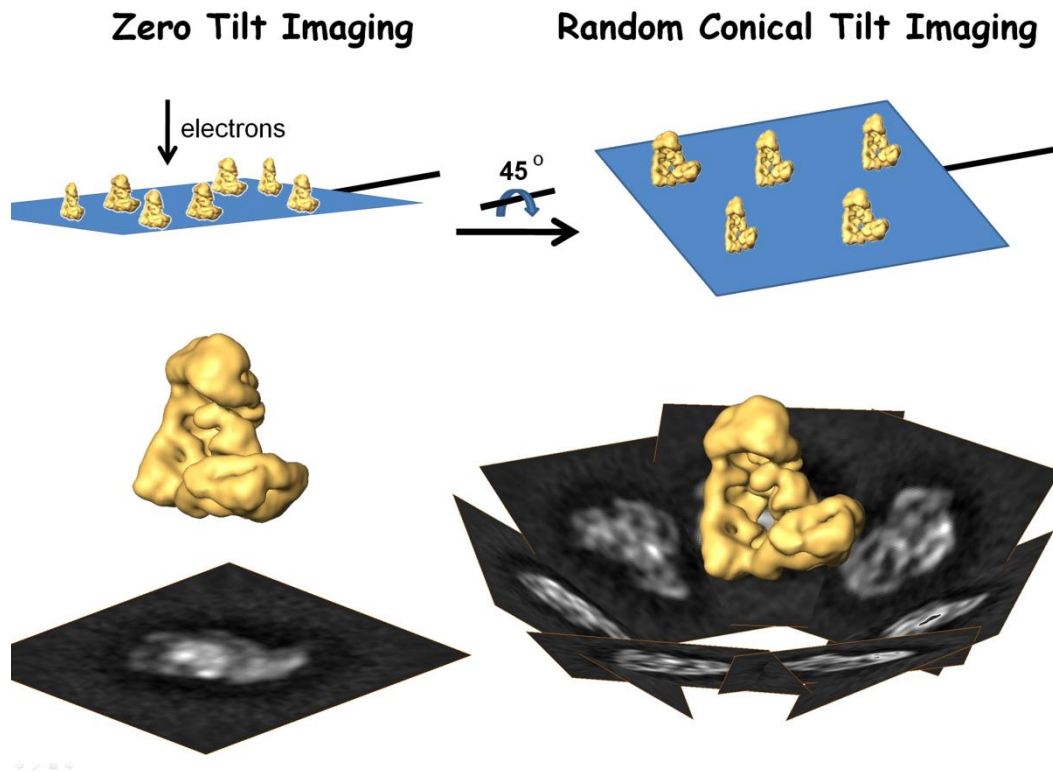
Figure 7 The random conical tilt (RCT) method. EM grids containing monodisperse single particles of a macromolecular complex (here shown for the APC complex [38]) are recorded in the zero-tilt view (upper left) and after tilting the specimen holder by a certain degree (45° in this example, upper tight). From the untilted view, class averages are calculated after reference-free alignment and classification (lower left). Tilted views of particles which are contained in one class are used to reconstruct a RCT 3D (lower right). For the reconstruction of the RCT 3Ds, euler angles γ of each image correspond to to the experimental tilt angle (45° in this example), and random azimuth rotations correspond to euler angles α determined in the previous alignment step. Figure courtesy of Holger Stark.

1.3.2.8   3D Maximum-likelihood alignment (3DMLA)

RCT reconstructions offer a bias-free model that can, in principle, reflect the distinct conformation of a biomacromolecular complex which is represented by the average of the untilted class from which projections are chosen. Since large datasets allow reconstructions of several hundreds to thousands of RCT 3D volumes, potentially all conformational and assembly sub-states can be resolved within the complete set of 3D RCT reconstructions. In order to discriminate distinct structural features from reconstruction errors and noise in 3D RCT reconstructions, RCT reconstructions representing the same structural state of a complex need to be averaged. The need of averaging of RCT 3D reconstructions in order to increase the low SNR and top overcome the missing cone problem in order to gain reliable structural models from a heterogeneous dataset been first recognised by Sander et al. ([39]). A prerequisite for the classification and averaging of 3D RCT reconstructions is,

like for 2D images, registration of all reconstruction with respect to each other. Sander et al. have developed a maximum likelihood-based alignment of three-dimensional volumes. In this approach, alignment is implemented as alignment via polar coordinates [15] with one further averaging step in the z-dimension of the three-dimensional volume. In the current implementation, mass centering of all input 3D volumes is presumed, and alignment is restricted to optimization of rotational parameters. During 3D maximum likelihood alignment, the cross-correlation coefficients of all three-dimensional volumes to the reference volume with respect to all rotational parameters are determined and stored. As mentioned earlier, maximum-likelihood refinement approaches can be used to reduce initial reference bias significantly [32, 34]. In the current implementation of the 3DMLA algorithm, weighting of individual three-dimensional volumes is done according to CCCs assigned to individual parameters and an estimated SNR. For electron microscopy, the SNR can be estimated from the ccc_i between an image i and a noisy reference according to [40]:

$$SNR \approx \frac{ccc_i^2}{1 - ccc_i^2} \qquad (33)$$

Prior to alignment, single particle electron microscopy images are normalized to a common standard deviation. The SNR of an average image obtained from images i_{norm} with individual SNR_i can be maximized by weighting each image i_1 by an individual weigthing factor w_i For the average obtained from normalized images after weighting, thus,

$$i_{1..n} = \sum_{i=1}^{n} \frac{w_i(s + n_i)}{\sigma\sqrt{1 + \frac{1}{SNR_i}}} \qquad (34)$$

holds true (s, n denote the image's signal and noise portion, respectively, σ denotes the standard deviation of the noise). The overall SNR of the weighted sum can thus be estimated by

$$SNR_{i..n} = \frac{\left(\sum_{i=1}^{n} \frac{w_i}{\sqrt{1 + \frac{1}{SNR_i}}}\right)^2}{\sum_{i=1}^{n} \frac{w_i^2}{SNR_i + 1}} \qquad (35)$$

After each round of alignment, the best ccc_i is chosen to determine the translational parameters of each three-dimensional volume and to estimate the noise in the respective

volume. A new reference is created by averaging all aligned three-dimensional volumes weighted by a factor w_i which is obtained my maximizing (), and the next round of the alignment is started.

The protocol is usually iterated over 40 rounds with a decreasing sampling of Euler space from 15° in the first iteration up to 2° in the last iteration. The robustness of the approach has been demonstrated by the successful reconstruction of several macromolecular complexes, even if the initial model was Gaussian white noise, had the wrong handedness or symmetry or when a reference from another complex was used [39]. Thus, 3D maximum-likelihood alignment with subsequent MSA offers a bias-free approach to access populations of highly heterogenous biomacromolecules, even if the structural differences comprise major parts of the molecule and no prior structural or conformational information exists.

1.3.2.9   Choice of methodology

Given a wide spectrum of methods to approach structurally heterogeneous electron microscopy datasets, the choice of a suitable methodology has to be made for each dataset corresponding to its intrinsic features. If prior knowledge of the structure exists, and if structural flexibility can be presumed to occur only within a relatively small and well-defined area of the complex, image processing techniques which rely on sampling with replacements [30, 35] offer a reliable approach. Here, single projections or class averages from which angular orientations are known are repeatedly shuffled in order to obtain initial reference structures which represent heterogeneity of the dataset by randomly containing a higher number of projections from a distinct conformation of particles present in the dataset. 2D Projections of different conformers can be obtained by bootstrap analysis ([30], see chapter 1.3.2.3), optionally followed by local 2D MSA ([29], see chapter 1.3.2.2) and/or 3D MSA ([35], see chapter 1.3.2.6). Initial references can be refined using supervised classification ([28], see chapter 1.3.2.1) or maximum-likelihood refinement in 2D or 3D space ([32, 33], see chapter 1.3.2.5). Presumptions for having a dataset to which high-resolution reference 3Ds are available and in which conformational flexibility is observed only aver a limited range of the particle hold true for most ribosomal complexes. Indeed, successful application of sampling with replacements to create initial datasets have been predominantly reported for ribosomal structures (for review, see [4]).

If no prior knowledge on the structure and structural heterogeneity exists, RCT reconstructions provide a reliable source of initial models ([36], see chapter 1.3.2.7). A

recent implementation of a 3D maximum likelihood alignment helps to overcome methodological restriction imposed by the low signal-to-noise ration and missing cone in single RCT reconstructions ([39], see chapter 1.3.2.8). Averaging of RCT 3Ds after 3D maximum-likelihood alignment will thus be the method of choice if no information about the structure or underlying structural heterogeneity of the target complex exists.

1.4   The GraFix protocol

The biochemical procedures used for high-purity isolation of biological complexes are subject to extensive experimental optimization and vary greatly among the complexes to be purified. Nevertheless, a universal biochemical sample preparation procedure for single particle electron cryomicroscopy, termed GraFix, has been established in the Stark group [5]. The GraFix method is applicable to a large range of biological complexes irrespective of their purification scheme or the buffer systems used. Samples prepared by the GraFix protocol can be used for both negatively stained room temperature and cryo grid preparation as well as for unstained cryo preparation. The GraFix preparation method combines the gradient sedimentation of biomacromolecular complexes (Gra) with weak chemical fixation (Fix) in a centrifugation step in which samples are centrifuged into a gradient containing uni-directional density and fixative concentration gradients. The schematics of the GraFix setup are shown in Figure 8.
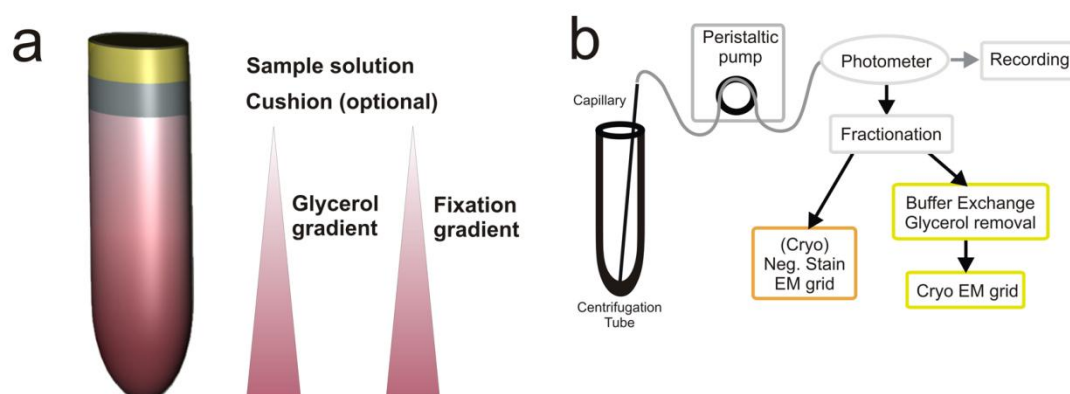
Figure 8 Schematics of the GraFix setup. (a) The GraFix gradient is prepared such that the fixation reagent is added to the denser glycerol solution. (b) GraFix gradients are fractionated from bottom to top, then either used directly for negative-stained electron microscopy or purified away from glycerol for unstained cryo grid preparation. Reprinted with permission from [5].

Attempts to stabilize complexes by weak fixation in mild buffers in which they are well preserved very often lead to aggregation of complexes due to unspecific inter-particle cross-linking which can be promoted, as an example, by low salt concentration. Fixation by chemical cross-linking during ultracentrifugation offers the advantage that weak,

unspecific interactions which promote aggregate formation are disrupted by the pressure that is imposed on the particles. Buffer systems that are incompatible to the fixative used during analysis are exchanged during the GraFix process without requiring alternations to the general protocol. The buffer conditions in the GraFix gradient tube can be adjusted towards the requirements of the purified complex with respect to pH, ionic strength and other factors, whereas buffering agents which are incompatible to the fixative (e.g. Tris as a buffering agent when glutaraldehyde is used as a fixative) can be replaced by a compatible buffer agent (e.g. HEPES as a buffering agent when glutaraldehyde is used as a fixative). The original sample buffer will diffuse into the gradient very slowly, whereas the high molecular weight complexes rapidly migrate over the length of the gradient, thus being protected by buffer-fixative background reactions. Optionally, a fixative-free cushion with a low concentration of crowding agent can be placed on top of the GraFix gradient. After the centrifuge run, fractions are collected and their absorbance are recorded in a photometer. Fractions with intact particles may either be used for negative stain or cryo negative stain grid preparation, or unstained cryo grids may be prepared after buffer exchange.

Macromolecular complexes treated with the GraFix protocol prior to electron microscopy profit significantly from an increased biochemical and structural stability [5]. Generally, a considerable reduction of aggregated and degraded particles on the EM grids can be observed in micrographs of GraFix-treated particles in comparison to micrographs recorded from untreated particles (see Figure 9 a, b). The general improvement of quality over the whole particle population of GraFix treated sample becomes even more obvious when comparing the particles after alignment, multivariate statistical analysis and classification [5]. As it can be seen in Figure 9 (c, d), the number of good class averages in which all domains of the complex are fully defined increases about five-fold with respect to class averages which were obtained from untreated sample following the same image processing procedure [5]. When comparing the mean spectral signal-to-noise ratio (SNR) [41] of all class averages, an increase in signal by 2-5 fold can be observed over a broad range of spatial frequencies [5]. Furthermore, an improved binding behaviour of GraFix-treated particles to carbon film as it is used for electron cryomicroscopy can be observed as well as a more isotropic distribution of projection angles which are exhibited by the particle population [5]. Improving the biochemical quality for single-particle cryomicroscopy, the application of the GraFix protocol is the prerequisite for accessibility

of certain biomacromolecular complexes to three-dimensional electron microscopy analysis. The protocol does not only apply to complexes which are too labile to maintain their structural integrity throughout the standard purification and preparation process. When abundance of the biomacromolecular complex is very low, utilizing the increased binding to electron microscopy carbon film may provide an amount of particles on the EM grid which is sufficient for high-resolution structural studies [5].
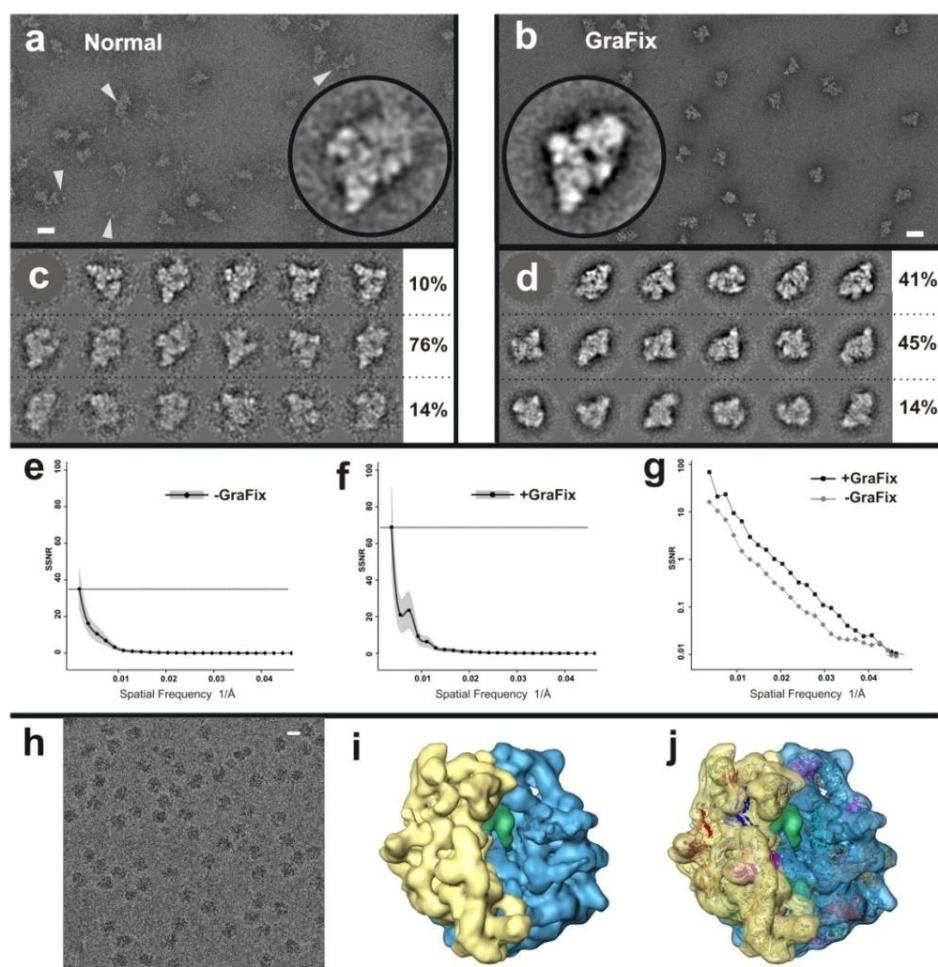
Figure 9 Effect of GraFix-based sample preparation on B complex spliceosomes and the 70S ribosome. (a,b) Uranyl formate–stained electron microscopic raw image of spliceosomes prepared by a conventional glycerol gradient (a) or GraFix (b). Scale bars, 40 nm. Arrowheads, smaller broken parts and flexible elements. Insets, similarly oriented spliceosomal class average. (c,d) Class averages obtained from a set of 5,000 raw images of non-GraFix-prepared (c) or GraFix-prepared (d) samples. The average number of class members is 15 images. Class averages were sorted vertically with respect to contrast and structural definition. GraFix treatment (d) generates computed class averages with much improved contrast (top and middle; 86% of images), as compared to samples prepared by the conventional method (c), where only B10% of class averages (top) show relatively well defined structural features. (e,f) Quantitative analysis of signal quality in class averages of non-GraFix-prepared (e) and GraFix-prepared (f) spliceosomes (gray shading, s.d.). (g) Logarithmic representation of data in e and f. (h) Unstained cryo-EM raw image of the 70S E. coli ribosome initiation complex after GraFix preparation and buffer exchange. (i) Three-dimensional reconstruction of the ribosomal 70S initiation complex at 15 A° resolution. Blue, 50S subunit. Yellow, 30S subunit. Green, P-site tRNA. (j) Fit of the 70S E. coli X-ray structure into the electron microscopy density map. Reprinted with permission from [5].

Structural heterogeneity and flexibility persists within the particle population of most biomacromolecular complexes even after the GraFix protocol has been applied. Even though structural and conformational heterogeneity may be reduced by stabilizing the complexes and confining them to several distinct conformational subs-states by fixation [5], we observed structural heterogeneity after application of the GraFix protocol. Thus, structural heterogeneity has to be further approached by image processing methods.

1.5   Macromolecular complexes studied in this work

In order to assess the performance of biochemical and image processing techniques which aim for the disentanglement of structural heterogeneity in electron microscopy datasets, several macromolecular complexes have been examined.

The 50S ribosomal subunit of *Thermotoga maritima* represents a ribosomal complex which possesses, like other ribosomal complexes, a high structural integrity of the core complex. Furthermore, three-dimensional structures of ribosomal references have been obtained at atomic resolution and may thus be used as an initial reference (for example, see [42]). A wealth of structural and biochemical data exists on a highly flexible domain, the ribosomal L7/L12 stalk (for a more recent example, see [43]).Thus, we attempt to resolve the structural heterogeneity of the complex by using sampling with replacement techniques like local 2D MSA and subsequent 3D MSA (see chapters 1.3.2.2, 1.3.2.6)

The *Saccharomyces cerevisiae* translation initiation factor 3 (eIF3) represents a cytosolic complex which is dynamically assembled. Since no prior information is available either on the structure of the yeast complex or on the structural dynamics, we chose RCT 3D maximum-likelihood alignment and subsequent 3D-MSA ([35, 36, 39], see chapters 1.3.2.7, 1.3.2.8, 1.3.2.6) to determine initial reference structures. Additionally, the effect of the GraFix protocol [5] on stabilization of the complex, concurrently limiting its structural heterogeneity, was assessed.

V-ATPases from *Thermus thermophilus* are transmembrane domains from which previous 3D-EM studies have been published. All previous studies, however, have suffered from instability and structural heterogeneity of the complex. Thus the V-ATPase is an ideal complex for assessing the performance of a tandem approach using a combination of GraFix [5] and RCT 3D maximum-likelihood alignment and subsequent 3D-MSA ([35, 36, 39], see chapters 1.3.2.7, 0, 1.3.2.6).

All macromolecular complexes which have been studied in this work are described in detail in the following.

1.6 The 50S Ribosomal subunit of *Thermotoga maritima*

The hyperthermophilic microorganism *Thermotoga maritima* is a non-spore-forming, rod-shaped eubacterium which grows optimally at a temperature of 80 °C. According to phylogeny analysis based on 16S rRNA sequence, *T .maritima* is one of the deepest and most slowly evolving lineages in Eubacteria [44]. The complete genome of *T. maritima* has been sequenced recently, suggesting extensive lateral gene transfer between archea and *T. maritima* [45]. Among other bacteria, *T. maritima* has been used as a model organism to study the ribosomal L7/L12 stalk [43, 46-48]. The ribosomal stalk comprises the ribosomal proteins L11 and L10 and a region of the ribosomal 23S rRNA to which both proteins bind. A complex is formed by L10 and multiple copies of the ribosomal proteins L7 and/or L12. Proteins L7 and L12 are identical except for an acetylated N-terminus of L12 as compared to L7. L12 (and likewise L7) are composed of a N-terminal dimerization module connected via a flexible hinge region to the globular C-terminal domain (CTD) [49]. As in other thermophilic bacteria, an extended ribosomal L7/L12-L10 stalk has been described, comprised of six copies of L7/L12 instead of four copies as observed in most mesophilic bacteria [43, 50]. The prevalence of a heptameric $(L7/L12)_6/L10$ complex seems to be universal in archeae [51]. In bacteria, L12 has been shown to be required for binding of elongation factor Tu (EF-Tu), elongation factor G (EF-G), initiation factor 2 (IF2) and release factor 3 (RF3) to the ribosome and subsequent ribosome-stimulated GTP hydrolysis (for review, see [52]). It has been further demonstrated that that the GTPase activity of EF-G is increased in the presence of isolated L12 protein [53]. Additionally, specific mutations in the c-terminal domain (CTD) of L12 have been shown to affect the binding of the EF-Tu●GTP●aminiacyl-tRNA ternary complex to the ribosome [54]. Structural comparison of the CTD of L12 and elongation factor Ts (EF-Ts) have, in combination with binding data, led to proposition of a model of transient binding of the L12 CTD to the G-domain of EF-Tu [54, 55]. The exchange rate between fluorescence-labeled L12 and excess unlabeled L12 on the ribosome remained slow, regardless whether EF-G was present or not, suggesting EF-G being recruited by L12 which is bound to the ribosome [43]. Thus, it is nowadays believed that highly mobile CTDs of L7/L12 promote recruitment of translation factors to the ribosome and stabilize the active GTPase

conformation of translation factors bound to the ribosome, stimulating their GTPase activity [43]. Based on the X-ray structure of L10 from *T. maritima* to which three dimers of the N-terminal domain (NTD) are bound, a model for the organization of bacterial stalks has been proposed [43](see Figure 10) A high degree of flexibility is observed for the ribosomal L7/L12 stalk [43]. Due to this high flexibility, the ribosomal stalk is absent or heavily truncated in most of the current X-ray structures and cryo-EM reconstructions of ribosomal complexes. Resolving its conformational sub-states is not only necessary to fully understand the biological function of a macromolecular complex, but also a prerequisite for the reconstruction of high-resolution structural information. Since structural flexibility of macromolecules from thermophilic organisms is reduced at lower temperatures [56], the elongated ribosomal stalk of *T.maritima* offers a promising target for structural studies. When comparing the 23S rRNA of *T. maritima* with known prokaryotic and archaeal sequences, several regions of additional sequence inserts, which probably form ribosomal expansion segments (ES), can be observed. Ribosomal expansion segments have been observed in the cryo-EM maps of eukaryotic ribosomes compared to x-ray structures of prokaryotic [57, 58] and archaeal [59] ribosomes.

Additional densities have been observed mainly in the periphery of the ribosome [60-65]. This finding is most probably due to the structural and functional evolutionary conservation of the ribosomal core among all species. One possible biological function of these ES might involve the recruitment and binding of factors to the ribosome. As one example, ribosomal scaffold protein RACK1 is proposed to provide a platform for signaling proteins on the ribosome [66]. The essential ES 27 of the large RSU rRNA [67] has been suggested to be involved in recruitment and binding of factors such as Sec61 to the peptide exit tunnel [68]. Assigning the precise biological role of ES is, however, impaired by the lack of knowledge of either potential binding partners or structural flexibility of certain parts of the ribosome. By applying reconstruction techniques which take structural heterogeneity of biomacromolecules into account, we hope to gain new insights into structural and functional implications of the ES.
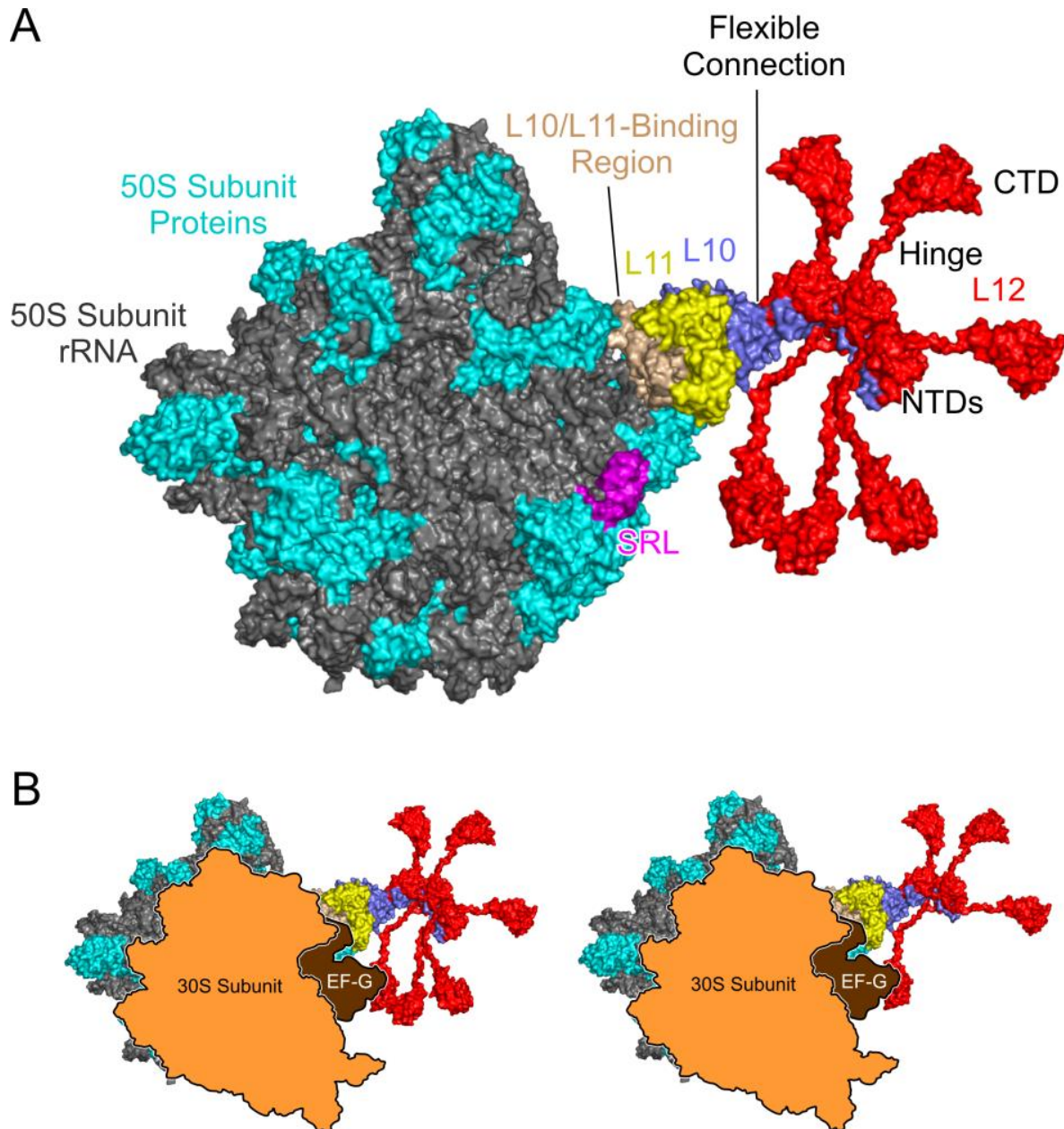
Figure 10 Model for the Organization of Bacterial Stalks. (A) Surface view of a *T. maritima* L10-(L12)6 complex modeled onto the *Haloarcula marismortui* 50S ribosomal subunit (50S subunit proteins, cyan; 23S and 5S rRNA, gray; L10/L11 binding region of 23S rRNA, beige; L11, yellow; L10, purple; L12, red; sarcin-ricin loop [SRL], magenta). The L12 hinge helices are depicted in a randomcoil structure. Orientation of L12 CTDs and hinge regions is arbitrary. (B) The same model as in (A) (left panel) or with a shortened, *E. coli*-type L10-(L12)4 complex (right panel), onto which outlines of the 30S subunit (orange) and of a bound factor (EF-G, brown) are overlaid. Reprinted from [43] with permission.

1.7   The Eukaryotic Initiation Factor 3 (eIF3) complex

Translation initiation in eukaryotes depends on the interplay of a large number of eukaryotic initiation factors (eIFs). At least twelve different eIFs form a dynamic network which stimulates recruitment of the initiator Met-tRNA$_i^{Met}$ and mRNA to the 40S ribosomal subunit, and scanning of the mRNA until the first AUG start codon is reached (for review, see [69]). Regulating initiation, eIFs are the focal point of translational control of gene expression in response to stress, hormones and developmental regulatory factors [70, 71]. The largest eukaryotic initiation factor, eIF3, was one of the first initiation factors to be identified. Mammalian eIF3 is comprised of 13 subunits which are designated eIF3a-m [72-77]. The molecular architecture of the budding yeast eIF3 is much simpler, it contains only five orthologs of five mammalian eIF3 subunits eIF3a, eIF3b, eIF3c, eIF3g and eIF3i [78]. All five subunits of the yeast eIF3 complex have been found to be essential for translation *in vivo* [79-85]. Remaining associated to the 40S ribosomal subunit (RSU) throughout translation initiation, eIF3 forms the scaffold for the interactions of the 40S RSU and the eIFs. Current models assume most of the reactions in the initiation pathway to be stimulated by eIF3. First, the assembly of the eIF2●GTP●Met-tRNA$_i^{Met}$ ternary complex (TC) and its binding to the free 40S RSU is stimulated by eIF1 and eIF1A, eIF3 and eIF5, yielding the 43S pre-inititation complex (PIC) [86-90]. Recruitment of mRNA bound to the cap-binding protein eIF4E and its interaction partners in the eIF4 complex as well as the poly(A)-binding protein (PABP) is stimulated by interactions between eIF4A and the 43S PIC, yielding the 48S PIC. In mammals, interactions between eIF3 and eIF4G have been reported to be involved in this step [91-94]. In the subsequent scanning for the first AUG codon, eIF5 stimulates GTP hydrolysis by the TC while eIF1 promotes scanning [95] and prevents the release of inorganic phosphate from the eIF2●GDP●P$_i$ complex at non-AUG codons [96]. Both eIF1 and eIF5 interact with eIF3 [72, 78, 79, 97], mutations in the N-terminal domain of eIF3c in yeast have been reported to alter the specificity of AUG scanning [98]. In mammals, eIF3 and the eIF4F●m^{7}GTP complex seem to confine eIF5-dependent GTP hydrolysis in the 48S PIC in absence of an AUG start codon at the P-site [99]. Owing to its implications in scanning and AUG recognition, eIF3 is believed to function in the reinitiation of GCN4 and other gene-specific translational control mechanisms. As an example, eIF3 seems to be remaining bound to the 60S RSU via a binding of eIF3g to the ribosomal protein L18 which is mediated by the transactivator protein TAV of plant caulimoviruses [100, 101]. In this way, TAV●eIF3 might be

transferred back to the 40S RSU during translation termination and release of the 60S

RSU. The immediate back transfer would enable it to regenerate a PIC which is capable of

scanning the mRNA to the next AUG, a process during which TC is rebound [100, 101].
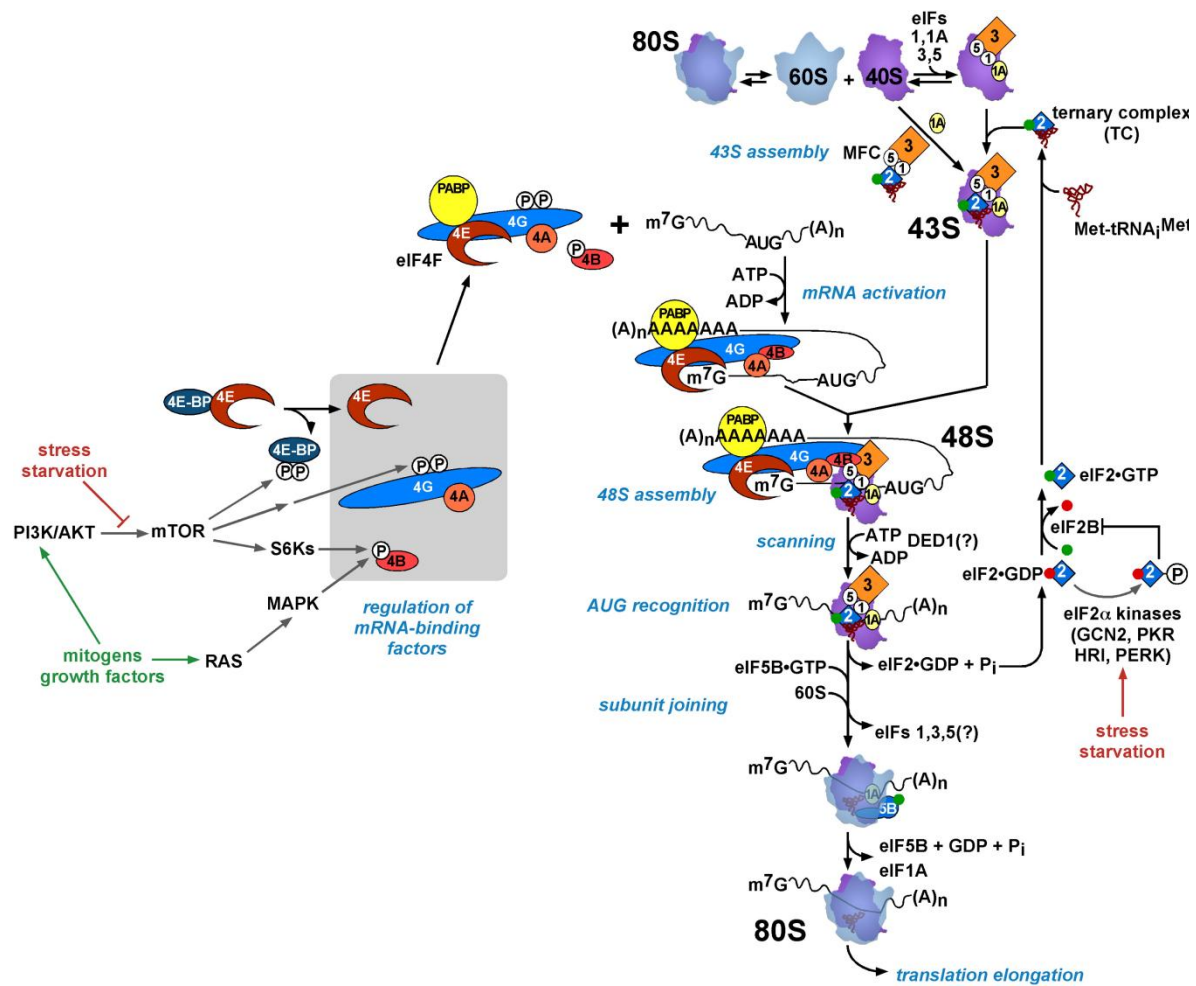
Figure 11 Eukaryotic Cap-Dependent Translation Initiation and Its Regulation. eIFs 1, 1A, and 3 promote dissociation of 80S ribosomes and, together with eIF5 and TC (eIF2•GTP•Met-tRNAi), assemble the 43S PIC. In yeast, these eIFs form a multifactor complex (MFC), which could bind 40S subunits as a unit. mRNA is activated by binding of eIF4F (eIF4E•eIF4G•eIF4A) to the cap and PABP to the poly(A) tail, circularizing the mRNA. The 43S PIC binds near the cap, facilitated by eIF3/eIF5 interactions with eIF4G/eIF4B, and scans the leader for AUG in an ATP-dependent (and possibly DED1-stimulated) reaction, with partial hydrolysis of the eIF2-bound GTP in the TC to eIF2•GDP•Pi. AUG recognition triggers eIF1 dissociation from the 40S platform (not depicted), allowing release of Pi and eIF2•GDP. Joining of the 60S subunits, with release of other eIFs, is catalyzed by eIF5B-GTP, and GTP hydrolysis triggers release of eIF5B•GDP and eIF1A, to yield the final 80S initiation complex. Under stress or starvation conditions, TC formation is reduced by eIF2α phosphorylation and eIF4F assembly is blocked by 4E-BP binding to eIF4E. Phosphorylation by mTOR dissociates 4E-BP from eIF4E. mTOR also promotes eIF4G and eIF4B phosphorylation either directly or via S6Ks. Mitogens and growth factors promote these phosphorylation events by activating mTOR via PI3K/Akt signaling or RAS/MAPK signaling. Reprinted with permission from [102].

Recently, eIF3 has been proposed to provide a regulation point of translation initiation by

acting as a docking site to the mammalian target of rapamycin (mTOR), a protein kinase.

Being recruited to eIF3, mTOR phosphorylates and activates S6 kinase 1 (S6K1) as a

response to adequate nutrient levels or insulin signalling. Being phosphorylated, S6K1 is released from the PIC and in turn phosphorylates ribosomal protein S6 and eIF4B which promotes eIFB4 recruitment and translation [103].

Generally, the eIF3 complex does not seem to have intrinsic catalytic activity and might not directly interact with tRNA$_i^{Met}$ or decoding sites on the ribosome [69]. Yet it is versatile in functionality throughout the process of translational initiation. It is currently believed that eIF3 acts as a scaffold for recruitment and organization for several proteins and protein complexes during translation initiation [69]. A cryo-EM reconstruction of human eIF3 and eIF3 interacting with the internal ribosomal entry site (IRES) of the hepatitis C virus (HCV) has been used to model the position of eIF3 on the 40S RSU [104] based on the position of the IRES determined independently in previous experiments [105]. In this model, human eIF3 binds on the solvent side of the 40S RSU [104]. The binding of mammalian eIF3 to the solvent side of the 40S RSU has previously been reported from low-resolution electron microscopy studies of a mammalian 40S●eIF3 complex [106]. In the binding model of mammalian eIF3, an arm-like feature of eIF3 is suggested to cover the small subunit protein S15/rpS13 [104]. An interaction between S15/rp13 to helix 34 (h34) of the large ribosomal subunit is one of two intersubunit contacts which have been reported to be critical for the assembly of the 80S ribosome [64, 107, 108], suggesting that eIF3 to inhibit intersubunit joining by covering the relevant binding site. In yeast, deletion studies have shown that binding of eIF3 to the 40S RSU is impaired by deleting the N-terminal or C-terminal domains of eIF3c and the N-terminal domain of eIF3a, even if the other components of eIF3 are intact [69]. It has been further shown that a subcomplex consisting of the N-terminal half of eIF3a, eIF3c and eIF5, which binds to the N-terminal domain of eIF3c, can bind the 40S RSU *in vitro* and *in vivo*. Together with experiments showing that the C-terminal domain of eIF3a is required for binding if the eIF3-eIF5 connection is disrupted by mutation [69], these findings support the notion that the N-terminal and C-terminal domains of eIF3c together with the N-terminal domain of eIF3a are among regular binding sites in the interaction of yeast eIF3 and the 40S RSU. Taken together, there might be a multiplicity of binding sites which are accounting for a stable and functional interaction of eIF3, the 40S RSU and other factors. Generally, little is known about the spatial distribution of subunits of both yeast and mammalian eIF3. Knowledge on locations of the subunits within the overall complexes is a prerequisite for the further understanding of structure and function of eIF3.

1.8   The Vacuolar ATPase (V-ATPase) of *Thermus thermophilus*

Adenosine triphosphate (ATP) is the energy carrier for a majority of biochemical reactions within living organisms. Being required in large amounts, ATP is synthesized by ATPases, a family of membrane-bound ion pumps. ATPases can be classified in to three distinct sub-families characterized by different functions and architectures: F-ATPases, V-ATPases and A-ATPases. F-ATPases are referred to as ATP synthetases; they synthesize ATP in a process driven by the electro-chemical gradient across membranes which is created by respiratory chain complexes or during photosynthesis [109]. Thus, F-type ATPases are found in the inner membrane of mitochondria, thylakoid membranes of chloroplasts, and membranes of bacteria [109-112]. F-ATPases are comprising two major domains: the water-soluble F_1 domain and the transmembrane F_0 domain. F-type ATPases are among the best understood ATPases. Crystal structures have been solved for parts of the F_1 domain from mitochondria of bovine heart [113] and rat liver [114] as well as from spinach chloroplasts [115]. A crystal structure including parts of F_0 has been obtained from complexes derived from yeast mitochondria [116], revealing a proteolipid ring which consists of 10 copies of subunit c . In the crystal structure of F_1, a hexagonal arrangement of alternating α and β subunits around a central cavity can be observed. A single copy of subunit γ is bound inside the cavity, reaching out towards F_0 by about 35 Å to form the central stalk. Subunit ε is bound to subunit γ, together, they make contact with the transmembrane ring of c subunits. Laterally connected a subunits are connected to the $\alpha_3\beta_3$ domain via a peripheral stalk comprising b and δ subunits [117]. Functionally, the F-ATPase acts as a rotary motor enzyme. Driven by the proton motive force (PMF), protons are translocated between the interfaces of subunits c and a, causing rotation of subunits c together with γ and ε. The rotation of subunits γ and ε inside the $\alpha_3\beta_3$ hexameric ring enables binding of Mg-ADP and inorganic phosphate to the catalytic sites on the β subunit interface to the α subunit. Subsequently, the product ATP is released. The peripheral b,δ stalk organizes the correct spatial arrangement of F_1 and F_0 throughout the process (for review, see [118]). This subunit design principle is comparable to those observed among V-ATPases and A-ATPases [119]. In some prokaryotes, F-ATPase may function as a proton pump which establishes, in an ATP-dependent manner, an electrochemical transport across the plasma membrane in order to drive secondary transport processes [119]. Dual functionality comprising ATP synthesis driven by the PMF and ATP-driven proton transport along membranes is presumed for most archaeal ATPases (A-ATPases) [119].

Structurally, A-ATPases share a higher degree of similarity with eukaryotic V-ATPases [120, 121] whereas the functional design as reversible ATP synthetases resembles those of F-ATPases [122, 123]. Yet, A-ATPases are poorly characterized in comparison to F-ATPases and V-ATPases, probably due to difficulties in obtaining high yields of protein for biochemical or structural analyses. Electron microscopy studies have revealed similarities between the overall structure of A-ATPase from *Methanococcus janaschii* [124] and the prokaryotic A/V-Type ATPase [125]. Generally, A/V-ATPases found in some bacteria [126-128] are hypothesized to be a result of lateral gene transfer from Archeae [129, 130]. V-ATPases are believed to function almost exclusively as ATP-driven proton pumps in cells, They can be found in endomembranes of cells such as those of clathrin-coated vesicles, chromaffin granules, endosomes, lysosomes, synaptic vesicles, Golgi derived vesicles and in tonoplasts and vacuoles of plants. Their proton-pumping action is vital to a large number of intra- and intercellular processes like receptor-mediated endocytosis, protein trafficking, pH maintenance, storage of metabolites and neurotransmitter release [131-135]. A reversal of the proton-pumping reaction has been described *in vitro* [136, 137]. Thus, the functional preference seems not to be due to fundamental differences between F-ATPases and V-ATPases. The subunit architecture of V-type ATPases is related to those of F-type ATPases although the nomenclature differs. V-ATPases are organized in a cytosolic V_1 and a membrane bound V_0 domain homologous to F_1 and F_0 in F-ATPases. The V-ATPase subunits A and B are homologous to the F-ATPase subunits β and α, respectively [138, 139]. The proteolipid *c* domains of F_0 and V_0, respectively, are most conserved albeit displaying different numbers of transmembrane helices, and are thus believed to have evolved from a common ancestor gene by duplication [130, 140]. The simplest subunit architecture has been described for the V-ATPase from *Thermus thermophilus*s see (Figure 12a). Here, the V_1 comprises a rotor built up from a hexameric A_3B_3 ring arranged around a rotor of subunits *D* and *F*. Subunit *a* and subunit *d*, together with the proteolipid *c* domain form the V_0-domain (subunits *a*, *d* and *c* being homologous to F-ATPase F_0 subunits *a, γ* and *c*, respectively). V_1 and V_0 are connected by at least two peripheral stalks comprising subunits *E* and *G* (for review, see [119]). The presence of a subunit *H* has been reported in eukaryotic V-ATPases (see Figure 12b), being the only subunit that is not required for assembly [141]. Yet, removal of subunit *H* from the intact V-ATPase leads to loss of Mg-ATPase activity [142-144].

Crystal structures of various V-ATPase subunits from different organisms have been solved (for review, see [119]).
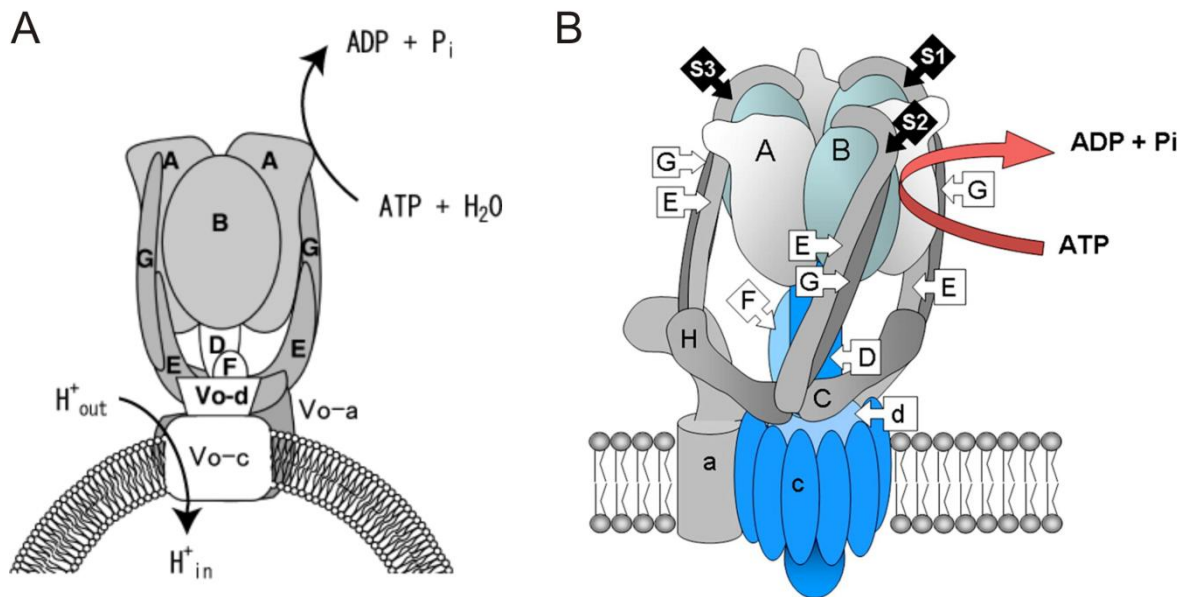
Figure 12 Subunit architecture of V-ATPases. (A) Schematic representation of the subunit architecture model of the V-ATPase of *Thermus thermophilus*. In the current model, a centrals stalk formed by subunits D and F and two peripheral stalks comprising subunits G and E are connecting the A3B3 ring of V1 with the V0 domain comprising proteolipid c domain, subunit a and subunit D (also sometimes referred to as V0-d). Reprinted from [137] with author´s permission. **(B)** Subunit architecture model for eukaryotic V-ATPases. The nomenclature is equivalent to (A), peripheral stalks comprising subunits E and G are lebeled S1-S3, respectively. Additionally, subunit H is present and interconnects S2 and S3. Reprinted with permission from [145].

Recent electron microscopy reconstructions of V-type ATPases provide further insight into the functional roles and interplay of the different subunits. In different reconstructions of complexes from the bovine brain, different stoichiometries of either one [146] or two [147, 148] peripheral stalks from subunits G,E have been reported. The visibility of either one or two peripheral stalks in a plant V-ATPase was dependent on addition of Adenylyl-imidodiphosphate (AMP-PNP), a non-hydrolysable ATP analogue [149]. From complexes examined in presence of AMP-PNP, two peripheral stalks could be visualized [149]. Besides less well defined stalk densities, the reconstruction from the AMP-PNP-devoid complexes revealed a change in the tilt angle of V_1 versus V_0 [149]. Reversible dissociation of V_1 from V_0 has been reported earlier to coincide with a silencing of Mg-ATPase and proton translocation activities [150, 151]. The observed tilting of V1 might display an intrinsic structural regulatory feature towards the dissociation of V_1 at low ATP levels [149]. In a previous electron microscopy reconstruction of the V-ATPase from Thermus thermophilus, two peripheral stalk subunits were asymmetrically attached to the

V_1 subunit [152]. While most electron microscopy reconstructions of V-ATPases have relied on angular reconstitution as a reconstruction method, a more recent study has used the random conical tilt (RCT) technique to unambiguously determine the handedness of a yeast V-ATPase in a bias-free approach [153]. In the subsequent high-resolution reconstruction, three peripheral E, G stalks can be observed [153]. The three stalks, extending from the top of V_1 towards V_0, are interconnected by three elongated arms oriented parallel to the plane of the membrane [153]. Arm1 is anchored to the membrane and interconnects stalks 1 and 2. Arm2, which is located on top of arm1, makes a connection to stalk1. The determined position of arm2 is consistent with the localization of subunit H in bovine V-ATPases [147], and the crystallographic model of subunit H [154] could be accommodated within the observed density [153]. Arm3 connects stalk2 and stalk3 (for comparison, please refer to Figure 12b). Since a weaker density was observed for arm3 in the averaged reconstruction, a higher degree of mobility is proposed for this domain. Its position agrees with localization of subunit c by antibody labelling experiments in *Saccharomyces cerevisiae* [119]. Different conformations of the *EGc* sub-complex in solution as proposed by small angle X-Ray scattering (SAXS) analysis have led to a model for the regulation of V-ATPase activity in which the stability of the *EGc* sub-complex might be altered by cellular signalling events such as phosphorylation [155, 156]. In turn, decreased binding of sub-complexes *EG* and c eventually lead to dissociation of subunit c from the *EG* stalk, which might result in a reduced stability of the V_1. If further ATP hydrolysis occurs, the remaining stalk connections may be insufficient to buffer the system's torque, causing V_1 to dissociates [153]. The prevalence of three peripheral stalk in eukaryotes has been confirmed by a recent study of the cryo-electron microscopy structure of the tobacco hornworm *Manduca sexta* V-ATPase [145]. The overall architecture as interpreted by fitting of crystal structures into the electron microscopy reconstruction is similar as described in yeast [153] (see Figure 12). The cryo-electron microscopy reconstruction at 16.5 Å [145] has the highest of all reported resolutions, and is the first reconstruction of a V-ATPase in its natively hydrated state. The observation of three peripheral stalks connected by a network of arms at the interface of V_1 and V_0 has been observed. As in the negative stain reconstruction, a solvent accessible [153] cavity is observed in the V_0 domain within the axis of the central *F,D* stalk. Additionally, the cryo structure reveals another cavity in the V_1 domain, again found in the axis of the central stalk [145]. Thus, it is concluded that in contrast to F_1 in F-ATPases, the central *F,D* stalk

makes only superficial contact with V_1, probably explaining the dynamic nature of the V_1V_0 assembly in V-ATPases. The collar interconnecting stalks 1-3 is proposed to be comprised of subunits *c* and *H* based on biochemical data [157-159] and docking of crystal structures of both subunits [154, 160]. The interconnecting collar is proposed to play an extended role in the rapid but reversible detachment of V_1 from V_0 [150, 151]. As proposed earlier [153], cellular signalling events such as phosphorylation [155, 156] might lead to detachment of subunit *c* from the membrane, allowing subunit *H* to switch from a position in which it is bound to subunit *c* to subunit *F* of the central stalk, inhibiting ATP hydrolysis in V_1 [161].

Recent electron microscopy studies of eukaryotic V-ATPases [145, 153] challenge the finding of two peripheral stalks in *Thermus thermophilus* [152]. Generally, an increase of subunit interactions is observed as an adaptation to hyperthermophilic environments as in the habitat of *Thermus thermophilus* [162]*,* proposing a potential role for a third peripheral stalk. The asymmetric distribution of peripheral stalks in the previous reconstruction from *Thermus thermophilus* [152] leaves uncertainty about whether the presence of a third stalk might have remained unresolved, either due to high flexibility or reconstruction errors e.g. in *euler angle* assignment during angular reconstitution. The observed cavity in V_0 and V_1 in the axis of the central stalk raises the question how a mechanically stable interaction that can still rapidly disassemble, is established. Applying image processing techniques which allow the resolution of structural sub-population of V-ATPases under given biochemical conditions will expectedly not only allow the retrieval of a higher resolution of the reconstruction, but also allow insights into mechanisms which are as yet not understood.

2 Material and Methods

2.1 Materials

2.1.1 Software

Software	Source
ActivePerl	www.activestate.com/activeperl/
AmiraDev 4.1	TGS Europe, Merignac Cedex, France
assemble0.3	http://www.bioinformatics.org/assemble/wiki/index.php/Main_Page
BOXER from EMAN Suite	http://blake.bcm.tmc.edu/eman/eman1/
ImageJ	rsbweb.nih.gov/ij/
IMAGIC-5	Image Science, Berlin, Germany
NumPy	http://numpy.scipy.org
Python 2.5.1	http://www.python.org/download/releases/2.5.1/
RNAstructure 4.5	en.bio-soft.net/rna/RNAstructure.html
SciPy	http://numpy.scipy.org
SIGNATURE	www.brandeis.edu/~jzchen/signature
SQL Server 2005	Microsoft Corporation, Redmond, WA, USA
TVIPS EM-MENU	Tietz Video System, Gauting, Germany

Table 1 Software used during this thesis

2.1.2 Chemicals

Chemical	Supplier
Acetic acid ethyl ester, Ultrapure grade	Fluka-Riedel-de-Haen, Seelze, Germany
Acetobutyrate cellulose (Triafol)	Sigma-Aldrich, Seelze, Germany
Acetone	Merck, Darmstadt, Germany
Acetyl Coenzyme A Sodium Salt	Sigma-Aldrich, Seelze, Germany
Adenosintriphosphat (ATP)	Amersham Biosciences, Piscataway, NJ, USA
Alkylbenzyldimethylammonium 0.5% (osvan solution)	Sigma-Aldrich, Seelze, Germany
Dithiothreitol (DTT)	Roth, Karlsruhe, Germany
Ethanol	Merck, Darmstadt, Germany
Ethylenediaminetetraacetic acid (EDTA)	Sigma-Aldrich, Seelze, Germany
Glutaraldehyde	Electron Microscopy Sciences, Hattfield, PA, USA
Glycerol	Merck, Darmstadt, Germany
Glycine	Merck, Darmstadt, Germany
Histidine	Sigma-Aldrich, Seelze, Germany
Hydrofluoric acid	Fluka-Riedel-de-Haen, Seelze, Germany
Imidazole	Sigma-Aldrich, Seelze, Germany
Magnesium chloride	Merck, Darmstadt, Germany
Millipore water	Millipore GmbH, Schwalbach, Germany

N-2-Hydroxyethylpiperazin-N-2-ethansulfonic acid (HEPES)	Sigma-Aldrich, Seelze, Germany
Potassium chloride	Merck, Darmstadt, Germany
Potassium hydroxide	Sigma-Aldrich, Seelze, Germany
Sodium chloride	Merck, Darmstadt, Germany
Uranyl formate	Polyscience Inc., Warrington, PA, USA

Table 2 Chemicals used during this thesis

2.1.3   Laboratory materials

Item	*Source*
Carbon rods, highest grade	Ringsdorff Werke GmbH, Bonn, Germany
Centrifuge Tubes	Beckmann, Palo Alto, CA, USA
Copper EM grids, 3.05 mm, 300 mesh, square	Plano, Wetzlar, Germany
Developer Kodak D19 for EM negatives	Kodak GmbH, Stuttgart, Germany
EM negatives (8.3 cm x 10.2 cm)	Kodak Electron Image film SO-163, Kodak GmbH, Stuttgart, Germany
Ethane (liquid)	Messer, Sulzbach, Germany
Filter paper 90 mm, No. 4 and 5	Whatman, Dassel, Germany
Glass slides (76 x 26 mm)	Gerhard Menzel GmbH + Co. KG, Braunschweig, Germany
Mica, 75 x 25 mm	Plano, Wetzlar, Germany
Nitrogen (liquid)	Messer, Sulzbach, Germany
PD MiniTrap G10 columns	GE Healthcare, Munich, Germany
Pierce Centrifuge Column (0.8 ml)	Thermo Fisher Scienctific Inc., Rockford, IL, USA
Teflon preparation block (volume: 25µl – 200µl)	Manufactured by the precision engineering team, MPI Göttingen
Wetting agent for EM negatives	Agepon, Agfa, Köln, Germany
Electron microscopy grids QUANTIFOIL R3.5/1 Cu 200 mesh	Quantifoil Micro Tools GmbH, Jena, Germany
Sterile filters 0.2 µm	Millipore, Billerica, MA, USA

Table 3 Laboratory materials used for this thesis

2.1.4   Special equipment

Item	*Source*
Chart Recorder Bromma 2210	Amersham Pharmacia Biotech, Freiburg, Germany
Cryo electron microscope CM200 FEG	Philips, Eindhoven, The Netherlands
Cryo EM specimen holder	Gatan GmbH, München, Germany
Desktop Computer Dell Dimension 8400	Dell GmbH, Frankfurt, Germany
Edwards E12E vacuum coating unit	BOC Edwards, Kirchheim, Germany
Fraction Collector FRAC-200	Amersham Pharmacia Biotech, Freiburg, Germany
Glow-discharging apparatus	Manufactured by the precision engineering team, MPI Göttingen

Gradient forming instrument	Gradient Master, BioComp Instruments, Inc., Fredericton, NB, Canada
Graphics Board Nvidia Quadro FX 450	PNY Technologies Europe, Mergnac Cedex, France
Heraeus Desk Centrifuge Biofuge Fresco	Heraeus Sepatech GmbH, Osterode, Germany
Peristaltic Pump LKB-Pump P-1	Amersham Pharmacia Biotech, Freiburg, Germany
Room temperature EM specimen holder	Philips, Eindhoven, The Netherlands
Sorvall Centrifuge Evolution RC	Thermo Electron, Langenselbold, Germany
Sorvall Rotor SA-300	Thermo Electron, Langensolbold, Germany
Sorvall Rotor TH660	Thermo Electron, Langenselbold, Germany
Sorvall Ultracentrifuge Discovery 90	Thermo Electron, Langenselbold, Germany
TemCam F415 (slow scan 4K x 4K CCD camera)	Tietz Video Systems, Gauting, Germany
Vitrobot Mark IV	FEI, Eindhoven, The Netherlands
Windows computer cluster consisting of 32 computer nodes of the following configuration:	Sysgen GmbH, Bremen, Germany
Mainboard Tyan Thunder S2877G2NR	
-2x AMD OPTERON 270 2.0 GHz, Dual Core	
19" Rackmount-Case: Chenbro RM117-02, 1HE	
-400 Watt Power Supply EPS P1M-6400P	
-2GB Memory, Kingston DDRAM / PC400 (4x512MB) ECC reg.	
-80GB Harddisk Seagate BARRACUDA 7200.7 7200 rpm	

Table 4 Special equipment used during this thesis

2.1.5 Buffers

All buffers were prepared using Millipore-filtered water and were subsequently sterile filtered using 0.2 micron filters.

Buffer	*Composition*
eIF3 cryo exchange buffer	20 mM HEPES adjusted with potassium hydroxide to pH 7.5 , 50 mM sodium chloride
eIF3 GraFix buffer, 40% Glycerol	50 mM HEPES adjusted with potassium hydroxide to pH 7.5 , 150 mM sodium chloride, 40% glycerol, 0.15% glutaraldehyde
eIF3 GraFix buffer, 10% Glycerol	50 mM HEPES adjusted with potassium hydroxide to pH 7.5 , 150 mM sodium chloride, 10% glycerol
eIF3 sample buffer	50 mM HEPES adjusted with potassium hydroxide to pH 7.5 , 150 mM sodium chloride
TAKM7	50 mM Tris adjusted with HCl to pH 7.5, 20 mM Ammonium chloride, 40 mM potassium chloride, and 10 mM magnesium acetate
V-ATPase GraFix buffer, 10% Glycerol	50 mM HEPES adjusted with potassium hydroxide to pH 8.0 , 150 mM sodium chloride, 10% glycerol, 0.05% dodecyl-ß-d-maltoside
V-ATPase GraFix buffer, 40% Glycerol	50 mM HEPES adjusted with potassium hydroxide to pH 8.0 , 150 mM sodium chloride, 40% glycerol, 0.05% dodecyl-ß-d-maltoside, 0.15% glutaraldehyde
V-ATPase sample buffer	20 mM Tris adjusted with hydrochloric acid to a pH of 8.0, 0.1 mM EDTA, 0.05% Triton-X 100, 500 mM sodium chloride

Table 5 Buffers used during this thesis.

2.2 Biochemical methods

2.2.1 Isolation and purification of biomacromolecular complexes

The 50S ribosomal subunit of *Thermotoga maritima* was purified as previously described [163], except for using a French press for opening of cells. Samples of *Thermotoga maritima* 50S large ribosomal subunit were kindly provided by Dr Ute Kothe, Institute of Physical Biochemistry, University of Witten/Herdecke, Germany. Large ribosomal subunits were provided in $TAKM_7$ buffer (see Table 5). Complexes of eIF3 were expressed in *E.coli* in standard expression vectors using constructs with different combinations of subunits eIF3a-g. Subunits of eIF3 were purified following standard purification schemes. In a final step, eIF3 complexes were purified on a gel filtration column. Fractions were collected while light absorbance at a wavelength of 280nm was recorded. Peak fractions were used for GraFix treatment (see chapter 2.2.2, Table 6). Purified samples of *Saccharomyces cerevisiae* eIF3 were kindly provided by Sohail Khoshnevis, Department of Structural Biology, University of Göttingen, Germany. Samples of eIF3 were provided in eIF3 sample buffer (see Table 5). Intact V-ATPase from *Thermus thermophilus* was isolated as previously described [164]. Samples of *Thermus thermophilus* V-ATPase were kindly provided by Dr Christoph Gerle, Department of Biophysics, Division of Biological Sciences, Graduate School of Science, Kyoto, Japan. Samples of V-ATPase were provided in V-ATPase sample buffer (see Table 5).

2.2.2 GraFix preparation of biomacromolecular complexes

The GraFix protocol was applied to *Saccharomyces cerevisiae* eIF3 andV-ATPase from *Thermus thermophilus*. Both samples were run on a GraFix gradient with a gradient of glycerol (Merck, Darmstadt, Germany) ranging from 10-40% glycerol (top to bottom) and a glutaraldehyde (Electron Microscopy Sciences, Hattfield, PA, USA) gradient ranging from 0-0.15% glutaraldehyde (top to bottom). Glutaraldehyde was added to GraFix gradient solutions immediately before preparation of the gradients. Gradients were prepared in a gradient forming instrument (Gradient Master, BioComp Instruments, Inc., Fredericton, NB, Canada) using pre-programmed parameters for the formation of a 10-40% (w/v) glycerol gradient. Prior to centrifugation, GraFix gradients were stored at 4°C for one hour. 200 μl of sample was loaded on top of the gradient and gradients were centrifuged in a Sorvall Centrifuge Evolution RC (Thermo Electron, Langenselbold,

Germany) using a Sorvall Th660 rotor (Thermo Electron, Langenselbold, Germany) according to conditions listed in Table 6.

Sample	Time/hrs	Speed/g	Temperature/°C
eIF3	16	217144	4
V-ATPase	18	239402	4

Table 6 Ultracentrifugation of GraFix gradients during this thesis

After centrifugation, gradients were fractionated from the bottom and stored at 4°C. In case of eIF3, samples were stored for 24 hours prior to subsequent electron microscopy analysis.

2.3   Preparation of samples for single-particle electron microscopy

2.3.1   Preparation of grids for cryopreparation of samples on carbon foil

For testing purposes, self-made electron microscopy grids were prepared as follows: Light microscopy object slides (Menzel, Braunschweig, Germany) were cleaned with absolute ethanol (Merck, Darmstadt, Germany) and incubated in osvan solution (alkylbenzyldimethylammonium 0.5%, Sigma-Aldrich, Seelze, Germany) to render object slides surfaces hydrophobic. Object slides were immersed into Millipore-filtered water (Millipore, Schwalbach, Germany) and removed carefully to avoid residual water on the slide surfaces. After slides were incubated in an atmosphere of elevated air humidity on a cooled aluminium block to promote condensation of small water droplets on the slides, 1 ml triafol (acetobutyrate cellulose, Sigma-Aldrich) in acetic acid ethyl ester 0.5 % (Ultrapure grade, Fluka Riedel-de-Haën, Seelze, Germany) was pipetted over the glass slide. After evaporation of acetic acid ethyl ester, a thin but solid triafol film remained on the object slides, leaving small holes where water droplets have been condensing. Overhanging foil was removed generously and foil was checked for adequate hole size under a light microscope. Foils suitable for further processing were immersed in hydrofluoric acid 0.1% (Fluka) for 30 seconds to allow ablation of the triafol film. The film was floated on the surface of Millipore-filtered water (Millipore) by slowly immersing the object slide at an angle of approximately 45°. Copper grids (Agar, Essex, UK) were cleaned in ethanol in an ultrasonic water bath and then placed onto the film with their dull side facing the triafol film. Triafol films completely covered with copper grids were taken out by slowly immersing an object slide on which a paper sticker was fixed into the water while pushing a free margin of the triafol film under the water surface. Doing this, the side on which the paper of the sticker is fixed has to point towards the triafol film with the

copper grids on it, ensuring an orientation in which the whole triafol film would be able to completely adhere to the object slide. Once the triafol film has been completely submersed and adhered, the object slide is turned to face the side with the copper towards the water surface and then taken out of the water. Grids with triafol film on them were dried and afterwards sputtered with carbon as described above. Here, direct sputtering was applied to the copper grids. Subsequently, Triafol was removed by incubating the carbon-coated copper grids on a filter paper (Whatman, Oxon, UK) soaked with acetic acid ethyl ester in a closed glass vial overnight. Alternatively, grids were submersed in acetic acid ethyl ester for 5 seconds and then air-dried. Grids were checked by light microscopy for complete removal of triafol after ethyl acetate treatment. For the production of continuous carbon films, mica (Plano G250-1, 25x75 mm) were freshly split and placed into a Edwards E12E vacuum coating unit (BOC Edwards, Kirchheim, Germany) equipped with special carbon rods (Ringsdorff Werke GmbH, Bonn, Germany). Indirect carbon coating was achieved by placing mica into a metal basket which prevented direct carbon aggregation on mica (custom-made). Sputtered carbon was reflected on mica by object slides positioned at the basket's side. Continuous carbon foil was floated on the surface of Millipore-filtered water as described for triafol films. Copper grids with holey carbon film were placed onto the floated continuous carbon foil with the carbon-coated side of the grid facing towards the continuous carbon foil. Grids were taken out of the water as described for triafoil and air dried. Prior to sample application, grids were rendered hydrophilic in a glow discharger (custom-made, 5-10 discharges for around 2 seconds each) 45-60 min before sample application.

For high-quality preparation of electron microscopy grids of the final sample used for three-dimensional reconstruction, Quantifoil grids (Quantifoil Micro Tools GmbH, Jena, Germany) were used. Quantifoil grids are shipped with triafoil support film. Carbon sputtering and cleaning with ethyl acetate was done before floating of continuous carbon film as previously described.

2.3.2 Preparation of negatively stained samples

For estimation of sample quality of different fractions obtained after ultracentrifugation and for assessment of optimal sample concentration on the carbon foil, samples were negatively stained with uranyl formate [165]. Approximately 0.5 mg uranyl formate were dissolved in 200-300 µl of Millipore-filtered water and centrifuged for a minimum of 30 minutes at 4°C to pellet uranyl formate crystals which would hamper electron microscopy

analysis. 5 ml of sample was pipetted on the carbon-coated surface of glow-discharged grids and incubated for 30 seconds. Residual sample was removed by blotting from the sides with filter paper (Whatman). 5 µl of negative stain solution was applied and incubated for 30 seconds. Stain solution was then removed by blotting as for the sample solution, and grids were subsequently quick-dried under a light bulb.

2.3.3   Preparation of unstained cryo samples

2.3.3.1   Preparation of unstained cryo samples from the *Thermotoga maritima* 50S large ribosomal subunit

Samples of *Thermotoga maritima* 50S large ribosomal subunit were vitrified for unstained cryo analysis [166]. 5 µl of sample were pipetted on a glow-discharged carbon-coated grid and incubated for 1 min. Excess buffer was removed by blotting with a filter paper (Whatman) from the front. When excess buffer stopped migrating on the filter paper, grids were immediately plunged into liquid ethane cooled by liquid nitrogen via a guillotine (custom made). Grids were quickly transferred into liquid nitrogen where they were stored for further electron microscopy analysis.

2.3.3.2   Preparation of unstained cryo samples from eIF3

GraFix fraction 15 out of 26 of eIF3 was submitted to a buffer exchange procedure via dead-volume gel filtration using PD MiniTrap G-10 colums (GE Healthcare, Munich, Germany). The standard gravity protocol was applied and the eluate was collected in fractions of four drops. Fraction number four out of six was used for the preparation of unstained cryo samples [166]. 5 µl of sample were pipetted on a glow-discharged carbon-coated grid and incubated for 2 min in the Vitrobot *Mark IV* (FEI, Eindhoven, The Netherlands) at 4°C and a relative humidity (RH) of 90%. Subsequently, grids were blotted in the Vitrobot using the following blot parameters: blot force 0, blot time 5s, blot total 1. Grids were transferred to liquid nitrogen where they were stored for further electron microscopy analysis.

2.4   Electron microscopy analysis

2.4.1   Transmission electron microscopy

Transmission electron microscopy images of all complexes were recorded in a CM200 FEG electron microscope (Philips, Eindhoven, The Netherlands) at acceleration voltages of

160 kV and 200 kV. Images were recorded either on Kodak Electron Image films (SO-163, Kodak GmbH, Stuttgart, Germany) and subsequently digitized with a step size of 10.58 μm corresponding to a pixel size of 2.133 Å on the object scale on a Heidelberg Prime Scan drum scanner (Heidelberg Druckereien, Heidelberg, Germany), or recorded on a 4K x 4K Charge Coupled Device (CCD) camera (Tietz Video Systems, Gauting, Germany) with 2-fold binning. Parameters of transmission electron microscopy for the samples studied in this thesis are summarized in Table 7.

Sample	Exposure Conditions	Acceleration Voltage (kV)	Magnification x1000 fold	Post-magnification x1000 fold	Recording Device
Thermotoga maritima 50S large ribosomal subunit	unstained cryo in holes	200	50	50	Kodak electron SO-163 film
Thermotoga maritima 50S large ribosomal subunit	unstained cryo on foil	200	115	161	CCD camera
Saccharomyces cerevisiae eIF3	room temperature negative stain	160	88	122	CCD camera
Thermus thermophilus V-ATPase	room temperature negative stain	160	88	122	CCD camera

Table 7 Transmission electron microscopy of samples studied in this thesis

2.4.2   Processing of raw images

Images recorded in spot series with CCD camera were combined in a large image (micrograph) using the ImageJ software. Single particles were selected from micrographs and digitized films using semi-automated routines [167, 168]. Selected particles were extracted from the micrograph as small single images which are appended in a single file. Images were corrected for CTF and astigmatism using multivariate statistical analysis and subsequent classification of individual power spectra [9]. Prior to further analysis, a soft circular mask was applied to the images and images were bandpass filtered to exclude very high and very low spatial frequencies. Additionally, pixel intensities were normalized to a common standard deviation. Extraction and preparation of images was done using the IMAGIC-5 Software [19].

2.4.3   Image processing

Image processing of electron microscopy data was carried out using the IMAGIC-5 software [19]. Alignment of images to reference templates was performed on a non-dedicated, dynamic compute cluster [169], using alignment via transformation to polar

coordinates [16]. After obtaining a set of initial models, standard routines for the refinement of electron microscopy data were followed (see chapter 1.2.1). Methodologies for obtaining initial models are summarized in Table 8, and explained in detail in the Results section.

Sample	Method for obtaining initial model
Thermotoga maritima 50S large ribosomal subunit	Angular reconstitution using low-pass-filtered crystal structure of *Thermus thermophilus* 50S large ribosomal subunit (PDB ID: 2j03, [42]) as an initial reference (see chapter 1.2.4)
Saccharomyces cerevisiae eIF3	3D Maximum-likelihood alignment of RCT 3Ds, subsequent 3D MSA (see chapter 1.3.2.8)
Thermus thermophilus V-ATPase	3D Maximum-likelihood alignment of RCT 3Ds, subsequent 3D MSA (see chapter 1.3.2.8)

Table 8 Methods for obtaining initial models in image processing of electron microscopy data in this thesis

2.5   The MaverickTilt Software

To obtain bias-free initial references which reflect structural heterogeneity in the macromolecular complexes which have been recorded in the electron microscope, 3D MSA and classification of random conical tilt reconstructions (RCT 3Ds) can be done after 3D maximum-likelihood alignment of RCT 3Ds [39]. 3D classes represent conformational- and assembly substates of macromolecular complexes present in the sample population. By averaging RCT 3Ds which belong to the same class, the low signal-to-noise ratio of single RCT reconstructions can be increased. Additionally, missing structural information in the *missing cone* in each single RCT can be replaced by averaging several RCT 3Ds where the missing cones are non-overlapping ([39], for review, see chapter 1.3.2.8). To resolve the complete range of conformational and assembly sub-states in the particle population, a large number of RCT 3Ds is desired.

The calculation of RCT 3Ds requires the selection of image pairs on two images representing once an untilted view (zero-tilt image) and a tilted view (tilted image, see Figure 7) of the same specimen area. Due to technical difficulties in the image acquisition process which has to be done separately for both images, the two images are in most cases at least partly non-overlapping. This results in the fact that particles which are present in the zero-tilt image of the specimen area do not necessarily appear on the tilted image of the specimen area and vice versa. Currently, correlating pairs of particles on the untilted specimen micrograph and the tilted specimen micrograph, termed *tilt pairs*, have to be

manually selected. An example for manually selected *tilt pairs* via the *cowPicker* software engineered by Mario Lüttich can be seen in Figure 13.

The process of manual selection of tilt pairs is tedious and time-consuming, as for each micrograph, more than 1,000 particles have to be selected in some cases. The manual selection of tilt pairs is currently the most severe bottleneck in the routine processing of reconstructing random conical tilt (RCT) data.
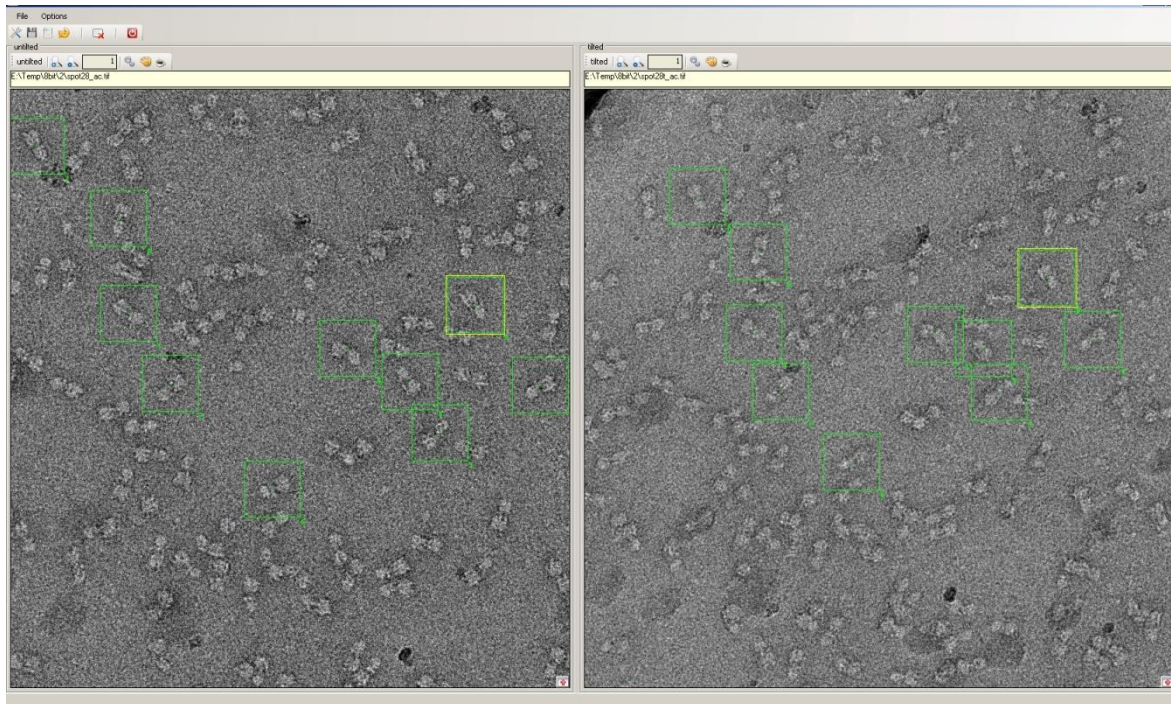
Figure 13 Tilt pairs of negatively stained V-ATPase. Screenshot of the *cowPicker* software displaying an untilted micrograph (left) and the corresponding area in the micrograph of the tilted specimen (right). Particle coordinates have been manually selected as center coordinates of the green boxes around the selected particles.

Currently, no automated routines for the selection of *tilt pairs* exist. Semi-automated approaches assist the selection of the tilted particles by predicting the position of the tilted particle during manual selection of the untilted correlate (i.e. [170], cowPicker software). Although facilitating the tilt pair selection process, all untilted particles have to be manually selected. Additionally, the prediction of the tilted correlate is inaccurate and has to be manually corrected in most cases.

To overcome the bottleneck of manual selection of tilt pairs, the MaverickTilt software has been developed. It automatically correlates center coordinates of particles in the untilted micrograph with center coordinates of particles in the tilted micrograph in order to establish tilt pairs which can be used for the calculation of RCT 3D volumes. Particle center coordinates in either micrographs can be obtained seperately using semi-automated

routines [167, 168]. The MaverickTilt software determines tilt pairs without any user input or supervision being required. MaverickTilt uses an iterative approach, in which an initial set of correlates is determined by transposing all particle coordinates into a feature space which is invariant to the image transformation introduced by tilting of the specimen holder. Using pattern recognition algorithms, local patterns of points are compared in invariant feature space in order to find correlating point sets. After an initial set of reference coordinates is determined, it can be used to locally predict the locations of tilted particles from the locations of untilted particles. For the prediction of particle coordinates on the tilted micrograph, features which are invariant to image transformations applied by tilting of the sample area are used. Local prediction of tilted particle center coordinates is iterated over the full range of particles until the number of identified tilt correlates no longer increases.

In the following, the underlying algorithms used by the MaverickTilt software are described.

2.5.1 Basic algorithms

The geometric operation which leads to the projection of the tilted image from the untilted image is referred to as affine transformation. Generally, affine transformations of an image consist of a linear transformation (rotation, scaling or shear) followed by a translation. The affine transformation that describes the projection of a two-dimensional point set during RCT analysis comprises a rotation around all three-dimensional axes and a translation. Using homogenous coordinates, the affine projection of the two-dimensional point set of center coordinates from the untilted image to its position in the tilted image can be described by the following transformation matrix:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta_1 & \sin\theta_1 & 0 \\ -\sin\theta_1 & \cos\alpha\theta_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta_2 & \sin\theta_2 & 0 \\ -\sin\theta_2 & \cos\alpha\theta_2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta_3 & \sin\theta_3 & 0 \\ -\sin\theta_3 & \cos\alpha\theta_3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (36)$$

where t_x and t_y denote translations along the x- and y axis and θ_n denotes the angle of rotation around the axis of dimension n. The transformation matrix in (36) describes the affine transformation of $x \rightarrow x'$ and $y \rightarrow y'$, respectively, which projects points from the zero-tilt image to the tilted image in RCT analysis. The particle center coordinates (denoted by x,y and $x'y'$, respectively) are referred to hereafter as 'points'.

By minimizing the least square error for equation (36), the parameters that determine the underlying affine transformation can be estimated. Since the dataset is partly non-overlapping in case of electron microscopy tilt pairs, (36) cannot be formulated *ab initio*.

Conventional approaches (as used, for example by XMIPP [170] and the cowPicker software) use least-square minimization of (36) using parameters that have been obtained from manually selected tilt pairs. Using these parameters, (36) is used to predict the location of center coordinates of a particle on the tilted micrograph from its untilted correlate.
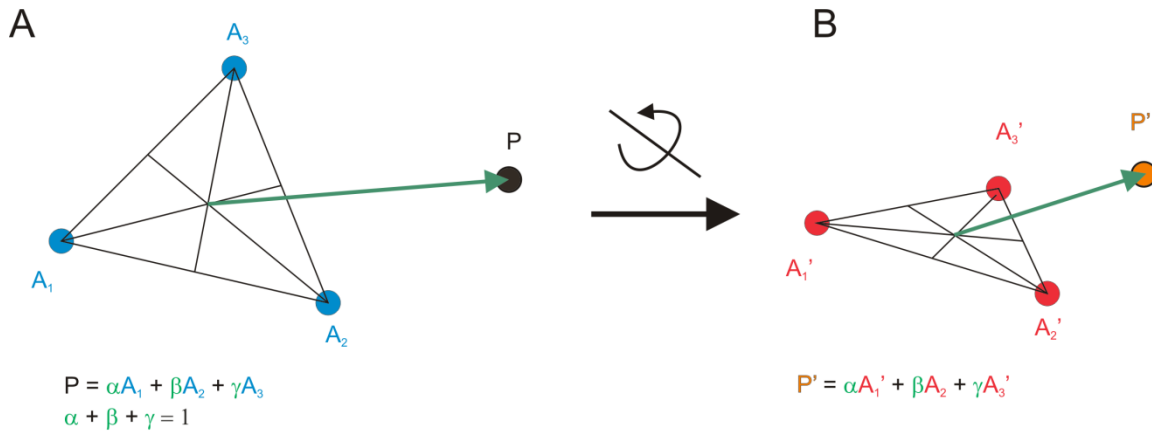
Figure 14 Use of homogenous barycentric coordinates to estimate tilt pairs. Point sets $A = \{A_1A_2A_3P\}$ and $B = \{A'_1A'_2A'_3P'\}$ are related by an affine transformation (arrow). If $A_1A_2A_3P$ and $A'_1A'_2A'$ are known, P' can be predicted by calculating homogenous barycentric coordinates α, β and γ from $A_1A_2A_3P$ according to (37), (see A), α, β and γ obtained from $A_1A_2A_3$ (see A) can be used to predict P' from $\{A'_1A'_2A'_3\}$. (see B).

If x and y denote center coordinates of a particle in EM, non-correlating errors for x,y and x',y', respectively, can be observed. These errors are mostly introduced in software-assisted particle selection where selection algorithms fail to deliver accurate results due to low SNR in the image. This coordinate error, combined with additional topology of the EM specimen may result in least square fitting of (36) with false or suboptimal parameters. Even when applied to a small, fully overlapping subset of center coordinates, the approach to determine parameters of (36) fails to deliver an accuracy that is necessary to predict particle center coordinates in the tilted image from their zero-tilt counterparts reliably.

Instead of trying to solve (36), the MaverickTilt software uses homogeneous barycentric coordinates (HBC). Considering a triangle $A_1A_2A_3$ and a point P lying in the plane of the triangle, P can be expressed as

$$P = \alpha A_1 + \beta A_2 + \gamma A_3 \tag{37}$$

where

$$\alpha + \beta + \gamma = 1 \tag{38}$$

α, β and γ are called homogenous barycentric coordinates of P with respect to A_1, A_2, A_3 and can be uniquely obtained. The HBCs are affine invariant, i.e. preserved under any

combination of affine transformations. If point sets $A_1A_2A_3P$ and $A'_1A'_2A'_3P'$ are related by the same affine transformation, P' can be determined by

$$P' = \alpha A'_1 + \beta A'_2 + \gamma A'_3 \tag{39}$$

if $A_1A_2A_3P$ and $A'_1A'_2A'_3$ are known since α,β,γ can be determined from $A_1A_2A_3P$ (4). Thus, the position of any point on the tilted micrograph can be determined from position of its affine partner on the untilted micrograph if the coordinates of three other affine point pairs are known. The use of HBCs for the determination of a point in an affine point set is illustrated in Figure 14.

For the estimation of a point in the tilted micrograph, a point set of minimum three points has to be known from which their affine partners are known so that $A_1 \rightarrow A'_1$, $A_2 \rightarrow A'_2$, $A_3 \rightarrow A'_3$ are projected by the same underlying affine transformation. The set of points from which the affine partners are known is referred to hereafter as reference point set.

2.5.2   Determination of the initial reference point set

For the determination of an initial reference point set $A_1A_2A_3$ and $A'_1A'_2A'_3$, respectively, MaverickTilt supports two operational modes: The reference point set can be either determined by the user (manual referencing mode) or be determined automatically (auto referencing mode). For the auto referencing mode, the program first determines a set of HBCs relative to every point for both the untilted and the tilted micrograph. For each point, the HBCs of the N points with the smallest Euclidian distance to the selected point are determined using the centroids (barycenters) of three triangles built from the selected point and the four points with the smallest Euclidian distance. Triangle centroids are affine invariant, thus their HBCs are preserved. In this way, every coordinate point is represented by a discrete point set in which each point is represented by its HBC α and β (γ can be obtained from α and β according to (38) and is thus not needed in the further representation). Given its new coordinates α and β, every point is represented within an affine invariant point set. Being represented in the homogenous barycentric coordinate space, point sets from both the untilted and the tilted micrograph are transposed to an isometric feature space. Thus, they can be compared irrespective of their affine transformation.

For the comparison of point sets, the *Hausdorff distance* is widely used. Given two point sets $A = \{a_1, a_2,..., a_m\}$ and $B = \{b_1, b_2,..., b_n\}$, the Hausdorff distance is defined as

$$H(A, B) = \max\left(h(A, B), h(B, A)\right) \tag{40}$$

where

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \tag{41}$$

(see also Figure 15). For the comparison of point sets with occlusions and outliers, i.e. in partially overlapping point sets, the *modified Hausdorff distance* has been proposed [171], in which $h(A, B)$ is defined as

$$h(A, B) = \frac{1}{m} \sum_{a \in A} d(a, B) \tag{42}$$

where

$$d(a, B) = \min_{b \in B} \|a - b\| \tag{43}$$

In order to include correspondence information of the point sets *A* and *B*, the *enhanced Hausdorff distance* was implemented [172]. Here,

$$h(A, B) = \frac{1}{m - \psi} \sum_{a \in A} d(a, B) \tag{44}$$

where ψ is the number of cases in which more than one point of *A* has the smallest Euclidian distance to a distinct point in *B*. *d* in (44) is obtained according to (43).

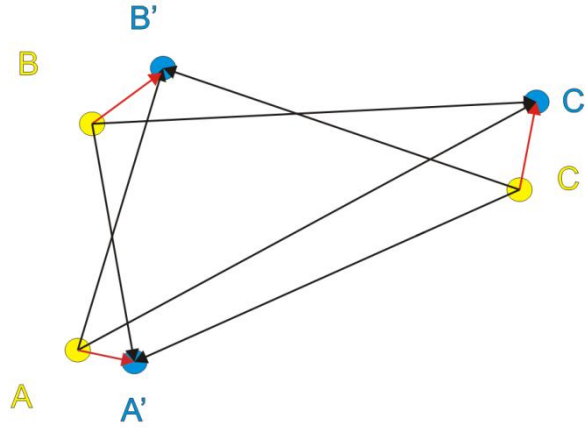**Figure 15 The Hausdorff distance.** When comparing point sets *P = {A, B, C}* (yellow) and *P' = {A'B'C'}* (blue) the Hausdorff distance is the sum of the shortest distances of any point in *P* to any point in *P'*, marked with a red arrow. In this example, the Hausdorff distance corresponds to

$$h(A, B) = \|AA'\| + \|BB'\| + \|CC'\|.$$

.

For obtaining an initial reference set of point pairs correlated by the same affine transformation, the *enhanced* Hausdorff distances of all point sets from the zero-tilt image comprising every point and the N closest points are calculated against all point sets in the tilted image. All points are represented by their HBCs α_n and β_n. For each point set from the untilted micrograph, the point set from the tilted micrograph with the smallest enhanced Hausdorff distance is stored (see Figure 16).

A Calculation of closest centroids and deriving barycentric coordinates

untilted micrograph

$A_1 = \alpha_1 C_1 + \beta_1 C_2 + \gamma_1 C_3$ where $\alpha_1 + \beta_1 + \gamma_1 = 1$
$A_2 = \alpha_2 C_1 + \beta_2 C_2 + \gamma_2 C_3$ where $\alpha_2 + \beta_2 + \gamma_2 = 1$
$A_3 = \alpha_3 C_1 + \beta_3 C_2 + \gamma_3 C_3$ where $\alpha_3 + \beta_3 + \gamma_3 = 1$
$P = \alpha_4 C_1 + \beta_4 C_2 + \gamma_4 C_3$ where $\alpha_4 + \beta_4 + \gamma_4 = 1$

Affine invariant coordinate representation:

$A_1^{aff} = (\alpha_1, \beta_1)$
$A_2^{aff} = (\alpha_2, \beta_2)$
$A_3^{aff} = (\alpha_3, \beta_3)$
$P = (\alpha_4, \beta_4)$

tilted micrograph

$A_1' = \alpha'_1 C_1' + \beta'_1 C_2' + \gamma'_1 C_3'$ where $\alpha'_1 + \beta'_1 + \gamma'_1 = 1$
$A_2' = \alpha'_2 C_1' + \beta'_2 C_2' + \gamma'_2 C_3'$ where $\alpha'_2 + \beta'_2 + \gamma'_2 = 1$
$A_3' = \alpha'_3 C_1' + \beta'_3 C_2' + \gamma'_3 C_3'$ where $\alpha'_3 + \beta'_3 + \gamma'_3 = 1$
$P' = \alpha'_4 C_1' + \beta'_4 C_2' + \gamma'_4 C_3'$ where $\alpha'_4 + \beta'_4 + \gamma'_4 = 1$

Affine invariant coordinate representation:

$A_1'^{aff} = (\alpha'_1, \beta'_1)$
$A_2'^{aff} = (\alpha'_2, \beta'_2)$
$A_3'^{aff} = (\alpha'_3, \beta'_3)$
$P'^{aff} = (\alpha'_4, \beta'_4)$

B Calculation of Hausdorff distance between affine invariant point sets

Hausdorff distance $= \left| A_1^{aff} A_1'^{aff} \right| + \left| A_2^{aff} A_2'^{aff} \right| + \left| A_3^{aff} A_3'^{aff} \right| + \left| P^{aff} P'^{aff} \right|$
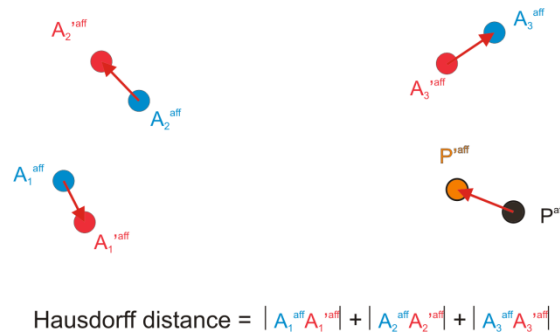
Figure 16 Determination of the initial reference point set. (A) Every point in each micrograph is represented as an affine invariant point set. The point set is derived by constructing the centroids of triangles constructed with the closest points (four in this example) and determining the HBCs of every point with respect to the triangle centroids. Each point is represented by its HBCs α_n and β_n in a new affine invariant coordinate system. (B) The Hausdorff distance between affine invariance point sets is calculated from each point on the untilted micrograph P versus each point from the tilted micrograph P', and the best result for each point is stored. Additionally, the pairing information $A_n \rightarrow A'_m$ obtained from calculating the best distances in the Hausorff distance $A_n^{aff} \rightarrow A'_m{}^{aff}$ is stored.

As a result of the calculation of the enhanced Hausdorff distance, for every point and the N closest points on the untilted micrograph, potential matches on the tilted micrograph are proposed. Since every point is theoretically sampled with N-fold redundancy, correct matches are likely to be proposed more often by the comparison of neighbouring points. Every point set is scored according to how often proposed matches within the point set are also proposed by the comparison of other point sets. In a user-defined way, point sets being proposed minimally O times are used for further processing. To exclude false positives, for every remaining point set, a single iteration of the routine for searching surrounding affine pairs (see chapter 2.5.2) is started. The closest points of the point set for which most affine pairs could be correlated will be chosen as the initial reference set.

2.5.3   Iterative detection of tilt pairs

With a given initial set of affine reference pairs, a fuzzy centroid of the point set in the untilted micrograph with points $(x_1, x_2, ..., x_k)$ is calculated by

$$C_{fuzzy} = \frac{\frac{1}{d_1}x_1 + \frac{1}{d_2}x_2 + \cdots + \frac{1}{d_k}x_k}{k} \tag{45}$$

where

$$d_n = \|x_n - C\| \tag{46}$$

and

$$C = \frac{x_1 + x_2 + \cdots + x_k}{k} \tag{47}$$

The distinct unknown point P on the untilted micrograph which is closest to c_{fuzzy} is selected. Using three known points $A_1A_2A_3$ closest to P, the HBCs of P are calculated according to (37). By using the corresponding affine partners $A´_1A´_2A´_3P$ in the tilted micrograph, the homogenous barycentric coordinates are used to predict the position of the target point in the tilted micrograph (see Figure 14). The point with the smallest Euclidian distance to the predicted point is chosen and its homogenous HBCs with respect to the reference points used for its projection are calculated. The prediction error Err_D (Euclidian distance between predicted and actual point) as well as the barycentric error Err_B (square difference between homogenous barycentric coordinates used for prediction and those by which the actual point is represented) are calculated. If the errors lie within a predefined range, the retrieved affine pair is stored. The process is repeated until all points in the untilted micrograph are processed.

Certain parameters which have not been numerically specified in the text can be user-defined during the process, these parameters are summarized in Table 9. The overall workflow of the MaverickTilt software is illustrated in
Figure **17**.

Name of variable	Description
N	Number of points sampled to define an affine invariant point set for every point in the automated search for initial reference points
0_T	Threshold for minimal number of occurences for the proposal of an affine pair for further processing
Err_B	Threshold for square difference between homogenous barycentric coordinates used for prediction and those by which the actual point is represented. The error is calculated as $(\alpha'-\alpha)^2+(\beta'-\beta)^2+(\gamma'-\gamma)^2$
Err_D	Threshold for euclidian distance between predicted and actual point)

Table 9 User-defined variables in the MaverickTilt software

The accuracy of the affine pair detection is very much depending on the reference points which are used to define the homogenous barycentric coordinates. The prediction works best if the reference points are close to the target points because errors introduced by grid topology are smaller on short distances. In order to avoid false positives, MaverickTilt uses relatively strict parameters for accepting coordinates after their prediction error and barycentric error have been calculated. Thus, the number of successfully detected affine pairs drops with increasing target point distance from the centroid of the population of known point pairs. In order to increase the number of successfully correlated affine pairs, it is necessary to iterate the procedure of affine pair detection. In each iteration, the initial population of known affine pairs being able to be used as reference points is increased. The process is iterated until the number of detected affine pairs no longer increases.

The software has been implemented using Python 2.5.1 and the NumPy and SciPy packages (see Table 1).
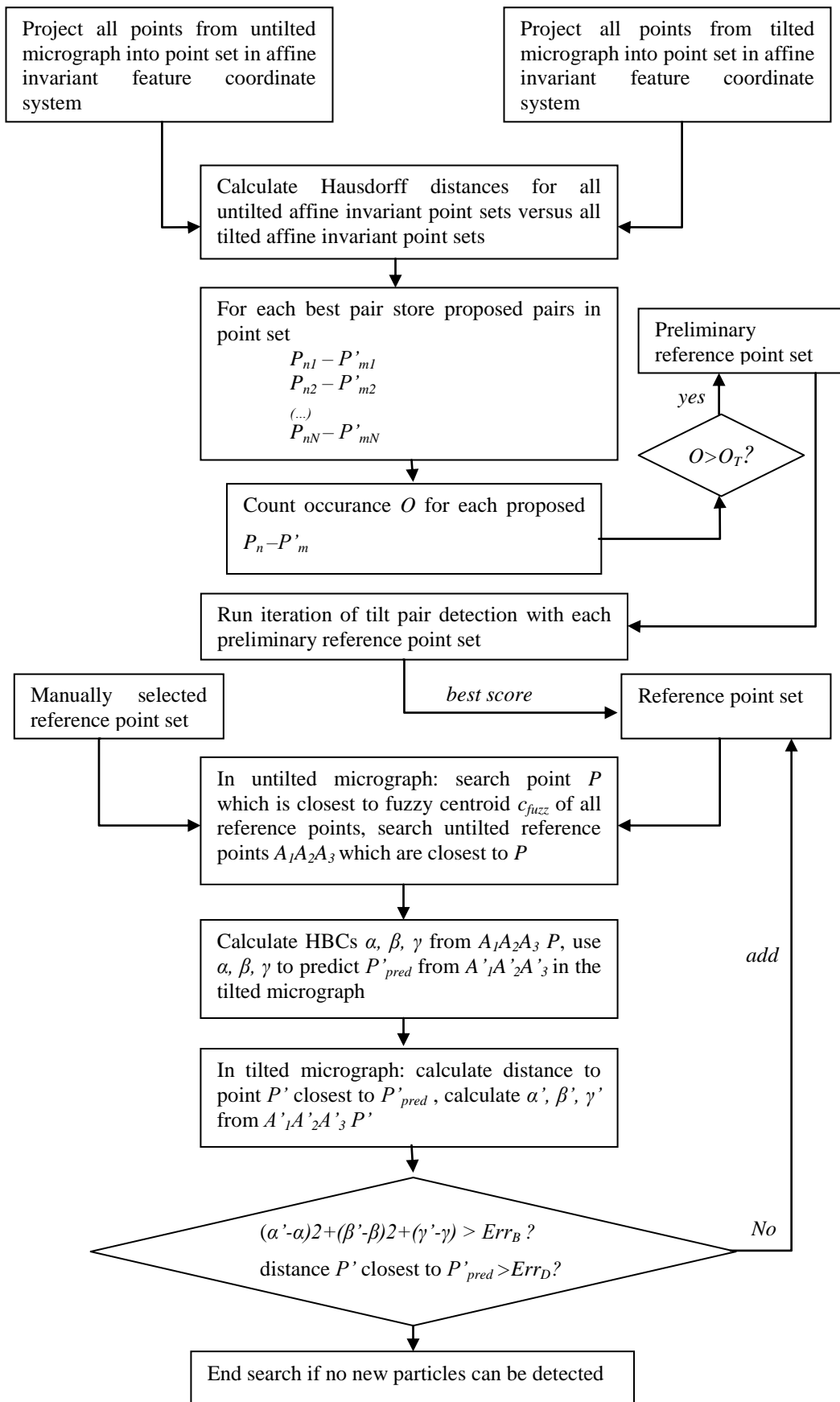
Figure 17 Work flow of the MaverickTilt software

3 Results

3.1 The MaverickTilt software

The objective of the MaverickTilt software is to correlate particles which belong to the same *tilt pair* from a micrograph of the untilted specimen and a micrograph of the specimen tilted around any number of axes by an experimental *tilt angle*. This can be done in two operational modes which can be used-defined: either in a manually referenced mode where the user specifies at least four tilt pairs manually (referred to hereafter as the *manual* mode) or in a fully automatic mode in which no reference pairs are given (referred to hereafter as *auto* mode). As input, MaverickTilt requires center coordinates of particles in each untilted and tilted micrograph in the PLT file format, which can be obtained from the standard output of a number of semi-automated particle selection programs [167, 168]. For the manual mode, the user may provide reference tilt pairs which are stored in the XML format supported by Mario Lüttich´s cowPicker software. MaverickTilt will output coordinates of *tilt pairs* in the XML format readable by the cowPicker software and convertible to the PLT file format. For convenience, the MaverickTilt software was designed to operate on all PLT files in a specified folder which have a given root file name that has to be different for coordinate files from the tilted and untilted micrographs but same root suffix number for corresponding tilt pair coordinates.

3.1.1 Benchmarking

To assess accuracy, yield and speed of the MaverickTilt program, center coordinates from tilt pair micrographs which were manually selected were used as a reference data set. Center coordinates from both untilted and tilted micrographs were converted to the PLT format in a way that the location of center-coordinates is the same in PLT files for the untilted and tilted micrograph. This information is irrelevant to the MaverickTilt software but can be used for automatic supervision of correctness of predicted tilt pairs. In total, 43 reference point sets were used for evaluation, containing from 10 to nearly 1300 particle center co-orrdinates.

Several sets of parameters for user-defined variables N, 0_T , Err_B, Err_D (see Table 9) were assessed for accuracy and yield, where accuracy was given the highest priority.

After experimental parameter optimization, values for N, 0_T , Err_B, and Err_D were defined as stated in Table 10.

Name of variable	*Value after optimization*
N	70
0_T	4
Err_B	0.75
Err_D	2*difference between best and second best estimate for P'

Table 10 Optimization of user-defined variables in the MaverickTilt software

3.1.1.1   Benchmarking using automated acquisition of initial reference point sets

The results of the benchmark test of the software running in the auto mode with optimized parameters are listed in Table 11. On average, the MaverickTilt software automatically detects 99.70 % (σ = 1.71 %) of the initial reference point sets correctly, the subsequent average of correctly predicted tilt pairs is 99.76% (σ = 1.09 %). Even though parameters are optimized for correct prediction of tilt pairs and not for retrieval of a high number of particles, the software captures on average 69.24% of all tilt pairs present in the sample. Test set #28 (being marked in red in Table 11) shows a distinctively bad performance. Strikingly, the MaverickTilt software is able to retrieve only 2.89% of tilt pairs present in the micrograph. The retrieval of such few tilt pairs will result in a very small output file. A low number of retrieved particles (i.e. small size of output file) allows automatic recognition of datasets on which the MaverickTilt software did not perform optimally (outliers). Excluding the single outlier, the statistics on identification of correct initial reference point sets further improve. Here, the MaverickTilt software correctly determines the initial reference point sets with a probability of 99.96% (σ = 0.29%), failing for a correct correlation of all points in a single case (see Table 11). The average percentage of correctly predicted tilt pairs from the whole dataset improves to 99.92% (σ = 0.19%), indicating the stability of the algorithm even when the initial reference point set is not fully correctly predicted. For processing of all test datasets, the MaverickTilt software required 7h 29 min on a single Intel® Xeon® CPU E5345 (2.33 GHz), thus requiring 10 min 26 sec for the calculation of a dataset on average.

Number of test set	Number of particle coordinates contained in set	Percentage of correctly predicted initial reference set	Percentage of correctly correlated tilt pairs	Percentage of retrieved particles
1	314	100.00	100.00	46.50
2	552	100.00	99.75	73.55
3	801	100.00	100.00	50.19
4	746	100.00	100.00	54.83
5	451	100.00	100.00	72.51
6	714	100.00	100.00	52.66
7	402	100.00	100.00	62.44
8	451	100.00	99.62	58.09
9	193	100.00	100.00	69.95
10	317	100.00	100.00	74.76
11	953	100.00	100.00	58.66
12	1067	100.00	99.01	56.61
13	196	100.00	100.00	81.63
14	159	100.00	100.00	76.73
15	145	100.00	100.00	76.55
16	160	100.00	100.00	78.13
17	253	100.00	100.00	52.17
18	190	100.00	100.00	65.79
19	10	100.00	100.00	60.00
20	282	98.11	99.53	75.89
21	83	100.00	100.00	46.99
22	222	100.00	100.00	61.71
23	247	100.00	100.00	80.57
24	266	100.00	99.53	80.83
25	319	100.00	100.00	81.50
26	270	100.00	100.00	66.30
27	228	100.00	100.00	67.11
28	485	88.89	92.86	2.89
29	374	100.00	100.00	81.02
30	341	100.00	100.00	77.13
31	404	100.00	100.00	76.49
32	355	100.00	100.00	70.42
33	283	100.00	100.00	84.10
34	279	100.00	100.00	83.51
35	332	100.00	100.00	85.84
36	366	100.00	100.00	78.14
37	377	100.00	100.00	76.66
38	1144	100.00	100.00	82.60
39	1067	100.00	99.88	78.35
40	1037	100.00	100.00	83.51
41	1124	100.00	100.00	86.21
42	1282	100.00	99.71	79.80
43	637	100.00	99.77	68.13
Average		**99.70**	**99.76**	**69.24**
Average w/o outlier		**99.96**	**99.92**	**70.82**

Table 11 Evaluation of the MaverickTilt software performance in the auto mode on noise-free test data

3.1.1.2   Identification of false positives

In its current implementation of the *auto* mode, the program initiates with a correlation of four tilt pairs which have the best score in the initial search for start-up reference pairs. Since this score is not evaluated globally, potentially false positives might be selected. In order to evaluate the program´s susceptibility for the selection of false positives, the software was forced to compare each untilted micrograph point set with an unrelated tilted micrograph point set. The MaverickTilt software was run in the parameter configuration stated in Table 10).

On average, the MaverickTilt software correlates an amount of 1.65% (σ = 0.95%) of sample points contained in the input data set as false positives (for details, see Appendix 2). This corresponds to an average of 6 particles per dataset (σ = 4). In the worst cases, the MaverickTilt software false positively correlated 4.82% of the dataset (corresponding to the initially retrieved reference point set of 4 points within the respective dataset) or an amount of 29 false positive tilt pairs (corresponding to 3.04% of the dataset). In this case, false positives can be effectively excluded from further processes by permitting only retrieved datasets which contain more than 5% of the input data and at least 30 particles. This value serves as an initial suggestion, however, datasets of even a slightly larger size will not be of any further use and will be discarded in any case.

3.1.1.3   Benchmarking using manually selected initial reference point sets

In order to test the *manual* mode, a set of eight tilt pairs which are closest to the fuzzy centroid (45) of all particles in the untilted micrograph coordinate set were extracted. This set is believed to represent typical behaviour of the user to manually select tilt pairs which are located in the center of the micrograph in proximity of the densest particle cloud.

Additionally to eight correlated tilt pairs per dataset as initial input, parameter Err_B was relaxed to from 0.75 to 3 in order to increase the number of retrieved particles in the first iterations of the algorithm (compare Figure 17, Table 10). All other parameters remained unchanged. Starting with eight correct pairs, the Maverick recovered on average 62.5% (σ = 37.48%) of the input dataset with an average accuracy of 99.74% (σ = 0.84%, for details, see). Compared to the results of the benchmarking test for the auto mode, these results seem to be inferior. The comparably poor performance of the *manual* mode might be due to the experimental setup: because the auto referencing mode searches for an optimal starting point by running one iteration of the correlation algorithm, the test reference sets have been obtained using an optimizing methodology. The ability of a human user to

identify good starting points by manually selecting particles from an area with little contamination and overlapping particles might not have been faithfully reproduced in our approach. This could also explain why a relaxation of Err_B to a value of 3 was necessary to retrieve a satisfying amount of particles. Remarkably, the *manual* mode captures over 87% of all particles present in the input dataset (compared with the single best retrieval rate of 86,21% for the auto mode, see Table 11) with an accuracy of 100% for more than 40% of all datasets used in the benchmark test (for detail, see Appendix 1). The population statistics, however, suffer from numerous cases where either a low percentage of input data is correlated or where the accuracy of correlation is lower than 100%.

3.1.1.4 Performance on noisy data

In all previously described benchmark tests, data sets which contain only manually selected pairs which represent a fully correlating data set have been used. Being manually selected, particle coordinates can be presumed to be well-centered. Additionally, the datasets are fully overlapping, i.e. for each particle coordinate in the untilted micrograph, there is an affine partner in the tilted micrograph. In practice, this ideal representation of the input coordinates cannot be expected. As a result of semi-automated acquisition of particle center coordinates [167, 168], the determined coordinates deviate from the true center of the particle. Thus, errors for the particle coordinates which are uncorrelated in the untilted and tilted micrograph can be expected. This *de-centering noise* is represented by adding or subtracting a value according to the noise from the x- and y-coordinates of the particle center. Since datasets are non-overlapping, each micrograph will contain a distinct number of particles to which no affine partner exists. This scenario becomes even more severe when recording tilt pairs under unstained cryo conditions: when tilting the micrograph, the contrast of contaminations will increase, making them more susceptible to being detected only on the tilted micrograph by semi-automated or automated particle selection routines. Particle coordinates which do not have an affine partner or are representing contaminations are referred to hereafter as *noise coordinates.*

To assess the performance of the MaverickTilt software under conditions of noise being present, the noise-free data was overlaid independantly with two kinds of noise. *De-centering noise* was simulated by adding uniformly distributed noise to the particle center coordinates. The noise was represented by adding or subtracting a random value in the range (0, n_{cent}) from each x- and y-coordinate in the dataset. *Noise coordinates* were simulated by randomly generating an amount of n_{coos} as many points as in the original

dataset which were represented by x- and y-coordinates which were randomly chosen between 0 and the maximal x-and y-coordinate values from the dataset. To assess susceptibility of the MaverickTilt software for errors from de-centering noise and noise coordinates, both types of noise were assessed independently using parameters of the MaverickTilt software as stated in Table 10. Additionally, outliers were excluded by admitting only retrieved datasets which contain more than 5% of the input data and contain minimally 30 particles (see chapter 3.1.1.1). For all remaining datasets, the average percentage of correctly determined initial reference point sets ($C(init)$) as well as the average correctness of tilt pair correlation ($C(all)$) was determined. Additionally, the percentage of retrieved particles from the sum of all particles in the data set ($R(all)$) was determined as well as the percentage of false positives ($R(false)$) contained in the retrieved tilt pairs. The result of analysis are summarized in Table 12 and Table 13.

The MaverickTilt program delivers reliable results even when the center coordinates of the particles are shifted from the true center up with maximal shift of 20 pixels (see Table 12), predicting tilt pairs correctly with an accuracy of minimally 99.56%. In the range of a maximal de-centering-noise between 10 and 20, however, the recovery of particles from the dataset drops dramatically from around 40% of retrieved tilt pairs to about 10% (see Table 12 Graph B). This result is probably due to the fact that in order to exclude false positives, particle correlations which are not in agreement with threshold parameters Err_B and Err_D (see Table 10) are discarded in large amount. An increase of the recovery rate for a maximal de-centering noise of 40 pixel correlates with an increase of predicting false positives; around 2% of all recovered particles were false positives.

Performance of the MaverickTilt program drops more dramatically when introducing *noise coordinates*. After adding 10% of the initial points as random coordinates, the average rate of correct prediction drops to 99% (see Table 13). The performance in both identifying correct particles and retrieving tilt pairs further decreases. Thus, the exclusion of false positive particles should be a major concern in the acquisition of particle center coordinates.

n_{cent} (pixel)	C(init)	$\sigma_{C(init)}$	C(all)	$\sigma_{C(all)}$	R(all)	R(false)
5	99.73	1.37	99.66	1.24	57.01	0.19
10	99.71	1.33	99.84	0.39	44.80	0.21
20	99.46	2.61	99.56	0.94	11.12	0.35
40	97.38	6.69	97.03	1.88	14.15	2.35
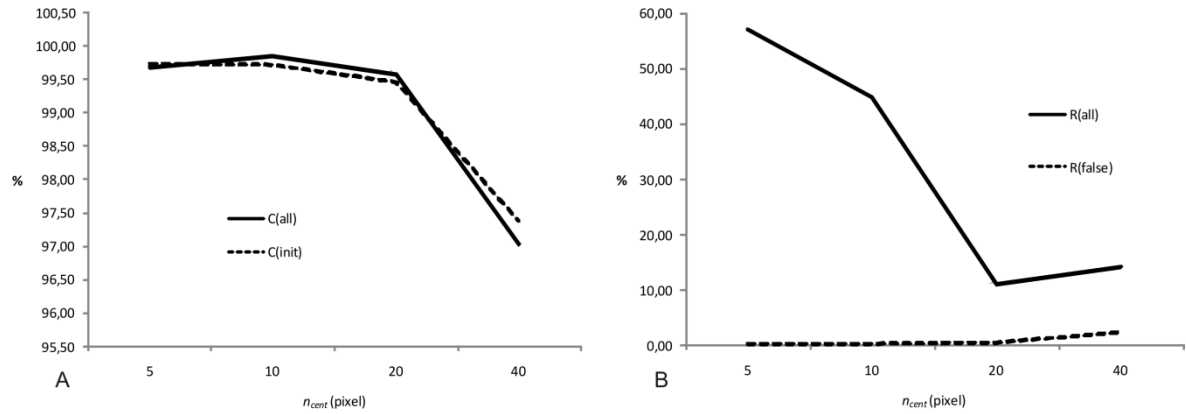
Table 12 Performance of the MaverickTilt in presence of *de-centering noise*

n_{coos}	C(init)	$\sigma_{C(init)}$	C(all)	$\sigma_{C(all)}$	R(all)	R(false)
0.1	98.50	3.78	99.00	1.51	48.92	0.87
0.2	98.68	3.92	98.89	0.94	39.37	1.10
0.4	97.38	6.69	97.03	1.88	14.15	2.35
0.6	94.10	9.25	95.80	2.43	11.70	3.72

Table 13 Performance of the MaverickTilt in presence of *noise coordinates*

3.2 The 50S Ribosomal subunit of *Thermotoga maritima*

For the examination of structural heterogeneity within the large ribosomal subunit of *Thermotoga maritima*, Local 2D MSA analysis was done followed by a subsequent MSA and classification in 3D space (see Figure 6, chapters 1.3.2.2, 1.3.2.6). Additionally, the high structural integrity observed in ribosomal complexes suggested a reconstruction using conventional approaches in which the complete dataset was merged into a single reconstruction. As a result of conventional reconstruction, we obtained a three-dimensional structure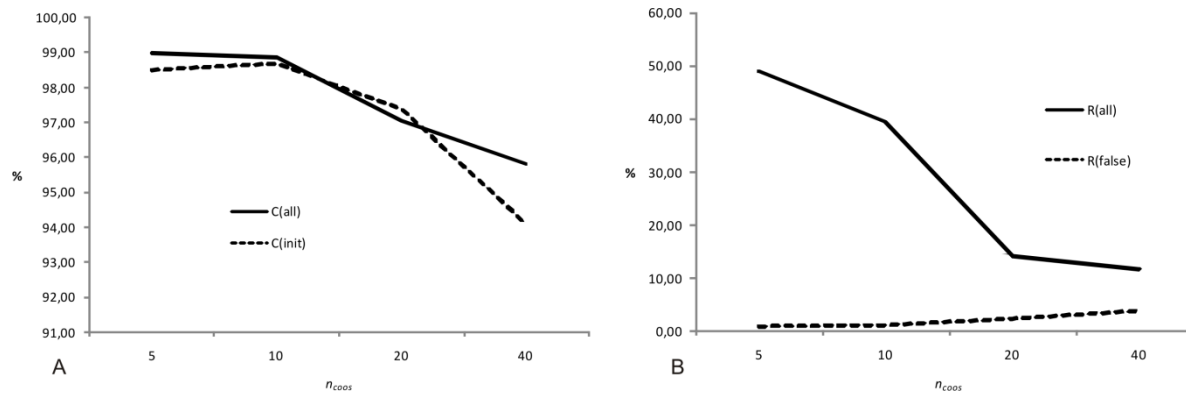 of the large ribosomal subunit of the hyperthermophilic eubacterium *Thermotoga maritima* at a resolution of 17 Å (0.5 FSC). In order to compare the structure of the *T. maritima* 50S RSU to known structures of the large ribosomal subunit of other bacteria, multiple sequence alignment of the 23S and 16S rRNA of extremphilic and mesophilic archaea and bacteria was done, yielding the hyperthermophilic eubacterium *Thermos thermophilus* as the evolutionary closest relative of all compared organisms in terms of sequence similarity of the 23S rRNA. The crystal structure of the *Thermus thermophilus* large ribosomal subunit [42] was manually docked into the cryo-EM map (see Figure 20 A). In the cryo-EM reconstruction, an extended ribosomal L7/L12 stalk becomes visible as well additional densities which can be attributed to three expansion segments (see Figure 20).

3.2.1 The L7/L12 stalk of T. maritima

The ribosomal L7/L12 stalk of *T. maritima* has been previously reported to comprise six copies of the L7/L12 protein [43, 50]. In both reconstructions of 50S RSUs, regardless of whether they were prepared with or without a carbon support film on the cryo grid, six N-terminal domains of the L7/L12 protein can be accommodated in the cryo-EM density accounting for the ribosomal stalk (Figure 18).

Nevertheless, different preferential orientations of the large ribosomal subunit can be observed between 3D reconstructions obtained from particles either prepared on a carbon support film or in vitreous ice over holes on a holey carbon grid. A preferred orientation of the 50S RSU in the crown view is presumably adopted by a preferential binding of the 50S RSU to the carbon support film via the relatively plane intersubunit region (Figure 19).

This preferred orientation further limits the overall flexibility of the ribosomal stalk in the plane perpendicular to the intersubunit region, a reason for which we chose the particles prepared on carbon support film for further analysis towards the visualization of the C-terminal domains of the ribosomal stalk.

The C-terminal domains of the ribosomal stalk are highly flexible [43], and thus absent in structures which are determined by techniques that depend on averaging of the whole dataset. In order to resolve flexible states of the C-termini, the stalk region had to be analyzed separately. This was done by submitting images aligning to the same projections of the final 3D to a classification in which only the stalk region was considered [173]. Locally classified projection averages were randomly combined for the reconstruction of 2500 3D volumes. For the reconstruction of each 3D, we used 18 randomly chosen projection
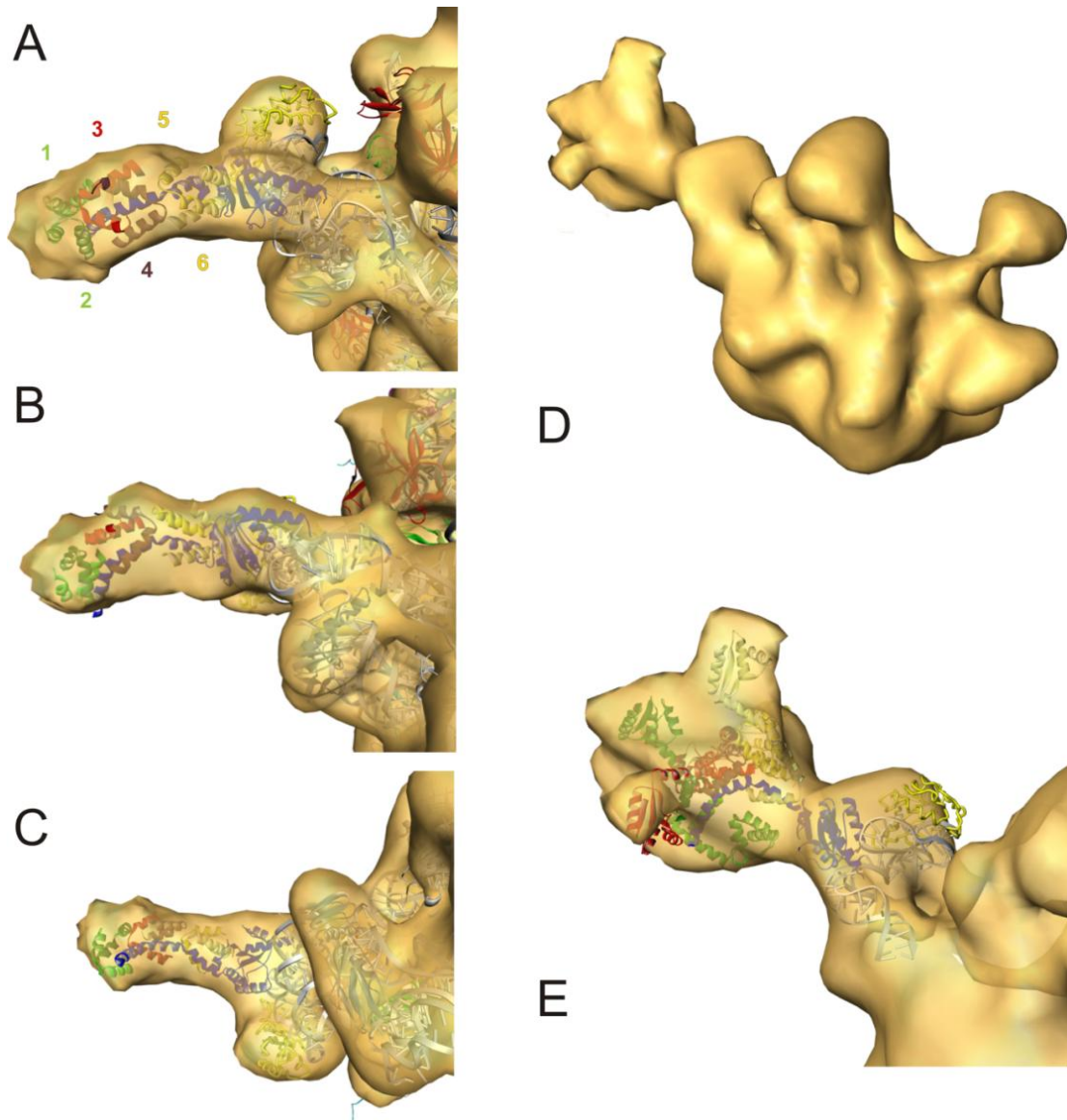
Figure 18 L7/L12 stalk of Thermotoga maritima comprising six copies of L7/L12. The crystal structure of the Thermotoga maritima L10- L7/L12 complex with a stoichiometry of 1:6 (L10:L7/L12) can be accomodated in the cryo-EM reconstruction accounting for the ribosomal stalk (A, B, C). To retrieve additional structural information in the area of the C-terminal domains of L7/L12 which are not resolved in the crystal structure, a large set of 3Ds was reconstructed from random sets of projection averages which were separated by local MSA of the stalk region. Subsequently, classification of all reconstructions was done by 3D-MSA. In (D), a representative 3D class average is shown, revealing significantly more density in the stalk region. This additional density provides enough space to accomodate the C-terminal domains of six copies of L7/L12 (E).

Figure 19 Reconstruction and angular orientation of the T. maritima 50S RSU in cryo in vitreous ice and on carbon support film. Crown view representation of the 50S RSU of T. maritima reconstructed from cryo-EM on carbon support film (A) and in vitreous ice (B). Angular distribution of projections in the overall particle population of particles recorded 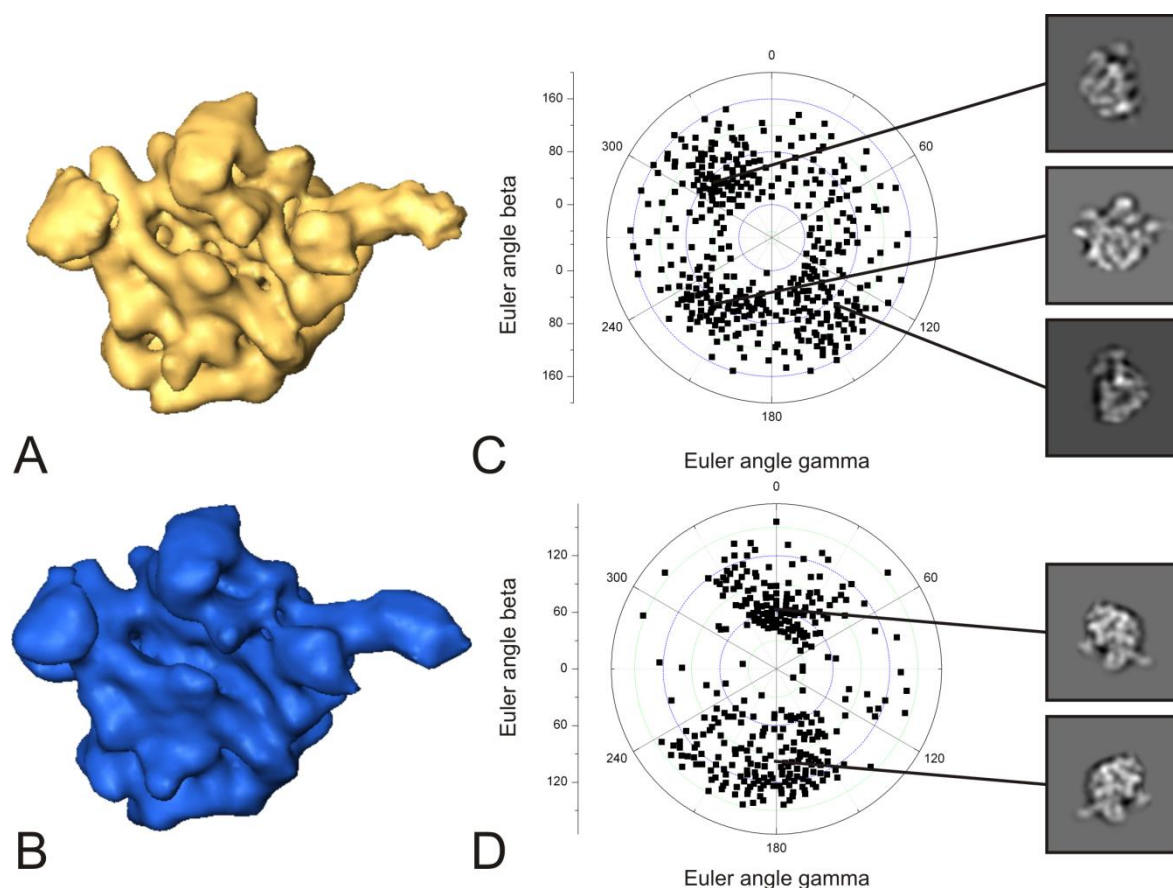under cryo conditions on carbon support film (C) and over holes (D). Projection averages on the right linked to centers of clouds in the graphs depicting the angular orientation represent preferred orientations in the respective datasets.

averages. 3Ds were submitted to MSA and subsequent classification in 3D space (see Figure 11, chapter 1.3.2.6). 3D class averages in which the signal-to-noise ratio was improved in the averaging process contained approx. 50 averaged 3Ds. In one of the obtained 3D volumes, an extended density of the stalk region could be observed. Within the observed additional density, the crystal structure of 6 c-terminal domains of L7/L12 (PDB ID: 1DD4, [174]) could be accommodated by manual docking, revealing the overall volume of the full-length L7/L12 stalk and indicating possible conformations of the c-termini of L7/L12 (Figure 18 D, E). In the observed volume, the c-termini of L7/L12 are almost symmetrically distributed in their orbital radius.

3.2.2   Expansion segments of the *T. maritima* large ribosomal subunit

In the cryo-EM reconstruction of the *T. maritima* 50S RSU, three expansion segments (ES) become visible (Figure 20). These ES have not been reported in any bacterial structure of the large ribosomal subunit.

23S rRNA sequences contributing to the expansion segments in *T. maritima* were identified by alignment of the 23S rRNA sequences of *T. maritima* and *T. thermophilus*. When comparing the secondary structures of the 23S rRNAs of *T. thermophilus* and *T.maritima*, the sequences identified in the previous alignment can be located at the expansion segments visible in secondary structure models of the 23S rRNA of *T. maritima* [175] (see Figure 21). Furthermore, the border regions of the expansion segments can be located in the crystal structure of the *T. thermophilus* large ribosomal subunit, and all of them could be located at the corresponding expansion segments of the overlayed cryo-EM reconstruction of the *T. maritima* 50S RSU (Figure 20). Based on the location of the sequences contributing to the expansion segments, the sequences were grouped into expansion segments (ES) 1, 2 and 3 (Figure 20). To test whether the corresponding 23S rRNA sequences of ES1, ES2 and ES3 may account for the extra densities observed in the cryo-EM reconstructions, we simulated secondary structures of the expansion segments identified by alignment and modeled 3D volumes of the calculated helical regions.
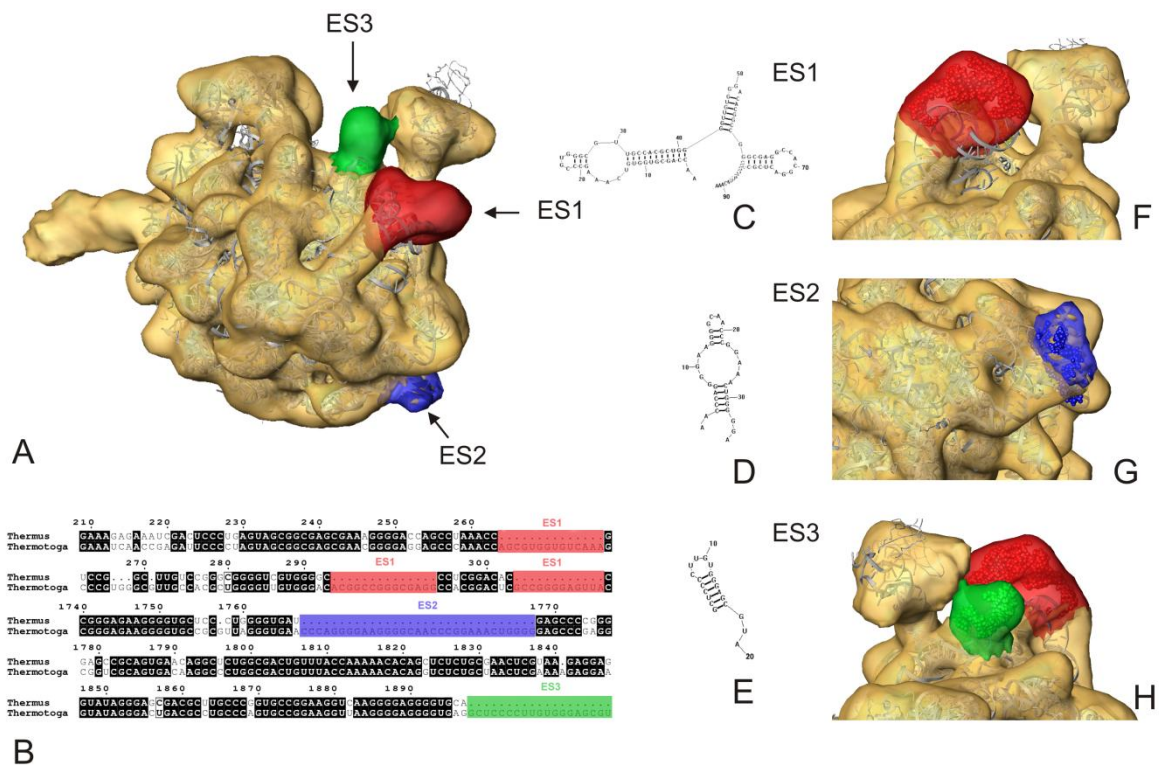
Figure 20 Three expansion segments within the 23S rRNA of Thermotoga maritima as compared to Thermus thermophilus can be identified. Expansion segments (ES) of the 23S rRNA of Thermotoga maritima have been identified by comparing the 3D cryo-EM reconstruction of the 50S ribosomal subunit ofThermotoga maritima with the crystal structure of the 50S ribosomal subunit of Thermus thermophilus (A). Furthermore, the sequences of the 23S rRNA of both species have been compared by an alignment using ClustalW (B). Sequences neighbouring the expansion segments in the alignment were identical to sequences bordering the expansion segments found in the three-dimensional structure. Secondary structure models of the rRNA covering all expansion segments were built for ES1 (red, figure C), ES2 (blue, figure D) and ES3 (green, figure E) using the RNAstructure 4.5 software. Space-filling models of RNA helices identified in the expansion segments were modeled in 3D using the assemble0.3 software. Space-filling RNA models were manually docked into cryo-EM densities accounting for ES 1 (F), ES 2 (G) and ES 3 (H).
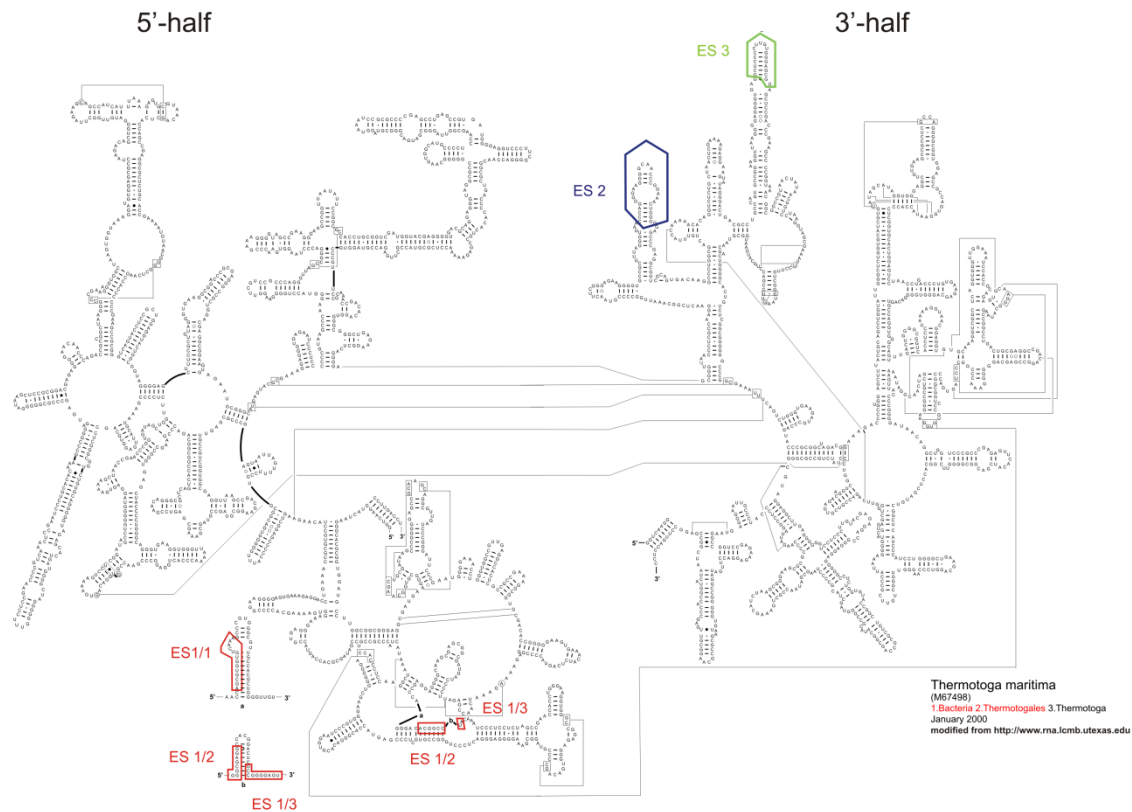
Figure 21 Secondary structure diagram of the large ribosomal subunit of *T. maritima*. 23S rRNA expansion segments compared to the 23S rRNA as identified by ClustalW alignment. Expansion segment (ES) 1 is labeled in red, ES2 is labeled in blue and ES3 is labeled in green. Modified from http://www.rna.lcmb.utexas.edu [175].

Since ES1 is spanning three regions of homologous sequence in *T. thermophilus*, the sequences which were unique in *T. maritima*

were concatenated in order to construct a space-filling RNA model of ES1. Manual docking of the space-filling models of the 23S rRNA expansions sequences shows that these rRNA sequences are likely to account for the additional densities observed, confirming the expansion segments in the large ribosomal subunit of *T. maritima* to be comprised of additional RNA elements (summarized in Figure 20)

3.2.3   Flexibility of the L1 stalk in *T. maritima*

The ribosomal L1 stalk is known to be one of the most flexible parts of the ribosome [176, 177]. This flexibility is innately linked to the biological function of L1 to assist in the translocation of tRNA as previously suggested [178].

To assess its flexibility in the 50S RSU of a hyperthermophilic organism in which decreased flexibility due to increased structural rigidity can be expected [162], 3D MSA as described for the L7/L12 stalk was used in order to identify possible significant conformational
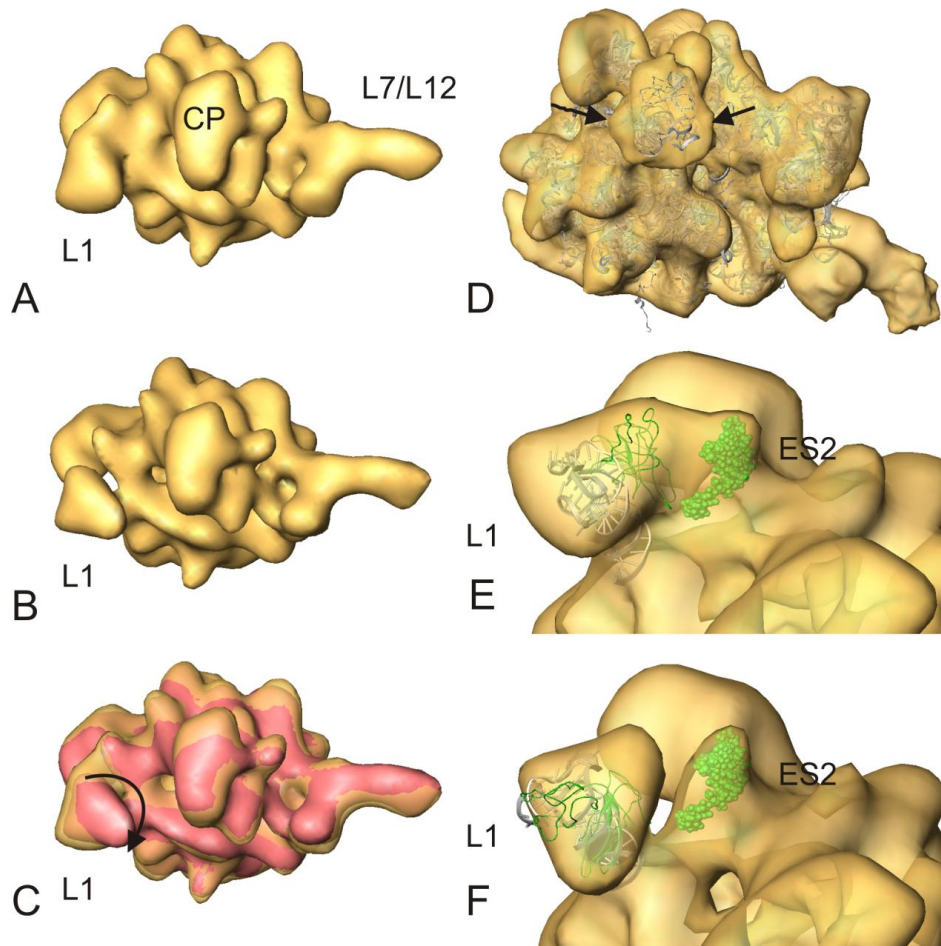
Figure 22 Flexibility of the L1 stalk in *T. maritima*. CP: central protuberance, ES2: Expansion segment 2. From 3D-MSA of the L1 region as described for the L7/L12 stalk, two distinct orientations can be seen as shown in (A) and (B). When overlaying the two different conformations, it can be seen that the density originally located at expansion segment 3 flaps towards the peptidyl transferase center (C). In the averaged overall structure of the 50S RSU, additional densities of L1 when comparing to the crystal structure of L1 from *T. thermophilus* can be observed as indicated by the arrows (D). These additional densities are most probably attributed to unresolved flexibility of L1 in the averaged overall structure. It can be seen that L1 alternatively makes contact (E) or not (F) with ES3 depending on the adopted conformation of L1.

sub-states. Mainly two different conformations of L1 were observed (Figure 22). In these two conformations, the density of L1 flaps towards the peptidyl transferase center (Figure 22 C). Similar motions of L1 have been described previously [48, 176, 178-180]. When comparing the density representing L1 in the final structure which was reconstructed by averaging all conformational sub-states, excess densities compared to the L1 crystal structure of L1 from *T. thermophilus* become visible (Figure 22D). These excess densities disappear when resolving the conformational sub-states of L1 in *T. maritima* (Figure 22 E, F). It is further noticeable that L1 binds to ES3 in the conformation which is distal to the petidyl transferase center.

3.3   Translation initiation factor 3 (eIF3) from *Saccharomyces cerevisiae*

3.3.1   GraFix preparation of eIF3 complexes

Biochemical purification eIF3 from *Saccharomyces cerevisiae* was optimized for high purity and yield and included a gel filtration purification step. Treatment of purified eIF3 complexes with the GraFix protocol [5] was a prerequisite for obtaining particles suitable for EM analysis. In order to further stabilize eIF3 complexes prior to preparation of EM grids, GraFix-treated samples were incubated for 24 hours in order to allow additional cross-linking. The effect of GraFix treatment on purified eIF3 complexes is demonstrated in Figure 23. Even after incubating GraFix-treated samples for 24 hours, no noticeable aggregate formation due to unspecific inter-molecule cross-linking was observed (see Figure 23 B).
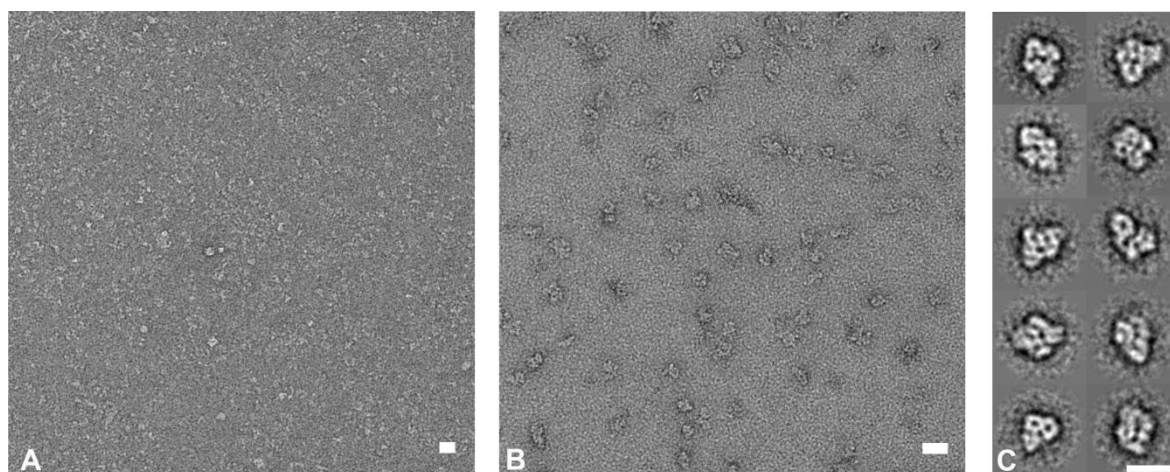
Figure 23 Effect of GraFix treatment of eIF3 after purification. A) EM micrograph of negatively stained eIF3 directly after purification. Fractionated, structurally heterogeneous particles are prevalent. B) EM micrograph of negatively stained eIF3 after GraFix treatment. Complexes are intact and homogenous in size. C) Class averages of eIF3 selected from untilted micrographs of GraFix-treated sample. The complexes are integer and well-defined, indicating structural homogeneity of complexes. Scale bar : 20 nm.

3.3.2   RCT analysis of negatively stained eIF3 complexes

Tilt series of negatively stained eIF3 complexes were recorded in order to obtain bias-free initial references for refinement. Upon tilting the specimen in the electron microscope, shadow-like features appeared on the complexes at the side of the tilt direction (see Figure 25 C). This phenomenon is referred to as the *shadowing effect*. Usually, negative stain is evenly distributed around the specimen, accumulating at the borders of the specimen (see Figure 24 A). When the specimen is tilted, the area of exposure is reduced for the stain accumulation at the border of the particle facing opposite the tilt direction. Upon interaction with the sample, the electron beam has to pass a thicker layer of stain at this

position, resulting in reduced transmission at the same position (see Figure 25 B). Thus, a shadow-like feature can be observed particle borders facing opposite of the tilt direction.

This shadowing effect will become more severe if the particles have been only weakly stained. In this case, the amount of stain on the particle border facing the tilt direction may be insufficient to compete with the shadowing effect on the opposite side in terms of contrast formation, leading to a projection image which is weakly defined on the side of the tilt direction and sharply defined on the opposite side. This unilateral contrast strongly impairs subsequent alignment and reconstruction procedures. High-pass filtering of tilted single particle images can help to suppress the artifacts introduced by the shadowing effect, but runs at the cost of suppressing information in the low spatial frequency domain, i.e. the overall shape of the molecule. The consequences of the shadowing effect and high pass filtering are demonstrated in Figure 25.
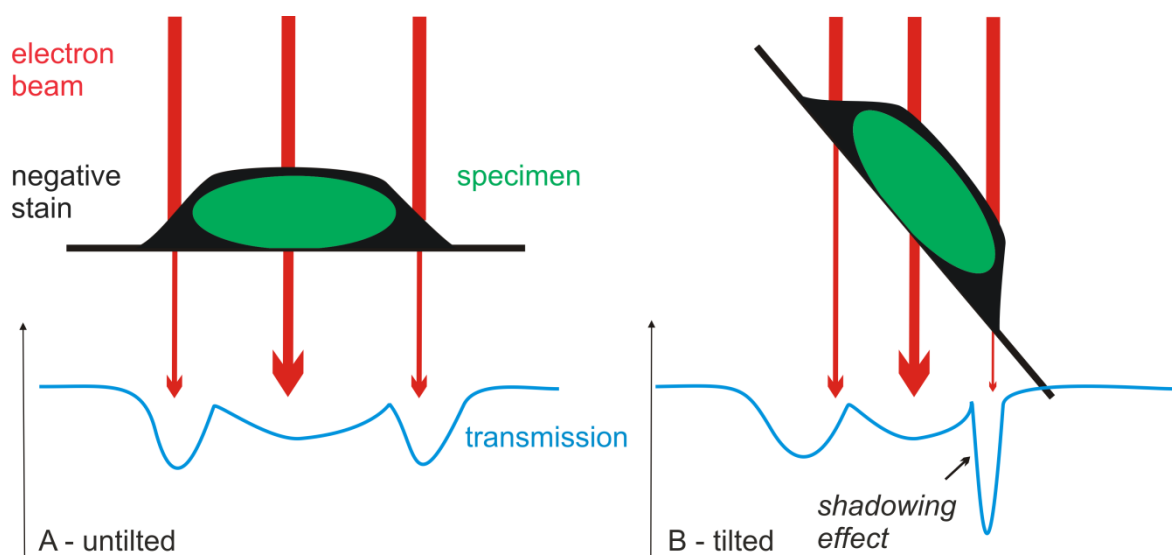
Figure 24 Unilateral shadowing effect in tilt exposures of negatively stained specimen. A) Under untilted exposure conditions, stain accumulations on the border of the specimen result in equivalent reduction of electron transmission. B) When tilting the specimen, the area of exposure is condensed for the stain accumulation at the border which faces opposite the tilt direction. The electron beam has to pass a thicker layer of stain, thus, transmission of electrons is unilaterally reduced.

After band-pass filtering single particle images of eIF3 recorded from a tilted micrograph, shadowing effects could be observed for the majority of particles (see Figure 25 A). Since all images have been tilted in the same direction, subsequent RCT reconstructions were dominated by a 'shadow cone' on one side of the particle. This feature became even more prominent after averaging all RCT volume after 3D maximum likelihood alignment [39]. In the average three-dimensional volume, the mere feature of the reconstruction is a hollow
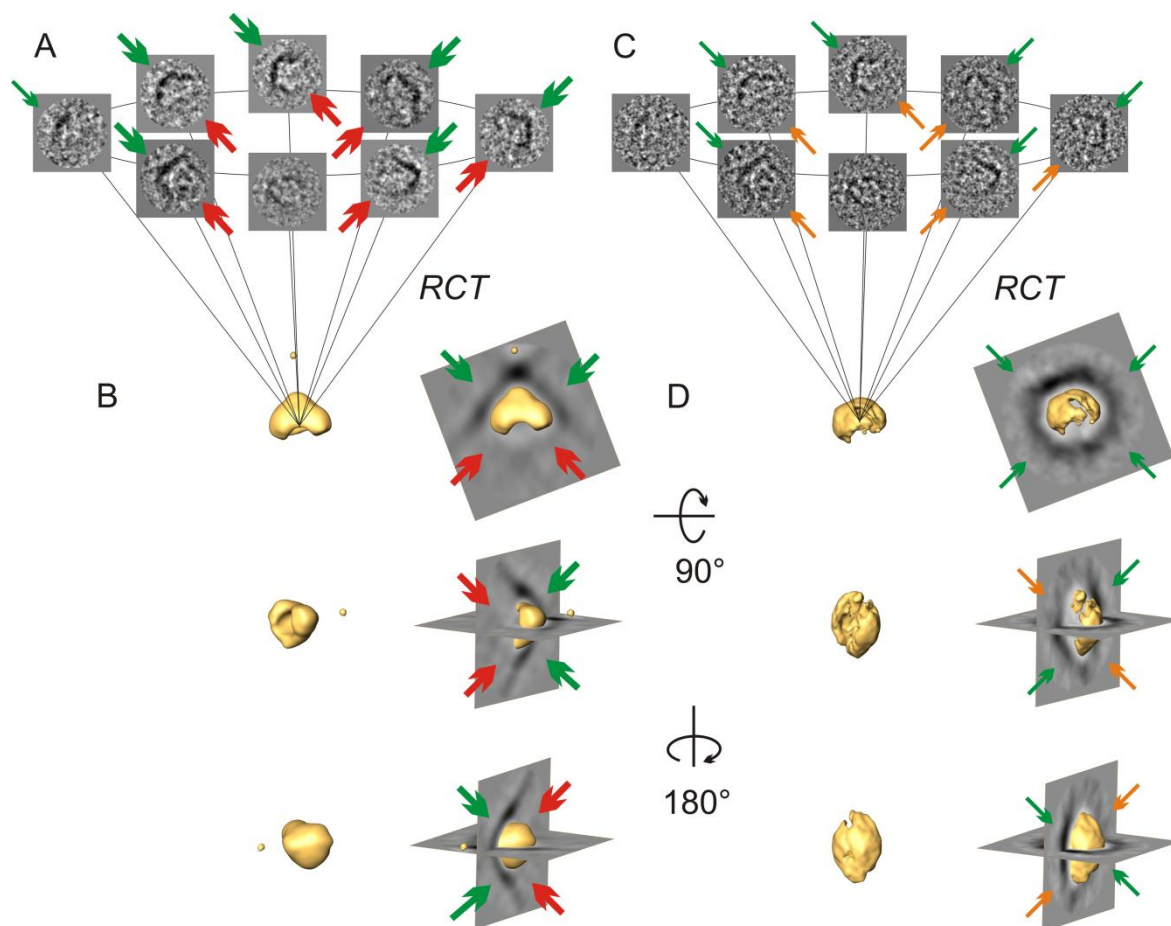
Figure 25 Effect of negative stain and filtering on random conical tilt reconstructions. A) Band-pass filtered images of tilted eIF3 complexes show a strong shadowing effect leading to strong contrast at one side of the particle (indicated by green arrows) and low contrast at the opposing side (indicated by red arrows), leading to a unilateral definition of the particle boundaries. B) The average of all RCT reconstructions after 3DML alignment show a good definition on one side of the reconstructed 3D (green arrow) and a poor definition on the opposite side (red arrow). The presence of a shadowing cone in the average impairs 3DML alignment, resulting in a almost featureless average which represents the unilateral contrast in the single particle images. C) High-pass filtered images of the tilted eIF3 complex. The shadowing effect has been reduced upon high pass filtering, decreasing the contrast difference between well-defined particle borders (green arrows), and the opposite borders being less well defined (orange arrows). D) In the average RCT reconstruction obtained after 3DML alignment, the particle borders are better defined than in the reconstruction from the band-pass filtered image. The loss of information in the low spatial frequency domain is reflected by under-representation of prominent low-resolution features which contain information about the particle shape. Instead, fine details corresponding to the higher spatial frequency domain are overrepresented.

cone which is well defined at one side (see green arrow in Figure 25B) and contains almost no information on the opposite side (see red arrows in Figure 25B).

The observed three-dimensional volume features represent the three-dimensional reconstruction error introduced by shadowing effects. By applying a high-pass filter, the shadowing effect can be reduced (see Figure 25 C). Since coarse image features which define the overall shape of the molecule are represented in the same low spatial frequency domain, information about the overall particle shape is reduced as well. High-pass filtering of images for the reconstruction of RCT 3Ds resulted in three-dimensional volumes which

were more evenly defined. In the average random conical tilt reconstruction obtained after 3D maximum likelihood alignment, the effect of shadowing is significantly reduced, with the borders of the particle being defined in all directions (see Figure 25D).
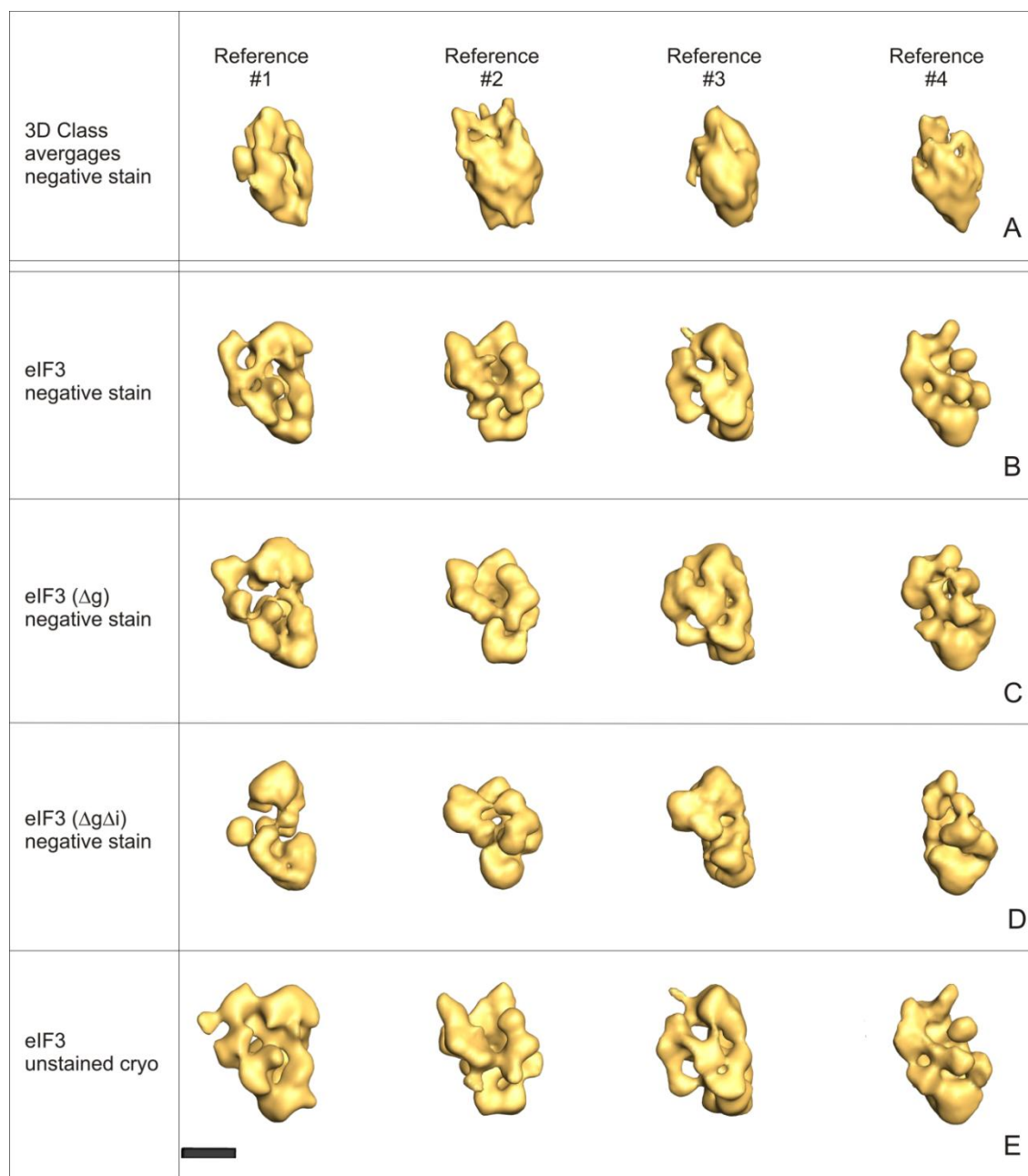
Figure 26 Refinement of eIF3 and eIF3 sub-complexes. (A) Selected 3D class averages used as initial references. Initial references for refinement of eIF3 complexes were obtained from RCT reconstructions of high-pass filtered tilted images. 3D class averages were calculated after 3DML alignment, 3DMSA and classification. (B) 3D volumes reconstructed from untilted negative stain images of eIF3. 3D volumes shown in (B) are the result of supervised classification, each volume represents the refined model of the reference above in the same column. (C), (D) Refined models from negative stain datasets of eIF3(Δg) and eIF3(ΔgΔi), respectively. 3D volumes in (C) and (D) are results of supervised classification, using refined models reconstructed from the unstained negative stain dataset (B) as initial references. Each volume represents the refined model of the reference above in the same column. (E) Reconstruction of eIF3 from an unstained cryo dataset. Reconstructions from (B) were used as initial models for supervised classification. Scale bar: 10 nm.

3.3.3   Refinement of three-dimensional reconstructions

3D class averages of RCT 3D reconstructions after 3D maximum-likelihood alignment were used as references for refinement of the same dataset. In each 3D class average, around 10 RCT 3D volumes were averaged. For refinement of 3Ds from untilted negative data, images were treated by MSA and subsequently classified [29] after the first round of alignment during supervised classification (see chapter 1.3.2.1). Subsequently, angular orientation of class averages was determined using angular reconstitution [181]

As references for angular reconstitution, three-dimensional volumes were reconstructed from averages of all images aligning to the same reference using the euler angles of the reference image (projection matching). In this step, initial bias from the starting model is reduced. The calculated 3D volumes were used for iterative refinement by supervised classification (see chapter 1.3.2.1).

In order to determine the locations of subunits, sub-complexes comprising eIF3a,b,c,i were obtained by assembling complexes without eIF3g (eIF3(Δg)). eIF3a,b,c sub-complexes were assembled in absence of eIF3g,i (eIF3(ΔgΔi). Zero-tilt negative stain images were recorded from both sub-complexes. Both subcomplexes were refined following the protocol described for the eIF3 complexes, using the refined structures of the eIF3 complexes as initial references for supervised classification. In parallel, an unstained cryo dataset of eIF3 was refined following the same procedure. The results of iterative refinement of eIF3 and the (eIF3(Δg) and eIF3(ΔgΔi) subcomplexes are shown in Figure 26. It can be seen that all reconstructions are well-defined. Three-dimensional structures from sub-complexes resemble the reference volume, but are lacking some parts of the molecule (see Figure 26 C,D). As the number of subunit decreases in sub-complexes, so does the volume of the reconstructions, as expected (see Figure 26 transition from C to D). There is, however, a significant amount of heterogeneity among the three-dimensional volumes which are reconstructed from the same dataset (represented by rows in Figure 26). This heterogeneity is first observed in the reconstructions of negatively stained eIF3 complexes which are obtained from 3D class averages (see Figure 26 B). Since these reconstructions were used for the refinement of all other datasets, the heterogeneity propagates throughout the volumes obtained for all other complexes (see Figure 26 C-E) The structural heterogeneity could be explained either by dynamic rearrangement of subunits or by errors in the reconstitution procedures. Since the class averages which were initially obtained for the negatively stained particles obtained from untilted view of eIF3

are relatively homogeneous (see Figure 23 C), we do not expect radical rearrangements of eIF3 in absence of any other interaction partners. Furthermore, the crude structural features which would display such a large conformational flexibility are not present in the initial references (see Figure 26 A). Instead, high spatial frequencies (i.e. finer structural features) are overrepresented. To exclude the possibility that the observed structural heterogeneity is an artefact of the refinement progress, cross-validation of single reconstructions obtained from the dataset can be done.

3.3.4   Cross-validation of refined three-dimensional reconstructions

During the refinement progress, images of a dataset are sorted into classes which represent the references to which the images aligned best (supervised classification, see chapter 1.3.2.1). For cross-validation, all images which have been sorted into a group representing one reference structure are submitted to MSA and subsequent classification [29]. Class averages were used as input for the cross-validation process (compare Figure 27 I1-I4). Subsequently, angular reconstitution ([181], see chapter 1.2.4) was done with class averages using the three-dimensional reconstructions which represent all *other* classes in the dataset as a reference (see Figure 27 R1-R4). After determining the angular orientation of all class averages by angular reconstitution, a 3D structure is calculated. This 3D structure is used as a reference for another round of angular reconstitution, and a new 3D is calculated. This process is repeated several times. If datasets are consistent, each input dataset will reproduce the structure which onto which it was refined even with the wrong reference given in the first iteration of angular reconstitution. Reproduction of the reference structure, on the other hand, indicates inconsistency of datasets grouped by supervised classification and ambiguity in the assignment of angular orientations. The result of cross-validation is shown in Figure 27. It can clearly be seen that 3D volumes obtained after cross-validation represent the references rather than the input used, indicating inconsistency of the previous supervised classification. The inconsistency of the results obtained by supervised classification is most probably due to the weak definition of coarse structural features in the initial reference 3Ds (see Figure 26 A), which is, in turn, a result of high-pass filtering of tilted single particle images used for the reconstruction of initial reference 3Ds (see Figure 25 C). Thus, it is necessary to obtain new initial references for supervised classification. To obtain reliable initial models from RCT reconstructions, it is necessary to overcome the artefacts introduced by the negative stain shadowing effect which can be observed in exposures of tilted negatively stained specimen

(see Figure 24). Reliable initial models can be gained either by using a negative stain preparation with thicker stain or by recording tilt series under unstained cryo conditions.
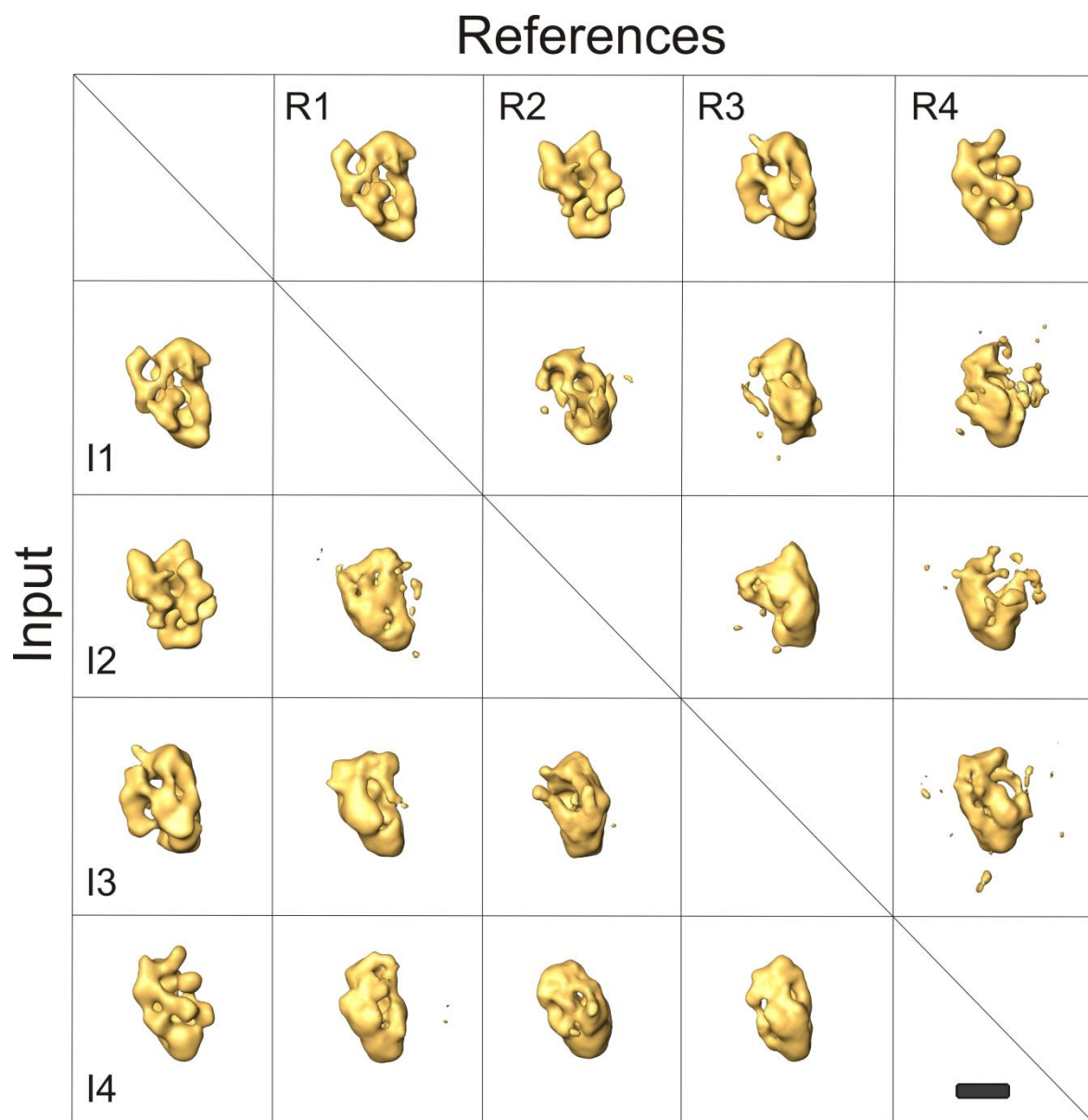
Figure 27 Cross-validation of eIF3 complexes reconstructed from untilted negative stain images. Images grouped by supervised classification are reconstructed using 3Ds of other groups as a reference. (I1-I4) As input, classes of images obtained after supervised classification were used. Each class is represented by the 3D to which images of the class aligned best. (R1-R4) For each input class, 3Ds from all other classes were used as an initial reference for 4 iterations of angular reconstitution and 3D reconstruction. Unlabeled 3D volumes show results of the cross-validation procedure, the row of each volume represents the input class, columns indicate the reference used in the first round of angular reconstitution. Scale bar: 10 nm.

3.4   The V-ATPase of *Thermus thermophilus*

The structure of the intact V-ATPase of Thermus thermophilus was reconstructed from negatively stained complexes which were treated by the GraFix protocol ([5], see chapter 1.4). For obtaining random conical tilt reconstructions (RCT 3Ds) ([36], see chapter 1.3.2.7) to determine initial references for refinement, a total of 45,256 tilt pairs were correlated from a dataset of 110,845 particles from untilted micrographs and 108,997 from tilted micrographs using the MaverickTilt software (see chapter 2.5). From this dataset, particles were sorted after MSA and classification ([29], see chapter 1.2.3) by discarding all images that were contained in classes which were aggregated or which represented disrupted complexes. After sorting, a total of 41,221 tilt pairs and 88,886 particles from untilted micrographs were used for further processing, which corresponds to a relative amount of more than 80% of intact particles. In total, 650 RCT 3Ds were calculated from class averages, using on average 63 images of tilted complexes for the reconstruction of a single RCT 3D volume. All RCT 3D volumes were submitted to 3D maximum-likelihood alignment ([39], see chapter 1.3.2.8), and subsequently classified by MSA and subsequent classification in 3D space ([29], see chapter 1.3.2.6). Class averages were reconstructed using on average 18 RCT 3Ds. 3D class averages were visually inspected, five 3D class averages displaying distinctly different features were selected as initial references for iterative refinement by supervised classification (see chapter 1.3.2.1) using all particles that were obtained from untilted micrographs. The overall refinement process is illustrated in Figure 28. After supervised classification, the dataset was split almost evenly between all reconstructions that have current resolutions ranging from 26-29 Å (0.5 FSC, [26], see chapter 1.2.6). Refinement statistics of the *Thermus thermophilus* V-ATPase are summarized in Table 14.

Reconstruction	Colour code	Number of particles used for reconstruction	percentage of complete dataset	Current resolution/Å (0.5 FSC)
1	yellow	15602	17.55	26.91
2	pink	19637	22.09	27.49
3	blue	10766	12.11	29.29
4	red	16631	18.71	27.81
5	green	26250	29.53	26.03
total		88886		

Table 14 refinement statistics of *Thermus thermophilus* V-ATPase
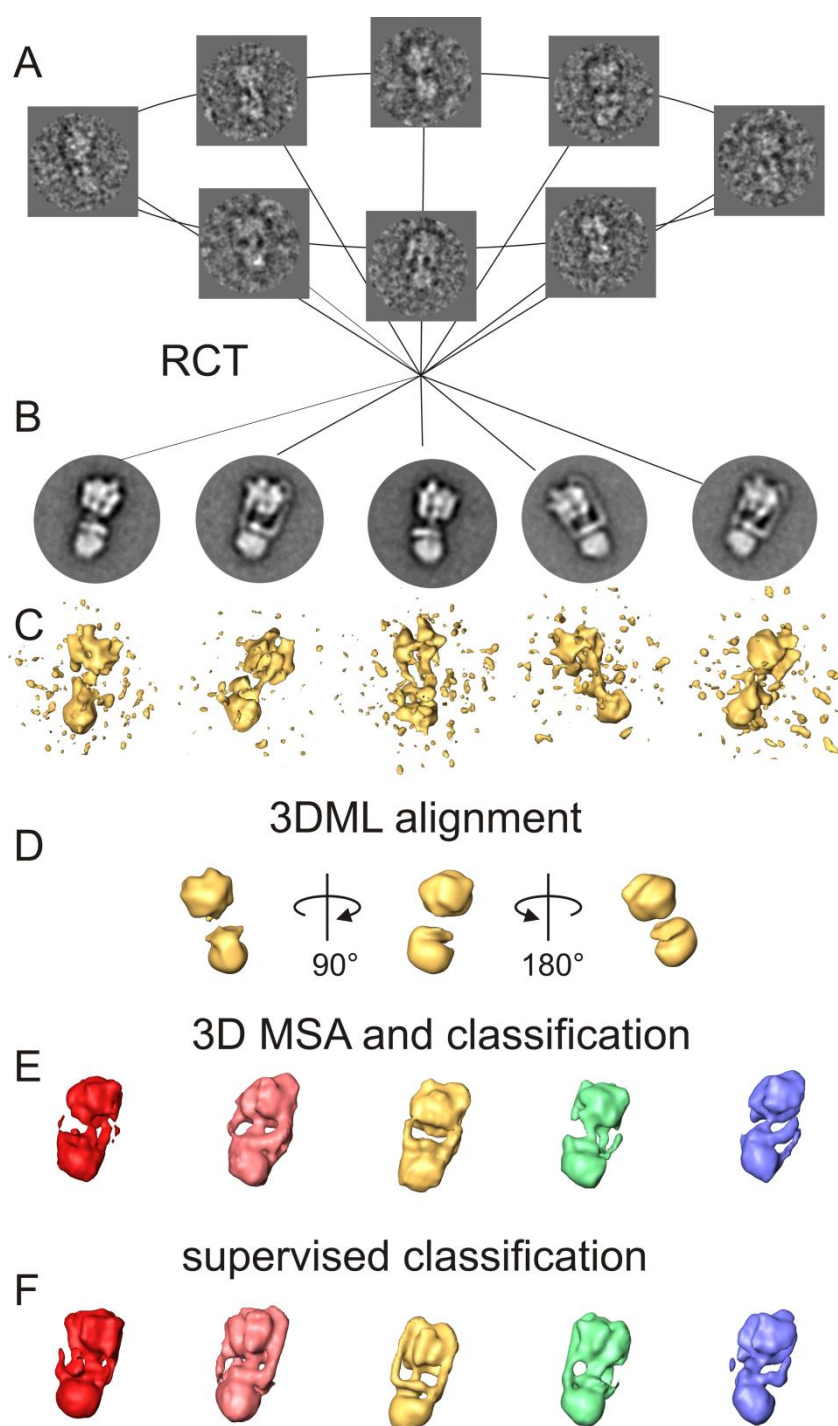
Figure 28 Refinement of V-ATPase from *Thermus thermophilus*. (A) Tilted negatively stained single particle images of the V-ATPase from Thermus thermophilus were used to reconstruct RCT 3Ds from class averages of untilted negatively stained single-particle images (B). RCT 3Ds (C) were registered by 3D maximum-likelihood (3DML) alignment. The consensus structure obtained by averaging all RCT 3Ds after 3DML alignment is shown in (C). Five 3D class averages were selected after 3D MSA and classification of the aligned RCT 3Ds (E) and used as initial references for supervised classification. Three-dimensional reconstructions obtained after supervised classification are shown in (F). Scale bar for 3D reconstructions: 10 nm.

For illustration of the location of subunits within the electron microscopy reconstruction, X-ray structures of known subunits related to the *Thermus thermophilus* vATPase have been manually docked into the EM map (see Figure 29).
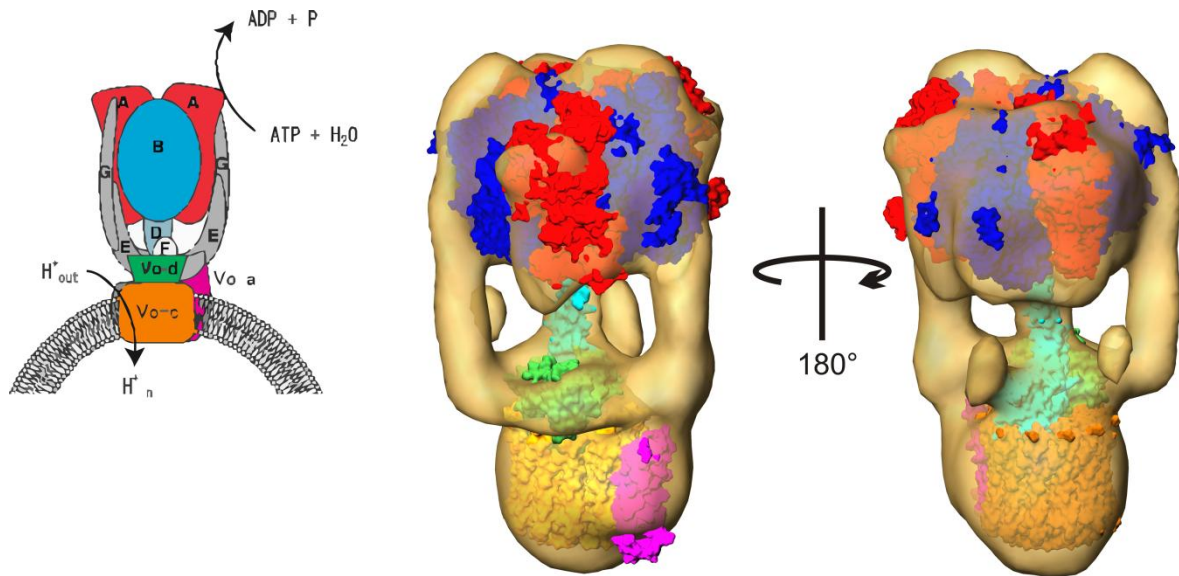
Figure 29 Illustration of subunit localization within the EM reconstruction of the Thermus thermophilus V-ATPase. Left: a schematic representation of the subunit architecture is given together with the colour code used for the docking of X-ray structures (figure modified from [137] with author´s permission). Right: manual docking of x-ray structures from ATPase subunits. Blue, *Methanosarcina mazei go1* subunit A ([182], PDB ID:2C61), red, *Pyrococcus horikoshii* subunit B (PDB ID: 2DM9), cyan, *Bos taurus* central stalk of the F(1)-ATPase, a presumed functional homologue to the *Thermus thermophilus* subunit D ([183], PDB ID: 1E79), green, *Thermus thermophilus* subunit V_0-d ([184], PDB ID 1R5Z), orange, *Escherichia coli* subunit V0-c and V0-a (magenta), ([185], PDB ID: 1C17). Docked X-ray structures do not necessarily represent actual conformational states.

From the final set of three-dimensional reconstructions, different tilt angles of the water-soluble V_1-domain versus the membrane-bound V_0-domain can be observed. Additionally, the number of observed peripheral stalks varies between one and four in all reconstructions.

3.4.1   Central cavities in the reconstruction of V-ATPases

In all final reconstructions of V-ATPases, central cavities can be observed in the V_1 head domain at sub-noise density thresholds. Visualization of these cavities in reconstructions of negatively stained complexes indicate solvent accessibility of these cavities, since the penetration with heavy metal stain is a prerequisite for the reconstruction of interior features of a macromolecular complex. Additionally, solvent-accessible cavities are observed within the membrane-bound V_0-domain when the threshold for three-dimensional visualization is lowered to a level at which not all features of the reconstruction are fully visible. The necessity of lowering the threshold for visualization of cavities within V_0 indicates a relatively small volume or weak accessibility by heavy metal stain. The observation of cavities is demonstrated in Figure 30.
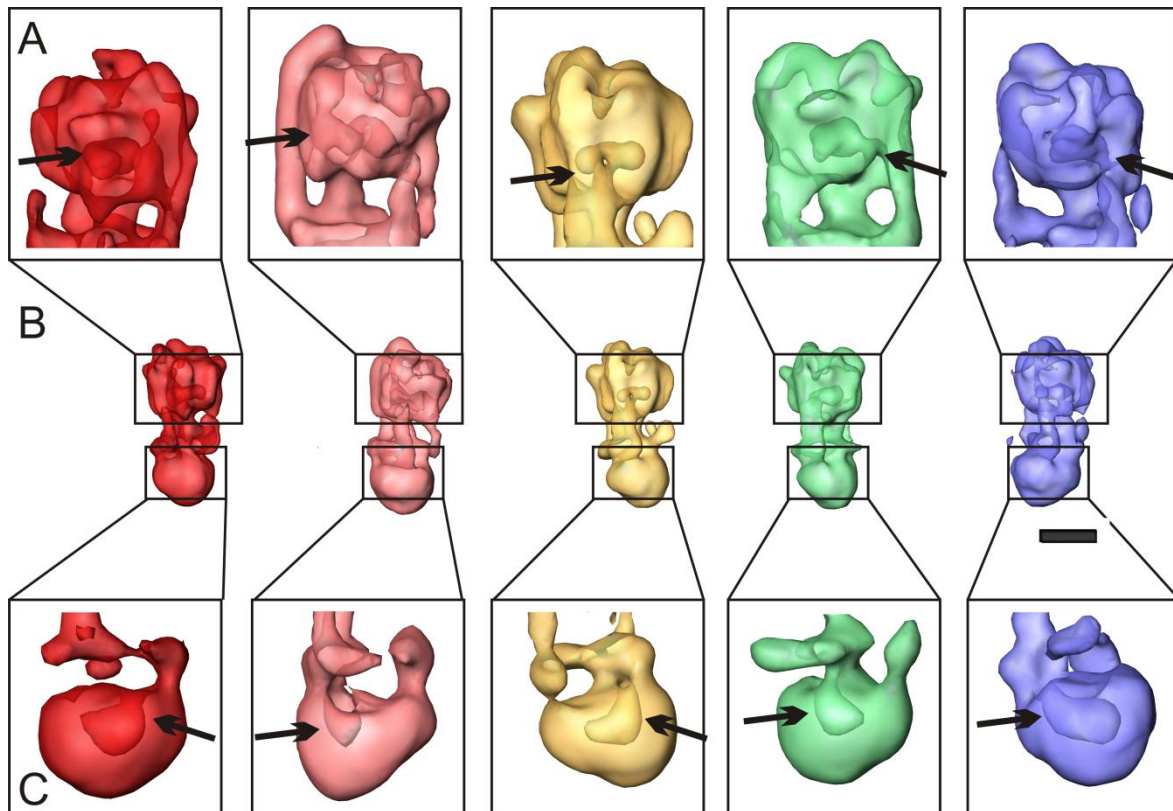
Figure 30 Internal cavities in reconstructions of V-ATPases. (A) Enlarged transparent views of the V_1-domain (B) show cavities in every reconstruction (marked by arrows). Cavities are located at the lower part of V_1, presumably at the interface of the central subunit D. A connection between the cavity in V_1 and the exterior is made by a small tunnel in the direction of a large peripheral stalk. (C) When lowering the density threshold for displaying of complexes, internal cavities become visible in the transmembrane domain V_0. Cavities are connecting to the exterior via a small tunnel towards the central stalk, reaching towards V_0-a on the other side. Scale bar in (B): 10 nm.

Internal cavities have been observed in previous reconstructions of a V-ATPase from *Saccharomyces cerevisiae*, where an internal cavity was observed in the transmembrane V_0 domain [153]. Here, however, no connection between the internal cavity and the exterior were shown. In a more recent reconstruction of the *Manduca sexta* V-ATPase, similar observations of internal cavities have been described both in V_1 and V_0 which make connections towards the peripheral stalks and the central stalk, respectively [145].

Here, we describe for the first time the observation of internal cavities which are connected to the exterior in *Thermus thermophilus*, suggesting a similar architecture with respect to anchoring of the central stalk as previously described in eukaryotes.

3.4.2 Conformational heterogeneity of stalk connections

In a previous reconstruction of the *Thermus thermophilus* V-ATPase, a total of two peripheral comprising subunits E and G stalks were observed [152]. As in other reconstructions the central stalk formed by subunits D and F extended from V_0 towards V_1

Figure 31 Observation of different stochiometry and conformation of peripheral stalks. All 3Ds in the same row represent the same view. The position of V_0-d is marked with an asterisk. Peripheral stalks are labelled with numbers 1-4, the font size of the numbering indicating their relative thickness. In (A), a total of four peripheral stalks can be observed, all stalks being relatively thin as compared to the other reconstructions. A slight kink of V_1 with respect to V_0 can be observed in (A) as well. In (E), only two r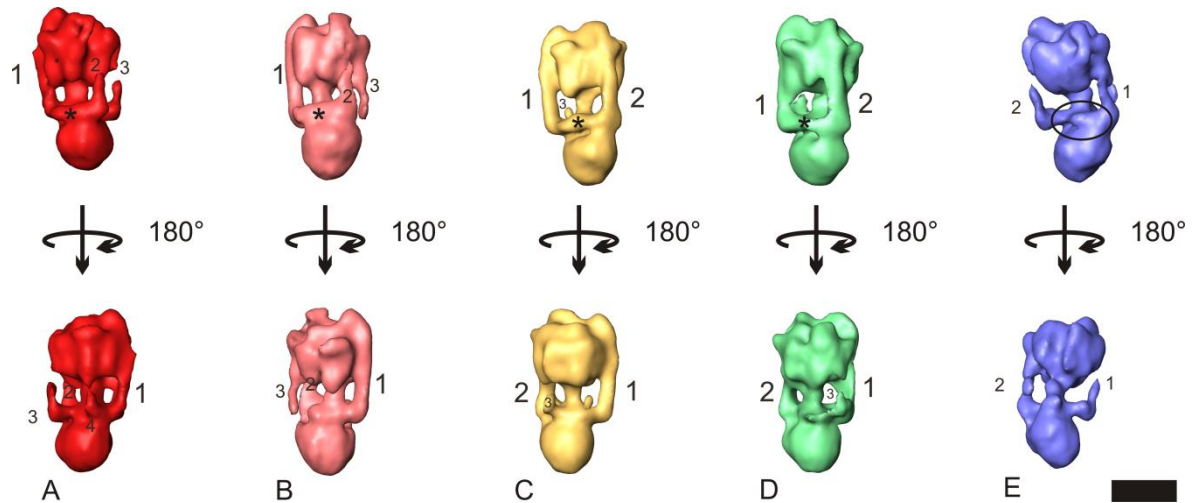elatively thin stalks are observed, together with a kinked V_1 domain. Furthermore, subunit V_0-d seems to be absent in this reconstruction (missing domain enclosed by circle) .In (B), one major stalk is observed together with two minor stalks, which are close to each other and in close proximity to subunit V_0-d. In (C) and (D), two major stalks are observed adjacent to subunit V_0-d and a very small minor stalk is observed opposite of V_0-d. Scale bar: 10 nm.

relatively perpendicular to the membrane (for example, see [145, 152, 153]). When calculating the consensus structure of all RCT 3Ds after 3D maximum-likelihood alignment, neither the central stalk nor any peripheral stalks can be visualized (see Figure 28 D). This finding indicates a high degree of structural heterogeneity and disorder for all stalks, since averaging leads to a total loss of features. In our final reconstructions obtained after supervised classification, however, we observe different numbers of peripheral stalks, ranging from a total of two to four. Remarkably, the observed peripheral stalks differ significantly in thickness, showing up to two thicker peripheral stalks and up to two thinner peripheral stalks, the connection from V_0 to V_1 is partly disrupted in the latter. Moreover, thick stalks were always observed at the same position, which is adjacent to a density attributed to subunit V_0-d (see Figure 29). The stalks with a relatively thick appearance will be referred to as *major stalks* hereafter, the smaller stalks will be termed *minor stalks* in subsequent descriptions. The observations of different numbers of peripheral stalks in different reconstructions are illustrated in Figure 31.

It is noteworthy that the presence of predominantly minor peripheral stalks correlates with a tilting of the V_1 domain versus the V_0 domain (see Figure 31 A, E). In the most severe case, the presence of only two minor stalks coincides with a massive kinking of V_1 as observed in a reconstruction from which V_0-d seems to be absent (see Figure 31 E). The

observation of correlations between presence and thickness of stalks and flexibility of the V_1 domain versus V_0 suggests examination of flexibility of V_1 in detail.

3.4.3   Flexibility of the cytosolic V_1 domain

Flexibility of the cytosolic V_1 domain of V-ATPases has been reported in previous works in which the complex was studied by EM [149, 152]. In this work, we present three-dimensional reconstructions in which flexibility of the V_1 domain is fully reproduced without impairing the three-dimensional reconstruction process. To evaluate the structural flexibility exhibited by the V_1 domain, tilt angles were measured with respect to V_0 by measuring the angle between the lower delineation of V_1 and the line along the cavity between V_0-c and V_0-d, which is believed to be parallel to the membrane. Furthermore, the angle of the central stalk versus the line running parallel to the membrane was measured. For angular measurement, each reconstruction was rotated to a view where the most extreme kinking of V_1 versus V_0 was observed. The analyses of tilting of V1 versus V_0 are summarized in Figure 32. Generally, tilting of V_0 is observed solely in the directions of the two major stalks (depicted by the two columns attached to V_1 in the center diagram of Figure 32). The observed tilt angle is $\alpha = 22°$ when one of the major stalks is disrupted (compare red and pink reconstructions in Figure 32 and Figure 31 A, B) and increases to $\alpha = 27°$ when only a single thin connection is seen at the position of a single major stalk and the other stalk is disrupted (compare blue reconstruction in Figure 32 and Figure 31 E). Thus, the reciprocal correlation of peripheral stalk integrity and flexibility of V_1 is further underlined.
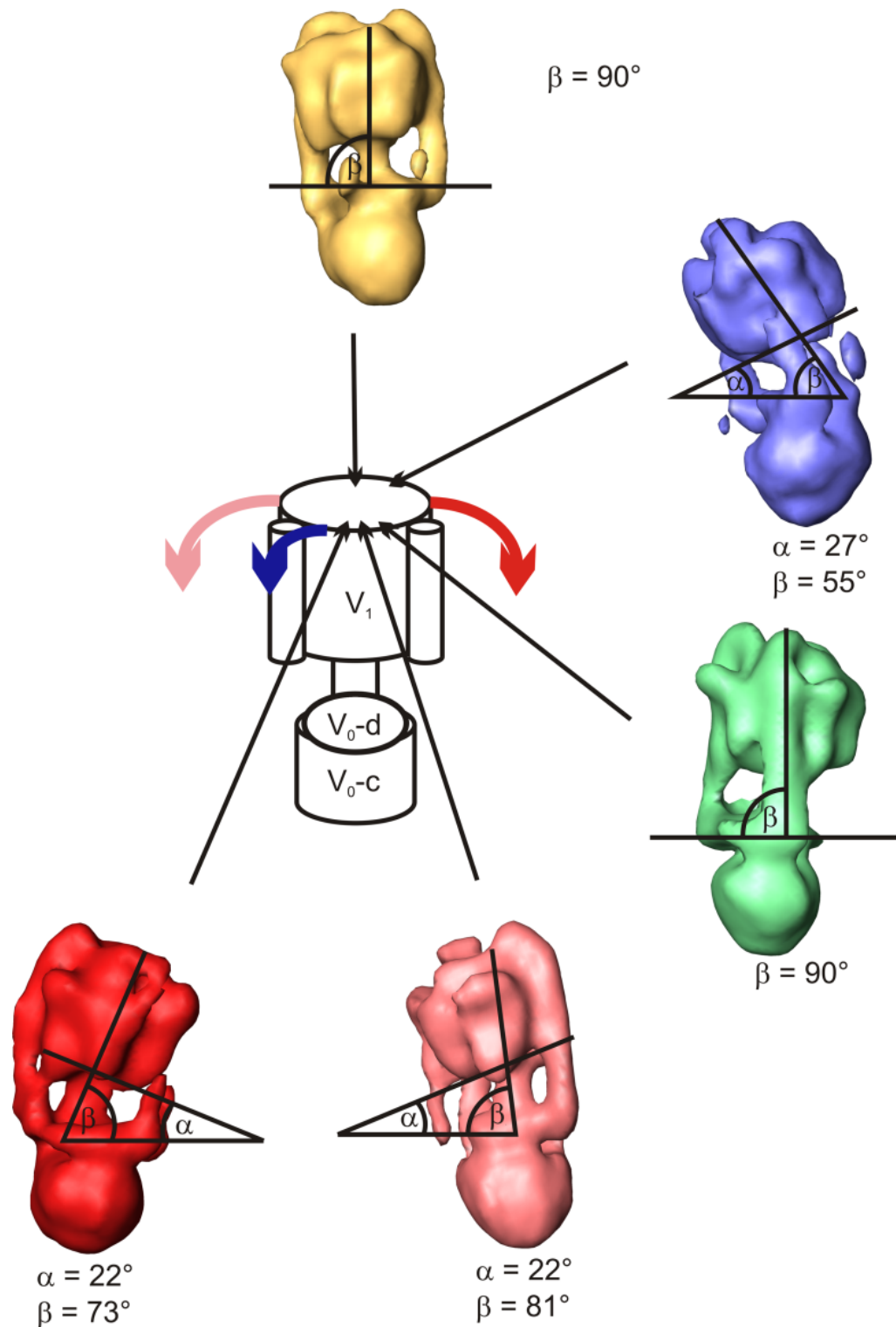
Figure 32 Measurement of tilt angles between V_1 and the central stalk versus V_0. Each reconstruction was rotated into the direction from which it displayed the most extreme tilt angle. Absolute orientation of every 3D is indicated by an arrow indicating the point of view on the complex representation in the middle. Tilt angles α indicate angles between the lower delineation of V_1 and the membrane, angles β denotes angles between the central stalk and the membrane. Coloured arrows in the central model indicate tilt directions for each reconstruction

4 Discussion

4.1 The MaverickTilt software

4.1.1 Applicability of the software

The MaverickTilt software determines correlation between particles on an untilted and a tilted micrograph based on their given center coordinates. The basic algorithm which is used by the MaverickTilt software is the transformation of all particles into an affine invariant feature space. Thus, the MaverickTilt software will correlate particles which have been projected to a correlating micrograph by any affine transformation, i.e. rotation of any angle around any number of axes, translation, scaling and shear. Although having been developed to overcome the particle selection bottleneck in the process of random conical tilt (RCT, see [36], chapter 1.3.2.7), the software can be used for the collection of orthogonal tilt reconstruction (OTR, [186]) data without any limitations. Theoretically, the software can also be used for the correlation of features (e.g. gold particles) throughout a series of tomograms.

4.1.2 Implementation notes

In its current implementation, the MaverickTilt requires 7,5 hrs for the correlation of 43 datasets. Depending on the size of the dataset, a human user would require weeks to months to manually collect this amount of data, thus, a significant speedup in the overall process of RCT analysis has been achieved. Currently, the MaverickTilt software is implemented in *Python*, a high-level programming language that makes use of only a single CPU. Implementing the program in a lower-level language (e.g. *Fortran, C*) will further increase the speed of the program. Additionally, the problem of identifying the initial reference point sets which is the rate-limiting step, is *embarrassingly parallel* (for review, see [187]). Parallelization will allow real-time correlation of tilt pairs in automated high-throughput electron microscopy setups.

4.1.3 Performance of the MaverickTilt program on noisy data

The MaverickTilt has been demonstrated to be highly accurate in the correlation of tilt pairs (99,7% correct predictions in noise-free data). The reliability of the MaverickTilt

software in presence of *de-centering noise* and noise *coordinates* has been assessed beyond the levels of experimental data. Semi-automated routines for particle selection [167, 168] are relatively accurate in determining particle centring coordinates, resulting in a subsequent correction shift during alignments of usually no more than 10 pixels. The rate of false positive particle coordinates is usually negligible in negative stain preparations, however, the problem might become more severe in cryo datasets where contaminations are abundant. In tests on exoerimental data, MaverickTilt failed in this case rather than delivering false positives. To make cryo datasets with a high level of background contamination accessible to automated tilt pair correlation, we propose an approach for de-noising of particle coordinate data.

4.1.4   De-noising of particle coordinate datasets

As previously described, *de-centering noise* and *noise coordinates* occurs during semi-automated particle selection [167, 168]. During iterative alignment and classification of particles, particles are aligned to centered references, thus, the centering error is corrected during alignment (see chapter 1.2.2). Multivariate statistical analysis (MSA) and subsequent classification will sort particles according to their orientation, where particles representing contamination (representing false positive *noise coordinates*) are likely to be grouped into either the same class, or into classes which are generally more heterogeneous. Visual inspection of classes for either the representation of contaminants or for heterogeneity allows identification of particles that represent *noise coordinates* without requiring inspection of the whole dataset. The methodology for de-noising of particle coordinate datasets is thus proposed to be implemented as follows: First, all particles selected by the particle detection software are used for several iterations of reference-free alignment and classification ([188], see chapter 1.2.1). Subsequently, all particle coordinates are corrected by the shift proposed during the last iteration of the alignment, correcting the dataset for *de-centering noise*. Particles where the proposed shift exceeds a certain threshold value cam be discarded. Finally, particle quality is assessed by the cross-correlation coefficient (CCC) determined during alignment (see chapter 1.2.2) and by visual inspection of class averages reconstructed after alignment. Particles which correlate to the alignment templates with a CCC below a user-defined threshold or which were grouped in a class in which contaminants are represented can be discarded, allowing correction for the reduction of *noise coordinates*. The implementation of a software which

automates the progress of particle coordinate after visual inspection of averaged classified images is currently in progress.

4.1.5 Application to experimental data

The MaverickTilt software has been used to correlate tilt pairs in experimental tilt data sets of negatively stained *Thermus thermophilus* V-ATPase and *S.cerevisiae* eIF3. For eIF3, 11,025 particles were retrieved from 16 micrographs and used for subsequent analysis. From 45 micrographs tilt pairs of the V-ATPase, 45,256 tilt pairs could be retrieved. Micrographs from which particle coordinates where correlated were visually inspected for errors. An example for the output of the MaverickTilt software is given in Figure 33.
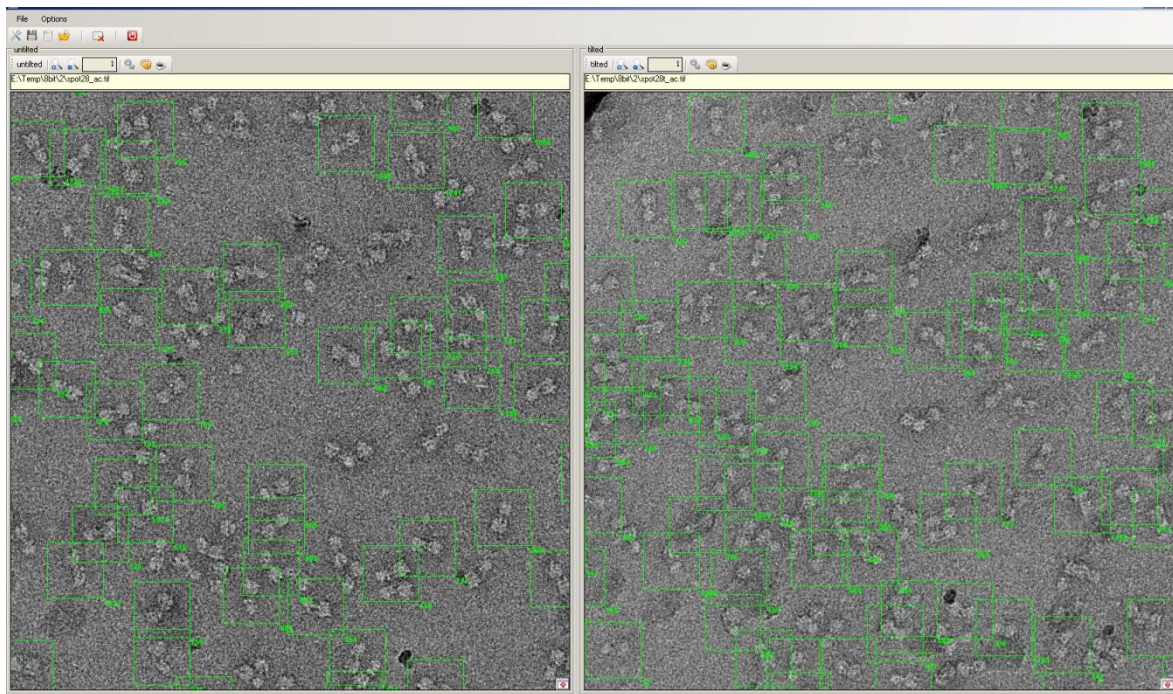
Figure 33 Correlation of experimental data with the MaverickTilt software. Screenshot of the cowPicker software displaying the output of the MaverickTilt software on negatively stained micrographs of V-ATPase.

From all selected *tilt pairs*, random conical tilt reconstructions (RCT, see chapter 1.3.2.7) were done. An example gallery of RCT reconstructions from the correlated dataset of the V-ATPase is shown in Figure 34.

As seen shown in the examples in Figure 34, all three-dimensional volumes faithfully represent the structure of the V-ATPase, demonstrating the reliability of tilt pair correlation by the MaverickTilt software and its applicability to experimental data.
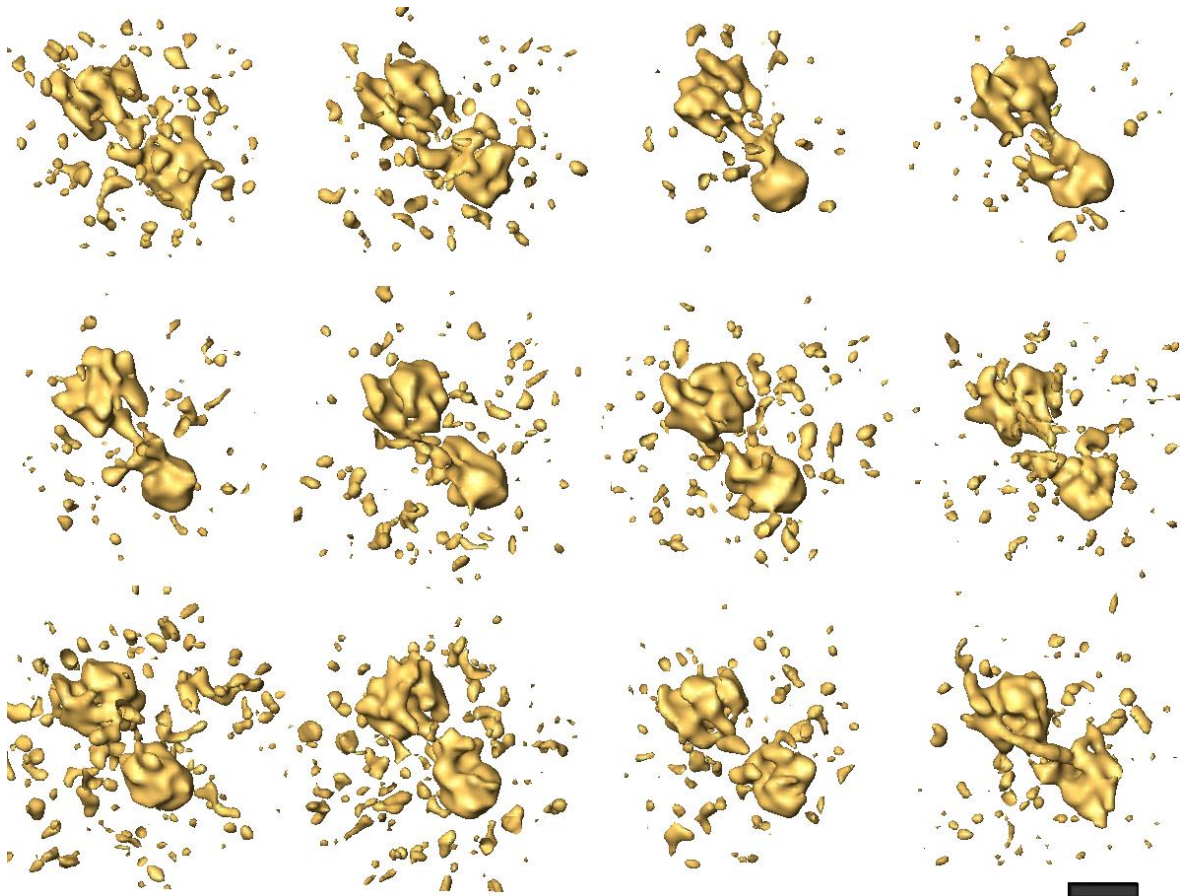
Figure 34 Example RCT reconstructions of V-ATPase from *Thermus thermophilus*. Tilt pairs for RCT reconstructions were correlated using the MaverickTilt software. The structure of the V-ATPase is faithfully represented in all volumes reconstructed from the dataset. Scale bar: 10 nm.

4.2   The 50S Ribosomal subunit of *Thermotoga maritima*

4.2.1   The extended ribosomal stalk of *T. maritima*

In previously reported three-dimensional structures of the large ribosomal structures obtained either by x-ray crystallography or cryo-EM reconstruction, the ribosomal L7/L12 stalk is either absent or truncated. This finding has been attributed to the high degree of flexibility of the ribosomal stalk, which leads to a loss of signal in averaging processes. Within the large ribosomal subunit of *T. maritima*, however, an extended ribosomal L7/L12 stalk can be visualized in the reconstruction. In the averaged, higher-resolution structure, the N-terminal domains of six copies of L7/L12 are clearly visible. In order to visualize the highly flexible C-terminal domains of L7/L12, local classification of the stalk region followed by classification of a high number of randomly reconstructed 3Ds was done. This method is particularly useful for complexes in which the flexible parts are known and representing only a minor portion of the overall complex. It is computationally undemanding compared to

other approaches and can be easily implemented. Using this approach, it was possible to reconstruct densities which are marking preferential positions of the flexible C-terminal domains of L7/L12. Six copies of the C-terminal domains of L7/L12 could be accommodated in the corresponding density, confirming the exhaustive representation of preferential conformational states of the L7/L12 C-terminal domains.

The visualization of the highly flexible ribosomal stalk of *T.maritima* can be, in part, ascribed to a preferential orientation of the large ribosomal subunit in the "crown view" in the particle population which was deduced from the dataset recorded from particles prepared on a carbon support film. This preferential orientation is adopted because the plane subunit interface of the 50S RSU preferentially binds to the carbon support film, thus limiting stalk movements in the plane perpendicular to the subunit interface plane. This notion is supported by the finding that the overall stalk structures in the dataset created from particles in cryo without carbon support film is slightly bent towards the ribosomal A-site. Given the biological function of the ribosomal stalk in delivering factors to the A-site, this finding is not surprising. Nevertheless, it demonstrates the effect of carbon film on limiting flexibility of macromolecular complexes upon preferential binding to the film.

Rigidity of enzymes and macromolecular complexes is generally observed in hyperthermophilic organisms (for review, see [162]), which might additionally contribute to improved visualization of the L7/L12 stalk in T. maritima. The observation of six copies of L7/L12 on the ribosomal stalk of archaea [51], and thermophilic [50] or halophilic [51] eubacteria. The prevalence of a heptameric (L7/L12)$_6$/L10 complex seems to be a more generalized feature of extremophilic organisms. As one mechanism of protein stabilization in thermophilic environments, increasing intersubunit interactions and oligomerization of proteins has been observed [162]. In this sense, additional binding sites for factors as offered by two additional copies of L7/L12 might constitute an adaptation towards thermophilicity. Furthermore, a cooperative effect of L7/L12-L10-L11 for binding to the 23S rRNA has been described, with increasing affinity and stability of the binding as the number of copies of L7/L12 in the complex raised from four to six [189]. Both factors, the increased number of copies of L7/L12 on the ribosomal stalk in combination with the reduced flexibility of the stalk, can presumably be contributed to the thermophilic environment of the host organism and significantly facilitate the visualization of the ribosomal stalk.

4.2.2   Expansion segments of the large ribosomal subunit

When comparing the large ribosomal subunits of *T. maritima* and *T. thermophilus*, three 23S rRNA expansion segments can be found, one of which has not been reported in any ribosomal structure so far. The expansion segments termed ES1, ES2 and ES3 are all located in the periphery of the large ribosomal subunit, as can be expected from previous findings highlighting the evolutionary and structural conservation of the ribosomal core [60-65]. The largest of the three expansion segments, ES1, can also be observed when comparing the *E.coli* 23S rRNA to the *T. thermophilus* 23S rRNA. ES1 of *T. maritima*, however, is much larger than ES1 found in *T. thermophilus*. Since the optimal growth temperature of *T. maritima* (~80°C) is higher than that of of *T. thermophilus* (~60°C), an implication of an enlarged ES1 for adaptation to a thermophilic environment can be hypothesized. Due to the absence of binding partners in the presented structure, the precise biological function cannot be determined.

The second expansion segment, ES2 is an expansion in the helix 63 of the 23S rRNA which has been previously reported in eukaryotic ribosomal structures [68, 190]. The suggested function of this expansion segment found in eukaryotes has been to coordinate non-ribosomal factors (like chaperones, modifying enzymes or SRP) at the peptide exit tunnel [68]. It may provide additional binding sites to the translocation apparatus or SRP, which is shortened in eukaryotes. The existence of additional binding sites may promote and stabilize binding in the thermophilic environment of *T. maritima*. Yet, the expansion segment is much smaller than in eukaryotes. Since the large ribosomal subunit of *T. maritima* has been proven to be quite rigid, and since a space-filling RNA model of the helical region of the expansion segment could be completely accommodated in the additional density, we believe that there is no further flexibility of the expansion segment which could be visualized in full-length.

ES3 is unique to any ribosomal structure known so far. This uniqueness, together with the absence of any external binding factor in the structure, makes the assignment of the biological function of this expansion segment particularly challenging. It can be observed, however, that the ES3 establishes contact with the highly flexible L1 arm, which is believed to be involved in translocation of the tRNA from the P- to the E-site [48, 178] and/or involved in the release of tRNAs from the E-site [179]. In order to get further insights into the biological role of ES3, analysis of conformational sub-states of L1 was done.

4.2.3 Conformational flexibility of L1 in the large ribosomal subunit

Extensive flexibility of L1 in ribosomes has been reported previously [48, 176-179]. Recent work supports the notion that L1 assists EF-G in translocating the tRNA from the P-site to the E-site of the translating ribosome [178]. L1 is hypothesized to support EF-G(GDP●P$_i$) in the direction of the inherent Brownian motion tRNA anticodon-mRNA codon complex into the forward direction. This may be achieved by blocking of the reverse path by EF-G(GDP●P$_i$) and by L1 biasing the stochastic movement of the hybrid P/E-site tRNA towards the E-site [178]. The flexible movement of L1 that is necessary to exert this function could be detected at the level of the isolated 50S RSU of *T. maritima*. The finding that binding of L1 to the ES3 occurs in one of the conformational states of L1 which have been reported in this study adds further evidence to the implication of ES3 in adapting the ribosome to high temperatures. We thus conclude that ES3 may serve as a scaffold to limit the flexibility of L1 to provide rigidity in order to kinetically control the function of L1 in the hyperthermophilic environment.

Generally, the precise role of the expansion segments can be further explored only in presence of binding partners to which they potentially bind. The stable and highly rigid ribosome of T. maritima, however, offers a promising target for further structural investigations.

4.3 Translation initiation factor 3 (eIF3) from *Saccharomyces cerevisiae*

For the reconstruction of eIF3 and eIF3 sub-complexes, random conical tilt reconstructions (RCT 3Ds) were obtained from high-pass filtered single particle images of the eIF3 holocomplex (see Figure 25). RCT 3Ds were averaged after mutual registration via 3D maximum-likelihood alignment [39] and subsequent MSA and classification [29] in 3D space. The 3D structure of negatively stained eIF3 holocomplex was refined using supervised classification (see chapter 1.3.2.1), and refined 3D structures were used for refinement of negatively stained sub-complexes and an unstained cryo dataset.

Application of the GraFix protocol to eIF3 was required in order to obtain complexes which displayed sufficient structural homogeneity and integrity to be suited for further EM analysis. For obtaining high amounts of highly pure eIF3 complexes, a sophisticated purification scheme had to be applied, including a gel filtration step for purification of pre-assembled eIF3 complexes. Gel filtration imposes a considerable amount of mechanical stress on the complex, leading to disruption of labile complexes. Recent observations made

in our group suggest that instability of complexes and susceptibility to disruption after gel filtration is a common feature of biomacromolecular complexes. Disruption of complexes after gel filtration severely limits the applicability of sophisticated purification schemes of complexes for structural investigations. Here, we show that application of the GraFix protocol enables structural investigations by EM without the need of altering optimized purification schemes. This does not only underline the applicability of the GraFix protocol, but also demonstrates GraFix treatment to be an indispensible step in order to make certain complexes accessible to high-resolution structural examinations in the first place.

In order to obtain RCT 3Ds which represent structural features of the complex rather than reproducing reconstruction errors introduced by the *shadowing effect* (see Figure 24), high-pass filtering of single particle images was necessary (see Figure 25). High-pass filtering of negatively stained single particles recorded with a tilted specimen holder can be applied in order to reduce shadowing effects. Caution has to be taken, however, that structural information in the low spatial frequency domain is not suppressed to a level which renders coarse structural information inaccessible for reconstruction algorithms. To assess reliability of the obtained reconstructions which are used as an initial reference for the subsequent refinement procedure, several checkpoints have to be passed.

4.3.1 Checkpoints for the evaluation of initial models obtained from averaged RCT reconstructions

After 3D maximum likelihood alignment of RCT 3Ds, a consensus structure of all registered RCT volumes was calculated [39]. Within this consensus structure, features which are displayed by untilted class averages which have been used for RCT reconstruction should be recognizable. When using band-pass filtered tilted images for RCT reconstruction, a featureless cone was obtained as consensus structure after alignment (see Figure 25 B). Since no comparability to previously determined class averages (see Figure 23) exists, high-pass filtering was applied to images used for RCT reconstructions (see Figure 25 C). Here, the overall shape resembled the class averages, but more detailed structural features were not discernible, thus giving a first indication of the poor suitability of averaged RCT structures for refinement.

3D RCT class averages used as initial references, generally, should resemble the consensus structure, but still be clearly discernible in terms of structural sub-states defined by conformational flexibility of different assembly states. Classification of registered RCT 3Ds resulted in few class averages which were reproducing the overall shape of the

consensus structure while still displaying distinguishable features. As a test, RCT 3D class averages obtained this way were used as initial references for supervised classification (see Figure 26).

Refinement by supervised classification is known to suffer from bias introduced by initial references (for review, see [32]). Initiating supervised class with incorrect references will, most likely, reproduce the incorrect input model [32]. Thus, cross-validation is required to verify the results of supervised classification. During supervised classification, images are grouped with respect to preferential alignment to projections from a reference (see chapter 1.3.2.1). If the structure represented by the preferential reference of a group can be reproduced by using a reference from another group, supervised classification can be presumed to have sorted images correctly. If the false positive reference is reproduced, sorting by supervised classification is demonstrated to be unspecific, and the result has to be discarded.

Supervised classification of datasets from eIF3 and eIF3 sub-complexes resulted in 3D structures which were well-defined but heterogeneous within one dataset (see Figure 26). Cross-correlation identified grouping by supervised classification to be unspecific (see Figure 27). This non-specificity is observed because initial references did not contain sufficient information to faithfully describe different conformations contained within the dataset. This information is necessary for correct separation of particles, which is, in turn, a prerequisite for obtaining high-resolution structures.

The fact that mid-resolution reconstruction which were not displaying any obvious flaws were obtained from an initial reference set of insufficient quality underlines the importance of the correctness of initial references for refinement of electron microscopy data.

At the current stage, new initial models have to be obtained to promote the determination of structures from the currently available datasets. In order to circumvent the technical problems we faced because of negative stain-induced *shadowing effects,* unstained cryo tilt datasets will be used for the determination of initial reference 3Ds.

4.4 The V-ATPase of *Thermus thermophilus*

4.4.1 GraFix preparation of the the V-ATPase of *Thermus thermophilus*

One of the structurally most intriguing features of V-ATPases is that they comprise two major domains: The membrane-bound proteolipid V_0 domain and the cytosolic, water-soluble V_1 domain. Biochemical procedures for isolation and purification of intact

complexes for EM need to satisfy requirements for the purification of both water-soluble and transmembrane proteins. If these requirements are not equally met, V_1 will dissociate from V0, resulting in a disrupted complex. Current protocols for the preparation of ATPases for EM can be regarded as a balanced compromise between the biochemical demands imposed by both the cytosolic and transmembrane domain of the complex, respectively. In previous preparations of V-ATPases, however, the presence of disrupted particles on the EM grid has been observed to be significant and can be observed directly at the level of raw electron micrographs preparation ([152], Figure 6 therein, [153], Figure 1 B therein and [145], Figure 1 (b) therein, [149], Figure 1 B therein). The number of particles used in previous EM studies of negatively stained V-ATPases, varied between 10,000 [149], 12,000 [152] and 16,000 [153] particles. In a recent study based on unstained cryo preparations, around 16,000 particles were selected from 135 micrographs [145]. Given the number of particles from previous studies all lie within the same range (between 10,000 and 16,000), they can be supposed to represent the upper limit for the number of intact particles that can be retrieved from an average dataset.

After applying the GraFix protocol, we were able to select a minimum of 108,997 particles from only 45 micrographs (untilted view). Sorting of single particle images from V-ATPases after MSA and classification ([29], see chapter 1.2.3) by discarding all images which were contained in classes which represented disrupted or aggregated complexes, a minimum of 80% of all particles were found to be structurally fully intact. From the set of tilt pairs correlated by the MaverickTilt software, less than 10% of particles were discarded altogether from both the tilted and untilted views, indicating more than 90% of particles to be structurally intact on the EM grid. This high percentage of intact V-ATPases is unmatched by any previous electron microscopy preparation (compare [152], Figure 6, [153], Figure 1, [145], Figure 1 (b), and [149], Figure 1 B), and can be attributed to the use of the GraFix protocol. Given the high structural integrity observed in our dataset, both image quality and image statistics are provided in a way that is needed in order to attain a high-resolution structure of the *Thermus thermophilus* V-ATPase.

4.4.2   Validation of refined reconstructions

The importance of correct starting models for faithfully resolving heterogeneous sub-states in a population of complexes within an EM dataset has been demonstrated for the eIF3 complex (see chapter 4.3.1). As a result of supervised classification, well-defined three-dimensional reconstruction might emerge from the dataset even when information on

structural features and structural heterogeneity is not contained in the initial reference set obtained from aligned and averaged random conical tilt reconstructions (RCT 3Ds). For the V-ATPase of *Thermus thermophilus*, validation of each step of the refinement process is facilitated by previous work from which the structure of the complex is already known at low resolution [152], allowing the structural assessment of refinement intermediates.

As an initial critical step, visual inspection of negatively stained complexes which were recorded from a tilted specimen holder has to be done. In the case of V-ATPase, no shadowing effect, i.e. unilateral contrast in single particle images, could be observed (see Figure 28 A, Figure 34 right box). Single RCT 3Ds are well defined in all directions. Furthermore, all important features of intact V-ATPases can be recognized already at the level of single RCT 3Ds (see Figure 28 C, Figure 34). In the RCT 3D consensus structure which is obtained by averaging all RCT 3Ds after maximum likelihood alignment, V_1 and V_0 can be clearly recognized (see Figure 28 D). In the representation of V_0, subunit V_0-d is resolved as a protruding edge, indicating faithful reproduction of the asymmetry observed for the V-ATPase. From the consensus structure after 3D maximum-likelihood alignment, all peripheral stalks as well as the central stalk are absent. This absence can be explained by a high degree of flexibility of these domains, which is resolved in subsequent analysis. 3D class averages display the full range of features expected for low-resolution reconstructions of V-ATPases from *Thermus thermophilus*, including a central stalk and peripheral stalks (see Figure 28 E). Furthermore, 3D class averages can be unanimously distinguished from each other by variations in the positioning of the central stalk, the number of peripheral stalks and the positioning of the V_1 domain with respect to V_0. Three-dimensional reconstructions obtained after supervised classification strongly resemble the references which were used as input, except for showing more detail in the reconstruction (see Figure 28 F).

To further validate the refinement, cross-validation was done was described for the eIF3 complex (see chapter 3.3.4). The results of cross-validation of final reconstructions from *Thermus thermophilus* V-ATPase are shown in Figure 35. Here, it can be observed that reconstructions obtained after cross-validation preserve detailed structural features (e.g. peripheral stalks) only if present in the reference. False positive structural features (e.g. reconstruction of additional stalks which are present in the reference volume but not in the input dataset) are rarely observed. Thus, the sorting of the dataset by supervised classification and subsequent refinement is robust.
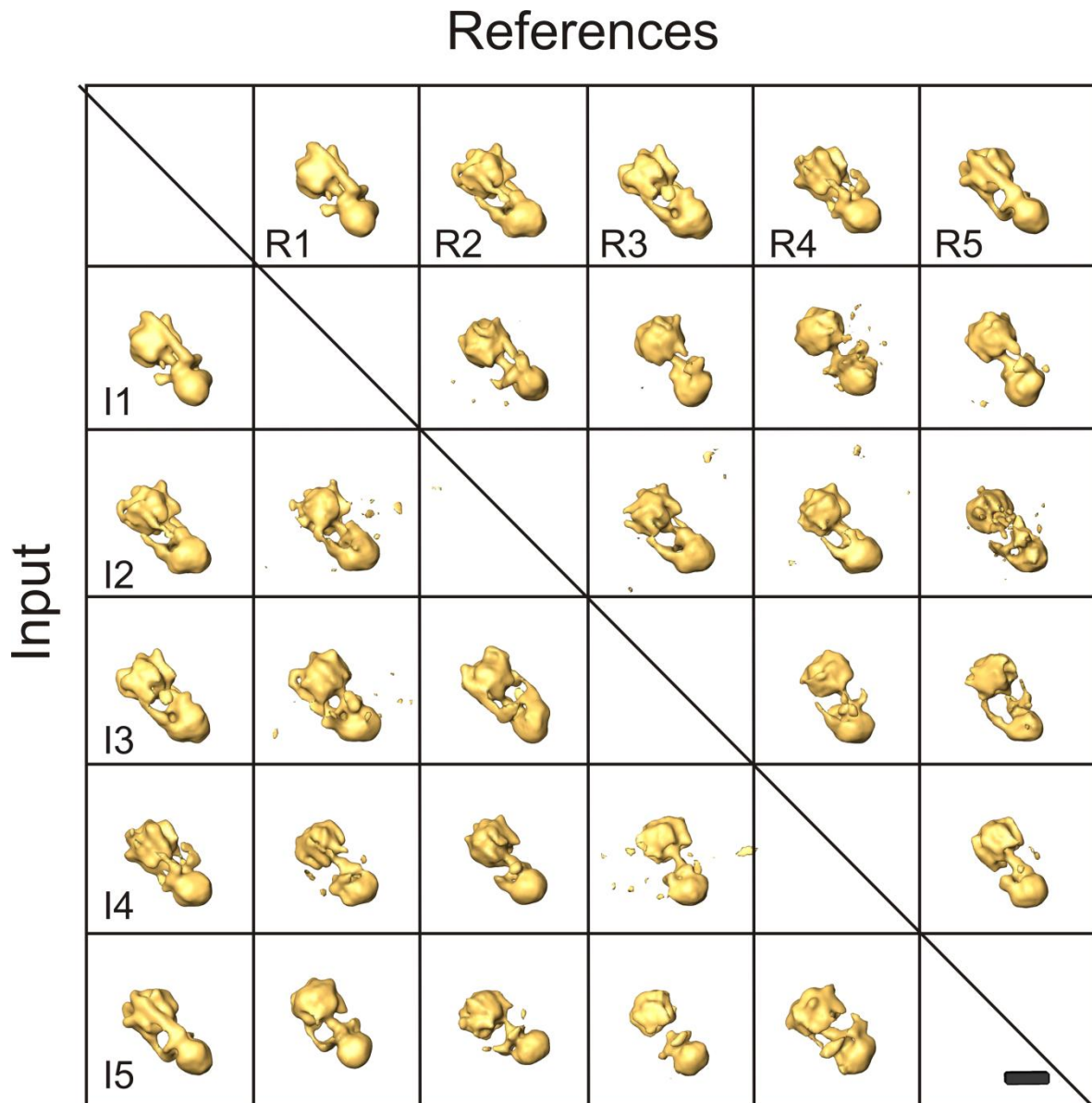
References

Figure 35 Cross-validation of refined *Thermus thermophilus* V-ATPase reconstructions. Images aligning best to input references (I1-I5) were reconstructed by iterative angular reconstitution, using references (R1-R5) as indicated in the scheme as reference in the first iteration. In reconstructions obtained during cross-validation, fine structural features (e.g. peripheral stalks) are only recovered if present in the reference structure. False positive reconstruction of features is rarely observed. Scale bar: 10 nm.

Sorting of particles during supervised classification can potentially occur according to preferential orientations of particles which were represented by the initial references. To exclude the possibility that initial references used for supervised classification represent preferred orientations rather than conformational sub-populations in the dataset, the angular orientation of all particles after supervised classification was checked. To assess the angular orientation of particles, all datasets sorted by supervised classification were treated by MSA and subsequent classification [29] in order to reduce bias from the previous alignment step.
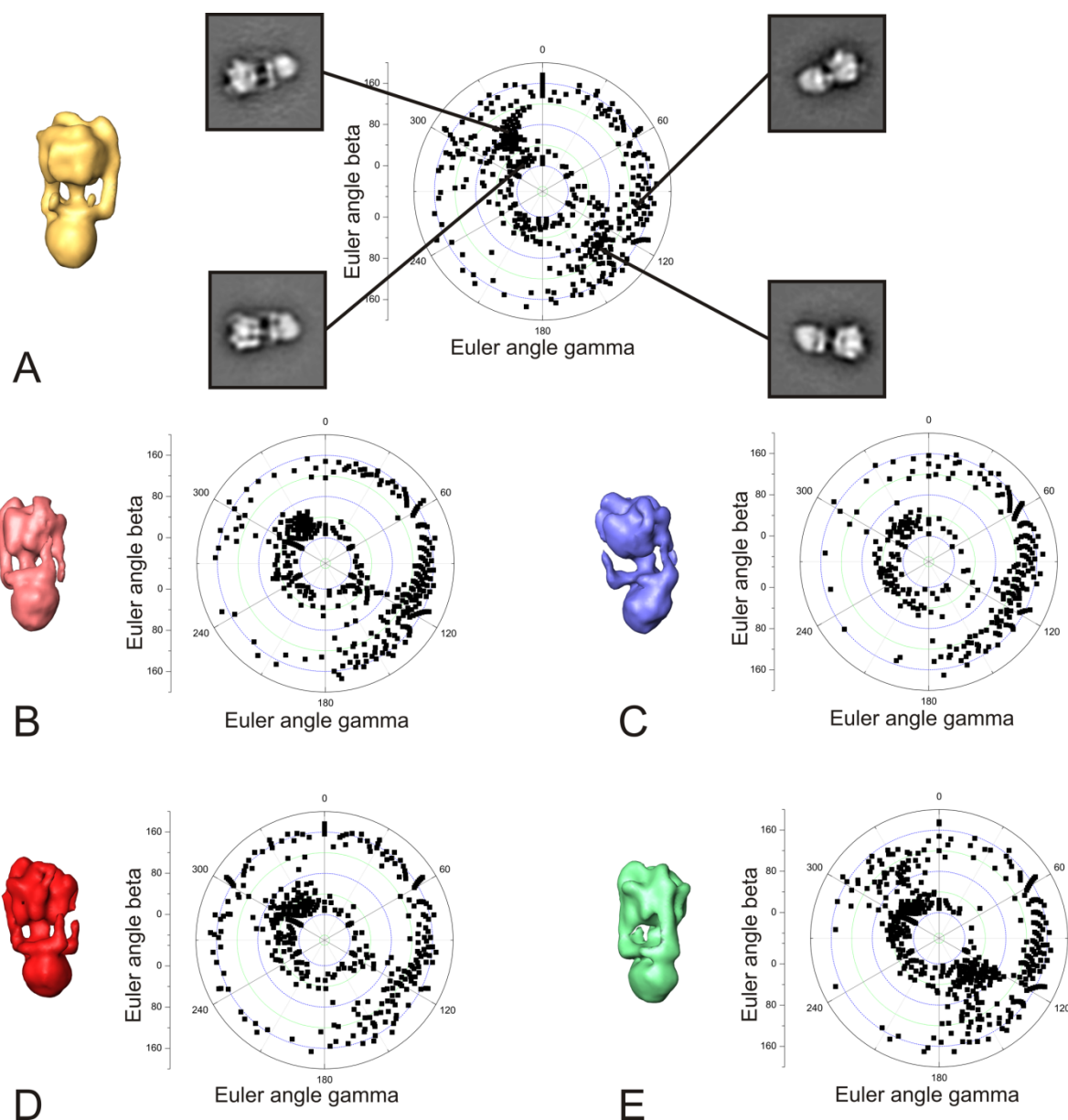
Figure 36 Angular distribution of class averages obtained from datasets after supervised classification.
Euler angles β and γ of class averages were obtained by angular reconstitutions. 3D reconstructions to the left of each plot were obtained from the dataset of which Euler angles are plotted. Particle orientations represented by Euler angles in clusters of the plots are shown in (A). Particles are predominantly represented by views rotating around the axis of the central stalk which is parallel to the grid surface. Distributions of angular orientations are similar for all datasets, indicating no bias introduced by sorting for preferential orientation during supervised classification.

Classes were averaged, and the angular orientations of all class averages were determined by angular reconstitution [181], using the 3D volume from the previous refinement step as a reference. Euler angles β and γ of all class averages were plotted and plots were compared (Euler angle α is normalized between all projections during alignment and can thus be neglected). As can be seen in Figure 36, the distributions of Euler angles represented by class averages from all datasets are very similar, thus, no bias due to sorting for preferred angular orientations has occurred during supervised classification.

4.4.3 Flexibility of the cytosolic V_1 domain

Flexibility of the cytosolic V_1 domain of V-ATPases has been reported in previous studies in which the complex was examined by EM [149, 152]. In a previous reconstruction of a plant V-ATPase, the prevalence of a flexible V_1 domain was induced by the absence of adenylyl-imidodiphosphate (AMP-PNP), a non-hydrolysable ATP analogue [149]. The three-dimensional map obtained from particles prepared in absence of AMP-PNP had a low resolution and showed few structural features which are characteristic for V-ATPases. In contrast, those features were present in a reconstruction from complexes to which AMP-PNP was added [149], indicating severe impairment of the overall reconstruction process by flexibility of the V_1 domain. For the reconstruction of the V-ATPase of *Thermus thermophilus*, consequently, particles displaying a tilted V_1 domain at the level of 2D projections were discarded prior to three-dimensional reconstruction analysis [152]. It is likely, however, that structural heterogeneity cannot be determined in all views at the level of 2D images. A significant amount of complexes might still represent conformational states in which V_1 adopts different orientations, even after sorting complexes from a distinct view. Indeed, an initial reconstruction for which only 4200 single particle images were used did not improve in resolution after another 8100 particles were included in the dataset [152]. The fact that even a two-fold increase in image statistics does not lead to an improvement in the final 3D map of the *Thermus thermophilus* V-ATPase indicates that conventional reconstruction schemes relying on averaging of the whole dataset face their limits at the current resolution levels.

Observing an increased flexibility of the V_1 subunits under nucleotide-free conditions suggests the structural dynamics of V_1 to correlate with physiologically functional states [149, 151]. Resolving the structural heterogeneity of V_1 will not only help to attain a high-resolution reconstruction, but may also provide insight into functional conformations the V-ATPase can adopt.

By resolving structural heterogeneity using aligned averages of RCT 3Ds, we were able to capture flexibility of V_1 and to make full use of the whole range of all selected particles for the determination of high-resolution three-dimensional structures.

4.4.4 Functional implications of structural heterogeneity

Structural heterogeneity of V-ATPases with respect to the orientation of V_1 versus V_0 has been observed in all EM studies of V-ATPases as well as dissociation of V_1 from V_0. Reversible dissociation of V_1 from V_0 is believed to be the main regulatory switch for

controlling the activity of V-ATPases [150, 151, 153, 156]. Flexibility of V_1 has been observed to become predominantly in a nucleotide-free environment, suggesting a tilted V_1 subunit to be a functional intermediate before dissociation of V_1 [149].

Concomitantly with tilting of V_1 with respect to V_0 at different angles, varying stochiometry and intensity for the peripheral stalks of the V-ATPase could be visualized (see Figure 31, Figure 32). The observation of a central cavity in the V_1 domain which is directly located above the central stalk ([145], Figure 30 A) suggests that the centrals stalk does not penetrate deeply into V_1, unlike in F-ATPases where the central stalk formed by subunit γ extends thoughout a central cavity formed by the $\alpha_3\beta_3$ ring which forms the F_1 domain [183]. Thus, structural integrity of the V-ATPases during continous rotation of V_1 is, at least in large part, provided by the interaction of V_1 with the peripheral stalks. Indeed, a minimum of two peripheral stalks has been described in EM reconstructions (for example, see [122, 145, 149, 153]) from V-ATPases compared to F-ATPases, where only one peripheral stalk has been observed [191]. Thus, the interactions between V_1 and the peripheral stalks have been proposed to be the focal point of regulation for the reversible dissociation of V_1 from V_0 [184]. Subunit V_0-d of *Thermus thermophilus* (termed subunit C in other nomenclature) has been suggested to be involved in regulation of reversible dissociation [184]. As one regulatory event, phosphorylation of V_0-d has been described [155, 156], and release of V_0-d after phosphorylation has been discussed as a regulatory event in subunit dissociation [145, 151]. In our structure, V_0-d makes contact to the centrals stalk and the two major peripheral stalks. Remarkably, V_0-d is not resolved in the reconstruction which shows only two minor peripheral stalks and which displays the most extreme tilting of V_1 versus V_0 (see Figure 31 E). Taken together, we observe tilting of V_1 versus V_0, which can be presumed to be a structural intermediate of V_1 release from the complex. The observation of two thoroughly defined major stalks and the appearance of only a very small third minor stalk correlates with the observation of an untilted V_1 domain which is perpendicular to the membrane plane (see Figure 31C, D) In reconstructions in which larger tilting angles of V_1 versus V_0 are observed, we observe reduced densities for at least one major stalk and increased density for additional minor stalks in two out of three reconstructions (see Figure 31 A, B). In the reconstruction which displays the most extreme tilt angle, two minor stalks are observed at positions where major stalks are normally located. Additional, subunit V_0-d seems to be absent from the reconstruction (see Figure 31E). In all of our reconstructions, the total mass representing peripheral stalks can

be presumed to be roughly the same, with the same amount of mass either being distributed among two major stalks (C, D), one major stalk and two minor stalks (B) or four minor stalks (A). The only exception from this observation is represented by the reconstruction which represents the most extreme tilting of V_1 versus V_0 and where subunit V_0-d is absent (E). Here, only two minor stalks are present at the position where major stalks are normally expected. Based on this observations, we propose that a peripheral stalk in the fully functional V-ATPase of *Thermus thermophilus* is formed by subunits in the stochiometry of E_2G_2 rather than EG, and that two E_2G_2 peripheral stalks stabilize the functional conformation of V-ATPases. Subunits E and G from the Archaeon *Thermoplasma acidophilum* have been demonstrated to form homodimers and higher polymers in solution, yet, if they are allowed to interact at equimolar amounts, the formation of a monodisperse EG heterodimer was observed [192]. Given the principle ability to form homodimeres, a E_2G_2 arrangement on the V-ATPase as scaffold is possible.
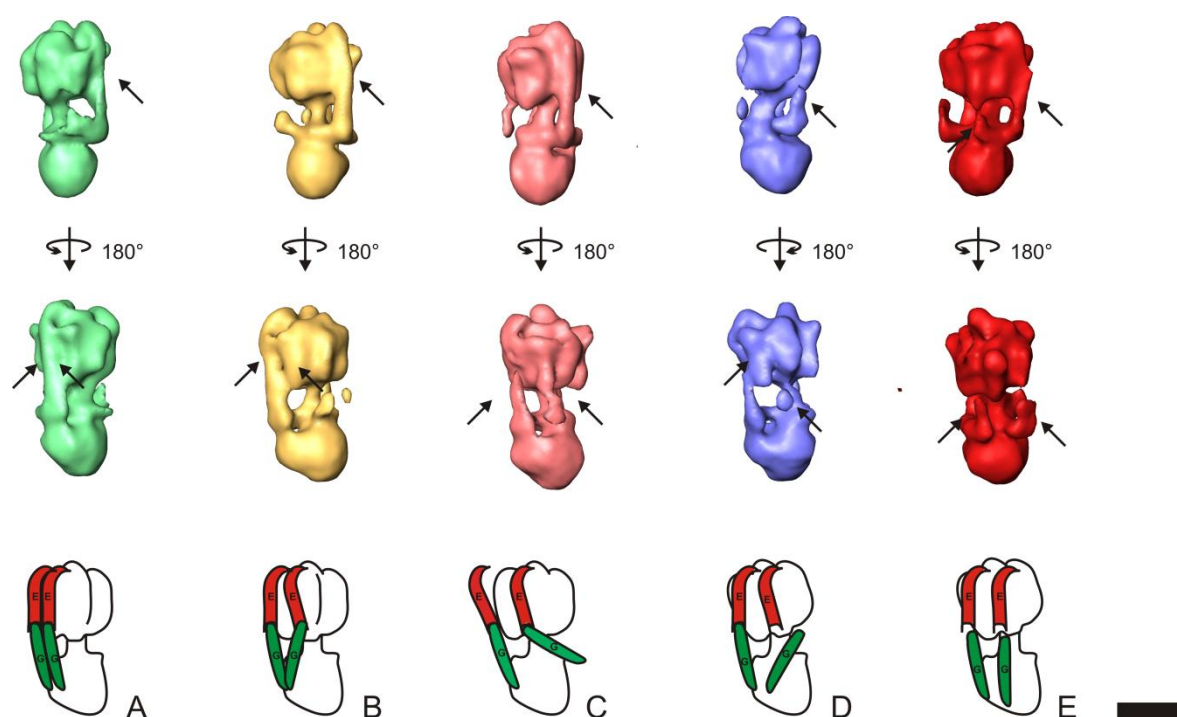
Figure 37 Hypothetical conformational states of E2G2 peripheral stalks. Views in the same column represent same reconstructions. Hypothetical conformational states of a tetrameric E_2G_2 peripheral are suggested based on observations made in 3D EM reconstructions above the model. Note that with decreased integrity of peripheral stalks, an increased tilting of the V_1 head domain is observed.

Tandem-MS of V-ATPases from *Thermus thermophilus* suggested the prevalence of two EG heterodimers [193]. In this MS analysis, however, V_0 was completely absent from the complex, presumably after being dissociated during initial droplet stages of electrospray

[193]. It is possible, however, that during this step, additional copies of EG complexes were stripped from the complex, leaving only two EG complexes attached to V_1. In fact, the prevalence of two EG complexes (representing a minor stalk) has been demonstrated to be the minimal mounting of intact V-ATPases in our reconstructions (see Figure 31). In our reconstructions, different conformations of E_2G_2 versus EG can be proposed, which are depicted in Figure 37. Since we currently do not have any biochemical data at hand which can be directly use to validate our proposal, the initial model should be consumed with care, indicating new directions for biochemical experiments which might validate these findings rather than attempting to postulate a new working model.

5 Conclusions

5.1 Three-dimensional electron microscopy of macromolecular complexes

In this work, three-dimensional structures of the large ribosomal subunit of *Thermotoga maritima*, the eukaryotic initiation factor 3 (eIF3) of *Saccharomyces cerevisiae* and the vacuolar V-Type ATPase from *Thermus thermophilus* are presented. Using three-dimensional reconstruction techniques which allow the resolution of structural heterogeneity within the single particle electron microscopy dataset, structural features which were unresolved in previous reconstructions could be visualized in the current reconstructions of the large ribosomal subunit of *Thermotoga maritima* and the vacuolar V-Type ATPase from *Thermus thermophilus*. In the reconstruction of the *Thermotoga maritima* large ribosomal subunit, the ribosomal L7/L12 stalk, which is absent in most ribosomal structures could be visualized in full length. Additionally, flexibility of the ribosomal L1 stalk could be demonstrated on isolated large ribosomal subunits. Reconstructions of the V-ATPase of *Thermus thermophilus* reveal additional peripheral stalks which could not be visualized in previous reconstructions of V-ATPases. Flexible sub-states of the cytosolic V_1 head domain could be visualized and correlated with observed stalk stochiometry. The disentanglement of structural heterogeneity provides intriguing new insights into the structure-function relationship of the examined complexes.

5.2 Applicability of the GraFix protocol

The GraFix protocol has been reported to improve properties of macromolecular complexes for EM studies [5]. The utilization of the GraFix protocol has been demonstrated to improve the structural integrity and binding behaviour of cytosolic complexes which are prevalent at mediocre quality [5]. In this work, we demonstrate and extend the scope of the applicability of the GraFix protocol. In the case of eIF3, the application of the GraFix protocol was a prerequisite to obtain viable complexes from a mixture of partly discrupted complexes after gel filtration (see Figure 23 B). After GraFix treatment, averages of highly stable and structurally homogenous complexes were obtained (see Figure 23 C). Responding to the fragility of the complex, the GraFix protocol could be

extended by an additional incubation in GraFix buffer after centrifugation without observing unspecific intermolecular cross-linking. By additional fixation during overnight incubation, even the most fragile complexes can be rendered viable for EM analysis.

The GraFix protocol was further demonstrated to be applicable to membrane complexes without any observed limitations. Stabilizing the V-ATPase by use of GraFix dramatically increased the number of intact particles on the EM grid, allowing to approach statistical image processing with an increased level of significance and reliability.

Encouraged by the extended applicability of the GraFix protocol, we anticipate improvement in reliability, attained resolution, and interpretability of a large number of future three-dimensional EM reconstructions.

5.3 Resolving structural heterogeneity in EM reconstructions

The necessity of abandoning the dogma of averaging all single particle images obtained from a complex into one final reconstruction is gaining recognition in the field of three-dimensional electron microscopy. Several new methods have been proposed to tackle heterogeneity in EM single particle datasets during refinement, and they all share the need for reliable initial models in which structural heterogeneity is at least partly resolved (see chapter 1.3.2.). Two principal methods are currently recognized in the field: Sampling with replacements and random conical tilt reconstruction (RCT).

For sampling with replacements, three-dimensional orientations of all projections have to be precisely known. Thus, a low-resolution map of the complex has to be available, and structural heterogeneity may comprise only a small fraction of the total complex volume. These presumptions hold true for most ribosomal complexes, so we could successfully use the approach to visualize the ribosomal L7/L12 stalk in full length.

If no initial structural information is available, RCT reconstructions can be used as initial references. The refinement of eIF3, however, has demonstrated that this methods faces its limits at low quality of the RCT data. Even with wrong initial models, a set of structurally integer reconstructions was obtained. This finding highlights the importance of quality checks and cross-validation to verify refinement results.

Being able to resolve structural heterogeneity comprising any part of the macromolecular assembly, averaging of RCT 3Ds after 3D maximum-likelihood alignment and classification in 3D space is a promising technique for creating initial 3D models. Its broad application which exploits the full statistical range of large datasets, however, has been

limited by the bottleneck of manual collection of *tilt pairs* needed for RCT reconstructions. By automating the process of *tilt pair* collection, the MaverickTilt software introduced in this work helps to overcome this bottleneck. Thus, a speed-up in obtaining initial reference sets for future reconstructions is anticipated. Since with the new software, large datasets can be correlated in a high-throughput approach, the reliability of the initial reference sets is further improved by increasing statistics by a higher number of RCT reconstructions which can be used for 3D maximum-likelihood alignment and subsequent classification in 3D space.

Considering ongoing advances in sample preparation and image processing techniques, we anticipate a wealth of biological and functional information to be retrieved in future three-dimensional electron microscopy reconstructions of macromolecular complexes. In concert with other complementary techniques such as X-ray crystallography and NMR, new findings will allow a deepening of our understanding of the complex network of biochemical processes that govern our life.

6 Bibliography

1. Sprangers, R. and L.E. Kay, *Quantitative dynamics and binding studies of the 20S proteasome by NMR.* Nature, 2007. **445**(7128): p. 618-22.
2. Szymczyna, B.R., et al., *Solution NMR studies of the maturation intermediates of a 13 MDa viral capsid.* J Am Chem Soc, 2007. **129**(25): p. 7867-76.
3. Ludtke, S.J., et al., *De novo backbone trace of GroEL from single particle electron cryomicroscopy.* Structure, 2008. **16**(3): p. 441-8.
4. Leschziner, A.E. and E. Nogales, *Visualizing flexibility at molecular resolution: analysis of heterogeneity in single-particle electron microscopy reconstructions.* Annu Rev Biophys Biomol Struct, 2007. **36**: p. 43-62.
5. Kastner, B., et al., *GraFix: sample preparation for single-particle electron cryomicroscopy.* Nat Methods, 2008. **5**(1): p. 53-5.
6. Reimer, L., *Transmission electron microscopy; physics of image formation and microanalysis.* 4 ed. Springer series in optical sciences; Vol. 36. 1997, Berlin; Heidelberg; New York: Springer Verlag. 584.
7. Wade, R.H. and J. Frank, *Electron microscope transfer functions for partially coherent axial illumination and chromatic defocus spread.* Optik, 1977. **49**(2): p. 81-92.
8. Frank, J., *The envelope of electron microscopic transfer functions for partially coherent illumination.* Optik, 1973. **38**(5): p. 519-536.
9. Sander, B., M.M. Golas, and H. Stark, *Automatic CTF correction for single particles based upon multivariate statistical analysis of individual power spectra.* J Struct Biol, 2003. **142**(3): p. 392-401.
10. Glaeser, R.M. and K.H. Downing, *Assessment of resolution in biological electron crystallography.* Ultramicroscopy, 1992. **47**(1-3): p. 256-65.
11. Saad, A., et al., *Fourier amplitude decay of electron cryomicroscopic images of single particles and effects on structure determination.* J Struct Biol, 2001. **133**(1): p. 32-42.
12. Thon, F., *Zur Defokussierunsabhaengigkeit des Phasenkontrastes bei der elektronenmikroskopischen Abbildung.* Z. Naturforschg., 1966. **21a**: p. 476-478.
13. *Cryomicroscopy and radiation damage. International Study Group for Cryo Electron Microscopy Workshop. Schloss Ringberg am Tegernsee, 9-14 November 1981.* Ultramicroscopy, 1982. **10**(1-2): p. 1-177.
14. van Heel, M., et al., *Single-particle electron cryo-microscopy: towards atomic resolution.* Q Rev Biophys, 2000. **33**(4): p. 307-69.
15. Joyeux, L. and P.A. Penczek, *Efficiency of 2D alignment methods.* Ultramicroscopy, 2002. **92**(2): p. 33-46.
16. Sander, B., M.M. Golas, and H. Stark, *Corrim-based alignment for improved speed in single-particle image processing.* J Struct Biol, 2003. **143**(3): p. 219-28.
17. Penczek, P., M. Radermacher, and J. Frank, *Three-dimensional reconstruction of single particles embedded in ice.* Ultramicroscopy, 1992. **40**(1): p. 33-53.
18. van Heel, M., *Multivariate statistical classification of noisy images (randomly oriented biological macromolecules).* Ultramicroscopy, 1984. **13**: p. 165-184.
19. van Heel, M., et al., *A new generation of the IMAGIC image processing system.* J Struct Biol, 1996. **116**(1): p. 17-24.
20. van Heel, M., *Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction.* Ultramicroscopy, 1987. **21**: p. 111-124.
21. Harauz, G., van Heel, M., *Exact filters for general geometry three-dimensional reconstruction.* Optik, 1986. **73**: p. 146-156.
22. Smith, S.W., *The Scientist and Engineer's Guide to Digital Signal Processing.* 1st edition ed. 1997: California Technical Pub. 626.

23. Frank, J., Penczek, O.A. Agrawal, R.K., Grassucci, R.A., Heagle, A.B., *Three-dimensional cryo-eletron microscopy of ribosomes.* Methods Enzymol, 1999. **317**: p. 276-291.

24. Saxton, W.O. and W. Baumeister, *The correlation averaging of a regularly arranged bacterial cell envelope protein.* J Microsc, 1982. **127**(Pt 2): p. 127-38.

25. Frank, J. and M. van Heel, *Correspondence analysis of aligned images of biological particles.* J Mol Biol, 1982. **161**(1): p. 134-7.

26. van Heel, M. and M. Schatz, *Fourier shell correlation threshold criteria.* J Struct Biol, 2005. **151**(3): p. 250-62.

27. Gavin, A.C., et al., *Proteome survey reveals modularity of the yeast cell machinery.* Nature, 2006. **440**(7084): p. 631-6.

28. Gao, H., et al., *Dynamics of EF-G interaction with the ribosome explored by classification of a heterogeneous cryo-EM dataset.* J Struct Biol, 2004. **147**(3): p. 283-90.

29. van Heel, M., *Multivariate statistical classification of noisy images (randomly oriented biological macromolecules).* Ultramicroscopy, 1984. **13**(1-2): p. 165-83.

30. Penczek, P.A., et al., *Estimation of variance in single-particle reconstruction using the bootstrap technique.* J Struct Biol, 2006. **154**(2): p. 168-83.

31. Hall, R.J., B. Siridechadilok, and E. Nogales, *Cross-correlation of common lines: a novel approach for single-particle reconstruction of a structure containing a flexible domain.* J Struct Biol, 2007. **159**(3): p. 474-82.

32. Scheres, S.H., et al., *Maximum-likelihood multi-reference refinement for electron microscopy images.* J Mol Biol, 2005. **348**(1): p. 139-49.

33. Scheres, S.H., *Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization.* Nat. Methods, 2007. **4**: p. 27-29.

34. Sigworth, F.J., *A maximum-likelihood approach to single-particle image refinement.* J Struct Biol, 1998. **122**(3): p. 328-39.

35. Simonetti, A., et al., *Structure of the 30S translation initiation complex.* Nature, 2008. **455**(7211): p. 416-20.

36. Radermacher, M., et al., *Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli.* J Microsc, 1987. **146**(Pt 2): p. 113-36.

37. Dube, P., et al., *The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry.* EMBO J, 1993. **12**(4): p. 1303-9.

38. Gieffers, C., et al., *Three-dimensional structure of the anaphase-promoting complex.* Mol Cell, 2001. **7**(4): p. 907-13.

39. Sander, B., et al., *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.* Mol Cell, 2006. **24**(2): p. 267-78.

40. Frank, J. and L. Al-Ali, *Signal-to-noise ratio of electron micrographs obtained by cross correlation.* Nature, 1975. **256**(5516): p. 376-9.

41. Unser, M., Trus, BL., Steven, AC., *A new resolution criterion based on spectral signal-to-noise ratios.* Ultramicroscopy, 1987. **23**(1): p. 39-51.

42. Selmer, M., et al., *Structure of the 70S ribosome complexed with mRNA and tRNA.* Science, 2006. **313**(5795): p. 1935-42.

43. Diaconu, M., et al., *Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation.* Cell, 2005. **121**(7): p. 991-1004.

44. Achenbach-Richter, L., et al., *Were the original eubacteria thermophiles?* Syst Appl Microbiol, 1987. **9**: p. 34-9.

45. Nelson, K.E., et al., *Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima.* Nature, 1999. **399**(6734): p. 323-9.

46. Maliuchenko, N.V., et al., *[Investigation of ribosomes of E. coli and T. maritima by atomic force microscopy].* Biofizika, 2006. **51**(3): p. 440-5.

47. Moens, P.D., M.C. Wahl, and D.M. Jameson, *Oligomeric state and mode of self-association of Thermotoga maritima ribosomal stalk protein L12 in solution.* Biochemistry, 2005. **44**(9): p. 3298-305.

48.   Valle, M., et al., *Locking and unlocking of ribosomal motions.* Cell, 2003. **114**(1): p. 123-34.

49.   Liljas, A. and A.T. Gudkov, *The structure and dynamics of ribosomal protein L12.* Biochimie, 1987. **69**(10): p. 1043-7.

50.   Ilag, L.L., et al., *Heptameric (L12)6/L10 rather than canonical pentameric complexes are found by tandem MS of intact ribosomes from thermophilic bacteria.* Proc Natl Acad Sci U S A, 2005. **102**(23): p. 8192-7.

51.   Maki, Y., et al., *Three binding sites for stalk protein dimers are generally present in ribosomes from archaeal organism.* J Biol Chem, 2007. **282**(45): p. 32827-33.

52.   Wahl, M.C. and W. Moller, *Structure and function of the acidic ribosomal stalk proteins.* Curr Protein Pept Sci, 2002. **3**(1): p. 93-106.

53.   Savelsbergh, A., et al., *Stimulation of the GTPase activity of translation elongation factor G by ribosomal protein L7/12.* J Biol Chem, 2000. **275**(2): p. 890-4.

54.   Kothe, U., et al., *Interaction of helix D of elongation factor Tu with helices 4 and 5 of protein L7/12 on the ribosome.* J Mol Biol, 2004. **336**(5): p. 1011-21.

55.   Wieden, H.J., W. Wintermeyer, and M.V. Rodnina, *A common structural motif in elongation factor Ts and ribosomal protein L7/12 may be involved in the interaction with elongation factor Tu.* J Mol Evol, 2001. **52**(2): p. 129-36.

56.   Tehei, M. and G. Zaccai, *Adaptation to high temperatures through macromolecular dynamics by neutron scattering.* FEBS J, 2007. **274**(16): p. 4034-43.

57.   Wimberly, B.T., et al., *Structure of the 30S ribosomal subunit.* Nature, 2000. **407**(6802): p. 327-39.

58.   Yusupov, M.M., et al., *Crystal structure of the ribosome at 5.5 A resolution.* Science, 2001. **292**(5518): p. 883-96.

59.   Ban, N., et al., *The complete atomic structure of the large ribosomal subunit at 2.4 A resolution.* Science, 2000. **289**(5481): p. 905-20.

60.   Doudna, J.A. and V.L. Rath, *Structure and function of the eukaryotic ribosome: the next frontier.* Cell, 2002. **109**(2): p. 153-6.

61.   Dube, P., et al., *Correlation of the expansion segments in mammalian rRNA with the fine structure of the 80 S ribosome; a cryoelectron microscopic reconstruction of the rabbit reticulocyte ribosome at 21 A resolution.* J Mol Biol, 1998. **279**(2): p. 403-21.

62.   Manuell, A.L., et al., *Composition and structure of the 80S ribosome from the green alga Chlamydomonas reinhardtii: 80S ribosomes are conserved in plants and animals.* J Mol Biol, 2005. **351**(2): p. 266-79.

63.   Morgan, D.G., et al., *A comparison of the yeast and rabbit 80 S ribosome reveals the topology of the nascent chain exit tunnel, inter-subunit bridges and mammalian rRNA expansion segments.* J Mol Biol, 2000. **301**(2): p. 301-21.

64.   Spahn, C.M., et al., *Structure of the 80S ribosome from Saccharomyces cerevisiae--tRNA-ribosome and subunit-subunit interactions.* Cell, 2001. **107**(3): p. 373-86.

65.   Spahn, C.M., et al., *Cryo-EM visualization of a viral internal ribosome entry site bound to human ribosomes: the IRES functions as an RNA-based translation factor.* Cell, 2004. **118**(4): p. 465-75.

66.   Nilsson, J., et al., *Regulation of eukaryotic translation by the RACK1 protein: a platform for signalling molecules on the ribosome.* EMBO Rep, 2004. **5**(12): p. 1137-41.

67.   Sweeney, R., L. Chen, and M.C. Yao, *An rRNA variable region has an evolutionarily conserved essential role despite sequence divergence.* Mol Cell Biol, 1994. **14**(6): p. 4203-15.

68.   Beckmann, R., et al., *Architecture of the protein-conducting channel associated with the translating 80S ribosome.* Cell, 2001. **107**(3): p. 361-72.

69.   Hinnebusch, A.G., *eIF3: a versatile scaffold for translation initiation complexes.* Trends in Biochemical Sciences, 2006. **31**(10): p. 553-562.

70.   Gingras, A.C., B. Raught, and N. Sonenberg, *Regulation of translation initiation by FRAP/mTOR.* Genes Dev, 2001. **15**(7): p. 807-26.

71. Harding, H.P., et al., *Regulated translation initiation controls stress-induced gene expression in mammalian cells.* Mol Cell, 2000. **6**(5): p. 1099-108.

72. Asano, K., et al., *A multifactor complex of eukaryotic initiation factors, eIF1, eIF2, eIF3, eIF5, and initiator tRNA(Met) is an important translation initiation intermediate in vivo.* Genes Dev, 2000. **14**(19): p. 2534-46.

73. Asano, K., et al., *Structure of cDNAs encoding human eukaryotic initiation factor 3 subunits. Possible roles in RNA binding and macromolecular assembly.* J Biol Chem, 1997. **272**(43): p. 27042-52.

74. Browning, K.S., et al., *Unified nomenclature for the subunits of eukaryotic initiation factor 3.* Trends Biochem Sci, 2001. **26**(5): p. 284.

75. Mayeur, G.L., et al., *Characterization of eIF3k: a newly discovered subunit of mammalian translation initiation factor elF3.* Eur J Biochem, 2003. **270**(20): p. 4133-9.

76. Morris-Desbois, C., et al., *The human protein HSPC021 interacts with Int-6 and is associated with eukaryotic translation initiation factor 3.* J Biol Chem, 2001. **276**(49): p. 45988-95.

77. Unbehaun, A., et al., *Release of initiation factors from 48S complexes during ribosomal subunit joining and the link between establishment of codon-anticodon base-pairing and hydrolysis of eIF2-bound GTP.* Genes Dev, 2004. **18**(24): p. 3078-93.

78. Phan, L., et al., *Identification of a translation initiation factor 3 (eIF3) core complex, conserved in yeast and mammals, that interacts with eIF5.* Mol Cell Biol, 1998. **18**(8): p. 4935-46.

79. Asano, K., et al., *Complex formation by all five homologues of mammalian translation initiation factor 3 subunits from yeast Saccharomyces cerevisiae.* J Biol Chem, 1998. **273**(29): p. 18573-85.

80. Greenberg, J.R., et al., *Nip1p associates with 40 S ribosomes and the Prt1p subunit of eukaryotic initiation factor 3 and is required for efficient translation initiation.* J Biol Chem, 1998. **273**(36): p. 23485-94.

81. Hanachi, P., J.W. Hershey, and H.P. Vornlocher, *Characterization of the p33 subunit of eukaryotic translation initiation factor-3 from Saccharomyces cerevisiae.* J Biol Chem, 1999. **274**(13): p. 8546-53.

82. Naranda, T., et al., *The 39-kilodalton subunit of eukaryotic translation initiation factor 3 is essential for the complex's integrity and for cell viability in Saccharomyces cerevisiae.* Mol Cell Biol, 1997. **17**(1): p. 145-53.

83. Valasek, L., et al., *Rpg1, the Saccharomyces cerevisiae homologue of the largest subunit of mammalian translation initiation factor 3, is required for translational activity.* J Biol Chem, 1998. **273**(33): p. 21253-60.

84. Verlhac, M.H., et al., *Identification of partners of TIF34, a component of the yeast eIF3 complex, required for cell proliferation and translation initiation.* EMBO J, 1997. **16**(22): p. 6812-22.

85. Vornlocher, H.P., et al., *A 110-kilodalton subunit of translation initiation factor eIF3 and an associated 135-kilodalton protein are encoded by the Saccharomyces cerevisiae TIF32 and TIF31 genes.* J Biol Chem, 1999. **274**(24): p. 16802-12.

86. Chaudhuri, J., D. Chowdhury, and U. Maitra, *Distinct functions of eukaryotic translation initiation factors eIF1A and eIF3 in the formation of the 40 S ribosomal preinitiation complex.* J Biol Chem, 1999. **274**(25): p. 17975-80.

87. Kolupaeva, V.G., et al., *Binding of eukaryotic initiation factor 3 to ribosomal 40S subunits and its role in ribosomal dissociation and anti-association.* RNA, 2005. **11**(4): p. 470-86.

88. Majumdar, R., A. Bandyopadhyay, and U. Maitra, *Mammalian translation initiation factor eIF1 functions with eIF1A and eIF3 in the formation of a stable 40 S preinitiation complex.* J Biol Chem, 2003. **278**(8): p. 6580-7.

89. Algire, M.A., et al., *Development and characterization of a reconstituted yeast translation initiation system.* RNA, 2002. **8**(3): p. 382-97.

90. Maag, D., et al., *A conformational change in the eukaryotic translation preinitiation complex and release of eIF1 signal recognition of the start codon.* Mol Cell, 2005. **17**(2): p. 265-75.

91. Imataka, H. and N. Sonenberg, *Human eukaryotic translation initiation factor 4G (eIF4G) possesses two separate and independent binding sites for eIF4A.* Mol Cell Biol, 1997. **17**(12): p. 6940-7.

92. Korneeva, N.L., et al., *Mutually cooperative binding of eukaryotic translation initiation factor (eIF) 3 and eIF4A to human eIF4G-1.* J Biol Chem, 2000. **275**(52): p. 41369-76.

93. Lamphear, B.J., et al., *Mapping of functional domains in eukaryotic protein synthesis initiation factor 4G (eIF4G) with picornaviral proteases. Implications for cap-dependent and cap-independent translational initiation.* J Biol Chem, 1995. **270**(37): p. 21975-83.

94. Morino, S., et al., *Eukaryotic translation initiation factor 4E (eIF4E) binding site and the middle one-third of eIF4GI constitute the core domain for cap-dependent translation, and the C-terminal one-third functions as a modulatory region.* Mol Cell Biol, 2000. **20**(2): p. 468-77.

95. Pestova, T.V. and V.G. Kolupaeva, *The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection.* Genes Dev, 2002. **16**(22): p. 2906-22.

96. Algire, M.A., D. Maag, and J.R. Lorsch, *Pi release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation.* Mol Cell, 2005. **20**(2): p. 251-62.

97. Fletcher, C.M., et al., *Structure and interactions of the translation initiation factor eIF1.* EMBO J, 1999. **18**(9): p. 2631-7.

98. Lomakin, I.B., et al., *Position of eukaryotic initiation factor eIF1 on the 40S ribosomal subunit determined by directed hydroxyl radical probing.* Genes Dev, 2003. **17**(22): p. 2786-97.

99. Majumdar, R. and U. Maitra, *Regulation of GTP hydrolysis prior to ribosomal AUG selection during eukaryotic translation initiation.* EMBO J, 2005. **24**(21): p. 3737-46.

100. Park, H.S., et al., *Eucaryotic initiation factor 4B controls eIF3-mediated ribosomal entry of viral reinitiation factor.* EMBO J, 2004. **23**(6): p. 1381-91.

101. Park, H.S., et al., *A plant viral "reinitiation" factor interacts with the host translational machinery.* Cell, 2001. **106**(6): p. 723-33.

102. Sonenberg, N. and A.G. Hinnebusch, *New Modes of Translational Control in Development, Behavior, and Disease.* Molecular Cell, 2007. **28**(5): p. 721-729.

103. Holz, M.K., et al., *mTOR and S6K1 mediate assembly of the translation preinitiation complex through dynamic protein interchange and ordered phosphorylation events.* Cell, 2005. **123**(4): p. 569-80.

104. Siridechadilok, B., et al., *Structural roles for human translation factor eIF3 in initiation of protein synthesis.* Science, 2005. **310**(5753): p. 1513-5.

105. Spahn, C.M., et al., *Hepatitis C virus IRES RNA-induced changes in the conformation of the 40s ribosomal subunit.* Science, 2001. **291**(5510): p. 1959-62.

106. Srivastava, S., A. Verschoor, and J. Frank, *Eukaryotic initiation factor 3 does not prevent association through physical blockage of the ribosomal subunit-subunit interface.* J Mol Biol, 1992. **226**(2): p. 301-4.

107. Culver, G.M., et al., *Identification of an RNA-protein bridge spanning the ribosomal subunit interface.* Science, 1999. **285**(5436): p. 2133-6.

108. Maivali, U. and J. Remme, *Definition of bases in 23S rRNA essential for ribosomal subunit association.* RNA, 2004. **10**(4): p. 600-4.

109. Junge, W., H. Lill, and S. Engelbrecht, *ATP synthase: an electrochemical transducer with rotatory mechanics.* Trends Biochem Sci, 1997. **22**(11): p. 420-3.

110. Boyer, P.D., *The ATP synthase--a splendid molecular machine.* Annu Rev Biochem, 1997. **66**: p. 717-49.

111. Fillingame, R.H., *Coupling H+ transport and ATP synthesis in F1F0-ATP synthases: glimpses of interacting parts in a dynamic molecular machine.* J Exp Biol, 1997. **200**(Pt 2): p. 217-24.

112. Senior, A.E., *ATP synthesis by oxidative phosphorylation.* Physiol Rev, 1988. **68**(1): p. 177-231.

113. Abrahams, J.P., et al., *Structure at 2.8 A resolution of F1-ATPase from bovine heart mitochondria.* Nature, 1994. **370**(6491): p. 621-8.

114. Bianchet, M.A., et al., *The 2.8-A structure of rat liver F1-ATPase: configuration of a critical intermediate in ATP synthesis/hydrolysis.* Proc Natl Acad Sci U S A, 1998. **95**(19): p. 11065-70.

115. Groth, G. and E. Pohl, *The structure of the chloroplast F1-ATPase at 3.2 A resolution.* J Biol Chem, 2001. **276**(2): p. 1345-52.

116. Stock, D., A.G. Leslie, and J.E. Walker, *Molecular architecture of the rotary motor in ATP synthase.* Science, 1999. **286**(5445): p. 1700-5.

117. Dunn, S.D., D.T. McLachlin, and M. Revington, *The second stalk of Escherichia coli ATP synthase.* Biochim Biophys Acta, 2000. **1458**(2-3): p. 356-63.

118. Junge, W., et al., *Inter-subunit rotation and elastic power transmission in F0F1-ATPase.* FEBS Lett, 2001. **504**(3): p. 152-60.

119. Wilkens, S., Z. Zhang, and Y. Zheng, *A structural model of the vacuolar ATPase from transmission electron microscopy.* Micron, 2005. **36**(2): p. 109-26.

120. Denda, K., et al., *Molecular cloning of the beta-subunit of a possible non-F0F1 type ATP synthase from the acidothermophilic archaebacterium, Sulfolobus acidocaldarius.* J Biol Chem, 1988. **263**(33): p. 17251-4.

121. Schafer, G. and M. Meyering-Vos, *F-type or V-type? The chimeric nature of the archaebacterial ATP synthase.* Biochim Biophys Acta, 1992. **1101**(2): p. 232-5.

122. Muller, V., C. Ruppert, and T. Lemker, *Structure and function of the A1A0-ATPases from methanogenic Archaea.* J Bioenerg Biomembr, 1999. **31**(1): p. 15-27.

123. Muller, V. and G. Gruber, *ATP synthases: structure, function and evolution of unique energy converters.* Cell Mol Life Sci, 2003. **60**(3): p. 474-94.

124. Coskun, U., et al., *Structure and subunit arrangement of the A-type ATP synthase complex from the archaeon Methanococcus jannaschii visualized by electron microscopy.* J Biol Chem, 2004. **279**(37): p. 38644-8.

125. Boekema, E.J., et al., *Visualization of a peripheral stalk in V-type ATPase: evidence for the stator structure essential to rotational catalysis.* Proc Natl Acad Sci U S A, 1997. **94**(26): p. 14291-3.

126. Speelmans, G., et al., *The F- or V-type Na(+)-ATPase of the thermophilic bacterium Clostridium fervidus.* J Bacteriol, 1994. **176**(16): p. 5160-2.

127. Takase, K., et al., *Sequencing and characterization of the ntp gene cluster for vacuolar-type Na(+)-translocating ATPase of Enterococcus hirae.* J Biol Chem, 1994. **269**(15): p. 11037-44.

128. Yokoyama, K., T. Oshima, and M. Yoshida, *Thermus thermophilus membrane-associated ATPase. Indication of a eubacterial V-type ATPase.* J Biol Chem, 1990. **265**(35): p. 21946-50.

129. Hilario, E. and J.P. Gogarten, *Horizontal transfer of ATPase genes--the tree of life becomes a net of life.* Biosystems, 1993. **31**(2-3): p. 111-9.

130. Hilario, E. and J.P. Gogarten, *The prokaryote-to-eukaryote transition reflected in the evolution of the V/F/A-ATPase catalytic and proteolipid subunits.* J Mol Evol, 1998. **46**(6): p. 703-15.

131. Finbow, M.E. and M.A. Harrison, *The vacuolar H+-ATPase: a universal proton pump of eukaryotes.* Biochem J, 1997. **324 (Pt 3)**: p. 697-712.

132. Futai, M., et al., *Luminal acidification of diverse organelles by V-ATPase in animal cells.* J Exp Biol, 2000. **203**(Pt 1): p. 107-16.

133. Nelson, N. and W.R. Harvey, *Vacuolar and plasma membrane proton-adenosinetriphosphatases.* Physiol Rev, 1999. **79**(2): p. 361-85.

134.   Nishi, T., S. Kawasaki-Nishi, and M. Forgac, *The first putative transmembrane segment of subunit c" (Vma16p) of the yeast V-ATPase is not necessary for function.* J Biol Chem, 2003. **278**(8): p. 5821-7.

135.   Stevens, T.H. and M. Forgac, *Structure, function and regulation of the vacuolar (H+)-ATPase.* Annu Rev Cell Dev Biol, 1997. **13**: p. 779-808.

136.   Hirata, T., et al., *Regulation and reversibility of vacuolar H(+)-ATPase.* J Biol Chem, 2000. **275**(1): p. 386-9.

137.   Toei, M., et al., *Dodecamer rotor ring defines H+/ATP ratio for ATP synthesis of prokaryotic V-ATPase from Thermus thermophilus.* Proc Natl Acad Sci U S A, 2007. **104**(51): p. 20256-61.

138.   Bowman, B.J., et al., *Isolation of genes encoding the Neurospora vacuolar ATPase. Analysis of vma-2 encoding the 57-kDa polypeptide and comparison to vma-1.* J Biol Chem, 1988. **263**(28): p. 14002-7.

139.   Zimniak, L., et al., *The cDNA sequence of the 69-kDa subunit of the carrot vacuolar H+-ATPase. Homology to the beta-chain of F0F1-ATPases.* J Biol Chem, 1988. **263**(19): p. 9102-12.

140.   Mandel, M., et al., *cDNA sequence encoding the 16-kDa proteolipid of chromaffin granules implies gene duplication in the evolution of H+-ATPases.* Proc Natl Acad Sci U S A, 1988. **85**(15): p. 5521-4.

141.   Ho, K., et al., *Cloning and expression of an inwardly rectifying ATP-regulated potassium channel.* Nature, 1993. **362**(6415): p. 31-8.

142.   Xie, X.S., et al., *Role of a 50-57-kDa polypeptide heterodimer in the function of the clathrin-coated vesicle proton pump.* J Biol Chem, 1994. **269**(41): p. 25809-15.

143.   Xu, T., E. Vasilyeva, and M. Forgac, *Subunit interactions in the clathrin-coated vesicle vacuolar (H(+))-ATPase complex.* J Biol Chem, 1999. **274**(41): p. 28909-15.

144.   Zhou, Z., et al., *Recombinant SFD isoforms activate vacuolar proton pumps.* J Biol Chem, 1999. **274**(22): p. 15913-9.

145.   Muench, S.P., et al., *Cryo-electron microscopy of the vacuolar ATPase motor reveals its mechanical and regulatory complexity.* J Mol Biol, 2009. **386**(4): p. 989-99.

146.   Gregorini, M., et al., *Three-dimensional reconstruction of bovine brain V-ATPase by cryo-electron microscopy and single particle analysis.* J Struct Biol, 2007. **158**(3): p. 445-54.

147.   Wilkens, S., T. Inoue, and M. Forgac, *Three-dimensional structure of the vacuolar ATPase. Localization of subunit H by difference imaging and chemical cross-linking.* J Biol Chem, 2004. **279**(40): p. 41942-9.

148.   Wilkens, S., E. Vasilyeva, and M. Forgac, *Structure of the vacuolar ATPase by electron microscopy.* J Biol Chem, 1999. **274**(45): p. 31804-10.

149.   Domgall, I., et al., *Three-dimensional map of a plant V-ATPase based on electron microscopy.* J Biol Chem, 2002. **277**(15): p. 13115-21.

150.   Sumner, J.P., et al., *Regulation of plasma membrane V-ATPase activity by dissociation of peripheral subunits.* J Biol Chem, 1995. **270**(10): p. 5649-53.

151.   Kane, P.M., *Disassembly and reassembly of the yeast vacuolar H(+)-ATPase in vivo.* J Biol Chem, 1995. **270**(28): p. 17025-32.

152.   Bernal, R.A. and D. Stock, *Three-dimensional structure of the intact Thermus thermophilus H+-ATPase/synthase by electron microscopy.* Structure, 2004. **12**(10): p. 1789-98.

153.   Diepholz, M., et al., *A different conformation for EGC stator subcomplex in solution and in the assembled yeast V-ATPase: possible implications for regulatory disassembly.* Structure, 2008. **16**(12): p. 1789-98.

154.   Sagermann, M., T.H. Stevens, and B.W. Matthews, *Crystal structure of the regulatory subunit H of the V-type ATPase of Saccharomyces cerevisiae.* Proc Natl Acad Sci U S A, 2001. **98**(13): p. 7134-9.

155.   Hong-Hermesdorf, A., et al., *A WNK kinase binds and phosphorylates V-ATPase subunit C.* FEBS Lett, 2006. **580**(3): p. 932-9.

156.   Voss, M., et al., *Stimulus-induced phosphorylation of vacuolar H(+)-ATPase by protein kinase A*. J Biol Chem, 2007. **282**(46): p. 33735-42.

157.   Inoue, T. and M. Forgac, *Cysteine-mediated cross-linking indicates that subunit C of the V-ATPase is in close proximity to subunits E and G of the V1 domain and subunit a of the V0 domain*. J Biol Chem, 2005. **280**(30): p. 27896-903.

158.   Jones, R.P., et al., *Defined sites of interaction between subunits E (Vma4p), C (Vma5p), and G (Vma10p) within the stator structure of the vacuolar H+-ATPase*. Biochemistry, 2005. **44**(10): p. 3933-41.

159.   Landolt-Marticorena, C., et al., *Evidence that the NH2 terminus of vph1p, an integral subunit of the V0 sector of the yeast V-ATPase, interacts directly with the Vma1p and Vma13p subunits of the V1 sector*. J Biol Chem, 2000. **275**(20): p. 15449-57.

160.   Drory, O., F. Frolow, and N. Nelson, *Crystal structure of yeast V-ATPase subunit C reveals its stator function*. EMBO Rep, 2004. **5**(12): p. 1148-52.

161.   Parra, K.J., K.L. Keenan, and P.M. Kane, *The H subunit (Vma13p) of the yeast V-ATPase inhibits the ATPase activity of cytosolic V1 complexes*. J Biol Chem, 2000. **275**(28): p. 21761-7.

162.   Vieille, C. and G.J. Zeikus, *Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability*. Microbiol Mol Biol Rev, 2001. **65**(1): p. 1-43.

163.   Rodnina, M.V. and W. Wintermeyer, *GTP consumption of elongation factor Tu during translation of heteropolymeric mRNAs*. Proc Natl Acad Sci U S A, 1995. **92**(6): p. 1945-9.

164.   Yokoyama, K., et al., *Rotation of the proteolipid ring in the V-ATPase*. J Biol Chem, 2003. **278**(27): p. 24255-8.

165.   Leberman, R., *Use of uranyl formate as a negative stain*. J Mol Biol, 1965. **13**(2): p. 606.

166.   Lepault, J. and J. Dubochet, *Electron microscopy of frozen hydrated specimens: preparation and characteristics*. Methods Enzymol, 1986. **127**: p. 719-30.

167.   Chen, J.Z. and N. Grigorieff, *SIGNATURE: a single-particle selection system for molecular electron microscopy*. J Struct Biol, 2007. **157**(1): p. 168-73.

168.   Ludtke, S.J., P.R. Baldwin, and W. Chiu, *EMAN: semiautomated software for high-resolution single-particle reconstructions*. J Struct Biol, 1999. **128**(1): p. 82-97.

169.   Schmeisser, M., et al., *Parallel, distributed and GPU computing technologies in single-particle electron microscopy*. Acta Crystallogr D Biol Crystallogr, 2009. **65**(Pt 7): p. 659-71.

170.   Sorzano, C.O., et al., *XMIPP: a new generation of an open-source image processing package for electron microscopy*. J Struct Biol, 2004. **148**(2): p. 194-204.

171.   Dubuisson, M.P. and A.K. Jain. *A modified Hausdorff distance for object matching*. in *Pattern Recognition, 1994. Vol. 1 - Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*. 1994.

172.   Gope, C. and N. Kehtarnavaz, *Affine invariant comparison of point-sets using convex hulls and hausdorff distances*. Pattern Recogn., 2007. **40**(1): p. 309-320.

173.   Klaholz, B.P., A.G. Myasnikov, and M. Van Heel, *Visualization of release factor 3 on the ribosome during termination of protein synthesis*. Nature, 2004. **427**(6977): p. 862-5.

174.   Wahl, M.C., et al., *Flexibility, conformational diversity and two dimerization modes in complexes of ribosomal protein L12*. EMBO J, 2000. **19**(2): p. 174-86.

175.   Cannone, J.J., et al., *The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs*. BMC Bioinformatics, 2002. **3**: p. 2.

176.   Agrawal, R.K., R.K. Lata, and J. Frank, *Conformational variability in Escherichia coli 70S ribosome as revealed by 3D cryo-electron microscopy*. Int J Biochem Cell Biol, 1999. **31**(1): p. 243-54.

177.   Trylska, J., V. Tozzini, and J.A. McCammon, *Exploring global motions and correlations in the ribosome*. Biophys J, 2005. **89**(3): p. 1455-63.

178.   Fei, J., et al., *Coupling of ribosomal L1 stalk and tRNA dynamics during translation elongation*. Mol Cell, 2008. **30**(3): p. 348-59.

179. Tama, F., et al., *Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy.* Proc Natl Acad Sci U S A, 2003. **100**(16): p. 9319-23.

180. Kurkcuoglu, O., et al., *The ribosome structure controls and directs mRNA entry, translocation and exit dynamics.* Phys Biol, 2008. **5**(4): p. 46005.

181. Van Heel, M., *Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction.* Ultramicroscopy, 1987. **21**(2): p. 111-23.

182. Schafer, I.B., et al., *Crystal structure of the archaeal A1Ao ATP synthase subunit B from Methanosarcina mazei Go1: Implications of nucleotide-binding differences in the major A1Ao subunits A and B.* J Mol Biol, 2006. **358**(3): p. 725-40.

183. Gibbons, C., et al., *The structure of the central stalk in bovine F(1)-ATPase at 2.4 A resolution.* Nat Struct Biol, 2000. **7**(11): p. 1055-61.

184. Iwata, M., et al., *Crystal structure of a central stalk subunit C and reversible association/dissociation of vacuole-type ATPase.* Proc Natl Acad Sci U S A, 2004. **101**(1): p. 59-64.

185. Rastogi, V.K. and M.E. Girvin, *Structural changes linked to proton translocation by subunit c of the ATP synthase.* Nature, 1999. **402**(6759): p. 263-8.

186. Leschziner, A.E. and E. Nogales, *The orthogonal tilt reconstruction method: an approach to generating single-class volumes with no missing cone for ab initio reconstruction of asymmetric particles.* J Struct Biol, 2006. **153**(3): p. 284-99.

187. Schmeisser, M., et al., *Parallel, distributed and GPU computing technologies in single particle electron microscopy.* to appear in Acta Crystallogr D., 2009.

188. Orlova, E.V., et al., *Structure of the 13-fold symmetric portal protein of bacteriophage SPP1.* Nat Struct Biol, 1999. **6**(9): p. 842-6.

189. Shcherbakov, D., et al., *Stability of the 'L12 stalk' in ribosomes from mesophilic and (hyper)thermophilic Archaea and Bacteria.* Nucleic Acids Res, 2006. **34**(20): p. 5800-14.

190. Nilsson, J., et al., *Comparison of fungal 80 S ribosomes by cryo-EM reveals diversity in structure and conformation of rRNA expansion segments.* J Mol Biol, 2007. **369**(2): p. 429-38.

191. Rubinstein, J.L., J.E. Walker, and R. Henderson, *Structure of the mitochondrial ATP synthase by electron cryomicroscopy.* EMBO J, 2003. **22**(23): p. 6182-92.

192. Kish-Trier, E., et al., *The stator complex of the A1A0-ATP synthase--structural characterization of the E and H subunits.* J Mol Biol, 2008. **375**(3): p. 673-85.

193. Esteban, O., et al., *Stoichiometry and localization of the stator subunits E and G in Thermus thermophilus H+-ATPase/synthase.* J Biol Chem, 2008. **283**(5): p. 2595-603.

Number of test set	Number of particle co-ordinates contained in set	Percentage of correctly correlated tilt pairs/per cent	Percentage of retrieved particles/ per cent
1	314	100.00	43.31210191
2	552	99.80	92.57246377
3	801	100.00	5.243445693
4	746	100.00	6.836461126
5	451	100.00	6.430155211
6	714	100.00	5.882352941
7	402	100.00	85.5721393
8	451	99.36	34.58980044
9	193	100.00	2.07253886
10	317	100.00	85.17350158
11	953	100.00	0.419727177
12	1067	100.00	1.030927835
13	196	100.00	89.28571429
14	159	100.00	89.93710692
15	145	100.00	86.20689655
16	160	100.00	86.875
17	253	100.00	2.766798419
18	190	94.95	52.10526316
19	10	99.42	86.31840796
20	282	99.12	40.42553191
21	83	100.00	10.84337349
22	222	99.38	72.97297297
23	247	100.00	1.619433198
24	266	99.57	87.59398496
25	319	100.00	86.8338558
26	270	100.00	87.40740741
27	228	100.00	73.68421053
28	485	100.00	0.824742268
29	374	99.71	91.17647059
30	341	100.00	87.97653959
31	404	99.71	85.64356436
32	355	100.00	90.14084507
33	283	100.00	88.69257951
34	279	100.00	96.05734767
35	332	100.00	91.26506024
36	366	100.00	91.2568306
37	377	100.00	89.38992042
38	1144	100.00	95.1048951
39	1067	97.87	83.50515464
40	1037	100.00	94.31051109
41	1124	100.00	95.99644128
42	1282	100.00	95.16380655
43	637	100.00	87.28414443
Average		99.74	62.50698667
sigma		0.84	37.48

Appendix 1 Detailed statistics of the benchmark test of the MaverickTilt *manual* mode

Number of test set	Number of particle co-ordinates contained in set	Percentage of retrieved particles/ per cent	Number of retrieved particles
1	314	2,866242038	9
2	552	0,724637681	4
3	801	0,49937578	4
4	746	0,804289544	6
5	451	1,55210643	7
6	714	0,56022409	4
7	402	1,243781095	5
8	451	1,10864745	5
9	193	2,07253886	4
10	317	1,261829653	4
11	953	3,043022036	29
12	1067	0,468603561	5
13	196	2,551020408	5
14	159	2,51572327	4
15	145	2,75862069	4
16	160	2,5	4
17	253	1,581027668	4
18	190	2,631578947	5
19	10	1,243781095	0,12
20	282	1,418439716	4
21	83	4,819277108	4
22	222	1,801801802	4
23	247	1,619433198	4
24	266	2,631578947	7
25	319	1,253918495	4
26	270	1,481481481	4
27	228	2,192982456	5
28	485	0,824742268	4
29	374	1,604278075	6
30	341	1,46627566	5
31	404	1,237623762	5
32	355	1,690140845	6
33	283	1,766784452	5
34	279	1,792114695	5
35	332	1,506024096	5
36	366	1,366120219	5
37	377	3,978779841	15
38	1144	0,611888112	7
39	1067	0,468603561	5
40	1037	1,446480231	15
41	1124	0,533807829	6
42	1282	0,390015601	5
43	637	1,098901099	7
Average		1,650896369	5,9098693
sigma		0,951263381	4,4082542

Appendix 2 Detailed statistics of the benchmark test of the MaverickTilt *auto* on false positive datasets

7   Curriculum Vitae

Florian Hauer, MSc (Molecular Biology)

Date of birth: May 6, 1980 in Karlsruhe, Germany

MPI for Biophysical Chemistry

Am Faßberg 11, 37077 Göttingen, Germany

Phone: +49(0)551 201 1306

eMail: fhauer@gwdg.de

University studies

10/2005-05/2009	**Doctoral studies**, IMPRS Molecular Biology, Göttingen, Germany
	PhD thesis "Three-dimensional electron microscopy of structurally heterogeneous biological macromolecules", Department of Three-dimensional Electron Cryomicroscopy, MPI for Biophysical Chemistry, Göttingen, Germany
10/2005-03/2007	**Integrated MSc Thesis**, IMPRS Molecular Biology, Göttingen, Germany. Title:"The three-dimensional structure of the SecYEG complex of Escherichia coli attached to the 50S Ribosomal Subunit as revealed by Cryo Electron Microscopy", Department of Three-dimensional Electron Cryomicroscopy, MPI for Biophysical Chemistry, Göttingen, Germany
10/2004-9/2005	**Graduate Studies**, Molecular Biology, IMPRS Molecular Biology, Göttingen, Germany
10/2003-05/2005	**DAAD research exchange project**, IIT Kharagpur, India
	Project title:"Comparative metabolite profiling of Tasar silkworm *Anthereae mylitta* and medical host plant *Terminalia arjuna*"
10/2000-9/2003	**Undergraduate Studies**, Molecular Bioltechnology, TU Dresden, Germany

Professional activities

05/2000-03/2003	**Free lance consultant** for new media and marketing, Dresden, Germany
06/1999-09/2000	**Male nurse**, Klinikum Karlsbad-Langensteinbach, Karlsbad, Germany

Extracurricular activities:

10/2005-05/2009	**Member of the PhD Committee**, MPI for Biophysical Chemistry, Göttingen, Germany
09/2007-09/2008	**Secretary** of the Max-Planck wide association of PhD students